International Journal of Research Publication and Reviews

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

A Survey on Sign Languages and Semaphoric Hand Gestures

Nezrin Naushad[a], *Anu Bonia Francis*[b]

[a]*Student, Department of Computer Science and Engineering (Mtech),RIT Engineering College Kottayam, kerala , India*
[b]*Professor , Dept of Computer Science and Engineering , RIT Engineering college , Kerala, India*

ABSTRACT

Sign Language is a language which allows mute people to communicate. The ease of communication offered by this language, however, disappears when one of the interlocutors, who may or not be mute, does not know Sign Language and a conversation starts using that language. There is an undeniable communication problem between the Deaf community and the hearing majority. Innovations in automatic sign language recognition try to tear down this communication barrier. Sign Language is a language which allows mute people to communicate. Sign Language (SL) is not commonly learned by non-mute people; thus, mute people have problems in communicating. The focus of this project is on this type of problem facilitating communication via SL by automating the transcription of SL without the need for a human translator. Here discusses a system that takes advantage of Convolutional Neural Networks. Hand gesture recognition is still a topic of great interest for the computer vision community. The process of hand gesture recognition is divided into hand location, hand segmentation and its classification. My method locates and extracts the hand to generate new images, uses them in a Convolutional Neural Network (CNN) to classify and then recognize hand gestures. Semaphoric are specific hand gestures that define a set of commands and/or symbols to interact with machines.

Keywords: Sign language, Convolutional Neural Network, Sign Language, Semaphoric Hand Gestures.

1. Main text

The goal of this project was to build a convolutional neural network able to classify which letter of the American Sign Language (ASL) alphabet is being signed, given an image of a signing hand. This project is a first step towards building a possible sign language translator, which can take communications in sign language and translate them into written and oral language. Such a translator would greatly lower the barrier for many deaf and mute individuals to be able to better communicate with others in day-to-day interactions. This goal is further motivated by the isolation that is felt within the deaf community. Loneliness and depression exist in higher rates among the deaf population, especially when they are immersed in a hearing world. Large barriers that profoundly affect life quality stem from the communication disconnect between the deaf and the hearing. Some examples are information deprivation, limitation of social connections, and difficulty integrating in society. Most research implementation for this task have used depth maps generated by depth camera and high-resolution images. The objective of this project was to see if neural networks are able to classify signed ASL letters using simple images of hands taken with a personal device such as a laptop webcam. This is in alignment with the motivation as this would make a future implementation of a real time ASL-to-oral/written language translator practical in an everyday situation. In recent years, the use of different types of controls besides mouse and keyboard has become common. A number of devices are available for specific tasks or applications, one interaction form that has gained popularity is the area of natural interactions between humans and machines. Web browsers, video games, Virtual Reality (VR) environments and a diverse set of tools have taken advantage of users' natural interactions, e.g., voice control, touchpads, haptic devices, cameras, etc. Immersing the user into the systems or environments in a natural way is the goal of this type of research. Natural interaction involves the use of a user's body without additional hardware. This is called Natural User Interface (NUI). By using sensors and capturing the diverse interactions of the user's body it is possible to recognize commands and perform required tasks in a system.Sign Language (SL) is not commonly learned by non-mute people; thus, mute people have problems communicating. Usually, people do not learn it if there is no mute person in their relation circles or if it is not required for their job. When they engage with a mute person

* *Corresponding author..*
E-mail address: nezrinnaushad9@gmail.com

the communication can be hard and tedious. The goal of this project is to use the Convolutional Neural Network (CNN) recognizing and transcribing into text SL images (gestures performed by a hand). Given the broad scope of this task I limited the scope of the study to American Sign Language (ASL) letters and number symbols.

1.1. Related Works

Sign Language Recognition involves a variety of Techniques from diverse areas. The first step of Sign Language Recognition is to capture the symbols Performed by the user. Methods for recognizing bodies and body parts have been widely studied and a diverse Set of applications created. Neural Networks (NN) are Often used as recognition models for image processing. Mubashira Zaman, Soweba Rahman, Tooba Rafique, Filza Ali, and Muhammad Usman Akram proposed" Hand gesture recognition using color markers," [1] in Which present markers on the hand other techniques to capture and track finger information for the Recognition of hand gestures. In this paper, a hand Gesture recognition method based on color marker Detection is presented, in this case, have used four Types of colored markers (red, blue, yellow and green) mounted on the two hands. With this posture the user can perform different gestures as zoom, move, draw, and write on a virtual keyboard. This implemented system provides more flexible, natural and intuitive interaction possibilities, and also offers an economic and practical way of interaction. of this is that Markers are not uncomfortable but they need a specific setup step and a constructed environment. This approach may not be suitable for the problem. Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao proposed "Recurrent convolutional neural networks for text classification," [2] in which this paper discuss about neural networks in the area of speech recognition. One of the drawbacks is Due to the spatiotemporal feature of the RCNN they are suitable for these kinds of tasks. Text classification is a foundational task in many NLP applications. Traditional text classifiers often rely on many human- designed features, such as dictionaries, knowledgebases and special tree kernels. In contrast to traditional methods, here introduce a recurrent convolutional neural network for text classification without human-designed features. In this model, here apply a recurrent structure to capture contextual information as far as possible when learning word representations, which may introduce considerably less noise compared to traditional window based neural networks. here also employ a max-pooling layer that automatically judges which words play key roles in text classification to capture the key components in texts. Here conduct experiments on four commonly used datasets. The experimental results show that the proposed method outperforms the state-of-the-art methods on several datasets, particularly on document-level datasets. Michael Egmont-Petersen, Dick de Ridder, and Heinz Handel's proposed "Image processing with neural network review" [3] in which this paper presents an approach that Networks (NN) are often used as recognition models for image processing. Neural network usually require much more data So better to use other algorithms that will deal with less data these are some limitations. Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang proposed" Robust part-based hand gesture recognition using kinect sensor" [4], Regarding hand gesture recognition and Sign Language problems. The Kinect has been used in diverse methods, concluding in high accuracy and real time systems. The recently developed depth sensors, e.g., the Kinect sensor, have provided new opportunities for human-computer interaction (HCI). Although great progress has been made by leveraging the Kinect sensor, e.g., in human body tracking, face recognition and human action recognition, robust hand gesture recognition remains an open problem.

Compared to the entire human body, the hand is a smaller object with more complex articulations and more easily affected by segmentation errors. It is thus a very challenging problem to recognize hand gestures. This paper focuses on building a robust part-based hand gesture recognition system using Kinect sensor. To handle the noisy hand shapes obtained from the Kinect sensor, here propose a novel distance metric, Finger-Earth Mover's Distance (FEMD), to measure the dissimilarity between hand shapes. As it only matches the finger parts while not the whole hand, it can better distinguish the hand gestures of slight differences. The extensive experiments demonstrate that our hand gesture recognition system is accurate (a 93.2 percentage mean accuracy on a challenging 10-gesture dataset), efficient (average 0.0750 s per frame), robust to hand articulations, distortions and orientation or scale changes, and can work in uncontrolled environments (cluttered backgrounds and lighting conditions). The superiority of the system is further demonstrated in two real-life HCI applications. LeJake Araullo, and Lewis Carterigh Ellen Potter proposed "The leap motion controller: a view on sign language" [5] This paper compared to other sensors it is small and portable. Some of the major limitations are: the area of capture restricts the user's movement and it may not be appropriate for sign language. Hand tracking gloves are other techniques to capture and track finger information for the recognition of hand gestures. Gloves are accurate in determining the exact position of each finger. But they are not comfortable for users. They can obstruct fingers' movements and the resulting hand gesture may not be correct. Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Gerald Penn proposed" Applying Convolutional Neural Networks concepts to hybrid NN-HMM model for speech recognition" [6]. In this paper, describeConvolutional Neural Networks (CNN) have showed success in achieving translation in variance for many image processing tasks. The success is largely attributed to the use of local filtering and max-pooling in the CNN architecture. In this paper, propose to apply CNN to speech recognition within the framework of hybrid NN-HMM model. Here propose to use local filtering and max-pooling in frequency domain to normalize speaker variance to achieve higher multi-speaker speech recognition performance. In this method, a pair of local filtering layer and max-pooling layer is added at the lowest end of neural network (NN) to normalize spectral variations of speech signals.

In experiments, the proposed CNN architecture is evaluated in a speaker independent speech recognition task using the standard TIMIT data sets. Experimental results show that the proposed CNN method can achieve over 10 percentage relative error reduction in the core TIMIT test sets when comparing with a regular NN using the same number of hidden layers and weights. The results also show that the best result of the proposed CNN model is better than previously published results on the same TIMIT test sets that use a pre-trained deep NN model. Adam Baumberg proposed" Reliable Feature Matching Across Widely Separated Views" [7] In this paper present a robust method for automatically matching features in images corresponding to the same physical point on an object seen from two arbitrary viewpoints. Unlike conventional stereo matching approaches, we assume no prior knowledge about the relative camera positions and orientations. In fact, in application this is the information wish to determine from the image feature matches. Features are detected in two or more images and characterized using affine texture invariants. The problem of window effects is explicitly addressed by the method -the feature characterization is invariant to linear transformations of the image data including rotation, stretch and skew. The feature matching process is optimized for a structure-from- motion application where wish to ignore unreliable matches at the expense of reducing the number of feature matches. Jinting Wu, Kang Li, Xiaoguang Zhao, and Min Tan, proposed" Unfamiliar Dynamic Hand Gestures Recognition Based on Zero-Shot Learning" [8], Most existing robots can recognize trained hand gestures to interpret user's intent, while untrained dynamic hand gestures are hard to be understood correctly.

1.2. Data Collection

The primary source of data for this project is the compiled dataset of American Sign Language (ASL) called the ASL Alphabet from Kaggle user Akash. The dataset is comprised of 87,000 images which are 200x200 pixels. There are 29 total classes, each with 3000 images, 26 for the letters A-Z and 3 for space, delete and nothing. This data is solely of the user Akash gesturing in ASL, with the images taken from his laptop's webcam. These photos were then cropped, rescaled, and labelled for use.

Methods

Step1.Data Pre-processing: The data preprocessing done using the PILLOW library, an image processing library, and sklearn. decomposition library, which is useful for its matrix optimization and decomposition functionality.

Step 2. Image Enhancement: A combination of brightness, contrast, sharpness, and color enhancement was used on the images. For example, the contrast and brightness were changed such that fingers could be distinguished when the image was very dark.

Step 3. Edge Enhancement: Edge enhancement is an image filtering technique that makes edges more defined. This is achieved by the increase of contrast in a local region of the image that is detected as an edge. This has the effect of making the border of the hand and fingers, versus the background, much clearer and more distinct. This can potentially help the neural network identify the hand and its boundaries.

Step 4. Image Whitening: ZCA, or image whitening, is a technique that uses the singular value decomposition of a matrix. This algorithm decorrelates the data, and removes the redundant, or obvious, information out of the data. This allows for the neural network to look for more complex and sophisticated relationships, and to uncover the underlying structure of the patterns it is being trained on. The covariance matrix of the image is set to identity, and the mean to zero.

Step 5. Hand Identification and Segmentation: To segment the hand, I had to choose the size of captured images, the users do not a vary in position, and thus, the hand is of the same size (only with small variations). The effective size for the problem was 100x100 images: these images enclose the hand and have room for small movements.

Step 6. Training: The training loop for the model is contained in train model the model is trained with hyperparameters obtained from a config file that lists the learning rate, batch size, image filtering, and number of epochs. The configuration used to train the model is saved along with the model architecture for future evaluation and tweaking for improved results. Within the training loop, the training and validation datasets are loaded as Data loaders and the model is trained using Adam Optimizer with Cross Entropy Loss. The model is evaluated every epoch on the validation set and the model with best validation accuracy is saved to storage for further evaluation and use. Upon finishing training, the training and validation error and loss is saved to the disk, along with a plot of error and loss over training.

Step 7. Classify Gesture: After a model has been trained, it can be used to classify a new ASL gesture that is available as a file on the filesystem. The user inputs the file path of the gesture image and the test data.py script will pass the file path to process data.py to load and preprocess the file the same way as the model has been trained.

Step 8. Training and Validation: The models are trained using Adam optimizer and Cross Entropy Loss. Adam optimizer is known for converging quickly in comparison with Stochastic Gradient Descent (SGD), even while using momentum. However, initially Adam would not decrease our loss thus we abandoned it to use SGD. Debugging Adam optimizer after our final. This is the main step in which put images into training set an test set to use the sequence model built by the Keras library.

Step 9. Neural Networks: Neural Networks, are based on connected perceptron's organized in different layers. The neurons in the same layer are not connected between them, they are connected to the next layer. The most common and simplest ones have three layers, the input layer (this layer has one neuron per each variable in the feature vector), the hidden layer (depending on the problem, the number of perceptron's vary), and the output layer (it contains one neuron per class in the dataset) which determines the class of the input data.

1.3. Convolutional Neural Networks: A convolutional layer is the first layer the input data faces when it enters into the CNN. The task of convolution in an image needs three elements: an image to convolve, a convolution filter and a convolution stride parameter. As an example, an image of 24x24 pixels as input; the convolution filter is a smaller image learned from training the network, let's say it is a 5x5 matrix. Starting from a corner of the image, the filter is mapped to a region of the image called a receptive field. Each pixel in the image is considered a neuron and the values of them are multiplied by convolution filter neurons; the resulting values are added and a single value is computed to store it in the feature map. After convolution is performed, the receptive field is shifted based on the stride parameter to cover the entire image. Multiple convolution filters lead to multiple feature maps, increasing the recognized features. The training of the CNN focuses on learning the value of neurons of convolution filters, called weights. At the very beginning, these values are randomized and after applying backpropagation algorithms, the network updates the weights to fit training data.

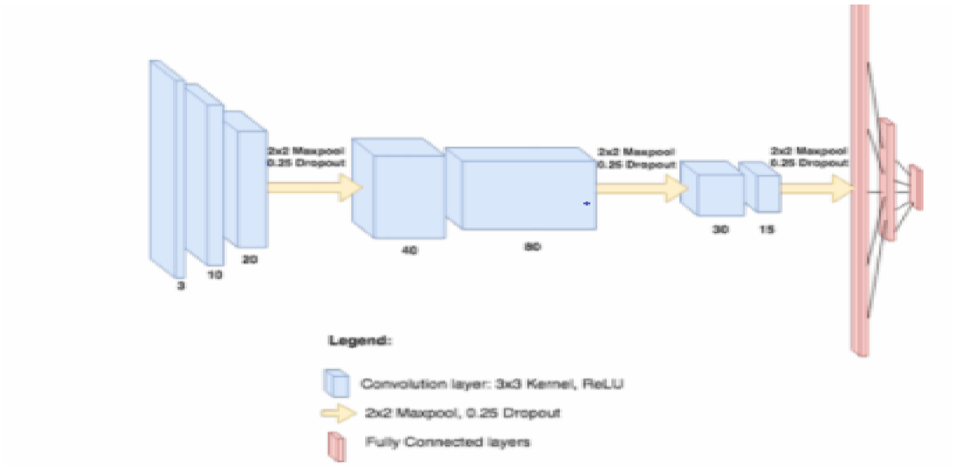

Fig -1: Model Architecture as implemented in Using Deep Convolutional Networks for Gesture Recognition in American Sign Language

1.4. Applications: The physical movement of the human hand produces gestures, and hand gesture recognition leads to the advancement in automated vehicle movement system. In this paper, the human hand gestures are detected and recognized using convolutional neural networks (CNN) classification approach. This process flow consists of hand region of interest segmentation using mask image, fingers segmentation, normalization of segmented finger image and finger recognition using CNN classifier. The hand region of the image is segmented from the whole image using mask images. The adaptive histogram equalization method is used as enhancement method for improving the contrast of each pixel in an image. In Section headings should be left justified, bold, with the first letter capitalized and numbered consecutively, starting with the Introduction. Sub-section headings should be in capital and lower-case italic letters, numbered 1.1, 1.2, etc., and left justified, with second and subsequent lines indented. All headings should have a minimum of three text lines after them before a page or column break. Ensure the text area is not blank except for the last page. In this paper, connected component analysis algorithm is used in order to segment the finger tips from hand image.

- Helps deaf and hard hearing people to exchange information between their own community and with other people.
- Deals from sign gesture acquisition and continues till text/speech generation.
- CNNs are able to automate the process of feature construction.
- Sign language recognition appertains to track and recognize the meaningful emotion of human made with fingers, hands, head, arms, face etc.

1.5. ***Conclusions***

The problem of Sign Language (SL) recognition using images is still a challenge. Similarity of gestures, user's accent, context and signs with multiple meanings lead to ambiguity. These are some reasons why previous work used limited datasets. This study provided information about the constraints of the problem. So, concluded that a dataset to train and evaluate the system must have sufficient gesture variations to generalize each symbol. The average method was based on [64], but in my case the focus was hand images and parallelized the method using the GPU to increase the efficiency.

1.6. Future scope

The Hand Gesture recognition is moving at tremendous speed for the futuristic products and services and major companies are developing technology based on the hand gesture system and that includes companies like Microsoft, Samsung, Sony and it includes the devices like Laptop, Hand held devices, Professional and LED lights. The verticals include where the Gesture technology is and will be evident are Entertainment, Artificial Intelligence, Education and Medical and Automation fields. And with lot of Research and Development in the field of Gesture Recognition Field, the use and adoption will become more cost effective and cheaper. It's a brilliant feature turning data into features with mix of technology and Human wave. Smart phones have been experiencing enormous amount of Gesture Recognition Technology with look and views and working to manage the Smartphone in reading, viewing and that includes what call touch less gestures. Google Glass has been also in the same cadre. And the Technology has also been embedded into smart televisions nowadays as well, which can easily control and managed by Voice and Hand options. In the medical fields Hand Gesture may also be experienced in terms of Robotic Nurse and medical assistance. As the Technology is always revolving and changing the future is quite unpredictable but have to be certain the future of Gesture Recognition is here to stay with more and eventful and Life touching experiences.

References

Exploiting Recurrent Neural Networks and LeapMotion Controller for the Recognition of Sign Languageand Semaphoric Hand Gestures Danilo Avola , IEEETransactions on multimedia, vol. 21, no. 1, January2019.

Image processing with neural network review.Pattern recognition, IEEE transactions Michael
EgmontPetersen, Dick de Ridder, and Heinz .2004[3] Robust part-based hand gesture recognition usingkinect sensor. Zhou Ren, Junsong Yuan, Jingjing Meng,and Zhengyou Zhang, IEEE transactions on multimedia,
2013.

The leap motion controller: a view on sign language,Leigh Ellen Potter, Jake Araullo, and Lewis Carter. IEEEtransactions ,2013.

Hand gesture recognition using color markers,Mubashira Zaman, Soweba Rahman, Tooba Rafique,Filza Ali, and Muhammad Usman Akram, 2016.

Recurrent convolutional neural networks for textclassification Siwei Lai, Lihong Xu, Kang Liu, and Jun
Zhao. 2015.

A deep neural framework for continuous signlanguage recognition by iterative training, IEEE
transactions on multimedia 2019.

Osama Abdel-Hamid, Abdel-Rahman Mohamed Jiang, and Gerald Penn. Applying convolutional neural networks concepts to hybrid nn-hmm model forspeech recognition. In 2012 IEEE internationalconference on Acoustics, speech and signal processing(ICASSP), pages 4277–4280. IEEE, 2012.

Anant Agarwal and Manish K Thakur. Sign languagerecognition using Microsoft kinect. In ContemporaryComputing (IC3), 2013 Sixth International Conferenceon, pages 181–185. IEEE, 2013.