

Live Sign Language Translation: A Survey

Aditya Dawda¹, Aditya Devchakke², Avdhoot Durgude³

¹Student, Dept. of Computer Science and Engineering, MIT School of Engineering, Maharashtra, India

²Student, Dept. of Computer Science and Engineering, MIT School of Engineering, Maharashtra, India

³Student, Dept. of Computer Science and Engineering, MIT School of Engineering, Maharashtra, India

Abstract - Languages are the basic need for communication, sharing of knowledge, emotions and thoughts. As language is a verbal, persons having hearing or speaking disabilities use various gestures to communicate with others. Thus, it is known as 'sign language', it has its own patterns and variables. It always becomes hard for a person who uses sign language for communication with others. Additionally, there exists various sign languages, mainly American Sign Language (ASL), British, Australian and New Zealand Sign Language (BANZSL), Chinese Sign Language (CSL), French Sign Language (LSF) and Japanese Sign Language (JSL). By using concepts of Human Computer Interaction (HCI) LSTM (Long Short Term Memory) networks were studied and implemented for classification of gesture data because of their ability to learn long-term dependencies. This paper presents the recommendation system/model to translate facial and hand gestures of Sign Languages into English Language. Most of the search was based on CNN or variants of CNN (Convolutional N), SVM, KNN and TensorFlow Library based models and images of gestures in black and gray form were feeded to these models for prediction of sign language. No facial or body pose were used in the search, so our study concluded most of the models were static in nature as they process only one frame and only one body part.

Key Words: American sign language, Computer Vision, Convolutional Neural Network, Deep Learning Model, Gesture recognition, LSTM, Media pipe Holistic.

1. INTRODUCTION

This paper studies the problem of vision-based Sign Language Translation (SLT), which bridges the communication gap between the deaf mute and normal people. It is related to several video understanding topics that targets to interpret video into understandable text and language. Sign Language is a gesture language which visually transmits sign patterns using hand-shapes, orientation and movements of the hands, arms or body, facial expressions and lip-patterns to convey word meanings instead of acoustic sound patterns. Different sign languages exist around the world, each with its own

vocabulary and gestures. A lot of research has been done with respect to Sign Language Translation. Some examples are ASL (American Sign Language) [1], GSL (German Sign Language), BSL (British Sign Language), and so on. There are approximately 6000 gestures of ASL for common words, using fingers to communicate obscure words or proper nouns. This language is commonly used in deaf communities, including interpreters, friends, and families of the deaf, as well as people who are hard of hearing themselves. However, these languages are not commonly known outside of these communities, and therefore communication barriers exist between deaf and hearing people. Hand gestures are nonverbal communication methods for these people. It becomes very difficult and time consuming to exchange the information or feelings for a person who uses sign language to communicate with other non-users. One can make use of devices which translate sign language into words, but this is not an effective way as it can have huge costs and need maintenance. The main aim of our research is to form an effective way of communication between the people using sign language and the English language.

Recognition varies from researcher to researcher based on the method they use to conduct their research and the model they built their application on [6]. This paper can help other researchers to have an idea of the methods used previously; they get to know about the pros of the earlier research and cons which previous research could not overcome. Further they can make use of this data to make changes in their system for achieving higher accuracy, also they will get an idea to make a plan for building models, making complex calculations in various ways in order to get the desired results.

The problem of developing sign language recognition ranges from data collection to building an effective model for image recognition. Using a camera can be ineffective as there are many noises associated with it which leads to a lot of image pre-processing while using a sensor-based approach can be costly and not feasible for every researcher. There are some classifiers which classify

alphabets of English but no phrases translation is done by them and vice-versa.

2. LITERATURE SURVEY

Purpose of reviewing literature is to find out different approaches and methods applied by the authors for translating sign language into English language. Various models are studied for image processing and image classification.

Following table presents the research of different authors in this field:

Table -1.1: Literature paper-1 [1]

Title	American Sign Language Recognition System: An Optimal Approach.
Author	Shivashankara S, Srinath S [1].
Publication	International Journal of Image, Graphics and Signal Processing, Aug 2018
Result/ Accuracy	ASL alphabets A-Z (26) and numbers 0-9 (10) are recognized with 93.05% accuracy.

Table -1.2: Literature paper-2 [2]

Title	Vision-based Portuguese Sign Language Recognition System.
Author	Paulo Trigueiros, Fernando Ribeiro, Luís Paulo Reis [2]
Publication	Springer Science and Business Media LLC, April 2014.
Result/ Accuracy	Sixteen models were compared based on precision, recall, accuracy, activation function, and AROC. The variance in most models was analogous. Accuracy: 74%, Precision: 78%, Recall: 68%.

Table -1.3: Literature paper-3 [3]

Title	A Comprehensive Analysis on Sign Language Recognition System
Author	Rajesh George Rajan, M Judith Leo [3]
Publication	International Journal of Recent Technology and Engineering (IJRTE), March 2019
Result/ Accuracy	Different Acquisition methods like SVM, HMM, SVM+ANN, PCA+KNN, MLP+MDC, Polynomial classifier and ANFIS for sign language recognition.

Table -1.4: Literature paper-4 [4]

Title	Intelligent Hand Cricket.
Author	Aditya Dawda, Aditya Devchakke[4]
Publication	Cyber Intelligence and Information Retrieval 2021, Springer
Result/ Accuracy	CNN based hand gesture recognition for playing hand cricket games with 97.76% training and 95.38% validation accuracy.

Table -1.5: Literature paper-5 [5]

Title	Conversion of sign language into text
Author	Mahesh Kumar N B [5]
Publication	International Journal of Applied Engineering Research (2018)
Result/ Accuracy	Recognizing Indian sign language with MATLAB, total of 10 images of each alphabet has been taken. It used Linear Discriminant Analysis (LDA), because LDA reduces dimensionality ultimately reducing noise, and achieving higher accuracy.

Table -1.6: Literature paper-6 [6]

Title	Hierarchical LSTM for Sign Language Translation
Author	Dan Guo, Wengang Zhou, Houqiang Li, Meng Wang [6]
Publication	The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)
Result/ Accuracy	Hierarchical-LSTM framework for sign language translation, which builds a high-level visual semantic embedding model for SLT. It explores visemes via online variable-length key clip mining and attention-aware weighting..
Conclusion	This study presents three distinct ways that enable the choice of selecting a solution in specific situations. K-Means, DBSCAN, and Hierarchical clustering all have certain types of drawbacks that render them unsuitable when used solely.

Table -1.7: Literature paper-7 [7]

Title	Sign language recognition system using CNN and Computer Vision
Author	MehreenHurroo, Mohammad ElhamWalizad [7]
Publication	International Journal of Engineering Research & Technology (IJERT), December 2020
Result/ Accuracy	CNN based model for classifying ASL with 10 alphabets having accuracy above 90% (with gloves), HSV color algorithm for detecting gesture. The dataset is self-created with 80:20 split

Table -1.8: Literature paper-8 [8]

Title	Improving Sign Language Translation with Monolingual Data by Sign Back-Translation
Author	Hao Zhou1 Wengang Zhou1, Weizhen Qi1 Junfu Pu1 Houqiang Li1[8]
Publication	CVPR 2021
Result/ Accuracy	In this paper they propose to improve the translation quality with monolingual data, which is rarely investigated in SLT. By designing a Sign BT pipeline, they converted massive spoken language texts into source sign sequences. The synthetic pairs are treated as additional training data to alleviate the shortage of parallel data in training.

Table -1.9: Literature paper-9 [9]

Title	Research of a sign language translation system based on deep learning
Author	Siming He [9]
Publication	Conference on artificial intelligence and advanced manufacturing ,2019
Result/ Accuracy	Faster R-CNN with an embedded RPN module is used with accuracy of 91.7%, he used 3D CNN for feature extraction and sign language recognition framework consisting of LSTM

Table -1.10: Literature paper-10 [10]

Title	Sign Language Translator
Author	Divyanshu Mishra, MedhaviTyagi, and AnkurVerma, Gaurav Dubey[10]
Publication	International Journal of Advanced Science and Technology (2020)
Result/	The classification model provides an

Accuracy	accuracy of 93% on the validation set and the translation of gestures to the English language is happening smoothly frame by frame. However, the system working on a computer is not flexible as the user has to carry the computer wherever he or she goes.
----------	--

Table -1.11: Literature paper-11 [11]

Title	Multi-Modality American Sign Language Recognition
Author	ChenyangZhang,YingliTian,Matt Huenerfauth [11]
Publication	Rochester Institute of Technology RIT Scholar Works,9-2016
Result/ Accuracy	This learns from labelled ASL videos captured by a Kinect depth sensor and predicts ASL components in new input videos, words are divided into 4 groups according to their frequencies. The combined accuracy is 36.07% with 99 lexical items

Table -1.12: Literature paper-12 [12]

Title	Real-Time American Sign Language Recognition with Faster Regional Convolutional Neural Networks
Author	S. Dinesh,S. Sivaprakash [12]
Publication	International Journal of Innovative Research in Science, Engineering and Technology, March 2018
Result/ Accuracy	They are able to interpret isolated hand gestures from the Argentinian Sign Language (LSA) with r-CNN model having accuracy of 93.33%

Table -1.13: Literature paper-13 [13]

Title	Recognition of Single-Handed Sign Language Gestures using Contour Tracing Descriptor
Author	Rohit Sharma, YashNemani, Sumit Kumar [13]
Publication	Proceedings of the World Congress on Engineering 2013 Vol II, WCE 2013, July 3 - 5, 2013, London, U.K.
Result/ Accuracy	They use classifiers (Support Vector Machines and k-Nearest Neighbors) to characterize each color channel after background subtraction. The change is using a contour trace, which is an efficient representation of hand contours. The accuracy of 62.3% is achieved using an SVM on the segmented color channel model.

Table -1.14: Literature paper-14 [14]

Title	Hand gesture recognition based on dynamic Bayesian network framework
Author	H suk,bongkee sin [14]
Publication	Korea university 2010
Result/ Accuracy	Dynamic Bayesian network-based inference is.10 isolated gestures, they attempt to classify moving hand gestures having 99% accuracy

Table -1.15: Literature paper-15 [15]

Title	SIGN LANGUAGE RECOGNITION: STATE OF THE ART
Author	Ashok K Sahoo, GouriSankar Mishra and Kiran Kumar Ravulakollu[15]
Publication	ARPN Journal of Engineering and Applied Sciences, VOL. 9, NO. 2, FEBRUARY 2014
Result/ Accuracy	Lifepoint Fingerspell Library data set is

Accuracy	used with 87.64 accuracy
----------	--------------------------

Table -1.16: Literature paper-16 [16]

Title	Sign Language Translation Using Deep Convolutional Neural Networks
Author	RahibH.Abiyev , Murat Arslan and John Bush Idoko [16]
Publication	KSII Transactions on Internet and Information Systems VOL. 14, NO. 2, Feb. 2020
Result/ Accuracy	Model is divided into three parts (SSD, Inception v3 and SVM) which are then integrated in one model with 99.90 % accuracy.

Table -1.17: Literature paper-17 [17]

Title	Sign Language Recognition using Convolutional Neural Networks
Author	Lionel Pigou, Sander Dieleman [17]
Publication	Ghent University, ELIS, Belgium,2015
Result/ Accuracy	Aims to classify 20 Italian gestures from the ChaLearn 2014 using Microsoft Kinect with accuracy of 91.7

Table -1.18: Literature paper-18 [18]

Title	Indian Sign Language Recognition System
Author	Yogeshwar I. Rokade, Prashant M. Jadav [18]

Publication	ResearchGate 2017
Result/ Accuracy	In this paper they worked on Indian sign language using ANN and it also supports vector machines. For feature extraction, central moments and HU moments are used, Artificial Neural Networks are used to classify the sign which gives average accuracy of 94% and SVM gives accuracy of 92%.

Table -1.19: Literature paper-19 [19]

Title	Sign language Recognition Using Machine Learning Algorithm
Author	Prof.Radha S. Shirbhate, Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar,Ms. Mayuri A. Khandge[19]
Publication	International Research Journal of Engineering and Technology (IRJET)2020
Result/ Accuracy	In this paper they have gone through an automatic sign language gesture recognition system in real-time, using different tools. Also, they proposed work expected to recognized the sign language and convert it into the text.

Table -1.20: Literature paper-20 [20]

Title	A Survey on Sign Language Recognition Systems
Author	Shruty M. Tomar, Dr.Narendra M. Patel,Dr. Darshak G. Thakore [20]
Publication	International Journal of Creative research Thoughts (IJCRT) 2021

Result/ Accuracy	In This they Study various classification techniques concludes that deep neural network (CNN, Inception model, LSTM) performs better than traditional classifiers such as KNN and SVM.
---------------------	--

Table -1.21: Literature paper-21 [21]

Title	Sign language recognition system based on prediction in Human-Computer Interaction
Author	Maher Jebali, Patrice Dalle, and Mohamed Jemni[21]
Publication	ResearchGate 2014
Result/ Accuracy	The integrated framework is used for head and hand tracking in videos of SL, it is in fact employed in prediction based SLR. Con-cerning the detection of different components,

3. OBSERVATIONS:

As the above table explains there are various methods/models to detect the sign language [3]. One thing which is frequently observed was the quantity and quality of data set used to train the model [15]. When the dataset size is increased the person reaches more to the desired accuracy. Quality of data set affects the model exponentially, where there is dataset made by web cam, data set of any higher resolution, data augmentation, data set with continuous videos changes the accuracy.

Out of the 21 papers reviewed most of the papers are based on CNN or CNN integrated models for image classification. Various approaches like convolutional neural network (CNN), Region based convolutional neural network(r-CNN) [12], K-nearest Neighbors (KNN) [13], Dynamic Bayesian network (DBN) [14], Support vector machines (SVM) [13][18], Matlab with Linear discriminant analysis (LDA) [5].

LSTM based model to capture the important features and discard unwanted features based on past features is used by the author of this paper to bridge the gap between the research as conventional methods process each frame independently and not based on the information generated from past steps.

The accuracy of 62.3% is achieved using an SVM on the segmented color channel model [13]. Artificial Neural Networks are used to classify the sign which gives average accuracy of 94% and SVM gives accuracy of 92% [18].

4. CONCLUSION:

Normally, for sign language translation, various projects have used CNN model, some of them also tried with Linear discriminant analysis (LDA) and achieved the output[5]. Here we have proposed a LSTM framework for sign language translation which works on long short-term memory concept. This model removes the unwanted features and remembers features which are important and updates features for further recognition. Thus, building a comprehensive sign language recognizer through LSTM instead of CNN, for phrases and alphabet detection, data has been gathered for processing using web cameras. Further various data augmentation techniques could be used to get the model working under numerous situations and achieving higher accuracy.

REFERENCES:

1. SsShivashankara&Dr.S.Srinath. (2018).” American Sign Language Recognition System: An Optimal Approach”. International Journal of Image, Graphics and Signal Processing. 10. 10.5815/ijigsp.2018.08.03.
2. Trigueiros Paulo, Ribeiro Fernando, Reis Luís. (2014). Vision-Based Portuguese Sign Language Recognition System. Advances in Intelligent Systems and Computing. 275. 10.1007/978-3-319-05951-8_57.
3. Rajan, Rajesh & Leo, Judith. (2019). “A comprehensive analysis on sign language recognition system”.
4. KingeAnuj, Kulkarni Nilima, Devchakke Aditya, Dawda Aditya, Mukhopadhyay Ankit. (2022).” Intelligent Hand Cricket”. 10.1007/978-981-16-4284-5_33.
5. Mahesh Kumar N B. “Conversion of Sign Language into Text”. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018) pp. 7154-7161.

6. Dan Guo, Wengang Zhou, Houqiang Li, Meng Wang. "Hierarchical LSTM for Sign Language Translation". The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18), Vol. 32 No. 1 (2018)
7. Mehreen Hurroo, Mohammad Elham, 2020, "Sign Language Recognition System using Convolutional Neural Network and Computer Vision". INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 09, Issue 12 (December 2020).
8. Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, Houqiang Li: Improving Sign Language Translation with Monolingual Data by Sign Back-Translation. CoRR abs/2105.12397 (2021)
9. S. He, "Research of a Sign Language Translation System Based on Deep Learning," 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM), 2019, pp. 392-396, doi: 10.1109/AIAM48774.2019.00083.
10. D. M. M. T. Ankur Verma, Gaurav Dubey, "Sign Language Translator", IJAST, vol. 29, no. 5s, pp. 246 - 253, Mar. 2020.
11. C. Zhang, Y. Tian and M. Huenerfauth, "Multi-modality American Sign Language recognition," 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 2881-2885, doi: 10.1109/ICIP.2016.7532886.
12. S. Dinesh, S. Sivaprakash, Real-Time American Sign Language Recognition with Faster Regional Convolutional Neural Networks, international Journal of Innovative Research in Science, Engineering and Technology, March 2018
13. Sharma, R., Nemani, Y., Kumar, S., Kane, L., & Khanna, P. Recognition of Single-Handed Sign Language Gestures using Contour Tracing Descriptor.
14. Suk, H. I., Sin, B. K., & Lee, S. W. (2010). Hand gesture recognition based on dynamic Bayesian network framework. Pattern Recognition, 43(9), 3059-3072.
<https://doi.org/10.1016/j.patcog.2010.03.016>
15. Sahoo, Ashok & Mishra, Gouri & Ravula Kollu, Kiran. (2014). Sign language recognition: State of the art. ARPN Journal of Engineering and Applied Sciences. 9. 116-134.
16. "Sign Language Translation Using Deep Convolutional Neural Networks," KSII Transactions on Internet and Information Systems, vol. 14, no. 2. Korean Society for Internet Information (KSII), 29-Feb-2020.
17. L. Pigou, S. Dieleman, P.-J. Kinderman's, and B. Schrauwen, "Sign language recognition using convolutional neural networks," in Lecture Notes in Computer Science, Zürich, Switzerland, 2015, pp. 572-578.
18. Rokade, Yogeshwar & Jadav, Prashant. (2017). Indian Sign Language Recognition System. International Journal of Engineering and Technology. 9. 189-196.
10.21817/ijet/2017/v9i3/170903S030.
19. Shirbhate, Radha S., Mr. Vedant D. Shinde, Ms. Sanam A. Metkari, Ms. Pooja U. Borkar and Ms. Mayuri A. Khandge. "Sign language Recognition Using Machine Learning Algorithm." Volume: 07 Issue: 03 Mar 2020
20. 1Shruty M. Tomar, 2Dr. Narendra M. Patel, 3Dr. Darshak G. Thakore. A Survey on Sign Language Recognition Systems. IJCRT, Volume 9, Issue 3 March 2021.
21. Jebali, Maher & Dalle, Patrice & Jemni, Mohamed. (2014). Sign Language Recognition System Based on Prediction in Human-Computer Interaction. Communications in Computer and Information Science. 435. 565-570. 10.1007/978-3-319-07854-0_98.

BIOGRAPHIES:

Aditya Dawda is from Thane city, Maharashtra, India. He is currently pursuing a B.Tech In Computer Science from MIT-ADT University, School of Engineering. He is one of the authors of the paper 'Intelligent Hand Cricket' that won the Best Paper Award at CIIR 2021(a Springer conference). He developed projects like Intelligent Hand Cricket, Energy Demand Series, Sign Language Translation. He is interested in MERN stack development and Artificial Intelligence.





Aditya Devchakke is from Pune City, Maharashtra, India. He is currently pursuing B.Tech from MIT-ADT University, School of Engineering, Pune, India. He has won the Best Paper Award at the Springer conference(CIIR 21). He has worked on projects such as Energy price prediction, mask detection, hand cricket game. He is interested in WebDevelopment, Data Analysis, Machine Learning and Deep Learning.



Avdhoot Durgude is from Pune City, Maharashtra, India. He is currently pursuing B.Tech from MIT-ADT University, School of Engineering, Pune, India. He has worked on projects such as Driver drowsiness detection system, Image segmentation. He is interested in Data Analysis, Machine Learning and Deep Learning.