A Primitive Action-driven Recognition Method for the Realization of Global Heterogeneous Sign Language Recognition

Takafumi Nakanishi,[1,2] Ayako Minematsu,[2]
Ryotaro Okada,[1,2] Osamu Hasegawa,[1,2]
Virach Sornlertlamvanich[1,2]

[1] Musashino University, Department of Data Science, Tokyo, Japan
takafumi.nakanishi@ds.musashino-u.ac.jp, ryotaro.okada@ds.musashino-u.ac.jp, osamu@ds.musashino-u.ac.jp, virach@musashino-u.ac.jp.
[2] Musashino University, Asia AI Institute, Tokyo, Japan
takafumi.nakanishi@ds.musashino-u.ac.jp, ayako.minematsu@ds.musashino-u.ac.jp, ryotaro.okada@ds.musashino-u.ac.jp, osamu@ds.musashino-u.ac.jp

We represent a primitive action-driven recognition method for realizing global heterogeneous sign language recognition. We should realize a method to recognize sign languages from various linguistic areas as easily as possible for global communication. However, most of the current sign language recognition methods realize specific sign language recognition for individual linguistic regions, and when we realize sign language recognition among multilingual regions, we should implement it in an ad hoc manner. To develop multilingual sign language recognition, it is necessary to realize a new method to handle various sign systems in a unified manner. This method defines common primitive actions of various sign language systems worldwide and describes what the combination of these primitive actions indicates in various sign language systems to realize sign language recognition. This method consists of multiple primitive action recognition modules and a primitive action composition module. Each primitive action recognition module recognizes each primitive action common to all sign languages. The primitive action composition module determines the actual sign meaning from the combination of recognition results from multiple primitive action recognition modules.

Keywords:
sign language recognition, primitive actions, global communication platform, global heterogeneous sign language, action-driven recognition

University of Maribor Press

DOI https://doi.org/10.18690/um.feri.5.2023.14
ISBN 978-961-286-745-4

1 Introduction

Communicating with diverse people from different linguistic backgrounds is becoming increasingly important. Sign language, mainly used by people who are deaf or hard of hearing for everyday communication, has established itself as its own language system. For all people to communicate naturally and easily with each other, it is important not only to translate between languages but also to recognize and work seamlessly with other methods, such as sign language. It is important to realize a global communication platform that helps all diverse people communicate.

So far, we have been studying sign language recognition methods [1][2][3][4]. This research focuses on methods to achieve sign language recognition with small training data. We have found it difficult to collect enough sign language video data. Therefore, applying machine learning methods to sign language recognition is generally difficult. In our research [1][2][3][4], recognition is realized by extracting time-series skeletal features from training data in advance, extracting time-series skeletal features from input videos, and computing similarity weighing. These methods [1][2][3][4] are realized in Japanese Sign Language and can be applied to other sign languages. However, the cost is too high to apply to many sign languages quickly.

According to the reference [5], there are more than 400 different sign languages worldwide, depending on the country, region, etc. For all people to communicate naturally and easily, it is necessary to recognize and compose capabilities for these 400+ sign languages that must be realized and seamlessly coordinated. We need to create a system that facilitates the application of recognition and composition to these various sign languages. Most current sign language recognition methods realize specific sign language recognition for individual linguistic regions. When we realize sign language recognition among multilingual regions, we should implement it ad hoc manner. To develop multilingual sign language recognition, it is necessary to realize a new method to handle various sign systems in a unified manner.

We represent a primitive action-driven recognition method for realizing global heterogeneous sign language recognition. This method defines common primitive actions of various sign language systems worldwide. It describes what the combination of these primitive actions indicates in various sign language systems to

realize sign language recognition. This method consists of multiple primitive action recognition modules and a primitive action composition module. Each primitive action recognition module recognizes each primitive action common to all sign languages. The primitive action composition module determines the actual sign meaning from the combination of recognition results from multiple primitive action recognition modules. When introducing a new sign language system, this method can be implemented simply by adding the combinations of primitive actions and their meanings to the knowledge base in the primitive action composition module. In other words, by realizing this method, it will be possible to integrate more than 400 sign language systems without implementing ad hoc recognition and synthesis systems for each. This method will realize a new global communication platform that avoids the communication divide and allows people to communicate freely in the current situation, where people communicate in many ways.

This paper uses HamNoSys (The Hamburg Sign Language Notation System) [6], a transcription system common to all signs, to realize multiple primitive action recognition modules. The HamNoSys is a transcription system for all sign languages with a direct correspondence between symbols and gesture aspects, such as hand location, shape, and movement. We can realize each primitive action recognition function according to each handshape chart in HamNoSys.

This paper makes the following contributions to the broader research field.

- We propose a new method—a primitive action-driven recognition method to realize global heterogeneous sign language recognition.
- To realize our method, we apply the HamNoSys [5] to multiple primitive action recognition modules.

This paper is organized as follows. In section 2, we present some related works of our method. Section 3 provides an overview of the existing study, HamNoSys [5]. Section 4 represents our primitive action-driven recognition method to realize global heterogeneous sign language recognition. In section 5, we describe some results of preliminary experiments. Finally, in section 6, we summarize this paper.

2 Related Works

Our previous works [1][2][3][4] presents sign language recognition methods. In these methods, recognition is realized by extracting time-series skeletal features from training data in advance, extracting time-series skeletal features from input videos, and computing similarity weighing. We have found it difficult to collect enough sign language video data. Therefore, applying machine learning methods to sign language recognition is generally difficult. Our previous paper [1][2][3][4] also described some related works for the realization of sign language.

The reference [7] surveys machine learning methods applied in sign language recognition systems. This reference [7] says that sign language involves the usage of the upper part of the body, such as hand gestures [8], facial expression [9], lip-reading [10], head nodding, and body postures to disseminate information [11] [12] [13]. We classify hand gestures and lip reading as verbal behavior. We classify head nodding and body postures to disseminate information as emotional behavior. We classify facial expressions as both verbal and emotional behavior.

According to reference [3], sign language recognition methods can be divided into two categories: continuous recognition of multiple sign words and discontinuous recognition. To realize continuous recognition, there are some works such as the method of hidden Markov model (HMM) and dynamic time warping (DTW) [14] or the methods using Random Forest, artificial neural network (ANN), and support vector machine (SVM) [15]. To realize non-continuous recognition, there are some works, such as the method of k-nearest neighbor (k-NN) [16], SVM [17], and sparse Bayesian classification of feature vectors generated from motion gradient orientation images extracted from input videos [18]. To realize sign language recognition for non-continuous and non-time-series data, there are some works such as the method of k-NN [19], similarity calculation using Euclidean distance [20], cosine similarity [19][21], ANN [22], SVM [23], and convolutional neural network (CNN) [24]. The reference [25] provides a research survey on recognizing emotions from body gestures. Their works solve the problem of some of these sign language recognition functions.

However, preparing enough training data for sign language recognition is necessary to realize these methods. Adequately preparing sign language videos and their labeled training data is often impossible. In addition, according to the reference [5], there are more than 400 different sign languages worldwide; depending on the country, region, etc., it is not realistic to implement a recognition system for more than 400 different sign languages in an ad hoc manner. We must realize a recognition platform that could easily and uniformly apply each sign language.

The concept of "primitive" is proposed by Kiyoki et al. [25] in the metadatabase system architecture. The metadatabase system connects several legacy databases. For connecting several legacy databases through the metadatabase system, each legacy database has some primitive functions.

Applying the reference [25], this method has the recognition function of each basic hand movement as each primitive action recognition module. It derives the meaning of sign language from the primitive action composition module that integrates them. Our method can be implemented simply by adding the combinations of primitive actions and their meanings to the knowledge base in the primitive action composition module for realizing the other sign language recognition system. In other words, by realizing this method, it will be possible to integrate more than 400 sign language systems without implementing ad hoc recognition and synthesis systems for each. This method will realize a new global communication platform that avoids the communication divide and allows people to communicate freely in the current situation, where people communicate in many ways.

3 HamNoSys (The Hamburg Sign Language Notation System)

This paper uses HamNoSys (The Hamburg Sign Language Notation System) [6], a transcription system common to all signs, to realize multiple primitive action recognition modules. The HamNoSys is a transcription system for all sign languages with a direct correspondence between symbols and gesture aspects, such as hand location, shape, and movement. We can realize each primitive action recognition function according to each handshape chart [27] in HamNoSys.

Figure 1 shows the HamNoSys handshape chart. The description of each handshape in Figure 1 comprises symbols for the basic forms and diacritics for thumb position and bending. By this approach, the handshape descriptions should include all handshapes used in sign language worldwide. HamNoSys can be applied internationally because it does not refer to nationally diversified finger figures.

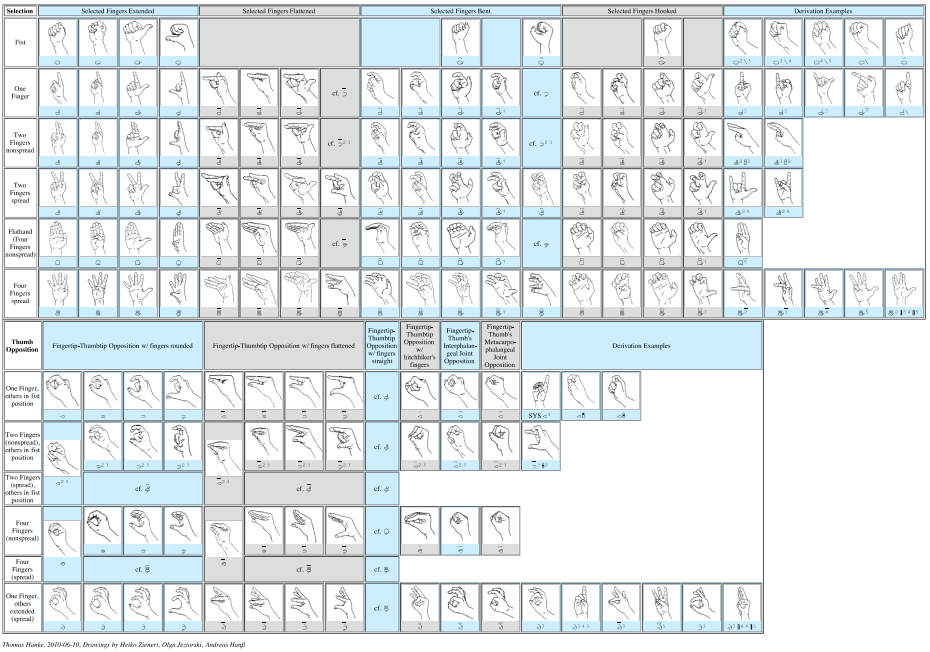

Figure 1: HamNoSys Handshape Chart
Source: [27].

We construct each primitive action recognition module that recognizes the shape of each finger appearing in Figure 1. All finger-expressed signs are combinations of finger shapes in Figure 1. We can potentially recognize all the world's sign languages by building a system that can recognize all these finger shapes. We only need to prepare as a knowledge base which combinations of finger shapes indicate what meaning. The primitive action composition module is the function that derives meaning using this knowledge base. In other words, our method can realize the recognition of various signs without training data from sign language videos simply by appending the knowledge base referenced by the primitive action composition module. This eliminates the need to collect sufficient video sign language media

content necessary to achieve sign language recognition. We can easily realize a unified global sign language recognition system. This method will realize a new global communication platform that avoids the communication divide and allows people to communicate freely in the current situation, where people communicate in many ways.

4 Primitive Action-driven Recognition Method

This section presents our new method— a primitive action-driven recognition method. This sign language recognition method can easily and uniformly apply to various sign language systems. This method will realize a new global communication platform that avoids the communication divide and allows people to communicate freely in the current situation, where people communicate in many ways.

4.1 Overview

The primitive action-driven recognition method consists of preprocessing (time-series skeletal feature extraction), multiple primitive action recognition modules, and a primitive action composition module, as shown in Figure 2.

The preprocessing extracts time-series skeletal feature data from the input sign language video data. The time-series skeletal feature data recognizes basic hand movements in multiple primitive action recognition modules.

The multiple primitive action recognition modules recognize basic hand movements. We construct each primitive action recognition module that recognizes the shape of each finger in HamNoSys [6][27], as shown in Figure 1. This module recognizes the basic action of each hand from the time-series skeletal features and converts it into symbols specified by HamNoSys called sign language spelling.

The primitive action composition module determines the actual sign meaning from the combination of recognition results from multiple primitive action recognition modules by using a sign language spelling knowledge base. When introducing a new sign language system, this method can be implemented simply by adding the combinations of primitive actions and their meanings to the sign language spelling knowledge base in the primitive action composition module.

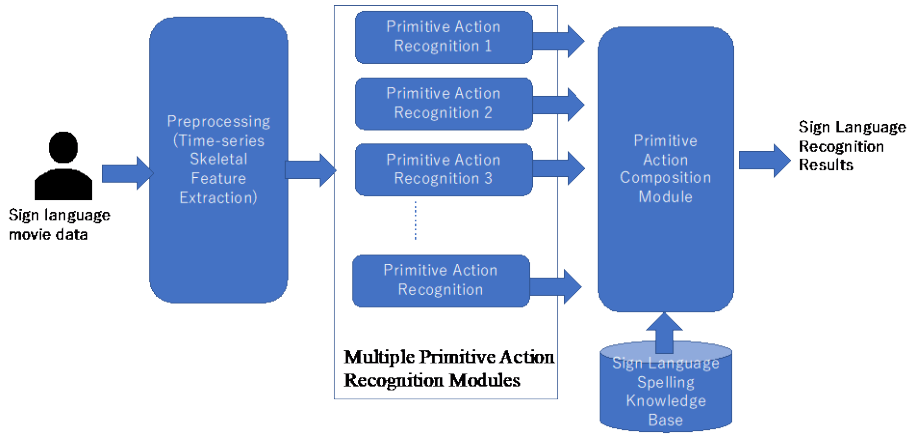

Figure 2: An overview of a primitive action-driven recognition method. This method consists of preprocessing (time-series skeletal feature extraction), multiple primitive action recognition modules, and a primitive action composition module. The multiple primitive action recognition modules recognize basic hand movements. The primitive action composition module integrates the recognition results of the multiple primitive action recognition modules according to the description in the sign language spelling knowledge base to infer the meaning of the input sign language.

Source: own.

The sign language spelling knowledge base consists of a sequence of symbols called sign language spellings defined by HamNoSys and a word. When users can use HamNoSys as a reference for sign spelling and the hand shapes that make up the sign language, they can easily add new words to this knowledge base. This is the most important feature of this method. Most previous sign language recognition methods required enough labeled video content representing sign language as training data. However, our works [1][2][3][4] have shown that it is difficult to collect enough sign language video content to apply existing machine learning methods.

Furthermore, according to the reference [5], to apply various sign languages worldwide, it is necessary to realize as many as 400 different sign languages into an integrated system. Furthermore, according to the reference [5], to apply various sign languages worldwide, it is necessary to realize as many as 400 different sign languages into an integrated system. An existing study used to implement a simplified knowledge base description method is HamNoSys [6][27]. The HamNoSys is a transcription system for all sign languages with a direct correspondence between symbols and gesture aspects, such as hand location, shape, and movement. The

handshape descriptions should include all handshapes used in sign language worldwide. HamNoSys can be applied internationally because it does not refer to nationally diversified finger figures. When we can create a sign language spelling knowledge base of sign language spelling and word pairs using HamNoSys for each of the various sign languages, we can realize an integrated sign language recognition system across the different sign languages.

4.2 Preprocessing (Time-series Skeletal Feature Extraction)

The preprocessing extracts time-series skeletal features representing both hands' positions each time from sign language video data. Figure 3 shows the detail of the time-series feature extraction modules.

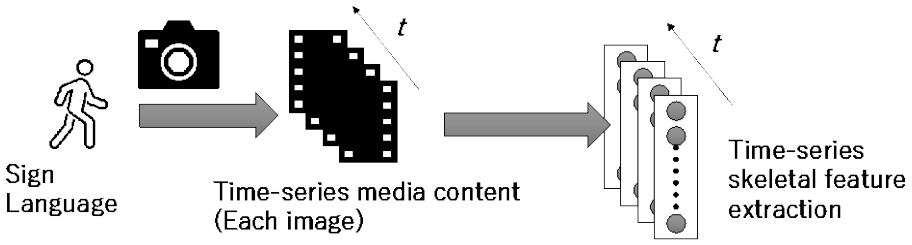

Figure 3: Preprocessing (Time-series Skeletal Feature Extraction)
Source: own.

First, it converts the input sign language video data into a set of images at each time as the time-series media content set. Next, it extracts features representing both hands' positions in each image. Through this process, we can obtain time-series multiple features at each time. In this paper, we apply Mediapipe [28] to feature extraction. The Mediapipe can extract hands, faces, arms, and body parts skeletal features. This paper uses landmarks of both hands' parts as features. The Mediapipe extracts each normalized position (x,y,z) data of 42 landmarks from each image. We can obtain 126 features each time as time-series features. Therefore, it generates a $126 \times t$ time-series feature matrix. This matrix shows the 126 features of the motion extracted from the sign language represented in the input video and their temporal variation.

4.3 Multiple Primitive Action Recognition Modules

The multiple primitive action recognition modules recognize basic hand movements, as shown in Figure 4. We construct each primitive action recognition module that recognizes the shape of each finger in HamNoSys [6][27].

From the time-series skeletal features extracted by preprocessing, all primitive action recognition modules are executed for each time, and the corresponding primitive actions are derived. In other words, this module assigns a single symbol by HamNoSys that represents the hand movement each time. We can obtain the representation of recognition results for each frame as a sequence of HamNoSys symbols. The symbol sequence extracted for each time (frame) have duplicates of the same symbol. The system deletes consecutive identical symbols.

Through these modules, we can obtain symbol sequences that appear in HamNoSys from time series skeletal feature data.

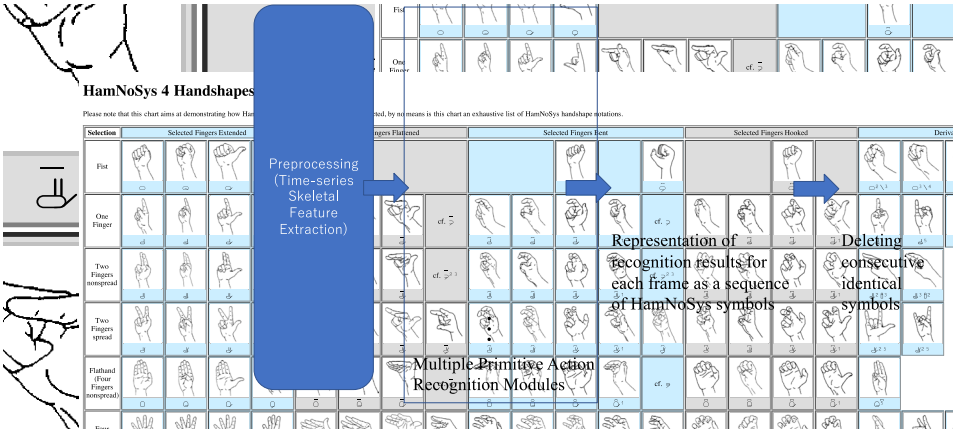

Figure 4: An overview of multiple primitive action recognition modules
Source: own.

4.4 Sign Language Spelling Knowledge Base

The sign language spelling knowledge base consists of a sequence of symbols called sign language spellings defined by HamNoSys and a word. When users can use HamNoSys as a reference for sign spelling and the hand shapes that make up the

sign language, they can easily add new words to this knowledge base. This is the most important feature of this method.

Figure 5 shows how to compose sign language spelling. In general, sign language consists of multiple hand gestures. The sign language spelling describes what handshapes are in what order. Figure 5 shows an example of the sign meaning hello in Japanese. The sign language meaning hello is performed by a hand gesture with the index and middle fingers raised, followed by a hand gesture with the index and middle fingers bent. Each hand shape is assigned a symbol that is determined within HamNoSys. The Sign language spelling is represented by one or more symbol sequences denoted by HamNoSys.

By applying the same methodology, creating a knowledge base for sign language recognition worldwide is possible. The HamNoSys set of finger shapes is common in the world's sign languages. By introducing such sign language spelling, building a knowledge base with simple descriptions is possible without creating or collecting new sign language videos.

The sign language spelling knowledge base consists of a sequence of symbols called sign language spellings defined by HamNoSys and a word. This knowledge base can be created for each different sign language system. Table 1 shows an example of the sign language spelling knowledge base.

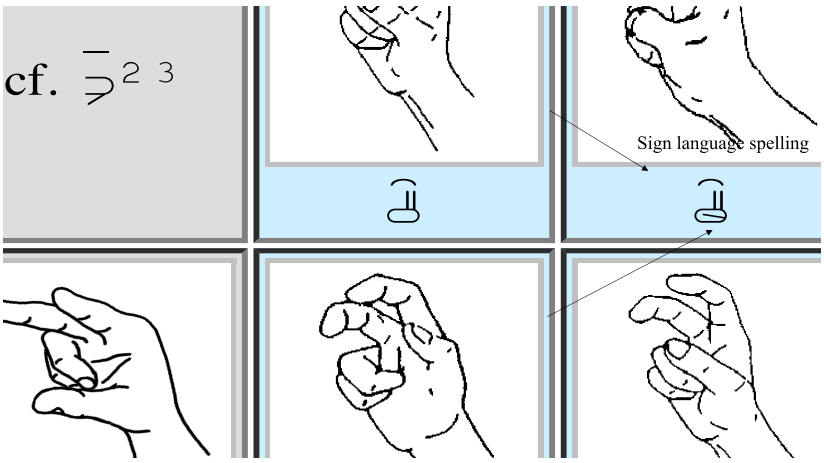

Figure 5: How to compose sign language spelling

Source: own.

Table 1: An example of the sign language spelling knowledge base. The sign language spelling knowledge base consists of a sequence of symbols called sign language spellings defined by HamNoSys and a word.

4.5 Primitive Action Composition Module

The primitive action composition module derives appropriate words by matching the symbol sequences extracted by the multiple primitive action recognition modules with each sign language spelling in the sign language spelling knowledge base.

The primitive action composition module must weigh the similarity between sign language spellings and symbol sequences. It is possible to derive a word that matches the sign by comparing the sequence of symbols extracted by the multiple primitive action recognition modules with the sign spelling in the sign language spelling knowledge base using the Levenshtein distance. In this paper, we apply Levenshtein distance to weigh the similarity between sign language spellings. Levenshtein distance is one of the edit distances and is defined as the minimum number of times required to transform one string into another.

5 Conclusion

This paper represented a primitive action-driven recognition method for the realization of global heterogeneous sign language recognition. This method defines common primitive actions of various sign language systems worldwide. It describes what the combination of these primitive actions indicates in various sign language systems to realize sign language recognition. This method will realize a new global communication platform that avoids the communication divide and allows people

to communicate freely in the current situation, where people communicate in many ways.

We apply our proposed method to create a new global communication platform and develop our method to bridge diverse communities. It is important to seamlessly bridge between diverse speaking communities (including the sign language speaker community). Our proposed method realizes the platform bridging diverse communities.

We will realize a new function that recognizes words and sentences in sign language as our future work. We will apply our method to the sign language of various countries. We need to establish a sign language segmentation method for this to work. Moreover, developing a co-editing environment for a sign language spelling knowledge base is necessary to realize this method, verify its effectiveness using large-scale data, and conduct experiments on subjects with native signers.

References

[1] Nitta, T., Hagimoto, S., Yanase, A., Okada, R., Sornlertlamvanich, V., Nakanishi, T. Realization for Finger Character Recognition Method by Similarity Measure of Finger Features, International Journal of Smart Computing and Artificial Intelligence, Vol. 6 No. 1, 2022.

[2] Hagimoto S, Nitta T, Yanase A, Nakanishi T, Okada R, Sornlertlamvanich V, Knowledge Base Creation by Reliability of Coordinates Detected from Videos for Finger Character Recognition, In proc. of 19th IADIS International Conference e-Society 2021, FSP 5.1-F144, 2021. p.169-176.

[3] Nitta T, Hagimoto S, Yanase A, Nakanishi T, Okada R, Sornlertlamvanich V. Finger Character Recognition in Sign Language Using Finger Feature Knowledge Base for Similarity Measure, In Proceedings of the 3rd IEEE/IIAI International Congress on Applied Information Technology (IEEE/IIAI AIT 2020), 2020.

[4] Nakanishi, T., Minematsu, A., Okada, R., Hasegawa, O., Sornlertlamvanich, V. Sign Language Recognition by Similarity Measure with Emotional Expression Specific to Signers, 32nd International Conference on Information Modelling and Knowledge Bases, 2022.

[5] SIL International (2018a). Sign Languages. https://www.sil.org/sign-languages

[6] Hanke, T. HamNoSys-representing sign language data in language resources and language processing contexts. In: Streiter, Oliver, Vettori, Chiara (eds): LREC 2004, Workshop proceedings: Representation and processing of sign languages. Paris: ELRA; 2004. pp. 1-6.

[7] Adeyanju I. A, Bello O. O, Adegboye M. A. Machine learning methods for sign language recognition: A critical review and analysis. Intelligent Systems with Applications, 2021 12, 200056.

[8] Gupta R, Rajan S. Comparative analysis of convolution neural network models for continuous Indian sign language classification, Procedia Computer Science, 171 2020, pp. 1542-1550.

[9] Chowdhry D.A, Hussain A, Ur Rehman M.Z, Ahmad F, Ahmad A, Pervaiz M. Smart security system for sensitive area using face recognition, Proceedings of the IEEE conference on

sustainable utilization and development in engineering and technology, IEEE CSUDET 2013, pp. 11-14.
[10] Cheok M.J, Omar Z, Jaward M.H. A review of hand gesture and sign language recognition techniques, International Journal of Machine Learning and Cybernetics, 10 (1) 2019, pp. 131-153.
[11] Butt U.M, Husnain B, Ahmed U, Tariq A, Tariq I, Butt M.A, Zia M.S. Feature based algorithmic analysis on American sign language dataset, International Journal of Advanced Computer Science and Applications, 10 (5) 2019, pp. 583-589.
[12] Rastgoo R, Kiani K, Escalera S. Sign language recognition: A deep survey, Expert Systems with Applications, 164 2021, Article 113794.
[13] Lee C.K.M, Ng K.H, Chen C.H, Lau H.C.W, Chung S.Y, Tsoi T. American sign language recognition and training method with recurrent neural network, Expert Systems with Applications, 167 2021, Article 114403.
[14] Huang, Y., Monekosso, D., Wang, H., Augusto, JC. A hybrid method for hand gesture recognition, 2012 Eighth International Conference on Intelligent Environments, Guanajuato, Mexico, June 2012. pp. 297–300.
[15] Yuan, S. et al. Chinese sign language alphabet recognition based on random forest algorithm. 2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT, June 2020. pp. 340–344.
[16] Izzah, A., Suciati, N. Translation of sign language using generic fourier descriptor and nearest neighbour. IJCI, vol. 3, no. 1, February 2014. pp. 31– 41.
[17] Raheja, JL., Mishra, A., Chaudhary, A. Indian sign language recognition using SVM, Pattern Recognit. Image Anal., vol. 26, April 2016. pp. 434–441.
[18] Wong, SF., Cipolla, R.Real-time adaptive hand motion recognition using a sparse bayesian classifier. Computer Vision in Human-Computer Interaction, Berlin, Heidelberg, 2005, pp. 170–179.
[19] Mahmud, I., Tabassum, T., Uddin, Md.P., Ali, E., Nitu, AM., Afjal, MI. Efficient noise reduction and HOG feature extraction for sign language recognition. 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), 2018. pp. 1–4.
[20] Hartanto, R., Susanto, A., Santosa, P.I., Real time static hand gesture recognition system prototype for Indonesian sign language. 2014 6th International Conference on Information Technology and Electrical Engineering, Yogyakarta, Indonesia, 2014. pp. 1–6.
[21] Anand, MS., Kumar, NM., Kumaresan, A. An efficient framework for Indian sign language recognition using wavelet transform. Circuits and Systems, vol. 07, no. 8, June 2016. pp. 1874–1883.
[22] Hasan,MM., Khaliluzzaman,Md., Himel,SA., Chowdhury, RT. Hand sign language recognition for Bangla alphabet based on Freeman Chain Code and ANN. 2017 4th International Conference on Advances in Electrical Engineering (ICAEE), Dhaka, September 2017. pp. 749–753.
[23] Athira, PK., Sruthi, CJ., Lijiya, A. A signer independent sign language recognition with co-articulation elimination from live videos: an Indian scenario. J. King Saud. Univ. - Comput. Inf. Sci., vol. 34, no. 3, March 2022. pp. 771–778.
[24] Aloysius, N., Geetha, M., A scale space model of weighted average CNN ensemble for ASL fingerspelling recognition. Int. J. Comput. Sci. Eng., vol. 22, no. 1, May 2020. pp. 154–161.
[25] Noroozi F, Kaminska D, Corneanu C, Sapinski T, Escalera S, Anbarjafari G. Survey on emotional body gesture recognition. IEEE transactions on affective computing, 12(02), 2021. p. 505-523.
[26] Kiyoki, Y., Hosokawa, Y., Ishibashi, N. A metadatabase system architecture for integrating heterogeneous databases with temporal and spatial operations." Advanced Database Research and Development Series 10, 2000. pp. 158-165.
[27] HamNoSys Handshapes,

[28] https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HamNoSys_Handshapes.pdf
[29] Mediapipe, https://google.github.io/mediapipe/