American Sign Language: A Survey

Nidhi Iche , Shanthi Guru

***Abstract*— Sign language is complex to understand but the complete language which involves the hand's movement, facial expressions, and body postures. Sign language is core communication media to the people which cannot speak. It is not a universal language means every country has its own sign language. Every country has its own grammar for their sign language, word orders and pronunciation. The issues arise when people try to communicate using their language with the people who are unaware of this language grammar. In this paper a survey is been made to highlight the various research works carried out by the researches and comparative analysis of those works in recognizing American Sign Language. An attempt is made to present sign language recognition approaches and various techniques, some constraints in gesture recognition, process.**

***Keywords*- America Sign Language, Gesture Recognition, Hand Gesture Recognition (HGR), Sensor Based, Vision Based.**

I. INTRODUCTION

Sign language is a way by which the gestures made by the user are used for communication. Human gestures are one of the powerful and efficient ways of interaction and express the feelings. The language allows the speech impaired people to communicate with rest of the world [2]. But issue with sign language is that it is only confined to the people who are deprived of speech. In order to communicate, normal people also have to be aware of the sign language.

Sign language is a language through which communication happens without the means of acoustic sounds. It relies on sign patterns like orientation and movements of the arm that aids in communication between individuals. There are very subtle differences in the gestures between similar signs. The ASL makes it challenging to capture and distinguish differences between distinct signs [8]. The deaf and dumb people are the communication part, to convey their thought with other deaf and dumb people and with other normal people [3].

- **What is American Sign Language?**

American Sign Language (ASL) is a complete and complex language that operates on the signs made with the hands and other movements, including facial expressions and different gestures of the body. It is the first language of many deaf all over the world, and one of several communication options available to deaf people. ASL is said to be the popular and fourth most commonly used language in the United States.

The system is designed to visually recognize all static gestures of American Sign Language (ASL) with a bare hand. Different users have different hand shapes and skin colors, making it more difficult for the system to recognize a gesture. The system combines five feature extraction algorithms for user independent and robust hand gesture recognition [7, 9]. The whole system works in four steps for gesture recognition such as image acquisition, pre-processing, feature extraction and feature recognition. Image frames taken by video camera interfaced with the computer are tested by our trained CNN [5]. The CNN is trained with sample images of our database and it recognizes ASL alphabets and numbers with accuracy in a real-time environment [10].

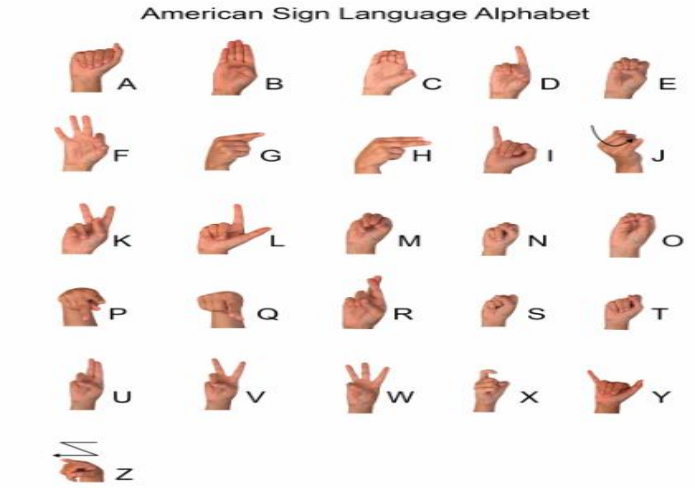


Figure 1: - 26 alphabets of American Sign Language.

II. LITERATURE SURVEY

Yifan Zhang, Congqi Cao, Jian Cheng, and Hanqing Lu [1] proposed “EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition " In this work, they have introduced up-to-date the largest dataset called EgoGesture for the task of egocentric gesture recognition with sufficient size, variation and reality, to successfully train deep networks. The dataset is more complex than any existing datasets as our data is collected from the most diverse scenes. Compared to gesture classificationin segmented data, the performance on gesture detection is far from satisfaction and has much more space to improve.

Ponlawat Chophuk, Kanjana Pattanaworapan, Kosin hamnongthai [2] proposed “Fist American Sign Language Recognition Using Leap Motion Sensor " in which the problem of the fist signs in American Sign Language is solved because their recognition is not perfect due to the Euclidian distances of the fingertip to palm position used in 3D of those signs are similar, so it is difficult to recognize them. We therefore propose a system of fist. American Sign Language recognition with a bare hand in depth plane by using 3D non-contact motion sensor. In this system, two patterns of the polygon area between the consecutive fingertip positions with palm position in depth plane by Shoelace formula are used to identify the fist sign language: 1) six triangles area, 2) one hexagon area, then a decision tree is applied to classify the alphabets. The results showed the 7 alphabets in fist ASL using the researchers hand. The accuracy of the method proposed is approximately 96.1%.Nevertheless, N and T alphabets were not perfect because the Euclidian distances of the fingertip to palm position proposed in 3D of those signs are similar, so it is difficult to recognize them.

Suraksha Devi, Suman Deb [3] proposed “Low cost tangible glove for translating sign gestures to speech and text in Hindi language " where primary objective of this work is to enable the speech impaired and the mute to have a seamless communication and independent living in the society. The proposed prototype and accompanying algorithm accomplished the initial objective. The glove is cost effective and is capable of translating sign gestures (conventional Indian Sign Language) into speech-text in real time using android application on the phone. The significant recording of hand movements and feature extraction is done by applying Principle Component Analysis. The glove prototype is independent of the surrounding light or any other kind of interference. As a result, precise and accurate recognition gesture is possible in less time. This paper emphasizes on the translation of sign language With the help of the glove; however, the glove could be used for various other applications such as for the virtual reality interaction, gaming, entertainment, education technology, robotics etc. This is an ongoing work; the of the algorithm and prototype are going on.

Peijun Bao, Ana I. Maqueda, Carlos R. del-Blanco, and Narciso Garca [4] proposed “Tiny Hand Gesture Recognition without Localization via a Deep Convolutional Network " Visual hand-gesture recognition is being increasingly desired for human computer interaction interfaces. In many applications, hands only occupy about 10 whereas the most of it contains background, human face, and human body. Spatial localization of the hands in such scenarios could be a challenging task and ground truth bounding boxes need to be provided for training, which is usually not accessible. However, the location of the hand is not a requirement when the criteria is just the recognition of a gesture to command a consumer electronics device, such as mobiles phones and TVs. In this paper, a deep convolutional neural network is proposed to directly classify hand gestures in images without any segmentation or detection stage that could discard the irrelevant not-hand areas. The designed hand-gesture recognition network can classify seven sorts of hand gestures in a user-independent manner and on real time, achieving an accuracy of 97.1complex backgrounds.

Md. Mohiminul Islam,Sarah Siddiqua, and Jawata Afnan [5] proposed “Real Time Hand Gesture Recognition Using Di_erent Algorithms Based on American Sign Language " In which they have applied different algorithms for feature extraction of hand gesture recognition system. This includes K convex hull for fingertip detection, pixel segmentation, eccentricity, elongatedness of the object. The experimental results show that K convex hull algorithm gives more accurate fingertip detection. Besides, other algorithms are applied to get greater accuracy in recognition system. Image frames taken by mobile video

camera interfaced with the computer are tested by our trained ANN.

Shaoqing Ren, Kaiming He Ross Girshick Jian Sun [6] proposed "Faster R- CNN: Towards Real-Time Object Detection with Region Proposal Networks " State-of-the-art object detection networks depend on region proposal algorithms to hypothesize object locations. Advances like SPPnet and Fast R-CNN have reduced the running time of these detection networks, exposing region proposal computation as a bottleneck. In this work, author introduce a Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, thus enabling nearly cost-free region proposals. An RPN is a fully-convolutional network that simultaneously predicts object bounds and objectness scores at each position. RPNs are trained end-to-end to generate highquality region proposals, which are used by Fast R-CNN for detection. With a simple alternating optimization, RPN and Fast R-CNN can be trained to share convolutional features.

Ross Girshick, Je_ Donahue, Trevor Darrell, and Jitendra Malik [7] proposed "Region-based Convolutional Networks for Accurate Object Detection and Segmentation " Object detection performance, as measured on the canonical PASCAL VOC Challenge datasets, plateaued in the _nal years of the competition. The best-performing methods were complex ensemble systems that typically combined multiple low-level image feature with high-level context. In this paper, we propose a simple and scalable detection algorithm that improves mean average precision (mAP) by more than 50approach combines two ideas: (1) one can apply high-capacity convolutional networks (CNNs) to bottom-up region proposals in order to localize and segment objects and (2) when labeled training data are scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, boosts performance significantly. Since we combine region proposals with CNNs, we call the resulting model a R-CNN or Region-based Convolutional Network.

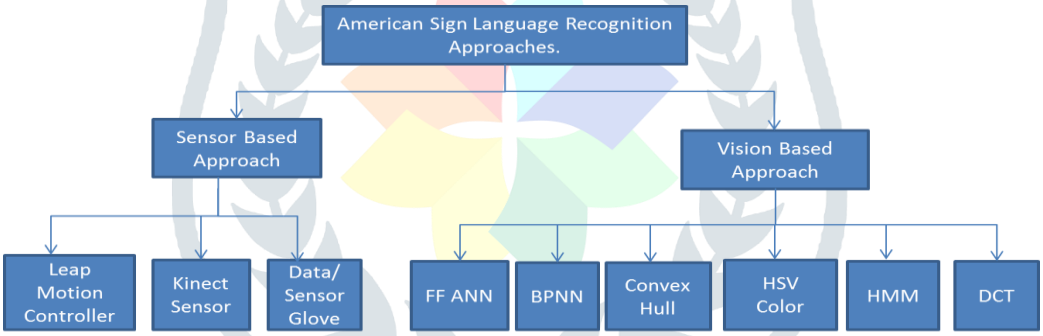
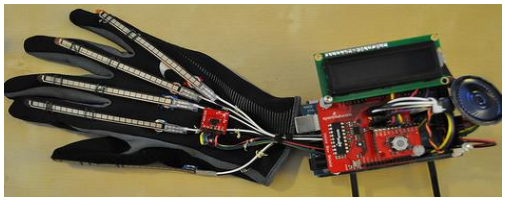


Figure 2: - American Sign Language Recognition Approaches.

SENSOR BASED APPROACHES

This approach uses different types of sensors and external hardware to collect the data of different gesture (motion) done. In Hand Gesture Recognition (HGR), multiple sensors are placed on the hand to make hand gesture, once hand gesture is done then the data is recorded for further analysis. Main drawback is that complex gestures cannot be performed. Further sensor based approaches are classified as follows:

Data/Sensor Glove approach: - Here signer needs to wear a sensor glove to detect hand gesture signal and finger motions. Data glove also contains flex sensor and accelerometer or gyroscope [4]. Flex sensor detects the bend signal also. Major drawbacks of this approach is that it leads to less natural behaviour as the signer needs to wear the sensor glove and sensor gloves are pretty expensive.



Figure(a): - Glove with flex sensor and accelerometer

Leap Motion Controller approach: **-** It is a small USB peripheral sensor, detects the movement of hand and translates the detected signal into computer commands It contains three infrared (IR) LED's and two IR cameras. Camera generates around 300 frames per second of reflected data.These signals are sending to the computer over USB cable for processing.

Advantages: **-** ASLR system using leap Motion Sensor, provides robust and economical result comparing to cyber glove sensor.

Figure (b): - Leap Motion Sensor

Kinect Sensor approach:- It is a motion sensor with Xbox 360 which has depth sensor consist of IR laser, RGB camera and Multi-array microphone. It supports 3D motion capture of complete body. It also recognizes facial movement and voice too. Cao Dong et al introduce a method for 24 ASL alphabet recognition. It is very useful in controlling industrial robots, factory floors, and hospital.

Figure (c): -Kinect Sensor

TECHNIQUES OF VISION BASED ASLR PROCESS

Feed Forward Artificial Neural Network (FF ANN):- It is dissimilar from Recurrent Neural Networks. In FF ANN, connections between the units do not form a cycle. [6]. The FF ANN was the simplest and foremost type of ANN invented. Here, the information moves in single direction (One/Forward Direction), forward from the input nodes to the output nodes through the hidden nodes.
Advantage: -37 ASL signs and numbers are recognized with 30 different feature vectors using FF ANN and back propagation NN algorithms [9].

Convex Hull: **-** Convex envelope (hull) is a set of X points in a Euclidean plane/space is the tiniest convex set, which contains X [11]. When X is a bounded plane subset, the convex hull may be pictured as the shape enclosed by a rubber band and stretched around X. The convex envelope may be defined as the intersection of all convex sets containing X.

HSV Color Model: **-** It is a perceptual color model that describes color based on 3 elementary features. The color Hue(H) is a basic feature of color ranging from 0 to 360 in a color space such as red, yellow, green. Saturation (S) defines the purity of the color in a color space. S specifies the amount of grey color, ranging up to 0to 1, where 0 is grey and 1 is primary color. Value (V) is the brightness of the color which changes its saturation ranging between 0 to 100% where 0 is totally black and 100 is pure white in color space.
Advantage: -This model permits us to state the skin pixel boundary only in terms of hue and saturation [14].

Morphological Operation: **-** In mathematics, Morphology is a tool for mining the useful image components. It contains representation and description of region shape, i.e convex hulls, skeletons, and boundaries. The language of mathematical morphology is set theory, and is used to apply directly on binary images. The two basic operations on sets are; an image, and the structuring element. [14], an erosion and dilation are the two components that are used to get a thin structure of the edge to detected binary image for ASL gesture recognition.

Hidden Markov Model (HMM):- HMM is a Statistical Markov Model (SMM), standard system that is assumed to be a Markov process with hidden (unobserved) states. In HMM, the state is not directly visible, but its output is visible and it is state dependent. A probability distribution of each state is dependent over the possible output tokens. So HMM, contains generated sequence of tokens that provides some information about the sequence of the states [15].

Advantage: - HMMs are very useful in speech, gesture recognition, musical score following, handwriting, partial discharges and bioinformatics.

Principal Component Analysis (PCA):- PCA is a statistical technique, uses an orthogonal transformation to translate the set of inspection of correlated variables to a set of values of linearly variables called principal. So that the first principal component has the biggest possible variance, and each subsequent component has possible under the limitation that it is orthogonal to the earlier components. The resultant vectors are an uncorrelated orthogonal set. PCA is precise towards relative scaling of the original variables and it can be used as a tool in exploratory data analysis for making predictive models as well.

Discrete Cosine Transform (DCT):- DCT is strongly correlate to the Discrete Fourier Transform (DFT) that uses only real numbers so the result is real. DCT inclines to focus information, making it beneficial for several sciences and engineering for lossy image and audio compression applications and serving in mineralising feature vector size for dissimilar applications as well. DCTs are approximately equivalent to DFTs of double the length, functioning on real data with even symmetry. In some variations the input and/or output data are moved by half a sample.

K-Nearest Neighbours (K-NN):- K-NN is the simplest and easiest machine learning algorithm, used for classification and regression purpose. This technique is based on some training samples that are present in the feature space. Distance between the test sample and training samples are calculated initially by using Euclidean distance. Later, least distance 'k' nearest neighbours is determined. [14] The test sample is classified based on the mainstream votes of the nearest neighbours. A non-labelled query image is simply allotted to the label of its k nearest neighbours in testing phase[17].

Support Vector Machines (SVMs):- SVM is a powerful supervised learning method, for analysing the data in classification and regression process. It is also used to classify the test data group as one of the multiple signs, based on the value of the feature. Each sample in a training set is marked as either belonging to classification and regression. SVM is a provision vectors and kernels are working for various learning tasks. Various tasks can perform in different areas by selecting suitable kernel methods [17].

III. RESULT ANALYSIS

Following is the Result Analysis of the average recognition rate of various techniques in Sensor Based and Vision Based approaches for American Sign Language Recognition.

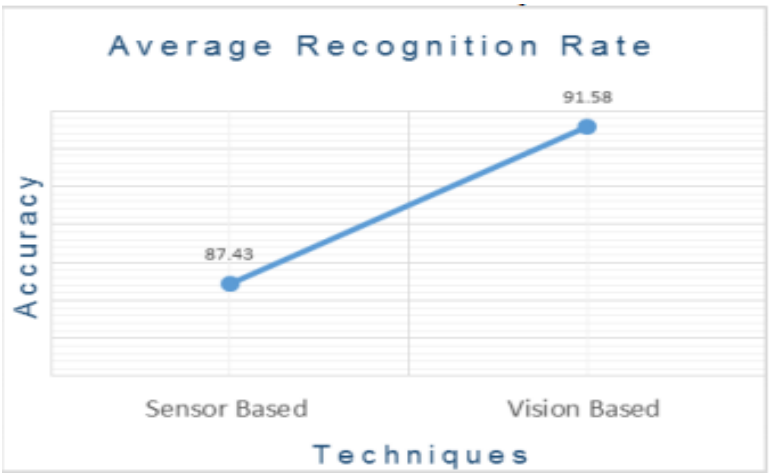


Figure 3: - Chart of the average recognition rate of sensor based and Vision based approach.

IV. CONCLUSIONS

The paper presents the survey of American Sign Language Recognition process, and its various constraints. Also in this paper, an attempt is made to present various sensor and vision based approaches for American Sign Language Recognition. The techniques like Hand Glove with flex sensors, Kinect sensor, Leap Motion Sensor, CGP, SOM, PCA, FF ANN, FDT, SVM, Naïve Bayes, k-NN, and many more are described. From this study it is clear that the vision based approach for gesture recognition has made outstanding progress in its field. One main benefit of vision based approach is that it can be implemented practically and there is no need to wear or use any sensors or hand gloves.

REFERENCES

[1] Yifan Zhang , Congqi Cao, Jian Cheng, and Hanqing Lu “EgoGesture: A New Dataset and Benchmark for Egocentric Hand Gesture Recognition ", IEEE Transactions on Multimedia 2018.

[2] Ponlawat Chophuk, Kanjana Pattanaworapan, Kosin hamnongthai “Fist American Sign Language Recognition Using Leap Motion Sensor ", IEEE 2018.

[3] Suraksha Devi, Suman Deb “Low cost tangible glove for translating sign gestures to speech and text in Hindi language ", IEEE 2017.

[4] Peijun Bao, Ana I. Maqueda, Carlos R. del-Blanco, and Narciso Garca “Tiny Hand Gesture Recognition without Localization via a Deep Convolutional Network ", IEEE 2017.

[5] Md. Mohiminul Islam, Sarah Siddiqua, and Jawata Afnan “Real Time Hand Gesture Recognition Using Di_erent Algorithms Based on American Sign Language", IEEE 2017.

[6] Shaoqing Ren, Kaiming He Ross Girshick Jian Sun “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks "IEEE 2016.

[7] Ross Girshick, Je_ Donahue, Trevor Darrell, and Jitendra Malik “Region-based Convolutional Networks for Accurate Object Detection and Segmentation", IEEE transactions on image processing 2015.

[8] Ching-HuaChuan, Eric Regina, Caroline Guardino, “American Sign Language Recognition Using Leap Motion Sensors”, 13th International Conference on Machine Learning and Applications, 2014, pp.541-544.

[9] S.H. Lee, M.K. Sohn, D.J. Kim, B. Kim, and H. Kim, "Smart TV interaction system using face and hand gesture recognition," in Proc. ICCE, Las Vegas, NV, 2013, pp. 173-174.

[10] S. Kim, G. Park, S. Yim, S. Choi and S. Choi, "Gesture-recognizing hand-held interface with vibrotactile feedback for 3D interaction," IEEE Trans. Consum. Electron., vol. 55, no. 3, pp. 1169-1177, 2009.

[11] D. W. Lee, J. M. Lim, J. Sunwoo, I. Y. Cho and C. H. Lee, "Actual remote control: a universal remote control using hand motions on a virtual menu," IEEE Trans. Consum. Electron., vol. 55, no. 3, pp. 1439-1446, 2009.

[12] A. SharmilaKonwar, B. Sagarika Borah, C. Dr.T. Tuithung, “An American Sign Language Detection System using HSV Color Model and Edge Detection”, International Conference onCommunication and Signal Processing, 2014, pp.743- 747.

[13] Thad Starner, Alex Pentland, “Real-Time American Sign Language Visual Recognition from Video Using Hidden Markov Models. Master's Thesis”, MIT, 1995.

[14] Lum Kin Yun, Tan TianSwee, RinaAnuar, ZuraimiYahya, AzliYahya, “Sign Language Recognition System using sEMG and Hidden Markov Model”, Recent Advances in Mathematical Methods, Intelligent Systems and Materials, 2013, pp.50-53.

[15] F. Erden and A. E. Çetin, "Hand gesture based remote control system using infrared sensors and a camera," IEEE Trans. Consum.Electron., vol. 60, no. 4, pp. 675-680, 2014.

[16] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," IEEE Trans. on Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 142-158, 2016.

[17] S. Nagarajan, T. S. Subashini, Image Based Hand Gesture Recognition using Statistical Features and Soft Computing Techniques, International Journal of Emerging Technology and Advanced Engineering, Vol 5, Issue 7, 2015, 476-482.