Utilizing LSTM Models for IMU-Based Hand Gesture Classification

Guotao Gou, Yiheng Zhang, Chenhan Zhang, Ziwei Huang, Dekai Feng, Yuanzhuo Zhang

***Abstract*—This course project explores the application of Inertial Measurement Unit (IMU) technology in predicting and classifying multiple hand gestures, with a focus on enhancing the user experience for wearable robotic devices. By employing advanced signal processing and machine learning techniques, we have developed a system capable of accurately interpreting various grasping intentions. The project leverages an LSTM (Long Short-Term Memory) model to process IMU data collected from participants, achieving a high level of accuracy in gesture recognition. The outcomes of this project are expected to contribute to the development of more intuitive and responsive wearable technologies, improving the interaction between users and their devices.**

***Keywords*—Intent Recognition, Long Short-term Memory (LSTM), Inertial Measurement Unit (IMU).**

I. Introduction

In the realm of human-computer interaction, the ability to control devices with natural hand gestures has the potential to revolutionize the way we engage with technology. The advent of wearable robotic devices has brought this possibility closer to reality, yet the challenge remains to create a system that can accurately interpret the user's intentions from their hand movements. This course project aims to address this challenge by utilizing IMU technology to predict and classify a range of hand gestures [1], [2].

The project's foundation lies in the IM900 IMU, a state-of-the-art sensing device that provides rich motion data with high temporal resolution. By strategically positioning the IMU on the forearm, we ensure that the collected data accurately reflects the user's arm movements without interference from hand motion. The data is then processed to calculate the trajectory and orientation of the arm, which are critical features for gesture classification.

Our approach to gesture recognition involves training an LSTM model, a type of deep learning model particularly suited for handling sequential data. The model is trained on a dataset where each gesture is labeled and represented by its corresponding arm movement features. Through rigorous testing and validation, the LSTM model has demonstrated a high degree of accuracy in classifying the gestures, even when faced with new, unseen data [3], [4]

The success of this project holds significant implications for the future of wearable technology. By improving the accuracy and responsiveness of gesture recognition, we can create devices that are more intuitive and user-friendly. This not only enhances the user experience but also opens up new possibilities for applications in various fields, from gaming and virtual reality to assistive technologies for individuals with motor impairments.

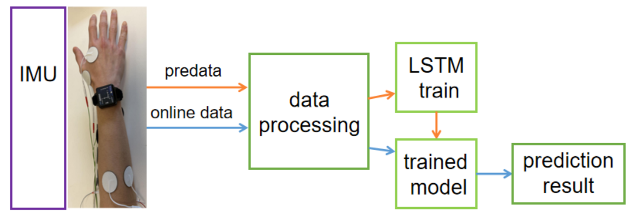

Fig. 1. General process of collecting IMU data to train LSTM model and predict the intention result.

In conclusion, this course project represents a step forward in the field of human-computer interaction, showcasing the potential of IMU technology and machine learning in creating more natural and efficient ways for users to interact with their devices [5], [6].

II. Material and Methods

A. System Hardware

In this study, we introduce the IM900, a state-of-the-art sensing device with robust capabilities, designed to interface seamlessly with personal PCs via BLE 5.0 Bluetooth technology. The IM900 boasts an impressive 8dbm transmission power and an extended communication range of up to 60 meters, ensuring reliable data transfer even over substantial distances. The device's reporting frequency is highly versatile, ranging from 0.5 Hz to 250 Hz; for the purpose of our experiment, we have selected a 30 Hz reporting frequency to optimize the balance between temporal resolution and data volume.

Our research involved the collection of comprehensive motion data sets using the IM900. This included tri-axial acceleration data, with a full-scale range of ±16g, allowing for the capture of both subtle and dynamic movements. Additionally, tri-axial angular velocity measurements were recorded, with a capacity of ±2000°/s, to accurately document the rotational aspects of limb motion. The inclusion of a tri-axial magnetometer, primarily utilized for automatic calibration, further enriches the data, ensuring high fidelity in the representation of the arm's orientation in space.

The IM900 was strategically positioned at the proximal end of the forearm to ensure unobstructed data collection. This placement was critical in guaranteeing that the device's readings were solely reflective of the arm's motion, without interference from the hand's movements.

B. Intent Recognition

1) Data Processing: The absolute acceleration data obtained from the IMU served as a reference for determining the initiation of arm movement, based on an analysis of experimental data. We established a threshold where an absolute acceleration greater than 0.5 m/s^2 indicated the start of movement, marking this as the starting point. Once movement commenced, we calculated the velocity in the x, y, and z directions by integrating the acceleration data from the starting point. A subsequent integration of these velocities provided us with the distances traveled in each direction, representing the trajectory of the arm. Similarly, by integrating the angular velocities in the x, y, and z directions from the starting point, we obtained the angles in each axis, representing the orientation of the arm. Such processing of the input data imbues it with practical physical significance, enhancing the accuracy of classification and making the data more comprehensible.

2) Deep Learning Process: We obtain the distances and angles relative to the starting point in the x, y, and z axes for each sample. These represent the trajectory and orientation of the arm, respectively. And the data are padded or truncated to make sure they have the same length. After that, they are inputted as features. Different labels are assigned to the data sets corresponding to different gesture intentions. In our case, gesture draw 1 are labeled as 0, gesture draw 2 are labeled as 1 and gesture draw 3 are labeled as 2. Fifty percent of the total data is used for testing, while the other fifty percent is used for training the model. And the model we use is LSTM (Long Short-term Memory) model.

LSTM model is a deep learning model suitable for classifying sequential data. In this study, we implemented a dual-layer Long Short-Term Memory (LSTM) network to handle time series prediction tasks. The model consists of an input layer that takes sequences of shape $(\text{length\_of\_data}, 6)$, where length_of_data is the sequence length and 6 is the number of features (trajectory and orientation in x, y and z plane). The first LSTM layer has 100 units and returns sequences to maintain temporal dependencies, followed by a dropout layer with a dropout rate of 0.2 to reduce overfitting. The second LSTM layer also has 100 units but does not return sequences, preparing the model to finalize output processing. This is followed by another dropout layer with the same dropout rate. The output layer is a densely connected layer with a softmax activation function, used to output the probability distribution over the target classes. The model uses L2 regularization with a coefficient of 0.001 in the output layer to prevent coefficient inflation. It is compiled with the Adam optimizer and categorical crossentropy loss, and it measures accuracy as a performance metric. The model is trained with a batch size of 10 for 100 epochs, using 50% of the data for testing to validate the model. Upon training completion, the model achieved a test accuracy of 97%. Finally, the model is saved in HDF5 format for subsequent use. This setup highlights the model's robustness in handling overfitting through dropout and L2 regularization, ensuring generalizability on unseen data.

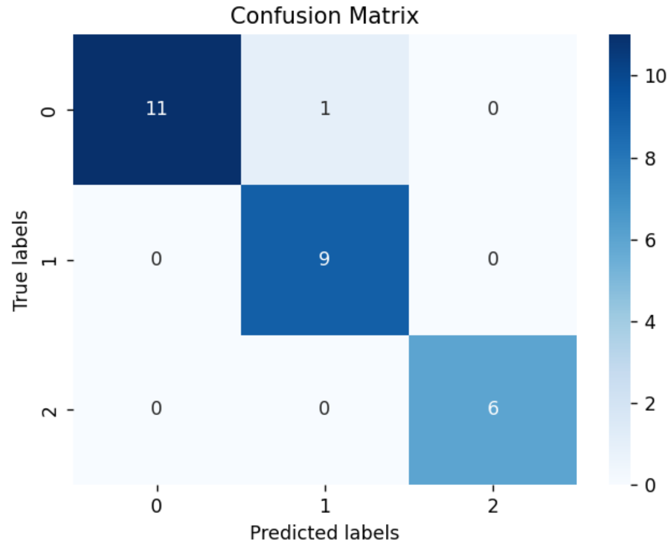

Fig. 2. Confusion matrix of gestures 1, 2 and 3 using LSTM model.

III. Results and Discussion

A. Data Collection

In current offline experiment part, a total of three participants were involved in the collection of 54 data sets. These data sets were evenly split to support both the training and testing of our deep learning model. Three distinct grasping intentions were carefully designed to capture the participants' actions when grasping different objects. The data collection for each set lasted approximately 3 seconds, encompassing the initiation of data capture, the hand gesture movement, and the final termination of the capture, with each set comprising around 90 data points.

Participants took a 10-second rest after each data set to prepare for the next one. Upon completing all the data sets for one grasping intention, they rested for 3 minutes before proceeding to the next grasping action data collection. The total data collection process was estimated to last about 35 minutes, taking into account the collection and rest times for all participants.

B. Classification Results

Figure 2 show confusion matrix of gestures 1, 2 and 3 using LSTM model.The labels 0, 1 and 2 represent gestures 1, 2, and 3, respectively.The training set and test set of the LSTM model are obtained by collecting the writing gestures 1,2 and 3 of three testers wearing IMU, and the sample size ratio of the test set and training set is 1:1.

From the table, it can be found that 11 of the 12 samples with a true gesture of 1 were correctly predicted by the model LSTM, and one was misjudged as a gesture of 2. The reason for the misjudgment may be that the features of gesture 1 are similar to those of gesture 2, or the subject's hand is unstable when writing gesture 1, and the accuracy of the prediction is still very high.In addition, the 9 samples with a real gesture of 2 and the 6 samples with a real gesture of 3 were all correctly judged by the model LSTM. In general, it can be found that the classification effect of model LSTM is very good through confusion matrix.

IV. Conclusion

Assistive devices to help people with disabilities to perform grasping manoeuvres are an effective way to help people with disabilities to increase their self-care ability in daily life.Therefore, we constructed models for recognising the arm movements of disabled people and controlling the assistive devices by using the arm movements as commands.

We chose a single object for training data and multiple objects for testing the model. Although the amount of training data for our model is small, the accuracy for the test data is 96.3%, which shows that the model out by this method has excellent generality.

With our model, we are able to accurately recognise arm movements and send the correct signals to the robotic arm. In the future, we hope to combine this model with robotic arm grasping to send commands through arm movements to control the robotic arm to grasp the corresponding objects, in order to achieve assisting people with disabilities to perform grasping movements completely.

References

[1] S. C. Colachis IV, M. A. Bockbrader, M. Zhang, D. A. Friedenberg, N. V. Annetta, M. A. Schwemmer, N. D. Skomrock, W. J. Mysiw, A. R. Rezai, H. S. Bresler *et al.*, "Dexterous control of seven functional hand movements using cortically-controlled transcutaneous muscle stimulation in a person with tetraplegia," *Frontiers in Neuroscience*, vol. 12, p. 208, 2018.

[2] D. A. Friedenberg, M. A. Schwemmer, A. J. Landgraf, N. V. Annetta, M. A. Bockbrader, C. E. Bouton, M. Zhang, A. R. Rezai, W. J. Mysiw, H. S. Bresler *et al.*, "Neuroprosthetic-enabled control of graded arm muscle contraction in a paralyzed human," *Scientific reports*, vol. 7, no. 1, p. 8386, 2017.

[3] Y. Wang, B. Metcalfe, Y. Zhao, and D. Zhang, "An assistive system for upper limb motion combining functional electrical stimulation and robotic exoskeleton," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 2, pp. 260–268, 2020.

[4] N. Bhagat, K. King, R. Ramdeo, A. Stein, and C. Bouton, "Determining grasp selection from arm trajectories via deep learning to enable functional hand movement in tetraplegia," *Bioelectronic Medicine*, vol. 6, pp. 1–8, 2020.

[5] Y. Zheng and X. Hu, "Elicited finger and wrist extension through transcutaneous radial nerve stimulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 9, pp. 1875–1882, 2019.

[6] D. Yang, Q. Huang, Z. Jiang, and L. Jiang, "Design of multi-channel electrical stimulator integrated with online impedance measurement," *Journal of Medical and Biological Engineering*, vol. 40, pp. 943–950, 2020.