IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

A Survey on Sign Language Interpretation using Machine Learning

Dipali Pawar[1], Aniruddha Jathar[2], Omkar Kodmur[3], Prathamesh Belurkar[4], Shrikant Bhandalkar[5]
[1]Professor, Department of Computer Engineering
[2,3,4,5]Student, Department of Computer Engineering
Zeal College of Engineering and Research, Pune, India

Abstract: *Every day we see many people with disabilities like the deaf, the dumb and the blind, etc. Sign language is one of the communication tools for the hard-of-hearing people community and common people community. But normal people find it hard to understand the sign language and gestures of the deaf and dumb. Many tools can be used to translate the sign language created by the disabled into a form that normal people can understand. The studies are based on various image acquisition, pre-processing, hand gesture segmentation, extraction of features, and classification methods. This paper aims to research and examine the methods employed within the SLR systems, and the classification methods used, and to propose the most promising technique for future research. Due to the latest advancement in classification methods, a few of the currently proposed works specifically contribute to classification methods, together with hybrid techniques and deep learning. This paper specializes in the classification strategies utilized in earlier Sign Language Recognition. Based on our review, HMM-based techniques were explored significantly in previous studies, which include modifications. Deep learning consisting of convolutional neural networks has become popular over the past five years.*

Keywords: SLR, Sign Language, Recognition, Computer Vision, Neural Networks, Deep Learning, HMM, CNN.

I. INTRODUCTION

This paper addresses the various algorithms and techniques that can be used to understand the sign language and hand gestures of hard-of-hearing people. The hand gesture recognition system is considered a method for more intuitive and efficient interaction with humans and computers. The program range includes virtual prototyping, an examination of sign language, and medical training [1][2].

Sign language is one of the communication tools for hard-of-hearing people communities and common people communities [1]. Currently, research work has mainly focused on the identification of static sign language signs from images or video sequences captured under controlled conditions. Signers in this category are required to wear a glove sensor or a shaded glove. The task will be standardized by wearing gloves during the segmentation process. The disadvantage of this approach is that during the device service, the participant will wear the sensor components together with the gloves [2].

Methods Based on Vision: Computer-based vision strategies are non-invasive and based on how people perceive their environmental knowledge [4]. Although developing a vision-based interface for general use is challenging, designing such an interface for a controlled environment is still achievable. Feature selection is crucial to the recognition of gestures since hand gestures are very unique in shape variation, textures, and motion [5][7].

II. THEORY

For static hand recognition, it is easy to recognize hand posture by extracting some features such as finger directions, fingertips, skin color, and hand contours. Such features are not always available and reliable due to lighting conditions and the background of the image. There are also many other non-geometric features, such as the silhouette, color, and textures, that are inadequate in recognition. Since it is not easy to define features clearly, the entire frame or transformed image is taken as the input, and features are chosen automatically and implicitly by the recognizer

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 1, June 2023

[6][9][10]. This paper aims to review and evaluate the approaches used in previous studies. It also aims to recommend the best method to investigate for future research. Majid and Zain (2013) were studied the development of Sign Language Recognition devices for various sign languages. They only studied the best 32 related publications up to 2012 [12].

III. LITERATURE SURVEY

Paper	Model	Optimizer	Accuracy
Indian sign language translator using gesture recognition algorithm.	CNN	Stochastic Gradient Descent	83.29
Indian sign language recognition: database creation, hand tracking and segmentation.	Faster R–CNN (CNN-LSTM)	SoftMax Function (CNN) CTC	42
AjuDennisan American Sign Language Alphabet Recognition Using Deep Learning.	3D CNN	Multilayer perceptron classifier	94.2
Recognition of isolated indian sign language gesture in real time.	CNN	multi-classification problem & SoftMax function	99.72
Optical flow hand tracking and active contour hand shape features for continuous sign language recognition with artificial neural networks.	CNN-LSTM	SoftMax function (CNN)	95
Selfie sign language recognition with multiple features on adaboost multilabel multiclass classifier.	CNN	Adaboost multilabel multiclass classifier	88
A video based Indian Sign Language Recognition System (INSLR) using wavelet transform and fuzzy logic.	Wavelet transform and Fuzzy logic	Movement analysis classifier	75
American sign language fingerspelling recognition in the wild.	CNN	SoftMax function (CNN)	98
American sign language recognition system: an optimal approach.	CNN	CNN classifier	87
American sign language recognition using machine learning and computer vision.	CNN	multi-classification problem & SoftMax function	86
Continuous Chinese sign language recognition with CNN-LSTM.	CNN	CNN	76
Classifier fusion based on Bayes aggregation method for Indian sign language datasets.	CNN	SoftMax function (CNN)	91

IV. METHODOLOGY

A review of the sign language recognition system

1. **Sign Language**

The word sign language is similar to the language phrase, many of both are spread around different world territories [1]. Similar to language, sign language evolves over a long period of period sign language grammar and vocabulary, so it is considered a legitimate language. Because no perception of hearing is needed to understand sign language and no voice is needed to produce sign language, it is the common language among the deaf.

Sign languages [3] are usually constructed by using simultaneous compilation associated with hand shapes, orientations, and moves of the hands, palms, or body, along with facial expressions to fluidly explicit a speaker's thoughts [13].

DOI: 10.48175/568

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.301
Volume 3, Issue 1, June 2023

2. Sign Capturing Methods

The signs must be captured to provide input for the sign language recognition system. To capture images of hand gestures through a Microsoft Kinect camera which handles single-hand signs, double-hand signs, and finger-spelling [4], Microsoft Kinect sensors to capture multimodal data [1][2][3], Microsoft Kinect (RGB-D) sensor handled by the Nui Capture Analyze application [4], front cameras and mobile cameras [9], Sony video cameras [2][3][10], and Cannon 600 D camera RGB videos [4] are used. Microsoft Kinect was initially designed as a Gaming console peripheral device. The three sensors, that is, RGB, audio, and depth allow movements to be detected and user faces/speeches to be recognized. Microsoft Kinect sensors use a variety of useful computer vision applications, including gesture recognition, motion recognition, robotics, and virtual reality [6].

3. Sign Language Recognition Techniques

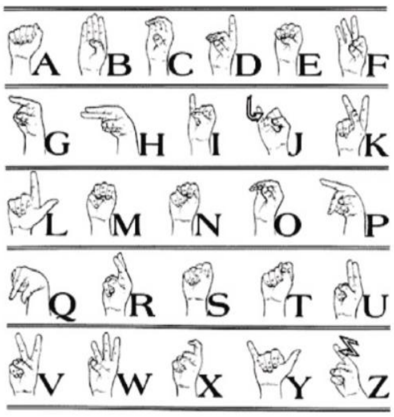
There are again two different approaches to vision-based sign language recognition: appearance-based and deep learning-based.Appearance-based approaches are modeled by a collection of 2D intensity images. In turn, gestures are modeled as a sequence of views. Appearance-based approaches attempt to infer the pose of the palm and the joint angles [1][4]. Appearance-based frameworks use images or videos as data sources. They straightforwardly interpret these videos/images. They do not utilize a spatial representation of the body. The parameters are usually derived straightforwardly from the images or videos using a template data repository.

Traditional Machine Learning-based Approaches: A famous AI approach is to regard a motion as the output of a stochastic process. Of this category of approaches, CNN has received the most consideration in the literature for classifying gestures [5].

4. Image Acquisition

The important element used in the sign language recognition method (SLR) as the input method is the camera. The input data for the SLR are in the form of a moving image that can easily be recorded by a camera. Nevertheless, some researchers use normal cameras to capture images [1][2][4][5]. Some researchers claim that they are using cameras and no gloves to reduce the challenge of using sensor-based gloves. Cameras usually support many video formats, so we need to define the default format and the format that we want to utilize by using Digitizer Configuration Format (DCF) file. Some researchers have used higher-quality cameras because the image of the web camera is blurred. A camera was used to capture 30 frames per second of real-time video and then analyzed frame by frame for dynamic gestures. The system uses a skin filter to extract the skin region and is then converted into HSV color space for each frame to an image.

There is also another device named Microsoft Kinect [1][2][5] that is used to capture images.



Sign Language Alphabets

Pre-processing

The image pre-processing step improves the system input to modify the image and videos. Median and Gaussian filters are some of the most frequently used methods to reduce noise in input images or videos. In research median filtering is exclusively used for the image pre-processing stage and morphological operations are also broadly used to remove

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 1, June 2023

unwanted information from the input. For instance, Badhe et al. [1] and Krishnaveni et al. [3] threshold the input image into binary and then K-means clustering with morphological operations in the pre-processing stage to remove noise. An adaptive histogram [4] is used to improve the contrast of input images acquired in different environments.

The segmentation approach can be contextual or non-contextual. Contextual segmentation considers the spatial connection between highlights, for example, edge recognition strategies, while a non-context-oriented division does not consider spatial relationships but bunches pixels dependent on global attributes. Skin detection also applies hand movement tracking with skin detection to produce more specific end results. Similarly, to skin detection, colored gloves are used to provide distinct characteristics to the hands, thus aiding in hand segmentation.

Skin color segmentation is processed in the RGB model, HSV model, HIS model, and YCbCrcolor models [13], while color segmentation follows difficulties because we may face sensitivity to illumination, cameras, and skin tone. The HSV color model is famous because the hue of the hand used differentiates palm from arm easily. In research [3][8][10], the Palms and faces of people were segmented according to HSV and YCbCrcolor models. Ahmed et al. [4] used the RGB color model to carry out hand skin color segmentation. Match and compare with the RGB color model used to find skin colorin a given image or video. According to research [2][3][5], the YCbCrcolor model was found to be useful for color segmentation under various light conditions.

Classification

Classification finds a function to determine the category to which the input data belongs. It can be two-category or multi-category. Factors including the classification construction method, the properties of the data to be classified, and the number of training samples all influence classification accuracy.

Softmax function [2][5][10][12][13]the Softmax classifier finds classes based on final output probabilities in an efficient manner, and it implies multi-classification networks. In Softmax, every class is assigned decimal probabilities in the multiclass problem, and probabilities add up to 1.0. The categorical cross-entropy loss function is mostly used with the softmax activation function. The output of the model is most effective because of its logarithmic output value. The softmax activation rescales the model output so that it has the right properties. Because of this, it is common to append a softmax function because of the final layer of the neural network.

Machine Learning Classifiers

We chose the Hidden Markov model (HMM) [4], which has the highest probability of generating the given collection and the sign with the highest probability. The method of popularity is to pick out the model from the set of fashions that could properly represent the phrase series. Support vector machines (SVMs) [1] are used to classify signs based on these features and linguistic elements. The SVM classifier is a multiclass classifier that searches for an ideal hyperplane as a decision function. Once trained on images containing some particular gestures, the SVM classifier can make decisions regarding the sign [5][8][10].

Deep Learning Based Approaches

Author	Dataset image Acquisition	Pre-processing	Model	Optimizer	Accuracy
Nikhil Kasukurthi and et all (2019) [3]	Intel Real sense P200 Depth camera	Squeeze Net	CNN	Stochastic Gradient Descent	83.29
B. Shi et al. (2018)	Web cam	Alex Net	Faster R–CNN (CNN-LSTM)	SoftMax Function (CNN) CTC	42
Huang and et all (2015)	Microsoft Kinect RGB video	HOG&GMM-HMM	3D CNN	Multilayer perceptron classifier	94.2

Wadhawan and et al. (2020)	Web Camera	Alex Net	CNN	multi-classification problem & SoftMax function	99.72
Su Yang and et all (2017)	Web Camera	RGB to HSV color space	CNN-LSTM	SoftMax function (CNN)	95

V. CONCLUSION

In this paper, we shared a quantitative study of different methods used in sign language recognition, covering 80 publications from 2010 until 2021. An analysis based on appearance-based SLR and vision-based SLR (deep learning) Each category was examined with roughly 40 papers. The following findings were observed in a study of papers:

A sign language recognition system has been developed from classifying only static signs and alphabets to a system that can effectively apprehend dynamic actions that come in continuous sequences of images.

Most papers published with vision-based approaches provide better results than appearance-based approaches. Researchers are currently paying more attention to making a large vocabulary for sign language recognition systems. Dataset availability and improvements in computing speed provide access to more training for given samples.

REFERENCES

[1] [1] P.C. Badhe, V. Kulkarni, Indian sign language translator using gesture recognition algorithm 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS), Bhubaneswar (2015), pp. 195-200.

[2] P.A. Nanivadekar, V. Kulkarni, Indian sign language recognition: database creation, hand tracking and segmentation. 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications, CSCITA, Mumbai (2014), pp. 358-363.

[3] Nikhil Kasukurthi, Brij Rokad, Shiv Bidani, AjuDennisan American Sign Language Alphabet Recognition Using Deep Learning, (2014).

[4] A. Nandy, J.S. Prasad, S. Mondal, P. Chakraborty, G.C. Nandi, Recognition of isolated Indian sign language gesture in real timeInf. Process. Manag. (2010), pp. 102-107.

[5] P.V.V. Kishore, M.V.D. Prasad, D.A. Kumar, A.S.C.S. Sastry, Optical flow hand tracking and active contour hand shape features for continuous signlanguage recognition with artificial neural networks. 2016 IEEE 6thInternational Conference on Advanced Computing (IACC), Bhimavaram (2016), pp. 346-351.

[6] G.A. Rao, P.V.V. Kishore, Selfie sign language recognition with multiple features on adaboost multilabel multiclass classifier, J. Eng. Sci. Technol., 13 (8) (2018), pp. 2352-2368.

[7] P.V.V. Kishore, P.R. Kumar, A video based Indian Sign Language Recognition System (INSLR) using wavelet transform and fuzzy logic, Int. J. Eng. Technol., 4 (5) (2012), p. 537.

[8] B. Shi, et al., American sign language fingerspelling recognition in the wild2018 IEEE Spoken Language Technology Workshop (SLT), Athens, Greece (2018), pp. 145-152.

[9] S.S. Shivashankara, S. Srinath, American sign language recognition system: an optimal approachInt. J. Image Graph. Signal Process. (2018).

[10] KshitijBantupalli, Ying Xie, American sign language recognition using machine learning and computer vision, Master of Science in Computer Science Theses, 21 (2019).

[11] M. Krishnaveni, V. Radha, Classifier fusion based on Bayes aggregation method for Indian sign language datasetsProcedia Eng., 30 (2012), pp. 1110-1118.

[12] Yang Su, Qing Zhu, Continuous Chinese sign language recognition with CNN-LSTM, Proc. SPIE 10420, Ninth International Conference on Digital Image Processing (ICDIP 2017) (21 July 2017), p. 104200F.