Handwritten Numerical Character Recognition using LSTM based on Dual Leap Motion Controllers

Noriaki KANEKO[†a)], *Non-member*, **Masakatsu OGAWA**[†], ***Senior member***

SUMMARY. With the increasing infection of the new coronavirus (COVID-19), people expect to introduce the contactless operation rather than the screen touch operation. To address this demand, we focused on the numerical input with contactless. Sine the hand tracking sensor, known as the product name of Leap Motion controller, has the tracking function of the movement of fingers and hands, it is useful for contactless operation. However, it is difficult for most people to communicate the numerical number by their own hands, because they do not know the sign language for hearing-impaired people. Therefore, it is necessary to recognize by the written characters used in daily life. Since the tracking of hand and finger movements differ depending on the location of the Leap Motion controller, we propose to place two Leap Motion controllers at different positions to recognize handwritten characters. We apply the 10 types from 0 to 9 as the handwritten characters and classify these characters by machine learning algorithms such as LSTM. As a result of the classification accuracy, applying the bidirectional LSTM classification achieves the maximum, 94.7%.
keywords: *Contactless Operation, Sign Language, Handwritten Characters, Hand Tracking Sensor, Leap Motion controller, LSTM*

1. Introduction

Demand for contactless operations is rapidly increasing to prevent the spread of the new coronavirus (COVID-19). Devices for contactless operation are being developed year by year, and there are many studies using them. In the conventional study aimed at popularizing contactless devices, the authors have developed the new contactless input interface using the hand tracking sensor known the product name as Leap Motion controller **[1]**. However, from the results of experiments on input accuracy and usability conducted in the same study, it is mentioned that contactless input interfaces that require complicated operations are unacceptable to people. Therefore, we have been researching a method for intuitively inputting information using a contactless device. In our previous study, we used the three-dimensional data (coordinates and vectors related to both hands and 10 fingers) obtained by Leap Motion controller to classify numerical sign language and handwritten numerical characters using algorithms such as the linear support vector machine (SVM) and Random Forest [2]. The classification accuracy of the numerical sign language for hearing-impaired people was high because of the stationary state, but the accuracy of the handwritten numerical characters was low because the hands and fingers are moving. We also applied Long Short-Term Memory (LSTM) and Bidirectional LSTM [3]. The feature for the machine learning in [2] is the statistics, but [3] is the time-series data. As a result, the classification accuracy was improved to about 93%.

In the conventional study [4], American sign language, which is a dynamic gesture, was classified by Bidirectional LSTM using 24 types of features obtained by the Leap Motion controller. This study improved the classification accuracy compared with the related studies. Still, because of the instability of the dataset, the author stated that a better measurement method had to be considered for classifying more complex dynamic gestures.

From these analysis results, we considered it difficult to measure written character perfectly with a single sensor due to the limited sensing angle. In this study, we propose to place dual Leap Motion controllers at different positions to recognize handwritten numerical characters and compare the accuracy with single Leap Motion controllers to confirm the effectiveness of the proposed method.

2. Leap Motion controller

Leap Motion controller is a device developed by ultraleap. It is a small contactless hand tracking sensor that can capture and record the movement of the subject's hand with an infrared camera and CMOS image sensor [5]. The size of the main body is 80 mm in width, 30 mm in depth, 11 mm in height, and weighs 45 g, and it is tiny and lightweight. The sensor tracks only 10 fingers and both hands and detects the range of the space from 25 mm to 600 mm above the device and the viewing angle of about 150 degrees, measures at a maximum of 100 fps. The sensor measures many data types, such as the coordinates of the center and fingertips of the hand, the unit vector, the moving speed, and the angle in the coordinate space constructed based on Leap Motion controller.

Furthermore, it is possible to output a 3D model of the subject's hand predicted from the acquired data in real-time shown in Fig.1. Leap Motion controller takes a picture with the built-in infrared camera. It converts the two-dimensional data of the subject's hand acquired by image processing into three-dimensional data in a virtual coordinate space. Therefore, if there is an obstacle between the hand and Leap Motion controller, or if the hand's position is inappropriate, the accuracy may significantly decrease. Therefore, a measurer should keep the hand and all fingers within the measurement range recommended by Leap Motion controller.

[†]The author is with Sophia University
[a)] E-mail: Kane0423@gmail.com

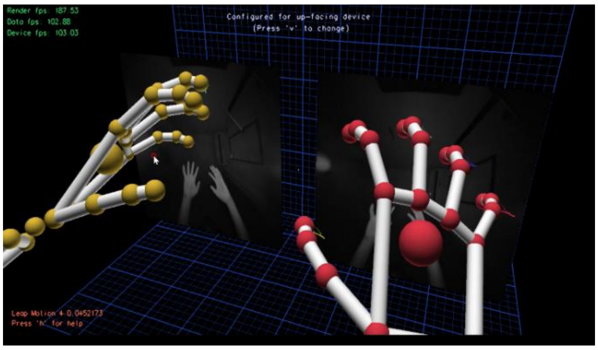

Fig. 1 3Drendering by Leap Motion controllers

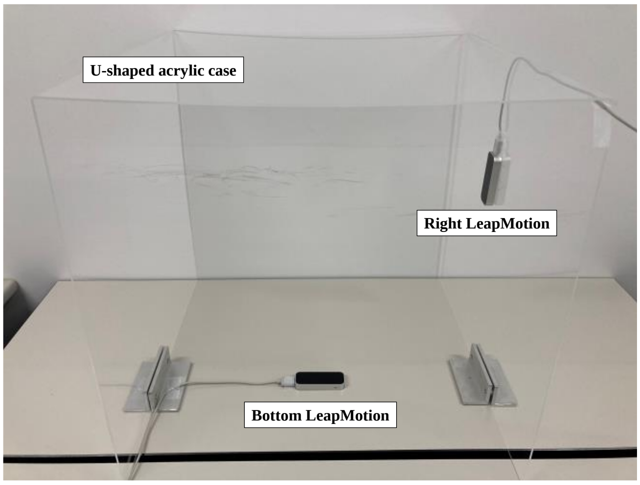

Fig. 2 U-shaped acrylic case

3. Measurement

One of the use cases of this study is the contactless input operation of an automatic teller machine (ATM) in banks. In this condition, the operation is in a predetermined space. To imitate this condition, we use a U-shaped acrylic case with internal depth dimensions: 500 mm x width: 500 mm x height: 500 mm shown in Fig.2. Additionally, to increase sensing angle by Leap Motion controller, we use dual controllers located at the center position of the bottom and the right side. We use two PCs with a connected Leap Motion controller to each. These PCs access Network Time Protocol (NTP) server for synchronization. The effect of using the U-shaped acrylic case is to limit the hand movement according to the sensing area of Leap Motion controllers. We assume a dynamic gesture that indicates the Arabic numerals "0" to "9" with the index finger as handwritten numerical characters. The reason for choosing Arabic numerals is that people are familiar with them and can understand them intuitively. In the measurement, the shape of the hand is fixed with only the index finger upright, and each number is expressed from the trajectory of the tip of the index finger.

For each number, measure the time series data 30 times at 40 fps. The data acquired by each Leap Motion are the coordinates of the center of the palm, the coordinates of each fingertip, the three-dimensional data of the vector from the center of the palm to each finger, and epoch time for synchronization. The time-series length of the data acquired by two controllers is 120.

Algorithm

```
for ( i = 0 ; i < BottomDataLength ; i++ )
    {
    X = 10000
    for ( j = 0 ; i < RightDataLength ; i++ )
        {
        if (
                X > Absolute( BottomTimestamps[i] – RightTimestamps[j] )
           )
           X = Absolute( BottomTimestamps[i] – RightTimestamps[j] )
           Register = j

    Combine BottomDataRow[i] with Right DataRow[Register]

        }
}
```

Fig. 3 Algorithm of data combining

Fig.3 shows the algorithms of data combination for synchronization based on measured time, i.e., epoch time. BottomDataLength and RightDataLength are the length of time series data in the bottom and right sidecontrollers, respectively. BottomTimestamps and RightTimestamps are the epoch time, BottomDataRow and RightDataRow are the data when the absolute difference in the epoch time is the minimum. The algorithm searches the data when the absolute difference is the minimum using the variable *X*.

4. Evaluation

For evaluation, we use three data set: measured data at single Leap Motion (right and bottom side) and measured data at dual Leap Motion controllers. Using these data sets, we compare the classification accuracy. We apply seven machine learning algorithms, SVM, Random Forest, Decision Tree, Extra Tree, Gaussian NB, LSTM, and Bidirectional LSTM, and evaluate the accuracy by 6-fold cross-validation. The input data of LSTM and Bidirectional LSTM consist of 11 types of features: the standardized coordinates of the center of the palm, the coordinates of each fingertip, and the vector from the center of the palm to each finger measured by the two Leap Motion controllers. Since the coordinates and vectors are composed of three components, x, y, and z, the actual number of features of the input data in classification by a single controller is 33, which comes from a calculation result of 11x 3. In the classification by dual controllers, the number of features is doubled to 66. In other machine learning algorithms, in addition to the variance of each coordinate, the standardized vector of average and variance, which are the statistics of the time series data acquired as features, are applied.

Figure 4 shows the structure of the LSTM learning model for evaluating data measured at dual controllers. LSTM is a kind of deep learning that improves Recurrent Neural Network (RNN), and in recent years, it has been attracting attention as a learning method for natural language processing and time-series data. RNN incorporates a loop-shaped neural network called cell, which enables effective learning for data with time-series changes. However, in RNNs, there was a concern about the vanishing gradient problem in which the gradient of the loss function becomes zero during repeated error backpropagation, and learning does not proceed. LSTM solves the vanishing gradient problem by using three mechanisms, forgetting gate, input gate, and output gate, and enables a more detailed analysis of long-term time-series data.

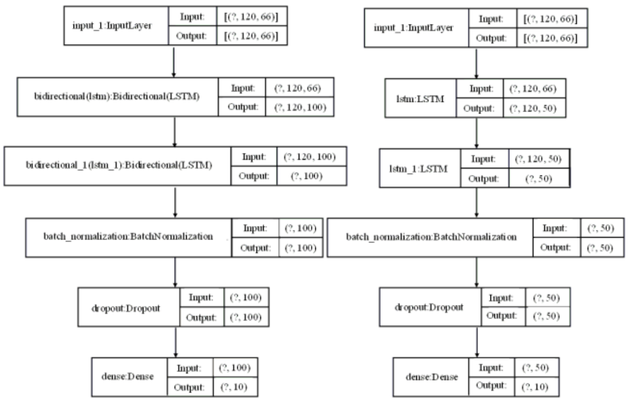

Fig. 4 Learning model of LSTM and Bidirectional LSTM.

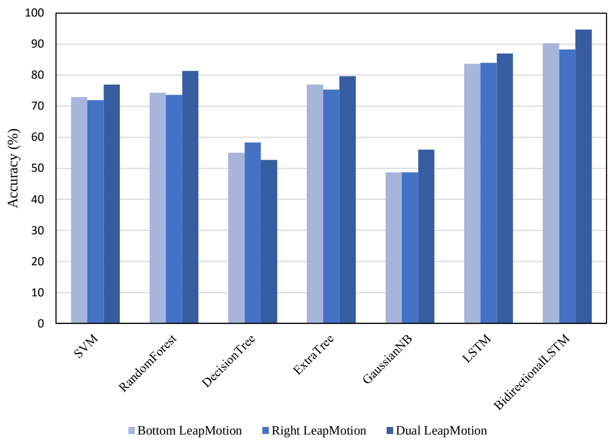

Fig. 5 Accuracy of each machine learning algorithm

Unlike LSTM, Bidirectional LSTM has a structure that performs recursive learning in both directions of time-series data. Using Bidirectional LSTM, we thought it is an effective machine learning method for classification handwritten numerical characters with a strong context by changing continuously.

5. Result and Discussion

Figure 5 shows the classification accuracy of each machine learning algorithm. Most machine learning algorithms in dual Leap Motion controllers have improved accuracy compared with a single controller.

Among the five machine learning algorithms excluding L STM and bidirectional LSTM, the accuracy with Random Forest was highest, 81.3%. Both LSTM and Bidirectional LSTM have higher accuracy than other machine learning algorithms. In particular, the accuracy with bidirectional LSTM using only the bottom Leap Motion achieves 90.33%. When using dual controllers, the accuracy improves to 94.67%, which achieves the highest accuracy in this study.

The accuracy of using dual controllers reduces only in the Decision Tree. In Gaussian NB, the accuracy in the single controller is the lowest accuracy among them, but using dual controllers, improves the accuracy. Even if using dual controllers, the degree of improvement depends on the accuracy of a single controller. This means that the accuracy remains low with dual controllers.

Table 1 summarizes the accuracy with bidirectional LSTM for each handwritten numerical character. It can be seen that

Table 1 Accuracy in Bidirectional LSTM

Labels	Bottom LeapMotion	Right LeapMotion	Dual LeapMotion
0	96.67	90	100
1	76.67	86.67	100
2	93.33	93.33	86.67
3	90	90	90
4	90	86.67	93.33
5	90	90	100
6	90	86.67	90
7	83.33	83.33	90
8	96.67	86.67	96.67
9	96.67	90	100

(%)

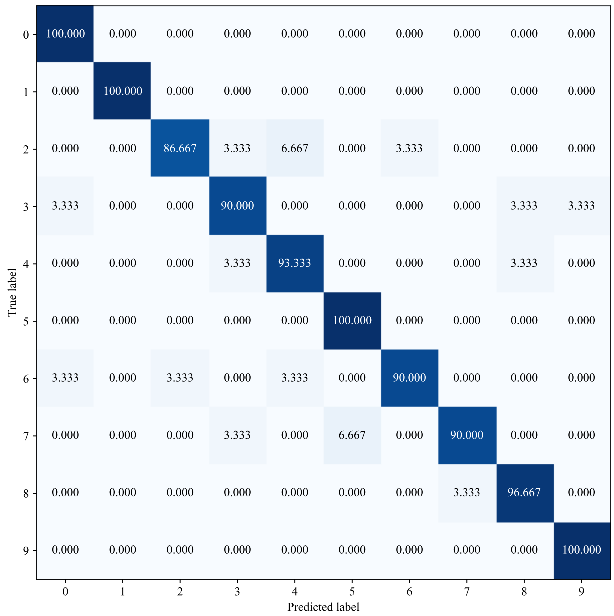

Fig. 6 Confusion Matrix of dual Leap Motion controllers in bidirectional LSTM

the accuracy improves in almost all labels when using dual controllers. Furthermore, the accuracy of the handwritten numerical characters "0", "1", "5", and "9" has reached 100% when using dual controllers. The results show that since the dual controllers increase the sensing angle, the proposed method achieves higher accuracy than a single controller.

Figure 6 shows the confusion matrix of bidirectional LSTM using dual controllers, which was the highest accuracy in this study. The figure shows that the bidirectional LSTM makes incorrect predictions for some data, and the incorrect prediction results do not have uniformity or reciprocity. We consider that the incorrect prediction in the confusion matrix was caused by Some unstable data about subtle movements of the hand-measured by using the controller.

In the previous study [4], an algorithm was used to properly end the measurement by conditional branching based on the features related to the speed of the fingertip that can be acquired by Leap Motion. We think building an appropriate algorithm give data stability.

Other challenging issue to further improve accuracy is a proposal of a new model in deep learning.

6. Conclusion

In this study, we focused on the contactless input without touching the screen. Our objective is the classification of numerical written characters that almost people use daily. To improve the classification accuracy subject to handwritten numerical characters from 0 to 9, we proposed the classification method of the characters with the dual hand-tracking sensor, known as the product name of Leap Motion. By using dual Leap Motion controllers, the accuracy of all machine learning algorithms except Decision Tree improved compared with a single controller. In particular, the bidirectional LSTM achieves the maximum accuracy of 94.7%. In this measurement, the number of subjects is only one. We want to increase the number of subjects in the future and evaluate it.

References

[1] K. Hosono, M. Sasakura, H. Tanabe, K. Takeshi, "A proposal of the character input method by gesture operation using Leap Motion." Proc. Annual Conference of the Japanese Society for Artificial Intelligence, vol. 28, pp. 1-4, 2014 (in Japanese).

[2] N. Kaneko, M. Ogawa, "Numerical Sing Language Identification by Machine Learning Using Leap Motion" Proc. IEICE Gen. Conf. '2021, B-15-14, 2021 (in Japanese).

[3] N. Kaneko, M. Ogawa, "Numerical Sing Language Identification by LSTM using Leap Motion based," Proc. IEICE Society Conf. '2021, B-15-52, 2021 (in Japanese).

[4] L. Yang, J. Chen, W. Zhu, "Dynamic Hand Gesture Recognition Based on a Leap Motion Controller Two-Layer Bidirectional Recurrent Neural Network," Sensors, Vol. 20, Issue7, 2106, 2020.

[5] ultraleap. "Motion/DEVELOPER API Overview," https://developerarchive.Leapmotion.com/documentation/csharp/devguide/Leap_Overview.html.