A COMPREHENSIVE ANALYSIS ON TECHNOLOGICAL APPROACHES IN SIGN LANGUAGE RECOGNITION

Kamaldeep Sharma[1][0000-0002-4869-8757], Brajesh Kumar[2][0000-0001-5101-7783],

Deepti Sehgal[3][0000-0001-8931-4273] and Ayan Kaushik[4][0000-0003-1012-7972]

[1] Lovely Professional University, Phagwara, Punjab, India
kamaldeep.16958@lpu.co.in
[2] Lovely Professional University, Phagwara, Punjab, India
brajeshkumarjha98@gmail.com
[3] Lovely Professional University, Phagwara, Punjab, India
deese2398@gmail.com
[4] Lovely Professional University, Phagwara, Punjab, India
ayan171299@gmail.com

Abstract. Communication is one of the imperative sides of the 21st century. We as human-being can't even imagine this world without communicating with each other. For honest communication to happen two persons should agree on a common language that each of them understands. Keeping this context in mind, there'll be a necessity for a translator in between a hearing/speech impaired person and a normal person. But the translators or consultants who understand sign language do not seem to be simply and commonly offered all the days, if offered conjointly, they charge tons. To help the social interaction of hearing-impaired - speech debilitated individuals with the society economical interactive communication tools are made. As the importance of sign language interpretation is increasing day by day so several triple-crown applications for linguistic communication recognition comprise new forms of real-time interpretation with progressed artificial intelligence and image processing approaches. In this paper, we gave an entire outline of various methodologies and processes rooted in deep learning and discussed technological approaches for linguistic communication interpretation, also achieving 95.7% accuracy with 2 layers of CNN model classifier for the sign gestures of 26 English alphabets with readily available resources. This paper also provides current analysis and advances in this area and tends to distinguish openings and difficulties for future exploration.

Keywords: Linguistic Communication, Deep Learning, Convolutional Neural Network, Image Processing, Sign Language Recognition

1 Introduction

Gesture-based communication is a remarkable kind of correspondence that regularly goes understudied. Deaf or speech-disabled people tend to communicate with others using gesture-based communication to convey their contemplations and emotions. Sign language is a little complicated language that involves the movement of hands, face, and different stances of the body. Each nation has its own local sign language [1]. Different sign languages have their own standard of language structure, different gestures for different words, and different pronunciations. Hearing/Speech Impaired individuals are not self-reliant henceforth prepared gesture-based communication translators or consultants are required during clinical, legitimate arrangements and instructional meetings. Though it's good to have access to translators yet it very well can be seen as problematic for various reasons: first and foremost, Hearing/speech impaired individuals face quandaries in expressing themselves, so an interpreter needs to understand their gestures and represent them on their behalf and secondly, the interpreter who is more than the judge of phonetic content acts as a medium between the signers and the non-signers [2]. Translators are masters of depiction but they are not commonly and easily offered all the days, if offered conjointly, they charge tons. The vast majority of ordinary people don't possess any knowledge of sign language. Thus, numerous corresponding hindrances exist for hearing/speech impaired gesture-based communication users.

The number of deaf people has recently reached four hundred million [3], as per the data provided by World Health Organization. Following this cause, the latest experiments are being intensified to make it easier for disabled people to communicate [3]. Though there are several sign languages concerning other sign languages, American Sign Language (ASL) is more spoken. In contrast, we see that ASL which is used in the USA and English-speaking Canada has a vast variety of lingos. There are 22 handshapes relate to the 26 alphabets in order and one can sign the 10 digits on a single hand.

With the wish to accelerate the capacity of hearing/speech deprived people to receive assistance in their own society without the aid of a skilled consultant in sign language. These kinds of virtual machines can address the unmet need for professional decoding facilities and increase the quality of living considerably.

The primary aim of this paper is to propose a two-layer CNN classifier model which can achieve significant accuracy with readily available resources and to provide a comprehensive analysis of different processes, methodologies for the understanding of sign language focused on deep learning and computational approaches. The recent developments in this field are further explored which helps to recognize prospects and obstacles for future research work.

2 Methodology

Decades after Human-Computer Interaction (HCI) became popular in the industry, Sign Language Recognition (SLR) has been explored. The standard and widespread methodology for the identification of vision-based SLR is depicted in Figure 1, which involves user input, pre-processing, extraction of features, classification based on various methodologies of deep learning, and then output predictions. Sign gesture illustration covers the spot to represent the key data. One of the important parts of the whole framework is pre-processing because it involves different stages.

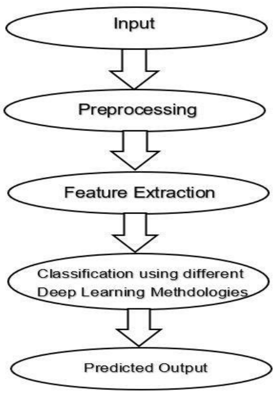

Fig. 1. General flow-diagram for Sign Language Recognition

2.1 Preprocessing

Creation of histogram. The very initial step is hand gesture recognition and for this, a threshold of the image is required. For the creation of the threshold of an image, a histogram is made. Graphical representation of the frequency of pixels in a grayscale image in which pixel value varies from 0 to 255 is known as histogram [5]. To get a good histogram of our hand with respect to surroundings, our hand should cover all those small squares to get a good histogram as shown in Figure 2. The histogram has a significant role in the calculation of the threshold value.

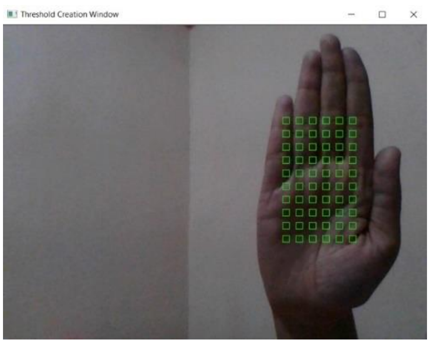

Fig. 2. Hand covering all green spots to create a good histogram

Separation of background and creation of the binary image. To facilitate the process of sign gesture recognition, the hand gesture region must be separated from the background. Hand and background have different pixel intensity hence, differences in their pixel intensity are used to separate them out from each other [6]. Pixels representing the hand region are assigned with a value so that those pixels could be identified. If pixel value exceeds a threshold value, it is assigned with a value that will represent that it is white, else it is assigned any other value which will represent that it is black. To remove our hand from the background, an adaptive threshold method is used and images are scaled to 128 x 128 pixels.

Contour Detection using OpenCV. All the continuous points along the boundary, having the same intensity or color are joined which forms a curve called contour. For shape analysis, object detection, and its recognition, contours proved to be a useful method [7]. Filtered binary images are preferred for better accuracy. Threshold or canny edge detection should be applied before finding contours [7]. Point to note that finding contours in OpenCV is like detecting a white entity on black background. Inbuilt OpenCV functions are used to identify the shape of the hand through the contour detection method. Different contour properties are required for the identification of gestures and to map with their meaning therefore outline is being drawn [6].

Image Labelling. After the creation of grayscale images, those grayscale gesture images are assigned a number that is in sequence and can be considered as a gesture identification number. Later these images are mapped with their corresponding meaning, this process is called image labeling. After this process, the labeled image is saved in the database.

Data Augmentation. While making gestures, restricting users to make gestures from one hand (either left hand or right hand) would not be convenient at all. If they are allowed to do so, it will affect the model accuracy. Therefore, data augmentation is one of the important aspects to get good accuracy. We can define data augmentation as the method by which slightly altered copies of pre-existing data are added to the already existing dataset in order to increase the amount of data which helps the model to learn from a different type of situation. This modification can be done by flipping images along a horizontal or vertical axis or by rotating images by few angles. It behaves as a regulator and helps to bring down the extent of overfitting while training deep learning or a machine learning model [8].

2.2 Feature Extraction

The process to pull out condensed and discriminative descriptors which are further used as training data for the classifier for a stronger recognition purpose and to enhance accuracy is known as feature extraction. One of the difficult issues that are being faced is to pull out the discriminative feature from different kinds of videos or images exposed to different lighting conditions. For the proposed model the Red-Green-Blue (RGB) input images are converted to grayscale images and gaussian blur is applied to minimize unwanted noise.

2.3 Model Training

The selection of a Deep Learning method among the pool of various deep learning methodologies which can give promising results over varied data is another considerable issue. Among various Deep Learning methodologies, Convolutional Neural Network (CNN) gives promising results while making predictions in vision-based SLR. The architecture of CNN which is being followed is shown in Figure 3.

128x128 pixels is the size of the input images. These are processed using 32 filter weights in the first convolutional layer (3x3 pixels each). This produces a 126x126 pixel representation for each of the filter-weights.

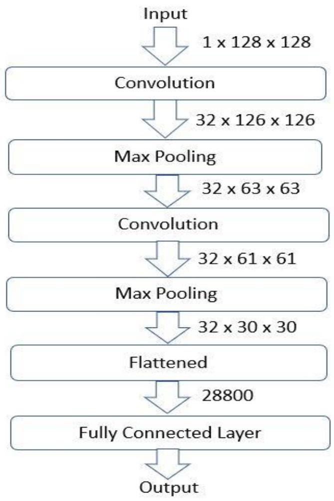

Fig. 3. CNN Architecture

Further, they are down-sampled using max-pooling of 2x2 resulting in images of 63x63 pixels. The input picture was subjected to max pooling with a pool size of (2, 2) and to the Rectified Linear Unit activation function. This decreases the number of parameters, lowering both the computing expense and the risk of overfitting. The resulting 63 x 63 images from the first pooling layer are then served to the second convolutional layer as an input. It is then processed in the second convolutional layer using 32 filter weights (3x3 pixels each). The resulting 60 x 60 pixels images are again down-sampled using a max pool of 2x2 and further reduced to 30x30 resolution of images. These images are fed into a 128-neuron which is a completely connected layer, and the output of the second convolutional layer is then reshaped into a 30x30x32 =28800-value array. This layer receives a 28800-value array as input. The second densely connected layer receives the contribution of all these layers. To prevent overfitting, a dropout layer is used with a value of 0.5. The output from the first densely connected layer is fed into a 96-neuron entirely connected layer. The second densely connected layer's output is fed into the final layer, which would have the same number of neurons as the number of groups were classified (alphabets + blank symbol).

The prediction layer calculates the likelihood of the picture falling into one of the groups. As a result, the output is scaled between 0 and 1, and the sum of each class's values equals 1. We were able to accomplish this by using the Softmax function. The prediction layer's contribution will initially be a little off from the actual value. To improve it, we used labeled data to train the networks. Cross-entropy is the performance measurement that is used in classification.

3 Results and discussions

3.1 Project Outcome

The primary goal was to improve the model's performance with readily available resources while also improving its accuracy. Two layers of the CNN model algorithm were used for this. In the first layer, a threshold was generated using the histogram obtained, and then a gaussian blur filter was used to extract features from the captured hand gesture. The data was then augmented to generate more inputs, allowing the model to identify movements from multiple angles and sides. The processed image input was then transferred to the Layer 1 CNN model, which yielded the same results for a particular set of symbols. The symbols for which the model did not make sufficient predictions are described below.

- For D: U, R
- For U: R, D
- For T: I, K, D
- For S: N, M

Then such sets were categorized using 4 different CNN-classifier designed specifically for each of them in layer 2. The principle behind layer 2 is that if the layer 1 model decides that the input belongs to one of those sets, the input would then proceed through the CNN Classifier in layer 2, which was designed explicitly to distinguish ambiguous symbols.

We obtained 85.4 percent accuracy in our model using just layer 1 of our algorithm, and 95.7 percent accuracy using a mix of layer 1 and layer 2 of our algorithm, which is higher than most existing study papers on American sign language. The confusion matrix of our model is shown below in Figure 4. We used the Adam optimizer to update the model in response to the performance of the loss function. Adam integrates the advantages of extensions of two stochastic gradient descent algorithms which are adaptive gradient algorithm (ADA-GRAD) and root mean square descent algorithm (RMSD).

						P	r	e	d	i	c	t	e	d		V	a	l	u	e	s					
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
	A	147	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	2	0	0
	B	0	139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0
	C	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	E	0	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	135	0	0	0	0	0	4	0	0	0	0	0	0	0	0	3	10	0	0	0
C	G	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
o	H	1	0	0	0	0	0	7	143	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
r	I	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
r	J	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	K	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	L	0	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0
t	M	0	0	0	0	0	0	0	0	0	0	2	0	152	0	0	0	0	0	0	0	0	0	0	0	0
	N	0	0	0	0	0	0	0	0	0	0	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0
V	O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	154	0	0	0	0	0	0	0	0	0	0
a	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0
l	Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	147	1	0	0	0	0	0	0	0
u	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0
e	S	0	0	0	0	1	0	0	0	0	0	0	0	0	0	10	0	0	0	133	0	0	0	0	8	0
s	T	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	151	0	0	0	0	0
	U	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0
	V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151	1	0	0
	W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	149	0	0
	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	148	0
	Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151
	Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 4. Confusion Matrix of the proposed model

3.2 Challenges faced

Initially, a dataset containing raw images of all ASL characters was to be compiled, however, no fitting dataset could be identified in either of the public dataset repositories. As a result, a new dataset was developed. The next issue was to choose the filter for feature extraction. Various filters, such as gaussian blur, binary threshold, canny edge detection, and others, were tested, with the gaussian blur filter proving to be the most efficient. Aside from that, there were problems with the model's accuracy due to lighting conditions; the model did not have promising accuracy when lighting conditions were poor. However, the advances and other techniques in this area are discussed in further sections which have high cost and the resources used in these methodologies are not easily available.

3.3 Object recognition techniques

In the case of single object recognition, for its clear recognition, 2-Dimensional features like Histogram of Oriented Gradient [10], Scale Invariant Feature Transform (SIFT) [9], Kernel descriptors [11] are used. All these methods are good only for single object detection. Therefore, these methods might not perform well when gestures involve multiple parts of the body as a part of the gesture. SIFT [9] can compute at the edges which will be invariant to scaling, rotation, the addition of noise. One of the advantages of depth images is that it is not get affected by illumination changes. Therefore, this advantage gives an edge to reduce the texture and color variability which occurs due to background, skin, and hair. Space-Time Occupancy Pattern (STOP) [12] is proposed to resolve the problems like occlusions and noise in-depth images. Space-Time Occupancy Pattern is focused on characterized space-time patterns of human gestures (which are 4-Dimensional in nature), to utilize the temporal and spatial contextual knowledge and permitting intra-class variations [3].

3.4 Sensor-based technologies

The category in which physical intervention of the user is required includes magnetic, visual, or acoustic sensor-based devices which need to be connected with hands or other body parts to locate their position. Few examples are: information gloves [13], accelerometer [15], digital camera sensor [14]. All these sensor-based technologies vary with different parameters like latency, resolution, accuracy, the vary of motion, price, and user comfort. In this category, individuals are required to wear a bulky gadget and to hold a load of wires or cables that link the device to a computer system. The ease and naturalness of user engagement are impeded by this. Also, when handling such instruments, a lot of precautions and calibration are required. Therefore, this class adds a layer of complexity to the user experience.

On the other hand, there is a category in which physical intervention is not required. This category gets much more appreciation than the previous one when Microsoft Kinect was first introduced. Google Tango Project [18], Microsoft Kinect [16], Leap motion controller [17] provides depth images or maps. As already discussed before, the advantage of depth images is that it does not get affected by illumination changes. Hence, the effect of illumination variability can be decreased to a larger extent. In the past recent years, it has been seen the increment in the studies carried out by researchers focused on human gesture-based recognition by utilising depth information provided by different sensing devices. Google Tango Project[18] is a limited-run experimental device which has a Microsoft Kinect-like vision along with innovative System-on-Chip (SoC) integrated inside. It has also a dedicated and very subtle 3D scanner. All these recent advances would make methods based on depth maps even more convenient and useful for signers [3].

3.5 Different sign languages across the globe

There is no specific standard for sign languages. Diverse gesture-based communications are utilized in numerous nations or locales. Table 1 shows the most commonly used sign languages and their respective geographical region in which they are being spoken.

3.6 Comparison of various deep learning methodologies

Deep neural network (DNN) based methodologies can take in reasonable contextual information from crude pictures or video fragments straightway, which is sturdy and adaptable. The comparison between different Deep Learning methodologies based on their accuracy is shown in Table 2.

Table 1. Different sign languages and their respective countries/regions

Sign Languages	Respective Countries/Regions	Reference
British Sign Language (BSL)	United Kingdom	[19]
British, Australian, and New Zealand Sign Language (BANZSL)	United Kingdom, Australia, New Zealand	[19]
French Sign Language (LSF)	France and French-speaking areas of Switzerland	[20]
American Sign Language (ASL)	United States, Canada, West Africa, and Southeast Asia	[21]
Irish Sign Language (ISL)	Ireland	[19]
Chinese Sign Language (CSL or ZGS)	China	[19]
Brazilian Sign Language (Libras)	Brazil	[19]
Indian Sign Language	India	[22]
Thai Sign Language	Thailand	[23]
German Sign Language (DGS)	Germany	[24]
Arabic Sign Language	Arab Middle East	[25]

Table 2. Deep Learning Methodologies and their accuracy

Various Deep Learning Methodologies	Accuracy (%)	Reference
Motion Sensor Information + SVM k-NN	79.83	[26]
Edge Orientation Histogram	88.26	[27]
DBN (3 RBM layers + 1 translation layer)	79	[28]
CNN	82	[29]
DNN + CRF	92	[30]
HMM + BLSTM-NN	96.2	[31]
PCA + HMM	89.10	[32]
Random Forest + Depth Comparison	90	[33]
Full Dimensional Feature + k-NN	99.6	[34]
SURF + SVM	97.13	[35]
AdaBoost + HAAR	98.7	[36]
GRNN + Ring Projection and Wavelet Transform	90.44	[37]
CNN + Convex Hull + Skin Detection	98.05	[38]
Edge Detection + Cross Correlation	94	[39]
Hybrid CNN-HMM	92.6	[40]

4 Conclusion and Future Scope

On the whole, this paper provides the different advances in sign language recognition. Experts in different disciplines have primarily performed past studies on sign language comprehension for numerous sign languages, reducing real-world utility. More than 25 studies were reviewed to get the knowledge of the past researches and then the different works based on sign language recognition using various Deep Learning methodologies and other technological approaches are discussed. This paper also highlights the various terms and processes involved in this area.

From the various studies, it can be depicted that there is no universal sign language. Sign language changes with concerning countries and regions. Therefore, when a deaf or speech-impaired person travels to another country or region, he/she might face the problem because the device in which SLR is integrated might not have the sign language of the region he/she belongs to. Also, Deep Neural Network (DNN) shows promising performance in various practical aspects. But It can learn good features from multi-modal histogram because these histograms can enhance the accuracy of recognition by providing more knowledge or features about sign gestures. So, further studies should be concerned about integrating all the sign languages used across the globe into one device with a convenient and customizable user interface to make it more robust and barrier breaker. In conjunction, in future research work, studying high-level functionality across deeper networks remains questionable to see the potential in terms of usability and accuracy development.

Conclusively, everyone is of value in society and has the right to communicate with others regardless of their physical impairments contemplating this, let us try to embrace hearing and speech impaired people in our everyday life and fostering them ahead in their life by incorporating innovation with their exigencies and other needs.

References

1. M. M. Islam, S. Siddiqua, and J. Afnan, "Real time Hand Gesture Recognition using different algorithms based on American Sign Language," *2017 IEEE Int. Conf. Imaging, Vis. Pattern Recognition, icIVPR 2017*, 2017, doi: 10.1109/ICIVPR.2017.7890854.
2. A. Young, R. Oram, and J. Napier, "Hearing people perceiving deaf people through sign language interpreters at work: on the loss of self through interpreted communication," *J. Appl. Commun. Res.*, vol. 47, no. 1, pp. 90–110, 2019, doi: 10.1080/00909882.2019.1574018.
3. L. Zheng, B. Liang, and A. Jiang, "Recent Advances of Deep Learning for Sign Language Recognition," *DICTA 2017 - 2017 Int. Conf. Digit. Image Comput. Tech. Appl.*, vol. 2017-Decem, pp. 1–7, 2017, doi: 10.1109/DICTA.2017.8227483.
4. "The 26 letters and 10 digits of American Sign Language (ASL). | Download Scientific Diagram." https://www.researchgate.net/figure/The-26-letters-and-10-digits-of-American-Sign-Language-ASL_fig1_328396430 (accessed Feb. 22, 2021).
5. https://www.geeksforgeeks.org/opencv-python-program-analyze-image-using-histogram/ (accessed Feb. 22, 2021).
6. K. Manikandan, A. Patidar, P. Walia, and A. B. Roy, "Hand Gesture Detection and Conversion to Speech and Text," *arXiv*, 2018.

7. “Contours - OpenCv” https://docs.opencv.org/3.4/d4/d73/tutorial_py_contours_begin.html (accessed Feb. 22, 2021).
8. “Data Augmentation | How to use Deep Learning when you have Limited Data — Part 2” https://nanonets.com/blog/data-augmentation-how-to-use-deep-learning-when-you-have-limited-data-part-2/ (accessed Feb. 22, 2021).
9. D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004, doi: 10.1023/B:VISI.0000029664.99615.94.
10. R. Rai, S. Shukla, and B. Singh, “Reactive Power Based MRAS for Speed Estimation of Solar Fed Induction Motor with Improved Feedback Linearization for Water Pumping,” *IEEE Trans. Ind. Informatics*, vol. 16, no. 7, pp. 4714–4725, 2020, doi: 10.1109/TII.2019.2950094.
11. L. Bo, K. Lai, X. Ren, and D. Fox, “Object recognition with hierarchical kernel descriptors,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1729–1736, 2011, doi: 10.1109/CVPR.2011.5995719.
12. A. W. Vieira, E. R. Nascimento, G. L. Oliveira, Z. Liu, and M. F. M. Campos, “STOP: Space-Time Occupancy Patterns for 3D action recognition from depth map sequences,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7441 LNCS, pp. 252–259, 2012, doi: 10.1007/978-3-642-33275-3_31.
13. M. Mohandes and S. O. Aliyu, “(12) United States Patent,” vol. 2, no. 12, 2017.
14. H. Hongo, M. Ohya, M. Yasumoto, Y. Niwa, and K. Yamamoto, “Focus of attention for face and hand gesture recognition using multiple cameras,” *Proc. - 4th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2000*, pp. 156–161, 2000, doi: 10.1109/AFGR.2000.840627.
15. X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, “A framework for hand gesture recognition based on accelerometer and EMG sensors,” *IEEE Trans. Syst. Man, Cybern. Part ASystems Humans*, vol. 41, no. 6, pp. 1064–1076, 2011, doi: 10.1109/TSMCA.2011.2116004.
16. K. Lai, J. Konrad, and P. Ishwar, “A gesture-driven computer interface using Kinect,” *Proc. IEEE Southwest Symp. Image Anal. Interpret.*, pp. 185–188, 2012, doi: 10.1109/SSIAI.2012.6202484.
17. C. H. Chuan, E. Regina, and C. Guardino, “American sign language recognition using leap motion sensor,” *Proc. - 2014 13th Int. Conf. Mach. Learn. Appl. ICMLA 2014*, pp. 541–544, 2014, doi: 10.1109/ICMLA.2014.110.
18. https://en.wikipedia.org/wiki/Tango_(platform) (accessed Feb. 22, 2021).
19. “A Guide to the Different Types of Sign Language Around the World.” https://k-international.com/blog/different-types-of-sign-language-around-the-world/ (accessed Feb. 22, 2021).
20. “French Sign Language” https://www.tradonline.fr/en/blog/french-sign-language-a-language-in-its-own-right/ (accessed Feb. 22, 2021).
21. “ASL - Wikipedia” https://en.wikipedia.org/wiki/American_Sign_Language (accessed Feb. 22, 2021).
22. “Welcome to Indian Sign Language Portal” https://indiansignlanguage.org/ (accessed Feb. 22, 2021).
23. “Thai sign-language - Wikipedia” https://en.wikipedia.org/wiki/Thai_Sign_Language (accessed Feb. 22, 2021).
24. “German sign-language-Wikipedia” https://en.wikipedia.org/wiki/German_Sign_Language (accessed Feb. 22, 2021).
25. “Arab sign-language family - Wikipedia.” https://en.wikipedia.org/wiki/Arab_sign-language_family (accessed Feb. 22, 2021).

26. A. Vaitkevičius, M. Taroza, T. Blažauskas, R. Damaševičius, R. Maskeliunas, and M. Woźniak, "Recognition of American Sign Language gestures in a Virtual Reality using Leap Motion," *Appl. Sci.*, vol. 9, no. 3, pp. 1–16, 2019, doi: 10.3390/app9030445.
27. J. R. Pansare and M. Ingle, "Vision-based approach for American Sign Language recognition using Edge Orientation Histogram," *2016 Int. Conf. Image, Vis. Comput. ICIVC 2016*, pp. 86–90, 2016, doi: 10.1109/ICIVC.2016.7571278.
28. L. Rioux-Maldague and P. Giguere, "Sign language fingerspelling classification from depth and color images using a deep belief network," *Proc. - Conf. Comput. Robot Vision, CRV 2014*, no. May 2014, pp. 92–97, 2014, doi: 10.1109/CRV.2014.20.
29. S. Ameen and S. Vadera, "A convolutional neural network to classify American Sign Language fingerspelling from depth and colour images," *Expert Syst.*, vol. 34, no. 3, 2017, doi: 10.1111/exsy.12197.
30. T. Kim *et al.*, "Lexicon-free fingerspelling recognition from video: Data, models, and signer adaptation," *Comput. Speech Lang.*, vol. 46, pp. 209–232, 2017, doi: 10.1016/j.csl.2017.05.009.
31. P. Kumar, H. Gauba, P. Pratim Roy, and D. Prosad Dogra, "A multimodal framework for sensor based sign language recognition," *Neurocomputing*, vol. 259, no. 2017, pp. 21–38, 2017, doi: 10.1016/j.neucom.2016.08.132.
32. M. M. Zaki and S. I. Shaheen, "Sign language recognition using a combination of new vision based features," *Pattern Recognit. Lett.*, vol. 32, no. 4, pp. 572–577, 2011, doi: 10.1016/j.patrec.2010.11.013.
33. C. Dong, M. C. Leu, and Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2015-Octob, pp. 44–52, 2015, doi: 10.1109/CVPRW.2015.7301347.
34. D. Aryanie and Y. Heryadi, "American sign language-based finger-spelling recognition using k-Nearest Neighbors classifier," *2015 3rd Int. Conf. Inf. Commun. Technol. ICoICT 2015*, pp. 533–536, 2015, doi: 10.1109/ICoICT.2015.7231481.
35. C. M. Jin, Z. Omar, and M. H. Jaward, "A mobile application of American sign language translation via image processing algorithms," *Proc. - 2016 IEEE Reg. 10 Symp. TENSYMP 2016*, pp. 104–109, 2016, doi: 10.1109/TENCONSpring.2016.7519386.
36. V. N. T. Truong, C. K. Yang, and Q. V. Tran, "A translator for American sign language to text and speech," *2016 IEEE 5th Glob. Conf. Consum. Electron. GCCE 2016*, 2016, doi: 10.1109/GCCE.2016.7800427.
37. Y. Fiagbe, "World Journal of Engineering," no. April, 2020.
38. M. Taskiran, M. Killioglu, and N. Kahraman, "A Real-Time System for Recognition of American Sign Language by using Deep Learning," *2018 41st Int. Conf. Telecommun. Signal Process. TSP 2018*, pp. 1–5, 2018, doi: 10.1109/TSP.2018.8441304.
39. A. Joshi, H. Sierra, and E. Arzuaga, "American sign language translation using edge detection and cross correlation," *2017 IEEE Colomb. Conf. Commun. Comput. COLCOM 2017 - Proc.*, 2017, doi: 10.1109/ColComCon.2017.8088212.
40. O. Koller, S. Zargaran, H. Ney, and R. Bowden, "Deep sign: Hybrid CNN-HMM for continuous sign language recognition," *Br. Mach. Vis. Conf. 2016, BMVC 2016*, vol. 2016-Septe, no. August, pp. 136.1-136.12, 2016, doi: 10.5244/C.30.136.