WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

SURVEY ON STATIC SIGN LANGUAGE RECOGNITION USING DEEP LEARNING

S. Chitra[1], R. Kokila[1], J. Suvithra[1], K. Krishnaveni[1], Nandhini A[2]

[1]IV year students, department of computer science and engineering shreenivasa engineering college, bommidi
[2]Asst. Prof. Department of computer science and engineering shreenivasa engineering college, bommidi

ABSTRACT

A system was advanced that will obey as a learning tool for starters in sign language that required hand detection. This system is based on a skin-color styling methodology, i.e., specific skin-color space edging the skin-color range is predetermined that will extract pixels (hand) from non–pixels (background). The images were cater into the model called the Convolutional Neural Network (CNN) for classification of images. keras was used for training of images. Provided with correct lighting condition and a uniform background, the system attained an average testing accuracy of 93.67%, of which 90.04% was featured to ASL alphabet recognition, 93.44% for number recognition and 97.52% for unchanged word recognition, thus surpassing that of other related studies. The route is used for fast calculus and is done in real time.

1. INTRODUCTION

Sign language recognition is a problem that has been addressed in investigation for years. However, we are still far from finding a complete solution available in our society. Among the works grow to address this problem, the majority of them have been based on basically two approaches: contact-based systems, such as sensor gloves; or vision-based systems, using only cameras. The dispatch is way budgeter and the boom of deep learning makes it more appealing.

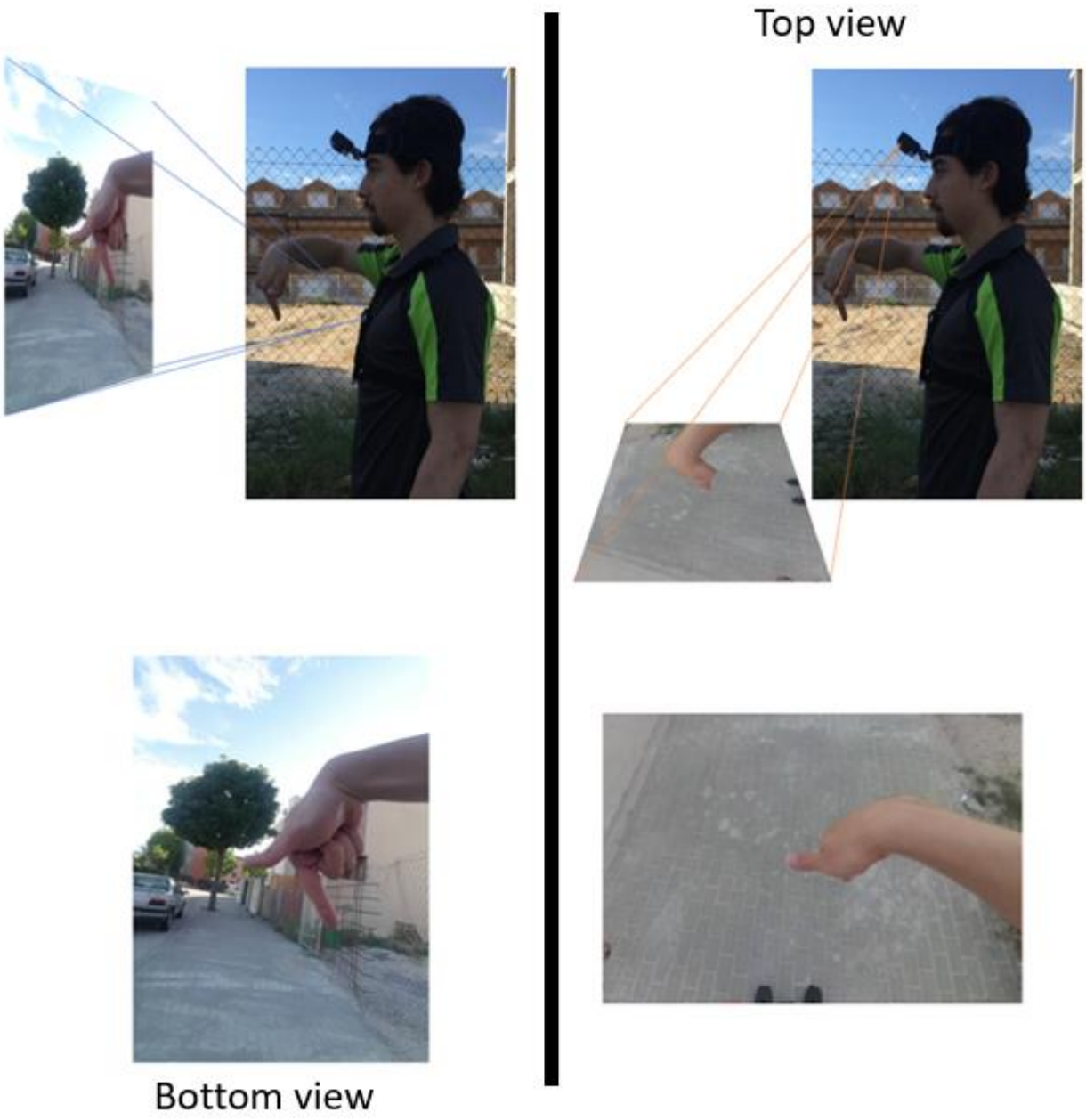
This post presents a prototype of a dual-cam first-person perception translation system for sign language using convolutional neural networks. The post is divided into three main parts: the system design, the dataset, and the deep learning model training and assessment.

2. VIEW SYSTEM

View is a key factor in sign language, and every sign language is intended to be understood by one person find in front of the other, from this perspective, a gesture can be completely observable. Viewing a gesture from another perspective makes it rural or almost hopeless to be recognize since every finger position and action will not be visible.

Trying to understand sign language from a first-view perspective has the same restriction, some gestures will end up looking the same way. But, this vagueness can be solved by locating more cameras in different positions. In this way, what a camera can't see, can be perfectly obvious by another camera.

The view system is collected of two cameras: a head-mounted camera and a chest-mounted camera. With these two cameras we obtain two different views of a sign, a top-view, and a bottom-view, that works collectively to identify signs.

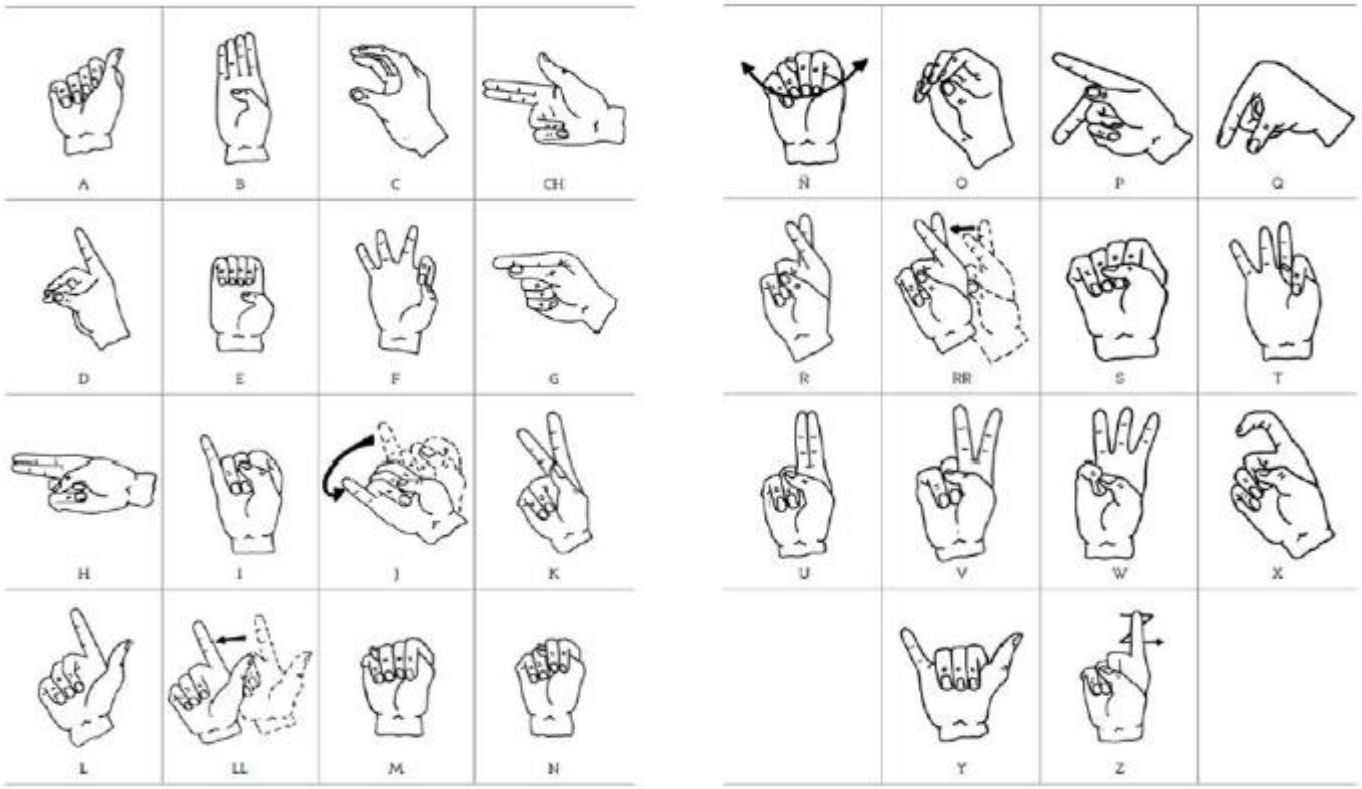


Sign comparable to letter Q in the Panamanian Sign Language from a top view and a bottom view standpoint (image by author)

Another benefit of this design is that the user will gain freedom. Something that is not accomplish in classical approaches, in which the user is not the person with damage but a third person that needs to take out a system with a camera and point a signer while the signer is performing a sign.

3. DATABASE

To conceive the first instance of this system is was used a dataset of 24 static signs from the Panamanian Manual Alphabet.



Panamanian Manual Alphabet

To model this problem as an image recognition problem, changed final such as letter J, Z, RR, and Ñ were keep because of the extra complexity they add to the solution.

4. DATA COLLECTION AND PREPROCESSING

To collect the dataset it was asked to four users to wear the view system and perform every signing for 10 seconds while both cameras were recording in a 640x480 pixel decision.

It was exact to the users to perform this process in three different scheme: indoors, outdoors, and in a green background scheme. For the indoors and outdoors scenarios the users were requested to move surrounding while performing the gestures in order to obtain images with different backgrounds, light sources, and positions. The green background scenario was voluntary for a data accrual process, we'll describe later.

After obtaining the videos, the architecture were extracted and reduced to a 125x125 pixel resolution.

From left to right: green background scenario, indoors and outdoors (image by author)

5. DATA AUGUMENTATION

Since the preprocessing before going to the complexity neural networks was clear to just rescaling, the background will always get passed to the model

To improve the generalization capacity of the model it was fake added more images with distinguishable backgrounds cut outing the green backgrounds. This way it is acquired more data without investing too much time.

Images with new backgrounds (image by author)

6. DATASET

This problem was classical as a multiclass classification problem with 24 classes, and the problem itself was divided into two subordinate multi-class classification troubles.

The approach to obvious which signing would be classified whit the top view model and which ones with the bottom view model was to select all the gestures that were too similar from the bottom view perspective as signing to be classified from the top view copy and the rest of gestures were going to be classified by the bottom view model. So basically, the top view model was used to crack obscure.

As a result, the dataset was divided into two parts, one for each style as displayed in the support table.

Model	Top View	Bottom View
Classes	B, C, D, E, F, G, H, I, L, O, P, Q, R, T, U, W, X	A, K, M, N, S, V, Y
Total	17	7

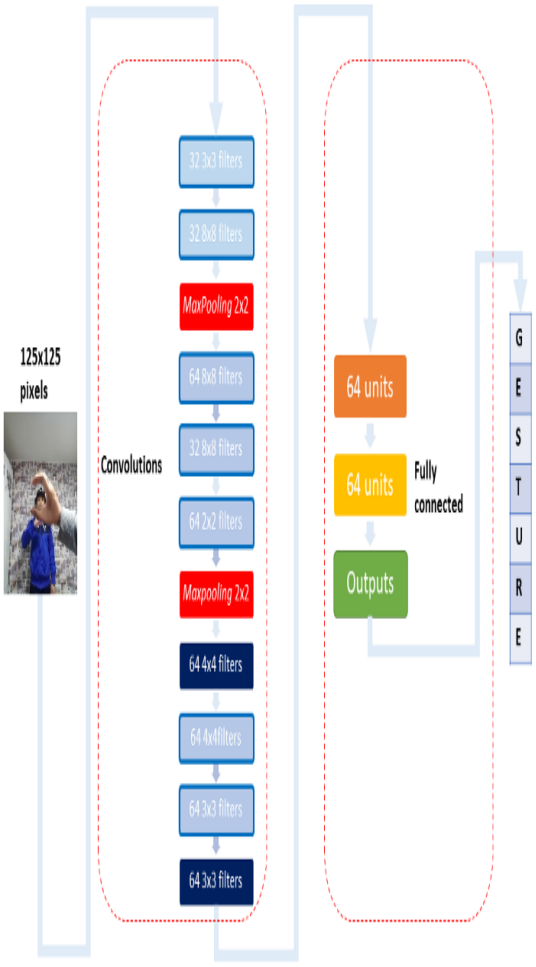
7. RETICULAR FORMATION MODEL

As state-of-the-art automation, convolutional neural networks was the alternative chosen for facing this problem. It was educate two models: one model for the top view and one for the bottom view.

8. AESTHETICS

The same convolutional neural network architecture was used for both, the top view and the bottom view models, the only sameness is the number of output units.

The construction of the convolutional neural web is shown in the following figure.



Convolutional neural network architecture

To improve the generalization capability of the models it was used dropout techniques between surface in the fully connected layer to improve model performance.

9. APPRAISAL

The models were evaluated in a test set with data communicate to a normal use of the system in indoors, in other words, in the background it appears a person acting as the observer, alike to the input image in the figure above (*Convolutional neural networks architecture*). The results are shown below.

Whereas, the model learned to categorize some signs, such as Q, R, H; in general, the results are kind of dispirit. It seems that the conception capability of the models wasn't too good. However, the style was also tested with real-time data expose the potential of the system.

The bottom view style was explore with real-time video with a green uniform background. I wore the chest-mounted camera capturing video at 5 frames per second while I was running the base view model in my laptop and try to fingerspell the word fútbol (my favorite sports in Spanish). The entries for every . letter were imitate by a click.

Bottom view model

	Precision	Recall
A	0,79	1,00
K	0,48	0,73
M	0,54	1,00
N	0,28	0,53
S	0,00	0,00
V	1,00	0,27
Y	0,00	0,00

Top view model

	Precision	Recall
B	0,69	0,73
C	0,39	1,00
D	0,50	0,73
E	0,52	1,00
F	1,00	0,65
G	0,65	1,00
H	1,00	0,67
I	0,67	0,65
L	0,65	1,00
O	0,67	0,93
P	0,00	0,00
Q	1,00	1,00
R	1,00	0,80
T	1,00	1,00
U	1,00	0,60
W	0,88	1,00
X	0,00	0,00

10. CONCLUSIONS

Sign language recognition is a hard problem if we opinion all the possible combinations of gestures that a system of this kind needs to comprehend and interpret. That being said, probably the best way to solve this problem is to disjoint it into simpler problems, and the system presented here would cognate to a possible solution to one of them.

The system didn't achieve too well but it was demonstrated that it can be built a priority-person sign language translation system using only cameras and convolutional neural networks.

It was perceived that the model tends to baffled several signs with each other, such as U and W. But imaging a bit about it, maybe it doesn't need to have a perfect performance since using an orthography corrector or a word predictor would boost the translation accuracy.

The next step is to inspect the solution and study ways to enhance the system. Some improvements could be carrying by collecting more better data, trying more convolutional neural network architectures, or redesigning the view system.

11. COMPLETE WORDS

I developed this project as part of my thesis work in university and I was encouraged by the feeling of working in something new. Although the results weren't too amazing, I think it can be a good starting point to make a better and superlative system.

REFERENCE

[1] S. Shahriar, A. Siddiquee, T. Islam, A. Ghosh, R. Chakraborty, A. I. Khan, C. Shahnaz, and S. A. Fattah, "Real-time american sign language recognition using skin segmentation and image category classification with convolutional neural network and deep learning," in Proc. TENCON 2018-2018 IEEE Region 10 Conference, 2018, pp. 1168-1171.

[2] JR. Daroya, D. Peralta, and P. Naval, "Alphabet sign language image classification using deep learning," in Proc. TENCON 2018-2018 IEEE Region 10 Conference, 2018, pp. 646-650.

[3] T. Kim, G. Shakhnarovich, and K. Livescu, "Finger-spelling recognition with semi-markov conditional random fields," in Proc. 2013 IEEE International Conference on Computer Vision, 2013, pp. 1521–1528.

[4] N. Pugeault and R. Bowden, "Spelling it out: Real-time asl finger-spelling recognition," in Proc. 2011 IEEE International Conference on Computer Vision Workshop, 2011, pp. 1114–1119.