IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 1, January 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001051

A Survey on Sign Language Recognition using Machine Learning

Sarah Taranum K*, Somya*, Srishti Nema*, Pratilipi Aich*, Dr. Sushila Shidnal#

Student, Dept. of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bangalore, India*

Professor, Dept. of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bangalore, India#

ABSTRACT:Communication between a specially abled person who is deaf and mute and a normal person is a challenging task. We are so used to communicating through speech but it is extremely important to remove the barriers of communication and facilitate smoother conversations. One of the ways of doing this is through Sign language. Thus, recognition of sign language through computerized systems and software development can be very helpful for real time communication. This paper reviews the different processes through which an assistive application can be made for facilitating communication with the people who are a part of the specially abled deaf and mute community. The main objective is to develop an image-based Sign Language recognizer software which can identify the hand gestures real-time, using Convolution Neural Networks.

KEYWORDS:Sign language Recognition, CNN, skin segmentation, machine learning, YCbCrcolor space, gesture recognition.

I. INTRODUCTION

Communication bridges the gap between people and allows them to share knowledge and grow. What would the human race be without this deep connection with each other, that communication brings. It might have been that we would have been just another species. We are as great as the link of communication that helps us to share the information. To develop that information together and benefit from it we have been and always will be as great as the extend of coming together. We are what our thoughts and ideas are, and together we bring together such ideas and thoughts. As communication becomes more robust and globally unified, we have not been able to make a parallel progress as it comes to the idea of Sign Language. Communication between people who are physically different and have difficulty in expression in the mainstream cannot and must not be held back. Their ideas are as unique as others or maybe more for the see and engage with world in a different way. Hence the platforms need to be as accessible to them as they are for others. Normally we're used to communicating through speech, but there are several other ways of communication. One of those ways is Sign Language, ordinarily used by the Deaf community. There is a very vague understanding of the sign language especially when it comes to the communication between the hearing and the hearing impaired. Finger-spelling or finger gestures form the foundation of sign language. Every finger gesture has a definitive meaning and rule, and is extremely useful to spell out words that don't have a proper sign assigned to them. For the effortless interaction involving the deaf and the mute community, sign language interpreters are highly preferred. A computerized system that can interpret the signed gesture and translate it into audio or plain text can be reasonable and immensely helpful for real time communication. Extremely talented individuals may have trouble with their hearing or speech, this shouldn't inhibit them from the workforce and accomplish what they desire. The cornerstone of this research is towards classification of finger gestures aiming for accuracy more than speed that can prove to be helpful for beginners especially children.

While there has been ample research around this concept, there are several challenges or concerns affecting sign language recognition. To name a few:

- There may be variations in the size of hand and speed of the hand gesture
- Video recognition gets affected by the environmental or background changes
- Occlusion of the hand poses problems
- Use of an add-on device
- Detecting the start and end of a sign must be automatic. Also, a sign may be affected by its preceding and succeeding signs
- Processing high resolution images or a large amount of data may hinder real time detection

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001051

The main purpose of this research is to study the previous works in the sign language interpretation area, explore the different methods and algorithms and an approach towards building a real-time Sign Language recognizer based on Convolutional Neural Networks.

II. LITERATURE SURVEY

A. Priyanka Mekala et al. (2011) [1]

A combinational neural network which extracts the features from the video stream obtained by using a camera for gesture recognition. By adopting the usage of MATLAB software, the recognized gestures are connected to the audio signals thereby facilitating a communication between the ordinary and disabled people. The camera consists of a camera sensor which senses the hand gestures when it is placed in front of the signer to capture the front view. Filtering of the unwanted noise from the image scene takes place and then we run a Background Subtraction process using the Running Gaussian Average method. In the gesture classification stage where the gestures using features computed from the video is recognized using a Combinational Neural Network. This approach gives a clear view of the parallelism between the feature extraction layer and the neural network layer.

B. Jie Huang et al. (2012) [2]

Microsoft Kinect is used as the input device which provides both color and depth video stream as well as tracks the movement of the user's body. The baseline method is taken as Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) used for pattern recognition. The trajectory and hand-shape features are extracted to train GMM-HMM. To effectively integrate the motion information for video analysis of sign language recognition, we use the 3D convolution in the CNN's convolution layers so that both spatial and temporal dimensions are captured. In order to capture the motion information, the feature maps present in the convolution layer are connected to multiple contiguous frames and the multiple 3D convolutions are applied with distinct kernels to the same location in the previous layer.

C. Vivek Bheda et al. (2019) [3]

This paper aims to explore real time Indian Sign Language recognition. The real time execution is done using video data rather than images. The proposed system consists of three stages: Preprocessing, Feature extraction and classification. Eigenvalues and Eigenvectors are used for features and Eigen value weighted the Euclidean distance-based classification technique is being used. The database consists of videos captured using a webcam where different letters of the alphabet were taken into consideration. Skin filtering algorithm has been used for detection of hand gestures. HSV color space is used for filtering the image frames. Histogram matching is done in the next step for different frames. The similarities in frames can be measured using a metric called Difference given by:

$$\text{Difference}(n) = \text{Hist}(n) - \text{Hist}(n-1).$$

The Euclidean distance between Eigen vectors is given by:

$$ED = \text{sqrt}\,(VT(n) - VD(n)).$$

D. B.D Nguyen and Hung Ngoc Do (2019) [4]

This model uses Caffe implementation of CNN using a depth sensor called Creative Senz3D camera of the resolution 320x240. We take 31 signs (alphabets and number) into consideration. Since the depth sensor cannot capture depth from black objects well a black wrist band is worn to obtain the depth voids around the wrist. Hand segmentation provides a simple and effective strategy of finding the connected components from the closest region of the depth image. The neural networks are trained and tested in five operating modes. From the data separation, in the first case the subjects are not separated and in the second case, the subjects are separated for training, validation and testing and involves many subjects. The former case gives the average result with respect to the three subjects while the latter gives an average result of training for four subjects and test with one subject and the performance increased to about 2-3%.

E. Omkar Vedak Prasad Zavre et al. (2019) [5]

This paper aims to recognize sign language using Pattern Recognition and gesture recognition techniques. The image-based gesture recognition is done in six different stages in the proposed methodology: Acquisition of Data,

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001051

Image preprocessing, Thresholding, Image Analysis and Contour detection. The RGB images captured through the webcam are then converted to YCbCr images. These are then further converted to binary images. The RGB image frames also need to be converted to HSV(hue, saturation, value) images. The process of thresholding helps in the generation of binary images from grayscale images. Blob detection is used to remove patches in the image which are formed due to collections of light or dark pixels. Contour detection is performed in the end to mark a clear boundary between the hand and the background for better recognition. Convex hull algorithms are used along with Gaussian filters to scale images and extract segmented contours. The recognition process is done by matching templates of hand by considering contour shapes.

F. KshitijBantupalli (2018) [6]

Google developed an Inception Model for image recognition which is used as the CNN model. The hand gestures are recorded by the camera and the videos are preprocessed to frame sequences and given to the CNN resulting in two outputs. 2048 size vectors are given by the global pool allowing the RNN to analyze more features. For long term dependencies the feature sequence is given to the LTSM where longer sequences of data have higher accuracy.For Gesture detection the Inception Model is utilized and all the convolutions are performed in parallel and the feature maps obtained are joined before going to the next layer. This operation is repeated several to attain a deeper network thereby retaining the final net-work. The outputs of Softmax Layer and Max Pooling layer are taken for the architecture and given to the RNN. Here the gesture classification process takes place where the LTSM classifies the gesture segments which are processed by the CNN into a gesture class using sequence data. By augmenting the test data set more information from the predictions were gathered.

G. Ashish S. Nikam and Aarti G. Ambekar (2016) [7]

This paper deals with creating an application for Sign language recognition and its conversion to English in the form of text and audio. The application developed takes the data from the webcam of the computer, then it is pre-processed using a combinatorial algorithm. The next step is recognizing the images by template matching. The proposed system takes place in two stages, Segmentation of hand and Recognition of hand sign. The feature descriptor used is histogram of oriented gradients (HOG) because of its ability to adapt to illuminations and rotations. The segmentation process is done using YCbCr color space and the classifier used is SVM classifier. The skin color is detected using the YCbCr model and edge detection is done by Canny edge detector. The feature extraction is done by HOG method. Template matching and recognition is done by using a SVM classifier. The Google text to speech API here is used to change the text output to audio.

H. Shadman Shahriar et al. (2018) [8]

Input images are obtained using a webcam. Data is processed using the skin color model. Human skin color is determined as follows 77<=Cb<=127 &130<=Cr<=173 using YCbCr perception color space. Background images having the same color as skin (door, wooden table) are removed by taking the absolute difference with the current image. The camera and other external noises distort the mask. Hence a denoising step is carried out followed by some morphological operations. A bounding box is used to extract the portion of interest for image feature extraction. The features are extracted via a trained CNN. Finally, a classifier trained via Deep Learning is used to classify the input image and determine the meaning of the hand gesture.

I. Byeongkeun Kang and Subarna Tripathi Truong Q (2015) [9]

This paper aims to develop a Sign Language Fingerspelling Recognition System using Image Processing, Supervised Learning and Deep Learning.There are two approaches to Sign language translation: vision-based and non-vision-based approach. The vision-based approach is more widely used and consists of software approach. The preprocessing of data is done by color models, specifically using the YCbCr model. Skin segmentation and detection is done using an integration of Recurrent Neural Network (RNN) layers to a Convolutional Neural Networks (CNN) model. Feature extraction and fingertips detection is done by convex hull algorithms. Principal Component Analysis(PCA) is used to change the dimension of feature vectors. Training and Classification is done by Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Hidden Markov Models (HMM).
These different methods are designed as a system which is given below:

- Input image → skin segmentation(feature detection) → crop and resize(feature extraction with dimensional adjustments) →
  1. HOG(Histogram of Gradients)+LBP(Linear Binary Pattern) descriptors → SVM classifier → output
  2. HOG descriptor→SVM classifier →output

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001051

3. LBP descriptor →SVM classifier →output
- CNN(end to end)
- CNN descriptor →SVM classifier →output

J. Joyeeta Singha and Karen Das (2013) [10]

General CNN architecture used which consists of multiple convolutional layers and the dense layers. Images were captured for each sign and by using the background- subtraction tools the backgrounds of each of the images were removed. Validation accuracy was high on single dataset but decreases when a mix of dataset is used. Premade dataset was also used to compare self-generated dataset's performance. A pipeline was developed such that it can be used by people to generate and add images to the dataset. By transforming their images just a few pixels the accuracy increased by 0.05. The performance increases with better representative initial training data and augmentation.

K. Murat Taskiran et al. (2019) [11]

This study proposes a CNN (Convolutional Neural Network) structure for feature extraction and classifier. Skin color detection and convex hull algorithms are used to establish hand position,result is then resized and given for classification to a trained neural network. ReLu is used as activation function. Scikit-Learn library is used to shuffle the array of image sample so as to split it for training and testing. The model is then created as a sequential network and starts the fitting process. The model and weights of the neural network are then fed into a real-time recognition algorithm. The algorithm consists of two parts - extracting hands bound convex hull point and classifying hand images with convolutional neural networks, running parallelly for better precision.To obtain skin region detection, image color space is converted from RGB to YCbCr. Pixel values are thresholded to create a mask. Then the image is processed with a bitwise AND gate with itself using the mask.Followed by pre-processing for noise reduction.Maximum connected area is calculated, and the bounding rectangle is obtained. Using OpenCV library functions, convex hull and convexity defections are calculated.

L. Lionel Pigou(B) et al. (2019) [12]

This paper considers a recognition system using the Microsoft Kinect, CNN and GPU acceleration. Automates the feature extraction using CNN. Performs 2 steps: Extract the features from frame sequences using CNN and Classification of the signs/gestures using ANN. The initial step is to crop the highest hand. If it is the left hand, then it is mirrored. This results in four video samples (32 frames of size 64x64). Dropout and data augmentation to reduce overfitting are applied to the CPU while the model works with GPU. It uses Nesterov's accelerated gradient descent (NAG) with a fixed momentum coefficient of 0.9, mini batches of size 20, learning rate of 0.003 with 5% decrease after every epoch. It is implemented using Python libraries Theano and PyLearn2 leading to faster implementation of 2D convolutions by Alex Krizhevsky.It also uses sliding window technique to spot gestures in video samples and to predict begin and end of the frames.

III. CONCLUSION

In this paper we have reviewed the different ways of Sign Language Recognition through techniques involving CNN and Deep Learning. We see how Deep Learning along with CNN can be used for Sign Language Recognition and compare the different models involving Supervised Learning, MATLAB and Back Propagation, GPU Acceleration and Real-time Skin Segmentation Algorithm. It gives a detailed analysis of the concepts of Sign Language Recognition along with the respective efficiency of different techniques.

REFERENCES

[1] Real-time sign language recognition based on neural network architecture, Priyanka Mekala, Ying Gao, Jeffrey Fan, AsadDavari, Dept. of Electrical and Computer Eng., Florida International University, FL, U.S.A, Electrical Eng. Dept., University of Wisconsin, Platteville, WI, U.S.A, Dept. of Electrical and Computer Eng., West Virginia University Institute of Tech, WV, U.S.A, Conference Paper April 2011.

[2] S. Chundru, "Ensuring Data Integrity Through Robustness and Explainability in AI Models," Transactions on Latest Trends in Artificial Intelligence, vol. 1, no. 1, pp. 1-19, 2020.

[3] Sign language recognition using 3d convolutional neural networks, Jie Huang, Wengang Zhou, Houqiang Li, and Weiping Li, University of Science and Technology of China, Hefei, China, 2012.

[4] R. Sugumar, A. Rengarajan and C. Jayakumar, Design a Weight Based Sorting Distortion Algorithm for Privacy Preserving Data Mining, Middle-East Journal of Scientific Research 23 (3): 405-412, 2015.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 1, January 2021 ||

DOI:10.15680/IJIRSET.2021.1001051

[5] Using Deep Convolutional Networks for Gesture Recognition in American Sign Language, Vivek Bheda and N. Dianna Radpour, Department of Computer Science, Department of Linguistics State University of New York at Buffalo, 2019.

[6] Deep Learning for American Sign Language Fingerspelling Recognition System, Huy B.D Nguyen and Hung Ngoc Do, School of Electrical Engineering, International University, Vietnam National University - Hochiminh City, Vietnam, 2019 26th International Conference on Telecommunications (ICT), April 8 - 10 , 2019, Hanoi, Vietnam.

[7] G. Vimal Raja, K. K. Sharma (2015). Applying Clustering technique on Climatic Data. Envirogeochimica Acta 2 (1):21-27.

[8] Sign Language Interpreter using Image Processing and Machine Learning, Omkar Vedak Prasad Zavre, Abhijeet Todkar, Manoj Patil, Department of Computer Engineering, Datta Meghe College of Engineering, Mumbai University, Airoli, India Assistant Professor, Department of Computer Engineering, Datta Meghe College of Engineering, Mumbai University, Airoli, India, International Research Journal of Engineering and Technology (IRJET), Volume: 06, Issue: 04 | Apr 2019.

[9] V. Balasubramanian and Sugumar Rajendran, "Rough set theory-based feature selection and FGA-NN classifier for medical data classification," Int. J. Business Intelligence and Data Mining, vol. 14, no. 3, pp. 322-358, 2019.

[10] American Sign Language Recognition using Deep Learning and Computer Vision, KshitijBantupalli, Department of Computer Science Kennesaw State University, Kennesaw, USA, Ying Xie, Department of Computer Science Kennesaw State University, Kennesaw, 2018.

[11] K. Thandapani and S. Rajendran, "Krill Based Optimal High Utility Item Selector (OHUIS) for Privacy Preserving Hiding Maximum Utility Item Sets", International Journal of Intelligent Engineering & Systems, Vol. 10, No. 6, 2017, doi: 10.22266/ijies2017.1231.17.

[12] Sign Language Recognition Using Image Based Hand Gesture Recognition Techniques, Ashish S. Nikam1 , Aarti G. Ambekar2 1P.G. Student, Dept. of Elect. & Telecom Dwarkadas J. Sanghvi College of engg. Mumbai, India, 2016 Online International Conference on Green Engineering and Technologies (IC-GET).

[13] Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning, Shadman Shahriar, Ashraf Siddiquee, Tanveerul Islam, Abesh Ghosh, Rajat Chakraborty, Asir Intisar Khan,Celia Shahnaz, Shaikh Anowarul Fattah, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology,Dhaka-1205, Bangladesh, Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018).

[14] Soundappan, S.J., Sugumar, R.: Optimal knowledge extraction technique based on hybridisation of improved artificial bee colony algorithm and cuckoo search algorithm. Int. J. Bus. Intell. Data Min. 11, 338 (2016)

[15] Real-time Sign Language Fingerspelling Recognition using Convolutional Neural Networks from Depth map, Byeongkeun Kang, Subarna Tripathi Truong Q. Nguyen Department of Electrical and Computer Engineering University of California, San Diego, 2015.

[16] Recognition of Indian Sign Language in Live Video, Joyeeta Singha, Karen Das Department of Electronics and Communication DBCET, Assam Don Bosco University Guwahati, Assam Department of Electronics and Communication DBCET, Assam Don Bosco University Guwahati, Assam, International Journal of Computer Applications, Volume 70– No.19, May 2013.

[17] G Jaikrishna, Sugumar Rajendran, Cost-effective privacy preserving of intermediate data using group search optimisation algorithm, International Journal of Business Information Systems, Volume 35, Issue 2, September 2020, pp.132-151.

[18] A Real-Time System For Recognition Of American Sign Language By Using Deep Learning, Murat Taskiran, Mehmet Killioglu, and NihanKahraman Electronics and Communications Engineering Yildiz Technical University,Istanbul, Turkey, 2019.

[19] Sasidevi Jayaraman, Sugumar Rajendran and Shanmuga Priya P., "Fuzzy c-means clustering and elliptic curve cryptography using privacy preserving in cloud," Int. J. Business Intelligence and Data Mining, Vol. 15, No. 3, 2019.

[20] G. Vimal Raja, K. K. Sharma (2014). Analysis and Processing of Climatic data using data mining techniques. Envirogeochimica Acta 1 (8):460-467.

[21] Sign Language Recognition Using Convolutional Neural Networks, Lionel Pigou(B), Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen ELIS, Ghent University, Ghent, Belgium, 2019.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details