Poster ID: 29

Rotation, Translation And Scale Invariant Sign Word Recognition Using Deep Learning

Abu Saleh Musa Miah Jungpil Shin, Md. Al Mehedi Hasan, Md Abdur Rahim and Yuichi Okuyama

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

ABSTRACT

Communication between people with disabilities and people who do not understand sign language is a growing social need and can be a tedious task. One of the main functions of sign language is to communicate with each other through hand gestures. To resolve Existing challenges of hand gesture recognition, we proposed a Rotation, Translation and Scale-invariant sign word recognition system using a convolutional neural network (CNN). We have followed three steps in our work: rotated, translated and scaled (RTS) version dataset generation, gesture segmentation, and sign word classification.

Firstly, we have enlarged a benchmark dataset of 20 sign words by making different amounts of Rotation, Translation and Scale of the original images to create the RTS version dataset.

Then we have applied the gesture segmentation technique. The segmentation consists of three levels, i) Otsu Thresholding with YCbCr, ii) Morphological analysis: dilation through opening morphology and iii) Watershed algorithm.

Finally, our designed CNN model has been trained to classify the hand gesture as well as the sign word. Our model has been evaluated using the twenty sign word dataset, five sign word dataset and the RTS version of these datasets and achieved high performance.

INTRODUCTION

Sign language is a nonverbal form of communication for the deaf and hard-of-hearing community. It is essential to help real-time communication among the deaf community, hard of hearing, and speech difficulties without the aid of an interpreter. According to the World Health Organization (WHO) report, 5% of the world population in 2018, 466 million people, have disabling hearing loss, and this is on the rise. The importance of sign language recognition has increased because of the high growth rate of the deaf and hard-of-hearing population globally and the extended use of vision-based application devices. In recent years, many researchers have proposed vision-based sign language recognition by utilizing inputs of the camera, such as 3d camera, web camera and stereo camera. The main cause of the vision-based approach's attractiveness is: it does not need any specialized device, and this method is affordable. At the same time, other domain needs different specialized devices such as power gloves, accelerometer, kinetic sensor, leap motion controller, and huge wire. However, SLR become difficult due to the complex background, uncontrolled environment, visual analysis of gesture, lighting illumination, finger occlusion, inter-class variation, constant fatigue, the similarity of the high intraclass, and complexities of different signs. Many researchers have utilized the segmentation approach considering the above problems of vision-based systems. To solve the problem, the significant contribution of the paper is as follows:

Firstly, we have developed a Rotated, Translated, and Scaled (RTS) version of the dataset with some specification to solve the RTS issue in the test or future images.

Secondly, we have designed a hybrid segmentation approach to reduce the redundant background of the images by combining Otsu Thresholding with YCbCr, Morphological analysis, and Watershed algorithm.

Finally, we propose a novel convolutional neural network (CNN) architecture model for extracting features.

Original Dataset

To evaluate the proposed model, two datasets was used and sample image of that two dataset shown in Figure 1 and 2.

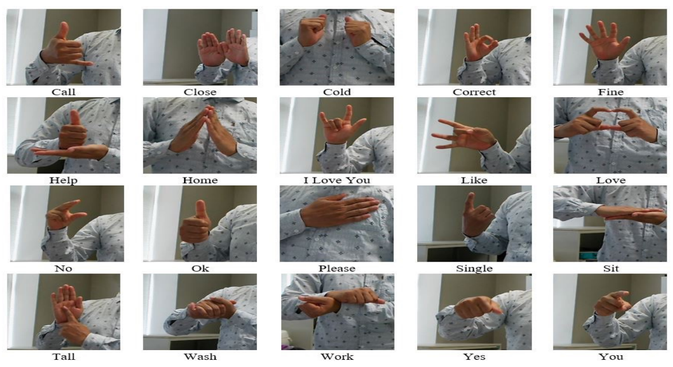

Figure 1. Twenty Sign Word Dataset.

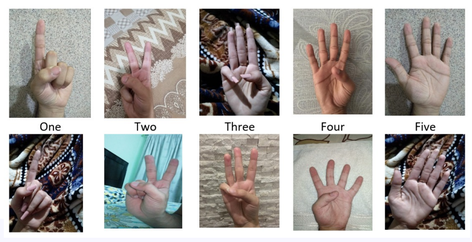

Figure 2 Five Sign Word Dataset.

RTS Dataset

In the Rotation technique, we randomly applied -30 to +30 degrees to rotate the image. Fig. 3. shows the Rotated images for -30 degree, 15 degrees and 30-degree angles. In the Translation technique, we used x-axis and y-axis translation randomly. We have selected a range between -40 to +40 pixels for translation which is 20% of the original images. In the Scaling technique, we have modified both directions of the object in the image to reduce and magnify the sign image. We have randomly selected scaling factor ranges between 0.2 to 1.5.

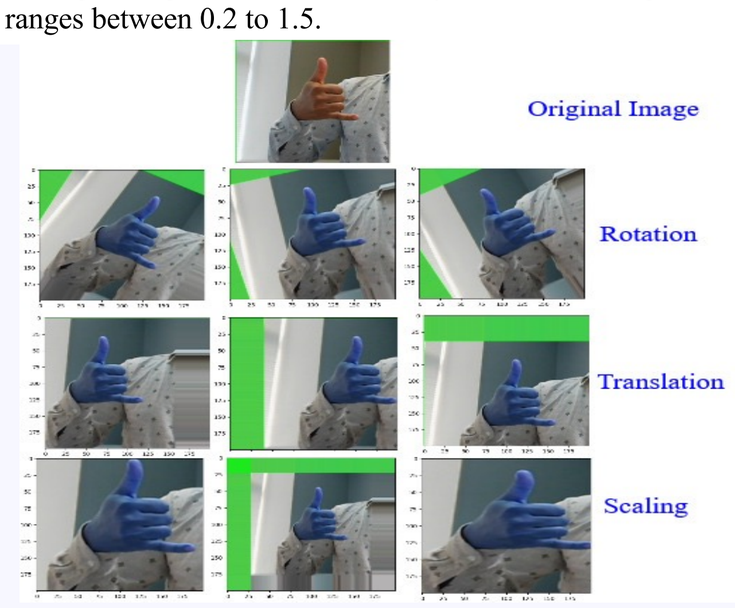

Figure 3. RTS process of an input image.

Sign Word Recognition Method

The conceptual structure of the proposed Sign word recognition system is demonstrated in Fig. 4. Following steps are followed for both training and test dataset. First, this structure divided the dataset into training and test sets. Second, the training dataset's Rotated, Translated, and Scaled (RTS) version is created. Third, a hybrid segmentation process is applied, which is consists of Otsu thresholding with YCbCr skin color segmentation, morphology analysis, and the watershed algorithm. Finally, a novel CNN architecture is developed to extract the feature map and classification.

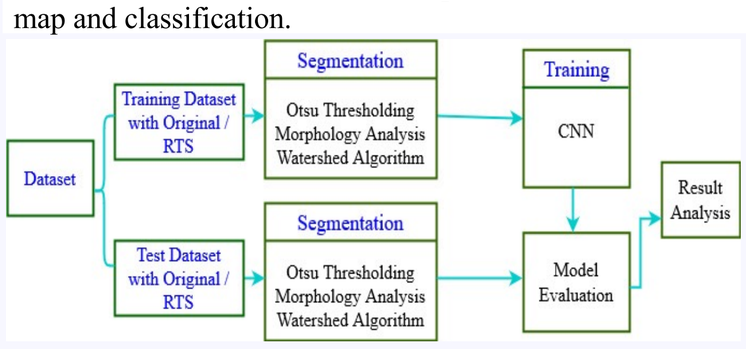

Figure 4. Proposed System

Segmentation Technique

The proposed segmentation process combined with three steps to overcome the problem: i) Combining Otsu Thresholding with YCbCr, ii) Morphological analysis: dilation through opening morphology and iii) Watershed algorithm. Fig. 5. presents the process of segmentation.

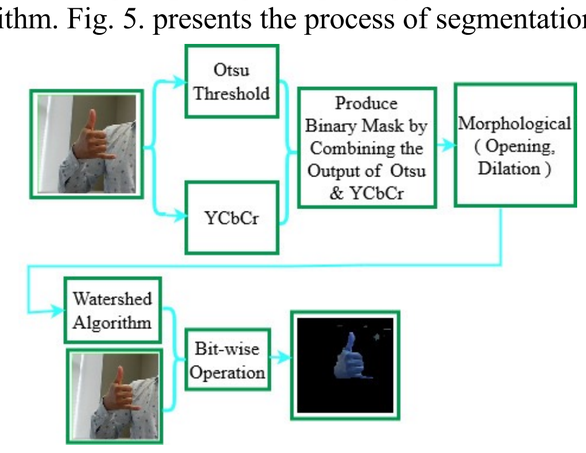

Figure 5. Segmentation process of an input image.

Convolutional Neural Network

This paper proposed a novel CNN-based classifier model for training and testing on the entire dataset. CNN is a widely used feature extraction and classification algorithm in the Sign language recognition domain.

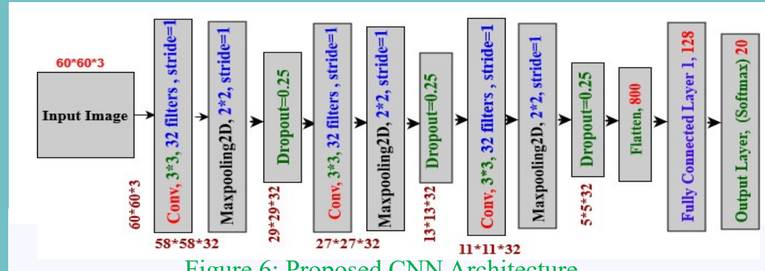

Figure 6: Proposed CNN Architecture.

Experimental Result

Training Dataset	Test Dataset	Precision	Recall	F1 score	Accuracy (%)
Twenty sign word	30% of Twenty sign word	98.00	98.00	99.00	98.80
Twenty sign word	30% RTS version of Twenty sign word	71.00	68.00	69.00	**68.00**
Five Sign Word	Original 15 test images	94.00	93.00	93.00	93.33
Five Sign Word	RTS version images were generated from 15 test images.	71.00	70.00	70.00	70.00

Table 1: performance (%) with the Original dataset

Training Dataset	Test Dataset	Precision (%)	Recall	F1 score	Accuracy (%)
RTS version of Twenty sign word	Twenty Sign Word Images	99.00	99.00	99.00	99.30
RTS version of Twenty sign word	RTS version of Twenty Sign Word	99.00	99.00	99.00	**99.10**
RTS version of Five sign word	Original 15 test images	99.00	99.00	99.00	**100**
RTS version of Five sign word	RTS version of 15 test images.	98.00	98.00	98.00	98.00

Table 2: performance (%) with RTS version of the 20 Sign

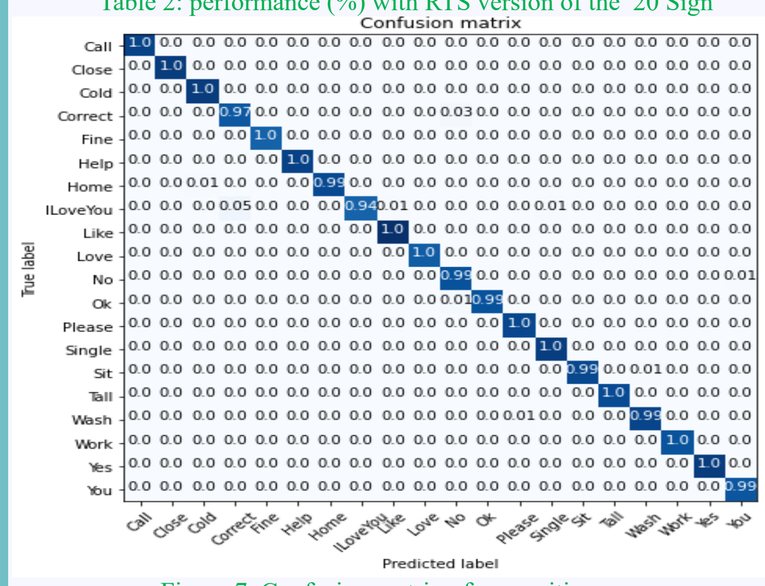

Figure 7. Confusion matrix of recognition accuracy

Method	Dataset	Number of Images	Segmentation	Test Set	Accuracy (%)
Rahim et al. [4],2019	20-Sign dataset	18000	yes	30% of theDataset	97.28
Mujahid et al. [33], 2021	5-Sign dataset	216	no	15 Test Images	97.68
Proposed (CNN Train and Test with RTS dataset)	RTS version dataset of [4]	190,000	yes	Same as [4]	99.30
	RTS version dataset of [33]	2160	yes	Same as [33]	100

Table 3: performance (%) with RTS version of the 20 Sign

Segmentation Technique

The proposed segmentation process combined with three steps to overcome the problem: i) Combining Otsu Thresholding with YCbCr, ii) Morphological analysis: dilation through opening morphology and iii) Watershed algorithm. Fig. 5. presents the process of segmentation.

Conclusions & Future Works

- A RTS with concatenated segmentation along with the feature fusion-based sign word recognition system was presented in this paper.
- We specify the range of rotation, translation and scaling factor for not changing semantic meaning with experiment
- To detect the hand gestures, we preprocessed input images using Otsu thresholding and YCbCr segmentation.
- Therefore, we used the proposed CNN model to extract features from segmented images,
- The results indicated that in the real-time environment, approximately 99% accuracy was achieved using trained features and the CNN classifier, and it led to better results than the state-of-the-art systems
- In future works may be work with more number of sign word that usually used by the deaf-mute community.

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/