International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 02 | Feb 2019 www.irjet.net p-ISSN: 2395-0072

American Sign Language Fingerspelling Recognition Using Convolutional Neural Networks from Depth map

Aleema Jose[1], Prameeja Prasidhan[2]

[1]Dept. Of Computer Science, St. Joseph's college (Autonomous) irinjalakuda, Kerala, India
[2]Lecturer, Dept. Of computer Science, St. Joseph's college (Autonomous) irinjalakuda, Kerala, India

---

Abstract - *American Sign Language (ASL) recognition is important for natural and useful communication between deaf and hearing majority community. To take a highly efficient initial step of the automatic fingerspelling recognition system using convolutional neural networks (CNNs) from depth map. We train CNNs for the classification of the 29 alphabets and space, delete, nothing using a subset of collecting depth data from multiple subject. We accomplish 99.9% perfection for recognized signers and 83.58% to 85.49% perfection for new signers. The processing time is 3 ms for the forecasting of a single image. The system accomplishes the highest perfection and speed. The trained model and dataset is available on our storage area.*

Key Words ASL, CNNs, depth map.

1. INTRODUCTION

American Sign Language recognition is very important for natural and useful communication between deaf and hearing majority community. Presently most of communication between these two communities highly depends on human based translation service. However this is inappropriate and expensive as human expertise is involved. Automatic sign language recognition focus on to understand the meaning of signs without the cooperation from experts. Then the sign can be translated into text based on the end user's needs. The sign language recognition is most important for the communication for deaf and hearing majority and it can providing equal chance to every person.

The data set contains thousands of hand pose and also, sign language has thousands of words that contains similar types of hand pose. The sign language recognition is can don by using hand fingers so the gestures recognition includes a small set of well specified gestures. Even the same sign have seriously different aspects for different signers and different point of view.

In this paper, we focus on the stable fingerspelling in American sign language it is small but it is relevant for the sign language recognition. This is small set of sign languages as shown in figure 1, is used to carrying names, address and so on. A large variations occur at the different signers.

Depth sensor enables to capturing addition information's and it provides and improves the accuracy and the processing time. Depth sensor and CNNs attain a real time and exact sign language recognition system.

2. RELATED WORKS

American Sign Language recognition is only considered as well specified hand gestures, some approaches are related to the sign language recognition. Suggest a human recognition system for human robot interaction using CNNs. Van den Bergh et al.proposed a hand gesture recognition system using Haar Wavelets and database searching. The system extracts features using Harr Wavelets and classifies input image finding the convenient match in the database. Different sign languages are used in different countries .The system is different from recognizing a single fingerspelling. The data set is corresponded to a single user.

ASL system which recognize a sentence of three to five words. To the system will recognize the 26 different alphabets. The system will recognize the ASL sign by applying CNNs.

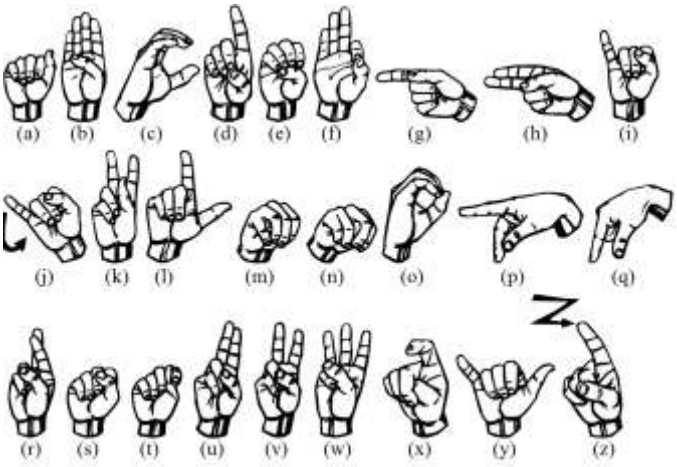

Fig-1: ASL fingerspelling alphabet [1].

3. METHOD

3.1. Data set

The data set consist of 1000 images and each of which 29 hand signs. We have only one class to represent. To collect dataset from different viewpoints the hand gesture is accessed while movie hand in the image plane it will check the represented sign to the dataset.

3.2. Hand segmentation

We assume the closest object from camera is the user hand.

Fig-2: captured image before preprocessing.[2].

The figure 2 show an example for the captured depth image. This is the real time and convenient to the hand segmentation. Figure 3 shows the segmented hand depth image sample s for 26 alphabets. We find a bounding box of hand region and scale it to 227 × 227[5].

The hand gesture only take only the portion of depth void around the hand it avoid the background. The 26 alphabet are captured by using the depth map the picture is like a black colored or gray color. A single sign have many data because it have some variations have occurs from the different signers.

Fig- 3: example of pre-processed dataset from A to Z [2]

The different signers have different consideration and different viewpoint. Figure 4 represent the different signers matching dataset.

Figure 4. The collected data from same meaning.[2]

4. CLASSIFICATION

a) Architecture: We use Caffe [3] implementation (CaffeNet)of the CNNs which is almost equivalent to Alex Net. The architecture consists of five convolution layers, fivemax-pooling layers, and three fully connected layers. Aftereach convolution layer or fully connected layer except thelast one, rectified linear unit layer is followed. For details,we will upload the architecture and also readers can refer toCaffeNet/Caffe and AlexNet .

b) Feature extraction: We extract a 4096-dimensionalfeature (final fully connected layer feature) vector from each pre-processed depth image using the aforementioned architecture. First, we subtract the mean image from each of the sample training/validation/test image. The featured extraction remove the unwanted background images .Then the mean subtracted image is forward-propagated to extract features.

c) Training: We train and test neural networks in five different operating modes.. One way to look at it is from the pre-training perspective and the second way is how we deal with the training/testing data separation for different subjects. In the former case, we categorize the operating modes into two categories, namely re-training and finetuning.For re-training, the model is re-trained from randomly generated weights using the collected fingerspelling data. In fine-tuning, we pre-train the CNNs using a largeILSVRC2012 classification dataset ; then we fine-tune the network weights for fingerspelling classification with the same architecture except the last layer which is replaced by 31 output classes. From the subjects' data separation perspective, in one case, we do not separate the subjects in training, validation, and testing and in the second scenario, we use data from different subjects for training, validation, and testing. The traing can be done by the captured images are used.

5. EXPERIMENTAL RESULTS

Our system achieves 99.9% accuracy when training, Fingerspelling recognition system will considered only six gestures or 24 signs respectively. The future enhancement we can add the face recognition to access the facial expression also. So it can easily identify the meaning of the expression.

6. CONCLUSION

The efficiency of using the CNNs and the depth sensor for ASL fingerspelling recognition system. Using the depth image and the CNNs achieves the real time performance and access the accurate image. This is very simple and

helpful for the normal communication for thr deaf and hearing majority community.

REFERENCES

1) A.S.L university. Fingerspelling.. http://www.lifeprint.com/asl101/fingerspelling

2) Real time sign language fingerspelling recognition using convolutional neural network from depth map. https://ieeexplore.ieee.org/document/7486481.

3) Y.jia caffe: an open source convolutional architecture for fat feature embedding, 2013.