applied sciences

MDPI

Article

Non-Touch Sign Word Recognition Based on Dynamic Hand Gesture Using Hybrid Segmentation and CNN Feature Fusion

Md Abdur Rahim , Md Rashedul Islam and Jungpil Shin *

School of Computer Science and Engineering, The University of Aizu, Aizuwakamatsu, Fukushima 965-8580, Japan

* Correspondence: jpshin@u-aizu.ac.jp; Tel.: +81-242-37-2704

Received: 30 July 2019; Accepted: 6 September 2019; Published: 10 September 2019

Abstract: Hand gesture-based sign language recognition is a prosperous application of human–computer interaction (HCI), where the deaf community, hard of hearing, and deaf family members communicate with the help of a computer device. To help the deaf community, this paper presents a non-touch sign word recognition system that translates the gesture of a sign word into text. However, the uncontrolled environment, perspective light diversity, and partial occlusion may greatly affect the reliability of hand gesture recognition. From this point of view, a hybrid segmentation technique including YCbCr and SkinMask segmentation is developed to identify the hand and extract the feature using the feature fusion of the convolutional neural network (CNN). YCbCr performs image conversion, binarization, erosion, and eventually filling the hole to obtain the segmented images. SkinMask images are obtained by matching the color of the hand. Finally, a multiclass SVM classifier is used to classify the hand gestures of a sign word. As a result, the sign of twenty common words is evaluated in real time, and the test results confirm that this system can not only obtain better-segmented images but also has a higher recognition rate than the conventional ones.

Keywords: human–computer interaction; hybrid segmentation; sign word recognition; convolutional neural network (CNN); support vector machine (SVM)

1. Introduction

According to the World Health Organization (WHO) report, 5% of the world population in 2018 [1], 466 million people, have disabling hearing loss (adults and children comprise 432 and 34 million, respectively), and this is on the rise. Sign language serves as a useful communication medium for communicating with this community and the rest of the community. However, the deaf community, hard of hearing, and those with deaf family members use different body parts for communication in parallel, such as head, arm, or body, finger and hand movement, and facial expression, to provide information. Therein, the hand gesture is a promising part in human–computer interaction and is used for the very practical application of sign language recognition. Sign language recognition is a challenging problem due to the various illumination, uncontrolled environments, complexities of different signs, finger occlusions, and the visual analysis of the gestures of the hand. Therefore, this study focuses on the recognition of the sign word, which is a form of communication tool using hands that helps to reduce the communication gap by introducing a machine interface between the deaf community and the rest of the community. In previous studies, many scholars used hand gestures as a communication medium between human and machine. In [2], the authors described a review of sign language recognition (SLR) based on sensor gloves, which used handshape and movement information. However, the gesture recognition system using hand gloves requires many wearable connections for capturing of hands and finger movements, which prevent the convenience and naturalism of human–computer interaction. This is very difficult and uncomfortable for everyday use.

On the other hand, non-wearable devices, such as Microsoft Kinect, RGBD depth camera, leap motion, webcam, and others, have recently become widespread because they does not require attaching sensors to people. However, SLR becomes difficult due to the variety of signs, self-occlusion, background noise, or variation of illumination. In this study, we propose a hybrid segmentation along with a feature fusion of the CNN method to solve the above problems. According to do this, the input of isolated signs word is collected from live video using a webcam, and the proposed hybrid segmentation (YCbCr and SkinMask segmentation) technique is applied for preprocessing the input of hand gestures. The segmented image is then used to extract the features using CNN, and the fusion features are provided for gesture recognition. To achieve this goal, this paper has major contributions as follows.

- Hand gesture recognition performance is sub-optimal due to the uncontrolled environment, perspective light diversity, and partial occlusion. Considering the challenges of recognizing the gesture of a sign word, this system proposes a hybrid segmentation strategy that can easily detect the gesture of the hand. Hybrid segmentation can be defined as the coordination of the techniques of two segmentations like YCbCr and SkinMask. YCbCr segmentation converts the input images into YCbCr, then performs binarization, erosion, and fills in the holes. SkinMask segmentation converts the input images into HSV, and the range of H, S, and V values is measured based on the color range of skin. Therefore, the segmented images are provided for feature extraction.
- We propose a two-channel strategy of the convolutional neural network, which would be an input YCbCr, and the other would be SkinMask segmented images. The features of segmented images are extracted using CNN, and then, a fusion is applied in the fully connected layer. Furthermore, the fusion feature is fed into the classification process.
- A multiclass SVM classifier is used to classify the hand gestures, and the system displays related text.

We structured this paper as follows. Section 2 explains the relevant research. In Section 3, we discuss the workflow and propose a model. Datasets and experimental results are described in Section 4. Section 5 gives the conclusions of this paper.

2. Related Work

Hand gesture recognition contributes to improving the development of human–computer interaction (HCI) strategies. There are many studies related to hand gestures using wearable and non-wearable sensors. We explain a brief discussion of that work in this section.

Wu et al. proposed an inertial and EMG sensor-based American Sign Language (ASL) recognition system [3]. The feature selection strategy was applied for fusing information and classified using the selected classification model. They tested 80 signs and achieved 96.16% and 85.24% accuracy for intra-class and its cross-session evaluation, respectively. In [4], the author proposed an adaptive segmentation technique for Chinese sign language recognition using EMG sensors. Moreover, the use of human–computer interaction (HCI) in terms of the recognition of sign gesture has increased through EMG signals [5,6]. However, the noise of the input signal data may be a concern for the recognition of signs. Researchers did not apply any filtering methods, fed the acquired signal directly into the system, and achieved less accuracy compared to the trained system. A wearable device is used to recognize an Arab sign language developed using a modified k-NN method for forty sentences with an eighty-word lexicon [7]. An SVM classifier-based sign interpretation system was designed and implemented using smart wearable hand devices [8]. This system utilized flex sensors, pressure sensors, and three-in-one-speed sensors for distinguishing the ASL alphabet. However, wearable technology is uncomfortable for daily life, until there is great progress.

In [9], the author presented an ASL alphabet recognition system based on multi-view augmentation and inference fusion using CNN. This method retrieved 3D data from a depth image and created more perspective views for effective training and reduced overfitting. However, the diversity of the image could not allow recognizing the real-time gesture of a specific sign. Otiniano Rodriguez et al. proposed an ASL recognition system using the Kinect sensor [10]. They achieved better results than single data systems, compared to RGB images, depth images, and performance on both in this system. Rahim et al. also introduced the Kinect sensor for HCI [11]. The area of hand and fingertips was identified with contour extremes and palm position, and the gesture of the hand was recognized by measuring skeletal data. In [12], the author provided an invariant framework that was able to detect the occluded gestures. The work dealt with a hand gesture recognition system using Kinect sensor skeleton data. However, there may still be concerns about distance, and it is not clear how the "no gesture" situation was recognized by the presented algorithm. In [13], the Kinect sensing device was used to perform the comprehensive application of continuous activity recognition, and it determined the sequence of activities from 3D skeleton information. However, the authors did not explain how these were conducted during the tracking activity. Shin et al. proposed Japanese and English character input systems based on the hand tapping of gestures [14]. However, the system required large computational time, and the input characters were hard to remember by the user's hand tapping gestures. In [15], the author presented a 3D motion sensor-based ASL recognition system. They used k-NN and SVM to classify the 26 letters of the English alphabet in ASL derived from the sensory data. Based on shape and texture characteristics, the LBG (Linde-Buzo-Gray) vector quantization was applied to solve the SLR system [16]. However, the author used RGB information for recognition of sign language in poor lighting; therefore, these methods failed. Meanwhile, a direct method was used to detect the gestures in the color of RGB space. For instance, the preprocessing and recognition accuracy was improved in an incoherent environment with skin-colored objects [17]. The depth sensors were affected by the fusion color, and the depth information may improve segmentation by exceeding each modulation limit [18].

Recently, CNN has shown outstanding performance in various categories and for recognition. Xiao Yan Wu [19] introduced a novel approach to recognize hand gestures using the two input channels of CNN. However, it still needs much information for the double channel of CNN research and development, adaptability to complex backgrounds, dynamic gesture recognition, and data with labels for training. In [20], the author introduced a novel system based on a CNN and the classical characteristic of features. A binary, depth, and grayscale method was implemented on different datasets and performed gesture recognition in real time. Agrawal et al. provided a progress report highlighting Indian sign language recognition [21]. They discussed different segmentation and tracking strategies, extraction features, and classified them into different categories with various strategies such as providing the dependency or interdependence of the signer, manual or non-manual, a device-based system, vocabulary size, and the freedom of isolated or continuous signs.

3. Method of Sign Word Recognition System

The proposed approach was implemented by following the workflow shown in Figure 1. This system recognizes the isolated sign word based on hand segmentation and the fusion feature of the input images. Input images were obtained from live video in a region of interest (ROI) area. The input image was then segmented using the proposed techniques and the features extracted to feed into the classifier.

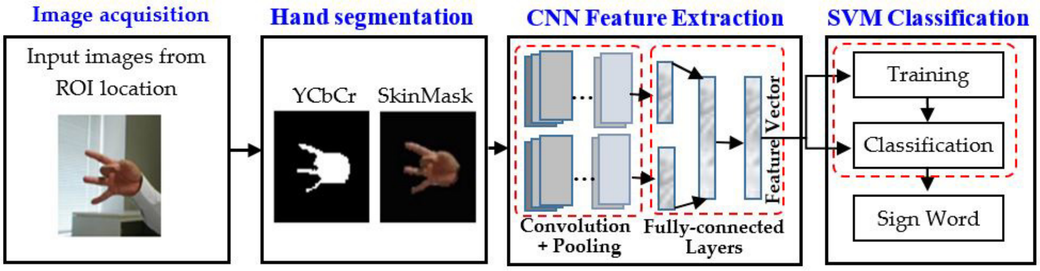


Figure 1. Proposed approach of the sign word recognition system.

3.1. Hand Segmentation Technique

Hand gesture segmentation is an essential part of sign word recognition. In this study, we propose hybrid segmentation techniques for segmenting hand from the input images. YCbCr and SkinMask were considered as a category of hybrid segmentation strategies, which were then integrated into a common vector.

3.1.1. YCbCr Segmentation

In this section, we segment the hand gesture from the RGB color space based on the chrominance-red (Cr) component of the YCbCr color space. The input image was converted from the RGB color space to a grayscale image of the YCbCr color space, which contained the components of luminance (Y) and the blue and red different components (Cb and Cr). The Cr component was extracted from YCbCr and used for further processing. Then, the extracted Cr image used the threshold method to process binary images. However, the converted image was exposed with two colors, black and white. The image grayscale value was 0-255; usually, zero is black, and 255 is white. In the segmentation process, we defined a threshold value of 128, which redefined the pixel values from 0–127 to zero and 128–255 to 255. Thereafter, we could process the grayscale image as a binary image. Then, erosion was applied to remove the boundary of the foreground pixels. As a result, the hole area became larger because the size of the foreground pixels may shrink. Finally, we achieved segmented images by filling the hole. Figure 2 represents the block diagram of the segmentation steps, and Figure 3 depicts the YCbCr segmentation process of an input image.

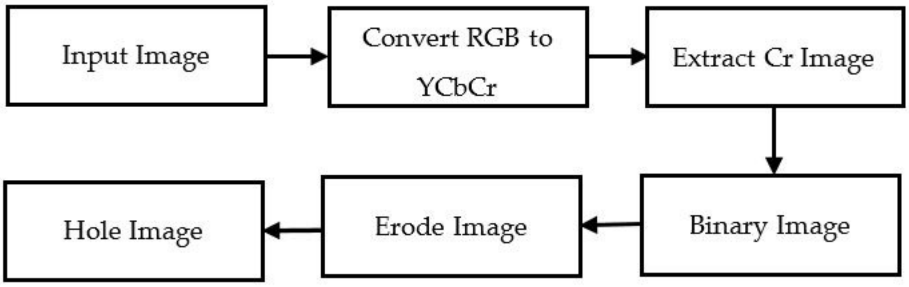


Figure 2. Block diagram of the YCbCr segmentation process.

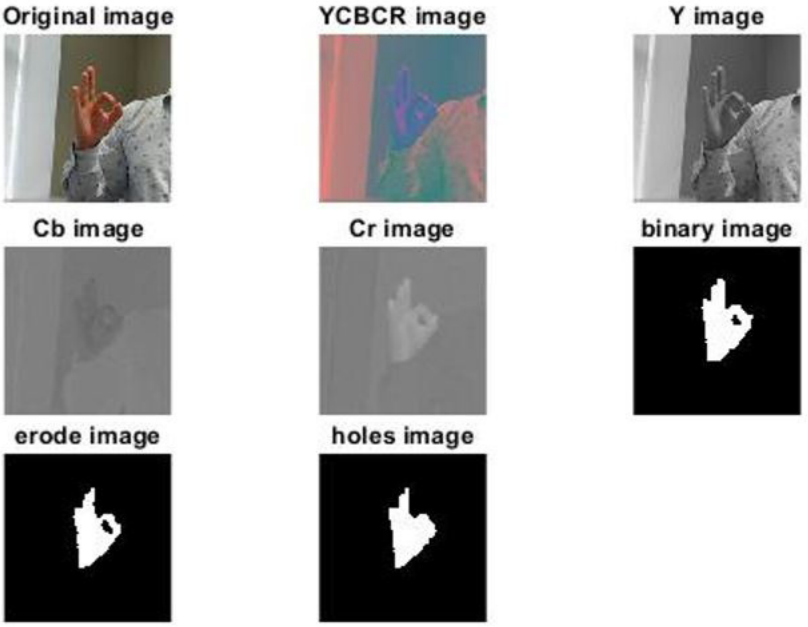


Figure 3. YCbCr segmentation process of an input image.

3.1.2. SkinMask Segmentation

To detect the hand, we converted the input image into HSV, which contained the hue (H), saturation (S), and value (V) components. The HSV value of each pixel was compared to the pixel quality of the skin and measured in a standard range, which depended on whether the pixel was a skin pixel or the value had a range of predefined threshold values. However, threshold masking was applied to determine which pixels should be presented as its dominant feature. Therefore, we implemented morphological processing (MP), which helped to remove noise and clutter from the image obtained in the output image. The MP created a new binary image in which pixels were only a non-zero value. Then, this method considered the connected region, which ignores small areas that are not possible at all. However, the area of the gesture's skin region was calculated by computing the number of pixels in the skin area of the region. Figure 4 shows the block diagram of the SkinMask process, and Figure 5 shows image preprocessing in a SkinMask segmentation technique.

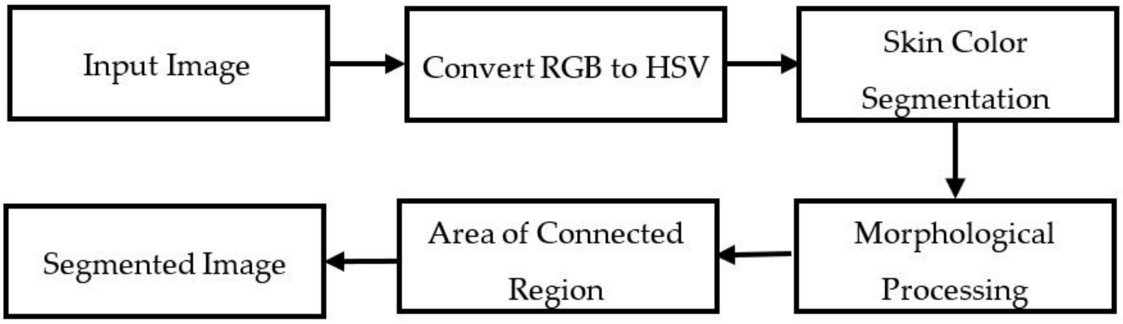


Figure 4. Block diagram of the SkinMask segmented process.

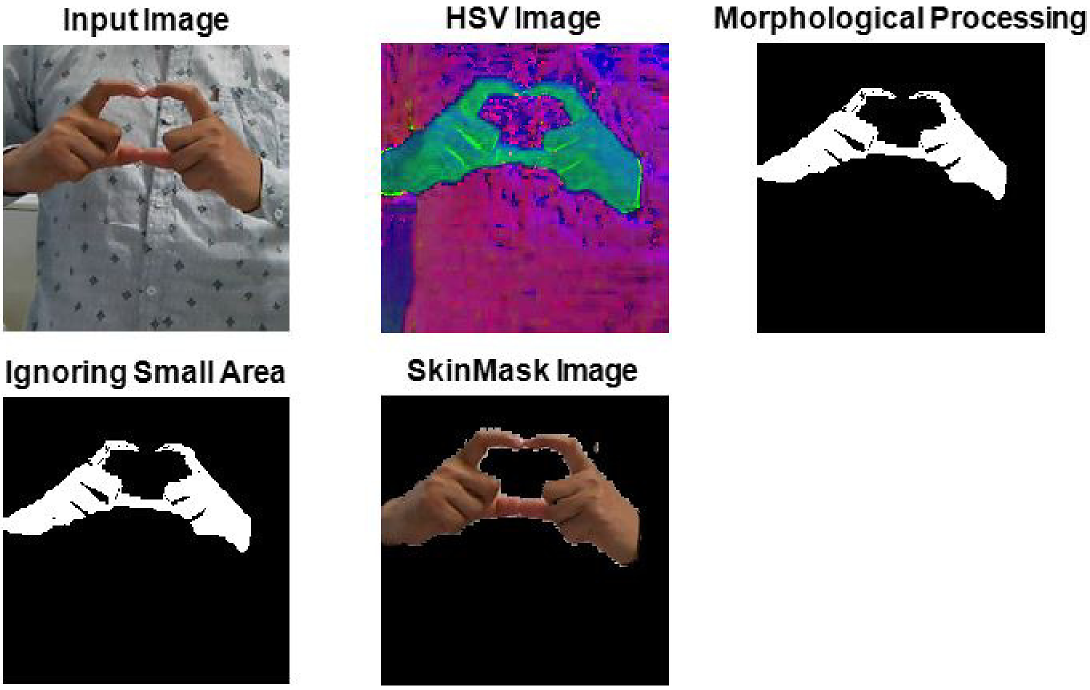


Figure 5. SkinMask segmentation process of an input image.

3.2. CNN Feature Extraction

As a popular class of machine learning techniques, the convolutional neural network (CNN) has expanded significantly in the technological advances of human–computer interaction [20]. The CNN described in Figure 6 included convolutional, pooling, fully-connected layer, activation function, and the classifier. In the proposed architecture, the convolution was executed in the input data using a filter or kernel to create a feature map. However, we conducted a convolution by sliding the input filter. Numerous convolutions were applied to inputs, where each activity used a different filter. Finally, we took a map of various features and kept them together as the ultimate output of the convolutional layer. The output of the convolution passed through the activation functions such as ReLU functions. To prevent feature shrinking of these maps, padding to an image was introduced. It added the layer of zero-value pixels around the input with zeros. After the level of convolution, a pooling layer was added within the CNN layer. The pooling layer can reduce the data layer while saving feature information. The most common pooling, maxpooling, was used, which reduces the size of feature maps and at the same time keeps important information. After the convolution and pooling layer, the classification was performed in the fully-connected layer. This fully-connected layer can only accept one-dimensional data that can convert 2D data to 1D data, and we used the Python Flatten function. Finally, the fully-connected layer integrated all the features and provided them to the multiclass SVM classifier.

For the training of the process, the two-channel CNN was used to recognize the gesture of the sign word. This architecture consisted of two relatively independent convolutional neural networks. The inputs of each channel were the YCbCr and SkinMask segmented gesture. The convolution layer and parameters were the same as each channel, but the weight was independent. After the pooling layer, two channels were connected to a fully-connected layer, and a full-scale map was implemented. The fully-connected hidden layer acted as a connector for two channels, which created the outputs of the SVM classifier.

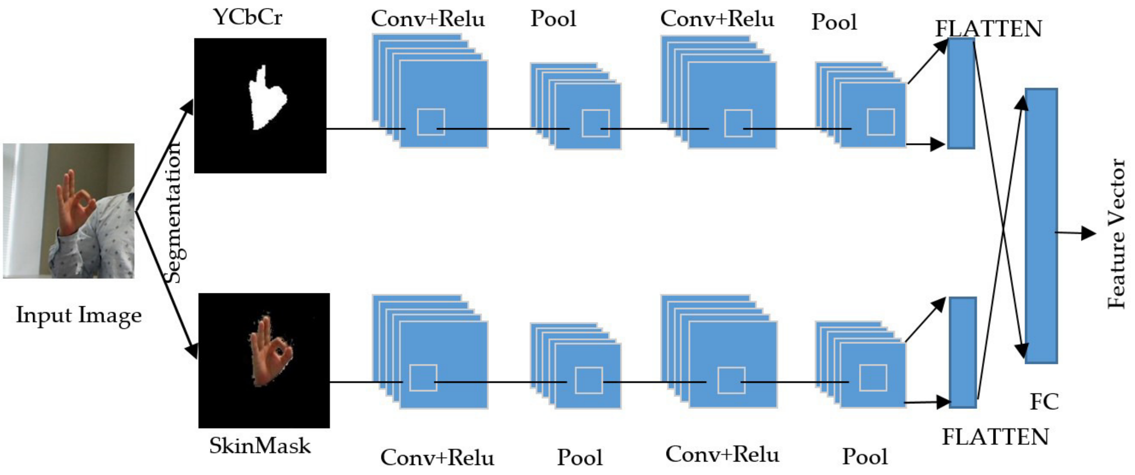


Figure 6. The architecture of the proposed feature extraction model.

3.3. SVM Classification

The support vector machine (SVM) technique creates a hyperplane for classification and regression in N-dimensional (the number of features) space [22]. This classification was done by a hyperplane, which was the largest classification for the nearest training data points of any category. In this study, we used multiclass SVM, which used labels from the feature vector. To create a binary classifier, we introduced one versus the rest (OVR), that allocated the classified class with the highest output function. In this case, the kernel function was invoked to classify non-linear datasets, which converted the lower-dimensional input space into a high-dimensional space. We selected the RBF (radial basis kernel) for the SVM's functionality, which can be the localization and limited response across the entire range of the main axis. Therefore, multi-class OVR SVMs were working in parallel, which separated one class from the rest, as shown in Equations (1) and (2). Using each support vector, this system measured the maximized output value of SVM for the classification. Figure 7 shows the OVR multiclass SVM approach to classify the gestures.

$$f_i(x) = w_i^t x + b_i \tag{1}$$

$$X \longmapsto arg \max_i f_i(x) \tag{2}$$

where the ith decision function classifies class i with positive labels and remaining classes as negative labels. $f_i(x)$ the N-dimensional vector and scalar is b_i, and X classifies the maximum output.

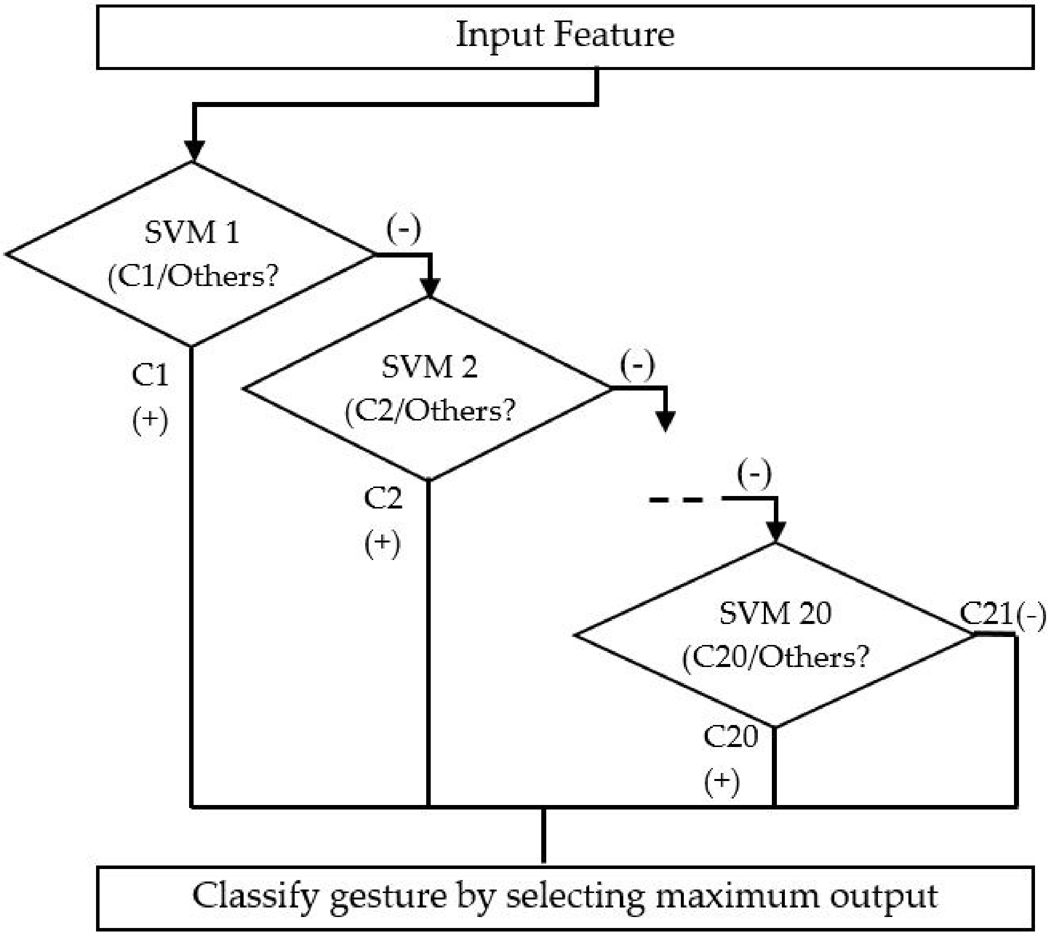


Figure 7. One-versus-rest multiclass SVM structure for classification.

4. Experimental Dataset and Simulation Results

In this section, a comprehensive evaluation if conducted of the proposed model for sign word recognition and real-time assessment, which converts the gesture of the sign word into text.

4.1. Dataset Description

To evaluate the proposed model, a dataset was constructed, and it is available online at this URL: https://www.u-aizu.ac.jp/labs/is-pp/pplab/swr/sign_word_dataset.zip. There were twenty isolated hand gestures (11 single-hand gestures and nine double hand gestures). The images of the dataset were collected with a pixel resolution of 200 × 200. Figure 8 depicts the example of the dataset images. To collect dataset images, three volunteers (mean age of 25) were requested to perform the gesture of the sign word. We collected three hundred images for each gesture sign from each individual. There was a total of 18,000 images of 20 isolated hand gestures. While capturing the images, we considered the complex background, different lighting illumination, different directions, and skin-like backgrounds.

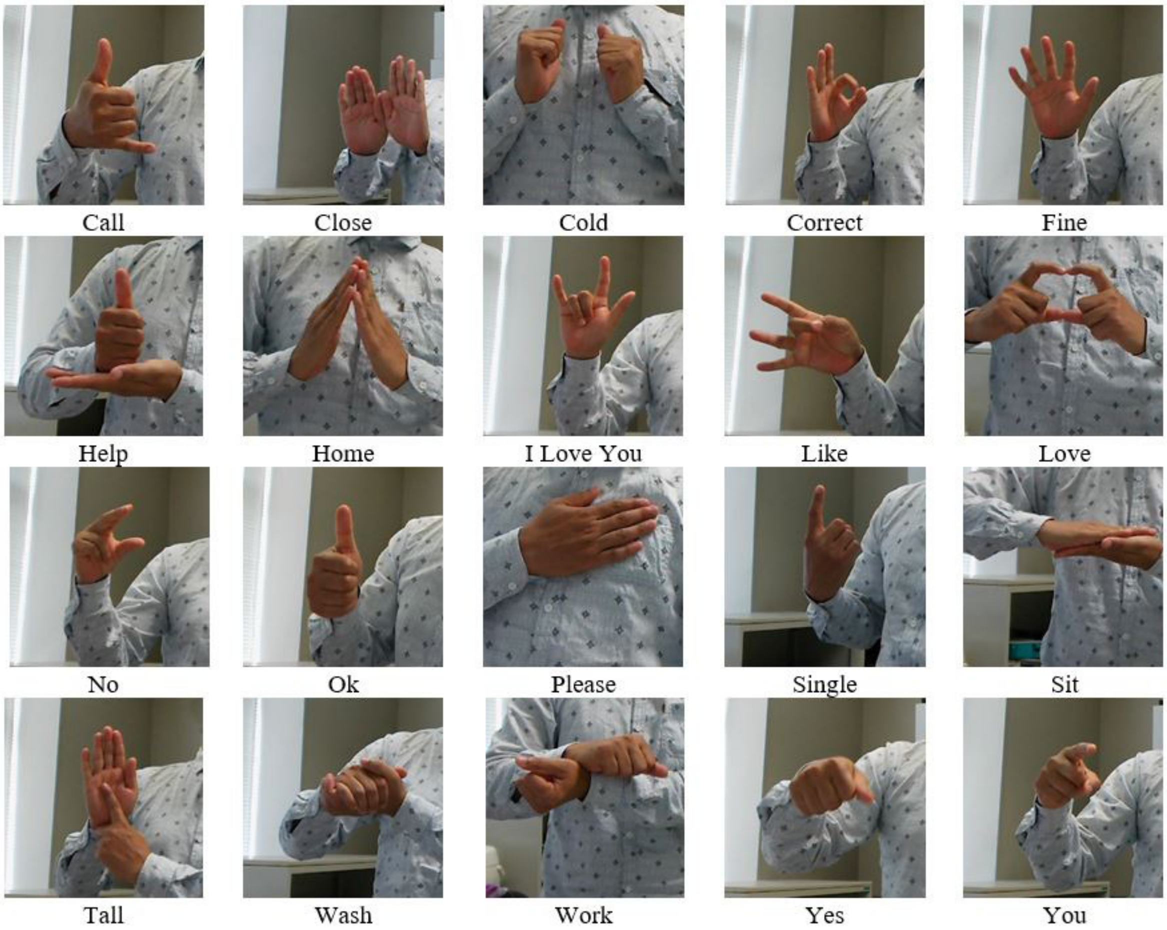


Figure 8. Example of the dataset images.

4.2. Hand Gesture Segmentation

For the model, we did not select the original image as the direct input, because the process completion time would be huge, and the results of classification would be affected by the redundant background. To avoid this situation, the hand gesture image was segmented, then used as the input of the model. In this study, we used hybrid segmentation methods for preprocessing dataset images. To identify the hands, we analyzed the Cr components of the YCbCr color space and identified the pixels that matched the color of the hand with a certain frame of HSV analysis of the SkinMask process. Therefore, the segmented images were fed into the feature extraction process. Figures 9 and 10 show the example of the segmented images of the YCbCr and SkinMask segmentation method. Furthermore, we considered the Sebastien Marcel Static Hand Posture Database [23] for comparison to our proposed method. Therefore, we applied our method to segment hand gestures and then transferred them to feature extraction and classification. Figure 11 shows the sample images of hand gestures in the dataset of [23] and segmented images using the YCbCr and the SkinMask segmentation method.

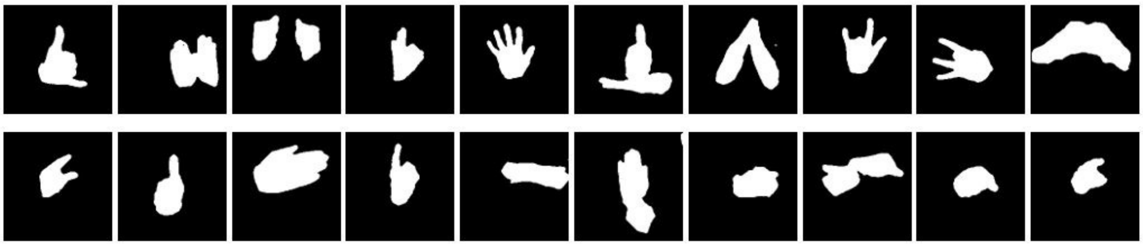


Figure 9. Example of the YCbCr segmented images from the datasets.

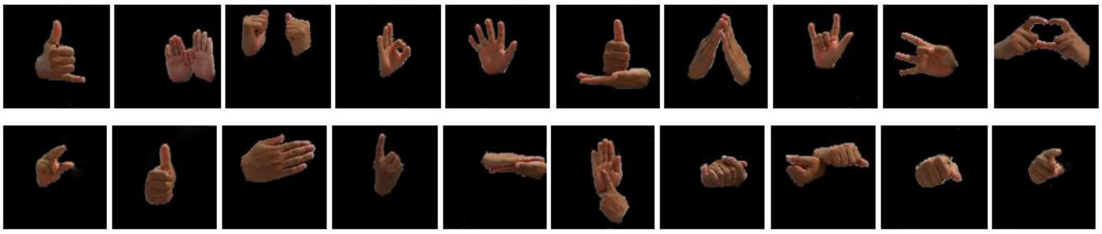


Figure 10. Example of the SkinMask segmented images from the datasets.

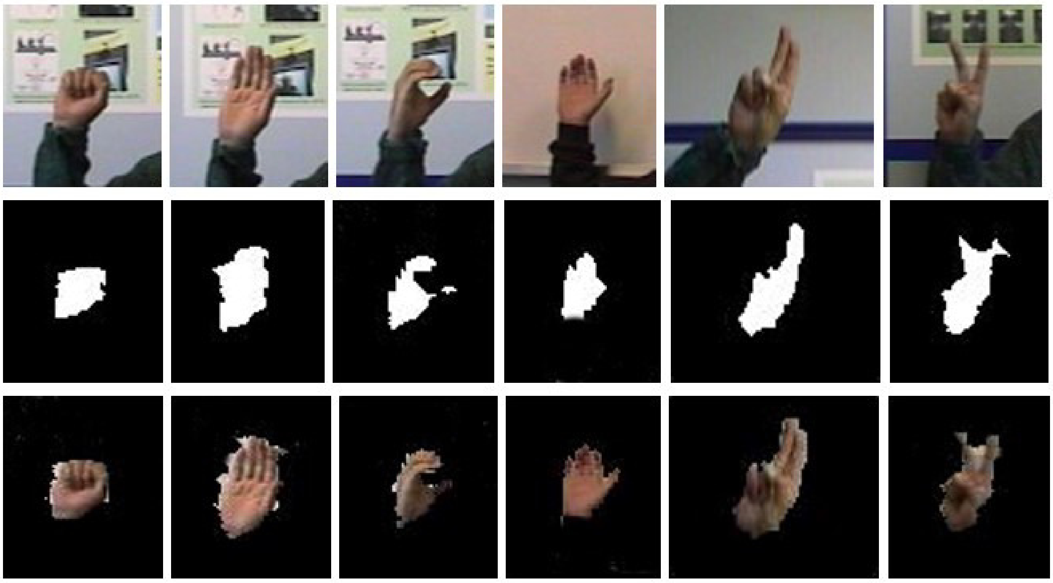


Figure 11. Sample images of hand gestures in the benchmark dataset, YCbCr image, and SkinMask segmented images (first row, second row, and third row, respectively).

4.3. Feature Extraction and Classification

To recognize the sign word gesture, we evaluated different signs of isolated hand gestures. The proposed CNN with feature fusion method was trained on the entire dataset. In this case, YCbCr and SkinMask segmented images were used as the inputs for the two channels of the proposed CNN model. We used a constructed dataset containing twenty classes, with 18,000 hand gesture images. We used 70% of the images for training purposes and 30% for testing. The features were extracted and fused at the second level of the fully-connected layer. A multi-class SVM classifier was used to classify the different hand gestures and interpret them as text. The average accuracy of the sign word recognition was about 97.28%. This system achieved the highest accuracy of the sign gestures on the "call", "correct", and "fine" gestures, respectively. Some sign gestures were apparently very similar in shape, leading to the misrecognition, which was then recognized by pose changes. For example, "OK" and "help" and "cold" and "work" are shown in Figure 12. The confusion matrix of Figure 13 shows the recognition accuracy of our model. In terms of accuracy, input feature sets were also fed with the softmax classification for performance comparison. Figure 14 shows the accuracy of the two classifications of softmax and SVM. Table 1 represents the accuracy of the comparison with state-of-the-art algorithms. As a comparison, we implemented different state-of-the-art algorithms under the same dataset, and the recognition accuracy is shown in Figure 15. The average recognition accuracy using the YCbCr + CNN [24] and SkinMask + CNN [25] methods was 96.58% and 96.11%, respectively. From the results, it can be said that the experimental evaluation of our proposed method improved the recognition accuracy. However, using the SVM classifier, our method achieved 97.28% accuracy, which was higher than the softmax classification. Furthermore, the considered databases were classified using the proposed model. The average classification accuracy is shown in Figure 16.

The main interface of sign word recognition is shown in Figure 17. The users were requested to perform hand gestures in the ROI location. The system recognized each gesture and interpreted it as text. Figure 18 represents a user simulation of real-time sign word recognition. The system evaluated each gesture during processing, which involved image pre-processing, feature extraction, and sign recognition. The total time to recognize each gesture was 26.03 ms, which was sufficient for real-time recognition at 30 fps. However, the frame rate was limited by the webcam to 33.33 ms per frame. We evaluated the average recognition accuracy in real-time of fifteen participants (mean age 22, male 13, and female 2) who performed five tasks (one task had twenty sign gestures). Figure 19 shows the average accuracy of hand gesture performances for all participants.

Table 1. Accuracy comparison with state-of-the-art systems (evaluated using our dataset).

Reference	Method	Reported Accuracy (%)	Classification Accuracy (%)	
			Softmax	SVM
[25]	Skin Model and CNN	95.96	95.26	96.11
[26]	CNN	91.7	93.61	94.7
[24]	YCbCr and CNN	96.2	95.37	96.58
Proposed	Hybrid segmentation and double channel of CNN	-	96.29	97.28

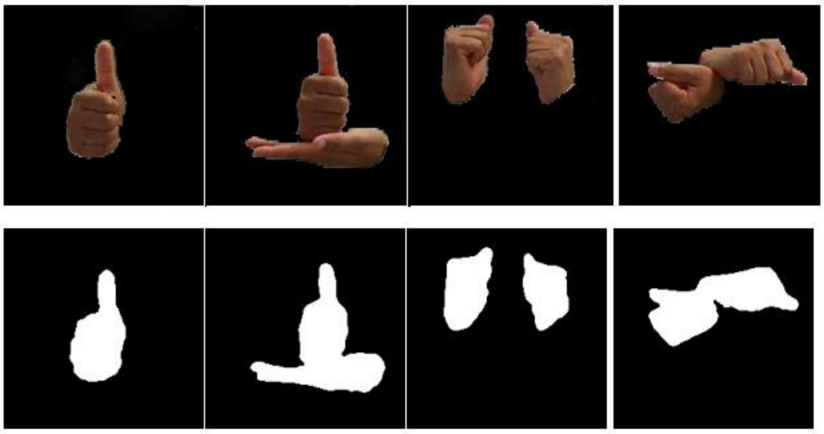
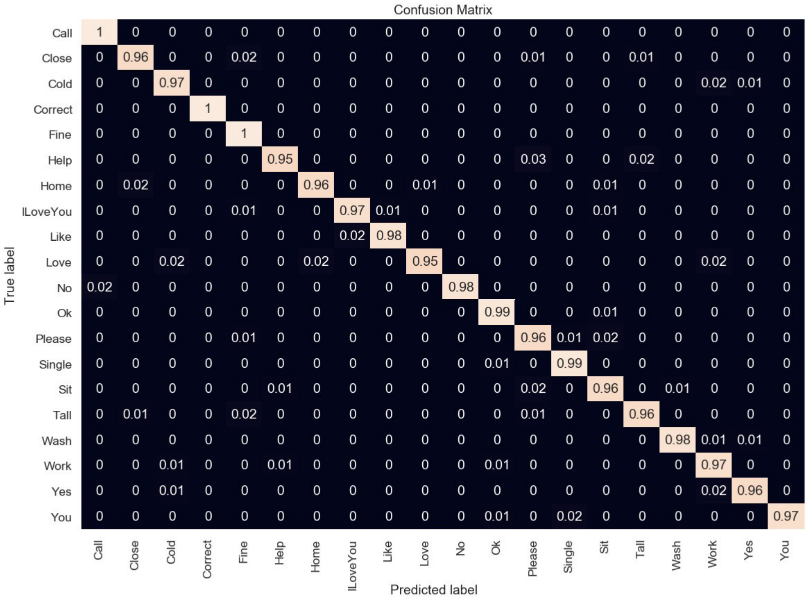


Figure 12. Misrecognized hand gestures.



	Call	Close	Cold	Correct	Fine	Help	Home	ILoveYou	Like	Love	No	Ok	Please	Single	Sit	Tall	Wash	Work	Yes	You
Call	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Close	0	0.96	0	0	0.02	0	0	0	0	0	0	0	0.01	0	0	0.01	0	0	0	0
Cold	0	0	0.97	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.01	0
Correct	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Fine	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Help	0	0	0	0	0	0.95	0	0	0	0	0	0	0.03	0	0	0.02	0	0	0	0
Home	0	0.02	0	0	0	0	0.96	0	0	0.01	0	0	0	0	0.01	0	0	0	0	0
ILoveYou	0	0	0	0	0.01	0	0	0.97	0.01	0	0	0	0	0	0.01	0	0	0	0	0
Like	0	0	0	0	0	0	0	0.02	0.98	0	0	0	0	0	0	0	0	0	0	0
Love	0	0	0.02	0	0	0	0.02	0	0	0.95	0	0	0	0	0	0	0	0.02	0	0
No	0.02	0	0	0	0	0	0	0	0	0	0.98	0	0	0	0	0	0	0	0	0
Ok	0	0	0	0	0	0	0	0	0	0	0	0.99	0	0	0.01	0	0	0	0	0
Please	0	0	0	0	0.01	0	0	0	0	0	0	0	0.96	0.01	0.02	0	0	0	0	0
Single	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.99	0	0	0	0	0	0
Sit	0	0	0	0	0	0.01	0	0	0	0	0	0	0.02	0	0.96	0	0.01	0	0	0
Tall	0	0.01	0	0	0.02	0	0	0	0	0	0	0	0.01	0	0	0.96	0	0	0	0
Wash	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.98	0.01	0.01	0
Work	0	0	0.01	0	0	0.01	0	0	0	0	0	0.01	0	0	0	0	0	0.97	0	0
Yes	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.96	0
You	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.02	0	0	0	0	0	0.97

Figure 13. Confusion matrix of recognition accuracy.

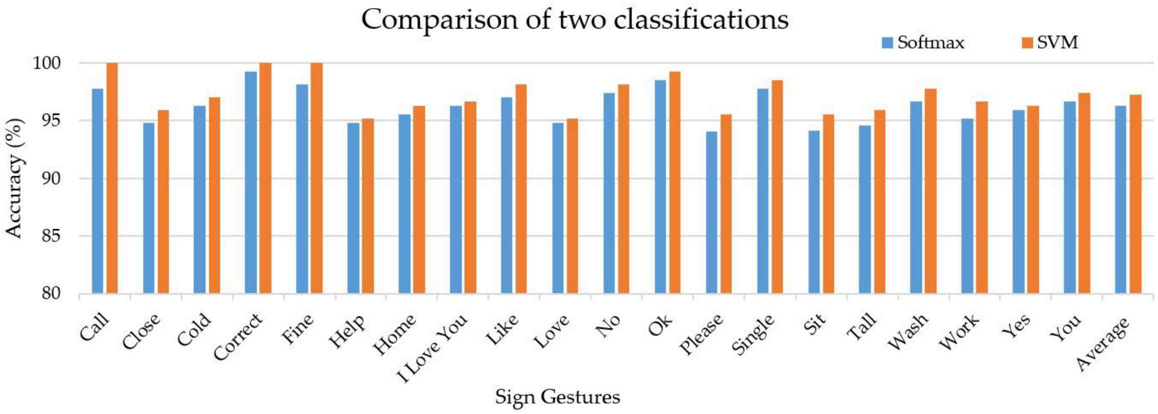


Figure 14. Performance evaluation of the two classifications, softmax and SVM.

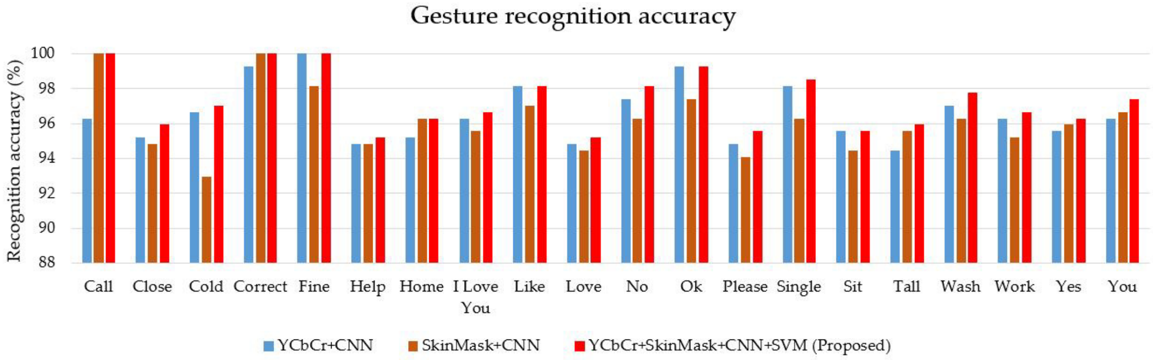


Figure 15. Comparison of average recognition of hand gestures.

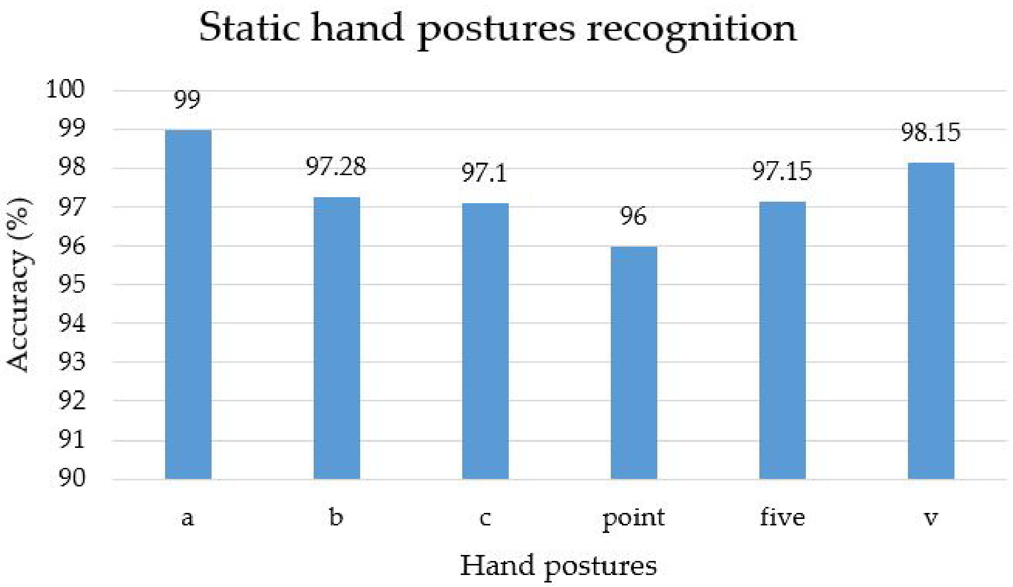
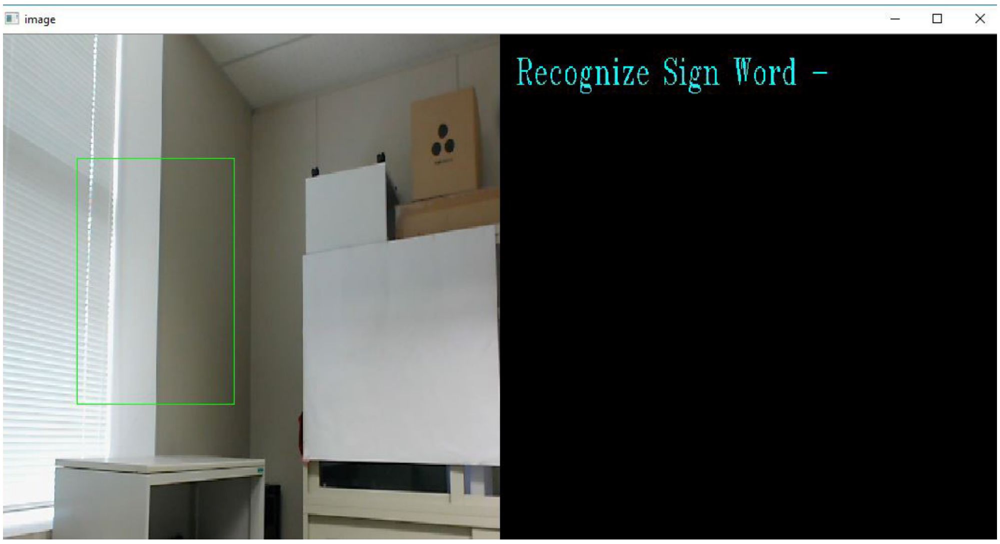
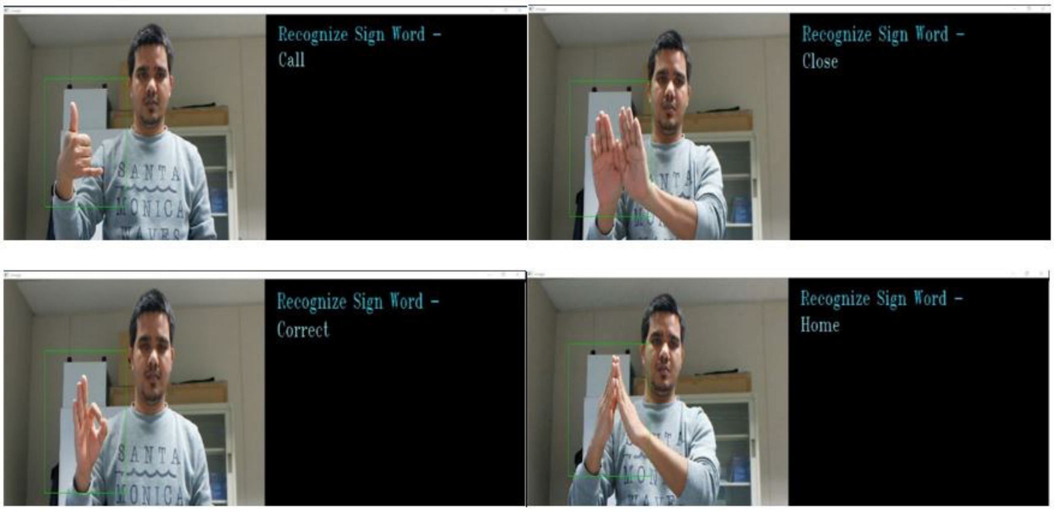


Figure 16. Recognition accuracy of static hand postures using benchmark dataset.



Recognize Sign Word -

Figure 17. Main interface of the proposed system.



Recognize Sign Word -
Call

Recognize Sign Word -
Close

Recognize Sign Word -
Correct

Recognize Sign Word -
Home

Figure 18. Simulation of the sign word recognition system.

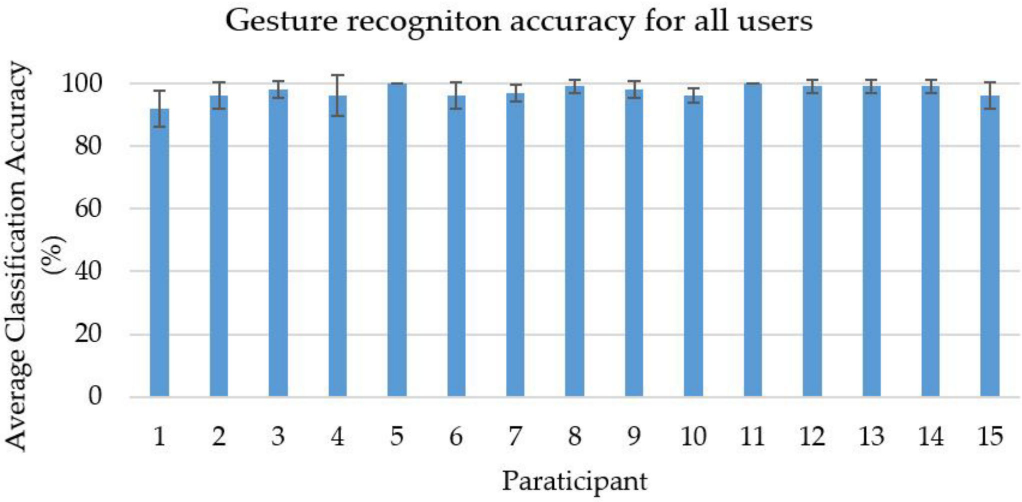


Figure 19. Average accuracy of the gesture recognition of all participants.

5. Conclusions

A hybrid segmentation along with the feature fusion-based sign word recognition system was presented in this paper. For this purpose, a model was proposed that was able to translate sign gestures into text. The proposed model included image preprocessing, feature extraction and fusion, and gesture classification. To detect the hand gestures, we preprocessed input images using YCbCr and SkinMask segmentation. Cr component analysis was considered in the YCbCr segmentation, and SkinMask identified the pixels that corresponded to the skin color of the hand. Therefore, we used the proposed model to extract features from segmented images, where YCbCr and SkinMask segmented images were the CNN's two-channel inputs. The feature fusion was accomplished in the fully-connected layer. At the level of classification, the multiclass SVM classifier was compiled by the hand gesture dataset created by the authors. The results indicated that in the real-time environment, approximately 97.28% accuracy was achieved using trained features and the SVM classifier, and it led to better results than the state-of-the-art systems.

Author Contributions: Conceptualization, M.A.R., M.R.I. and J.S.; methodology, M.A.R.; software, M.A.R. and M.R.I.; validation, M.A.R., M.R.I. and J.S.; formal analysis, M.A.R. and J.S.; investigation, M.A.R. and J.S.; resources, J.S.; data curation, M.A.R. and M.R.I.; writing–original draft preparation, M.A.R; writing–review and editing, M.A.R, M.R.I and J.S.; visualization, M.A.R.; supervision, J.S.; project administration, J.S.; funding acquisition, J.S. This paper was prepared the contributions of all authors. All authors have read and approved the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. World Health Organization (WHO). Available online: www.who.int/deafness/world-hearing-day/whd-2018/en (accessed on 20 July 2019).
2. Ahmed, M.A.; Zaidan, B.B.; Zaidan, A.A.; Salih, M.M.; Lakulu, M.M. A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017. *Sensors* **2018**, *18*, 2208. [CrossRef] [PubMed]
3. Wu, J.; Sun, L.; Jafari, R. A wearable system for recognizing American sign language in real-time using IMU and surface EMG sensors. *IEEE J. Biomed. Health Inform.* **2016**, *20*, 1281–1290. [CrossRef] [PubMed]
4. Li, Y.; Chen, X.; Zhang, X.; Wang, K.; Wang, Z.J. A sign-component based framework for Chinese sign language recognition using accelerometer and sEMG data. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 2695–2704. [PubMed]
5. Sun, Y.; Li, C.; Li, G.; Jiang, G.; Jiang, D.; Liu, H.; Zheng, Z.; Shu, W. Gesture recognition based on kinect and sEMG signal fusion. *Mob. Netw. Appl.* **2018**, *1*, 1–9. [CrossRef]
6. Gupta, H.P.; Chudgar, H.S.; Mukherjee, S.; Dutta, T.; Sharma, K. A continuous hand gestures recognition technique for human-machine interaction using accelerometer and gyroscope sensors. *IEEE Sens. J.* **2016**, *16*, 6425–6432. [CrossRef]
7. Tubaiz, N.; Shanableh, T.; Assaleh, K. Glove-based continuous Arabic sign language recognition in user-dependent mode. *IEEE Trans. Hum. Mach. Syst.* **2015**, *45*, 526–533. [CrossRef]
8. Lee, B.G.; Lee, S.M. Smart wearable hand device for sign language interpretation system with sensors fusion. *IEEE Sens. J.* **2018**, *18*, 1224–1232. [CrossRef]
9. Tao, W.; Leu, M.C.; Yin, Z. American Sign Language alphabet recognition using Convolutional Neural Networks with multiview augmentation and inference fusion. *Eng. Appl. Artif. Intell.* **2018**, *76*, 202–213. [CrossRef]
10. Rodriguez, K.O.; Chavez, G.C. Finger spelling recognition from RGB-D information using kernel descriptor. In Proceedings of the IEEE XXVI Conference on Graphics, Patterns and Images, Arequipa, Peru, 5–8 August 2013; pp. 1–7.
11. Rahim, M.A.; Shin, J.; Islam, M.R. Human-Machine Interaction based on Hand Gesture Recognition using Skeleton Information of Kinect Sensor. In Proceedings of the 3rd International Conference on Applications in Information Technology, Aizu-Wakamatsu, Japan, 1–3 November 2018; pp. 75–79.

12. Kumar, P.; Saini, R.; Roy, P.P.; Dogra, D.P. A position and rotation invariant framework for sign language recognition (SLR) using Kinect. *Multimed. Tools Appl.* **2018**, *77*, 8823–8846. [CrossRef]
13. Saini, R.; Kumar, P.; Roy, P.P.; Dogra, D.P. A novel framework of continuous human-activity recognition using kinect. *Neurocomputing* **2018**, *311*, 99–111. [CrossRef]
14. Shin, J.; Kim, C.M. Non-touch character input system based on hand tapping gestures using Kinect sensor. *IEEE Access* **2017**, *5*, 10496–10505. [CrossRef]
15. Chuan, C.H.; Regina, E.; Guardino, C. American sign language recognition using leap motion sensor. In Proceedings of the IEEE 13th International Conference on Machine Learning and Applications, Detroit, MI, USA, 3–6 December 2014; pp. 541–544.
16. Raut, K.S.; Mali, S.; Thepade, S.D.; Sanas, S.P. Recognition of American sign language using LBG vector quantization. In Proceedings of the IEEE International Conference on Computer Communication and Informatics, Coimbatore, India, 3–5 January 2014; pp. 1–5.
17. Pisharady, P.K.; Saerbeck, M. Recent methods and databases in vision-based hand gesture recognition: A review. *Comput. Vis. Image Underst.* **2015**, *141*, 152–165. [CrossRef]
18. D'Orazio, T.; Marani, R.; Renó, V.; Cicirelli, G. Recent trends in gesture recognition: How depth data has improved classical approaches. *Image Vis. Comput.* **2016**, *52*, 56–72. [CrossRef]
19. Wu, X.Y. A hand gesture recognition algorithm based on DC-CNN. *Multimed. Tools Appl.* **2019**, 1–13. [CrossRef]
20. Chevtchenko, S.F.; Vale, R.F.; Macario, V.; Cordeiro, F.R. A convolutional neural network with feature fusion for real-time hand posture recognition. *Appl. Soft Comput.* **2018**, *73*, 748–766. [CrossRef]
21. Agrawal, S.C.; Jalal, A.S.; Tripathi, R.K. A survey on manual and non-manual sign language recognition for isolated and continuous sign. *Int. J. Appl. Pattern Recognit.* **2016**, *3*, 99–134. [CrossRef]
22. Gholami, R.; Fakhari, N. Support Vector Machine: Principles, Parameters, and Applications. In *Handbook of Neural Computation*; Academic Press: Cambridge, MA, USA, 2017; pp. 515–535
23. Marcel, S.; Bernier, O. Hand posture recognition in a body-face centered space. In *International Gesture Workshop*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 97–100.
24. Flores, C.J.L.; Cutipa, A.G.; Enciso, R.L. Application of convolutional neural networks for static hand gestures recognition under different invariant features. In Proceedings of the 2017 IEEE XXIV International Conference on Electronics, Electrical Engineering and Computing (INTERCON), Cusco, Peru, 15–18 August 2017; pp. 1–4.
25. Lin, H.I.; Hsu, M.H.; Chen, W.K. Human hand gesture recognition using a convolution neural network. In Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE), Taipei, Taiwan, 18–22 August 2014; pp. 1038–1043.
26. Pigou, L.; Dieleman, S.; Kindermans, P.J.; Schrauwen, B. Sign language recognition using convolutional neural networks. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 572–578.