IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 19, 2021, accepted February 16, 2021, date of publication March 2, 2021, date of current version March 17, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3063203

Backhand-View-Based Continuous-Signed-Letter Recognition Using a Rewound Video Sequence and the Previous Signed-Letter Information

PONLAWAT CHOPHUK AND KOSIN CHAMNONGTHAI, (Senior Member, IEEE)
Department of Electronic and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
Corresponding author: Kosin Chamnongthai (kosin.cha@kmutt.ac.th)

This work was supported by the Petchra Pra Jom Klao Ph.D. Research Scholarship from the King Mongkut's University of Technology Thonburi under Grant 6/2560.

ABSTRACT In sign language, when signed letters are continuously spelled based on backhand view, a previous signed letter influences the trajectory of hand and fingers approaching the pause duration for signing the current signed letter. Since those varied trajectories are regarded as parts of the current signed letter, hand gesture during pause duration of the current signed letter is regarded as insufficient for recognition of the current signed letter. The previous signed letters, and trajectories of hand and fingers between the previous and the current signed letters should be included as data for classification. This paper proposes a method of backhand-view-based continuous-signed-letter recognition using a rewound video sequence with previous signed letter. In the method, a hand shape of previous signed letter and trajectories of finger joints moving from the previous signed letter to the current one are detected, features are then extracted, and finally, the features are classified for signed letter recognition. To evaluate performance of the proposed method, experiments with 10 participants were performed 20 times each, and the results revealed 96.07% accuracy approximately which were improved significantly from the conventional methods using forehand and backhand.

INDEX TERMS Continuous signed letter, rewound video, previous signed-letter, backhand view, LSTM.

I. INTRODUCTION

In general, a sign language, which is basically expressed by finger and hand gestures, consists of signed words and signed letters. The signed letter is being used for personal names and other proper nouns such as the name of cities, areas, regions, companies, and, brands. It is also used when introducing new technical terms and new words [1]. In fact, whereas the number of words, used for the conversation in our daily life is regarded as approximately 273,000 [2], the number of signed words is counted approximately 10,000 [3]. This means more than 200,000 words cannot be expressed by signed words, and these are required to be spelled based on signed letters. The sign letters in the American Sign Language (ASL) actually occupy 12-35% approximately [4] during full conversation. Moreover, hearing impaired teachers statistically utilize more than 50% of finger-spelled words during classroom lectures [5]. Therefore, signed letter recognition is regarded as important function in the development of automatic sign-language interpretation system.

To develop the function of signed letter recognition for the automatic sign-language interpretation system, practical implementation requires mobile function in which a sensor for capturing finger gestures should be installed on signer's body parts such as head, eye, chest, and so on. This raises up a research problem of signed letter recognition based on backhand view. Moreover, signed letters are isolated signed gestures which are normally assumed to be independent. However, signed letters are mainly used continuously for spelling a word in the sign language conversation so that a function of signed letter recognition requires continuous signed letters rather than isolated ones.

Research works related to sign language recognition including signed word and signed letter recognition have been done by many researchers. These research works can be divided into two groups; sensor-based [6] and

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tan.

vision-based approaches [7]. The sensor-based approach, where sensors are installed on parts of human body such as hands, fingers, and arms, takes advantage in accuracy, because positions of fingers and finger joints are accurately obtained. However, users may feel uncomfortable by wearable sensors. On the other hand, vision-based approach [8], [9] is more preference for users due to unwearable sensors. The research works in the vision-based approach can be divided into isolated [10]–[22] and continuous [23]–[45] groups. The methods in the isolated group are assumed to deal with isolated signed words or signed letters, and also are available against the continuous words and letters without any changes after segmentation. On the other hand, the methods in the continuous group target to recognize continuous words and letters which are partially changed depending on the neighboring words and letters. The methods in this group are technically divided into 2D and 3D data usage approaches. The 2D approach [23]–[42] uses less video and image data for speedy processing, but some words and letters which require depth information are not available. On the other hand, 3D approach [43]–[45] theoretically can deal well with the words and letters which consist of depth information of fingers. In this approach, Mittal *et al.* [43] proposed a modified model for continuous sign language using Leap Motion sensor. The model achieved high accuracy with isolated words, but slightly low accuracy in signing sentence. Fang *et al.* [44] proposed mobile sign-language interpretation system called DeepASL for recognizing backhand-based continuous words. This method processed a video sequence of continuous words, by recognizing each word individually, and achieved high accuracy. However, the beginning part of each word in a video sequence of continuous words is transformed from the previous word, and might change depending upon previous word. Warchol *et al.* [45] proposed a recognition of fingerspelling sequences in Polish sign language using clouds obtained from depth images. This method performed excellently in signing continuous signed letters because it recognizes a letter by using not only information of the current signed letter, but also post consecutive signed letter. Both mentioned methods are assumed to deal with forehand signs in which informative finger gestures can be seen clearly. However, in the case of our research problem of backhand-based continuous-signed-letter recognition, those informative finger gestures are almost occluded by hand and fingers so that features seen on the backhand must be considered.

In order to classify a signed letter in a video sequence of continuous signed letters based on the backhand view as our research problem, the authors of this paper therefore considered to utilize the information of not only the current letter, but also previous letter, which influences the transformation to the current letter. The transformation part of the current letter or word which is normally located in the beginning of a video sequence of letter, may be assumed to be changed depending on the previous letter therefore a recognizing signed letter should be trained with all possible previous letters. In the case of English language, since there exist 26 English letters with a couple of letters representing space and starting of the signed letter mode, recognition of a signed letter is possibly influenced by 28 previous letters. Therefore, there totally exist 28 patterns for recognition of a signed letter. Moreover, to sign a letter, it normally varies in signing speed even when signing the same letter, an appropriate time-independent classification tool was selected as the signed letter classifier in this paper.

This paper is organized as follows. Analysis of a backhand-based signed letter in a video sequence of continuous letters is described in Section II. Proposed continuous signed-letter recognition based on backhand view, and experiments and results are discussed in Sections III and IV, respectively. Discussion and conclusion are presented in Sections V and VI, respectively.

II. ANALYSIS OF A BCKHAND-BASED SIGNED LETTER IN A VIDEO SEQUENCE OF CONTINUOUS LETTERS

Backhand-view signed letters in American Sign Language include fist and non-fist signs, that can be categorized based on shape similarity into five groups: I, II, III, IV, and V, as shown in Fig. 1. To spell continuous signed letters based on a backhand, a signer normally uses a hand to sign letters for listeners in front of the signer so that informative gestures of fingers may be conveniently seen from the front side. While listeners in the front of the signer can see the informative finger gestures, a sensor of an interpretation system installed on the signer chest for mobility purpose is assumed to see backhand views. This means informative finger gestures are supposed to be hidden by the hand, and a visual sensor can track up only backhand data including backhand contour, backhand skin, and so on, as shown in Fig. 2 (a). Recently emerging Leap Motion sensor can excellently track 3D positions of finger tips and joints, as shown in Fig 2 (b). The differences of velocity of finger joints and tips between consecutive video frames during signing continuous signed letters (L_{t-1}, L_t, L_{t+1}) can be plotted as a graph, as shown

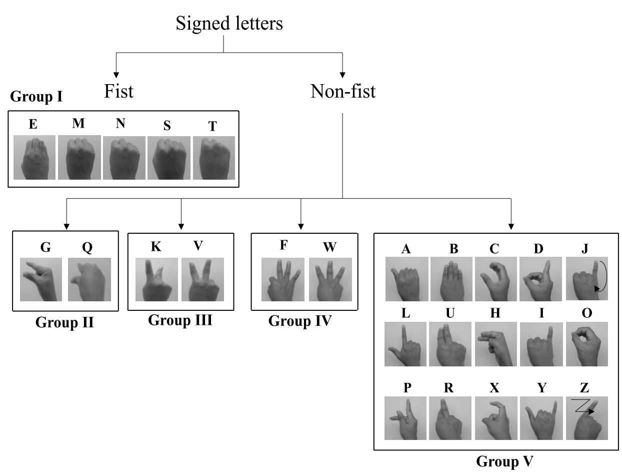

FIGURE 1. Signed-letter group.

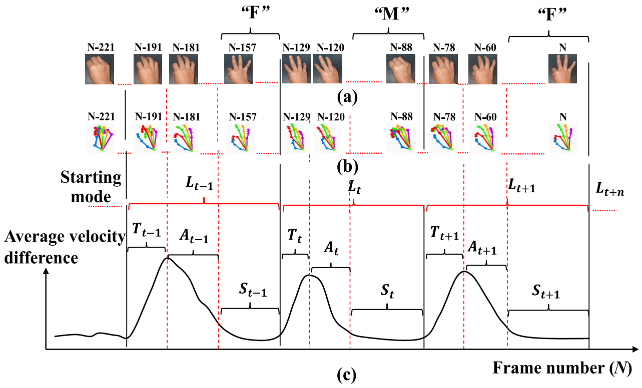

FIGURE 2. Definition of sign-letter components: (a) 2D image, (b) 3D positions of finger joint, and (c) Velocity difference between consecutive video frames.

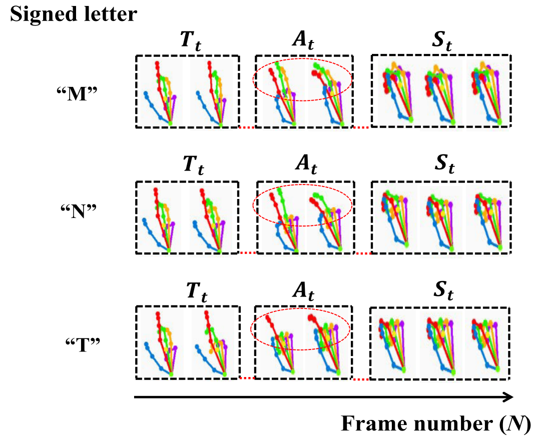

FIGURE 3. Difference appearing a rewound sequence.

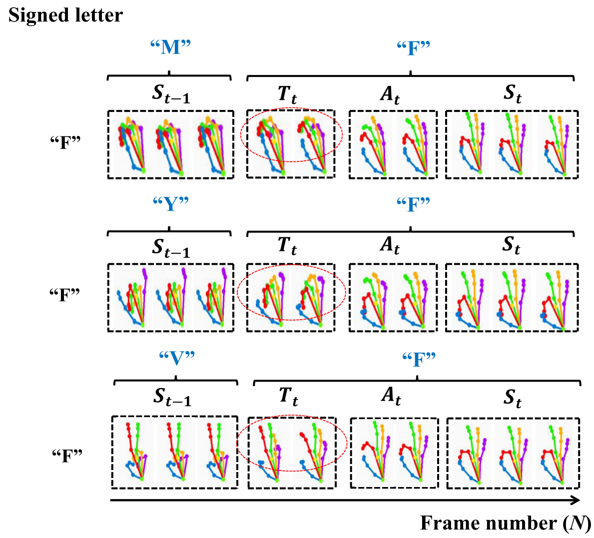

FIGURE 4. Difference of rewound sequence depending on previous signed letter.

in Fig. 2 (c). The graph analytically shows a transforming (T), an approaching (A), and a pause (S) component in each signed letter. Normally, the curve of the transforming component shows a slope in which the differences between a couple of consecutive video frames start from zero and increase until reaching the peak. Consequently, the curve of the approaching component starts from the peak and decrease until reaching zero. Finally, zero curve which is paused for a while shows pause component, and a signed letter based on the sign language dictionary is signed. Therefore, the pause component of a signed letter is regarded to be the most important part for signed letter recognition. Practically, the sign during a pause component of a signed letter is not always correctly caught up due to occlusion, and this may confuse the signed letter recognition. This means only a pause component is regarded as insufficient for signed letter recognition. We therefore consider rewinding a video sequence of a signed letter (L_t) from the pause period (S_t) back to the timing border with the previous signed letter (L_{t-1}). This concept is considered as solution to be able to differentiate even signed letters included in the same group which have similar backhand sign shapes. For instance, differences can be observed when rewinding sign letters "M", "N", and "T" which are included in the same Group I, as shown in the approaching components surrounded by red ellipses in Fig. 3. However, approaching (A_t) and transforming (T_t) components of the current signed letter (L_t) are observed to be varied depending on the pause component (S_{t-1}) of the previous signed letter (L_{t-1}), as shown in Fig. 4. For instance, the transforming and approaching components of "F" are obviously varied when it is signed following "M", "Y", and "V", as shown in Fig. 4. Therefore, to recognize a signed letter on backhand view, not only the current signed letter (L_t) including transforming (T_t), approaching (A_t), and pause (S_t) components, but also the pause component (S_{t-1}) of the previous signed letter (L_{t-1}) are required as basic concept of this paper.

Since the sign during pause component exactly shows the signed letter based on the sign language dictionary, most of the related works employed the information (I) of pause duration ($I(S_t)$) of the current signed letter to classify the current signed letter (L_t) as follows.

$$\text{Classify}\ (L_t) = \text{Classify}(I(S_t)) \tag{1}$$

While most of previous methods dealing with forehand and backhand views focus on information of the pause component of the current letter, our proposed approach, which is aimed to classify continuous signed letters based on backhand view, uses fused information between the pause component of the previous signed letter and the information of current signed letter including transforming, approaching, and pause components. The proposed approach can be mathematically expressed in term of information (I) by the conditional probability as follows.

$$\text{Classify}(I(L_t)) = \text{Classify}(I(S_{t-1}) + I(T_t) + I(A_t) + I(S_t) \tag{2}$$

$$I(T_t) = -\log P(T_t S_{t-1}) \tag{3}$$

then, our proposed method classify the current signed letter as follows.

$$\text{Classify}\,(I\,(L_t)) = \text{Classify}\left\{\log\left(\frac{1}{P(S_{t-1})}\right) + \log\left(\frac{1}{P(T_t|S_{t-1})}\right) + \log(\frac{1}{P(A_t)}) + \log(\frac{1}{P(S_t)})\right\} \tag{4}$$

where P represents probability.

Based on the mentioned equation, the transforming component (T_t) of the current signed letter (L_t) is influenced by the pause component (S_{t-1}) of the previous signed letter (L_{t-1}). Suppose the number of signed letters is 26, then the transforming component (T_t) of the current signed letter (L_t) is possibly varied by 26 patterns, as shown in Fig. 5. The transforming components (T_t) surrounded by red rectangles are varied based on the pause component (S_{t-1}) of the previous signed letter (L_{t-1}). These patterns of video subsequence including transforming (T_t) of the current signed letter (L_t) and pause component (S_{t-1}) of the previous letter (L_{t-1}) are proposed in this paper to train in a classifier.

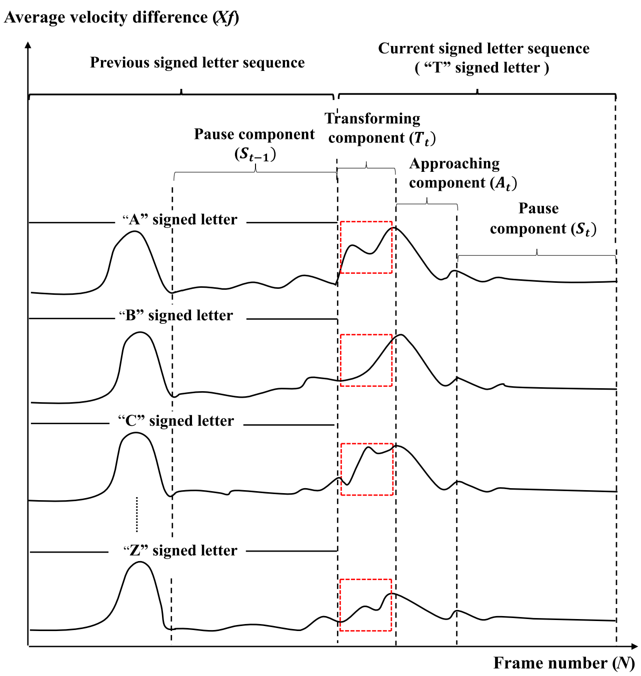

FIGURE 5. Transforming part of the current signed letter depending on the previous signed letter.

Moreover, signing speed normally varies based on signers and time even by the same signed letter (e.g. "T"), as shown by examples in Fig. 6. Although the same signed letter, it is not necessary to perform the same signing speed. This becomes a difficulty for classification, since the patterns representing feature vary in time. These patterns logically can be represented by series of state, which transits by probability, and described by a state transition diagram. Therefore, classifier of signed letter in our proposed method should be selected to be able to deal with time independent data such as HMM, LSTM, and so on.

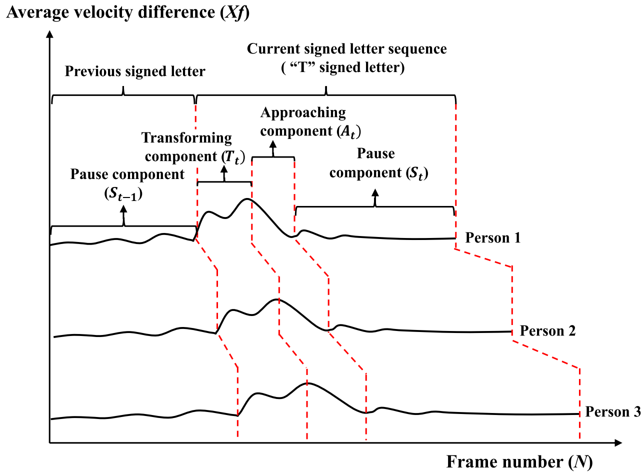

FIGURE 6. Same signed letter in different signing speed.

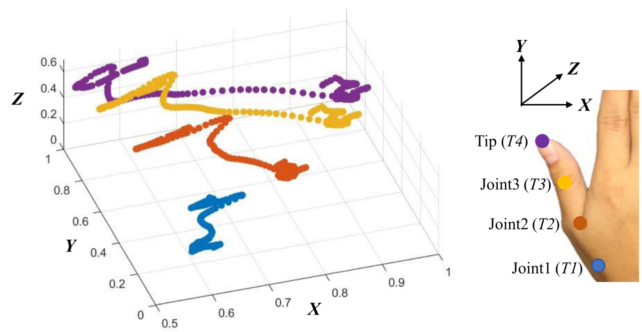

FIGURE 7. Trajectories of joints and tip of a thumb for "N" signed letter.

The mentioned average velocity differences are results obtained from 3D coordinates (X, Y, Z) of three joints and a tip in each finger, totally 20 coordinates for five fingers on a hand. Normally, these finger joints and tips move freely in 3D space for signing signed letters. If a signer signs a signed letter, the trajectories of joints and tips are considered to be performed in the same pattern, as shown by a sample of "N" signed letter in Fig. 7. Due to different sizes of human hands, the trajectories of joints and tips might be varied, practically it should be normalized in order to absorb the difference of hand sizes.

III. PROPOSED CONTINUOUS SIGNED-LETTER RECOGNITION BASED ON BACKHAND VIEW

Based on our basic concept mentioned above, the proposed method has been implemented, and overall proposed method with a flowchart is mentioned in subsection A. The details of all processes including signed letter sequence segmentation and pause duration detection, feature extraction, and classification are consequently explained in subsections B, C, and D, respectively.

A. OVERALL PROPOSED METHOD

As overall of the proposed method is implemented based on the mentioned basic concept, as shown in Fig. 8, the system starts by inputting a sequence of video frames representing continuous signed letters. These continuous signed letters are segmented into several signed letter sequences at the process of signed-letter sequence segmentation. The segmented video sequences representing signed letters are then used to detect pause duration of all signed letter sequences at the process of pause duration detection. Both of mentioned signed letter sequence segmentation and pause duration detection would be introduced in subsection B. Finally, the segmented video sequences representing signed and video frames during pause duration are used to extract features and classify in which details are described in subsections C and D, respectively, and the classified signed letters are the output as recognition results.

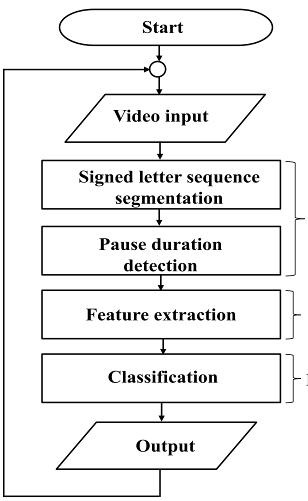

FIGURE 8. Flowchart of the proposed method.

B. SIGNED LETTER SEQUENCE SEGMENTATION AND PAUSE DURATION DETECTION

When video frames representing continuous signed letter sequences are used as the input into the system, as shown by an example of continuous signed letters in Fig. 9 (a), the video sequence practically can be segmented into several sequences representing signed letter, e.g. "mode starting", "F", "A", "N", and "Y". In the example, 3D position information of finger joints is obtained by a Leap Motion sensor, and colors in the Fig. 9 (a) differentiate fingers in a frame. If we take speed average of 15 joints and five tips (totally 20 points), it can be expressed in a velocity graph, as shown in Fig. 9 (b). It is observed that in the pause duration (S_i) where a finger gesture representing a signed letter (e.g. 'F', 'A', 'N', 'Y' in Fig. 9) based on the sign language dictionary, would be allocated in the back of a signed letter sequence, and signing a letter starts from the end of the pause duration of the previous signed letter, pass transformation duration (T_i) in which the average speed (v) is increased to reach the speed

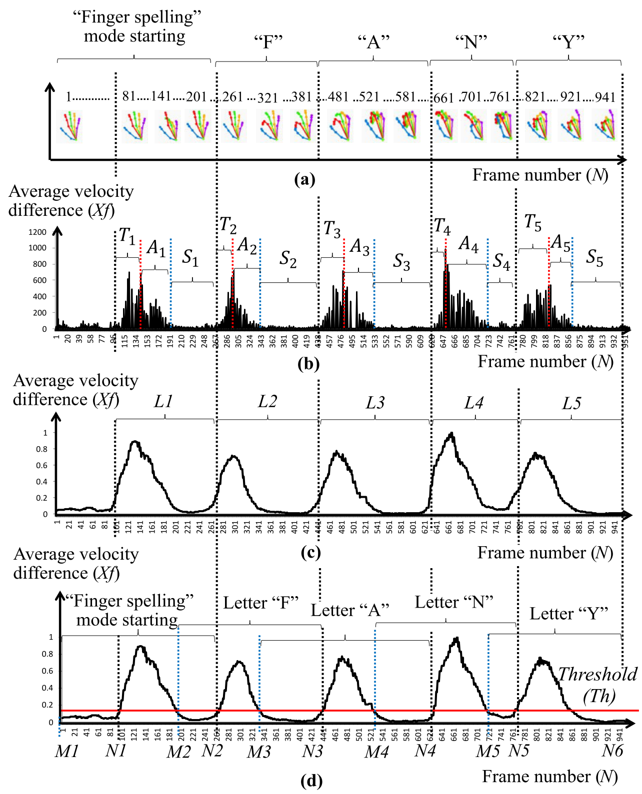

FIGURE 9. Proposed subsequence detection: (a) 3D data, (b) 3D velocity, (c) Smoothing (MSf), and (d) Pause duration detection(Ss).

peak, and then the average speed decreases in the duration called approaching (A_i) to approach the next signed letter in the pause duration. Since there exists many local peaks in the graph of a signed letter sequence and may cause a problem of missing the peak speed, smoothing process is needed to delete those local peaks, as shown in Fig. 9 (c). To detect the pause duration whose average speed theoretically becomes zero, is possibly detected by finding the starting frame of zero average speed as the border of the pause duration. However, practically there exist noises so that a threshold value should be set up based on the whole data of average speed, and the threshold would be used to suppress the noises, as shown in the following equations.

$$Th = \frac{p}{M}\sum_{i=1}^{M} |\Psi_i - \Psi_{i+1}| \tag{5}$$

where p is a parameter for segmentation.

$$\mu_i = \begin{cases} 0, & \Psi_i < Th \\ \Psi_i, & \Psi_i \geq Th \end{cases} \tag{6}$$

where $i = 1, \ldots, M$.

The algorithm of the proposed signed letter sequence segmentation and pause duration detection is depicted in the Algorithm 1.

C. FEATURE EXTRACTION

To extract feature for classification in the next process, effective features should be determined and fed to the classifier.

Algorithm 1 Signed Letter Segmentation and Proposed Subsequence Detection

```
STRAT
    // 3D Velocity of all joints in hand
REPEAT
    COMPUTE an average velocity calculation of all joints of consecutive frames as Xf
UNTIL frame number = N
    // Normalization
REPEAT
    COMPUTE normalization Xf into the range of 0 to 1 as Nf
UNTIL frame number = N
    // The moving average smoothing
COMPUTE Nf with zero padding as Zp
REPEAT
    COMPUTE Zp with the moving average among consecutive frames by n as MSf
UNTIL frame number = N
    // Adaptive thresholding
REPEAT
    COMPUTE the timing frame of borders of each word as Wn
    COMPUTE equation (5) to set thresholding (Th) of Wn
    COMPUTE equation (6) for suppressing noise to zero
UNTIL frame number = N
    // To find the timing frame of proposed subsequence
REPEAT
    COMPUTE the starting frame collection
        CASE starting frame from zero slope as Mn
UNTIL frame number = N
COMPUTE the timing frame of borders between sequences of Mn and Nn+1 to set proposed subsequence (Ss)
END
```

Since a Leap Motion sensor retrieves coordinates of all finger joints and tips of a hand, palm center, and wrist center, hence these data are useful and effective for classification. However, all coordinates may be moved in the 3D free space according to trajectories for signing signed letters in which the origin (0,0,0) is allocated at the sensor itself. These coordinates and trajectories may be varied based on distance from the hand to the origin while wrist is always moved together. If those trajectories are observed from the wrist, it seems to be patterned after each sign letters. Therefore, this paper proposes to convert the original origin to the wrist center as the new origin, and determine the coordinates of finger joints and tips based on the new origin at the wrist as features for classification. Practically, the coordinates of joints and tips of thumb, index, middle, ring, and pinky fingers, as shown in Fig. 10, are converted to the ones of the new origin at the wrist center using the equation (7). The algorithm for feature extraction is illustrated, as shown in Algorithm 2.

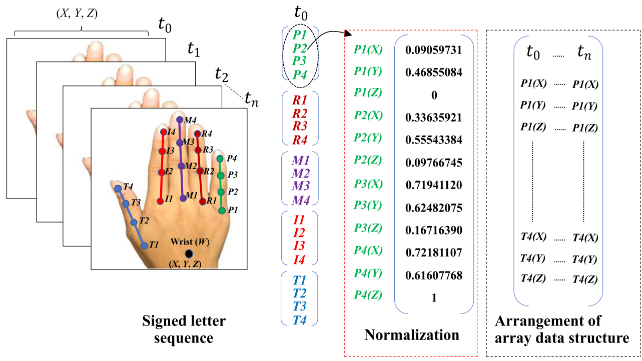

FIGURE 10. Coordinates of finger joint and tip as a classification feature.

Algorithm 2 Detection of Joint Trajectories

```
START
SET Fj = 3D hand joints, K = the number of signed letters, N = the number of frames of each signed letter
REPEAT
REPEAT
COMPUTE normalization Fj to the range of 0 to 1 as Nh
COMPUTE arrangement of Nh in array data structure for classification
UNTIL frame number = N
UNTIL signed letters = K
END
```

Moreover, since finger size may vary depending on person, coordinates of finger joint and tip would be differentiated among people. In order to solve the problem of different data, especially between training and testing done by different people, finger size should be registered in advance, and all coordinates should be normalized in practice.

$$P_{new}(i,j) = \sum_{i=1}^{N}\sum_{j=1}^{M}(P(i,j) - P(i,h)) \tag{7}$$

where P and P_{new} represent 3D coordinates (X, Y, Z) of joint and tip positions based on original and new origins, and i, j, M, N, and h stand for video frame, finger joints and tips, the total number of video frames, the total number of finger joint positions, and wrist joint position, respectively.

D. CLASSIFICATION

Since finger joint positions sensed by the Leap Motion sensor might be slightly changed due to different speed of finger motion for signing at the moment, those differences should be adjusted. In this paper, the designed template would quantize the sensed finger joint positions, absorb the differences occurring in different signing time, and create a series of binary codes representing a signed letter. Moreover, the signed letter dynamically appears in a sequence of video frames, the pattern in a signed letter sequence may transit in several states. To classify the signed letters, classification tools, which enable to deal with state transition, should be considered

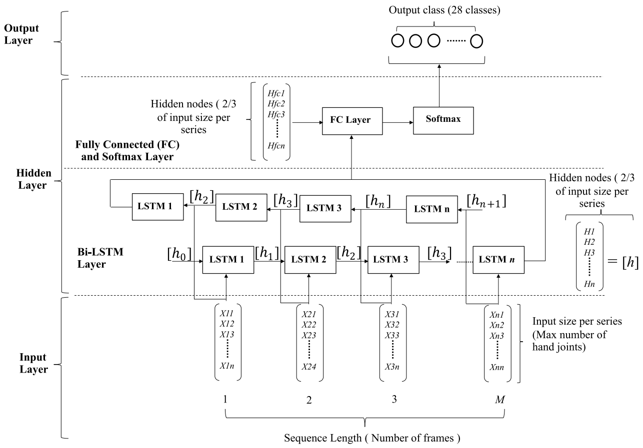

FIGURE 11. Long short-term memory (LSTM) networks.

TABLE 1. Classification specifications.

Layer	Parameter	Value
Input layer	Sequence length	longest
	Batch size	27
	Input size per series	60
	Input feature	1 dimension
Hidden layer	Bi-LSTM layer	longest
	Hidden node	40
	Activation function	Softmax
Output layer	LSTM model	Many to one
	Output class	28

by users to select as a classifier in their applications. This paper basically employs a series of quantized finger joint positions as input data, feed them to Long Short-Term Memory (LSTM) networks [46] for training in advance, and classify them based on the trained database into signed letters.

Users may set up number of input nodes based on binary digits of a series of the signed letter code, and number of classified signed letters as output nodes.

IV. EXPERIMENTS AND RESULTS

To evaluate performance of the proposed method, a Leap Motion sensor was installed on the chest of 10 participants, as shown by a photo in Fig. 12, and video clips of continuing

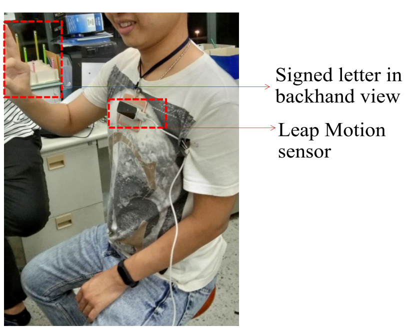

FIGURE 12. Photo of the experimental setup.

TABLE 2. Experimental specifications.

Devices /Participants	Specification
Computer system	Dell G3 Gaming w56691425TH CPU: Intel Core i7-8750H GPU: NVidia GeForce GTX 1050Ti Memory Size: 8 GB DDR4 Hard Disk Drive: 1 TB
Hardware	Leap Motion sensor
Video	200 frames per second
Participants	10 people
Experiment	Participants: 10 people 20 times/signed letter/person

TABLE 3. Accuracy comparison between proposed and conventional methods.

Type of signed letter	Data	Type of view	Method	Accuracy rate (%)
Isolated signed letter	2D	Forehand view	[10]	94.37
		Backhand view	[22]	80
	3D	Forehand view	[19]	93.81
Continuous signed letters	2D	Forehand view	[41]	84
			[42]	91
	3D	Forehand view	[45]	78.80
		Backhand view	Proposed method	96.07

two signed letters were recorded in a storage for evaluations. The specifications of the experimental system such as Leap Motion sensor, computer system, video, and so on are depicted in Table 2. Due to excellent ability of the Leap Motion sensor which can be used in both outdoor and indoor, experiments have been performed in environments in which background and light were not controlled. The participants were well trained in signing signed letters and using the experimental system before starting data collection. Recorded video clips of continuing signed letters were processed according to our proposed method, and experimental results in average accuracy rate are shown in Table 3. In that table, the experimental results of our proposed method are compared with conventional methods based on isolated signed letters using forehand in 2D [10] and 3D [19], and backhand in 2D [22], as shown in the 2^{nd} to 4^{th} rows, and continuous signed letters using forehand in 2D [41], [42] and 3D [45], as shown in the 5^{th} to 7^{th} rows. The experimental results of the proposed method based on backhand view in 3D is allocated in the bottom row. To show the improvement in accuracy of each signed letter, recognition results of the proposed method are compared with conventional methods, as shown in Table 4. The average accuracy rates of each signed letter performed by conventional methods using isolated signed letter in 2D for forehand [10], and backhand [22], and forehand in 3D [19] are shown in columns 2-4, and the

TABLE 4. Comparison of signed letter recognition accuracy.

Signed Letter	Recognition Accuracy (%)					
	Isolated signed letter			Continuous signed letters		
	2D		3D	2D	3D	
	Fore-hand	Back-hand	Fore-hand	Fore-hand	Backhand	
					Proposed Method	
	[10]	[22]	[19]	[42]	Accuracy	SD
A	≈ 95.0	60	90.82	≈88.0	98.24	0.30
B	≈97.5	98	100	≈98.5	96.56	0.11
C	≈97.0	62	100	≈92.5	99.60	0.04
D	≈97.0	90	89.67	≈92.5	98.08	0.14
E	≈95.5	80	99.00	≈98.5	98.80	0.25
F	≈97.0	88	100	≈98.5	96.64	0.10
G	≈95.5	88	90.00	≈95.0	98.80	0.13
H	≈95.5	88	72.17	≈83.0	99.60	0.06
I	≈97.0	98	100	≈94.0	95.44	0.10
J	-	-	99.67	≈89.0	91.04	0.25
K	≈95.0	-	99.50	≈94.5	98.00	0.10
L	≈98.0	90	99.83	≈98.5	98.24	0.03
M	≈92.0	34	95.17	≈77.0	94.48	0.42
N	≈88.0	54	93.67	≈90.5	94.16	0.16
O	≈92.5	94	99.33	≈98.5	95.04	0.32
P	≈93.5	82	86.67	≈82.0	98.32	0.27
Q	≈95.0	90	92.12	≈87.5	99.12	0.07
R	≈91.0	84	100	≈91.5	92.56	0.30
S	≈91.5	62	65.83	≈91.5	96.72	0.27
T	≈87.0	26	94.83	≈90.5	97.04	0.61
U	≈93.5	94	70.00	≈85.5	92.56	0.70
V	≈93.5	96	96.33	≈82.5	96.80	0.08
W	≈98.0	92	100	≈80.5	94.72	0.23
X	≈92.5	94	98.17	≈96.5	92.48	0.36
Y	≈97.0	96	100	≈93.5	92.40	0.60
Z	-	-	99.67	≈96.0	92.40	0.32
Average	**94.37**	**80**	**93.81**	**91**	**96.07**	**0.24**

ones with the method using continuous signed letters for forehand in 2D [42] are revealed in the 5^{th} column. Finally, the accuracy of each signed letter and its standard deviation performed by the proposed method are introduced in the 6^{th} and the last columns.

V. DISCUSSION

To solve the research problem of continuous backhand-signed-letters recognition, the pause duration of previous signed letter and rewound video of the current signed letter are used as state transition pattern for recognition in this paper. For instance, "M" and "N" signed letters would be recognized using not only pause (S_t) of the current signed letter as most of the conventional methods but also transforming (T_t) and approaching (A_t) components of the current signed letter with the pause component (S_{t-1}) of the previous signed letter, as shown in Figs. 13 (a) and (b). In case of signed letters in the same group (e.g. "M" and "N"), as explained

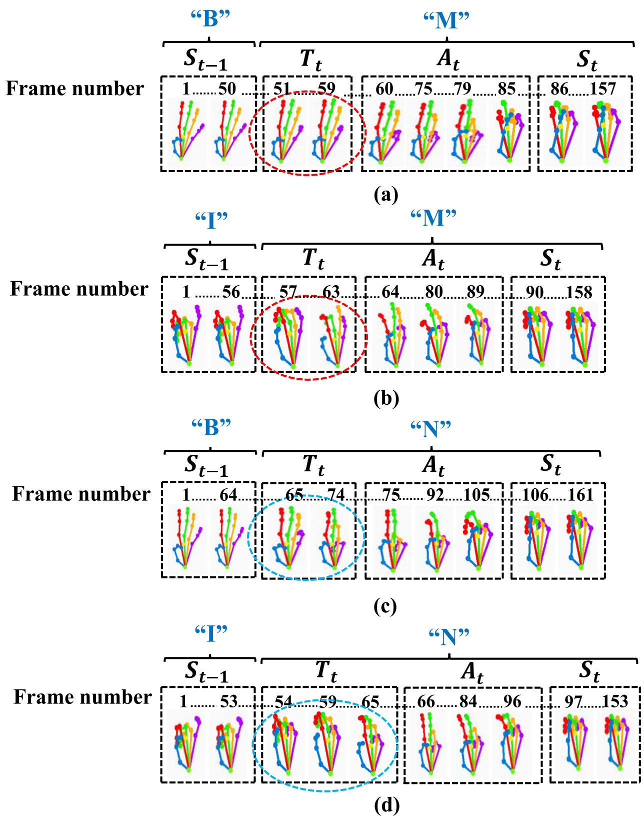

FIGURE 13. Examples of continuous signed letter sequence used in proposed method: (a) "M" signed letter, (b) "M" signed letter, (c) "N" signed letter, and (d) "N" signed letter.

in Fig. 1, it seems to be difficult to classify, as seen by pause components (S_t) of "M" and "N" in Figs. 13 (a) and (c). Compared with this, our proposed method adds not only information of rewound sequence including transforming (T_t) and approaching (A_t) components of the current letter (L_t) but also the pause component (S_{t-1}) of the previous signed letter (L_{t-1}). Comparison between rewound sequences of "M" and "N" shown in Figs. 13 (a) and (c) obviously reveals some differences in approaching and especially transforming components surrounded by ellipse. However, a couple transforming components (T_t) of the same current signed letters (L_t) are not necessary to be the same, as shown by examples of "M" and "N" signed letters in Figs. 13 (a) and (b), and Figs. 13 (c) and (d), respectively. Those transforming components (T_t) are observed to depend on the pause component (S_{t-1}) of the previous signed letter (L_{t-1}) which are "B" and "I", as shown in Fig. 13.

To analyze errors of the proposed method, a confusion matrix of 26 ASL signed letters reveals recognition results in per cent by correct ones on the diagonal line, and missed recognition distributed in other signed letters, as shown in Fig. 14. Some signed letters such as "J", "N", "U", "X", "Y", and "Z" faced many errors which should be especially considered for prevention. In analysis of those errors for solution strategy, some errors were found that the Leap

Prediction

Ground truth	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	98	0.1	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0.1	0	0.5	0	0	0	0	0	0.6	0
B	0.6	97	0	0	1	0	0.6	0	0.2	0	0.3	0	0	0	0	0	0	0	0	0	0.4	0.1	0.2	0	0	0
C	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0
D	0	0.2	0.2	98	0.1	0	0	0	0.4	0	0.1	0	0	0	0.3	0.1	0	0	0	0	0	0	0	0.5	0	0.1
E	0	0.1	0.2	0.1	99	0	0	0	0.1	0	0	0	0	0	0.2	0	0	0	0.6	0	0	0	0	0	0	0
F	0	0	0.1	0.1	1.4	97	0	0	0.2	0	0	0	0	0	0.6	0	0	0	0.7	0.2	0	0	0	0	0.1	0
G	0	0	0	0	0	0	99	0	0	0.4	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0
I	0.1	0.1	0	0	1.5	0.1	0	0	95	0.2	0	0	0.7	0.2	0.2	0.1	0	0	1.1	0.1	0	0	0	0	0	0.2
J	0	0	0	0	0	0	0.2	0	1	91	0	0.2	0.2	0.6	4.5	0.3	0.1	0	0.8	0	0	0	0	0	0	1
K	0	0	0	0	0	0	0	0	0	0	98	0.1	0.4	0.1	0.2	0	0	0.6	0.2	0.2	0.1	0.2	0	0	0	0
L	0	0.1	0.9	0.4	0	0	0	0	0.1	0	0	98	0	0.2	0.1	0	0	0	0	0.1	0	0	0	0	0	0
M	0	0	0	0	0.1	0	0	0	0	0	0	0	95	4.3	0.3	0	0	0	0.2	0.6	0	0	0	0	0	0
N	0	0	0	0	0.3	0	0	0	0.1	0	0	0	3.6	94	0		0	0	0.6	1.1	0	0	0	0.1	0	0
O	0	0	1	0.8	0.4	0.1	0	0	0.2	0.2	0.2	0.1	0.5	0.8	95	0	0	0	0	0.6	0	0	0	0	0	0.2
P	0	0	0	0	0	0	0.6	0.2	0	0.2	0	0	0	0	0.4	98	0	0	0	0.2	0	0	0	0	0.2	0
Q	0.1	0	0	0	0.3	0	0	0	0	0	0	0	0	0	0.1	0	99	0	0.2	0	0	0	0	0	0	0.2
R	0	0	0.1	0	0.3	0	0	0	0	0	0.3	0	0	0	0	0	0	93	0	0	6.2	0	0.6	0	0	0
S	1.4	0	0	0	0	0	0	0	0.1	0	0	0	0.4	0.3	0.1	0	0	0	97	1	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0.1	0	0	0	0.8	1.2	0.1	0	0	0	0.6	97	0	0	0	0.2	0	0.1
U	0	0.1	0	0	0.2	0	0	0	0	0.1	0	0	0	0	0	0	0	6.2	0	0	93	0.1	0.5	0	0	0.3
V	0	0	0	0.1	0.6	0	0	0	0	0	0.1	0	0	0.2	0	0	0	0	0	0	0.6	97	1.4	0	0	0.4
W	0	0.9	0	0.1	0.2	0	0	0	0	0.1	0.2	0	0.3	0.1	0.6	0.1	0	1.1	0	0	0.3	0.8	95	0.6	0	0
X	0	0	1.8	2.9	0	0	0	0	0.1	0.3	0	0	0	0	1.6	0.8	0	0	0	0	0	0	0	93	0	0.1
Y	6	0	0.1	0	0.2	0.3	0	0	0.3	0	0	0	0.1	0	0.6	0	0	0	0	0	0	0	0	0	92	0
Z	0	0	0.2	0.6	0.2	0.1	0	0	0	1.1	0.3	0.1	0	0.2	0.9	0.9	0.2	0.2	1.8	0.2	0.1	0	0	0.2	0.1	92

FIGURE 14. Confusion matrix of experimental results based on ASL signed letters.

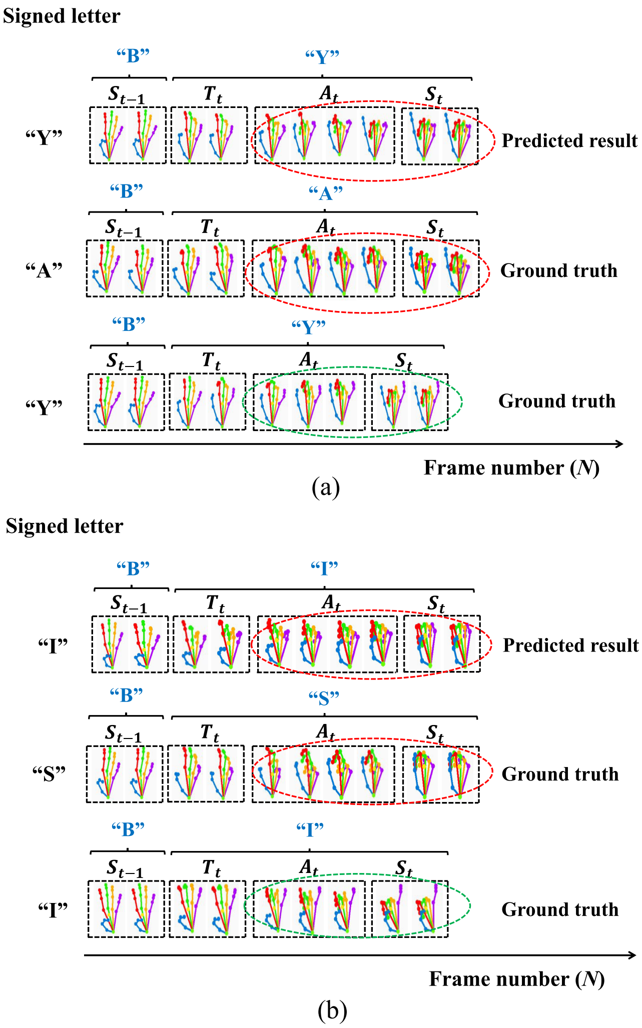

FIGURE 15. Error example caused by sensor.

Motion sensor could not retrieve joint positions correctly due to occlusion. Examples of this kind of device error are shown in Fig. 15. Joint positions and finger trajectories of "Y" and "I" signed letters were not correctly sensed and estimated by the Leap Motion, as shown in the first rows of

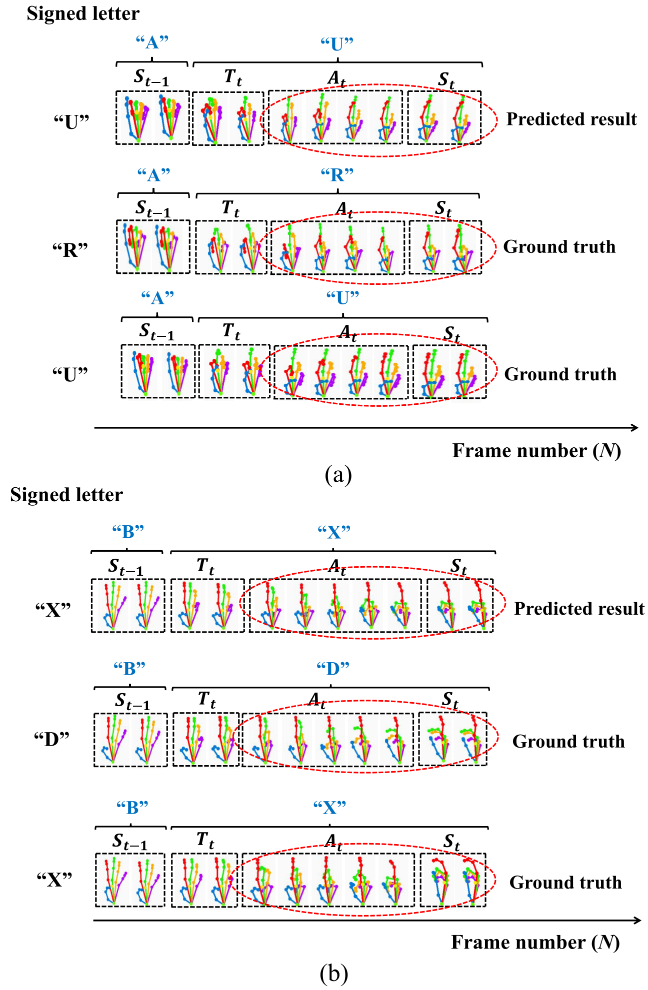

FIGURE 16. Error example caused by sensor.

Figs. 15 (a) and (b), respectively which originally should be as ground truth at the bottom rows. Unfortunately, it was similar to "A" and "S", respectively, and misclassified, as shown in the second rows of Figs. 15 (a) and (b), respectively, as the signing hand position was varied depending on person and some key finger gestures were occluded in some angles against the Leap Motion sensor. To solve this problem as future work, stereo Leap Motion sensors and depth sensor should be considered as tools for sensing hidden gestures. Similarly, some couples of similar signed letters such as "U" and "R", and "X" and "D" and so on, as shown in Figs. 16 (a) and (b), respectively, which originally hard to classify, also faced occlusion problem due to some angles against the Leap Motion sensor. To solve this problem of originally similar signed letters, not only usage of stereo sensor and depth sensor, as mentioned in Fig. 15, but also the post signed letter should be considered for solution. Moreover, since the Leap Motion sensor was fixed on the participant chest during the experiments, finger joint positions were allocated in the available scope. In some practical environments, users cannot avoid the movement and vibration of the sensor which might influence the recognition results. The sensor movement and vibration may affect the available sensing scope, and it is possible for the signing hand to be out of the scope. The problems of vibration and movement should be also considered as future work.

VI. CONCLUSION

This paper proposed a continuous signed-letter recognition method based on backhand view using rewound video sequence and previous signed letter information. It was found that most of backhand-view signed letters look similar with other signed letters due to occlusion of informative finger gesture in the front view. If video sequences approaching the signed letters were rewound, differences between similar signed letters could be observed. However, the rewound video sequences were actually varied depending on the previous signed letters. The different previous signed letters were found to influence even the same signed letters. Moreover, the signing speed might be varied all the time even from the same signed letter so that time independence pattern could be considered for recognition. Our proposed method therefore utilized rewound video sequence of the current signed letter with the information of the previous signed letter, and classified the signed letter based on time independence. The proposed method was implemented based on Leap Motion sensor and LSTM as hand sensor and time-independent classifier, respectively. The experiments with well-trained 10 participants revealed significant improvement in signed-letter recognition accuracy compared with conventional methods using forehand and backhand views.

ACKNOWLEDGMENT

The authors are thankful to the Setsatian School and Thung Mahamek School for kind cooperations in collecting hand video dataset.

REFERENCES

[1] M. M. Taylor, "Interpretation skills: English to American sign language," Interpreting Consolidated, Lexington, KY, USA, Tech. Rep. 324490974487, 2017.

[2] *OED Terminology*. Accessed: Sep. 14, 2020. [Online]. Available: https://public.oed.com/how-to-use-the-oed/glossary/

[3] *American Sign Language*. Accessed: Sep. 20, 2020. [Online]. Available: http://www.lifeprint.com/asl101/fingerspelling/fingerspelling.htm

[4] C. Padden and D. C. Gunsauls, "How the alphabet came to be used in a sign language," *Sign Lang. Stud.*, vol. 4, no. 1, pp. 10–33, 2003.

[5] S. Baker, "The importance of fingerspelling for reading," Vis. Lang. Vis. Learn. Sci. Learn. Center, Washington, DC, USA, Tech. Rep., 2010. [Online]. Available: https://www.academia.edu/4381952/The_Importance_of_Fingerspelling_for_Reading

[6] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan, M. M. Salih, and M. M. B. Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," *Sensors*, vol. 18, no. 7, p. 2208, Jul. 2018.

[7] S. Joudaki, D. B. Mohamad, T. Saba, A. Rehman, M. Al-Rodhaan, and A. Al-Dhelaan, "Vision-based sign language classification: A directional review," *IETE Tech. Rev.*, vol. 31, no. 5, pp. 383–391, Sep. 2014.

[8] A. S. Al-Shamayleh, R. Ahmad, M. A. M. Abushariah, K. A. Alam, and N. Jomhari, "A systematic literature review on vision based gesture recognition techniques," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 28121–28184, Nov. 2018.

[9] B. K. Chakraborty, D. Sarma, M. K. Bhuyan, and K. F. MacDorman, "Review of constraints on vision-based gesture recognition for human–computer interaction," *IET Comput. Vis.*, vol. 12, no. 1, pp. 3–15, Feb. 2018.

[10] Y. Hu, H.-F. Zhao, and Z.-G. Wang, "Sign language fingerspelling recognition using depth information and deep belief networks," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 06, Jun. 2018, Art. no. 1850018.

[11] L. K. S. Tolentino, R. O. S. Juan, A. C. Thio, M. A. B. Pamahoy, J. R. R. Forteza, and X. J. O. Garcia, "Static sign language recognition using deep learning," *Int. J. Mach. Learn. Comput.*, vol. 9, no. 6, pp. 821–827, Dec. 2019.

[12] S. Ameen and S. Vadera, "A convolutional neural network to classify American sign language fingerspelling from depth and colour images," *Expert Syst.*, vol. 34, no. 3, Jun. 2017, Art. no. e12197.

[13] J. P. Sahoo, S. Ari, and D. K. Ghosh, "Hand gesture recognition using DWT and f-ratio based feature descriptor," *IET Image Process.*, vol. 12, no. 10, pp. 1780–1787, Oct. 2018.

[14] P. Nakjai and T. Katanyukul, "Hand sign recognition for thai finger spelling: An application of convolution neural network," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 131–146, Feb. 2019.

[15] B. Xie, X. He, and Y. Li, "RGB-D static gesture recognition based on convolutional neural network," *J. Eng.*, vol. 2018, no. 16, pp. 1515–1520, Nov. 2018.

[16] Y. Liao, P. Xiong, W. Min, W. Min, and J. Lu, "Dynamic sign language recognition based on video sequence with BLSTM-3D residual networks," *IEEE Access*, vol. 7, pp. 38044–38054, 2019.

[17] J. Joy, K. Balakrishnan, and M. Sreeraj, "SignQuiz: A quiz based tool for learning fingerspelled signs in indian sign language using ASLR," *IEEE Access*, vol. 7, pp. 28363–28371, 2019.

[18] V. Ranga, N. Yadav, and P. Garg, "American sign language fingerspelling using hybrid discrete wavelet transform-Gabor filter and convolutional neural network," *J. Eng. Sci. Technol.*, vol. 13, no. 9, pp. 2655–2669, 2018.

[19] T.-W. Chong and B.-G. Lee, "American sign language recognition using leap motion controller with machine learning approach," *Sensors*, vol. 18, no. 10, p. 3554, Oct. 2018.

[20] W. Tao, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using convolutional neural networks with multiview augmentation and inference fusion," *Eng. Appl. Artif. Intell.*, vol. 76, pp. 202–213, Nov. 2018.

[21] W. Aly, S. Aly, and S. Almotairi, "User-independent American sign language alphabet recognition based on depth image and PCANet features," *IEEE Access*, vol. 7, pp. 123138–123150, 2019.

[22] K. Pattanaworapan, K. Chamnongthai, and J.-M. Guo, "Signer-independence finger alphabet recognition using discrete wavelet transform and area level run lengths," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 658–677, Jul. 2016.

[23] N. Aloysius and M. Geetha, "Understanding vision-based continuous sign language recognition," *Multimedia Tools Appl.*, vol. 79, nos. 31–32, pp. 22177–22209, Aug. 2020.

[24] R. S. Rokade and D. D. Doye, "Spelled sign word recognition using key frame," *IET Image Process.*, vol. 9, no. 5, pp. 381–388, May 2015.

[25] T. Kim, J. Keane, W. Wang, H. Tang, J. Riggle, G. Shakhnarovich, D. Brentari, and K. Livescu, "Lexicon-free fingerspelling recognition from video: Data, models, and signer adaptation," *Comput. Speech Lang.*, vol. 46, pp. 209–232, Nov. 2017.

[26] T. E. Jerde, J. F. Soechting, and M. Flanders, "Coarticulation in fluent fingerspelling," *J. Neurosci.*, vol. 23, no. 6, pp. 2383–2393, Mar. 2003.

[27] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "SubUNets: End-to-end hand shape and continuous sign language recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3075–3084.

[28] T. Kim, W. Wang, H. Tang, and K. Livescu, "Signer-independent fingerspelling recognition with deep neural network adaptation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 6160–6164.

[29] M. Hassan, K. Assaleh, and T. Shanableh, "Multiple proposals for continuous arabic sign language recognition," *Sens. Imag.*, vol. 20, no. 1, p. 4, Dec. 2019.

[30] O. Koller, R. Bowden, and H. Ney, "Automatic alignment of hamnosys subunits for continuous sign language recognition," in *Proc. LREC*, 2016, pp. 121–128.

[31] H. Wang, X. Chai, and X. Chen, "A novel sign language recognition framework using hierarchical Grassmann covariance matrix," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2806–2814, Nov. 2019.

[32] I. Papastratis, K. Dimitropoulos, D. Konstantinidis, and P. Daras, "Continuous sign language recognition through cross-modal alignment of video and text embeddings in a joint-latent space," *IEEE Access*, vol. 8, pp. 91170–91180, 2020.
[33] K. Tripathi and N. B. G. C. Nandi, "Continuous Indian sign language gesture recognition and sentence formation," *Procedia Comput. Sci.*, vol. 54, pp. 523–531, Jan. 2015.
[34] J. Huang, "Video-based sign language recognition without temporal segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 2257–2264.
[35] P. Chophuk, K. Pattanaworapn, and K. Chamnongthai, "Consideration of a selecting frame of finger-spelled words from backhand view," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2019.
[36] S. Yang and Q. Zhu, "Continuous Chinese sign language recognition with CNN-LSTM," in *Proc. 9th Int. Conf. Digit. Image, Int. Soc. Opt. Photon.*, vol. 10420, 2017, Art. no. 104200F.
[37] R. Cui, H. Liu, and C. Zhang, "Recurrent convolutional neural networks for continuous sign language recognition by staged optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7361–7369.
[38] Z. Yang, Z. Shi, X. Shen, and Y.-W. Tai, "SF-net: Structured feature network for continuous sign language recognition," 2019, *arXiv:1908.01341*. [Online]. Available: http://arxiv.org/abs/1908.01341
[39] Q. Xiao, X. Chang, X. Zhang, and X. Liu, "Multi-information spatial–temporal LSTM fusion continuous sign language neural machine translation," *IEEE Access*, vol. 8, pp. 216718–216728, 2020.
[40] O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Enabling robust statistical continuous sign language recognition via hybrid CNN-HMMs," *Int. J. Comput. Vis.*, vol. 126, no. 12, pp. 1311–1325, Dec. 2018.
[41] R. S. Rokade and D. D. Doye, "Spelled sentence recognition using radon transform," in *Proc. Sci. Inf. Conf.*, Aug. 2014, pp. 351–354.
[42] P. K. Athira, C. J. Sruthi, and A. Lijiya, "A signer independent sign language recognition with co-articulation elimination from live videos: An Indian scenario," *J. King Saud Univ. Comput. Inf. Sci.*, May 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S131915781831228X
[43] A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri, "A modified LSTM model for continuous sign language recognition using leap motion," *IEEE Sensors J.*, vol. 19, no. 16, pp. 7056–7063, Aug. 2019.
[44] B. Fang, J. Co, and M. Zhang, "DeepASL: Enabling ubiquitous and non-intrusive word and sentence-level sign language translation," in *Proc. 15th ACM Conf. Embedded Netw. Sensor Syst.*, 2017, pp. 1–13.
[45] D. Warchoł, T. Kapuściński, and M. Wysocki, "Recognition of fingerspelling sequences in polish sign language using point clouds obtained from depth images," *Sensors*, vol. 19, no. 5, p. 1078, Mar. 2019.
[46] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *Eur. Phys. J. Special Topics*, vol. 228, no. 10, pp. 2313–2324, Oct. 2019.

PONLAWAT CHOPHUK received the B.Eng. degree in electronic engineering and the M.Eng. degree in biomedical engineering from the King Mongkut's Institute of Technology Ladkrabang, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree in electrical and information engineering technology with the King Mongkut's University of Technology Thonburi. His current research interests include computer vision, image processing, human–computer interaction, and machine learning.

KOSIN CHAMNONGTHAI (Senior Member, IEEE) received the B.Eng. degree in applied electronics from The University of Electro-Communications, in 1985, the M.Eng. degree in electrical engineering from the Nippon Institute of Technology, in 1987, and the Ph.D. degree in electrical engineering from Keio University, in 1991. He is currently a Professor with the Department of Electronic and Telecommunication Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi. His research interests include computer vision, image processing, robot vision, signal processing, and pattern recognition. He is also a member of IEICE, TESA, ECTI, AIAT, APSIPA, TRS, and EEAAT. He is also the Vice President-Conference of APSIPA Association (2020-2021). He has served as an Editor for *ECTI E-Magazine* from 2011 to 2015, and an Associate Editor for *ECTI-EEC Transactions* from 2003 to 2010 and *ECTI-CIT Transactions* from 2011 to 2016. He has served as the Chairman for the IEEE COMSOC Thailand from 2004 to 2007 and the President for the ECTI Association from 2018 to 2019.