



Survey Paper on Sign Language Recognition

FIZAN MOHAMMED SHAREEF¹, LAVANYA², SOURABH SHETTY³,

VANSHIKA S HEGDE⁴, Dr.SREEJA RAJESH⁵

Student, Information Science & Engineering, MITE, Moodbidri, India¹⁻⁴

Associate Professor, Information Science & Engineering, MITE, Moodbidri, India⁵

Abstract: To improve communication between those who are deaf or hard of hearing and the hearing population, sign language recognition is a developing field of study within computer vision and pattern recognition. The methodology, difficulties, and most recent advancements in sign language recognition are all covered in this survey paper's succinct and innovative review. The study discusses the distinctive features of sign language, looks at the basic building blocks of recognition systems, assesses the datasets and metrics that are currently available, and provides developments in deep learning and multimodal integration. In order to promote inclusivity and accessibility for the deaf and hard of hearing communities, it ends by proposing potential avenues for future research.

Keywords: Artificial intelligence, feature classification, machine learning, supervised learning, Mediapipe, SVM.

I. INTRODUCTION

People who are deaf or hard of hearing use sign language, a visual and spatial language, to communicate. It serves as a bridge for successful communication and the sharing of ideas and feelings between the hearing population and the deaf community. However, the lack of communication between sign language users and non-signers might restrict deaf people's social and educational options.

The automatic interpretation and understanding of sign language motions using computer vision and pattern recognition algorithms has emerged as a potential research field to address this problem. Sign language recognition systems aim to enable real-time translation of sign language into spoken or written language, facilitating effective communication between sign language users and non-signers.

The development of accurate and robust sign language recognition systems is essential for promoting accessibility, inclusivity, and equal opportunities for the hard of hearing communities. These systems have the potential to empower individuals with hearing impairments by enabling them to participate fully in various domains, including education, employment, and social interactions.

However, despite the progress made, several challenges persist in sign language recognition. These challenges arise from the intricacies and variations in sign languages, the dynamic nature of hand gestures, occlusion issues, and variations in lighting conditions. Overcoming these challenges requires innovative approaches and interdisciplinary collaboration between researchers, machine learning, linguistics, and deaf studies.

This survey paper aims to provide an overview of techniques, methodologies, and challenges in sign language recognition. By examining the existing literature, analyzing key components of recognition systems, discussing datasets and evaluation metrics, and reviewing recent advancements, this survey paper aims to consolidate the knowledge in the field for future research directions. Ultimately, the goal of this survey paper is to contribute to the advancement of sign language recognition technology, promoting inclusivity, accessibility, and improved communication for the deaf and hard of hearing communities.

II. PURPOSE OF THIS PAPER

The purpose of a survey paper on sign language recognition is to provide a comprehensive and consolidated overview of the field. It serves multiple important purposes, including:

1. Consolidation of knowledge: A survey study pulls together a variety of research publications, methodology, strategies, and developments in the recognition of sign language. It aids scholars and practitioners in developing a thorough awareness of the state, accomplishments, and limitations of the area by providing a cogent summary of the extant literature.



2. Identification of research gaps: Through the survey, gaps and shortcomings in the existing research can be identified. These gaps may include unexplored areas, unresolved challenges, or opportunities for improvement.
3. Methodologies are assessed and compared: Hand tracking, gesture segmentation, feature extraction, and classification are some of the different steps in the recognition of sign language. Examining and contrasting the various approaches used in each component is possible while writing a survey study. Researchers can learn more about the efficacy of different methods and pinpoint fruitful areas for further investigation by evaluating the advantages and disadvantages of various methodologies.
4. Exploration of challenges and limitations: Sign language recognition presents unique challenges due to the visual nature of sign language, dynamic hand gestures, variations across different sign languages, and environmental factors. A survey paper examines these challenges, highlighting the difficulties faced by existing systems and the limitations of current approaches. Understanding these challenges is crucial for driving innovation and developing more robust and accurate sign language recognition systems.
5. Finding new patterns and future research directions: Since the field of sign language recognition is continuously developing, a survey report is helpful in finding these developments. Researchers might learn about prospective directions for additional research and development by looking at current developments, such as the incorporation of deep learning methods or the use of multimodal information.

Overall, the purpose of a survey paper on sign language recognition is to provide a comprehensive and up-to-date overview of the field, facilitate knowledge consolidation, highlight research gaps and challenges, compare methodologies, improving the accuracy, efficiency, and accessibility of sign language recognition systems.

III. LITERATURE SURVEY

In a paper by B. Joksimoski [1], presents a comprehensive survey of technological advancements in the field of sign language recognition, visualization, and synthesis. Sign language serves as a vital means of communication for the hard of hearing community, and recent developments in computer graphics, computer vision, neural networks, and hardware have opened up new possibilities in this domain. By following the PRISMA methodology and employing a Natural Language Processing toolkit, were systematically identified and analyzed. The synthesis of nearly 2000 papers revealed emerging trends, highlighting the significant impact of image processing and deep learning on improving performance metrics in sign language-related tasks. Furthermore, the paper identifies common threads and research gaps, for future research directions. This survey aims to consolidate knowledge, promote further advancements, and enhance communication and inclusivity for individuals using sign language.

To address the issue of handling transition sections between neighbouring signals in large- vocabulary continuous sign language recognition (SLR), G. Fang, W. Gao, and D. Zhao [2] present transition-movement models (TMMs) in their work. They present a temporal clustering approach that enhances k-means by dynamically clustering transition motions using time warping as a distance metric. They also offer an iterative segmentation approach for automatically segmenting transition portions and training the TMMs using a bootstrap procedure. The clustered TMMs have significant generalisation skills, making them suitable for continuous SLR with a big vocabulary. The Viterbi search technique can be used to accurately recognise continuous sign language sentences by combining TMMs and sign models. Overall, this method makes a significant contribution to raising the precision and efficiency of continuous sign language recognition systems.

An autonomous Australian sign language (Auslan) recognition system is presented by Holden, EJ., Lee, G., and Owens [3]. It tracks multiple target objects (faces and hands) in image sequences and extracts features for sign phrase recognition. The system uses a mix of motion cues and the snake algorithm to recognise the contour of the foreground moving item in order to get over the occlusion issues brought on by the overlapping face and hand regions. By establishing correspondences between simple geometric features in consecutive frames, accurate tracking is achieved. The system utilizes invariant features that capture relative geometrical positioning, shapes, and motion directions of the target objects, ensuring robust recognition. This research makes significant strides in developing an automated Auslan recognition system, contributing to the advancement of sign language recognition technology.

The survey conducted by A. Er-Rady, R. Faizi, R. O. H. Thami, and H. Housni [4] aims to explore the advancements in sign language recognition over the past decade. The authors provide a comprehensive review of the state-of-the-art components of an Automatic Sign Language Recognition (ASLR) system, starting from feature extraction and extending



to sign recognition. By examining the latest developments in each stage of the ASLR pipeline, the survey offers insights into the progress made in the field. The paper encompasses a wide range of topics, including various techniques and methodologies employed for feature extraction, modeling of sign language gestures, and the use of machine learning algorithms for classification. This survey serves as a valuable resource for researchers and practitioners interested in understanding the recent advancements and future directions of sign language recognition technology.

Srujana Gattupalli, Amir Ghaderi, and Vassilis Athitsos [5] present a paper introducing a new dataset specifically designed for human pose estimation in the sign language recognition (SLR) domain. By doing user-independent trials on their dataset, the authors assess the efficacy of two deep learning-based pose estimation techniques. They also investigate the use of transfer learning strategies and show how transfer learning can improve posture estimate accuracy. Future research in the topic will benefit greatly from the dataset and the conclusions reached using these techniques. To develop human posture estimation methods for sign language recognition, researchers can use this dataset as a benchmark and take advantage of the study's learnings.

Agarwal and M. K. Thakur [6] present a method for understanding sign language that makes use of depth images obtained using a Microsoft Kinect® camera. The authors create a separate depth and motion profile for each sign language gesture using computer vision techniques. The feature matrix created from these profiles is utilised to train a multi-class Support Vector Machine (SVM) classifier. By contrasting the outcomes with those obtained using current methods, the performance of the suggested system is assessed. The numerals 0 to 9 are represented by sign language movements in the dataset used in this study. By utilising depth photographs and combining computer vision methods with machine learning algorithms for sign language detection, this study advances the field.

Oscar Koller, Hermann Ney, and Richard Bowden [7] address the challenge of robustly modeling mouth shapes in sign language recognition using deep convolutional neural networks (CNNs). Annotating sign language mouth shapes is a difficult task, resulting in a scarcity of publicly available annotations. To overcome this limitation, the authors leverage related information sources as weak supervision. The human visual focus during sign language communication is primarily on the face, particularly the mouth shapes that exhibit natural patterns with significant variability. Surprisingly, the majority of current research on the understanding of sign language ignores the face and fails to specifically highlight mouth forms. Through the presentation of many contributions, this paper introduces developments in sign language recognition. First, they suggest a method for weakly supervised convolutional neural network training without the use of explicit frame labels. They also offer a strategy for incorporating the outputs of a neural network classifier into a Hidden Markov Model (HMM) framework. The categorization performance of mouth shapes is improved significantly by their methodology, outperforming the most recent cutting-edge methods. This study emphasises the need of taking lip shapes into account when recognising sign language and shows encouraging outcomes when utilising deep learning techniques.

In their publication [8], S. N. Sawant and M. S. Kumbhar describe a sign language recognition system that uses MATLAB to identify 26 motions from the Indian Sign Language. Four modules make up the system: feature extraction, sign recognition, sign-to-text conversion, and sign-to-voice conversion. Pre-processing and hand segmentation are also included. The pre-processing and segmentation module segments the hands using image processing methods. For recognition purposes, several features, such as Eigen values and Eigen vectors, are extracted. The technology known as Principal Component Analysis (PCA) is used to recognise gestures. After that, the gestures are translated into text and audio formats. The suggested approach intends to lower obstacles to verbal communication between those who can communicate verbally and those who are deaf or mute. By giving a way to decipher signs, language gestures, this system facilitates communication between individuals with hearing and speech impairments and the general population.

Vogler, C., and Metaxas [9] introduce a framework for recognizing American Sign Language (ASL) in their paper. Developing scalable systems for sign language poses challenges in creating fundamental components for signs and handling simultaneous events, such as hand movement and changes in handshape. Handling these simultaneous events can be particularly complex due to the potential for a large number of combinations. To address this, the authors loosely adopt the Movement-Hold model, which breaks down signs into constituent phonemes, serving as the foundational building blocks. They also explore integrating handshape into this breakdown and discuss the most effective representation for handshape. The framework proposed in this paper offers insights into addressing the challenges of sign language recognition, specifically for ASL, by providing a structured approach to analyzing the phonetic and handshape components of signs.

In the paper by Raheja, J.L., Mishra, A., and Chaudhary [10], the focus is on Indian sign recognition techniques in real-time scenarios. The authors employ a multi-step approach that involves preprocessing the captured video by converting it to the HSV color space and performing segmentation based on skin pixels. Additionally, depth information is utilized



to enhance the accuracy of the results. Hu-Moments and motion trajectory are then extracted from the image frames, and gesture classification is carried out using Support Vector Machine (SVM). The proposed system is evaluated using both a webcam and MS Kinect. This type of system holds promise in aiding the teaching and communication of individuals with hearing impairments by effectively recognizing and interpreting hand gestures in real-time situations.

In the paper by Lu, J., Nguyen, M., and Yan, W.Q. [11], the authors investigate state-of-the-art deep learning methods for sign language recognition. They propose Capsule Network (CapsNet) as a means to achieve this goal, which demonstrates positive results. Furthermore, they propose Selective Kernel Network (SKNet) with attention mechanism to extract spatial information. Recognizing sign language from digital videos in real-time is a new challenge in this research field due to its importance as a means of communication.

In the study conducted by Rodríguez-Moreno I, Martínez-Otzeta JM, Goienetxea I, and Sierra B [12], a recognition system for Argentinian Sign Language (LSA) is presented. The system utilizes hand landmarks extracted from videos in the LSA64 dataset to differentiate between different signs. Various features are extracted from the signals created using the values of hand landmarks. These signals are initially transformed using the Common Spatial Patterns (CSP) algorithm, which is commonly used for EEG systems and serves as a dimensionality reduction technique. The features extracted from the transformed signals are then utilized as input for different classifiers, including Random Forest (RF), K-Nearest Neighbors (KNN), and Multilayer Perceptron (MLP).

The importance of American Sign Language (ASL) is stressed in the study by D. Naglot and M. Kulkarni [13], along with a suggested classification method and effective outcomes. To facilitate interpretation and comprehension, ASL primarily employs one hand for motions. The "Leap Motion Controller" (LMC), a 3D non-contact motion sensor capable of tracking and identifying hands, fingers, bones, and finger-like objects, is used to capture the signs. The suggested system builds a classification model from an input feature set using a Multi-Layer Perceptron (MLP) neural network and the Back Propagation (BP) technique. To recognise distinct ASL alphabets, the MLP neural network is used, specifically taking into account 26 different alphabets. The programme is honed using a dataset consisting of a total of 520 samples, with 20 samples for each alphabet.

The paper by C. Vogler and D. Metaxas [14] introduces an approach for continuous recognition of American Sign Language (ASL) using 3D data of arm motions as input. Computer vision techniques are employed to extract 3D object shape and motion parameters, while the Ascension Technologies' "Flock of Birds" is utilized interchangeably to accurately obtain 3D movement parameters of ASL sentences. The selected ASL sentences come from a vocabulary of 53 signs with varied sentence structures. These parameters serve as features for hidden Markov models (HMMs). To improve recognition results and address coarticulation effects, the authors experiment with two approaches. The first approach involves training context-dependent HMMs inspired by speech recognition systems, while the second approach models transient movements between signs, drawing inspiration from ASL phonology.

In the paper by K. Kudrinko, E. Flavin, X. Zhu, and Q. Li [15], the authors discuss the use of sign language as a primary form of interaction for individuals who are Deaf, deafened, hard of hearing, or non-verbal. Communication barriers arise when interacting with individuals who do not understand or use sign language. Advancements in technology and machine learning have led to innovative approaches for gesture recognition. The paper focuses on analyzing studies that employ wearable sensor-based systems for classifying sign language gestures. A comprehensive review of 72 studies conducted from 1991 to 2019 is presented, highlighting trends, best practices, and common challenges. Attributes such as sign language variation, sensor configuration, classification method, study design, and performance metrics are analyzed and compared. The findings from this literature review can aid in the development of user-centered and robust wearable sensor-based systems for sign language recognition.

The paper by Lubo Geng, Xin Ma, Haibo Wang, Jason Gu, and Y. Li [16] proposes a novel feature descriptor for sign language recognition. It combines hand shape features extracted from depth images with spherical coordinate (SPC) features extracted from 3D hand motion trajectories to form the final feature representation. This representation effectively incorporates spatial and temporal information to depict the kinematic connectivity among the hand, wrist, and elbow for recognition. It also mitigates issues such as illumination changes and cluttered backgrounds that can interfere with other methods. The authors evaluate the effectiveness of their combined feature using a self-built dataset consisting of 320 instances. Experimental results with different feature representations demonstrate the superior performance of Extreme Learning Machine (ELM) compared to Support Vector Machine (SVM).

The survey paper by S. Mitra and T. Acharya [17] provides an overview of gesture recognition, with a specific focus on hand gestures and facial expressions. The paper discusses various applications involving hidden Markov models, particle



filtering and condensation, finite-state machines, optical flow, skin color, and connectionist models in detail. It also highlights existing challenges and potential avenues for future research in the field of gesture recognition.

In the work by K. Bantupalli and Y. Xie [18], the goal is to develop a vision-based application that facilitates sign language translation to text, thereby enhancing communication between signers and non-signers. The proposed model processes video sequences and extracts both temporal and spatial features from them. Spatial features are recognized using Inception, a Convolutional Neural Network (CNN), while temporal features are trained using a Recurrent Neural Network (RNN). The American Sign Language Dataset is utilized for training and evaluation purposes.

In the paper by Z. Wang et al. [19], the authors explore the use of smartphones as active sonar sensing systems to identify hand movements. The smartphone's speakers emit ultrasonic signals, and the microphone on the same phone receives the modified echo affected by hand movements. The paper provides a comprehensive survey on the characteristics of studies utilizing the active sonar sensing system for hand gesture recognition. It begins by reviewing existing research on hand gesture recognition based on acoustic signals. The characteristics of ultrasonic signals and the fundamental principles of hand gesture recognition are then described. The paper focuses on typical methods employed in these studies and presents a detailed analysis of signal generation, feature extraction, preprocessing, and recognition methods.

In the paper by N. Adaloglou et al. [20], a comparative experimental assessment of computer vision-based methods for sign language recognition is conducted. The study implements the latest deep neural network methods in this field and performs a thorough evaluation using multiple publicly available datasets. The objective is to gain insights into sign language recognition, with a particular focus on mapping non-segmented video streams to glosses. The study introduces two new sequence training criteria, known from the fields of speech and scene text recognition, and extensively discusses various pretraining schemes. Additionally, a new RGB+D dataset for the Greek sign language is created, which is the first sign language dataset providing three annotation levels (individual gloss, sentence, and spoken language) for the same set of video captures.

IV. CONCLUSION

In this survey, we have focused on the application of Artificial Intelligence (AI) and machine learning techniques, particularly supervised learning and Support Vector Machines (SVM), in sign language recognition using the Mediapipe framework. Through our analysis, we found that feature classification plays a critical role in accurately interpreting sign language gestures. By leveraging AI and machine learning algorithms, specifically SVM, we witnessed significant advancements in sign language recognition, enabling more effective communication for the hard of hearing community. These findings demonstrate the potential of AI-based technologies in bridging the communication gap and improving accessibility, highlighting the importance of ongoing research and development in this field.

REFERENCES

- [1] B. Joksimoski et al., "Technological Solutions for Sign Language Recognition: A Scoping Review of Research Trends, Challenges, and Opportunities," in *IEEE Access*, vol. 10, pp. 40979- 40998, 2022, doi: 10.1109/ACCESS.2022.3161440.
- [2] G. Fang, W. Gao and D. Zhao, "Large-Vocabulary Continuous Sign Language Recognition Based on Transition-Movement Models," in *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans*, vol. 37, no. 1, pp. 1-9, Jan. 2007, doi: 10.1109/TSMCA.2006.886347.
- [3] E.J. Holden, G. Lee, and R. Owens, "Australian sign language recognition," *Machine Vision and Applications*, vol. 16, no. 4, pp. 312-320, 2005. doi: 10.1007/s00138-005-0003-1.
- [4] A. Er-Rady, R. Faizi, R. O. H. Thami, and H. Housni, "Automatic sign language recognition: A survey," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 1-7, 2017, doi: 10.1109/ATSIP.2017.8075561.
- [5] S. Gattupalli, A. Ghaderi, and V. Athitsos, "Evaluation of Deep Learning based Pose Estimation for Sign Language Recognition," in *Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '16)*, Article 12, pp. 1-7, 2016.
- [6] A. Agarwal and M. K. Thakur, "Sign language recognition using Microsoft Kinect," 2013 Sixth International Conference on Contemporary Computing (IC3), pp. 181-185, 2013, doi: 10.1109/IC3.2013.6612186.
- [7] O. Koller, H. Ney, and R. Bowden, "Sign language recognition by means of statistical models," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, pp. 85-91, 2015.
- [8] S. N. Sawant and M. S. Kumbhar, "Real-time Sign Language Recognition using PCA," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, pp. 1412-1415, 2014, doi:



10.1109/ICACCCT.2014.7019333.

- [9] C. Vogler and D. Metaxas, "Handshapes and Movements: Multiple-Channel American Sign Language Recognition," in *Gesture-Based Communication in Human-Computer Interaction, GW 2003, Lecture Notes in Computer Science*, vol. 2915, pp. 275-284, Springer, Berlin, Heidelberg, 2003, doi: 10.1007/978-3-540-24598-8_23.
- [10] J.L. Raheja, A. Mishra, and A. Chaudhary, "Indian sign language recognition using SVM," *Pattern Recognition and Image Analysis*, vol. 26, no. 3, pp. 434-441, 2016.
- [11] J. Lu, M. Nguyen, and W.Q. Yan, "Sign Language Recognition from Digital Videos Using Deep Learning Methods," in *Geometry and Vision*, M. Nguyen, W.Q. Yan, and H. Ho (eds), *Communications in Computer and Information Science*, vol. 1386, Springer, Cham, 2021.
- [12] I. Rodríguez-Moreno, J.M. Martínez-Otzeta, I. Goienetxea, and B. Sierra, "Sign language recognition by means of common spatial patterns: An analysis," *PLoS ONE*, vol. 17, no. 10, e0276941, 2022, doi: 10.1371/journal.pone.0276941.
- [13] D. Naglot and M. Kulkarni, "Real-time sign language recognition using the Leap Motion controller," 2016 *International Conference on Inventive Computation Technologies (ICICT)*, pp. 1- 5, 2016, doi: 10.1109/INVENTIVE.2016.7830097.
- [14] C. Vogler and D. Metaxas, "Adapting hidden Markov models for ASL recognition by using three-dimensional computer vision methods," 1997 *IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, vol. 1, pp. 156-161, 1997, doi: 10.1109/ICSMC.1997.625741.
- [15] K. Kudrinko, E. Flavin, X. Zhu, and Q. Li, "Wearable Sensor-Based Sign Language Recognition: A Comprehensive Review," in *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 82-97, 2021, doi: 10.1109/RBME.2020.3019769.
- [16] L. Geng, X. Ma, H. Wang, J. Gu, and Y. Li, "Chinese sign language recognition with 3D hand motion trajectories and depth images," in *Proceedings of the 11th World Congress on Intelligent Control and Automation*, pp. 1457-1461, 2014, doi: 10.1109/WCICA.2014.7052933.
- [17] S. Mitra and T. Acharya, "Gesture Recognition: A Survey," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311-324, May 2007, doi: 10.1109/TSMCC.2007.893280.
- [18] K. Bantupalli and Y. Xie, "American Sign Language Recognition using Deep Learning and Computer Vision," 2018 *IEEE International Conference on Big Data (Big Data)*, pp. 4896-4899, 2018, doi: 10.1109/BigData.2018.8622141.
- [19] Z. Wang et al., "Hand Gesture Recognition Based on Active Ultrasonic Sensing of Smartphone: A Survey," in *IEEE Access*, vol. 7, pp. 111897-111922, 2019, doi: 10.1109/ACCESS.2019.2933987.
- [20] N. Adaloglou et al., "A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition," in *IEEE Transactions on Multimedia*, vol. 24, pp. 1750-1762, 2022, doi: 10.1109/TMM.2021.3070438.