An AI-Based Framework for Translating American Sign Language

Subjects: Computer Science, Artificial Intelligence

Contributor: Vijayendra D. Avina , Md Amiruzzaman , Stefanie Amiruzzaman , Linh B. Ngo , M. Ali Akber Dewan

Communication is an essential part of life, without which life would be very difficult. Each living being in the world communicates in their own way. American Sign Language (ASL) is a sign language used by deaf and hearing impaired people in the United States and Canada, devised in part by Thomas Hopkins Gallaudet and Laurent Clerc based on sign language in France.

ASL   deep learning   image   translation   video

1. Introduction

Communication is an essential part of life, without which life would be very difficult. Each living being in the world communicates in their own way. We, as human beings, usually communicate by speaking a language. However, there is an exception; for example, people with deaf symptoms and and hearing impairment. They use signs to communicate among themselves (i.e., deaf to deaf or deaf to impaired hearing). Over a period of time, these signs became a language. Just like all other languages, American Sign Language (ASL) also has its own syntax and semantics [1][2]. One must follow its syntax and semantics to communicate correctly and efficiently. Also, for communication to be successful, it is important to understand what is being communicated. Most people who do not have these disabilities are not aware of these signs, how to use them, or the meaning of the different signs. This could be because of lack of knowledge. As a result, they struggle to communicate with deaf and hearing-impaired people.

ASL has its own grammar and culture, which differ from place to place. Hence, there are many versions of sign language available in the world. French Sign Language (LSF), British Sign Language (BSL), and ASL are a few of the well-known sign languages. Based on the location, different signs are used to say different words. Therefore, it is also important to understand which signs to use in which area.

There are some existing studies that focus on identifying this mapping. People in the past have developed wearable devices that help to identify ASL signs [3][4]. Also, using the Convolution Neural Network (CNN) model and deep learning methodologies, there are a few research studies exploring ways to identify ASL. Most of these studies mainly focused on identifying fingerspelling, which is nothing but identifying signs for the English alphabet [5][6][7]. However, ASL has a vast variety of signs for different words, and little work has been conducted on the identification of these signs.

2. An AI-Based Framework for Translating American Sign Language to English and Vice Versa

Over the past several years, a good number of research studies has been conducted on interpreting ASL. Thad Starner et al. proposed sign language recognition based on Hidden Markov Models (HMM) [5]. This study used a camera to track hand movement to identify hand gestures. They extracted the features from the hand movements and fed them into four-state HMM to identify the ASL words in sentences. They evaluated their work by using a webcam or desk-mounted camera (second-person view) and a wearable camera (first-person view) for a 40-word lexicon. Similarly, Gaus and Wong [8] used two real-time hidden Markov model-based systems that were used to recognize ASL sentences by using a camera to track the user's hands. The authors used word lexicon, and in their system they used a desk-mounted camera to observe the user's hands.

In [6], Qutaishat et al. proposed a method that does not require any wearable gloves or virtual markings to identify ASL. Their process is divided into two phases—feature extraction and classification. At the feature-extraction phase, features are extracted using Hough transformation from the the input images. These features are then passed as input to the neural network classification model. Their work was mainly focused on recognizing static signs. Several studies, such as [7][9][10][11] used the CNN model to classify ASL alphabets. In a separate study, Garcia et al. [7] used the transfer learning concept and developed the model using the Berkeley version of GoogLeNet. Most of these works concentrated on recognizing the ASL fingerspelling corresponding to the English alphabet and numbers [5][6][10]. Furthermore, Rahman et al. [9] used a CNN model to recognize ASL alphabets and numerals. Using a publicly available dataset, their study mainly focused on improving the performance of the CNN model. The study did not involve any human interaction to assess the accuracy of the approach. A similar work was found in [12], where the authors used an ensemble classification technique to show performance improvement. In a separate study, Kasapbasi et al. [13] used a CNN model to predict American Sign Language Alphabets (ASLA), and Bellen et al. [14] focused on recognizing ASL-based gestures during video conferencing.

In a study, Ye et al. [15] used a 3D recurrent convolutional neural network (3DRCNN) to recognize ASL signs from continuous videos. Moreover, they used a fully connected recurrent neural network (FC-RNN), which captured the temporal information. The authors were able to recognise ASL alphabets and several ASL words. In [10][15], the authors used 3D-CNN models to classify ASL. In [10], authors developed a 3D-CNN architecture which consists of eight layers. They used multiple feature maps as inputs for better performance. The five features which they considered are color-R, color-G, color-B, depth, and body skeleton. They were able to achieve better prediction percentages compared to the GMM-HMM model. In [6], Munib et al. used images of signers' bare hands (in a natural way). Their goal was to develop an automatic translation system for ASL alphabets and signs. This study used Hough transform and neural network to recognize the ASL signs.

In [15], authors proposed a hybrid model, and it consisted of the 3D-CNN model and the Fully Connected Recurrent Neural Network (FC-RNN). The 3D-CNN model learns RGB, motion, and depth channel whereas FC-RNN captures the temporal features in the video. They collected their own dataset consisting of sequence videos and sentence videos. They achieved 69.2% accuracy. However, the use of 3D-CNN is a resource-intensive approach.

Jeroen et al. [16] proposed a hybrid approach to recognize sign language using statistical dynamic time wrapping for time wrapping and wrapped features are classified by separate classifiers. This approach relied mainly on 3D hand motion features. Mahesh et al. [17] tried to improve the performance of traditional approaches by minimizing the CPU processing time.

These existing previous works focus on building applications that enable communication between deaf people and hearing people [17]. However, creating an app requires a more precise design. One has to think of memory usage and other operations to enable a smooth user experience. Dongxu Li et al. [18] worked on gathering the word-level ASL Dataset and an approach to recognize them. In their work, they concluded that more advanced learning algorithms are needed to recognize the large dataset created by them. In [11][19], authors developed a means to convert from ASL to text. They used the CNN model to identify the ASL and then they converted the predicted label to text. They mainly concentrated on generating the text for fingerspelling instead of word-level signs. Garcia and Viesca [7], focused on classifying alphabet handshape correctly for letters a–k instead of all types of ASL alphabets. Another work presented in [20] detected ASL signs and converted to audio, and authors of [21] focused on constructing a corpus using the Mexican Sign Language (MSL).

References

1. Kuhn, J.; Aristodemo, V. Pluractionality, iconicity, and scope in French Sign Language. Semant. Pragmat. 2017, 10, 1–49.

2. Liddell, S.K. American Sign Language Syntax; Walter de Gruyter GmbH & Co KG: Berlin, Germany, 2021; Volume 52.

3. Kudrinko, K.; Flavin, E.; Zhu, X.; Li, Q. Wearable sensor-based sign language recognition: A comprehensive review. IEEE Rev. Biomed. Eng. 2020, 14, 82–97.

4. Lee, B.; Lee, S.M. Smart wearable hand device for sign language interpretation system with sensors fusion. IEEE Sens. J. 2017, 18, 1224–1232.

5. Starner, T.; Weaver, J.; Pentl, A. Real-time american sign language recognition using desk and wearable computer based video. IEEE Trans. Pattern Anal. Mach. Intell. 1998, 20, 1371–1375.

6. Munib, Q.; Habeeb, M.; Takruri, B.; Al-Malik, H.A. American sign language (ASL) recognition based on Hough transform and neural networks. Expert Syst. Appl. 2007, 32, 24–37.

7. Garcia, B.; Viesca, S.A. Real-time American sign language recognition with convolutional neural networks. Convolutional Neural Netw. Vis. Recognit. 2016, 2, 8.

8. Gaus, Y.F.A.; Wong, F. Hidden Markov Model-Based gesture recognition with overlapping hand-head/hand-hand estimated using Kalman Filter. In Proceedings of the 2012 Third International

Conference on Intelligent Systems Modelling and Simulation, Kota Kinabalu, Malaysia, 8–10 February 2012; pp. 262–267.

9. Rahman, M.M.; Islam, M.S.; Rahman, M.H.; Sassi, R.; Rivolta, M.W.; Aktaruzzaman, M. A new benchmark on american sign language recognition using convolutional neural network. In Proceedings of the 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 24–25 December 2019; pp. 1–6.

10. Huang, J.; Zhou, W.; Li, H.; Li, W. Sign language recognition using 3d convolutional neural networks. In Proceedings of the 2015 IEEE international conference on multimedia and expo (ICME), Turin, Italy, 29 June–3 July 2015; pp. 1–6.

11. Thakar, S.; Shah, S.; Shah, B.; Nimkar, A.V. Sign Language to Text Conversion in Real Time using Transfer Learning. In Proceedings of the 2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 7–9 October 2022; pp. 1–5.

12. Chung, H.X.; Hameed, N.; Clos, J.; Hasan, M.M. A Framework of Ensemble CNN Models for Real-Time Sign Language Translation. In Proceedings of the 2022 14th International Conference on Software, Knowledge, Information Management and Applications (SKIMA), Phnom Penh, Cambodia, 2–4 December 2022; pp. 27–32.

13. Kasapbaşi, A.; Elbushra, A.E.A.; Omar, A.H.; Yilmaz, A. DeepASLR: A CNN based human computer interface for American Sign Language recognition for hearing-impaired individuals. Comput. Methods Progr. Biomed. Update 2022, 2, 100048.

14. Enrique, M.B., III; Mendoza, J.R.M.; Seroy, D.G.T.; Ong, D.; de Guzman, J.A. Integrated Visual-Based ASL Captioning in Videoconferencing Using CNN. In Proceedings of the TENCON 2022-2022 IEEE Region 10 Conference (TENCON), Hong Kong, 1–4 November 2022; pp. 1–6.

15. Ye, Y.; Tian, Y.; Huenerfauth, M.; Liu, J. Recognizing american sign language gestures from within continuous videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2064–2073.

16. Lichtenauer, J.F.; Hendriks, E.A.; Reinders, M.J. Sign language recognition by combining statistical DTW and independent classification. IEEE Trans. Pattern Anal. Mach. Intell. 2008, 30, 2040–2046.

17. Mahesh, M.; Jayaprakash, A.; Geetha, M. Sign language translator for mobile platforms. In Proceedings of the 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 13–16 September 2017; pp. 1176–1181.

18. Li, D.; Rodriguez, C.; Yu, X.; Li, H. Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 1459–1469.

19. Patil, P.; Prajapat, J. Implementation of a real time communication system for deaf people using Internet of Things. In Proceedings of the 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, India, 11–12 May 2017; pp. 313–316.

20. Santon, A.L.; Margono, F.C.; Kurniawan, R.; Lucky, H.; Chow, A. Model for Detect Hand Sign Language Using Deep Convolutional Neural Network for the Speech/Hearing Impaired. In Proceedings of the 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), Jakarta, Indonesia, 16–17 November 2022; pp. 118–123.

21. Trujillo-Romero, F.; García-Bautista, G. Mexican Sign Language Corpus: Towards an automatic translator. ACM Trans. Asian-Low-Resour. Lang. Inf. Process. 2023, 22, 1–24.