Journal of Imaging Science and Technology® 66(4): 040403-1–040403-20, 2022.

A Dual-channel Artificial Neural Network Decision Fusion Framework Incorporated with Deep Learning of Inertial Measurement Unit Sensor-based Spectrum Images for Hand Gesture Intention Cognition

Ing-Jr Ding, Ya-Cheng Juang, and Bing-Tsan Lin
Department of Electrical Engineering, National Formosa University, Huwei Township, Yunlin County 632, Taiwan
E-mail: ingjr@nfu.edu.tw

Abstract. *The inertial measurement unit (IMU) is a popular sensor device, which is mainly employed to acquire body or hand gesture action information for performing specific recognition tasks. We present a dual-channel artificial neural network (ANN) recognition decision hybridization scheme incorporated with deep leaning of IMU-based spectrogram images for cognition of several common hand gesture intention categorization actions focused on the 6-axis IMU sensing data (containing 3-axis accelerometer and 3-axis gyroscope information) and the 6-axis IMU derived spectrogram images. In this hand gesture intention cognition approach, both symmetric and asymmetric ANN structures are considered for intention action classifications. The proposed dual-channel ANN decision fusion framework contains one ANN recognition channel with inputs of "6-axis IMU raw data" and the other ANN recognition channel with inputs of "IMU spectrogram image derived-critical deep learning features". Recognition decisions estimated from either of these two ANN recognition channels form the fusion framework. Three fusion schemes on dual-channel ANN recognition decisions are presented in this study, channel output layer accumulation, same channel candidate output and same-or-dual channel candidate output. In this study, the well-known deep learning neural network, visual geometry group- convolution neural network (VGG-CNN), is employed to carry out deep learning computations on IMU-based spectrogram images, from which, the critical deep learning feature of each spectrogram image can then be extracted and used as an input for the dual-channel ANN. For recognition performance comparisons, hand gesture intention recognition by the traditional VGG-CNN deep neural network approach (i.e. recognition of IMU spectrogram images using typical deep learning of the CNN model) is also performed. Experiments on classifications of six hand gesture intention actions show that the presented dual-channel ANN decision fusion incorporated with deep learning of IMU spectrum images has competitive performances, reaching better recognition accuracy than traditional CNN deep learning.* © *2022 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2022.66.4.040403]

1. INTRODUCTION

Body language through specific body action variations is an effective nonverbal cue and plays an important role in human communication. Along with speech communication and facial expressions, body action information can further be combined to accurately infer the specific intention behavior of the person [1]. In body language-based human intention expression, hand gesture actions is likely to be the most representative and effective [2]. In daily life, various common hand gesture actions have been frequently used to reveal corresponding specific human intention behaviors of "anger", "confidence", "anxiety", "joy", "tension" and "inspiration". Similar to matured speech recognition [3–7], face recognition [8, 9] and fingerprint recognition techniques [10, 11] widely used in real world applications, hand gesture action-based intention recognition, belonging to biometric characteristics recognition similarly, will undoubtedly became an indispensable and creative human machine interface (HMI) application.

With rapid development in the field of contact and contactless sensor devices, increased studies on hand gesture recognition with the specific type of sensor data have been proposed in the recent years [12–37]. According to variances of the employed sensor and the acquired data, hand gesture recognition can be primarily categorized into RGB image-based [12–14], 3-dimensional (3-D) space data-based [15–20], depth image-based [21–23], surface electromyography (sEMG)-based [24–31] and inertial measurement unit (IMU)-based recognition systems [27–37] where the first three types are contactless categorizations and the last two approaches belong to the contact class. For categorization of hand gesture recognition with IMU sensing data, the sensor device with IMU raw data sensing is generally equipped with (or embedded in) smart phones, smart watches or the smart wearable sport\health bracelets. Compared with contactless hand gesture recognition by the CMOS image sensor, the 3D sensor (also known as the RGB-D sensor) or the infrared depth sensor, contact hand gesture recognition using the wearable IMU motion sensor is more suitable for use in applications of human activity measurements, action analysis and recognition in the specific sport device and specified common hand gesture intention behavior recognition in this study due to the property of arbitrary sensing without inconvenient restrictions of the limited sensing region.

Studies on hand gesture recognition with the IMU data [27–37] can be further divided into IMU pattern recognition by typical pattern classifiers such as the artificial

Received Nov. 10, 2021; accepted for publication Feb. 26, 2022; published online Apr. 1, 2022. Associate Editor: Jia-Shing Sheu.
1062-3701/2022/66(4)/040403/20/$25.00

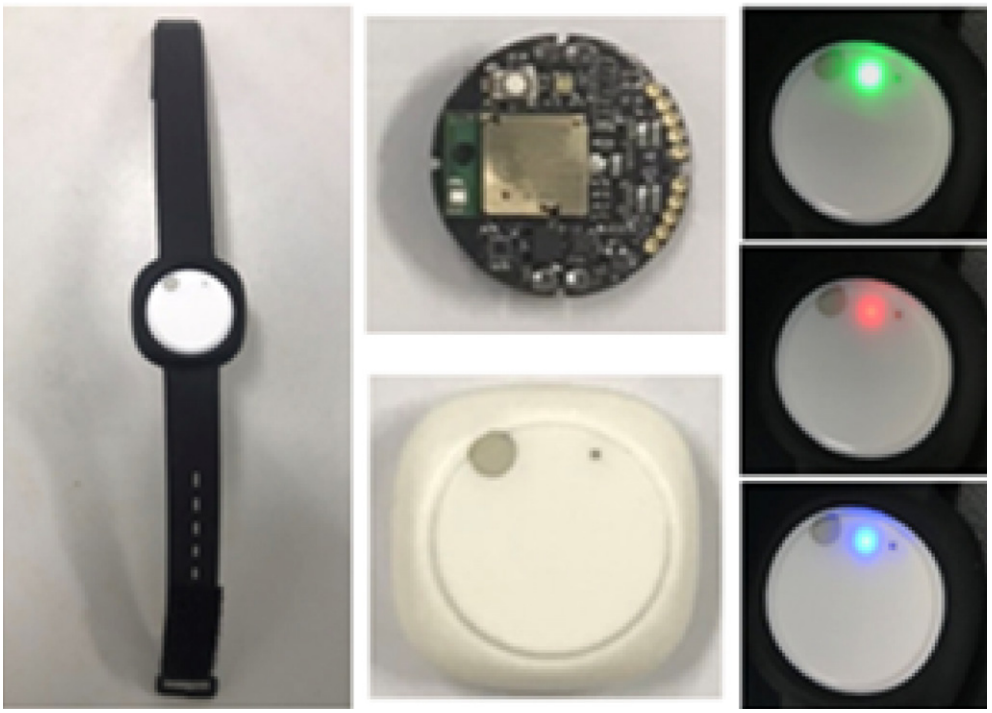


Figure 1. The MbientLab wearable bracelet device utilized in this work to acquire 3-axis accelerometer and 3-axis gyroscope motion data for constructing an efficient ANN recognition scheme with 6-axis IMU raw data for hand intention action recognition [38].

neural network (ANN) [34], IMU pattern recognition by deep neural network (DNN) with deep learning [35–37], and integrated or combined recognition of the IMU data and other data with different sensing modalities [27–33]. In the work of [34], only the accelerometer sensor data is employed directly to classify hand gestures by ANN. That the IMU signal is considered to be further combined with the sEMG signal for hand gesture recognition by ANN or other classifiers can be seen in these studies of [27–31]. In addition, IMU signal is also explored to be integrated with other modalities of sensor data, such as acoustic voice data [32] or mechanomyography data [33], to construct a multi-modality hand gesture recognition system. On the other hand, in addition to typical ANN, well-known DNN techniques with the framework of deep learning neural networks, such as the typical convolutional neural network (CNN), the typical recurrent neural network (RNN) and the long short term memory (LSTM)-type RNN, have also been explored in the task of hand gesture recognition with IMU data [35–37].

For the above mentioned IMU-based hand gesture recognition studies, very limited studies have focused on using deep learning model-derived features of the IMU raw signal on hand gesture recognition. Researchers have explored to derive visual geometry group (VGG)-CNN deep learning features from the Leap Motion Controller (LMC) 3-D image data for hand gesture intention-based identity recognition in the previous study [20]. The work of [20] only considers the the derived VGG-CNN deep learning feature directly for template match-based classification without any consideration of data fusion or model fusion designs on recognition. In this paper, a dual-channel ANN recognition hybridization scheme to make fine fusion of the wearable 6-axis IMU raw and its deep learning features by decision fusion for accurately cognizing various common hand gesture intention actions is proposed, which will be detailed in the following sections.

2. HAND GESTURE INTENTION RECOGNITION BY ANN CLASSIFICATION WITH 6-AXIS IMU RAW DATA

This work adopts the wearable bracelet device developed by MbientLab [38] to construct a hand intention action recognition system (see Figure 1). As shown in Fig. 1, the main sensor platform in the wearable bracelet device, called MetaMotionC (MMC), contains an accelerometer, a gyroscope and a temperature sensor. It is to be noted that in the current version of MMC [38], there are more abundant motion and environment sensors equipped on the board to provide real-time data acquisition of human body and environment data. In this study, for acquiring the continuous-time motion data of a specified hand intention action, mainly the 3-dimensional accelerometer data and the 3-dimensional gyroscope data, only the accelerometer and the gyroscope sensor of MMC are used. It's also noted that although the MMC includes some software development kits (SDKs) and application program interfaces (APIs) for rapid implementing an application, this work only employs the SDK of data acquisition of 6-axis inertial measurement unit (IMU) raw data (i.e. 3-axis accelerometer and 3-axis gyroscope raw data) to obtain hand motion data to further

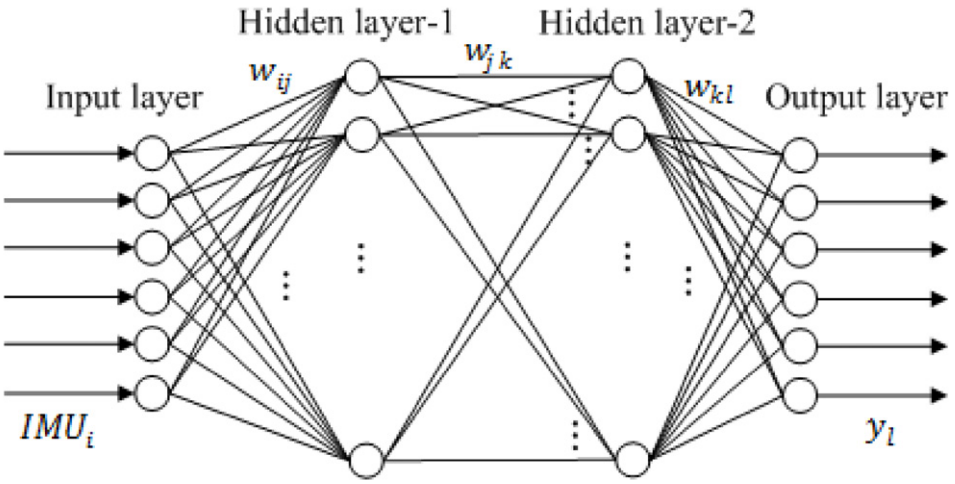


Figure 2. Illustrations of the ''symmetric'' artificial neural network structure employed in this work (the same neural node number set in each of two hidden layers).

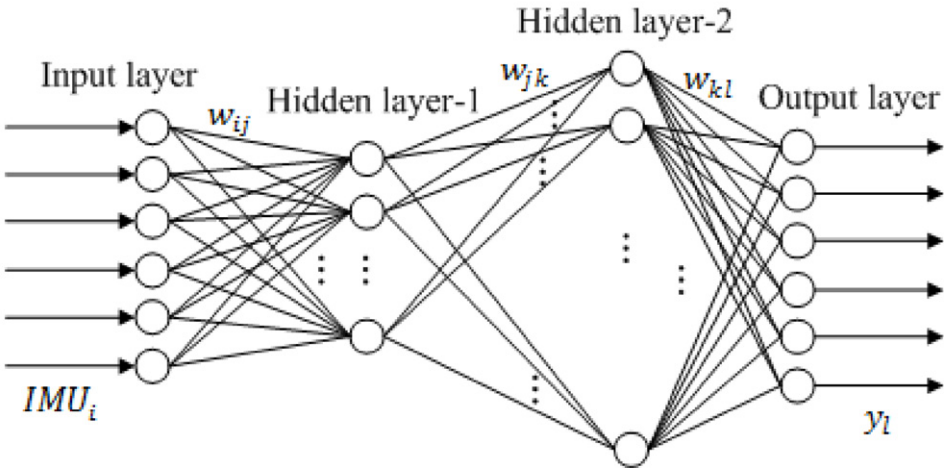


Figure 3. Illustrations of the ''asymmetric'' artificial neural network structure employed in this work (different neural node numbers set in each of two hidden layers).

analyze and categorize the specific hand intention action behavior.

In this study, the ANN classification scheme is utilized to be the classifier for categorizing different types of hand intention actions on the obtained or modified hand action data. Totally, 6 different classes of hand gesture actions to indicate specific human intention behaviors are defined in the developed recognition system. In this system, hand intention action recognition using ANN classifications with 6-axis IMU raw data, two different types of ANN structures are taken into account, which are "symmetric" and "asymmetric" ANN structures. Figures 2 and 3 illustrate the symmetric ANN and the asymmetric ANN network topologies respectively. During the phases of ANN model establishment (model training) and ANN model simulation (model test), basic principles of ANN calculations are adopted in this work, as shown in the following equations, Eqs. (1)–(4).

$$h_j = f_1\left(\sum_{i=1}^{n} w_{ij}\cdot IMU_i - \delta_j\right) \tag{1}$$

$$h_k = f_2\left(\sum_{j=1}^{m} w_{jk}\cdot h_j - \delta_k\right) \tag{2}$$

$$y_l = f_3\left(\sum_{k=1}^{o} w_{kl}\cdot h_k - \delta_l\right) \tag{3}$$

$$Error = \sum_{l=1}^{p}\left(Expectation_l - y_l\right). \tag{4}$$

Note that when performing ANN model establishment in the training phase, the well-known back propagation (BP) algorithm is utilized. All the input hand gesture

action data are made forward-propagating by Eqs. (1)–(3). Equation (4) shows that the error signal, *Error* is obtained by accumulation of output difference values of each of all p nodes of the output layer, $Expectation_l - y_l$, $l = 1, 2, \ldots, p$ ($p = 6$ in this work, denoting 6 different hand gesture intention categorizations). It's noted that $Expectation_l$ and y_l in Equation (4) denote the network expectation output and the network output of the lth node in the output layer respectively. In the BP approach, the item *Error* derived by Eq. (4) can be multiplied by the factor of the learning rate, *Rate*. The formed item, $Error \cdot Rate$, can then be used to adjust the connection weight parameter of the ANN model in the way of output-to-input back propagation.

Focusing on these two different ANN network topologies, this work evaluates the recognition accuracy in different settings of various nodes in hidden layer-1 and hidden layer-2. For evaluations of symmetric ANN, sets of (hidden layer-1, hidden layer-2) are (5, 5), (6, 6), (7, 7)..., and (20, 20). As for the asymmetric ANN topology, totally 20 sets of (hidden layer-1, hidden layer-2) are evaluated, which are (5, 25), (6, 24), (7, 23),..., (25, 5). Note that symmetric and asymmetric ANN structures with the best recognition accuracy ((13, 13) in symmetric ANN and (10, 20) in asymmetric ANN) will then be further considered in the design of dual-channel ANN recognition hybridizations additionally incorporated with the deep learning feature ANN recognition channel (i.e. the complete ANN structure of *6-13-13-6* and *6-10-20-6* after filling the input layer with 6 nodes of the IMU data and the output layer with 6 hand gesture intention class nodes), which will be detailed in the following sections.

3. DUAL-CHANNEL ANN RECOGNITION HYBRIDIZATIONS OF 6-AXIS IMU RAW DATA AND ITS IMU SPECTROGRAM IMAGE-DERIVED DEEP LEARNING FEATURES FOR HAND GESTURE INTENTION COGNITION

As mentioned in the previous section, two different model types of ANN calculations with 6-axis IMU raw data are initially established for classification of various hand gesture intention actions. Taking into consideration the advanced ANN model with feature information learning that is categorized into deep learning models on hand gesture intention recognition, the popular VGG-16 CNN deep learning model is employed in this study. For thoroughly considering VGG-16 CNN computations, a fundamental method to use the VGG-16 CNN deep learning model with 6 axis IMU-derived spectrogram images for classifying hand gesture intention actions is initially presented, following which, the dual-channel ANN recognition approach to properly hybridize both of the ANN classification decision of 6-axis IMU raw data and that of the derived VGG-16 CNN deep learning features by three different decision fusion designs is presented for hand gesture intention recognition.

3.1 *Hand Gesture Intention Recognition by the VGG-16 CNN Deep Learning Model with 6 Axis IMU Raw-Derived Spectrogram Images (Typical CNN Recognition on Input Images)*

As mentioned before, the 6-axis IMU raw data derived from 3-axis accelerometer and 3-axis gyroscope sensors of the MMC platform in certain continuous-time period can be used to represent motion variances of the specific hand gesture intention action. Such 6-axis IMU raw collected during continuous-time period is viewed as the time-domain motion data stream, which can further be transformed into the frequency-domain information for feature learning and classifications of the specific deep learning model (typical VGG-16 CNN employed in this study). In this work, the time domain-based IMU raw data stream is further transformed into the frequency-domain characteristics, called the IMU raw-derived spectrogram image. Such spectrogram information is estimated by using the fast Fourier transformation (FFT) calculation on the 6-axis IMU raw data stream. The 6 axis IMU raw-derived spectrogram image is properly resized to a fixed-size of 224 by 224 in order to match the input specification of the VGG-16 CNN model. The VGG-16 CNN deep learning method essentially belongs to the VGGNet convolution neural network, which is also popular and known as the VGGNet-series CNN (generally, VGG-CNN for simplicity) model [39]. The main property of VGG-CNN is that on pattern recognition, a satisfactory classification performance can be achieved by adjusting the depth (level) of the feature learning model. There are six different types of structures involved in VGG-CNN categorization models, each of which corresponds to one specific model configurations of "type A", "type A-LRN", "type B", "type C", "type D" and "type E". Each of these six VGG-CNN model configurations has the specific number of convolution and pooling calculations. The VGG-16 CNN model adopted in this work is categorized into the class of "type D". The VGG-16 CNN contains 13 convolution layers (compounded with pooling computations of 5 layers) and 3 fully connected (FC) layers. As depicted in Figure 4, VGG-16 CNN contains two main phases, 13 process layers for extraction of deep learning features and 3 process layers for categorizations of the derived deep learning feature. When an IMU-based spectrum with the size of 224 by 224 completes the deep learning task of accurate extractions of image characteristics by a series of computations of the first 13 layers of VGG-16 CNN, the input image of the IMU-based spectrum can then be represented by a data vector with the dimension of 4096 (i.e. the deep learning feature of 4096D). The 14th layer, i.e. the starting layer of these 3 fully connected layers, with 4096 nodes will be utilized to transmit the feature data of 4096D for further classifications of gesture intention actions.

Different from recognition calculations of conventional VGG-16 CNN in Fig. 4, the extracted deep learning feature after computations of 13 process layers has 4096 dimensions, and such 4096D deep learning features will be independently separated for further gesture intention classifications using

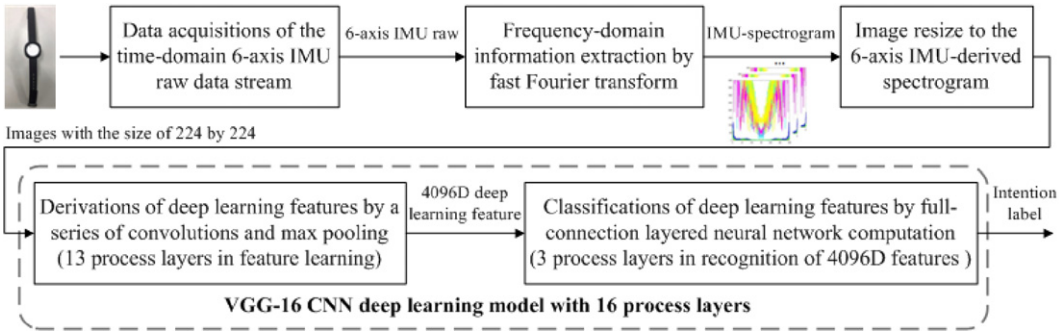


Figure 4. Fundamental deep learning by the VGG-16 CNN model with inputs of 6 axis IMU-derived spectrogram images for hand intention action recognition.

the symmetric ANN (or the asymmetric ANN) mentioned in Section 2. It's noted that main differences between the deep learning model VGG-16 CNN in Section 3.1 and the ANN model in Section 2 are (1) an additional ability of thorough learning of input image characteristic provided by VGG-16 CNN (a series of convolution and pooling calculations of the previous 13 layers) and (2) the fixed pattern classifier with an invariable design of three fully connected layers (i.e. FC-1 of the 14th layer, FC-2 of the 15th layer and classifications of the 16-layer) in VGG-16 CNN and the flexible pattern classifier in ANN that can be designed in a manner of self-definitions (e.g., the symmetric structure of two hidden layers containing the same node number in each layer and the asymmetric structure of two hidden layers containing the different node number in each layer designed in this work, as shown in Figs. 2 and 3). A scheme to be able to simultaneously retain the advantage of the typical VGG-16 CNN model on deep learning of input data (without considerations of the invariable final three layers) and the merit of the classical ANN model on a design of flexible and self-defined classifications will be expected to be performed in hand gesture intention recognition with outstanding recognition accuracy. Based on this line of thought, a dual-channel ANN recognition framework, which can intelligently hybridize two recognition decisions estimated from two self-defined ANN recognition channels of the 6-axis IMU raw data and the VGG-16 CNN extracted deep learning feature data is presented in this work, and is detailed in later sections.

3.2 *Hand Gesture Intention Recognition using ANN Classifications with Principal Component-Extracted Critical Deep Learning Features*

As mentioned, the 4096D deep learning feature can be extracted after a series of convolution and max pooling computation of the first 13 layers of VGG-16 CNN. For real-time recognition on hand gesture intention categorizations, such VGG-16 CNN extracted 4096D deep learning features will be made by data reduction. Well-known principal component analysis (PCA) will be utilized herein to derive the significant data characteristics of the 4096D deep learning features, i.e. principal components of 4096D features. The main calculation of the PCA approach is to first establish the eigenspace of the new dimension-reduced feature vector that has the smallest value of the reconstruction error, following which, an eigen-decomposition estimate is then done to obtain the eigen-value in each axis of the constructed new space for completing data transform in the new space. By calculations of Eq. (5) to achieve the minimum reconstruction error, the 4096D deep learning feature vector from VGG-16 CNN will be transformed into a new feature vector with only 80 principal components (the 80-D deep learning feature used hereafter, for simplicity).

$$\min_{M,PC_{DL}} MSE\{DL_{4096D} - M \cdot (PC_{DL})'\}, \tag{5}$$

where *MSE* denotes mean square error computations, DL_{4096D} is the 4096D deep learning feature; PC_{DL} denotes the new feature vector with only principal components in the constructed eigenspace; M represents the transformation matrix to perform data transform between two different data space.

As the work of hand gesture intention recognition by ANN classification with 6-axis IMU raw data (see Section 2), the 80-D deep learning feature will then be classified and evaluated for recognition accuracy by the symmetric and the asymmetric ANN (see Figure 5). Similar to ANN evaluations on above-mentioned ANN classification with 6-axis IMU raw data, ANN classification with 80-D deep learning features is also evaluated for recognition accuracy by symmetric ANN structures of (5, 5), (6, 6), (7, 7)..., and (20, 20), and asymmetric ANN structures of (5, 25), (6, 24), (7, 23),..., (25, 5). Note that in the asymmetric ANN structure, (10, 20) denoting 10 and 20 nodes in the first and the second hidden layer of the ANN respectively has the best recognition performance, which is the same with ANN classification with 6-axis IMU raw data; for evaluations of the symmetric ANN, the (9, 9) structure that represents 9 nodes contained in each of two hidden layers of the ANN achieves the best performance on hand gesture intention recognition. With the input layer of 80-D deep learning features and the output layer of 6 hand gesture intention categorizations, ANN structures of *80-10-20-6* and *80-9-9-6* are used in the recognition channel of 80-D deep learning feature ANN, which will then be further made to ANN decision fusion with those of the recognition channel of 6-axis IMU raw ANN.

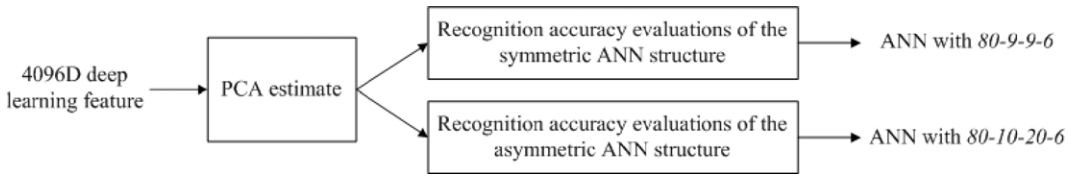


Figure 5. ANN classification calculations using the input of VGG-16 CNN derived deep learning feature information with data reduction of principal component analysis.

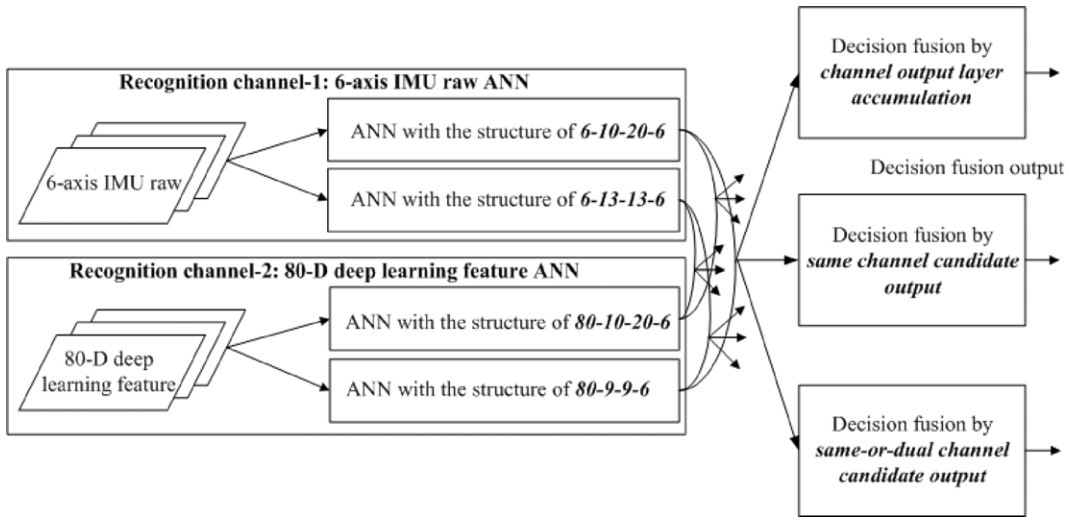


Figure 6. The presented dual-channel ANN recognition framework for intelligently hybridizing recognition decisions estimated from two different ANN channels, ANN with 6-axis IMU raw and ANN with 80-D deep learning features, on hand gesture intention recognition.

3.3 *Dual-Channel ANN Recognition Hybridizations of Original 6-Axis IMU Raw and its 80-D Critical Deep Learning Features by Decision Fusion for Hand Gesture Intention Recognition*

The presented dual-channel ANN recognition framework is illustrated in Figure 6. As shown in Fig. 6, one recognition channel is the 6-axis IMU raw ANN with the asymmetric structure of the input layer of 6 nodes, the hidden layer-1 of 10 nodes, the hidden layer-2 of 20 nodes and the output layer of 6 nodes (i.e., the 6-10-20-6 structure, for simplicity) or the symmetric structure of the input layer of 6 nodes, the hidden layer-1 of 13 nodes, the hidden layer-2 of 13 nodes and the output layer of 6 nodes (i.e., the 6-13-13-6 structure, for simplicity); another recognition channel is the 80-D deep learning feature ANN with the asymmetric structure of the input layer of 80 nodes, the hidden layer-1 of 10 nodes, the hidden layer-2 of 20 nodes and the output layer of 6 nodes (i.e., the 80-10-20-6 structure, for simplicity) or the symmetric structure of the input layer of 80 nodes, the hidden layer-1 of 9 nodes, the hidden layer-2 of 9 nodes and the output layer of 6 nodes (i.e., the 80-9-9-6 structure, for simplicity). Three decision fusion approaches, "channel output layer accumulation", "same channel candidate output" and "same-or-dual channel candidate output", for making recognition channel decision fusion on outputs of recognition channel-1 (6-axis IMU raw ANN) and outputs of recognition channel-2 (80-D deep learning feature ANN) can derive a total of 12 different recognition strategies on hand gesture intention recognition and are as follows:

- Combined *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by decision fusion of channel output layer accumulation,
- Combined *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by decision fusion of channel output layer accumulation,
- Combined *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by decision fusion of channel output layer accumulation,
- Combined *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by decision fusion of channel output layer accumulation,
- Combined *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by decision fusion of same channel candidate output,
- Combined *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by decision fusion of same channel candidate output,
- Combined *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by decision fusion of same channel candidate output,
- Combined *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by decision fusion of same channel candidate output,

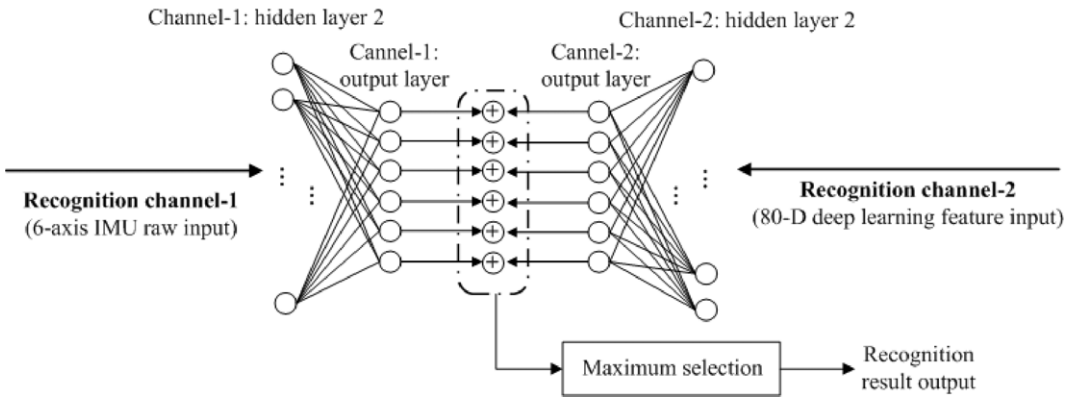


Figure 7. Illustrations of presented decision fusion of dual-channel ANN recognition using channel output layer accumulation for simultaneous considerations of 6-axis IMU raw and its deep learning features on hand gesture intention recognition.

- Combined *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by decision fusion of same-or-dual channel candidate output,
- Combined *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by decision fusion of same-or-dual channel candidate output,
- Combined *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by decision fusion of same-or-dual channel candidate output, and
- Combined *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by decision fusion of same-or-dual channel candidate output.

The presented three decision fusion approaches for hybridizing two different recognition channels of the 6-axis IMU raw and the 80-D deep learning feature are detailed in the following.

(1) Decision fusion of dual-channel ANN recognition by the *channel output layer accumulation* approach

The presented channel output layer accumulation approach is illustrated in Figure 7. As can be seen in Fig. 7, 6 node output values derived from the ANN output layer of recognition channel-1 and another 6 node output values derived from the ANN output layer of recognition channel-2 are considered to make the final recognition decision. In channel output layer accumulation, each node output value of recognition channel-1 is accumulated by its corresponding node output value of recognition channel-2; finally, a maximum operation is used on these 6 accumulated node output values for deriving the recognition result. Equations (6) and (7) show the calculations of presented channel output layer accumulation on decision fusion of dual-channel ANN recognition. Note that in Eqs. (6) and (7), n denotes the number of hand intention action categorizations, and n is equal to 6 (i.e. 6 classes for recognition test) in this work. The recognition result is the hand gesture intention classification label indicated by the max of all 6 accumulated output layered node values.

$$\begin{aligned}&\text{Accumulated output layered node}_i\\&\quad=\text{Output layered node(6 axis IMU raw)}_i\\&\qquad+\text{Output layered node(80D deep learning feature)}_i,\\&\qquad\quad i=1,2,\ldots,n.\end{aligned}\tag{6}$$

$$\begin{aligned}&\text{Recognition result}=\text{argmax}_{i=1,2\ldots,n}\ \text{Accumulated output}\\&\quad\text{layered node}_i.\end{aligned}\tag{7}$$

Note that in Eq. (6), the item Output layered node (6 axis IMU raw)$_i$ denotes the classification score of the ith categorization gesture intention derived from the corresponding ith node of the output layer of Channel-1 ANN recognition with the input of gesture actions represented by 6-axis IMU raw; another item Output layered node (80D deep learning feature)$_i$ in Eq. (6) is the classification score of the ith categorization gesture intention estimated from the corresponding ith node of the output layer of Channel-2 ANN recognition where the CNN deep learning feature, the 80D deep learning feature, is employed as the input data.

(2) Decision fusion of dual-channel ANN recognition by the *same channel candidate output* approach

Figure 8 illustrates the presented same channel candidate output approach on decision fusion of dual-channel ANN recognition. As shown in Fig. 8, the max node value index of all 6 nodes in the output layer of recognition channel-1 and recognition channel-2 is separately derived by a maximum operation, i.e., the candidate of channel-1 and the candidate of channel-2. If both candidates of channel-1 and channel-2 are the same, the recognition decision derived from the 6-axis IMU raw recognition channel and the 80-D deep learning feature recognition channel denotes the same hand gesture intention class. In this situation, the recognition result output is set as the classification label of the candidate of channel-1 (or the candidate of channel-2). Conversely, if the classification label of the candidate of channel-1 is different from that of the candidate of channel-2, the

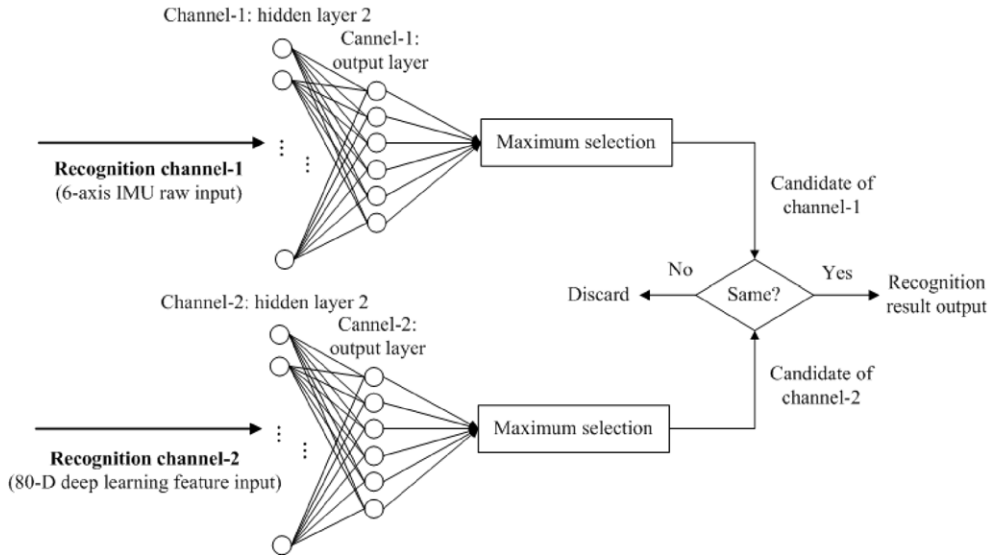


Figure 8. Illustrations of presented decision fusion of dual-channel ANN recognition using same channel candidate output for simultaneous considerations of 6-axis IMU raw and its deep learning features on hand gesture intention recognition.

hand gesture intention categorization recognized by two channels of 6-axis IMU raw and 80-D deep learning features is obviously different from each other. In this case, such input data for hand gesture intention action recognition calculation is perhaps substandard and therefore neglected directly, i.e. no recognition result is outputted to this input data. The rationale behind the presented same channel candidate output approach is that the recognition result will be outputted by the recognition system only in the case of reliable recognition result obtained.

(3) Decision fusion of dual-channel ANN recognition by the *same-or-dual channel candidate output* approach

The above-mentioned same channel candidate output approach is essentially "conditional recognition calculation". Although the same channel candidate output approach is expected to have an extremely satisfactory recognition accuracy due to only the reliable recognition result given, such type of conditional recognition calculation can occasionally encounter a problem of waiting for recognition results, i.e. possibly various times of input data given by the same user to obtain only one time of recognition result. To overcome this problem, the same-or-dual channel candidate output approach is presented. Figure 9 depicts the estimate procedure of the presented same-or-dual channel candidate output method. It can be obviously seen in Fig. 9 that the main difference between the same channel candidate output approach and the same-or-dual channel candidate output method occurs only in the situation when the candidate of channel-1 and the candidate of channel-2 are not the same. In such a situation, different from the same channel candidate output approach, presented same-or-dual channel candidate output herein will simultaneously give two recognition results to the user, both classification labels of channel-1 and channel-2 candidates. When one of both classification labels of channel-1 and channel-2 candidates is same as the classification label of the input hand gesture intention action categorization, recognition to this input data will be viewed to be correct recognition. The same-or-dual channel candidate output approach on decision fusion of dual-channel ANN recognition will give the recognition result in each input action data where the additional user verification task is required to make a correct choice only in the case of two different classification labels of channel-1 and channel-2 candidates.

4. EXPERIMENTS AND RESULTS

Experiments on hand gesture intention recognition with the MbientLab wearable bracelet device are made in a laboratory office environment. The 6-axis IMU raw information including 3-axis accelerometer and 3-axis gyroscope motion data is obtained from the wearable bracelet device and is then transmitted to the smart phone device for further signal analysis of the raw data as shown in Figure 10. As depicted in Fig. 10, when making a wireless connection between the wearable bracelet device with the MMC sensing platform and the specified smart phone device, the MMC sensing platform will first be detected, following which, a configuration setting task to select the desired type of sensing data to perform data acquisition on the sensor platform will then be done. As can be seen in Fig. 10, although the MMC sensing platform contains three different sensor types, accelerometer, gyroscope and temperature sensors, only accelerometer and gyroscope sensors are used and enabled in this work. After making a successful configuration, a session is then established to represent a new raw data capture mission. All types of sensing platforms developed by the MbientLab including MMC employed in this work use a csv-type file for dynamically recording the captured data during continuous-time period.

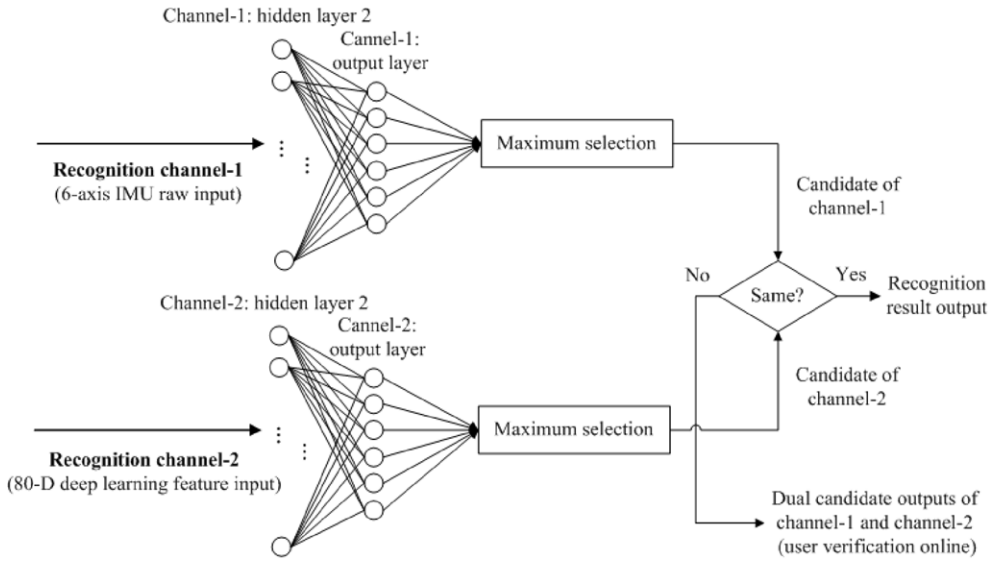


Figure 9. Illustrations of presented decision fusion of dual-channel ANN recognition using same-or-dual channel candidate output for simultaneous considerations of 6-axis IMU raw and its deep learning features on hand gesture intention recognition.

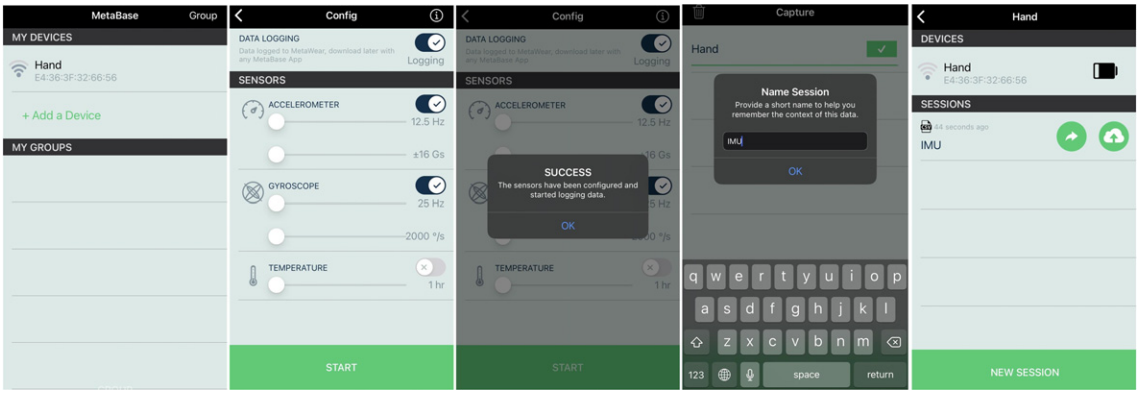
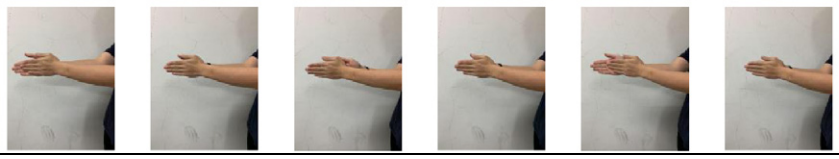


Figure 10. Connection, configuration and data capture settings on the specified smart phone device to be matched with the MMC sensing platform on the MbientLab wearable bracelet device.

Table I shows continuous-time actions of the defined six different hand gesture intention action categorizations of hand gesture intention recognition. These six different categorization actions, labeled as "Label-1", "Label-2", "Label-3", "Label-4", "Label-5" and "Label-6", represent "waving the arm from top to bottom", "putting the hand on the waist", "rubbing palms of both hands back and forth", "clapping both hands (both palms)", "crossing both arms on the chest" and "making a fist and beating the chest and then pointing the index finger at the opponent", respectively. Note that these operated actions of hand gesture intention categorizations, Label-1, Label-2, Label-3, Label-4, Label-5 and Label-6 denote common human intention behavior classes of "Anger", "Confidence", "Anxiety", "Joy", "Tension" and "Inspiration", respectively. These six defined hand gesture intention actions in this study are frequently seen and used to express an individual's specific social intention behavior in the daily life.

During establishment of the database containing the above-mentioned six different categorizations of hand gesture intention actions, three persons are requested to wear the MbientLab wearable bracelet with the 6-axis IMU raw motion data to make each of indicated continuous-time hand gesture actions. Each of these three hand gesture action-making persons make 50 actions for each categorization of hand gesture intention actions, a half for ANN model training (or the VGG-16 CNN model training) and the other half for recognition rate evaluations in the recognition test phase. Totally, 900 hand gesture intention actions are captured in the database, 150 actions for each of the defined six specified hand gesture intention categorization actions. In the collected database, 450 actions with 75 actions contained in each of the six hand gesture intention categorizations are employed as the training data to establish the recognition model, and the other 450 actions are used as the test data for performance evaluation of constructed models.

Table I. Six different continuous-time hand gesture intention actions with the 6-axis IMU raw sensing wearable bracelet defined in this work.

Hand intention	Six different categorizations of continuous-time hand intention actions with the 6-axis IMU wearable band
Label-1 ('Anger' intention)	
Label-2 ('Confidence' intention)	
Label-3 ('Anxiety' intention)	
Label-4 ('Joy' intention)	
Label-5 ('Tension' intention)	
Label-6 ('Inspiration' Intention)	

Tables II and III show the experimental results of hand gesture intention recognition by symmetric and asymmetric ANN classification with 6-axis IMU raw data. Observed from averaged recognition accuracy on classifications of these six hand gesture intention action types, asymmetric ANN with the structure of *6-10-20-6* is slightly superior to the symmetric ANN with the *6-13-13-6* structure by 3.34%. However, it seems that hand gesture intention recognition by ANN classification with 6-axis IMU raw data is not very satisfactory, only 60.67% of *6-10-20-6* ANN with 6-axis IMU raw and 57.33% of *6-13-13-6* ANN with 6-axis IMU raw achieved.

In the phase of hand gesture intention recognition by the typical VGG-16 CNN deep learning model approach with 6 axis IMU raw-derived spectrogram images, frequency-domain spectrogram images of each (Type-1, Type-2, Type-3, Type-4, Type-5 and Type-6) of these six continuous-time hand gesture intention categorization actions with time-domain sample signals of 6-axis IMU raw are depicted in Figures 11, 12, 13, 14, 15 and 16, respectively. Note that in this work of hand gesture intention recognition, as mentioned, each action categorization has recorded 75 actions for recognition test. In each of Figs. 11–16, for simplicity, only the recent ten test actions for each type (totally, 75 test actions recorded for each categorization) are shown. Each gesture intention action is represented by the acquired sample signals of 6-axis IMU raw from the wearable bracelet in certain continuous-time period, and such numerous collected IMU raw samples is FFT-transformed to an unique IMU-spectrogram. In Figs. 11–16, each image of the IMU-spectrogram represents the corresponding action with the continuous-time 6-axis IMU raw data. It can be clearly seen in Figs. 11–16, each of ten IMU-spectrograms that belong to the same action categorization has the similar pixel property. IMU-spectrograms categorized into different types are apparently distinct with each other, and such well-separated pattern categorization images will contribute significantly in typical VGG-16 CNN recognition and further extractions of the established deep learning feature from VGG-16 CNN. Table IV lists the averaged recognition performance of VGG-16 CNN with the 6 axis IMU raw-derived spectrogram image. As can be seen in Table IV, the recognition accuracy reaches 75.78%, which

Table II. Recognition accuracy of *6-13-13-6* ANN classifications with 6-axis IMU raw data.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	27	21	3	3	7	14
Type-2	2	61	4	1	6	1
Type-3	1	4	60	5	3	2
Type-4	3	11	11	29	15	6
Type-5	0	8	12	13	41	1
Type-6	9	7	4	8	7	40
Average recognition accuracy: 57.33%						

Table III. Recognition accuracy of *6-10-20-6* ANN classifications with 6-axis IMU raw data.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	32	13	3	3	8	16
Type-2	4	65	1	2	0	3
Type-3	1	3	56	6	8	1
Type-4	1	6	13	35	12	8
Type-5	0	5	7	14	44	5
Type-6	13	4	4	3	10	41
Average recognition accuracy: 60.67%						

Table IV. Recognition accuracy of VGG-16 CNN deep learning model with the RGB image information of 6 axis IMU-derived spectrogram images.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	65	5	0	1	0	4
Type-2	3	39	8	16	8	1
Type-3	2	0	62	10	1	0
Type-4	0	1	6	65	3	0
Type-5	4	3	8	4	55	1
Type-6	10	3	0	0	7	55
Average recognition accuracy: 75.78%						

is more acceptable than those of ANN classifications with time-domain signals of the 6-axis IMU raw data.

Tables V and VI present the results of ANN classifications with principal component-extracted critical deep learning features. As mentioned before, the 4096D deep learning feature extracted from calculations of VGG-16 CNN can further be derived from critical 80-D deep learning feature data by PCA estimates. For the use of critical 80-D deep learning features, in comparisons of symmetric ANN with the *80-9-9-6* structure and asymmetric ANN with the *80-10-20-6* structure, symmetric ANN is slightly better than asymmetric ANN on the recognition performance, which is apparently different to the comparison result of symmetric and asymmetric ANN classifications with 6-axis IMU raw data. However, in ANN classifications with the critical 80-D deep learning feature data, neither the symmetric nor the asymmetric ANN has the standard recognition performance.

Tables VII–X show the average recognition accuracy of the channel output layer accumulation approach that is used for decision fusion of dual-channel ANN recognition of the 6-axis IMU raw and its critical 80-D critical deep learning features, *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features, *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features, *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep

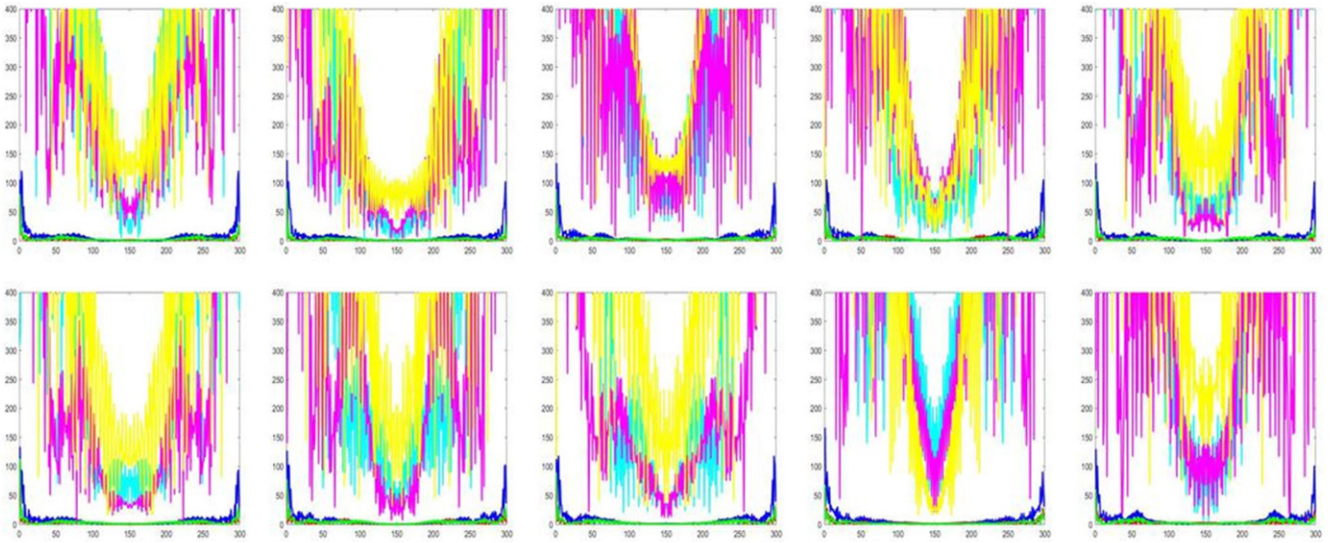


Figure 11. IMU-based spectrogram images (frequency-domain spectrograms) of "Type-1" hand gesture intention actions where 6-axis IMU raw data is transformed by FFT calculations (for simplicity, only the previous ten Type-1 test actions shown, totally 75 Type-1 actions in the test action database).

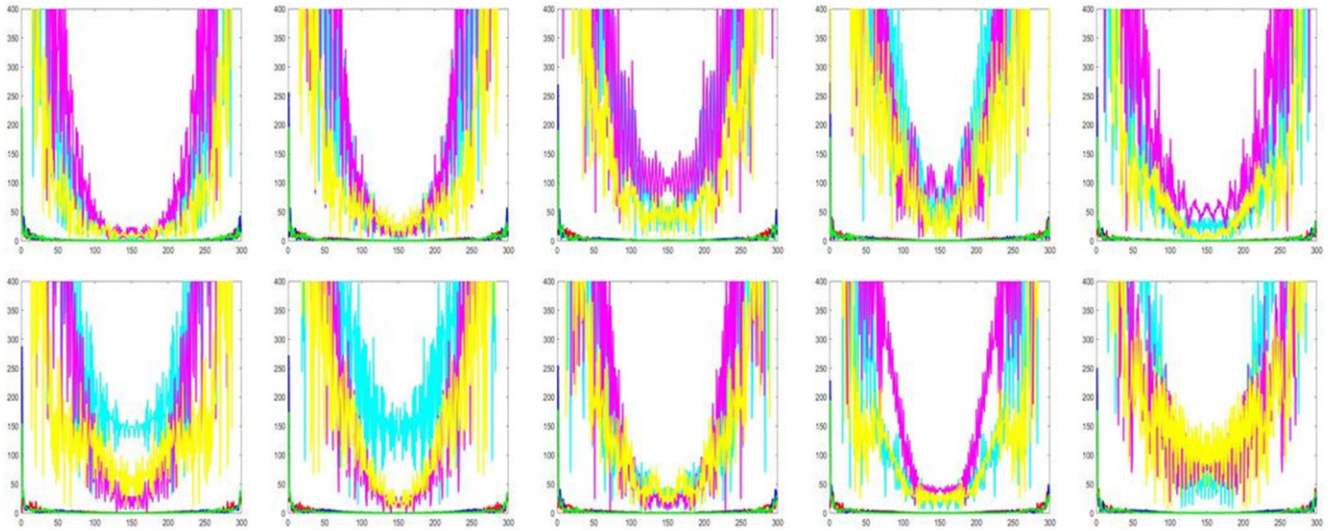


Figure 12. IMU-based spectrogram images (frequency-domain spectrograms) of "Type-2" hand gesture intention actions where 6-axis IMU raw data is transformed by FFT calculations (for simplicity, only the recent ten Type-2 test actions shown, totally 75 Type-2 actions in the test action database).

learning features, and *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features, respectively. The recognition performances among these different dual-channel ANN recognition hybridizations by

Table V. Recognition accuracy of *80-9-9-6* ANN classifications with 80-D deep learning features.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	49	11	5	6	0	4
Type-2	7	33	11	11	8	5
Type-3	2	8	41	16	7	1
Type-4	2	8	11	46	6	2
Type-5	11	14	1	7	34	8
Type-6	6	10	7	11	2	39
Average recognition accuracy: 53.78%						

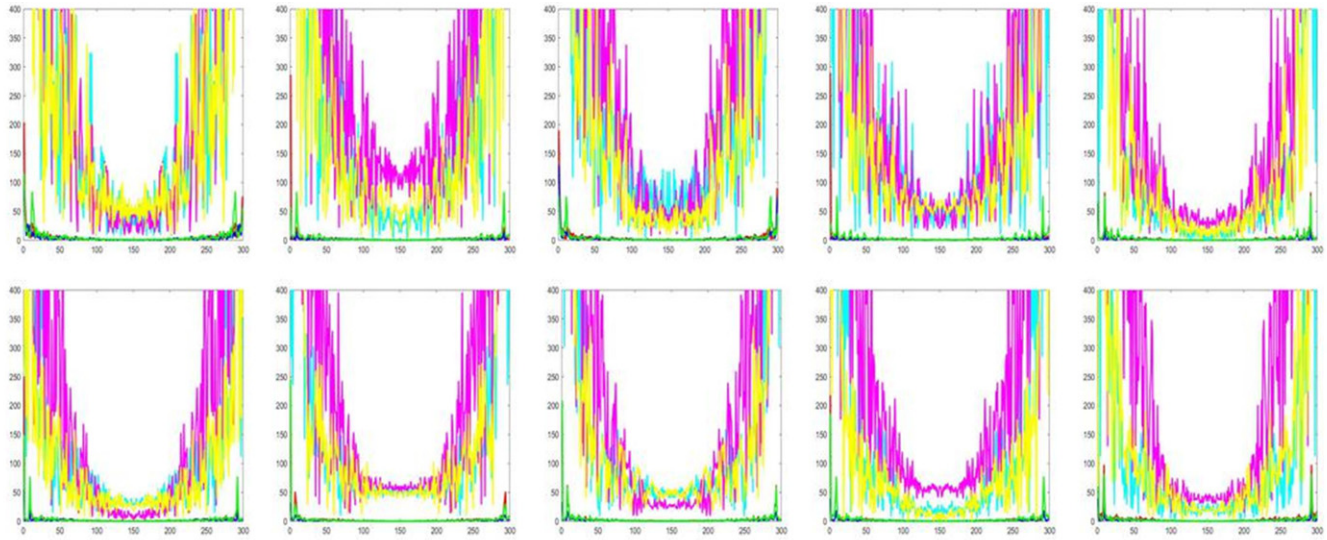


Figure 13. IMU-based spectrogram images (frequency-domain spectrograms) of "Type-3" hand gesture intention actions where 6-axis IMU raw data is transformed by FFT calculations (for simplicity, only the recent ten Type-3 test actions shown, totally 75 Type-3 actions in the test action database).

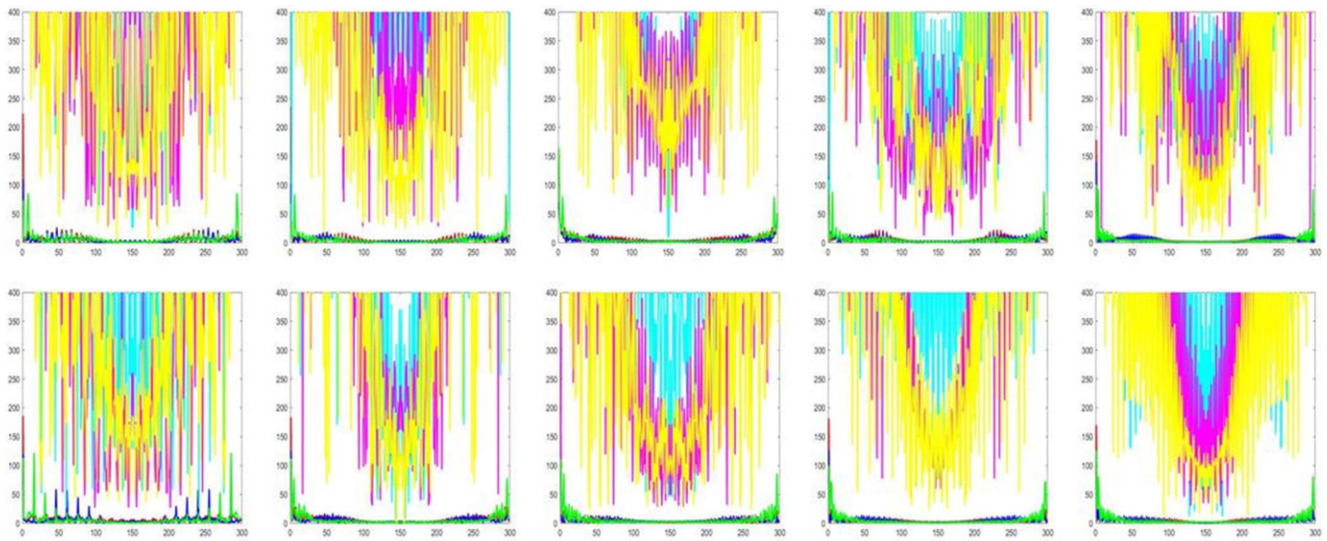


Figure 14. IMU-based spectrogram images (frequency-domain spectrograms) of "Type-4" hand gesture intention actions where 6-axis IMU raw data is transformed by FFT calculations (for simplicity, only the recent ten Type-4 test actions shown, totally 75 Type-4 actions in the test action database).

Table VI. Recognition accuracy of *80-10-20-6* ANN classifications with 80-D deep learning features.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	46	14	6	4	0	5
Type-2	4	38	9	7	9	8
Type-3	3	4	38	15	8	7
Type-4	3	10	10	38	8	6
Type-5	4	26	5	1	30	9
Type-6	6	11	7	8	2	41
Average recognition accuracy: 51.33%						

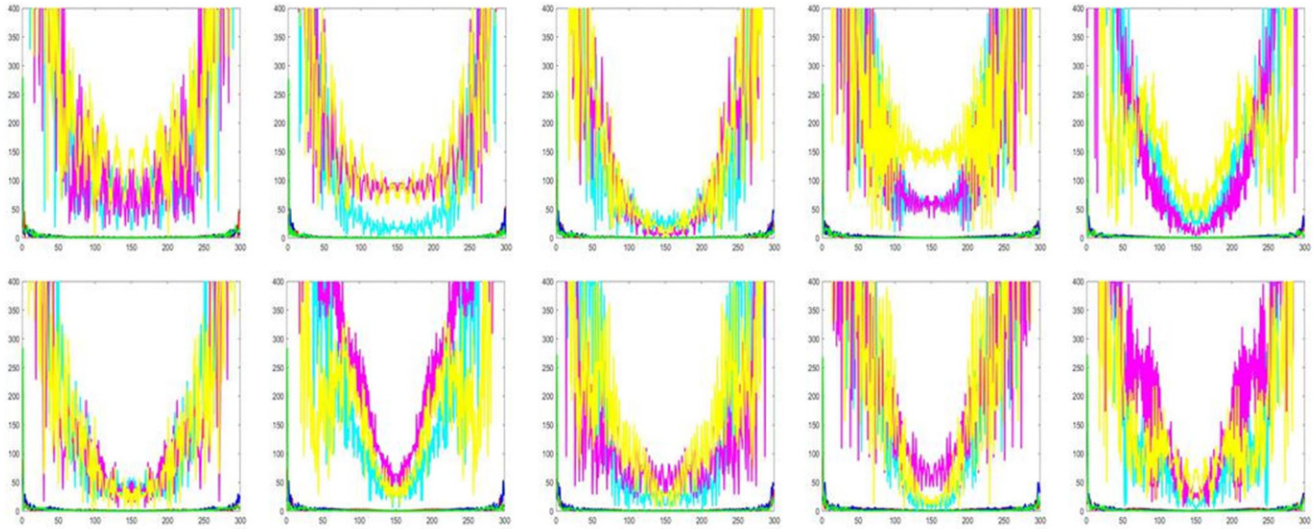


Figure 15. IMU-based spectrogram images (frequency-domain spectrograms) of "Type-5" hand gesture intention actions where 6-axis IMU raw data is transformed by FFT calculations (for simplicity, only the recent ten Type-5 test actions shown, totally 75 Type-5 actions in the test action database).

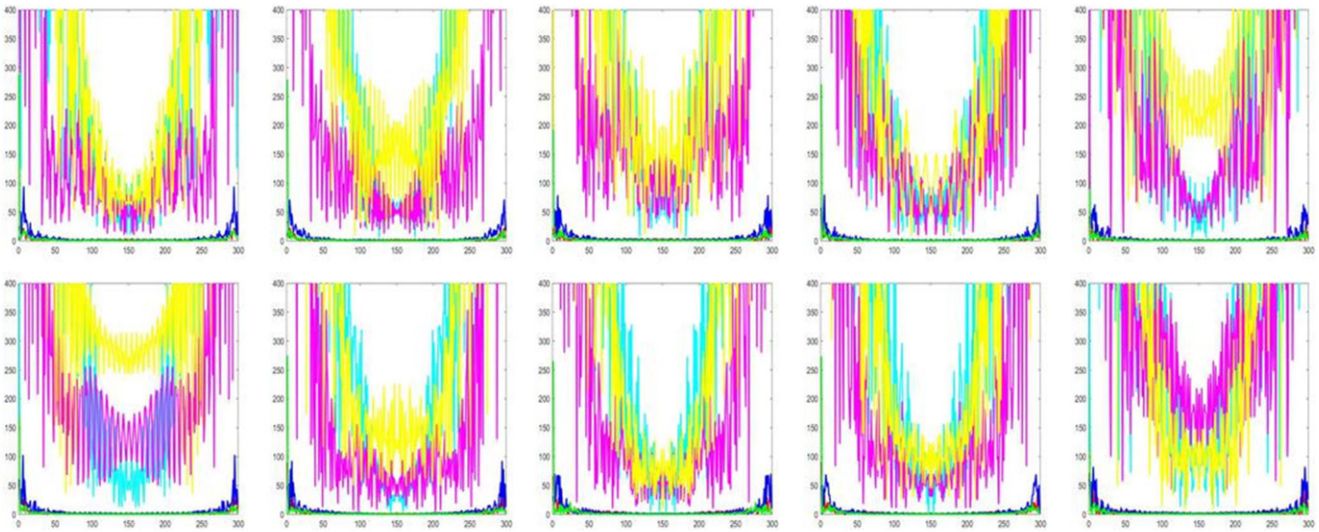


Figure 16. IMU-based spectrogram images (frequency-domain spectrograms) of "Type-6" hand gesture intention actions where 6-axis IMU raw data is transformed by FFT calculations (for simplicity, only the recent ten Type-6 test actions shown, totally 75 Type-6 actions in the test action database).

channel output layer accumulation are 66.67%, 64%, 62.44% and 60.67%. By properly incorporating the ANN decision estimate of the ANN recognition channel of critical 80-D critical deep learning features to the recognition output, almost all of these recognition outcomes are better than those of the single channel ANN recognition with 6-axis IMU raw data and those of the single channel ANN recognition with 80-D critical deep learning features. Although the recognition accuracy of decision fusion by the channel output layer accumulation approach is still lower than that of typical VGG-16 CNN with 6 axis IMU-derived spectrogram images, such presented decision fusion scheme is effective in recognition accuracy increase to single channel ANN without any considerations on model fusion.

Experimental results of dual-channel ANN recognition hybridizations by the same channel candidate output approach are listed in Tables XI–XIV. The averaged recognition accuracy of combinations of the IMU raw ANN channel and the critical deep learning feature ANN channel, combined *6-10-20-6* ANN and *80-10-20-6* ANN, combined *6-10-20-6* ANN and *80-9-9-6* ANN, combined *6-13-13-6* ANN and *80-10-20-6* ANN, and combined *6-13-13-6* ANN and *80-9-9-6* ANN, are 88.69%, 87.15%, 84.91% and 88.17%, respectively. The best of these four different hybridizations by the same channel candidate output on recognition performance is the ANN model fusion of the asymmetric ANN with IMU raw and the asymmetric ANN with critical deep learning features. Note that all averaged recognition performances of ANN recognition channel hybridizations by the presented same channel candidate output approach are obviously superior to 75.78% of the approach of VGG-16 CNN with IMU-derived spectrogram images, which strongly

Table VII. Recognition accuracy of combined *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by the channel output layer accumulation approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	54	7	2	1	2	9
Type-2	5	56	3	3	3	5
Type-3	2	2	53	9	7	2
Type-4	2	5	9	44	11	4
Type-5	3	11	3	6	45	7
Type-6	6	6	5	6	4	48
Average recognition accuracy: 66.67%						

Table VIII. Recognition accuracy of combined *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by the channel output layer accumulation approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	52	8	3	2	1	9
Type-2	8	45	5	3	8	6
Type-3	0	4	52	8	9	2
Type-4	3	6	6	45	9	6
Type-5	9	3	5	7	45	6
Type-6	5	5	2	7	7	49
Average recognition accuracy: 64%						

Table IX. Recognition accuracy of combined *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by the channel output layer accumulation approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	52	11	2	2	1	7
Type-2	3	52	6	2	9	3
Type-3	0	2	50	9	9	5
Type-4	3	8	8	40	12	4
Type-5	4	7	9	8	39	8
Type-6	5	6	5	7	4	48
Average recognition accuracy: 62.44%						

reveals the significant effectiveness of the presented decision fusion scheme.

Tables XV–XVIII depict the averaged recognition rates of hand gesture intention recognition by the same-or-dual channel candidate output approach to hybridize recognition decisions of dual-channel ANN recognition. By this presented decision method, the recognition rates become 78.89% of combined *6-10-20-6* ANN and *80-10-20-6* ANN, 79.78% of combined *6-10-20-6* ANN and *80-9-9-6* ANN, 78.67% of combined *6-13-13-6* ANN and *80-10-20-6* ANN, and 78% of combined *6-13-13-6* ANN and *80-9-9-6* ANN, all of which also performs more competitive than that of typical VGG-16 CNN with IMU-derived spectrogram images, those of ANN classifications with 6-axis IMU raw data and those of ANN classifications with 80-D critical deep learning features, as the same channel candidate output approach.

Finally, for clearly illustrating the effectiveness of various hand gesture intention recognition approaches, a recognition result summary is shown in Table XIX. Note that in Table XIX, for presented three different decision fusion

Table X. Recognition accuracy of combined *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by the channel output layer accumulation approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	45	8	5	3	4	10
Type-2	6	48	7	3	6	5
Type-3	1	6	49	7	9	3
Type-4	3	7	7	40	15	3
Type-5	5	5	9	8	41	7
Type-6	4	7	2	7	5	50
Average recognition accuracy: 60.67%						

Table XI. Recognition accuracy of combined *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by the same channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	17	2	1	0	0	2
Type-2	1	36	0	0	0	0
Type-3	0	0	28	0	0	0
Type-4	0	1	2	19	4	0
Type-5	0	2	0	0	26	0
Type-6	2	1	1	0	0	23
Average recognition accuracy: 88.69%						

Table XII. Recognition accuracy of combined *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by the same channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	20	1	1	1	0	2
Type-2	1	29	0	0	0	0
Type-3	0	1	29	1	1	0
Type-4	0	1	4	24	3	0
Type-5	0	1	0	2	29	0
Type-6	0	1	1	0	1	25
Average recognition accuracy: 87.15%						

approaches for hybridizations of dual ANN recognition channels in this study, only the hybridized framework with the best recognition accuracy is chosen for performance comparisons. It can be inferred from Table XIX, that single-channel ANN recognition approaches with 6-axis IMU raw data and critical deep learning features have apparently only the substandard performance on recognition, approaching to 60.67% and 53.78%. The typical VGG-16 CNN recognition approach that uses the IMU-derived spectrogram images for classification calculations performs still a little unsatisfactory, achieving 75.78%. Compared with the conventional single-channel ANN recognition approach and the typical VGG-16 CNN method, presented dual-channel ANN recognition hybridizations of 6-axis IMU raw and 80-D deep learning features by decision fusion of the same channel candidate output scheme and the same-or-dual channel candidate output scheme have apparently more acceptable performances with a significant recognition accuracy increment; the best is 88.69% of the same channel candidate output approach, followed by 79.78% of the same-or-dual channel candidate output approach. Although decision fusion by the channel output layer accumulation scheme is not superior to typical VGG-16 CNN on recognition accuracy, it is still much more competitive than conventional

Table XIII. Recognition accuracy of combined *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by the same channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	14	4	0	0	0	2
Type-2	0	33	1	0	0	0
Type-3	0	0	29	0	1	0
Type-4	0	4	2	12	2	0
Type-5	0	1	0	0	25	0
Type-6	3	3	0	1	0	22
Average recognition accuracy: 84.91%						

Table XIV. Recognition accuracy of combined *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by the same channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	18	2	0	1	0	2
Type-2	0	28	1	0	0	0
Type-3	0	1	33	0	1	0
Type-4	0	0	2	19	1	0
Type-5	0	1	0	2	28	0
Type-6	0	3	1	2	0	23
Average recognition accuracy: 88.17%						

Table XV. Recognition accuracy of combined *6-10-20-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by the same-or-dual channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	61	5	1	1	3	4
Type-2	2	67	1	2	0	3
Type-3	1	1	66	3	3	1
Type-4	0	3	6	54	10	2
Type-5	0	3	7	12	48	5
Type-6	7	1	3	1	4	59
Average recognition accuracy: 78.89%						

single-channel ANN recognition. Note that for 88.69% of the same channel candidate output approach, it can also been further observed the recognition performance of each of six different categorizations of gesture intention actions from the confusion matrix listed in Table XI. Observed from Table XI, the intention action of Type-3 performs best, achieving perfectly complete recognition, followed by the Type-2 action with only one wrong recognition result, and the action of Type-4 performs relatively imperfect, totally 7 incorrect recognition outcomes appeared. It is also noted that as mentioned in Section 3.3, in the design of the same channel candidate output approach, the recognition result of gesture intention action labels can be sent out from the system (i.e. a recognition decision made by the system) only in the situation that recognized labels estimated from each of two recognition channels are the same. It can be seen in Table XI (also see Tables XII–XIV) that each type action has inconsistent numbers of recognition decision outputs (e.g., 19 of Type-1, 37 of Type-2, 28 of Type-3, 26 of Type-4, 28 of Type-5 and 27 of Type-6 in Table XI). For recognition results of the other recognition methods (Tables V, VI, VII, VIII, IX, X, XV, XVI, XVII and XVIII), the recognition decision number of each type action in the confusion matrix keeps a consistent value of 75. From the viewpoint of

Table XVI. Recognition accuracy of combined *6-10-20-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by the same-or-dual channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	61	4	2	2	1	5
Type-2	2	69	1	1	0	2
Type-3	0	2	68	3	1	1
Type-4	0	1	7	57	7	3
Type-5	0	2	7	12	49	5
Type-6	12	1	2	0	5	55
Average recognition accuracy: 79.78%						

Table XVII. Recognition accuracy of combined *6-13-13-6* ANN with 6-axis IMU raw and *80-10-20-6* ANN with 80-D deep learning features by the same-or-dual channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	59	7	0	0	3	6
Type-2	1	66	3	1	3	1
Type-3	1	0	69	3	1	1
Type-4	1	6	5	55	8	0
Type-5	0	5	12	11	46	1
Type-6	5	3	2	4	2	59
Average recognition accuracy: 78.67%						

Table XVIII. Recognition accuracy of combined *6-13-13-6* ANN with 6-axis IMU raw and *80-9-9-6* ANN with 80-D deep learning features by the same-or-dual channel candidate output approach.

Input test hand gesture types	Six different hand gesture classifications on recognition outputs					
	Label-1	Label-2	Label-3	Label-4	Label-5	Label-6
Type-1	58	6	1	1	3	6
Type-2	1	66	3	1	4	0
Type-3	0	1	68	3	1	2
Type-4	2	3	6	56	7	1
Type-5	0	4	12	11	47	1
Type-6	6	4	3	5	1	56
Average recognition accuracy: 78%						

Table XIX. Recognition accuracy comparisons on hand gesture intention recognition of different categorization approaches of single channel ANN recognition\VGG-16 CNN without any fusion and dual-channel ANN recognition hybridizations by different decision fusion schemes.

Single channel ANN recognition with 6-axis IMU raw data or 80-D deep learning features \typical VGG-16 CNN recognition			Dual-channel ANN recognition hybridizations of 6-axis IMU raw and 80-D deep learning features by different decision fusion methods		
6-10-20-6 ANN with 6-axis IMU raw data	VGG-16 CNN with 6 axis IMU-derived spectrograms	80-9-9-6 ANN with 80-D deep learning features	The channel output layer accumulation approach	The same channel candidate output approach	The same-or-dual channel candidate output approach
60.67%	75.78%	53.78%	66.67%	88.69%	79.78%

hand gesture intention action recognition with the high performance requirement of reliably-decided recognition outcomes, dual-channel ANN recognition decision fusion by the same channel candidate output approach has the best competitiveness.

5. CONCLUSIONS

In this study, a dual-channel ANN recognition hybridization scheme with considerations of both IMU raw data and its derived critical deep learning feature information is proposed for classifications of several common hand gesture intention actions. Three different decision fusion approaches, channel output layer accumulation, same channel candidate output and same-or-dual channel candidate output, are presented for recognition channel hybridizations. Experimental results demonstrate the effectiveness of the developed scheme on recognition accuracy of hand gesture intention actions, a significant improvement on ANN recognition with IMU raw or its critical deep learning features and typical VGG-CNN deep neural network recognition.

ACKNOWLEDGMENTS

This research is partially supported by the Ministry of Science and Technology (MOST) in Taiwan under Grant MOST 109-2221-E-150-034-MY2.

REFERENCES

[1] Y. Kim, T. Soyata, and R. F. Behnagh, "Towards emotionally aware AI smart classroom: current issues and directions for engineering and education," IEEE Access **6**, 5308–5331 (2018).

[2] N. Weibel, S. Hwang, S. Rick, E. Sayyari, D. Lenzen, and J. Hollan, "Hands that speak: an integrated approach to studying complex human communicative body movements," *Proc. 49th Hawaii Int'l. Conf. on System Sciences (HICSS)* (IEEE, Piscataway, NJ, 2016), pp. 610–619.

[3] I. J. Ding and S. K. Lin, "Performance improvement of kinect software development kit–constructed speech recognition using a client-server sensor fusion strategy for smart human computer interface control applications," IEEE Access **5**, 4154–4162 (2017).

[4] I. J. Ding and J. Y. Shi, "Kinect microphone array–based speech and speaker recognition for the exhibition control of humanoid robots," Comput. Electrical Eng. **62**, 719–729 (2017).

[5] I. J. Ding and C. M. Ruan, "A study on utilizations of 3d sensor lip image for developing a pronunciation recognition system," J. Imag. Sci. Technol. **63**, 50402-1–50402-9 (2019).

[6] I. J. Ding, C. T. Yen, and D. C. Ou, "A method to integrate GMM, SVM and DTW for Speaker Recognition," Int. J. Eng. Technol. Innov. **4**, 38–47 (2014).

[7] I. J. Ding and S. K. Lin, "A wireless sensor network-speech recognition scheme using deployments of multiple kinect microphone array-sensors," Proc. Eng. Technol. Innov. **3**, 25–27 (2016).

[8] R. He, J. Cao, L. Song, Z. Sun, and T. Tan, "Adversarial cross-spectral face completion for NIR-VIS face recognition," IEEE Trans. Pattern Anal. Mach. Intell. **42**, 1025–1037 (2020).

[9] F. Liu, Q. Zhao, X. Liu, and D. Zeng, "Joint face alignment and 3D face reconstruction with application to face recognition," IEEE Trans. Pattern Anal. Mach. Intell. **42**, 664–678 (2020).

[10] M. P. Yankov, M. A. Olsen, M. B. Stegmann, S. S. Christensen, and S. Forchhammer, "Fingerprint entropy and identification capacity estimation based on pixel-level generative modelling," IEEE Trans. Inf. Forensics Secur. **15**, 56–65 (2020).

[11] V. Anand and V. Kanhangad, "PoreNet: CNN-based pore descriptor for high-resolution fingerprint recognition," IEEE Sensors J. **20**, 9305–9313 (2020).

[12] T. Grzejszczak, M. Kawulok, and A. Galuszka, "Hand landmarks detection and localization in color images," Multimedia Tools Appl. **75**, 16363–16387 (2016).

[13] Z. Hu and X. Zhu, "Gesture detection from RGB hand image using modified convolutional neural network," *Proc. 2nd Int'l. Conf. on Information Systems and Computer Aided Education (ICISCAE)* (IEEE, Piscataway, NJ, 2019), pp. 143–146.

[14] S. Zhang, W. Meng, H. Li, and X. Cui, "Multimodal spatiotemporal networks for sign language recognition," IEEE Access **7**, 180270–180280 (2019).

[15] P. V. V. Kishore, D. A. Kumar, A. S. C. S. Sastry, and E. K. Kumar, "Motionlets matching with adaptive kernels for 3-D indian sign language recognition," IEEE Sensors J. **18**, 3327–3337 (2018).

[16] T. Y. Pan, C. Y. Chang, W. L. Tsai, and M. C. Hu, "Multisensor-based 3D gesture recognition for a decision-making training system," IEEE Sensors J. **21**, 706–716 (2021).

[17] C. Yoo, S. Ji, Y. Shin, S. Kim, and S. Ko, "Fast and accurate 3D hand pose estimation via recurrent neural network for capturing hand articulations," IEEE Access **8**, 114010–114019 (2020).

[18] I. J. Ding and Z. G. Wu, "Two user adaptation-derived features for biometrical classifications of user identity in 3D-sensor-based body gesture recognition applications," IEEE Sensors J. **19**, 8432–8440 (2019).

[19] I. J. Ding and M. C. Hsieh, "A hand gesture action-based emotion recognition system by 3D image sensor information derived from leap motion sensors for the specific group with restlessness emotion problems", *Microsystem Technologies*, publication on-line first, May 13, 2020.

[20] I. J. Ding, N. W. Zheng, and M. C. Hsieh, "Hand gesture intention-based identity recognition using various recognition strategies incorporated with VGG convolution neural network-extracted deep learning features," J. Intell. Fuzzy Syst. **40**, 7775–7788 (2021).

[21] N. Dawar and N. Kehtarnavaz, "Real-time continuous detection and recognition of subject-specific smart TV gestures via fusion of depth and inertial sensing," IEEE Access **6**, 7019–7028 (2018).

[22] W. Aly, S. Aly, and S. Almotairi, "User-independent american sign language alphabet recognition based on depth image and pcanet features," IEEE Access **7**, 123138–123150 (2019).

[23] L. Wang, J. Liu, and J. Lan, "Feature evaluation of upper limb exercise rehabilitation interactive system based on kinect," IEEE Access **7**, 165985–165996 (2019).

[24] I. J. Ding, R. Z. Lin, and Z. Y. Lin, "Service robot system with integration of wearable myo armband for specialized hand gesture human–computer interfaces for people with disabilities with mobility problems," Comput. Electr. Eng. **69**, 815–827 (2018).

[25] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A fully embedded adaptive real-time hand gesture classifier leveraging HD-sEMG and deep learning," IEEE Trans. Biomed. Circuits Syst. **14**, 232–243 (2020).

[26] Y. Zhang, Y. Chen, H. Yu, X. Yang, and W. Lu, "Learning effective spatial–temporal features for sEMG armband-based gesture recognition," IEEE Internet Things J. **7**, 6979–6992 (2020).

[27] S. Pareek, H. Manjunath, E. T. Esfahani, and T. Kesavadas, "MyoTrack: realtime estimation of subject participation in robotic rehabilitation using sEMG and IMU," IEEE Access **7**, 76030–76041 (2019).

[28] S. Jiang, B. Lv, W. Guo, C. Zhang, H. Wang, X. Sheng, and P. B. Shull, "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via sEMG and IMU sensing," IEEE Trans. Indust. Inform. **14**, 3376–3385 (2018).

[29] J. G. Colli-Alfaro, A. Ibrahim, and A. L. Trejos, "Design of user-independent hand gesture recognition using multilayer perceptron networks and sensor fusion techniques," *Proc. IEEE 16th Int'l. Conf. on Rehabilitation Robotics (ICORR)* (IEEE, Piscataway, NJ, 2019), pp. 1103–1108.

[30] S. Shin, Y. Baek, J. Lee, Y. Eun, and S. H. Son, "Korean sign language recognition using EMG and IMU sensors based on group-dependent NN models," *Proc. IEEE Symposium Series on Computational Intelligence (SSCI)* (IEEE, Piscataway, NJ, 2017), pp. 1–7.

[31] S. P. Y. Jane and S. Sasidhar, "Sign language interpreter: classification of forearm EMG and IMU signals for signing exact english," *Proc. IEEE 14th*

Int'l. Conf. on Control and Automation (ICCA) (IEEE, Piscataway, NJ, 2018), pp. 947–952.

[32] A. Menshchikov, D. Ermilov, I. Dranitsky, L. Kupchenko, M. Panov, M. Fedorov, and A. Somov, "Data-driven body-machine interface for drone intuitive control through voice and gestures," *Proc. 45th Annual Conf. of the IEEE Industrial Electronics Society* (IEEE, Piscataway, NJ, 2019), pp. 5602–5609.

[33] Y. Ma, Y. Liu, R. Jin, X. Yuan, R. Sekha, S. Wilson, and R. Vaidyanathan, "Hand gesture recognition with convolutional neural networks for the multimodal UAV control," *Proc. 2017 Workshop on Research, Education and Development of Unmanned Aerial Systems (RED-UAS)* (IEEE, Piscataway, NJ, 2017), pp. 198–203.

[34] R. M. Stephenson, G. R. Naik, and R. Chai, "A system for accelerometer-based gesture classification using artificial neural networks," *Proc. 39th Annual Int'l. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, Piscataway, NJ, 2017), pp. 4187–4190.

[35] P. Koch, N. Brügge, H. Phan, M. Maass, and A. Mertins, "Forked recurrent neural network for hand gesture classification using inertial measurement data," *Proc. IEEE Int'l. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, Piscataway, NJ, 2019), pp. 2877–2881.

[36] K. Suri and R. Gupta, "Classification of hand gestures from wearable IMUs using deep neural network," *Proc. Int'l. Conf. on Inventive Communication and Computational Technologies (ICICCT)* (IEEE, Piscataway, NJ, 2018), pp. 45–50.

[37] E. Abraham, A. Nayak, and A. Iqbal, "Real-time translation of indian sign language using LSTM," *Proc. 2019 Global Conf. for Advancement in Technology (GCAT)* (IEEE, Piscataway, NJ, 2019), pp. 1–5.

[38] https://mbientlab.com/, retrieved on November 2021.

[39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Proc. Int'l. Conf. on Learning Representations (ICLR)* (New York, NY, 2015), eprint arXiv:1409.1556.