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.

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).

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..

Introduction

According to the World Health Organization (WHO) report, 5% of the world population in 2018, 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. Therefore, we developed a non-touch system for communication between the deaf community and the rest of the community using hand gestures. 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.

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.

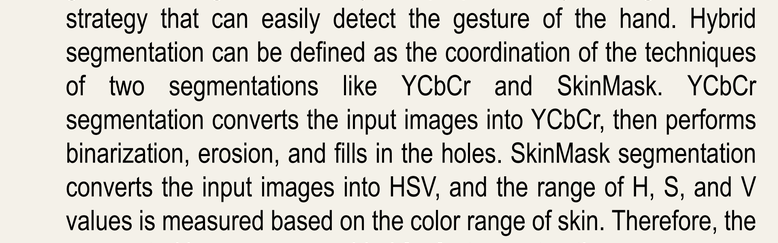

Figure 1. Proposed approach of the sign word recognition system.

Hand Segmentation Technique

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.

YCbCr Segmentation

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.

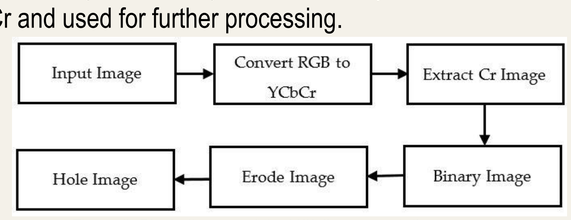

Figure 2. Block diagram of the YCbCr segmentation process.

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.

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.

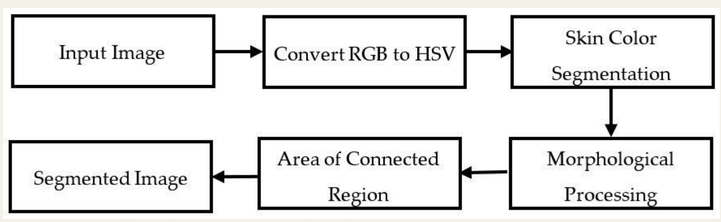

Figure 3. Block diagram of the SkinMask segmented process.

CNN Feature Extraction

The CNN described in Figure 4 included convolutional, pooling, fully connected layer, activation function, and the classifier.

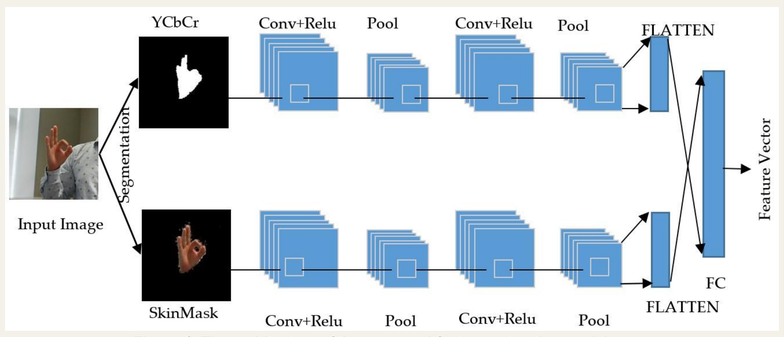

Figure 4. The architecture of the proposed feature extraction model.

SVM Classification

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).

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

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

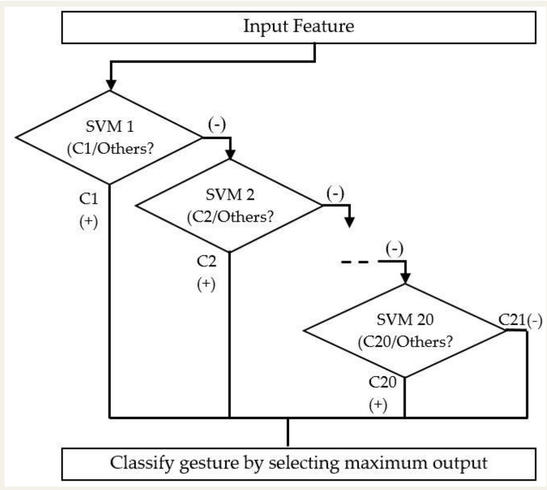

Figure 5. One-versus-rest multiclass SVM structure for classification.

Experimental Dataset and Simulation Results

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.

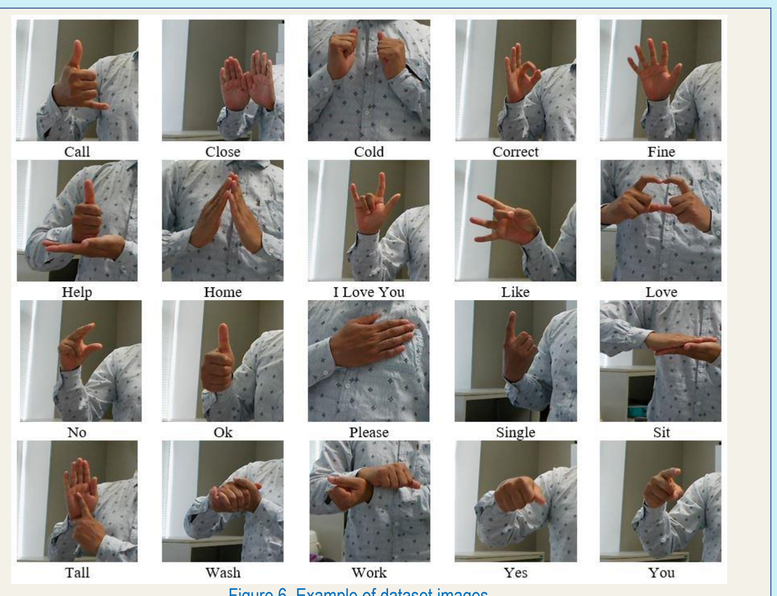

Figure 6. Example of dataset images.

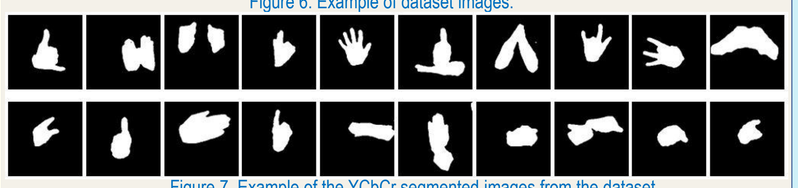

Figure 7. Example of the YCbCr segmented images from the dataset.

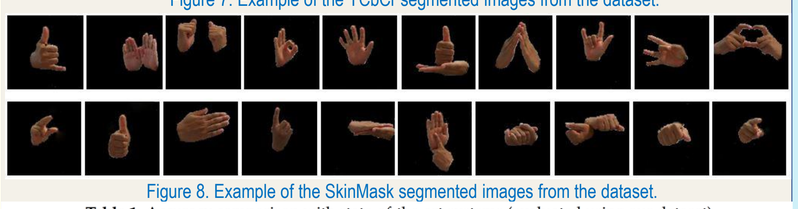

Figure 8. Example of the SkinMask segmented images from the dataset.

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

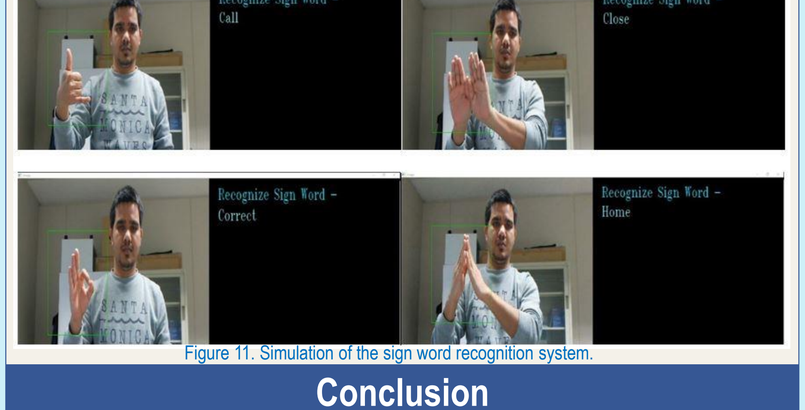

Figure 9. Comparison of average recognition of hand gestures.

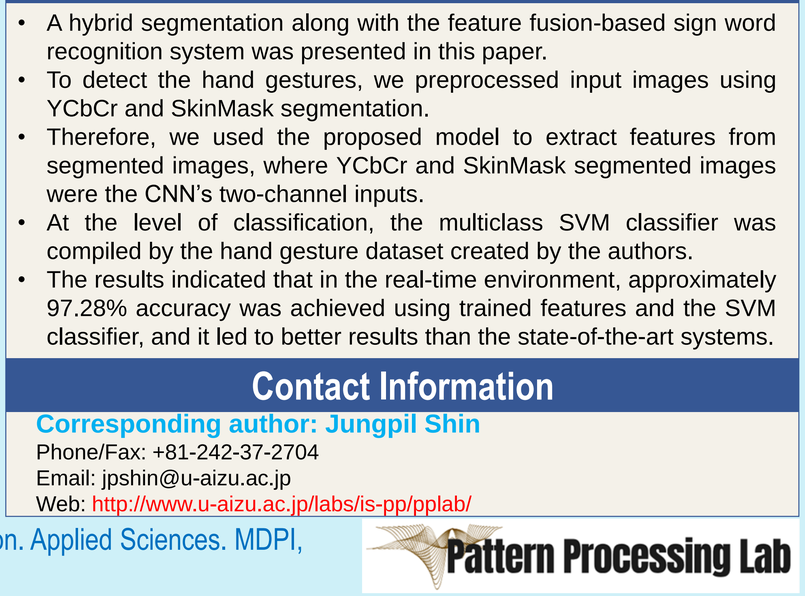

Figure 10. Confusion matrix of recognition accuracy.

Figure 11. Simulation of the sign word recognition system.

Conclusion

- A hybrid segmentation along with the feature fusion-based sign word recognition system was presented in this paper.
- To detect the hand gestures, we preprocessed input images using YCbCr and SkinMask segmentation.
- 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.
- 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.

Contact Information

Corresponding author: Jungpil Shin

Phone/Fax: +81-242-37-2704

Email: jpshin@u-aizu.ac.jp

Web: http://www.u-aizu.ac.jp/labs/is-pp/pplab/

Cite Rahim MA, Islam MR, Shin J. Non-Touch Sign Word Recognition Based on Dynamic Hand Gesture Using Hybrid Segmentation and CNN Feature Fusion. Applied Sciences. MDPI, 2019 Sep; 9(18): 3790.