KNOWLEDGE BASE CREATION BY RELIABILITY OF COORDINATES DETECTED FROM VIDEOS FOR FINGER CHARACTER RECOGNITION

Shinpei Hagimoto *, Takuma Nitta *, Ari Yanase *, Takafumi Nakanishi *, Ryotaro Okada *, Virach Sornlertlamvanich *

*Department of Data Science Musashino University **

*3-3-3 Ariake Koto-ku Tokyo 135-8181, Japan **

ABSTRACT

In this paper, we propose a novel knowledge base creation method by reliability of coordinates detected from videos for finger character recognition. A sign language communication is a crucial method for deaf or hearing-impaired people. One of the most important issues is the preparation of video as a training data set for finger character recognition. Unfortunately, preparing a large enough video data set for finger character recognition is a labor intensive and time-consuming task. Instead, we are proposing a finger character knowledge base consisting of video data and the designed metadata. Our method enables the creation of a knowledge base and metadata with high accuracy from a small amount of training data. By realizing our method, we can obtain a high accuracy metadata for finger character to use in the knowledge base. As a result, our proposed method can realize an accurate and robust finger character recognition system by consulting the knowledge base.

KEYWORDS

finger character recognition, sign language, finger character, knowledge base creation, reliability of coordinates, video metadata.

1. INTRODUCTION

About 5% of the world's population is handicapped by hearing loss or other deafness. For the hearing impaired, sign language and finger character are very important communication tools. However, one of the key problems is that very few people can understand and communicate with a sign language or finger character. One of the ways to support communication among the hearing impaired is through an automatic translation system of sign language and finger character. By realizing an automatic translation system, it will be possible to build an environment that promotes their social advancement and interaction.

In general, it is necessary to prepare a large corpus of training data to implement an automatic translation system. However, sign language and finger characters vary from country to country and from region to region. That makes it hard to collect large enough image data sets for training Japanese finger character. In our previous work (Nitta et al, 2020), we realized a finger character recognition system with a small amount of training data. The recognition system with a small amount of training data has not yet been able to achieve a quality of accuracy at an applicable level. To implement an automatic Japanese finger character translation system, it would be ideal to build a knowledge base from a small amount of training Japanese finger character images and use them as training data to get acceptable recognition accuracy.

In this paper, we propose a new method to create a knowledge base using the reliability of finger joint coordinates detected from finger character videos. The reliability is a value in the range [0,1] that shows how accurate the coordinates are. We will improve the recognition accuracy by creating a knowledge base with features extracted more accurately using the reliability of the coordinates. Our method enables the creation of a knowledge base and metadata with a high accuracy from a small amount of training data. By realizing our method, we can obtain high accuracy metadata for finger character to use in the knowledge base. By creating a knowledge base with our method, we realize accurate and robust finger character recognition.

The contribution of our method is as follows:

- We propose a knowledge base method instead of trained model method for finger character recognition. The proposed method is proven to compensate for the shortage of training image data sets of Japanese finger characters. It is difficult to create a knowledge base with small amount of data though, in our previous work, we successfully utilized the knowledge base for Japanese finger character recognition.
- We evaluate the knowledge base creation method.
- We implement a finger character recognition system.

This paper is organized as follows. In Section 2, the related works are discussed. In Section 3, we present our previous work, which is the basis of this work, and the proposed method of creating a knowledge base using reliability of finger joint coordinates. In Section 4, we present the results of experiments using the created knowledge base. Finally, in Section 5, we summarize the proposed method and evaluation results.

2. RELATED WORK

In this Section, as in our previous work (Nitta et al, 2020), we first present feature extraction and its pre-processing, which are the necessary components for finger character recognition, and knowledge base creation.

First, we present the existing methods for feature extraction and pre-processing of finger characters, which are necessary components for finger character recognition. The first group of methods is to directly extract features using a glove with multiple sensors (Fang et al, 2004; Kong and Ranganath, 2008). By extracting the features directly, it is possible to obtain very accurate data. The accurately extracted data contributes greatly to improve the accuracy of recognition, while it is inconvenient for people to wear gloves for communication in their daily life. The second group of methods is to extract features indirectly using a depth sensor (Cem et al, 2012). The depth sensor can extract the changes of the coordinate information of X-Y-Z axes of the body skeleton including hands and fingers due to the motion of finger characters. It is effective for objects with three-dimensional motion because the depth sensor can also extract depth coordinates. The third group of methods, similar to the second group of methods, is to extract features indirectly using a device equipped with a monocular camera. There are several ways to extract features from RGB images or videos. The first one is to use the color information of the image or video (Gu et al, 2012). This method is highly depending on the shape of the hand, skin color, background, etc. The second one is to use image recognition algorithms such as convolutional neural networks (CNNs) with image recognition preprocessing (Konstantinidis et al, 2018). This method requires a huge amount of image data. The third one is to extract the motion vector of an object using the optical flow algorithm (Konstantinidis et al, 2018; Lim et al, 2016). Furthermore, there is another one to extract the coordinates and joint angles of features from videos and images using deep learning that extracts skeletal coordinates (Simon et al, 2017). Our method, proposed in our previous work (Nitta et al, 2020), obtains finger character features by combining a camera and a deep learning model for extracting finger joint coordinates.

Concerning the knowledge base construction, the Mathematical Model of Meaning (Kiyoki et al, 1994; Kitagawa and Kiyoki, 1993) is applied to create the representation of multimedia data as vectors and determine the contextual semantic measure. In our previous work (Nitta et al, 2020), we extracted feature coordinates from trimmed finger character videos using OpenPose (Cao et al, 2017), and stored and created a knowledge base with the average values of the same coordinates as representative values. In this paper, the knowledge base is automatically created by the limited features using the reliability of the coordinates obtained. In OpenPose, we can obtain the reliability of the coordinates along with the coordinates, and it is used for data cleansing, etc.

3. KNOWLEDGE BASE CREATION METHOD BY RELIABILITY OF COORDINATES FOR RECOGNITION

In this Section, we present our proposed method for the knowledge base creation by coordinates reliability. We have proposed a finger character recognition method in a sign language by similarity measure using a

finger character feature knowledge base. This method consists of the knowledge base stored with features obtained from a small amount of training data. This method realizes finger character recognition by similarity measures using the knowledge base. This method consists of two systems— a knowledge base creation system, and a recognition system as shown in Figure 1. The knowledge creation system consists of three modules, namely the annotation, the feature extraction, and the knowledge base construction. The recognition system consists of three modules, namely the feature extraction, the data screening, and the similarity measure. We perform the dimensionality reduction and anomaly detection to reduce the computational cost. After that, we perform finger character recognition by similarity measures using the knowledge base and the extracted features. In this paper, we focus on the knowledge base creation system of our system. Our method extracts metadata by using the reliability of finger joint coordinates obtained from finger character. By realizing our method, we can obtain a high accuracy metadata for finger character to use in the knowledge base. By creating a knowledge base with our method, we realize accurate and robust finger character recognition.

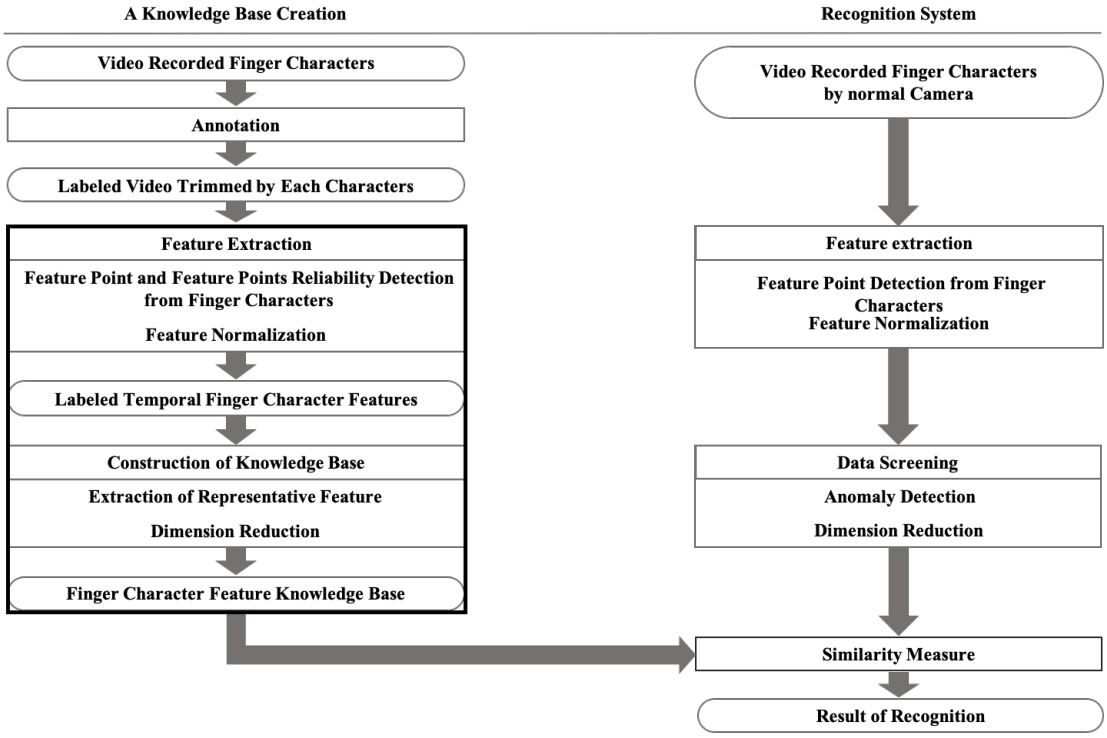

Figure 1. The overview of our recognition system. Our considered knowledge base creation is marked with a bold line.

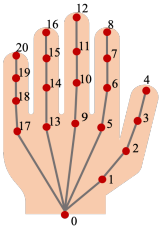

Figure 2. The right-hand skeleton joints and numbers employed our proposed method. The coordinates of these joints are extracted by OpenPose. Each joint is represented as "key01, key02, …, key21".

3.1 Reliability of coordinates

In this Section, we present about the reliability of the coordinates. We apply OpenPose to detect joint coordinates for deep learning model. We obtain the coordinates by using OpenPose as a feature of the finger character. OpenPose can extract the coordinates of 135 joints of a person's body in the image, and our method uses the joint coordinates of 21 of the right hands. Joints extracted from OpenPose and the assigned numbers to the joints are shown in Figure 2. Finger character features are 63-dimensional data using the X and Y axis coordinates and reliability of the coordinates of 21 joints of the right hand extracted from OpenPose. The reliability is a value between 0 and 1 defined in OpenPose. The closer the value is to 1, the higher the reliability of coordinate values and the more correct the coordinates are. There is one reliability for each of the 21 joint coordinates of the hand. A matrix C constructed by the data obtained from one finger character video is shown in Figure 3. The matrix C consists of X axis coordinates k_i_x and Y axis coordinates k_j_y and reliability of the coordinates k_l_r by each time.

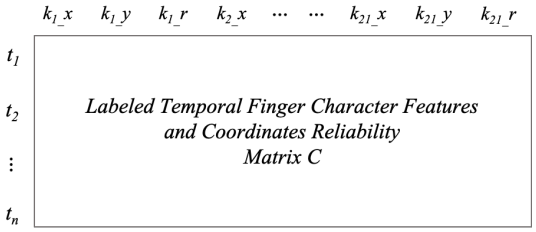

Figure 3. Labeled temporal finger character feature and coordinates reliability matrix C. The matrix C is obtained from a single finger character video. k_i_x is the coordinate of x, k_j_y is the coordinate of y, and k_l_r is the reliability of the coordinates. t is a temporal series.

3.2 Knowledge base creation method

In this Section, we present the knowledge base and methods to create the knowledge base using the reliability of the finger joint coordinates. In Section 3.2.1, we present the definition of the knowledge base in our proposed method. In Section 3.2.2, we present the outline of the creation of a knowledge base. In Section 3.2.3, 3.2.4, 3.2.5 and 3.2.6, we present methods for creating each of them.

3.2.1 knowledge base

In this Section, we present the definition of the knowledge base in our proposed method. A knowledge base is a matrix of features of 41 static finger characters without movement out of 46 Japanese finger characters. The structure of a knowledge base R is shown in Figure 4 (a). We stored the extracted coordinate features from matrix C using the reliability in the proposed method in matrix R. There are 42 features for each character, and a knowledge base has 42 columns and 41 rows. As in our previous work (Nitta et al, 2020), the knowledge base S for recognition consists of three knowledge bases R for robustness. The knowledge base S is shown in Figure 4 (b).

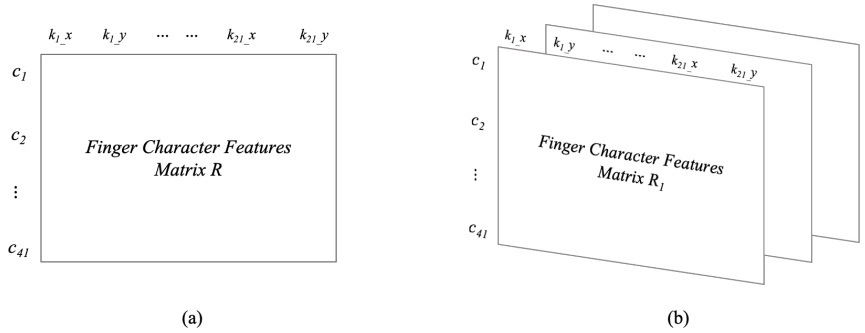

Figure 4. (a) The structure of Finger Character Features Matrix R. (b) The structure of the knowledge base S.

3.2.2 Outline

In this Section, we present the outline of the knowledge base creation. In our previous work (Nitta et al, 2020), the knowledge base consists of the average value of the same coordinates obtained from a single video as the representative value of that character. In this paper, we create a knowledge base by a new method of extracting the representative values. As shown in section 3.1, OpenPose can extract the reliability of each coordinate as well as the coordinate. We extract the features with more correct coordinates using the reliability and store them as representative values in a knowledge base. We also created a matrix Γ by extracting only the reliability from matrix C and used it to create the knowledge base. The matrix Γ is shown in Figure 5.

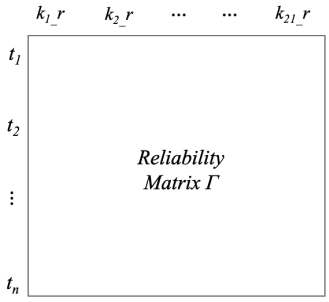

Figure 5. Matrix Γ consist of the value of reliability.

3.2.3 Knowledge base creation by the highest average value

In the first method, the features from the matrix C with the highest average of the value of reliability t_j_ave in the matrix Γ for each character are the representative values of that character. When the reliability of each coordinate is high, the average of the values of the reliability of the coordinates at that time is also high, and the features can be considered to be more correctly obtained. In this method, we extract the representative value of each character and store it in a knowledge base. The example of the representative value of c_1 in this method is shown in Figure 6.

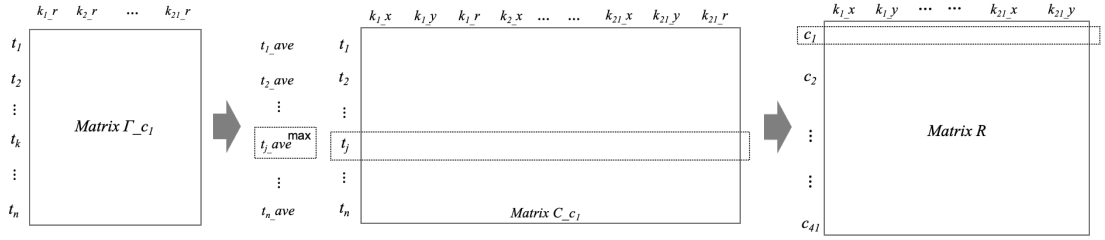

Figure 6. Knowledge base creation by the highest average value.

3.2.4 Knowledge base creation by the highest minimum value

In the second method, the features from the matrix C with the highest minimum values of reliability t_k_min in the matrix Γ for each character are used as the representative values for that character. The higher the minimum value of reliability, the higher the reliability of the coordinates at that time, and the more correctly the feature is obtained. In this method, we extract the representative value of each character and store it in a knowledge base. The example of the representative value of c_1 in this method is shown in Figure 7.

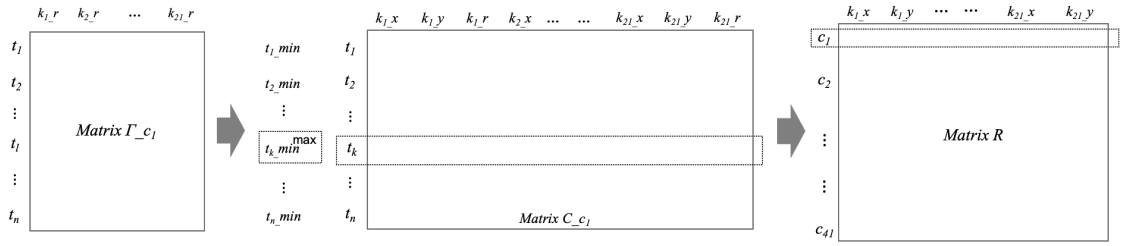

Figure 7. Knowledge base creation by the highest minimum value.

3.2.5 Knowledge base creation by the average of averages

In the third method, we first find the average value of the reliability t_l_ave in the matrix Γ of each character, as in the first method. Then, we extract the average of the averages. We extract all the coordinates from the matrix C when the average of the reliability is higher than the overall average A. Then, we average all the extracted coordinates for each same joint coordinate, and the coordinate of the average is the representative value. By using the coordinates with higher averages of reliability than a set value, we can use only the coordinates that have been obtained more correctly. The extracted representative values for each character are stored in the knowledge base. The example of the representative value of c_1 in this method is shown in Figure 8.

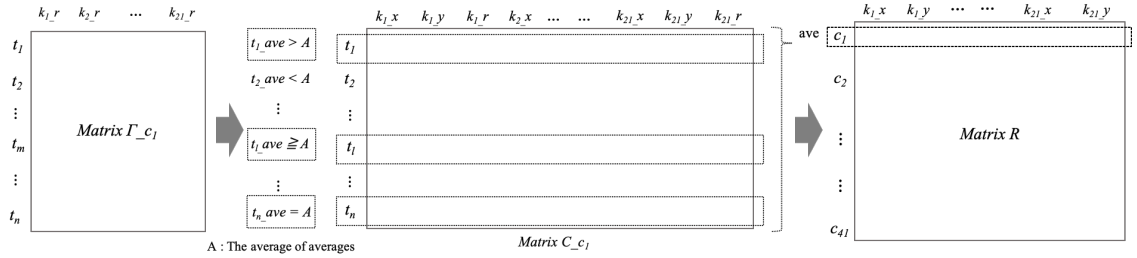

Figure 8. Knowledge base creation by the average of averages.

3.2.6 Knowledge base creation by the sum of the maximum and minimum values

In the fourth method, the feature from the matrix C with the highest value for sum of the maximum and minimum values of the reliability $t_m max + t_m\_min$ in the matrix Γ of each character is the representative value of that character. The higher the sum of the maximum and minimum values of the reliability, the higher the reliability of the coordinates at that time, and the more correctly the feature is obtained. The representative value of each character is obtained by this method and stored in the knowledge base. The example of the representative value of c_1 in this method is shown in Figure 9.

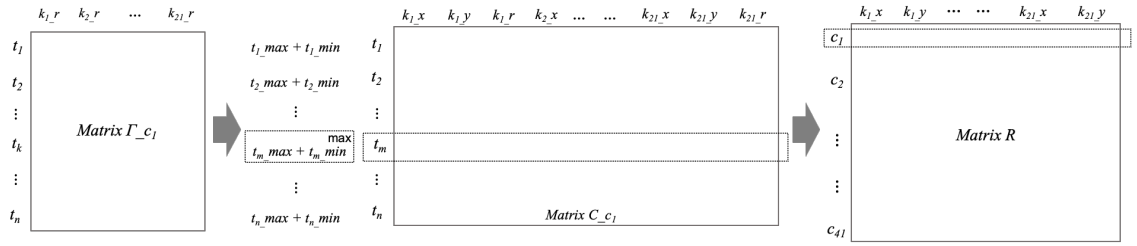

Figure 9. Knowledge base creation by the sum of the minimum and maximum values.

4. EXPERIMENTS

Table 1. The environment of experiment.

OS	Mac OS Catalina
RAM	16GB
CPU	2.3GHz Quad-Core Intel Core i5
OpenPose	Ver. 1.6.0
Camera	Monocular camera on iPad Pro 2nd generation

4.1 Experiment environment

All the experiments are performed in the environment shown in Table 1. We perform recognition using the recognition method shown in Figure 1, while changing the knowledge base we created. In addition, we utilize the same data as in our previous work for experimental data and test data for accuracy evaluation. The data consists of photographs of static finger characters taken from three research members and two experimental supporters. There are 41 static finger characters in each set, and there are 5 sets.

4.2 Experiment result

In this Section, we present the experimental results. We evaluate the recognition accuracy with the same test data as in our previous study. We perform similarity measure between the input data and each of the four types of knowledge bases as explained in Subsection 3.2.3 - 3.2.6. Also, we calculate the average of the accuracy for each knowledge base. The results and average accuracy are shown in Table 2.

Table 2 shows the recognition accuracy differs depending on the knowledge base. Also, it shows that the average recognition accuracy is the highest when the knowledge base is created by Average-Ave method in explained in Section 3.2.5. However, it shows that the recognition accuracy of Person-1, 3, and 5 is higher when Highest-Ave or Highest-Min, and that of Person-2 and 4 is higher when Average-Ave. There is a 5~6% difference in recognition accuracy depending on how the representative values are stored in the knowledge base.

Table 2. The results of all experiments.

Types of Knowledge base	Person-1	Person-2	Person-3	Person-4	Person-5	average
Highest-Ave	87.50 %	63.41 %	60.98 %	68.75 %	62.50 %	68.63 %
Highest-Min	82.50 %	63.41 %	60.98 %	65.63 %	62.50 %	67.00 %
Average-Ave	85.00 %	68.29 %	58.54 %	71.86 %	60.00 %	68.74 %
Sum-of-Max-Min	82.50 %	63.41 %	58.54 %	68.75 %	60.00 %	66.64 %

Table 3. The results of experiments in our previous work.

	Person-1	Person-2	Person-3	Person-4	Person-5	average
Previous work	87.50 %	63.41 %	60.98 %	75.00 %	62.50 %	69.88 %

4.3 Discussion

In this Section, we discuss the results of the experiments and compare them with the results of our previous work. The experimental results of the previous work are shown in Table 3. As shown in Tables 2 and 3, if we average the recognition accuracy of all five evaluation data sets, we can consider that the knowledge base of our previous work has the highest accuracy. However, except for Person-4, all the others, all the knowledge bases perform as well as the previous work.

We focus on Person-2 and Person-4. There is a difference of about 5% in the accuracy of Person-2 and about 3% in Person-4 between the results of the Average-Ave knowledge base and the results of our previous work. In the case of Person-2, the previous work recognized "E" as "KE" and "SHI" as "MU" incorrectly, but the Average-Ave knowledge base recognized them correctly. Also, in the case of Person-4, "KI", which was correctly recognized by our previous work, but incorrectly recognized as "I" by the Average-Ave knowledge base. In both cases, the incorrectly recognized hands have similar shapes. These finger characters are shown in Figure 10.

Figure 10. The finger character images that affected the accuracy.

5. CONCLUSION

In this paper, we proposed a new knowledge base creation method using the reliability of coordinates detected from finger character videos for recognition. Our method enables the creation of a knowledge base and metadata with high accuracy from a small amount of training data. By realizing our method, we can obtain high accuracy metadata for finger character to use in the knowledge base. By creating a knowledge base with our method, we realize accurate and robust finger character recognition. Our method achieves the same level of accuracy when comparing to our previous work in the finger character recognition using the knowledge base created in our method.

In the future, we will continue to improve the accuracy of the recognition and evaluate the method of knowledge base creation by utilizing our method. We would also like to develop the recognition of dynamic finger characters, which was not achieved in this work. Also, we would like to use our method and system to recognize not only dynamic finger characters but also sign language that has various motions and gestures that have meanings other than sign language. In addition, we will develop a communication tool on mobile devices applying our method and system to realize finger character recognition.

REFERENCES

Cao, Z., Simon, Z., Wei, S. and Sheikh, Y., 2017. Realtime multi-person 2D pose estimation using part affinity fields. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, pp. 1302-1310.

Cem, K., Furkan, K., Yunus, K. and Lale, A., 2012. Hand pose estimation and hand shape classification using multi–layered randomized decision forests. *2012 Proceedings of the 12th European conference on Computer Vision – Volume Part VI*. pp. 852-863.

Fang, G., Gao, W. and Zhao, D., 2004. Large vocabulary sign language recognition based on fuzzy decision trees. *in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*. vol. 34, no. 3, pp. 305-314, May.

Gu, L., Yuan, X. and Ikenaga, T., 2012. Hand gesture interface based on improved adaptive hand area detection and contour signature. *2012 International Symposium on Intelligent Signal Processing and Communications Systems*. Taipei, pp. 463-468.

Kitagawa, T. and Kiyoki, Y., 1993. A mathematical model of meaning and its application to multidatabase systems. *Proceedings RIDE-IMS `93: Third International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase System*s. Vienna, Austria, pp.130-135.

Kiyoki, Y., Kitagawa, T. and Takanari, H., 1994. A metadatabase system for semantic image search by a mathematical model of meaning. *ACM Sigmod Record*. vol.23, no. 4, pp. 34-41, 1994.

Kong, W.W. and Ranganath, R., 2008. Signing Exact English (SEE): Modeling and recognition. *Pattern Recognition*. vol. 41, no. 5, pp. 1638-1652.

Konstantinidis, D., Dimitropoulos, K. and Daras, P., 2018. A deep learning approach for analyzing video and skeletal features in sign language recognition. *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*. Krakow, Poland, pp1-6.

Lim, K.M., Tan, A.-W.C. and Tan, S.-C., 2016. Block-based histogram of optical flow for isolated sign language recognition. *Journal of Visual Communication and Image Representation*. vol. 40, part B, pp. 538- 545.

Nitta, T., Hagimoto, S., Yanase, A., Nakanishi, T., Okada, R. and Virach Sornlertlamvanich, 2020. Finger Character Recognition in Sign Language Using Finger Feature Knowledge Base for Similarity Measure. *3rd IEEE/ IIAI International Congress on Applied Information Technology*. Tokyo, Japan.

Simon, T., Joo, H., Matthews, I. and Sheikh, Y., 2017. Hand keypoint detection in single images using multiview bootstrapping. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, pp. 4645-4653.