

Sign Language Detection

Under the guidance of
Prof. Pallavi. A. Pathare
Department of Computer Engineering
Sir Visvesvaraya Institute
Of Technology, Nashik

Miss. Aakanksha Mahesh Nandre,
aakankshanandre123@gmail.com,
Department of Computer Engineering
Sir Visvesvaraya Institute
Of Technology, Nashik

Mr. Kunal Ravindra Markali,
kunalmarkali2020@gmail.com,
Department of Computer Engineering
Sir Visvesvaraya Institute
Of Technology, Nashik

Mr. Samadhan Machindra Bahirat,
bahiratsamadhan@gmail.com
Department of Computer Engineering
Sir Visvesvaraya Institute
Of Technology, Nashik

Mr. Nikhil Suresh Gite,
nikhilgite89@gmail.com
Department of Computer Engineering
Sir Visvesvaraya Institute
of Technology, Nashik

Mr. Suyog Arun Mule,
mulesuyog13@gmail.com,
Department of Computer Engineering
Sir Visvesvaraya Institute
of Technology, Nashik

Abstract - This project focuses on the development of a sign language detection system using computer vision techniques. Sign language is a vital means of communication for the deaf and hard-of-hearing community, and this system aims to bridge the communication gap by automatically recognizing sign language gestures and translating them into text. The proposed system uses a camera to capture images of a person signing, and then processes the image frames to detect and recognize hand gestures. The system utilizes algorithms to classify the detected hand gestures and map them to corresponding words or phrases in sign language.

Keywords: Sign language, Detection, Recognition, Computer vision, Image classification, Performance evaluation, Accuracy.

I.INTRODUCTION

Over the past decades, one of the basic features of communication between human beings is characterized by human language and culturally inherited through the generations, and vary considerably during human societies and among them. Many people are disabled due to hearing impairment, and they lose the ability to communicate through human language, they are called deaf. Deaf people are limited in communication with a normal humans because they have lost their hearing. Therefore, people who are deaf with a disability need help to continue living. WHO also showed results to predict the number of deaf people up for the year 2020 about 466 million deaf worldwide and possibly for 2050 the number becomes 900 million people with hearing loss [1]. Hard of hearing refers to people with hearing loss ranging from mild to severe. People who are hard of hearing usually communicate through Sign Language.

Deaf/dumb and hearing-impaired people face problems in communicating with others. Sign Language helps a deaf person to communicate with others and dealing with the communication technology [2]. Therefore, this research explains an overview for Sign Language, gestures, and the most important techniques used to distinguish Sign Language and it translates to a way that can be understood using information technology and computer vision, that support interaction and communicate with users, it can be utilized to assist in the successful social integration of hearing 978-1-6654-1573-6/21/\$31.00 ©2021 IEEE Nidhal K. El Abbadi Computer Science Department Faculty of Education, University of Kufa, Najaf, Iraq nidhal.elabbadi@uokufa.edu.iq impaired or deaf individuals and to communicate with others in natural ways. This paper follows the structure: the first section explains the concept and importance of Sign Language Gestures and Fingerspelling. The second section Illustrates hand signs recognition techniques. the third section Illustrates survey for techniques background to apply sign language and last section explains conclusion for paper.

1.1Sign Language

Sign Languages (SL) use a visual method or manual movement to communicate meaning. Sign Language (SL) is the primary method of communication among the deaf or hearing-impaired with normal people in life to explain their ideas or feelings. SL includes a group of gestures, hand movements, facial expressions, or head movements to represent letters or symbols and each letter of the alphabet has a sign. There are more SL utilized around the world and enhanced from different groups of people interacting with each other, so there are between 138 and 300 different types of Sign Language used

around the world today. Some types for SL are American Sign Language (ASL), British Sign Language (BSL), Arabic Sign Language (ArSL) and et. [3].

Sign Language should not be confused with "body language", which is a type of nonverbal communication. Manual signs are put into four components: Handshape, Location, Movement, Orientation. Fig. 1 shows examples for signs alphabets. many researchers utilize datasets for signs in SL recognition by different methods such as data gloves to provide measures of finger configurations, hand orientations, and use cameras to record videos of the signer [4]. In the Arab world, deaf people use Arabic Sign Language, it is worth mentioning that each country has its own Sign Language that differs from other countries and there is an international and unified Arabic Sign Language (ArSL) used by the deaf that was developed by World Federation of the Deaf (WFD) [3,4].

Sign languages are important for the deaf and dumb as follows:

- SL helps deaf and mute people communicate with each other and transmits mutual feelings between them.
- SL works on the mental, verbal and indicative development of people with special abilities from deaf and dumb.
- SL helps to eliminate fear, depression and frustration in the deaf and dumb. It works on the development of social, cognitive, and cultural relations for individuals



Fig.1. Most common signs for Arabic alphabets(adaptedfromwebsite <http://innoopedia.blogspot.com/2015/07/signlanguage.html>)

A. Gestures

Gesture recognition is a process of distinguishing a kinematic expression for expressing different letters, numbers, or phrases, where it is used for the face, hands, head, and body with different meanings. For the purpose

of developing applications that use gestures or signs in the virtual world and also to improve systems that depend on Sign Language, gestures must be widely recognized. More specifically, hand gestures are of two types, namely

- Static Gestures: gestures are independent of time which these are not change for hand position during the gesturing period and require less computational complexity [5]. Static gestures are considered to be a single group of images that represent each image in a separate frame and is often used to distinguish spelling fingers in Sign Language [6].

- Dynamic Gestures: hand gestures that depend on time are known as dynamic, where the hand position changes continuously with time which is more complex but suitable for real-time environments, where the movement of the hand is tracked during a time and is not based on one frame only, but on many frames [5]. Data acquisition for a dynamic sign to take from the video is dependent on sequences frames of images as input [8]. Gestures consist of four basic components depending on the hand: the shape of the hand, the movement of the hand, the direction of the hand, and the position of the hand relative to the body [6]. The importance of using hand gestures for communication requires an interpreter is usually needed to translate Sign Language into natural language and vice versa. The goal of distinguishing the signal must be to provide an accurate and appropriate mechanism for converting signs into text or meaningful speech. Therefore, we need techniques to achieve this goal, as the recognition of gestures represents a great challenge. Through Sign Language for the deaf can easily access a variety of public information services [7].

B. Fingerspelling Finger's alphabet is a supportive method for Sign Language to facilitate the learning of deaf people. This way includes using one hand to represent the letters of the alphabet and it is rarely used separately, it is used in introducing some new terms and concepts, this method can also be used names that do not have a sign known to the deaf community. The alphabet of the fingers has a clearer meaning, that each of the twenty-eight letters of the alphabet has a specific movement using the hand, for example, the letter A has a recognizable vowel on its own, and so is the rest of the letters [8]. The finger-spelling is one method of the visual physical communication methods, which is based on drawing the shapes of the alphabets, that is writing them in the air instead of writing them on paper and therefore it expresses a manual method and represents written language [9]. The alphabetical method is considered one of the most important methods of communicating with the deaf, as it is an integral part of the communication process in general, and overall communication especially about words, names of people,

or addresses that do not have agreed signs. fingerspelling is suitable for reading and the fastest to bring words to the front of the eyes of the deaf [10]. Thus, the justifications for using the fingerspelling method include that training in this method will not require much effort from the hearing-impaired student because it will be an application of what he acquired through the writing skill of the alphabet, meaning that learning this method comes after training the student in lip reading and pronunciation exercises. And writing words and letters, and finally comes the training phase [11].

1.2 HAND SIGNS RECOGNITION TECHNIQUES

Several researchers have classified the signs recognition system into four steps after obtaining the input image from the camera(s), videos, and even a device equipped with a data glove. These procedures are Tracking of hand and Segmentation, Feature Extraction Methods, Classification, and Recognition as illustrated in Fig. 2 [12]. The next section of this paper explains for many researchers to apply these steps of sign recognition. For this mainly many techniques of sign, recognition is used such as, Principal component analysis (PCA), Hidden Markov Model (HMM), Speeded Up Robust Feature (SURF), Cognitive Computing, Convolutional Neural Network (CNN), Artificial neural networks (ANN).

A. Artificial neural networks (ANN)

Most of the researchers previously worked on sign recognition in SL by ANN, and it is also applicable to classified the classifier for matching gestures, that many researchers applied ANN also has been used for to extricate shape of hand [14]. An artificial neural network is a system to process information that has multiple high-performance characteristics with those of the neural networks biological. In general ANN algorithm consists of three variables affecting for process in the algorithm, which are function activation, bonding weight, and model of interconnection among different layers of neurons [15]. In [16] the researcher presented a system to using ANN to process translation for Sign Language and which is special applied to ASL words into English.

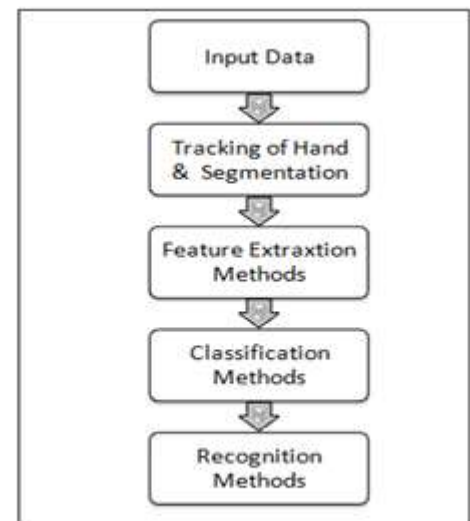


Fig. 2. most common processes for sign language

B. Convolutional Neural Networks (CNN) The convolutional neural network is a type of deep neural networks and consists of an input and an output layer, as well as multiple hidden layers, that are based on the common weights structure and characteristics of the fixed translation, it is also known as shift invariant or space invariant artificial neural networks, which has attracted much attention in recent studies. CNN is made up of neurons that have learnable weights and biases. Each neuron receives some inputs, performs a dot product, and optionally follows it with a non-linearity [16]. It can automatically extract high-level features from raw input features, which are much more powerful than human-designed features. Thus, it has brought significant improvements to many fields, for instance, image segmentation and recognition. In the contact map prediction field, CNN has also been employed in many works [18]. ConvNet architectures are most commonly applied to analyzing visual imagery, it consists mainly of several layers, the kernel for dimension reducing and activation function, in each layer number of filters used to extract the desired features. ConvNet architectures differ from one application to another.

C. K-NEAREST NEIGHBOR (K-NN) K-NN is a non-parametric statistical method whereby input data is classified by a majority vote of its neighbor. The data will be assigned to the class most common among its k nearest neighbors. Euclidean distance as in “(1)” is a commonly used similarity measure.

II. RELATED WORK AND LITERATURE SURVEY.

In this section, different techniques of methods of distinguishing gestures and signs for SL are studied and analyzed to translated into an understandable procedure for the period between (2016-2020).

(Haitham Badi, 2016) present a study to explain the advantage of two feature extraction techniques to process gestures distinguish problems based on hand contour and complex moments. The author suggested six fixed hand signals namely open and close, cut and paste, maximize and reduce, to develop a vision based on a fixed hand gesture recognition algorithm. The classification way and identify hand signs in image in SL have been using back-propagation learning algorithm, which was implemented in three phases: the first is preprocessing, second phase feature extraction, and the last phase classification by Artificial Neural Network ANN technique. The identification accuracy of hand contour was 71.30%, while the best recognition rate of complex moments was 86.90%. The drawback of this method was the suffering of a neural network from overfitting [20].

(Elakkiya, Selvamani, 2018) suggested a method to solve the ambiguities problem of the hand segmentation, movements epenthesis, and continuous signs in the video. This study using spatial and temporal features to improved dynamic programming by applied Hidden Markov Model with subunit multi-stream parallel which is known as (SMP-HMM) and minimum clustering for modeling series images. This research achieved results for accuracy up to 98% and 1.25s is considered an average time for recognition [22].

(Wenjin Tao, et.al., 2018) the authors presented a project for Sign Language is identify for (ASL) words distinguish using Leap Motion Controller (LMC), which is an option for data acquisition aim due to its high accuracy and low cost, that to extracted skeleton data and vision module. The final prediction was then produced to fuse two types of features which are to extract angle profiles from skeletal data and to extract visual features from two IR images, along with a neural network. The leave-one-out and the half-half are accuracy ratings are provided up to 80.1% and 99.7%, respectively [23].

(Zhou Z, Dai Y, Li W., 2018) introduced a new method for the recognition of isolated and continuous gestures based on the global template dynamic time warping (GTDTW) algorithm for gesture recognition with wearable gloves, which is obtained by the statistical way. GTDTW is used for gesture segmentation and also use for continuous gesture recognition. The accuracy of this method for recognition of isolated gestures was 98.8%, while the segmentation and recognition rate for continuous gestures was up to 95.6% [24].

(Salma Hayani, et.al., 2019) presented an automatic way to recognition Arabic sign language letters and numbers based on CNN. The proposed CNN consist of seven layers, where the first four layers focus on the feature extraction

and the rest three layers for classification. The recognition rate is about 90% [25].

P.K. Athira, et.al. (2019) suggest a design independent version vision-based signs recognition application which is identify gesture handed static and dynamic for one or double hand gestures from live video. The keyframe is extracting to decrease computation speed to a large range by using Zernike moments and enhanced process for co-articulation removal hand spelling alphabets after that which is extraction the feature vector from the gesture sequence and these features acquired to use for classification applying Support Vector Machine (SVM). the accuracy 91% for distinguished static gesture alphabets and 89% accuracy for single-handed dynamic words [26]. Rajat Sharma, et.al. (2020) presented a practiced model using a real camera to handle continuous images. Additionally, they have a linking sign identification and hand detection by SL which involve with each other alphabets, numbers, and grayscale images. The process implemented for each video frame is to extract relevant information by apply segmentation to isolate the signer's hands. Video frame segmentation using the geodesic active regions mode and depend on the Geodesic Active Contour (GAC) model. After the image filtering process, signs organize have been exercised CNN procedures. Feature extraction in CNN involves rearranging the scale of the assets required to accurately depict an overall arrangement of information. Thereafter, to reduce the susceptibility of the channels to influence noise and varieties can be "smoothened" by contributions from the convolution layer. The accuracy of the investigations was measured up to 99% accuracy. The drawbacks have impacting proper yield due to camera quality and also the speed of sign coordination with content may vary slightly. In any case, some weaknesses may differ in quality as evidenced by the light and prowess of the inclusion [27].

(M. M. Kamruzzaman, 2020) Suggested works on auto recognition of ArSL letters and convert each one to the corresponding voice. The training phase used a dataset of 100 images and the testing phase applied 25 images for each hand sign of 31 alphabets of ArSL. The architecture of the proposed CNN consists of two layers and two max pools, the system is trained with 100 epochs. The accuracy of this system is up to 90%. Others suggested more devices to consider to increase the accuracy by using Leap Motion or Xbox [28].

(Vincent Hernandez, et.al., 2020) presented a signal classification in real series. they suggested a new model called DeepConvLSTM, that is used the LeapMotion sensors to provide the input data to the different conventional machine learning which is designed to classify the American Sign Language. Authors improve the generalization of NN

by using simple data augmentation to prevent the overfitting, in addition to the suggestion of a kinematic model of the left and right forearm/hand/fingers/thumb. One of the limitations of this study that may reduce the classification performance is the high variation of performing movements between individuals. The accuracy of this method is 91.1% [29].

(Eman K., and Doaa R., 2020) build a model to translate the Arabic Sign Language to various meanings based on combining the deep learning algorithms and ontology. Ontology is implemented on the SL to solve many challenges of SL. The authors trained and tested the convolution neural network on several Arabic words and static alphabetic Arabic signs. The classification accuracy of this method was up to 88.87%. the drawback of this system is a high rate of incorrect recognition for some of the Arabic letters may reach up to 30% [30].

V CONCLUSION

No standard dataset is available on SL, so a new dataset is created for some research. In sign language recognition research, benchmark databases are available as a standard reference for future researches, such as Purdue RVL-SLLL [31], and RWTH-PHOENIX- Weather [32]. Also, TABLE I summarized the accuracy of Sign Language recognition methods, when using input images or videos. In TABLE II survey for the methods of Sign Language recognition to all research's studies in this paper.

Sign Language recognition methods are introduced, and we focused on methods for feature extraction and classification process to use identify the gestures for SL. Many recent algorithms (published in 2016-2020) presented and compared. Most papers presented in the last years focused on using the CNN algorithm. As future work, the System will be building to process dynamic real video and interpreter in real-time applications. The system will convert to produce a free application for deaf and speaking impaired people.

TABLE I ACCURACY OF SIGN LANGUAGE RECOGNITION

Ref	Method	Accuracy %
[20]	complex moments & ANN	86.90
[22]	Multi-Stream Parallel & HMM	98
[23]	IR images & neural network	80.1&99.7
[24]	GTDTW algorithm	95.6
[25]	ANN	86.90
[26]	ANN and C-mean	90
[27]	CNN	99
[28]	Deep learning & LeapMOTION	90
[29]	DeepConvLSTM and CNN	91.1
[30]	Ontology with CNN	98.06

TABLE II. SURVEY FOR THE METHODS OF SIGN

Ref.	Method	Advantages	Disadvantages	year
[20]	complex moments and ANN	a study to explain the advantage of two feature extraction techniques to process gestures distinguish problems based on hand contour and complex moments	The drawback of this method was the suffering of a neural network from overfitting	2016
[22]	HMM (SMP-HMM) and minimum entropy clustering for modeling units Sub-long video series	to solve ambiguities for hand segmentation and two adjacent signs in continuous video sequences	minimum clustering for modeling series images	2018
[23]	recognition using Leap Motion Controller (LMC).	to extract angle profiles from skeletal data and to extract visual features	words distinguish using Leap Motion Controller (LMC) is a high cost	2018
[24]	global template dynamic time warping (CTDTW) algorithm	applied to static and Dynamic gestures and the performance in real-time	for gesture recognition with wearable gloves.	2018
[25]	Using CNN	distinguish of Arab sign numbers and letters by the offline recognition system	CNN consist of seven-layer is taken a long time for processing	2019
[26]	LeapMotion or Xboxkinect.	design independent version vision-based signs recognition application	which is identify gesture handed static and dynamic for one or double hand gestures from live video.	2019
[27]	Using CNN	to reduce the susceptibility of the channels to influence noise and varieties	to accurately depict an overall arrangement of information.	2020
[28]	Using CNN	The different feature maps are combined to get the output of the convolution layer	Every image is converted as a 3D matrix by specified width, specified height, and specified depth.	2020
[29]	DeepConvLST M and CNN	Integrating convolution layers in a deep learning	the static neural network was used and it is better to use a dynamic neural network.	2020
[30]	combining ontology with deep learning	to semantic information (CNN)	snapshot of an image of the sign language static gesture that needs to be translated	2020

REFERENCES

- [1] M. Center, Deafness and hearing loss, Tech. rep., 445 World Health Organization (2020).
- [2] L. Ghazanfar, M. Nazeeruddin, A. Jaafar, A. Roaa, A. Rawan, ArASL: Arabic Alphabets Sign Language Dataset, Data in Brief, Volume 23, 2019, 103777.
- [3] Barnett S., Communication with deaf and hard-of-hearing people: a guide for medical education. Acad Med. 2002 Jul;77(7):694-700. doi: 10.1097/00001888-200207000-00009.
- [4] H. Cheng, Lu Yang, and Z. Liu, "Survey on 3D Hand Gesture Recognition", IEEE Transactions on circuits and systems for video technology, volume 26, Issue 9, September 2016, pages 1659-1673.
- [5] Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.
- [6] A. Mittal, P. Kumar, P.P. Roy, R. Balasubramanian, B.B. Chaudhuri, A modified LSTM model for continuous sign language recognition using leap motion, IEEE Sensor. J. 19 (16) (Aug. 2019) 7056–7063.
- [7] Hasan, M.M.; Mishra, P.K. Hand gesture modeling and recognition using geometric features: A review. Can. J. Image Process. Comput. Vis. 2012, 3, 12–26.
- [8] P. Garg, N. Aggarwal, Sofat, S. Vision-based hand gesture recognition. World Acad. Sci. Eng. Technol. 2009, 49, 972–977.
- [9] N. A. Ibraheem, R.Z. Khan, Survey on various gesture recognition technologies and techniques. Int. J. Comput. Appl. 2012, 50, 38–44.
- [10] S. Mitra, and T. Acharya. "Gesture Recognition: A Survey" IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol. 37 (3), pp. 311- 324, DOI: 10.1109/TSMCC. 2007. 893280.
- [11] Rajat Sharma et al. Sign Language Gesture Recognition, International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 7, Issue 2, June, pp. 14-19, 2020.
- [12] M. M. Kamruzzaman, Arabic Sign Language Recognition and Generating Arabic Speech Using Convolutional Neural Network, JOUR, SP - 3685614, VL - 2020.
- [13] J. S. Sonkusare, N. B. Chopade, Ravindra Sor and Sunil.L. Tade, "A Review on Hand Gesture Recognition System", International Conference on Computing Communication Control and Automation 2015, pages 790 794.
- [14] S.. K. Yewell and P.. K. Bharné, "Hand Gesture Recognition Using Different Algorithms Based on Artificial Neural Network", IEEE International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), 2016, pages 671-675.
- [15] P. Dhiman, M. Dhanda "Video Segmentation using FCM Algorithm", International Journal of Engineering Trends and Technology (IJETT), V36(2), 106-110 June 2016. www.ijettjournal.org.
- [16] Z. Zhang, L. Yang, Yefeng Zheng, Chapter 8 - Multimodal medical volumes translation and segmentation with a generative adversarial network, Editor(s): S. Kevin Zhou, Daniel Rueckert, Gabor Fichtinger, Handbook of Medical Image Computing and Computer-Assisted Intervention, Academic Press, 2020, Pages 183-204.
- [17] J. Appenrodt, A. Al-Hamadi, B. Michaelis (2010) Data gathering for gesture recognition systems based on single color-, stereo color- and thermal cameras. Int J Signal Process Image Process Pattern Recognit 3:37–50.
- [18] P. Subha Rajan, G. Balakrishnan, "Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People", International Conference on Communication Technology and System Design, 2011, pp. 861-868.
- [19] P. Dhiman, M. Dhanda, A Review on Various Techniques of Video Segmentation. IJIRST–International Journal for Innovative Research in Science & Technology. 2016 Apr;2(11).
- [20] H. Badi, Recent methods in vision-based hand gesture recognition. International Journal of Data Science and Analytics. 2016 Jul 1;1(2):77 87. Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. "librosa: Audio and music signal analysis in python." In Proceedings of the 14th python in science conference, pp. 18-25. 2015. [9] Giannakopoulos, Theodoros".