Proceedings of 2017 IEMEK Symposium on Embedded Technology
(ISET 2017)
Hotel Grand Haeundae, Busan, Korea, 18-*19 May*, *2017*.

CNN-based Korean Sign Language Recognition with IMU and EMG Sensors

Seongjoo Shin* Youngmi Baek* Sang Hyuk Son*

*Department of Information & Communication Engineering
DGIST, Daegu, Republic of Korea
E-mail: {sj_shin, ymbaek, son}@dgist.ac.kr

Abstract

Many hearing or disabled people are hard to communicate with the public. An automatic sign language recognition system enables them to comfortably communicate with people. Since it requires sophisticated and effective recognition and classification of various gestures defined in Korean sign language, we use a sensor fusion technology using IMU and EMG sensors to improve the accuracy of recognition and also employ CNN as a classification algorithm. Our algorithm includes how to identify between gestures and how to handle different length of gestures.

Keywords: Sign language, EMG, IMU, CNN, Sensor fusion.

1. Introduction

There are about 2.5 million people with disabilities in Korea [1]. 10% of them are hearing–disabled or speaking-disabled persons. Since most of the people don't know how to use sign language, a person with disability has difficulty in communicating whenever he asks for their helping. An automatic sign language recognition (SLR) system is useful for communication between people with and without disabilities. The system aims to recognize a given gesture as an input, and express its intention or meaning in natural language through auditory or visual tools. We have classified 30 of gestures defined in Korean sign language, which are distinguished by each movement and length. First of all, it is very important to determine the timing to end the given gesture and the timing to make the next gesture. To solve this problem, a basic posture as the status of no moving is defined as moving our arms and hands during the period from the end of the current gesture to the beginning of the next gesture. To measure the given movement of both hands and arms, we use inertial measurement unit (IMU). Electromyography (EMG) sensors are used to guess the shape of the hand. Finally, each gestures is classified using CNN (Convolutional Neural Network).

2. CNN-based Korean Sign Recognition

People have to move their arm to prepare next action after one movement. Movement between gestures can make gestures recognized as other gestures. So we define the default basic posture with no movement such as Figure 1. We also used the IMU's accelerometer to recognize hand and arm movements. We averages the last 10 sensor values on each axis of the accelerometer as in (1). If $ACC_{avg}(\mathrm{t})$ of one axis exceeds the threshold, it is the starting point. If $ACC_{avg}(\mathrm{t})$ of all axes is within the threshold for 10 times, it is the end point.

$$ACC_{avg}(t) = \frac{1}{10}\sum_{i=t-9}^{t} ACC(i) \quad (1)$$

People have a lot of biometric information. EMG is one of them. The EMG sensor measures the intensity of the EMG signal. The sensor outputs positive and negative values. The absolute value of this value means the magnitude of the force. So, we use the absolute value of EMG signal. Sensor fusion requires some pre-processing. First, each sensor has a different reading cycle. Therefore, it is necessary to equalize the period of each sensor. We downsample the period of the sensor to equal. Second, the range of values output by each sensor is different. We convert the values of each sensor into a certain range value [2].

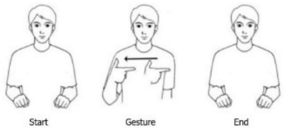

Figure 1. A basic posture

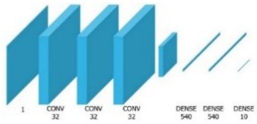

Figure 2. Proposed model of CNN

We used CNN to classify sign language. The architecture of a CNN is designed to take advantage of the 2D structure of an input image. We expressed the pre-processed sensor values as images. A dimension of the input data is the product of the length of the longest sample (180) and the number of sensor axes. For sample of short length, the remaining data is filled with zeros. These input data are used in the CNN model shown in figure 2. The model has three convolutional layers, one max-pooling and three dense layers. Generally, models with dropout have higher performance than models without dropout [3]. But this model also does not include dropout.

We used MYO device to measure movement. The arm band called MYO device consists of eight medical grade stainless steel EMG sensors and highly sensitive nine-axis IMU. The number of training data is 100 data for each class, total 3000 data. The number of test data is 20 data for each class, total 600 data. The experimental results using the CNN model are 99.6%. The main reason for the high accuracy is the data of one person. The value of the EMG signal varies from person to person. Each person has different muscle mass and subcutaneous fat layer thickness. This causes the accuracy to be lowered. However, only one male participated in our experiment.

3. Conclusions

This paper has exploited the deep learning and the sensor fusion technique in order to more accurately classify Korean sign language. We paid attention to two problems in sign language recognition. One is how to distinguish by difference in gestures. The other is that lengths of the gestures are different. The proposed model has classified 30 different sign language and got 99.6% accuracy. In the future, we plan to utilize many different people data to improve the performance of our model and develop an automatic recognition system. This system will notify us after recognizing sign language.

ACKNOWLEDGMENT

This research was supported in part by Global Research Laboratory Program (2013K1A1A2A02078326) through NRF, DGIST Research and Development Program (CPS Global Center) funded by the MSIP, and Institute for Information & communications Technology Promotion (IITP) grant funded by the Korean government (MSIP) (No. B0101-15-0557, Resilient Cyber-Physical Systems Research).

References

[1] http://kosis.kr/statHtml/statHtml.do?orgId=117&tblId=DT_11761_N003&conn_path=I2
[2] K. K. Pal and K. Sudeep, "Preprocessing for image classification by convolutional neural networks," in Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on, IEEE, 2016, pp. 1778--1781.
[3] H. Wu and X. Gu, "Towards dropout training for convolutional neural networks," Neural Networks, vol. 71, pp. 1--10, 2015.