Information Modelling and Knowledge Bases XXXIV
M. Tropmann-Frick et al. (Eds.)

© 2023 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA220490

Sign Language Recognition by Similarity Measure with Emotional Expression Specific to Signers

Takafumi NAKANISHI [a,b,1],Ayako MINEMATSU [b], Ryotaro OKADA [a,b], Osamu HASEGAWA [a,b] and Virach SORNLERTLAMVANICH [a,b]
[a] *Department of Data Science, Musashino University*
[b] *Asia AI Institute, Musashino University*

Abstract. Through technology, it is essential to seamlessly bridge the divide between diverse speaking communities (including the signer (the sign language speaker) community). In order to realize communication that successfully conveys emotions, it is necessary to recognize not only verbal information but also non-verbal information. In the case of signers, there are two main types of behavior: verbal behavior and emotional behavior. This paper presents a sign language recognition method by similarity measure with emotional expression specific to signers. We focus on recognizing the sign language conveying verbal information itself and on recognizing emotional expression. Our method recognizes sign language by time-series similarity measure on a small amount of model data, and at the same time, recognizes emotion expression specific to signers. Our method extracts time-series features of the body, arms, and hands from sign language videos and recognizes them by measuring the similarity of the time-series features. In addition, it recognizes the emotional expressions specific to signers from the time-series features of their faces.

Keywords. Sign Language Recognition, Similarity Measure, Emotional Behavior, Diverse Speaking Communities

1. Introduction

It has become increasingly important to understand how diverse people can work together to contribute to society in recent years. Smooth communication is the most necessary for diverse people to work together. In order to achieve smooth communication, we are focusing on the implementation of new global communication methods. In particular, we expect information technology to realize smooth communication across heterogeneous languages and sign languages.

Through technology, it is essential to seamlessly bridge the divide between diverse speaking communities (including the signer (the sign language speaker) community). In order to realize communication that successfully conveys emotions, it is necessary to recognize not only verbal information but also non-verbal information. In the case of sign language speakers, there are two main types of behavior: verbal behavior and

[1] Corresponding Author, Department of Data Science, Musashino University, 3-3-3 Ariake Koto-ku Tokyo 135-8181, Japan; E-mail: takafumi.nakanishi@ds.musashino-u.ac.jp.

emotional behavior. Verbal behavior requires technology to recognize the sign language itself, and emotional behavior requires technology for recognizing emotions. By realizing both behavior recognition, we realize semantic mutual understanding and emotional mutual understanding, which can utilize as a new global communication method. Therefore, it is essential to extract the features of verbal and emotional behavior and to realize a sign language recognition method that combines them.

Some signers say that sign language is less burdensome for emotional communication than written communication. Written communication is one of the ways of communication by sign language speakers with speakers of other languages. Written communication is a very effective way of conveying meaning. However, it is difficult to convey emotions to the other person through written communication alone. In order to solve these problems, it is essential to realize a sign language recognition and composition system that can reflect not only verbal behavior but also emotional behavior features. The realization of a place where diverse people can communicate smoothly requires a platform with a system that recognizes, composes, and expresses meanings and emotions. This platform can seamlessly bridge the divide between diverse speaking communities (including the signer (the sign language speaker) community). Our new method of global communication can be the core technology to realize this platform.

This paper presents a sign language recognition method by similarity measure with emotional expression specific to signers. We focus on recognizing the sign language conveying verbal information itself and on recognizing emotional expression. Our method recognizes sign language by time-series similarity measure on a small amount of model data, and at the same time, recognizes emotion expression specific to signers. Our method extracts time-series features of hands from sign language videos and recognizes them by measuring the similarity of the time-series features. In addition, it recognizes the emotional expressions specific to signers from the time-series features of their faces.

Our method consists of a hand gesture recognition function, a facial expression recognition function, and a recognition result integration function. The hand gesture recognition function enables recognizing sign language by time-series similarity measure. It is the ability to realize the semantic recognition that sign language represents. The facial expression recognition function recognizes facial expressions evolve to communicate intentions. The recognition result integration function integrates hand gesture recognition results and facial expression recognition function results.

The main features of this paper are as follows:

- We propose a new sign language recognition method by similarity measure from extracted features of the hands from sign language videos.
- We also propose a recognition method for emotional expression specific to signers from extracted features of the face from sign language videos.
- We position our proposed method as a new global communication method and develop our method to bridge over diverse communities.

This paper is organized as follows. In section 2, we present some related works of our method. Section 3 presents our proposed method, a sign language recognition method by similarity measure with emotional expression specific to signers. In section 4, we represent the result of similarity measures for recognizing sign language. In addition, in the section, we also present the extraction of emotional expression specific to signers. Finally, in section 5, we summarize this paper.

2. Related Works

The reference [1] provides a research survey on recognizing emotions from body gestures. This paper [1] assumes that body gesture-based emotion recognition generally consists of human detection, body pose detection, representation learning, and emotion recognition. This paper shows some research in each function.

We focus on verbal and emotional behavior to recognize sign language with emotion. Our method not only recognizes signs that convey actual meaning but also recognizes signer-specific expressions that express emotion. According to reference [2], sign language recognition alone can only recognize 35% of the content to convey, and non-verbal signals are essential for smooth communication. The reference [2] presents that the final message of an utterance is affected 35% by the actual words and 65% by non-verbal signals.

The reference [3] present a survey of machine learning methods applied in sign language recognition systems. This reference [3] says that sign language involves the usage of the upper part of the body, such as hand gestures [4], facial expression [5], lip-reading [6], head nodding and body postures to disseminate information [7] [8] [9]. 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 expression as both verbal and emotional behavior. In this paper, we present a new hand gesture and facial expression recognition method.

Our previous works [10][11] present finger character recognition in sign language using a finger feature knowledge base for similarity measures. Especially, our previous work [10] presents finger character feature extraction by combining a camera and a deep learning model for extracting finger joint coordinates. In this paper, we apply our previous works [10][11] and Mediapipe [12] to time-series feature extraction for sign language recognition. Existing methods for time-series feature extraction for sign language recognition are broadly classified into direct feature extraction using a glove with multiple sensors [13][14], indirect feature extraction using a depth sensor [15], and indirect feature extraction using a monocular camera [16][17][18]. In the case of methods that use gloves equipped with multiple sensors, while the method extracts accurate features, it can be inconvenient to communicate with gloves in daily life. In the case of methods that use depth sensors, it is also necessary to prepare depth sensors. In the case of methods using a monocular camera, the accuracy is lower than the first two methods, but we can smoothly introduce it in daily life. Recently, monocular camera-based methods such as Mediapipe [12] make extracted features high accuracy and ease.

The reference [19] creates two datasets: a synthetic motion dataset for model training and a dataset containing human annotations of real-world video clip pairs for motion similarity evaluation for human motion similarity measures. The reference [19] points out that measuring motion similarity has attracted less attention due to the lack of large datasets. In the case of realization of sign language recognition, it is difficult to apply the standard machine learning method due to the lack of large datasets like the above case.

Our method enables sign language recognition by similarity measure between the input sign language features and a miniature model data set. The sign language features represent some time series data. DTW [20][21][22] is a method to measure the similarity of time series data. DTW can robustly determine the distance between two sets of time-series data with different lengths, such as waveforms with different frequencies. Our previous work [23] proposes a similarity measure for time series semantic data based on

DTW [20][21][22]. It enables the calculation of the similarity of media content based on time-series changes.

3. Sign Language Recognition Method by Similarity Measure with Emotional Expression Specific to Signers

3.1. Overview of our proposed method

This paper presents a sign language recognition method by similarity measure with emotional expression specific to signers. We focus not only on recognizing the sign language itself but also on expressing emotions. Our method recognizes sign language by time-series similarity measure on a small amount of model data, and at the same time, recognizes emotion specific to signers.

Figure 1 shows an overview of our proposed method. Our method consists of a hand gesture recognition function, facial expression recognition function, and a recognition result integration function. The hand gesture recognition function enables recognizing sign language by time-series similarity measure. It is the ability to realize the semantic recognition that sign language represents. It recognizes sign language by time-series similarity measure on a small amount of model data. It extracts time-series features of the body, arms, and hands from sign language videos and recognizes them by measuring the similarity of the time-series features. The facial expression recognition function recognizes emotional behaviors other than sign language. This paper focuses on recognizing emotional behaviors as expressions of emotion-specific to signers. Especially, we realize facial expression recognition specific to signers in this paper. The recognition result integration function integrates hand gesture recognition results and facial expression recognition function results.

Our method extracts time-series features of the hands from sign language videos and recognizes them by measuring the similarity of the time-series features. In addition, it recognizes the emotional expressions specific to signers from the time-series features of their faces.

In section 3.2, we describe the realization method of the hand gesture recognition function. In section 3.3, we show the one of the solutions of facial expression recognition function focusing on the facial expression specific to signer. In section 3.4, we represent the recognition result integration function.

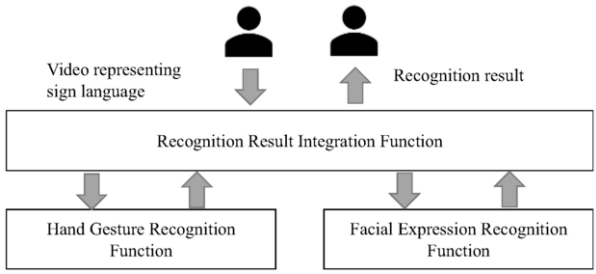


Figure 1. Overview of our proposed method.
Our method consists of a hand gesture recognition function, facial expression recognition function, and a recognition result integration function.

3.2. Hand gesture recognition function

3.2.1. Overview of hand gesture recognition function

Figure 2 shows a structure chart of the hand gesture recognition function. The hand gesture recognition function consists of metadata creation and recognition phases. The metadata creation phase takes as input a video of a sign language model and a label representing the semantic of the sign language, extracts time-series waveform features from the videos, and stores them in the sign language glossary. The recognition phase inputs a video of sign language, extracts time-series waveform features from the videos and measures waveform similarity between time-series features from the input video and stored waveform features in the sign language glossary. This function recognizes sign language by time-series similarity measure on a small amount of model data. It extracts time-series features of the body, arms, and hands from sign language videos. This function determines the semantic of a sign language by selecting the one with the highest similarity between the time-series waveform features extracted from the sign language example video, and the time-series waveform features extracted from the input sign language video.

There are two modules in the hand gesture recognition function: a time-series feature extraction module appearing in both the metadata creation phase and recognition phase and a waveform similarity measure module appearing in the recognition phase. The time-series feature extraction module extracts each normalized position (x,y,z) of 42 landmarks from both hands in each time as the time-series waveform features. The waveform similarity measure module calculates the similarity between each time-series waveform feature extracted from the time-series feature extraction module. We realize the waveform similarity by applying our previous method [23] based on DTW [20][21][22]. It enables the calculation of the similarity of media content based on time-series changes. Implementing two modules makes it possible to perform a similar search for sign languages as waveforms. In addition, it is possible to visualize time-series variance by time-series change of sign languages. Measuring the similarity of time-series waveform features extracted from each time-series feature extraction makes it possible to derive a similarity corresponding to the dynamism of the sign languages. The similar search determines semantic of the input sign language.

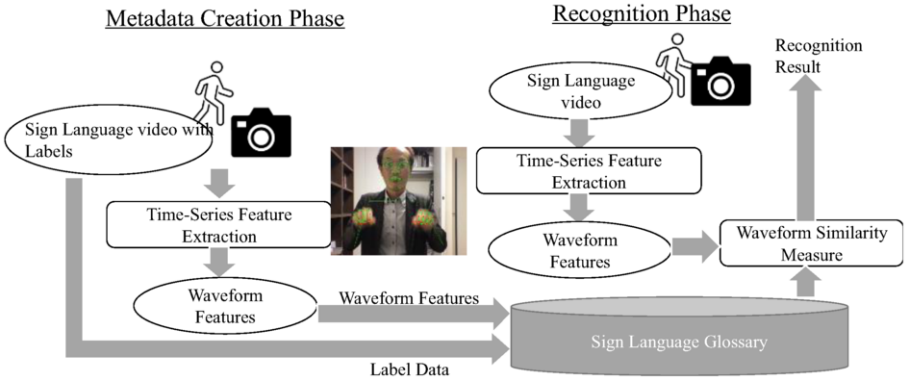


Figure 2. Structure chart of hand gesture recognition function.

3.2.2. Time-series feature extraction module in hand gesture recognition function

The time-series feature extraction module extracts waveform features representing both hands' positions each time from sign language video data.

Figure 3 shows the detail of the time-series feature extraction modules.

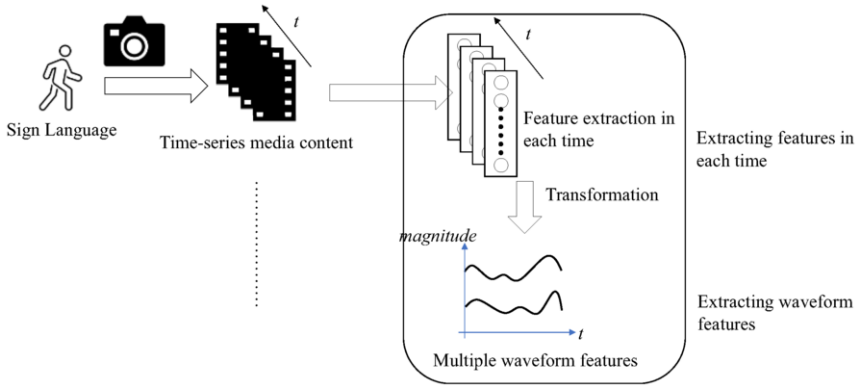


Figure 3. Time-series feature extraction module.

First, it converts to the input sign language video data into a set of images in each time as the time-series media content set.

Next, it extracts features representing both hands' positions in each image. By this process, we can obtain time-series multiple features at each time. In this paper, we apply Mediapipe [12] to feature extraction. The Mediapipe can extract hands, faces, arms, and body parts landmarks. 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 shown in Figure 4. This matrix shows the 126 features of the motion extracted from the sign language represented in the input video and their temporal variation. This matrix represents motion transitions for sign language.

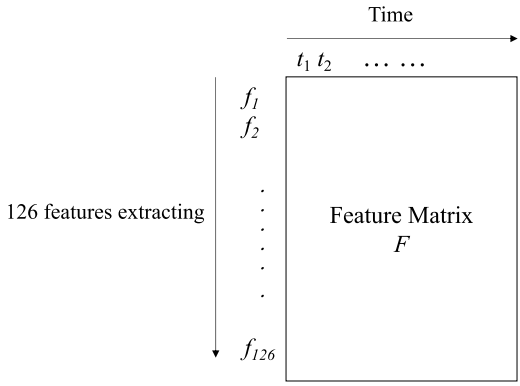


Figure 4. Time-series feature matrix.
The Mediapipe extracts each normalized position (x,y,z) data of 42 landmarks from each time. This matrix shows the 126 features extracted from the sign language represented in the input video and their temporal variation.

Finally, it creates multiple waveform features. It performs a moving average over each time-series feature extracted from Mediapipe, resulting in 126 waveform features. These waveform features represent time-series changes for recognition of sign language. Figure.5 shows the detail of this process.

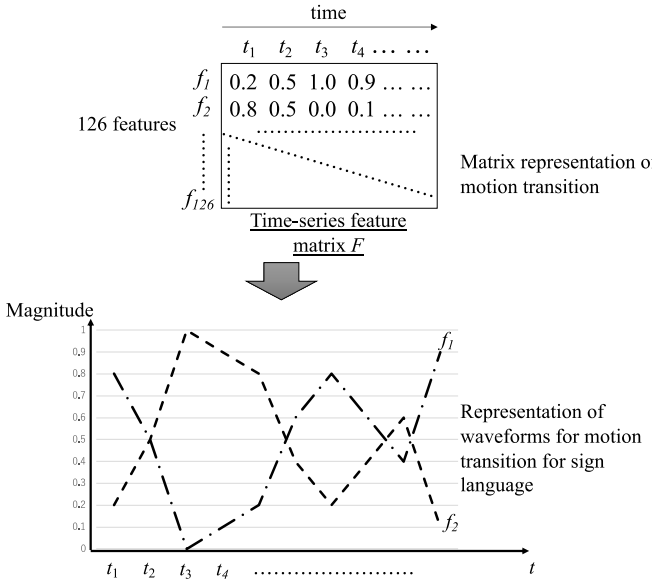


Figure 5. Waveform features creation.
This module generates 126 waveform features from the $126 \times t$ time-series feature matrix shown in Figure 4.

We can obtain 126 waveforms features that present sign language motion transitions by these processes.

In the case of the metadata creation phase in Figure 2, we prepare some sets of sign language videos with each label as a model data set. The time-series feature extraction module generates 126 waveforms features in each model data set and stores them in the sign language glossary. In the case of the recognition phase in Figure 2, we input a sign language video. the time-series feature extraction module generates 126 waveforms feature from the input video. In this method, all sign language videos are converted to 126 waveforms features by the time-series feature extraction module.

The time-series feature extraction module consists of the following three steps:

(1) Dividing video data according to a specified window size
It divides the input video data at regular intervals according to a specific window size to obtain time-series characteristics from the video data. By this step, we can obtain image data set.

(2) Creating feature matrix
It extracts 126 features representing both hands' positions in each image using Mediapipe. Therefore, it generates a time-series feature matrix shown in Figure 4. This matrix shows the 126 features extracted from the sign language represented in the input video and their temporal variation. This matrix represents motion transitions for sign language.

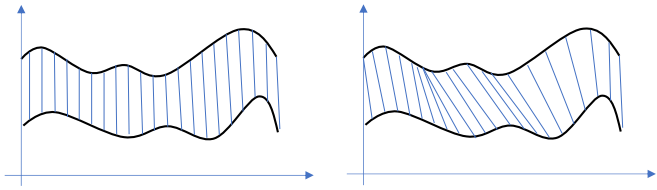
(3) Representing waveforms as motion transition

It represents multiple waveforms as motion transition. It performs a moving average over each time-series feature extracted from Mediapipe, resulting in 126 waveform features. These waveform features represent time-series changes for the recognition of sign language. When these values are plotted on the time axis, waveforms are visualized for each feature. We define the multiple waveforms as motion transition.

3.2.3. Waveform similarity measure module

The waveform similarity measure module is possible to show the relationship between each sign language based on time-series development by calculating the similarity between the motion transitions represented as waveforms. It realizes new time-series similarity by comparison with motion transitions represented by waveforms applying a signal processing technique— the DTW distance [20][21][22].

DTW is a pattern-matching technique used for one of the signal processing techniques, such as voice recognition, and so on. DTW can robustly determine the distance between two sets of time-series data with different lengths, such as waveforms with different frequencies. DTW finds the path where the two sets of time-series data are the shortest after calculating the distance between each point of the two sets of time-series data by brute force. Figure 6 shows the difference between Euclidean matching and dynamic time warping matching. DTW matching minimizes cumulative distance measurement consisting of local distances between aligned samples.



(a)Euclidean Matching (b)DTW Matching

Figure 6. Different matching of two similar time series. (a) is Euclidean matching. (b) is dynamic time warping matching. The DTW matching minimizes cumulative distance measurement consisting of local distances between aligned samples.

Using the waveform similarity measurement makes it possible to realize sign language retrieval corresponding to the time-series motion transition of sign language. Using the waveform similarity measurement makes it possible to realize sign language retrieval corresponding to the time-series motion transition of sign language. Using a search mechanism based on waveform similarity measure, the sign language recognition method enables sign language recognition even from a small amount of sign language model data. It is tough to create a large amount of sign language model data. We would like to provide a platform for heterogeneous signers to communicate. In order to realize this platform, it is necessary to create various sign language model datasets. Therefore, sign language recognition methods must be feasible with small sign language model data sets.

Figure 7 shows the realization methods for the waveform similarity measurement module. This module consists of two steps.

(1) Computing DTW distance in each waveform feature

The time-series feature extraction module outputs waveforms as motion transition from time-series media content. It computes the DTW distance in each waveform feature. This means computing the similarity of each single waveform.

(2) Composition of each DTW distance

It composes each DTW distance derived in (1). The composition value is a multiple waveform similarity value for motion transition.

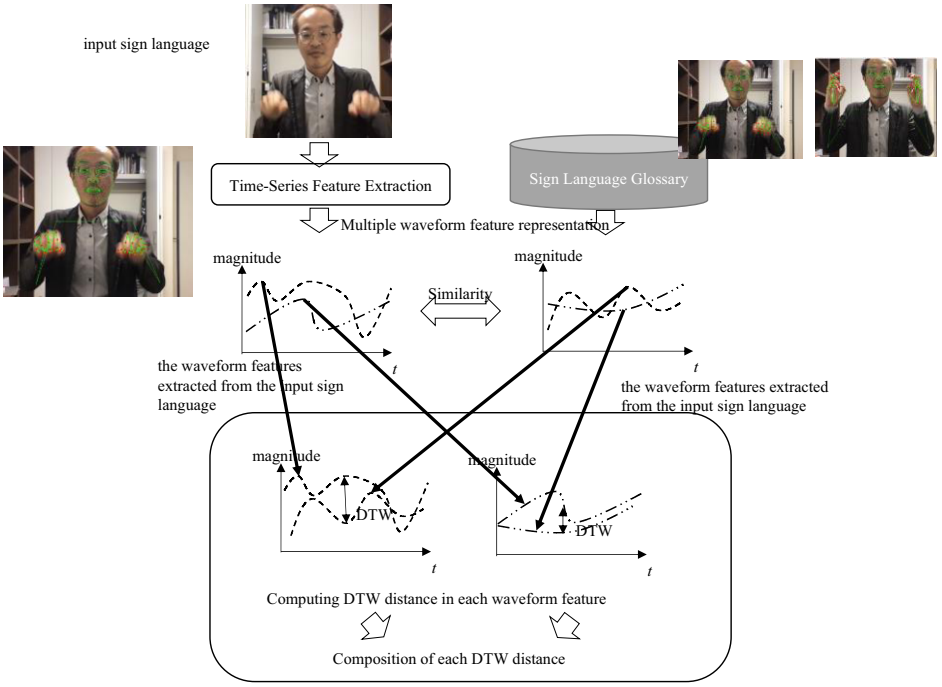


Figure 7. An overview of the waveform similarity measure module. The time-series feature extraction module outputs waveforms as motion transition from time-series media content. In order to compute similarity of motion transition, the module computes DTW distance in each waveform feature. Finally, the module composes each DTW distance. The composition value is a waveform similarity measurement for each motion transition.

3.3. Facial expression recognition function

3.3.1. Overview of facial expression recognition function

Figure 8 shows a structure chart of the facial expression recognition function. It consists of an event definition phase and recognition phase. The event definition phase defines the emotion to be recognized as an emotional event. Each event is stored in the emotion event meta database. The recognition phase inputs a video extracts time-series waveform features from the videos, and detects emotion events by the emotion event meta database.

There are two ways in the event definition phase: a rule based definition and a machine learning based definition. The rule based definition is used when a change in the features reliably defines the expression of the emotion. The machine learning based definition is used when the expression of emotion cannot be described by rules but can

be derived by learning from a large amount of model data. For emotional expressions such as standard facial expressions, we can use previous research such as facial expression recognition (e.g., Reference [24]). Our interviews with have confirmed the existence of several signer-specific emotional expressions. For example, a squinting expression indicates that the signer emphasizes the expression. By defining the emotional events of the signer's specific emotional expression, it is possible to convey the content of the sign language more accurately.

The recognition phase consists of a time-series feature extraction module and a emotion event detection module. The time-series feature extraction module extracts each normalized position (x,y,z) of landmarks from face, body, both foots, both arms and both hands in each time as the time-series waveform features. The emotion event detection module uses the emotion event meta database to detect emotion events.

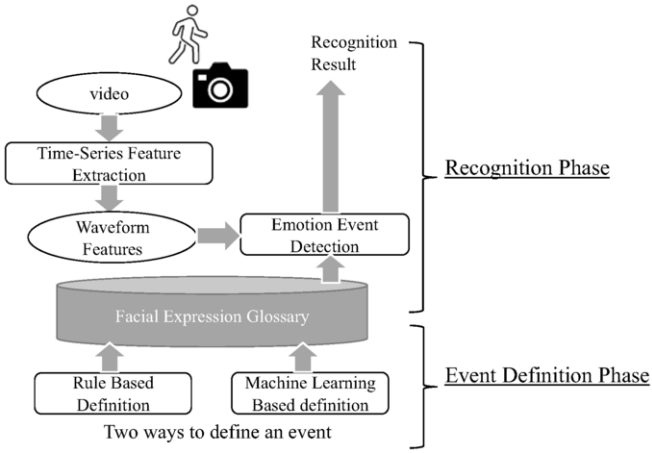


Figure 8. Structure chart of emotion recognition function.

3.3.2. Event definition phase

There are two ways in the event definition phase: a rule based definition and a machine learning based definition.

The rule based definition labels changes in time-series features. It is often possible to link changes in facial expressions and motion feature points to their semantic. When defining an event to detect that squinting implies emphasis, it should be possible to detect when a certain amount reduces the area occupied by the eyes.

The machine learning based definition consists of a training dataset consisting of time-series features and their labels. Unlike rule-based definitions, where changes in facial features and their meanings cannot be clearly labeled, definitions are made using large amounts of training data.

Here, in the case of event definitions for detecting signer-specific expressions, it is considered that the rule based definitions can be used in most cases. The signer-specific expressions of emotion are derived from interviews with signers and the research results by sign language experts. In other words, we need to realize a common understanding of emotional expression for signers on an ad hoc basis. We define a common understanding of each signer's specific emotional expression as each emotion event. In order to define as many signer-specific emotional expressions as possible as emotion events, it may be

necessary to implement a collaborative editing environment for emotion event definitions.

3.3.3. Time-series feature extraction module in facial expression recognition function

The time-series feature extraction module extracts each normalized position (x,y,z) of landmarks from face, body, both foots, both arms and both hands in each time as the time-series waveform features. In this paper, we apply Mediapipe [12] to feature extraction as same as the hand gesture recognition function. It uses all the hands, faces, arms, foots and body parts landmarks extracted by Mediapipe.

3.3.4. Emotion event detection

The emotion event detection module uses the emotion event meta database to detect emotion events. Emotion event outputs a recognition result only when an event is detected. For example, in the case of an emotional event indicating emphasis by squinting, the label "emphasis" is output as a recognition result only when the eyes are squinted. The waveform features extracted from the time-series feature extraction module described in section 3.3.3 are used as input for event detection.

3.4. Recognition result integration function

The recognition result integration function integrates hand gesture recognition results and facial expression recognition function results. For example, if an emotional event indicating emphasis by squinting is detected and a sign meaning "long time no see" is detected, the result "very long time no see" is output. However, there are cases where it is better to separate the detected verbal behaviors from the emotion event. For example, in the case of clear emotional events such as "happy" or "sad," the semantic may be inaccurate if mixed with the results of sign language detection.

Visualizing and presenting these recognition results to the user is future works.

4. Experiment

In this section, we implement the hand gesture recognition function shown in section 3.2 and a part of the facial expression recognition function shown in 3.3 and verify the recognition result.

4.1. Experiment 1 (Experiment on sign language glossary lookup using sign language video query)

4.1.1. Experiment environment

We prepared each two sign language videos, each for "good morning," "good afternoon," "good evening," and "long time no see," in Japanese sign language as model data are shown in Figure 9. The sign language videos shown in Figure 9 stored in the sign language glossary. We verified whether the waveform similarity metric correctly recognized the videos of "good morning," "good afternoon," "good evening," and "long time no see" taken from different angles from the prepared model data as query sign language videos. Figure 10 shows query sign language videos and their IDs.

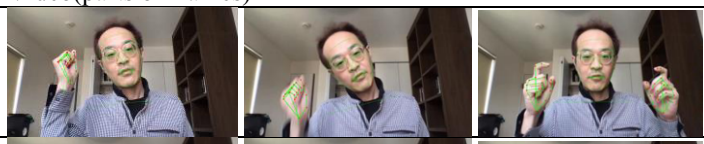
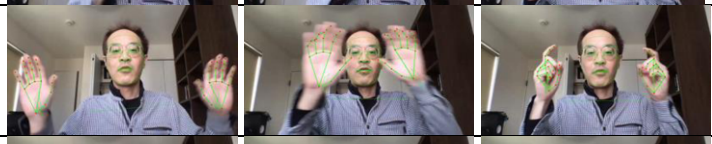
Video ID	Video(parts of frames)
Good morning1	
Good morning2	
Good afternoon1	
Good afternoon2	
Good evening1	
Good evening2	
Long time no see1	
Long time no see 2	

Figure 9. Each two sign language videos, each for "good morning," "good afternoon," "good evening," and "long time no see," as model data

4.2. Experimental results

We show waveform similarity measure results shown in Table 1 (in the case of "Good morning query"), Table 2 (in the case of "Good afternoon query"), Table 3 (in the case of "Good evening query"), and Table 4 (in the case of "Long time no see query"). The Waveform similarity measure shown in section 3.2.3 determines that the smaller the value, the higher the similarity.

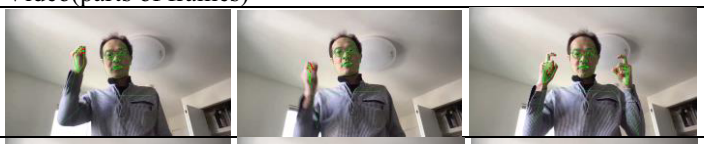
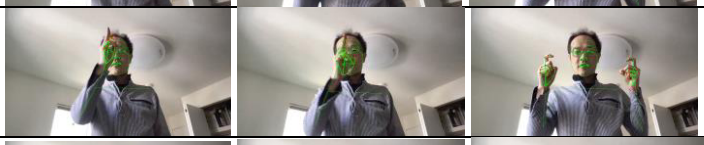
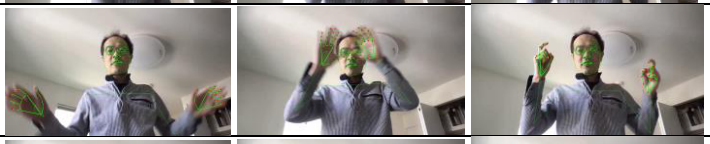
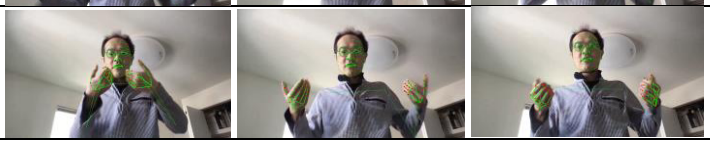
Video ID	Video(parts of frames)
Good morning query	
Good afternoon query	
Good evening query	
Long time no see query	

Figure 10. Query sign language videos and their IDs.

In the case of “Good morning query” shown in Table1, the system shows that the similarity between the input and “Good morning2” and “Good morning1” is high. Therefore, the sign language indicated by “Good morning query” is correctly recognized as “Good morning”.

In the case of “Good afternoon query” shown in Table2, the system shows that the similarity between the input and “Good afternoon1” and “Good afternoon2” is high. Therefore, the sign language indicated by “Good afternoon query” is correctly recognized as “Good afternoon”.

In the case of “Good afternoon query” shown in Table2, the system shows that the similarity between the input and “Good afternoon1” and “Good afternoon2” is high. Therefore, the sign language indicated by “Good afternoon query” is correctly recognized as “Good afternoon”.

In the case of “Good evening query” shown in Table3, the system shows that the similarity between the input and “Good evening2” and “Good evening1” is high. Therefore, the sign language indicated by “Good evening query” is correctly recognized as “Good evening”.

In the case of “Long time no see query” shown in Table4, the system shows that the similarity between the input and “Long time no see1” is high. Therefore, the sign language indicated by "Long time no see query" is correctly recognized as "Long time no see". However, the system indicates that the similarity of “Good evening1” is also high. The sign language for “Good evening” and “long time no see” has an arm-opening motion; therefore, the system measures high similarity.

The introduction of methods such as dimension deduction is one of the future works. This method tends to be computationally expensive. Moreover, it is necessary to create more sign language example data and verify the effectiveness of using them.

Table 1. Waveform similarity measure results in the case of "Good morning query"

	Data name	Waveform similarity measure
1	Good morning2	166.224711110
2	Good morning1	185.8336289461081
3	Good afternoon1	216.42646559481693
4	Good afternoon2	242.99949708402085
5	Good evening1	536.7560547591917
6	Good evening2	559.1061592315774
7	Long time no see1	586.2445960920552
8	Long time no see2	814.9661315709342

Table 2. Waveform similarity measure results in the case of "Good afternoon query"

	Data name	Waveform similarity measure
1	Good afternoon1	179.90706160093168
2	Good afternoon2	197.8562313570992
3	Good morning2	206.342194446534
4	Good morning1	223.68309196158623
5	Long time no see1	486.0481500020227
6	Good evening1	548.4300859502855
7	Good evening2	564.911794575454
8	Long time no see2	823.1264081739348

Table 3. Waveform similarity measure results in the case of "Good evening query"

	Data name	Waveform similarity measure
1	Good evening2	232.74890747326342
2	Good evening1	236.62293856350536
3	Long time no see1	491.2600778100607
4	Good morning1	492.3505294598326
5	Good morning2	504.9853604666803
6	Long time no see2	558.4608788777249
7	Good afternoon2	599.5930115504634
8	Good afternoon1	599.7239569006395

Table 4. Waveform similarity measure results in the case of "Long time no see query"

	Data name	Waveform similarity measure
1	Long time no see1	316.3062876378202
2	Good evening1	335.19070713716405
3	Long time no see2	336.5217383472059
4	Good evening2	391.3205649517722
5	Good afternoon1	427.5670674213245
6	Good afternoon2	428.2101874432208
7	Good morning1	438.1667977241975
8	Good afternoon1	599.7239569006395

4.3. Experiment 2 (Implementation of the facial expression recognition function)

4.3.1. Experimental environment

We implement the system detecting the expression “squinting”. As an expression of emotion unique to signers, an interview with two signers revealed that the expression “squinting” can mean emphasis. Figure 11 shows test data for the squinting expression detection. The image on the left side of Figure 11 shows the facial expression of a normal eye. The image on the right side of Figure 11 shows a squinted eye expression. The person in this image has narrow eyes even under normal facial expression, so it is not easy to distinguish between them.

Figure 11. Test data for the squinting expression detection.
The image on the left side shows the normal eye expression.
The image on the right side shows the squinted eye expression

4.3.2. Experimental results

We have implemented a system that detects eye regions and extracts squinting expressions by comparing the area of these regions. The Table 5 shows the values of the area of the detected eye region for the left and right images of Figure 11. The area of these regions in the case that the eyes are squinting is smaller. The system can recognize expressions that mean emphasis by squinting by detecting that the eye area is small for a certain amount of time in a row.

Table 5. the values of the area of the detected eye region for the left image (normal eyes) and right image (squinting eyes) of Figure 11

	Normal eyes	Squinting eyes
Left eye	0.00658263	0.00580797
Right eye	0.00667033	0.00585308

The results showed that it was possible to detect one of the emotional expressions unique to signers. We need to realize these signer-specific expressions on an ad hoc basis.

5. Conclusion

This paper presented a sign language recognition method by similarity measure with emotional expression specific to signers. We focus not only on recognizing the sign language itself but also on expressing emotions. Our method recognizes sign language by time-series similarity measure on a small amount of model data, and at the same time, recognizes emotion specific to signers.

We position our proposed method as a new global communication method and develop our method to bridge over diverse communities. It is essential to seamlessly bridge the divide between diverse speaking communities (including the signer (the sign language speaker) community). In order to realize communication that successfully conveys emotions, it is necessary to recognize not only verbal information but also non-verbal information. Our proposed method realizes the platform bridging over diverse communities.

We have implemented a basic sign language recognition and emotion expression recognition system using our method and verified its effectiveness.

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

References

[1] Noroozi F, Kaminska D, Corneanu C, Sapinski T, Escalera S, Anbarjafari G. Survey on emotional body gesture recognition. IEEE transactions on affective computing, 2021 12(02). p. 505-523.
[2] Brow K, Kinesics. Encyclopedia of Language and Linguistics 2nd Edition. Amsterdam, Netherlands: Elsevier Science, 2005.
[3] 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.
[4] 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.
[5] 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.
[6] 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.
[7] 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.
[8] Rastgoo R, Kiani K, Escalera S. Sign language recognition: A deep survey, Expert Systems with Applications, 164 2021, Article 113794.
[9] 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.
[10] 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.
[11] 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.
[12] Mediapipe, https://google.github.io/mediapipe/

[13] Fang G, Gao W, Zhao D. Large vocabulary sign language recognition based on fuzzy decision trees. in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. 2004 34(3), p. 305-314.
[14] Kong WW, Ranganath R, Signing Exact English (SEE): Modeling and recognition. *Pattern Recognition*. 2008 41(5), p. 1638-1652.
[15] Cem K, Furkan K, Yunus K, Lale A, 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. 2012 p. 852-863.
[16] Gu L, Yuan X, Ikenaga T. Hand gesture interface based on improved adaptive hand area detection and contour signature. 2012 International Symposium on Intelligent Signal Processing and Communications Systems. Taipei, 2012 p. 463-468.
[17] Konstantinidis D, Dimitropoulos K, Daras P. 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, 2018 p.1-6.
[18] Simon T, Joo H, Matthews I, Sheikh Y. Hand keypoint detection in single images using multiview bootstrapping. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017 p. 4645-4653.
[19] Park J, Cho S, Kim D, Bailo O, Park H, Hong S, Park J. A Body Part Embedding Model With Datasets for Measuring 2D Human Motion Similarity. IEEE Access, 9, 2021, p. 36547-36558.
[20] Sakoe H, Chiba S. Dynamic Programming Algorithm Optimization for Spoken Word Recognition, IEEE Transaction on Acoustics, Speech, and Signal Processing, Vol. ASSP-26, No. 1, 1978, pp. 43-49.
[21] Berndt D.J, Clifford J, Finding Patterns in Time Series: A Dynamic Pro-gramming Approach, Proceedings of Advances in Knowledge Discovery and Data Mining, AAAI/MIT, 1996, pp. 229-248.
[22] Rath T. M, Manmatha R. Word image matching using dynamic time warping, Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, 2003.
[23] Nakanishi N, Okada R, Nakahodo R, Semantic Waveform Measurement Method of Kansei Transition for Time-series Media Contents, International Journal of Smart Computing and Artificial Intelligence, International Institute of Applied Informatics, Vol. 5, No. 1, 2021, p. 51 – 66.
[24] Li S, Deng W. Deep facial expression recognition: A survey. IEEE transactions on affective computing, 2020