

Verification of the Usefulness of Personal Authentication with Aerial Input Numerals Using Leap Motion

Shun Yamamoto*, Momoyo Ito, Shin-ichi Ito, Minoru Fukumi

Tokushima University, Information Science and Intelligent Systems, 770-8506, Japan

ARTICLE INFO

Article history:

Received: 05 August, 2019

Accepted: 04 October, 2019

Online: 22 October, 2019

Keywords:

Aerial input numerals

Leap motion

Deep learning

CNN

Dilated convolution

Personal authentication

ABSTRACT

With the progress of IoT, everything is going to be connected to the network. It will bring us a lot of benefits however some security risks will be occurred by connecting network. To avoid such problems, it is indispensable to strengthen security more than now. We focus on personal authentication as one of the security.

As a security enhancement method, we proposed a method to carry out numeral identification and personal authentication using numerals written in the air with Leap motion sensor. In this paper, we also focus on proper handling of aerial input numerals to verify whether the numerals written in the air are helpful for authentication. We collect numerals 0 to 9 from five subjects, then apply three pre-processing to these data, learn and authenticate them by CNN (convolutional neural network) which is a method of machine learning. As a result of learning, an average authentication accuracy was 92.4%. This result suggests that numerals written in the air are possible to carry out personal authentication and it will be able to construct a better authentication system.

1. Introduction

With the progress of IoT, everything is going to be connected to the network. It will bring us a lot of benefits, however some security risks will be occurred by connecting network. To avoid such problems, It is indispensable to strengthen security more than now. As one of the security elements, We focused on personal authentication. As a current major personal authentication method, there are a method using card information and password or biometrics system using face, iris, vein and so on. Biometric authentication is possible with high accuracy, and it is difficult to duplicate authentication information. However, if biometric information is stolen, there is a big problems that it is difficult to change or the number of changes is limited. For method using card information and password, although there are some problems such as skimming and password leakage, it can be relatively easily dealt with by changing authentication information. Therefore, We considered that it is necessary to create a system that strengthens security by adding new elements to the existing authentication method using card information and password. Even if authentication information is stolen, this system can be easily recovered. As a method to realize this system, we propose a method input numerals in the air using Leap motion sensor and then perform numerals recognition and personal authentication from input handwriting.



Figure 1 : Leap motion sensor

For password authentication realization and further security enforcement, we proposed a method as following. First, numerals data are inputted using Leap motion. Then, these data are learned and identified with CNN which is one of deep learning methods. Moreover, we also assume that at the same time personal authentication will be carried out to enhance security. In this paper, we focus on personal authentication to verify whether the numerals written in the air have unique features between individual for authentication. We also focus on proper handling of aerial input numerals. Numerals written in the air have different features from numerals written in the paper. Therefore, we must try different pre-processing approaches to the data.

In the next subsection, we describe about related work to show the novelty in this paper.

* Shun Yamamoto, Email: c501837011@tokushima-u.ac.jp

1.1. Related work

First of all, this paper is an extended version of the work originally presented in 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS) [1]. In this work, we handled the data of three subjects measured by the Leap motion. Then we applied two pre-processings to these data. Finally they were learned by CNN. As a result, the average accuracy was 90.3%, False Rejection Rate(FRR) was 3.8% and False Acceptance Rate(FAR) was 5.9%. As an extension from this, in this paper, we carried out an addition of a pre-processing, increase of subjects, and change of the layer structure of CNN. By these extension, we could reach more effective way to handle the numerals written in the air, and we could confirm this layer structure change is useful for numerals which have fewer features between individual.

As a related work, there was a research conducted by Katagiri and Sugimura[2]. They used trajectory of writing in the air using a penlight and performed signature verification with DP matching on the data. As a result, FRR and FAR were both 3.6%. Moreover, they considered that it was difficult to duplicate the aerial input, which had individual features because the degree of freedom of input movement was large.

In the research by Hatanaka, et al. [3], they acquired signature data including finger identification information from 50 people by the Leap motion. Then, they also used DP matching to express a distance between features and carried out evaluation using the distance. Moreover, they experimented with or without screen feedback. When there were the feedback and no information of finger identification, the Equal Error Rate(EER) was 3.72%. When there was finger identification information, the EER was 2.67%. On the other hand, when there were no feedback and no finger identification information, the EER was 6.00%. Furthermore when there was information of finger identification, the EER was 5.11%. In response to this result, they considered that using of finger identification information could have a performance equivalent or superior to the related work [4]. However, in this research [3], subjects had to use several fingers when perform the measurement.

In the research by Nakanishi, et al. [5], they attempted to authentication of aerial handwritten signature only using information of fingertip coordinate without using the information of finger identification. They dealt with the time series data measured by the Leap motion and the subjects used only one finger for measurement. Then, another five people's forgery signature data were added to the kanji signature data for two kinds of people and the learning was done by CNN. As a result, the accuracy against two signatures were 97.0% and 95.9% respectively, FRR were 9.6% and 19.2%, and FAR were 0.8% and 0.1%. From these results, it indicated the possibility where the aerial input was useful even without using the information of finger identification. Assuming use in a real system, it is easy to use only one finger for users. Therefore, we attempt to do verification without the information of finger identification.

There is one another example using CNN and time series data [6]. In this paper, The feature extraction is carried out by CNN, as a conclusion, usefulness of using CNN for multivariate time series had been shown.

These researches indicate the usefulness of the aerial input for identification and authentication. These researches [2],[3],[4] dealt with hiragana or kanji in Japanese. We try to authenticate numerals in this paper. Numerals have relatively simpler shapes than hiragana or kanji. Therefore numerals data may not have big difference how to write. On the other hand, there was a result that the numerals written on a paper were identified in high accuracy (99.98%) [7].

Therefore, we try to verify that the numerals written in the air are useful for identification or authentication in this paper.

About numeral identification, there was our previous research [8]. In this research, we acquired 450 data from subjects by the Leap motion. We applied two pre-processings to these data, then these data were learned and classified by CNN. As a result, the average accuracy for identification was 93.4%. From this research, we could confirm the usefulness of numerals written in the air for identification. Hence, we try the personal authentication with the numerals written in the air as the next step in this paper.

2. Methods

In this paper, we measured numerals data using the Leap motion and then these data were learned and authenticated using CNN(Convolutional Neural Network). This flow is shown in Figure 2.

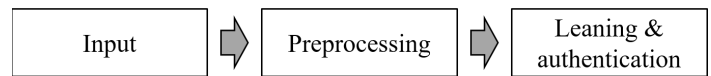


Figure 2 : Flow of the proposed method

When the Leap motion is used, a subject can write numerals with natural movement because it does not need special operation. Moreover, the introduction cost of the Leap motion is low and there was a result that characters which were measured by the Leap motion were useful for authentication. From these reasons, we use the Leap motion to collect data.

2.1. Input

The input data are measured by the Leap motion. Numerals 0 to 9 are the targets of measurement. These data consist of x, y, z coordinate values of the locus of the fingertip when subjects draw numerals. An example of the input data and its locus of x, y, z coordinate values are shown in Figure 3.

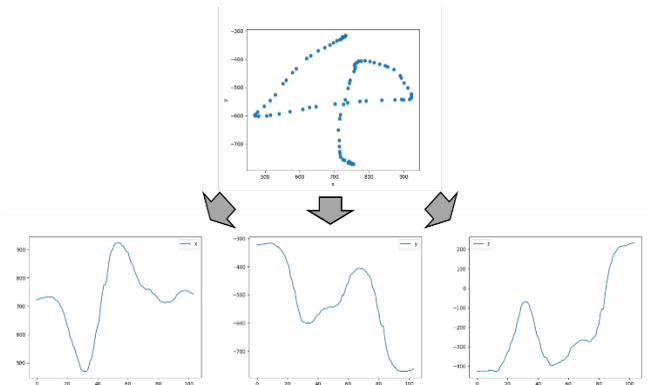


Figure 3 : Example of input data

2.2. Pre-processing

The data obtained directly from subjects are not suitable to learn in order to carry out the neural network learning. Therefore, we apply three pre-processings to raw measurement data then we reshape them to suitable. Of the three pre-processings, an outlier processing is the process newly added in this paper.

2.2.1 Adjustment of position to start writing

When we carry out the measurement, a inputting start coordinate values are not always same. From this factor, using such data directly for learning, they did not have good result for identification. To deal with this problem, we need to unify the writing start coordinate values. The i -th coordinate value before adjustment is defined as α_{before} , the i -th coordinate value after adjustment is defined as α_{after} , and the writing start coordinate value is defined as d . Moreover, we applied following formula to x, y, z coordinate values respectively for adjustment.

$$\alpha_{after}[i] = \alpha_{before}[i] - d \quad (1)$$

The data after application of this processing are shown in Figure 4. Through this adjustment, each starting point is adjusted to 0.

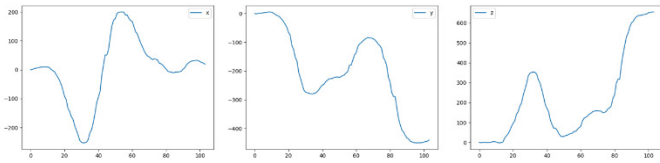


Figure 4 : Data after adjustment

2.2.2 Equalization of the data length

The measurement data have different data length for each measurement. To carry out learning by CNN using these data in later process, we have to unify length of the data. We set the longest data length in all measurement data as a length of the data after unified. Then, we use linear interpolation to unify the data length. The linear interpolation is a method to calculate a value corresponding to any coordinate which exists between two points. To carry out this processing, the following formulas are applied to i -th x, y, z coordinate values. In the following formulas, dst_size means data length after unification, $data_size$ means data length before unification, and t' means a value obtained by truncating the fractional part of t . The data after application of this processing are shown in Figure 5. In this paper, the number of data points was unified to 300.

$$scale = \frac{dst_size}{data_size} \quad (2)$$

$$t = \frac{i}{scale} \quad (3)$$

$$a = t - t' \quad (4)$$

$$x' = x_{[i]}(1 - a) + x_{[i+1]}a \quad (5)$$

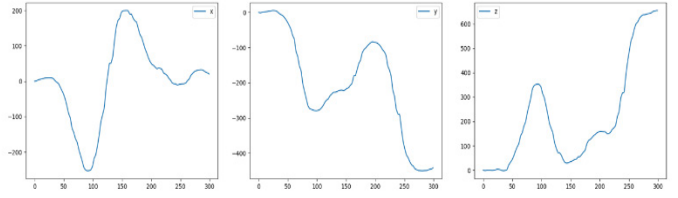


Figure 5 : Data after equalization using the linear interpolation

2.2.3 Outlier processing

When we carried out measurement, some outliers caused by the finger-sensor position relationship could occur. Therefore, it is necessary to remove such outliers. First, we use Interquartile range (IQR) to detect the outliers. We set upper limit and lower limit using IQR. Then, if any values exceed these limits, such values are detected as outliers. The IQR, upper limit, and lower limit can be calculated using the following formulas. 25percentile means 25 percent of data from the bottom and 75percentile means 75 percent of data from the bottom when data are arranged in ascending order.

$$IQR = 75percentile - 25percentile \quad (6)$$

$$Upper\ limit = 75percentile + 3.0 \times IQR \quad (7)$$

$$Lower\ limit = 25percentile - 3.0 \times IQR \quad (8)$$

Next, we remove the values which were detected by IQR. Finally we interpolate them by linear interpolation. These flow of processing is shown in Figure 6.

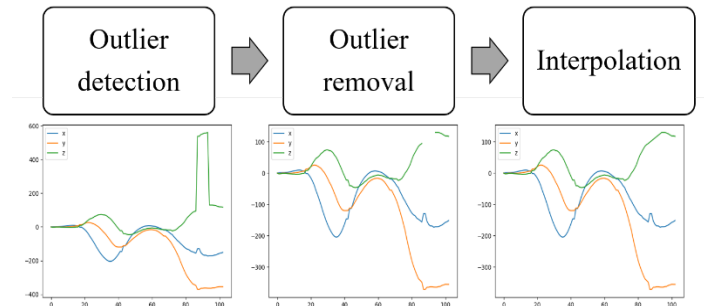


Figure 6 : Flow of outlier processing

By these processings, reduction of noises can be expected, we will be able to handle aerial input numerals better.

2.3. Learning & authentication

The preprocessed data are fed to CNN for authentication. The authentication (classification of subjects) with 2 classes separation(one subject or the others) is carried out.

2.4. Experiment

We measure the numerals 0 to 9 in order. Subjects carry out actual measurement after some practice of drawing. As shown in Figure 7, when we carry out the measurement, the Leap motion is placed on the desk and the subjects can see feedback on the screen.

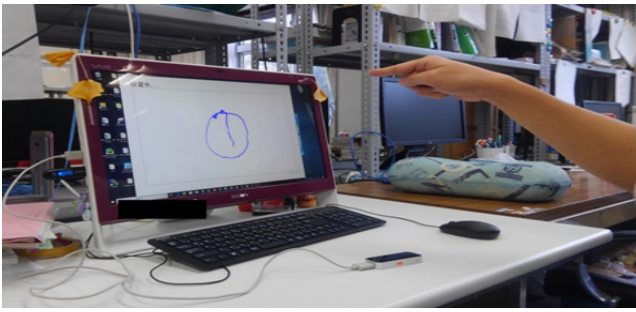


Figure 7 : Measurement scene

In the measurement, the subjects draw the numerals 0 to 9. The numerals 0 to 9 are defined as one set. 15 sets measurement a day is carried out. Measurement flow is shown in Figure 8.

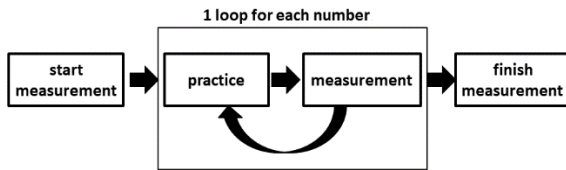


Figure 8 : Measurement flow

In this experiment, the subjects are four men and one woman. Measurement was carried out three days. These subjects are defined as A, B, C, D, and E. A target for authentication is the subject A. In the learning phase, a training data contains the two days data and a test data contains the rest one day data. Against these data, cross validation was carried out. The details of the data for validation are shown in Table 1.

Table 1 : Details of the data for cross validation

	Train	Test
Data1	Day1, Day2	Day3
Data2	Day1, Day3	Day2
Data3	Day2, Day3	Day1

Using these data, we carried out the learning and authentication by CNN. We used Adam as an optimization function and ReLU (rectified linear unit) as an activation function. Moreover, we set dropout rate of each layer 50%. The number of epochs was 300. The layer structure of CNN is shown in Figure 9.

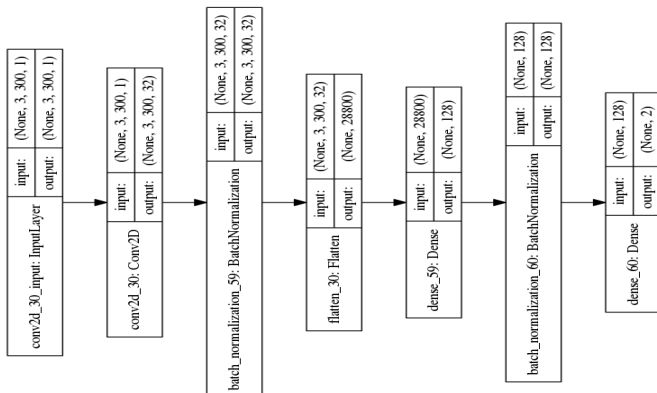


Figure 9 : Layer structure of CNN

As Figure 9 shows, the structure which was used in this experiment had just one convolution layer. Furthermore, the batch normalization and the early stopping were introduced to avoid the overfitting as much as possible. The kernel size was set to 3×2 to learn 3 channels(x, y and z coordinates) of input simultaneously. Moreover, when the convolution processing was carried out, we used a dilated convolution [9][10]. The dilated convolution can be carried out with making space for convolution. Using the dilated convolution, wider information can be processed. In the research by Luo, et al.[10], they referred dilated convolution is one of the methods to increase receptive field. Therefore, it can be expected it has the usefulness to authentication by numerals which have few feature differences between individual. This structure had fewer number of layers than common ones. There were two factors to this. First one was that the authentication target was numerals. Numerals have fewer features than other characters. The second factor was that the number of data was not sufficient for learning. From these factors, if there were more layers, excessive feature extraction could occur, and as the result, the overfitting may occur.

Table 2 : Experimental results 1

	Accuracy(%)		
	Data1	Data2	Data3
0	100.0	92.4	97.8
1	96.4	100.0	96.0
2	92.4	98.2	88.0
3	96.9	91.1	87.1
4	98.2	97.3	96.4
5	98.2	96.4	97.8
6	92.4	89.8	83.6
7	93.8	78.7	90.7
8	84.4	88.4	83.6
9	88.4	96.4	82.2

Table 3 : Experimental results 2

	Accuracy(%) (average)	FRR(%)	FAR(%)
0	96.7	0.0	5.4
1	97.4	0.5	3.7
2	92.9	1.5	10.4
3	91.7	0.2	13.6
4	97.3	1.5	3.0
5	97.5	1.0	3.2
6	88.6	2.5	16.5
7	87.7	0.7	19.8
8	85.5	5.2	19.0
9	89.0	0.7	17.5
average	92.4	1.4	11.2

3. Result

We carried out learning and authentication five times. The authentication results are shown in Tables 2 and 3. Moreover, we use FRR(False Rejection Rate) and FAR(False Acceptance Rate) as an evaluation criteria. FRR means a probability to reject a person who should be authenticated, and FAR means a probability to accept a person who should not be authenticated.

3.1. Discussion

We discuss about experimental results to verify the usefulness of the pre-processings and cause of misauthentication. From experimental results, we focus on numerals 5 and 8. These had the best and worst results. We considered that we could find out problems by comparing these numerals. First, comparison between subjects for data of "5" is shown in Figure 10.

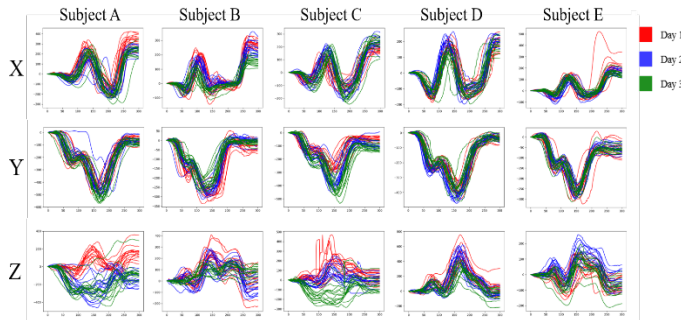


Figure 10 Comparison between subjects for data of "5"

The results for "5" were the best. From Figure 10, it can be seen that the x and y coordinate values have unique features between subjects. Even if the data had similar outlines, the scale of those data was different for each subject. However, in the z coordinate values, it can be seen that though some of data had unique features, data were sometimes unstable even within the same subject, and the outliers still remained. Therefore, we considered differences in the x and y coordinate values contributed to high accuracy. On the other hand, comparison between subjects for data of "8" which had the worst result is shown in Figure 11.

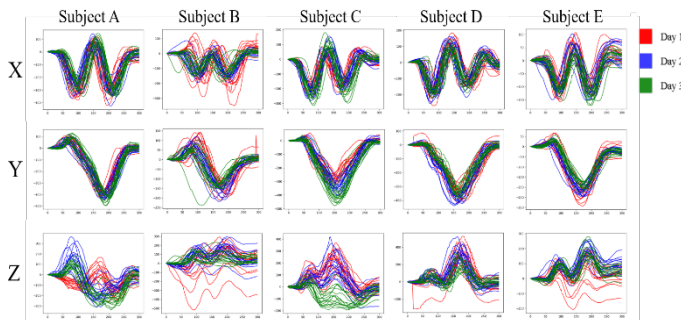


Figure 11: Comparison between subjects for data of "8"

From Figure 11, in the x and y coordinate values, it can be seen that the differences between subjects were relatively small compared to the case of "5". For the z coordinate values, they showed unique features for each subject and can be expected to be useful for authentication. However, some data trended differ even within the same subject. Thus, it can be considered the low accuracy was caused by such data. In the z coordinate values, such differences in data trended within the same subject can be seen not only in "8" but also in other numerals. Therefore, the misauthentication could be caused by the z coordinate values. However, if we removed the z coordinates, the authentication accuracy by numerals with relatively small difference between subjects in the x and y coordinate values became worse by reduction in subject's unique features. Accordingly, we will continue to use the z coordinates. However, we have to confirm

about a uniformity of the data within the same subject to use these data appropriately. Hence, we will conduct an additional experiment for confirmation of the uniformity as a future plan.

Next, we discuss about the usefulness of the pre-processings. We used three pre-processings, and focus on the outlier processing among them. The outlier processing was the most recently introduced process in our research and its utility has not been tested yet. Moreover, the FAR was worse than before applying the outlier processing. This process may have lost the features of the data. When we carried out comparison of data before and after applying this processing, we found some problems. Examples of data that have a problem process are shown in Figure 12.

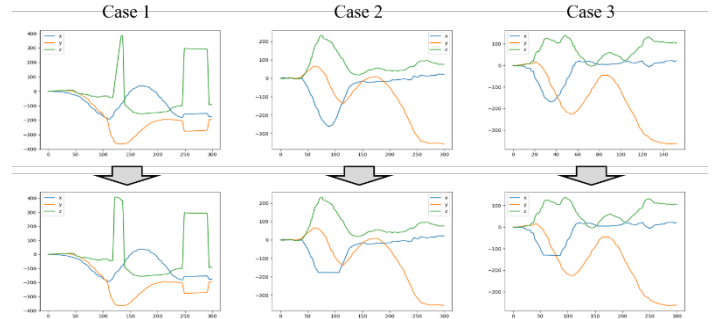


Figure 12 : Example of data that have a problem process

As shown in Figure 12, there were three cases. First, in the case 1, the outliers were not detected well although there were relatively long continuous outliers. If there were a lot of outliers, data distribution might be biased and outliers might not be detected. This distribution bias had a significant impact on the interquartile range. As a result, it can be considered this problem occurred.

Second, in the case 2, normal values were recognized as outliers. Therefore, a part of the data was deleted as outliers and an abrupt change occurred. Even if it looked like a natural change visually, it was regarded as outliers in case that the number of the data was small.

Third, in the case 3, the use of the linear interpolation for the outlier processing might be inappropriate. As shown in Figure 12, the linear interpolation sometimes caused unnatural interpolation.

However, the overall accuracy was improved by the introduction of the outlier processing because this processing was effective for some data. A successful example of outlier processing is shown in Figure 13.

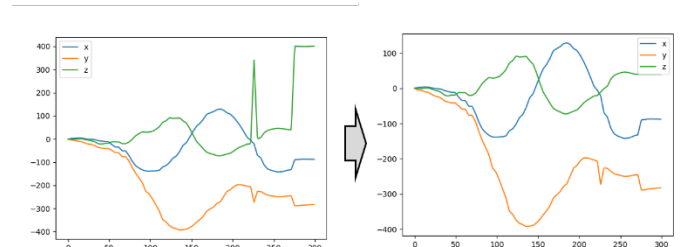


Figure 13 : Successful example of outlier processing

As shown in Figure 13, the outlier processing could reduce noises against some data. Noise generation is inevitable due to the nature of the experiment. Therefore, we need to solve the problems

mentioned in Figure 12. We will try to improve outlier processing by changing thresholds or trying another method as a future plan.

4. Conclusion

We proposed a method to carry out numeral identification and personal authentication by numerals inputted by the Leap motion sensor. In this paper, we focused on personal authentication and attempted discussion whether the aerial input numerals were useful for the authentication system. Moreover, we applied new one pre-processing and dilated convolution to numerals written in the air, verified these usefulness. We measured aerial input numerals data from five subjects by the Leap motion sensor and applied three pre-processings. And then, for these data, we carried out two classes separation(subject A or the others) by learning and authenticating using a Convolutional Neural Network with dilated convolution. As a result, average accuracy was 92.4%, FRR was 1.4%, and FAR was 11.2%. Compared to our previous paper[5], the overall accuracy has been improved. This result suggests that the aerial input numerals have enough features for authentication and the outlier processing and dilated convolution newly introduced this time are effective to aerial input numerals. Therefore we consider that we could show the usefulness of the aerial input numerals for authentication system. However, several problems, existence of biased z coordinate values in the same subject and problem of the outlier processing, were found by experiments and discussion. Against these problems, we will conduct additional experiment to collect more data, and then confirm trends of data. Moreover, we will try to carry out another method for outlier removal, find more effective way to handle aerial input numerals.

References

- [1] Shun Yamamoto, Minoru Fukumi, Shin-ichi Ito, Momoyo Ito, "Authentication of Aerial Input Numerals by Leap Motion and CNN", Proc. of 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), pp.189-193. <https://doi.org/10.1109/IOTAIS.2018.8600847>
- [2] Masaji Katagiri, Toshiaki Sugimura, "Personal Authentication by Signatures in the Air with a Video Camera" PRUM 2001-125, pp.9-16(2001) in Japanese
- [3] Issei Hatanaka, Masayuki Kashima, Kimionri Sato, Mutsumi Watanabe, "Study on Flexible Aerial Signature Individual Authentication System using the Finger Discrimination Information", Journal of the Institute of Image Information and Television, Vol70, No.6. pp.J125-J132(2016) in Japanese
- [4] Kumiko Yasuda, Daigo Muramatsu, Satoshi Shirato, Takashi Matsumoto, "Visual-Based Online Signature Verification Using Features Extracted from Video", Journal of Network and Computer Applications, 33, 3, pp.333-341(2010). <https://doi.org/10.1016/j.jnca.2009.12.010>
- [5] Kohei Nakanishi, Shin-ichi, Minoru Fukumi, Momoyo Ito, "Biometrics Authentication of Aerial Handwritten Signature Using a Convolutional Neural Network", Proc. of The 5th IIAE International Conference on Intelligent Systems and Image Processing 2017, 19-24. <https://doi.org/10.12792/icisip2017.007>.
- [6] Zheng Yi., Liu Qi., Chen Enhong., Ge Yong., Zhao J.Leon. (2014) "Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks." In: Li F., Li G., Hwang S., Yao B., Zhang Z. (eds) Web-Age Information Management. WAIM 2014. Lecture Notes in Computer Science, vol 8485. Springer, Cham. https://doi.org/10.1007/978-3-319-08010-9_33
- [7] Sajjad. S. Ahranjany, Farbod. Razzazi and Mohammad. H. Ghassemian, "A very high accuracy handwritten character recognition system for Farsi/Arabic digits using Convolutional Neural Networks," 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA), Changsha, 2010, pp. 1585-1592. <https://doi.org/10.1109/BICTA.2010.5645265>.
- [8] Shun Yamamoto, Minoru Fukumi, Shin-ichi Ito, Momoyo Ito, "Recognition of Aerial Input Numerals by Leap Motion and CNN", Proc. of SCIS and ISIS2018, pp.1177-1180. <https://doi.org/10.1109/SCIS-ISIS.2018.00185>.
- [9] Fisher Yu, Vladlen Koltun, "Multi-Scale Context Aggregation by Dilated Convolutions", Proc. of International Conference on Learning Representation (ICLR2016). arXiv:1511.07122,
- [10] Luo, Wenjie., Li, Yujia., Urtasun, Raquel., & Zemel, Richard. (2016). "Understanding the effective receptive field in deep convolutional neural networks". In *Advances in neural information processing systems* (pp. 4898-4906). arXiv:1701.04128