2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Low-Cost Wearable Hand Gesture Detecting System Based on IMU and Convolutional Neural Network

Pu-Fan Xu, Zi-Xuan Liu, Fei Li, and Hai-Peng Wang

Abstract—**In this paper, a low-cost wearable hand gesture detecting system based on distributed multi-node inertial measurement units (IMUs) and central node microcontroller is presented. It can obtain hand kinematic information and transmit data to the remote processing terminal wirelessly. To have a comprehensive understanding of hand kinematics, a convolutional neural network (CNN) model on the terminal is proposed to recognize and classify gestures and the modified Denavit-Hartenberg notation is used to acquire finger spatial locations. The experiment has not only completed a variety of gesture recognitions, but also captured and displayed the orientation and posture of a single finger. The prototype can be used in various occasions such as hand rehabilitation evaluation and human-computer interaction.**

I. Introduction

Capturing and analyzing hand kinematics is essential for medical rehabilitation, estimating recovery of hand function impairments [1], and human-computer interaction, aimed at virtual reality (VR) and game applications. Owing to the great desire of obtaining finger range of motion (ROM) conveniently and explaining datasets comprehensively, numerous wearable measurement monitors and data gloves are presented.

A wearable finger flexion monitor with bend sensors was raised by Lisa K Simone to replace a traditional measurement glove in-home and community activities [2]. Because the relative orientation of finger segments is measured by bend sensors placed across the joints, results can be easily affected by sensor displacements and a careful alignment is required before using it. Usually fixed on fabric gloves, it is inconvenient for repairment and replacement if there are sensor-failures. Some commercialized data gloves were emerged as well, such as 5DT Data Glove and CyberGlove®II. However, the unchangeable sensor-distance and unified glove-size limit their adaptabilities [3]. In addition, due to the high price, the application scenarios of those gloves are usually restricted to the academic use [3].

To solve the drawbacks of bend-sensor monitors and data gloves mentioned above, inertial measurement unit (IMU) is a good choice. Distributed on phalanxes without depending on the deformation, IMUs limit the displacement error of bend sensors. With a modular design, IMUs can support customization and replacement at the same time. Inertial sensors are widely used in human machine interface and clinical setting for hand function evaluation. Seong designed a natural user interface by tracking hand motion with Euler angels and recognizing click gesture with a machine learning method [4]. Henk G Kortier proposed an ambulatory system based on inertial sensors which can obtain fingertip position using forward kinematics with quaternions [5]. Combining these applications, a hand gesture detecting system can be constructed to classify gestures and calculate finger orientation simultaneously. Compared with machine learnings, convolutional neural networks (CNNs) are attractive as they can calculate high density datasets, deal with more gesture classes, and design features from raw data exactly [6], which suit the characteristics of hand kinematics and have the potential to be applied to a more sophisticated measuring platform in the future. As for figure positions, quaternions and Euler angels can be represented by modified Denavit-Hartenberg (D-H) notation, for its advantages in finger modeling, simulation, and coordinate transformation [7].

In order to eliminate the aforementioned limitations, a low-cost wearable hand gesture detecting system is proposed, which consists of a measuring platform, a hand gesture recognizer, and a finger position calculator and simulator. It has a modular structure that contains distributed multi-node IMUs with three-axis acceleration and three-axis angular velocity senor placed on each hand joint. In addition, to obtain hand kinematic information and transmit data to remote processing terminal wirelessly, a central node microcontroller with Bluetooth interface is installed on hand back. Specifically, a CNN model on the terminal realizes an accurate identification and classification of hand gesture. Besides, to acquire finger spatial locations and match up the sensor array, it is a new beneficial attempt to applied modified D-H notation and transformation matrix in modeling, simulation, and coordinate conversion of finger positions.

This work was supported by Natural Science Foundation of China (NSFC) (No. 61801262), the Fundamental Research Funds for the Central Universities (2242020R20010, 2242019K3DN06), and National undergraduate innovation and entrepreneurship training program (202010286023).

Pu-Fan Xu and Zi-Xuan Liu are with the School of Electronic Science and Engineering, Southeast University, 210096 Nanjing, China.

Fei Li is with the Institute of RF- & OE-ICs, Southeast University, 210096 Nanjing, China.

Hai-Peng Wang is with the School of Electronic and Information Engineering, Sanjiang University, 210012 Nanjing, China.

The corresponding authors of this paper are Fei Li (e-mail: lifei@seu.edu.cn) and Hai-Peng Wang (e-mail: wanghaipeng4213@126.com).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

II. SYSTEM AND METHODS

A. Hardware Design

1) Hardware investigation and overview

As a portable, efficient, and powerful sensor device, IMU can provide a better performance in evaluating hand gestures of the subjects by capturing the angle and acceleration information. The chosen IMU sensor units can faithfully reflect the hand dynamics without limitations of other sensors. It should be noted that a low-power consumption microcontroller is used to realize the multi-channel rapid data reception and transmission wirelessly.

The hand kinematic measuring platform of the intelligent low-cost wearable hand gesture detecting system was designed to have a distributed architecture, including three modules (Fig. 1): a distributed multi-node IMU array placed on each finger joint, a STM32 data processing module with wireless transceiver, and a remote processing terminal (personal computer).

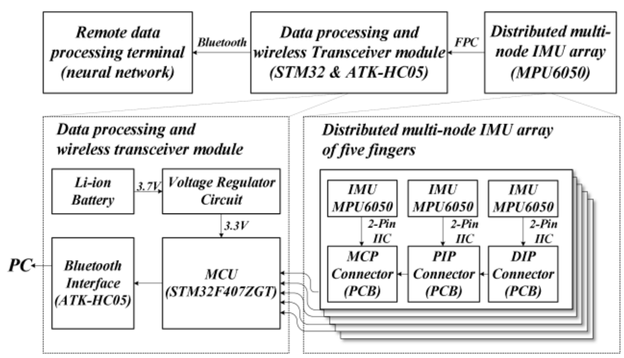


Figure 1. Block diagram of the proposed prototype

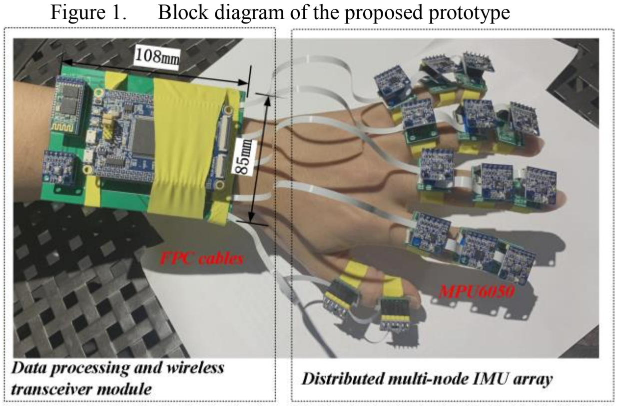


Figure 2. Photograph of the whole prototype that is divided into two parts corresponding to the two blocks in Fig 1 respectively.

2) Distributed multi-node IMU array

The sensor array mainly consists of fifteen integrated 6-axis motion tracking device (MPU6050) that combines a 3-axis gyroscope, 3-axis accelerometer, and a Digital Motion Processor™ (DMP) all in a $4\times4\times0.9\ \text{mm}^3$ package. Through Inter-Integrated Circuit (IIC) interface, MPUs output angles in the form of Euler angles. The outputs are processed by an on-board Motion Fusion™ which performs well in solving time drift through fusing gyroscope and accelerometer data.

Considering the number of wires and the high flexibility required by the system, nodes are linked with customized flexible printed circuit (FPC) cables, which is presented in the right dotted box in Fig. 2. The array can acquire distal interphalangeal (DIP), proximal interphalangeal (PIP) and metacarpal (MCP) joints angles of five fingers. To strengthen usability and simplicity of the device, a connector based on printed circuit board (PCB) is used to link the different number of pins at each joint (4-pin in DIP, 6-pin in PIP and 8-pin in MCP).

3) Data processing and wireless transceiver module

The module mainly includes two parts: a MCU (STM32F407ZGT) and a master-slave integrated Bluetooth module (ATK-HC05). Then the MCU acquires data from sensor array and transmits them to the processing terminal via Bluetooth interface.

B. Software Design

1) Hand gesture recognizer based on CNN

A CNN-based model is chosen as the hand gesture recognizer and built in PyCharm 2020.2 (Community Edition). The raw data are processed offline using MATLAB R2018b. Initially, invalid data are removed from the whole dataset when the value is zero in order to eliminate *no motion* data. Subsequently, the data segmentations are performed using 192-samples windows with increments of 50 samples. If the prepared data are denoted as $\forall X=\{x_1,x_2,\ldots,x_T\}$, the size of each subsegment x_t, $\forall t\in\{1,2,\ldots,T\}$, is $N\times192$, where N is the number of IMUs placed on the measuring platform. The corresponding gesture labels can be symbolized as $Y=\{y_1,y_2,\ldots,y_T\}$. The processed data are randomly shuffled and labels are standard scaled in order to train the recognizer. Eighty percent of the data are selected randomly as the training set and the remaining data are used for validation. The output gesture labels will be denoted as $\hat{Y}=\{\hat{y}_1,\hat{y}_2,\ldots,\hat{y}_T\}$.

The CNN used in the prototype is based on Alexnet [8] and tested for better performance on hand kinematic data empirically. The exact configuration of the CNN recognizer is shown in Fig. 3. There are five convolution layers in the model. The number of filter banks are 16, 32, 64, 64, and 16, respectively, and all the sizes of filter banks are 3. Same-padding is used for the input data for each convolution layer and activation function of all convolution layers is rectified linear unit (ReLU). To accelerate training, a batch normalization (BN) layer was added after the convolution. The pooling size is 2 and the number of moving rows is also 2 in the four max-pooling layers. The last BN layer is followed by a flatten layer and two fully connected (FC) layers. The second FC layer uses SoftMax as an activation function and is also the output layer with thirteen nodes corresponding to the number of the hand gestures.

The stochastic gradient descent with momentum (SGDM) algorithm is chosen as the training algorithm for the network with an initial learning rate of 0.001 and a momentum of 0.9. The loss function of the proposed CNN is cross-entropy loss:

$$L=-\left[y\log\hat{y}+(1-y)\log(1-\hat{y})\right] \tag{1}$$

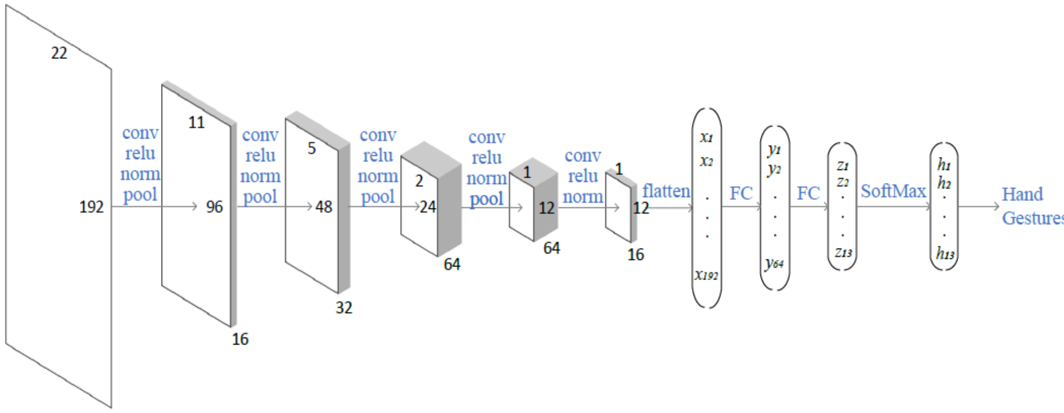


Figure 3 Architecture of the convolutional neural network (CNN).

where L is the loss function, y is the original label, and $\hat{y}$ is the result of the recognition. In addition, the minibatch size is 128 and the number of epochs is 60. To decrease overfitting, the dropout regularization with a rate set to 0.5 and the L2 regularization with a factor of 0.0001 are used.

2) Finger position calculator and simulator based on modified Denavit-Hartenberg (D-H) notation

Because the position of finger tips and the reconstruction of fingers are also valuable factors for hand kinematic assessment, finger positions are calculated and simulated at the same time using modified D-H notation. Modeling, calculation, and simulation for one finger are achieved using the data from the measuring platform mentioned above. A GUI program is written to display the dynamic movements and output the kinematics of one finger. All the procedures are achieved using MATLAB R2018b. In this section, the modeling method of the finger will be first mentioned, followed by the explanation of data processing steps.

Based on the specific explanation of modified D-H notation in [7], one finger can be modeled as an open-loop robot (while a hand must be modeled as a tree-structure robot) and the notation is shown in Fig. 4.

The parameters of one finger are shown in TABLE I.

TABLE I. THE PARAMETERS OF ONE FINGER

i	Parameters			
	a_{i-1}	φ_{i-1}	d_i	θ_i
1	0	φ_0	0	θ_1
2	a_1	0	0	θ_2
3	a_2	0	0	θ_3
4	a_3	0	0	0

a_{i-1} is the distance between O_{i-1} and z_i ; φ_{i-1} is the angle between z_{i-1} and z_i about x_{i-1} ; d_i :is the distance between O_i and x_{i-1} ; θ_i is the angle between x_{i-1} and x_i about z_i . Specifically, a_{i-1} is determined by the finger length and θ_i is calculated from the measuring platform data. a_0 is set to be zero to indicate the rotation of $Joint_0$. φ_0 can indicate the finger-tip absolute orientation.

The transformation matrix from $Frame_{i-1}$ to $Frame_i$ is ${}^{i}_{i-1}T$. The geometric model of the finger can be obtained by the successive multiplications of the transformation matrices:

$$ {}^{n}_{0}T = {}^{1}_{0}T\, {}^{2}_{1}T \cdots {}^{n-1}_{n-2}T\, {}^{n}_{n-1}T \qquad (2) $$

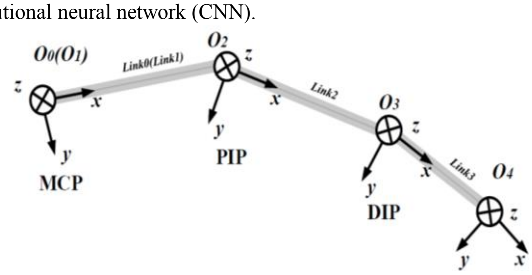


Figure 4. The model of one finger. A finger can be symbolized by four links, where $Link_0$ is the fixed base and $Link_3$ is the terminal link. $Joint_i$ connects $Link_{i-1}$ and $Link_i$. O_i is the fixed frame with respect to $Link_i$. z_i is the axis of $Joint_i$. x_i is defined on the common perpendicular of z_i and z_{i+1} .

The date transmitted from the measuring platform take the form of hexadecimal complement, consisting of roll angles, pitch angles and yaw angles of IMUs. All the angles are relative to the base. Data will be converted to decimal form at first, followed by calculating the differences of different IMU angles. This is aimed at gaining the relative angles, corresponding to the parameters of fingers mentioned above. In addition, the processed data are divided into groups and mean values are calculated in order to improve the accuracy and smoothness. Using Robotics Toolbox (RTB) for MATLAB, a finger can be modeled conveniently using the parameters. The transformation matric ${}^{n}_{0}T$ may be computed using fkine() method and kinematics, such as spatial coordinates and ZYZ Euler angles, are acquired with the help of RTB afterward.

III. EXPERIMENTS AND RESULTS

A. Data acquisition

In this study, the original angle data of hand kinematics is obtained by inertial sensor array (signals are sampled at 50Hz). The sensor collects the three-axis acceleration and angular velocity information at the same time, and then calculates the corrected three-axis Euler angle.

In order to test the classification accuracies of the CNN-based recognizer, the performance of the model is evaluated using the Ninapro database [9]. The first sub-database is named as DB1, containing a total of 53 gestures from 27 subjects. Thirteen finger gestures (including rest gesture) from Exercise A are chosen as the experiment data. In NinaproDB1, hand kinematics are captured by 22 sensors and represented by joint angles.

B. Results

1) Hand gesture recognition

To find a better CNN model, hyperparameters including decay rates and Nesterov momentum are optimized in the experiment. The decay rates are set to be 0, 0.001, 0.00001 and the Nesterov momentum is chosen to be used or not. With different combinations of these two factors, the best-performed set will be used in the final recognizer.

The classification accuracies of various parameters are shown in Table II. Since the raw data could be classified by picking one of 53 gestures randomly, the chance level accuracy is 1.89%. The optimal selection of parameter combination in this test is using Nesterov momentum without any learning rate decay, whose classification accuracy is 98.79%. Fig. 5 illustrates the changing process of loss and accuracy for both training and validation of the best set. The loss and accuracy curves are smooth without fluctuation, and close enough for the training and validation sets, illustrating the CNN model is suitable for the data of fingers can is well-trained.

TABLE II. CLASSIFICATION ACCURACIES OF VARIOUS PARAMETERS

Nesterov Momentum	Delay Rate		
	0	***0.001***	***0.00001***
Using	98.79%	97.76%	95.34%
Not Using	97.95%	95.53%	96.19%

2) Finger position calculation and simulation

Table III shows the spatial coordinates and ZYZ Euler angles of the fingertip (O_4) relative to the base frame (O_0) partly. The D-H notation is able to calculate the orientations of the finger tips conveniently and other kinetic results depending on the requirement. Fig .6a-6c display the dynamic movements, which are excerpted of the finger bending movement. Combining the accurate kinematic data and the visual simulations, evaluation of finger position can be made both accurately and intuitionally.

TABLE III. FINGER KINEMATICS

Spatial Orientations			Euler Angles (°)		
x	***y***	*z*	***R***	***P***	***Y***
5.00	0	0.01	–90.00	90.00	89.58
4.97	0	–0.24	–90.00	90.00	76.38
4.88	0	–0.52	–90.00	90.00	61.97
4.71	0	–0.82	–90.00	90.00	45.04
4.29	0	–1.38	–90.00	90.00	20.77
4.03	0	–1.78	–90.00	90.00	13.17

IV. DISCUSSION AND CONCLUSION

In this paper, a low-cost wearable IMU prototype which can recognize hand gestures and reconstruct finger positions is proposed. Combined with CNN model and modified D-H notation, the system can provide more reference criteria of hand rehabilitation evaluation, including hand kinematic information and classifications of hand gestures, which are usually analyzed separately. The prototype will be an important tool for reflecting the efficacy of treatments, remote monitoring system and development of somatosensory games. Meanwhile, more characteristic tests can be applied to improve the performance of the system, such as sensor resistance stability and repeatability. Further clinical research and market research are also planned to collect more data to improve the feasibility of the prototype.

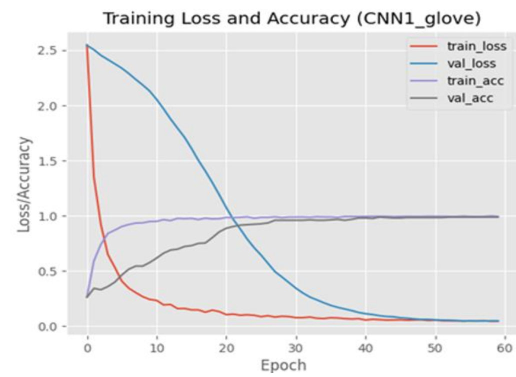
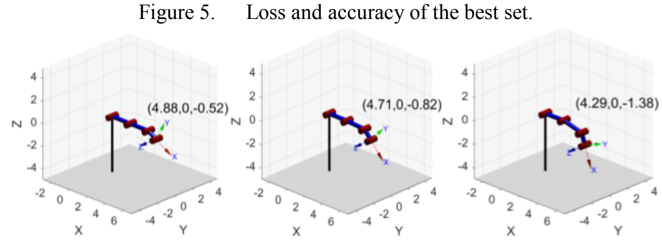


Figure 5. Loss and accuracy of the best set.



(a) (b) (c)

Figure 6. Dynamic movements in the GUI program.

REFERENCES

[1] N. P. Oess, J. Wanek, and A. Curt, "Design and evaluation of a low-cost instrumented glove for hand function assessment," *Journal of NeuroEngineering and Rehabilitation*, vol. 9, pp. 1-2, 2012.

[2] L. K. Simone and D. G. Kamper. "Design considerations for a wearable monitor to measure finger posture," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, pp. 2-4, 2005.

[3] R. Gentner and J. Classen, "Development, evaluation of a low cost sensor glove for assessment of human finger movements in neurophysiological settings," *Journal of Neuroscience Methods*, vol. 11, pp. 4-8, 2008.

[4] J. H. Seong and Y. Choi, "Design and Implementation of User Interface through Hand Movement Tracking and Gesture Recognition," in *International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 553-555, 2018.

[5] H. G. Kortier, V. I. Sluiter, and D. Roetenberg, "Assessment of hand kinematics using inertial and magnetic sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, pp. 4-10, 2014.

[6] A. R. Asif, A. Waris, S. O. Gilani, M. Jamil, H. Ashraf, M. Shafique, and I. K. Niazi, "Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG," *Sensors*, vol. 20, pp. 2-4, 2020.

[7] W. Khrcalil and J. F. Kleinfinger, "A new geometric notation for open and closed-loop robots," *IEEE International Conference on Robotics & Automation IEEE*, pp. 1174-1179, 1986.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, pp. 1098-1102, 2012.

[9] M. Atzori, A. Gijsberts, and C. Castellini, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific Data*, vol. 1, pp. 2-7, 2014.