



Review On Hand Gesture Recognition Using Machine Learning Techniques

¹Pravin Teli, ²Mayank Dhamal, ³Prajakta Kadam, ⁴Siddhesh Chinchmalatpure, ⁵Ashwini Pandagale

^{1,2,3,4}Final Year UG Student, ⁵Assistant Professor

Department of Computer Engineering, Ajeenkya D. Y. Patil School of Engineering, Lohegoan, Savitribai Phule Pune University, Pune, Maharashtra, India

¹sangitатели17@gmail.com, ²dhamal15mayank@gmail.com, ³prajaktakadam0610@gmail.com,
⁴siddheshchinchmalatpure@gmail.com, ⁵ashwini.pandagale@dypic.in

ABSTRACT-

Hand gestures are a form of non-verbal communication that can be used in a variety of areas, such as Deaf-mute communication, robot control, human-computer interaction (HCI), home automation and medical applications. Research papers based on hand gestures were adopted for studying for many different techniques, including those based on instrumented sensor technology and Computer vision. In other words, the hand signal can be divided into many categories such as posture and gestures as well as dynamic and static or a mixture of both. This document focuses on an overview of the hand gesture literature. techniques and presents their advantages and limitations under other circumstances. In addition, the performance of these methods is tabulated, with a focus on computer vision techniques dealing with similarity and difference points, manual segmentation technique used, classification algorithms and disadvantages, number and types of gestures, data set used, detection range (distance) and camera type used. This document is a high-level overview of hand gesture methods with a brief discussion of some possible applications.

Keywords- Hand Posture, Hand Gesture, Human Computer Interaction (HCI), Neural Network, Machine Learning, Image Processing .

I.INTRODUCTION

Based on the Who Health Organization (WHO) statistics, there are over 460 million individuals with hearing disorder disability. This variety has accrued to 900 million by 2050. in line with the world federation of deaf people, there are regarding three hundred sign languages (SLs) used round the world. Linguistic communication is that the bridge for communication between deaf and traditional people. it's outlined as a mode of interaction for the onerous of hearing people through a group of hand gestures, postures, movements, and facial expressions or movements that correspond to letters and words in our real life. to speak with deaf people, an interpreter is required to translate real-world words and sentences.[1] So, deaf individuals can perceive us or vice versa. Unfortunately, deaf people don't have a written type and have an enormous lack of electronic resources. Hand gesture recognition for human pc interaction is a district of active analysis in Artificial Intelligence. One in every of its primary goals is to make systems, which may determine specific gestures and use them to convey info or to manage a device. Though, gestures got to be modelled within the spatial and temporal domains, wherever a hand posture is that the static structure of the hand and a gesture is that the dynamic movement of the hand. There are essentially 2 sorts of approaches for hand gesture recognition: vision-based approaches and information glove approaches. This work focus is on making a vision-based system able to do period of time linguistic communication recognition. the rationale for selecting a system supported vision relates to the fact that it provides a less complicated and additional intuitive approach of communication between somebody's and a computer [2]. Being hand-pose one in every of the foremost vital communication tools in humans' daily life, and with the continual advances of image and video process techniques, analysis on human-machine interaction through gesture recognition light-emitting diode to the utilization of such technology terribly} very broad vary of applications, like bit screens, computer game consoles, virtual reality, medical applications, and linguistic communication recognition. though sign language is that the most natural approach of exchanging info among deaf individuals it's been determined that they're facing difficulties with traditional people interaction. linguistic communication consists of vocabulary of signs in barely a similar way as auditory communication consists of a vocabulary of words. Sign languages don't seem to be commonplace and universal and the grammars dissent from country to country [3].

II.MOTIVATION

Linguistic communication could be a system of interaction mistreatment visual gestures and signs as utilized by individuals with incapacity to listen to or speak. The human community has used this gesture-based language for his or her interaction once suggests that like audio, writing or writing is tough, however visual means are available. Gesture primarily based communication normally is used by someone after they would like to not speak, except for the deaf and dumb community, it is the sole method of communication. Human actions are a major indication of human interaction and of human activities otherwise referred to as non-verbal communication. The deaf and dumb community faces a great deal of issues whereas communication with others as it's not perpetually essential that the opposite individual or a gathering of people understand gesture-based communication. So, to beat this hindrance of communication between the two communities we tend to try to propose a system which can concentrate on mistreatment machine learning techniques to acknowledge an individual's linguistic communication gesture and consequently convert it to speech in order that others can understand and communicate with them.

III.LITERATURE SURVEY

Extensive studies have long past into the improvement of generation to attain gesture to speech conversion which may be achieved the use of one in every of 3 approaches: imaginative and prescient primarily based totally, sensor-primarily based totally or hybrid-primarily based totally that's a combination among imaginative and prescient-primarily based totally and glove-primarily based totally structures however it isn't always broadly used. These structures are followed to seize hand configurations wished for the Sign Language Recognition (SLR) [2]. Sensors-primarily based totally structures are essentially simply sensors established on positive kind of gloves used to gather enter facts for the gadget to begin the popularity manner. Various sorts of sensors are used to degree the bend angles for hands, the kidnapping among hands and the orientation of the wrist. The primary gain of glove-based structures over imaginative and prescient-primarily based totally structures is that gloves are greater immediately forward. They take away the want to manner uncooked facts into significant values with the aid of using at once reporting applicable and required enter facts in phrases of voltage values or resistances as in flex sensors' case to the computing tool. But on the alternative hand, imaginative and prescient-primarily based totally structures want to use unique monitoring and characteristic extraction algorithms to uncooked video streams, increasing the computational overhead [2]. Finger bending detection primary motion in SL associated with the hands is bending [4]. Flex sensors are the maximum utilized by researchers and builders consistent with the preceding papers. A flex sensor determines the quantity of finger curvature, primary based totally on resistive carbon elements. Many preceding works used flex sensors with hand orientation sensors to transform SL to textual content and speech. For instance, in [3], the proposed gadget interprets the Pakistan Sign Language (PSL) to speech with the assist of flex sensors, accelerometer, gyroscope and phone sensors, reaching a performance of 93.4% with alphabets and numbers. A comparable version become proposed in [4], a hand glove together with a few sensors and circuitry positioned at the arm of the deaf person, aiming to locate the extrude in gestures and convert them into textual content or speech. In [5], the authors used the equal glove version with ML to locate ASL, then show the gesture on an android software with accuracy 94.23%. Last however now no longer least, authors in [6] proposed a easy version of 5 flex sensors in aggregate with the gyroscope main to a success and correct translation of ASL to speech. In addition to the use of Python to construct a dependable and correct gadget for the messaging carrier to make the gadget person friendly. Some of the alternative used sensors in SLR are: Optical sensors which might be digital detectors that convert mild, or a extrude in mild into a digital signal, they may be used to degree the attitude of the finger curvature so that it will decide its form with the aid of using the quantity of mild passing through the channel which relies upon at the optic generation [7], so that after the finger is immediately, the density of the obtained mild may be very significant. The gain of this sensor is that it's miles appropriate for handicapped people whose hands can slightly carry out even very small motions. Tactile sensor is a sturdy polymer-thick movie tool whose resistance modifications whilst a pressure is applied, so with the aid of using calculating the quantity of pressure positioned at the finger, it permits us to decide whether or not the finger is curved or immediately. Capacitive contact sensors have emerged as one of the desired approaches for detecting SL. It consists of terminals that feature as emitting and receiving electrodes. In [8], the proposed version consisted of 8 sensors, at the tip of every finger and among the finger joints so that it will understand the posture of man or woman finger segments, to have greater unique analysing and with a Raspberry Pi because the processing unit. This version advantages from the contact sensor's virtual gain to translate all the 26 alphabets and 10 digits of ASL with an accuracy over 92%.

Vision-primarily based totally hand gesture popularity is a place of lively modern studies in laptop imaginative and prescient and ML. It calls for speedy and extraordinarily sturdy hand recording detection and gesture popularity in actual time [9]. Those structures use cameras as number one gear to achieve the important enter facts, MATLAB has been normally used to carry out photo processing with the aid of using the use of photo segmentation and with the assist of characteristic extraction algorithms, the gadget acknowledges the photo of the hand gesture captured with the aid of using the digital digicam and converts it to speech from a listing of pre-recorded tracks. After photo processing, ML and Template Matching may be used to get the favoured textual content to speech conversion. Image processing is an umbrella time period for lots capabilities that examine photos or convert one illustration of a photo into another. Nowadays photo processing is achieved withinside the virtual area changing the analog processing used withinside the past. In 2007, an automated imaginative and prescient-primarily based totally gadget with a custom designed colouration glove become presented, in which the colouration eases the extraction manner. At the end, an Artificial Neural Network (ANN) is used for classifying and translating gestures having an ordinary popularity charge over 90%. In 2008, used another technique primarily based totally at the pores and skin colouration, indicating that it is able to be used withinside the segmentation manner for isolating the human from a cluttered background. There is masses of preceding paintings that used this equal technique however they handiest range withinside the very last popularity accuracy. Although imaginative and prescient-

primarily based totally structures have a almost ideal popularity accuracy, their boundaries cannot be dismissed. The want of excessive specification digital digicam, confined subject of view of the shooting tool, excessive computational fees and the want for more than one camera to achieve sturdy effects are the primary boundaries to such structures.

IV.METHODOLOGY

Different methods and techniques of detecting hand gestures

1.Hand gestures based on the instrumented glove approach

Wearable glove-based sensors can be used to detect the movement and position of the hands. Additionally, through the use of glove-mounted sensors, they can easily provide accurate coordinates of palm and finger positions, orientation, and configurations [8]. However, this approach requires the user to be physically connected to the computer [11], which blocks the ease of interaction between the user and the computer. In addition, the price of these devices is quite high [13,14]. However, the modern glove-based approach uses touch technology, which is a more promising technology and is considered an industrial-grade haptic technology. haptic feedback that lets the user feel the shape, texture, movement, and weight of a virtual object through the use of microfluidic technology.

2. Hand gestures based on computer vision

The camera vision-based sensor is a common, suitable and applicable technique because it provides contactless communication between humans and machines [16]. Different configurations of cameras can be utilized and used, such as monocular, fisheye, TOF and IR [20]. However, this technique involves several challenges, including lighting variation, background noise, the effect of occlusions, complex background, processing time traded against resolution and frame rate and foreground or background objects presenting the same skin colour tone or otherwise appearing as hands [17,21].

2.1 Colour-Based Detection Using a Glove Marker

This method uses a camera to track the movement of the hand wearing a glove with markers of different colours. This method is used for interacting with 3D models, which allows some processing like zooming, moving, drawing and typing with a virtual keyboard with brilliant flexibility [6]. The colours on the glove allow the camera sensor to track and recognize the position of the palm and fingers, which can extract the geometric pattern of the hand shape [13,16]. The advantages of this method are its ease of use and its low price compared to glove sensor data [22]. However, it still requires the use of coloured gloves and limits the level of naturalness and spontaneity. Interaction with the HCI [19].

Skin colour detection is one of the maximum famous techniques for hand segmentation and is utilized in a huge variety of applications, which include item classification, degraded photo recovery, person motion tracking, video observation, HCI applications, facial recognition, hand segmentation and gesture identification. Skin colour detection has been executed the usage of techniques. The first approach is pixel primarily based totally pores and skin detection, wherein every pixel in a picture is classed into pores and skin or not, individually from its neighbour. The 2nd approach is vicinity pores and skin detection, wherein the pores and skin pixels are spatially processed primarily based totally on records which include depth and texture. Colour area may be used as a mathematical version to symbolize picture colour records. Several colour areas may be used in keeping with the utility kind which include virtual graphics, picture process applications, TV transmission and alertness of laptop imaginative and prescient techniques [19,16]

2.2 Recognition based on 3D models

The 3D model essentially depends on the 3D kinematic hand model, which has a high degree of freedom, with the estimation of the hand parameter obtained by comparing the input image with the two -dimensional appearance is projected by three-dimensional hand model. In addition, the 3D model represents the human hand characteristic as a pose estimate by forming a 3D or skeletal or solid model identical to the user's hand. Where the 3D model parameter is updated through the matching process. Where the depth parameter is added to the model to increase accuracy [20].

A study by Tekin et al. [24] proposed a new model for understanding interactions between hands and 3D objects using a single RGB image, where the single image is continuously trained with a neural network and displays a joint estimate of hand and object poses in 3D. Wan et al. [25] proposed estimating 3D hand pose from a single depth map using a self-monitoring neural network by approximating the surface of the hand with an array of spheres. based on simple RGB image. Where Graph Convolutional Neural Network (Graph CNN) was used to reconstruct a full 3D mesh for the surface of the hand. Another study by Taylor et al. [26] proposed a new human hand tracking system Combines the surface model with a new power function that continuously optimizes together via pose and matching and can track the hand several meters away from the camera. Malik et al. [27] proposed a novel CNN-based self-learning algorithm to segment the hand from a raw depth image and estimate the 3D hand pose estimation, including the structural limitations of the hand skeleton. Researcher presented a novel method to track a complex deformable object interacting with a hand. Chen et al. [28] Proposed SO (Hand Net) self-organizing hand network that achieves 3D hand pose estimation through semi-supervised learning. Where the end-to-end regression method was used to estimate 3D hand pose using a single depth image. Another study by Ge et al. [29] proposed a point-to-point regression method for 3D hand pose estimation in single-depth images. Wu et al. [30] Novelty Proposal for estimating hand pose by a single depth image that combines the detection-based method and the regression-based method to improve accuracy. Cai et al. [31] present a way to fit a sparsely labelled real dataset from a fully annotated synthetic dataset using low-cost depth imaging and using only RGB inputs for 3D co-predictions.

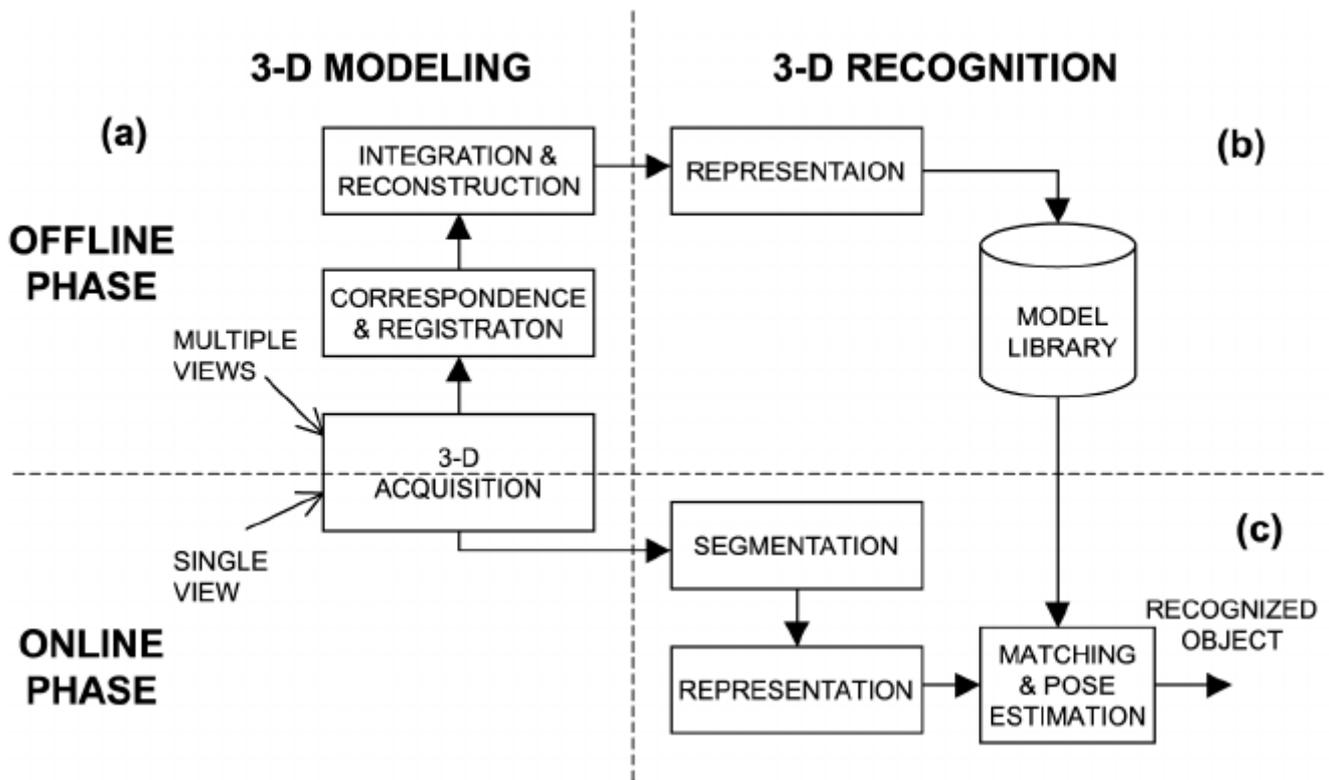


Fig 1. The 3D model-based recognition framework [41]

2.4 Skeleton-Based Image Recognition

The skeleton-based recognition specifies model parameters which may improve the detection of complicated options [16]. wherever the assorted representations of skeleton information for the hand model is used for classification, it describes geometric attributes and constraint and straightforward interprets features and correlations of data, so as to target geometric and data point features. the foremost common feature used is that the joint orientation, the house between joints, the skeletal joint location and degree of angle between joints and trajectories and curvature of the joints. Table four presents a group of analysis papers that use completely different segmentation techniques supported skeletal recognition to observe ROI. Hand segmentation victimization the depth detector of the Kinect camera, followed by location of the fingertips using 3D connections, geometrician distance, and geodesic distance over hand skeleton pixels to supply enlarged accuracy was projected in [18]. a brand-new 3D hand gesture recognition approach based on a deep learning model using parallel convolutional neural networks (CNN) to method hand skeleton joints' positions was introduced in [24], the proposed system incorporates a limitation wherever it works solely with complete sequence. The best viewpoint was calculable and also the purpose cloud of gesture remodeled employing a curve skeleton to specify topology, then Laplacian-based contraction was applied to specify the skeleton points in [29]. wherever the Hungarian algorithmic program was applied to calculate the match variant the skeleton point set, however the joint following data nonheritable by Kinect isn't correct enough that provides a result with constant vibration. a unique methodology supported skeletal options extracted from RGB recorded video of sign language, which presents difficulties to extracting accurate skeletal information as a result of occlusions, was offered in [21]. A dynamic hand gesture victimization depth and skeletal dataset for a skeleton-based approach was bestowed in [22], wherever supervised learning (SVM) used for classification with a linear kernel. Another dynamic hand gesture recognition using Kinect detector depth information for acquisition and segmentation that accustomed extract orientation feature, where the support vector machine (SVM) algorithmic program and HMM was utilized for classification and recognition to judge system performance where the SVM bring a decent result than HMM in some specification such pass on time, average recognition rate, was projected in [23]. A hybrid methodology for hand segmentation supported depth and colour information nonheritable by the Kinect detector with the facilitate of skeletal data were projected in [8,10]. during this methodology, the image threshold is applied to the depth frame and also the super-pixel segmentation method is employed to extract the hand from the colour frame, then the 2 results are combined for strong segmentation.

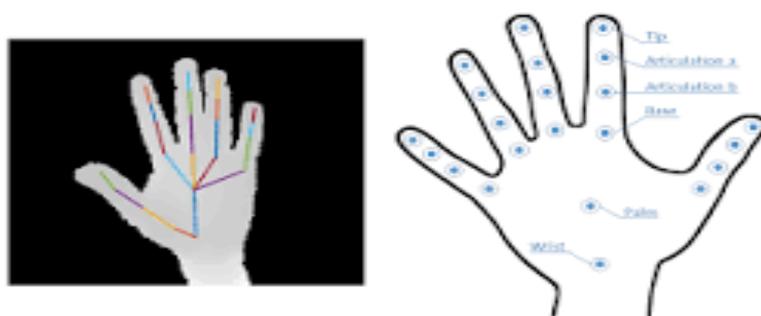


Fig 2. Example of skeleton recognition [40]

3. Recognition based on Deep Learning

Artificial Intelligence provides a good and reliable technique used in a variety of modern applications due to the use of a functional learning principle. Deep learning uses multiple layers to learn data and gives good prediction result. Most of the challenges this technique faces is a data set required to learn the algorithm that can affect time processing. The researcher proposed seven popular hand gestures, which they captured with a mobile camera, generating 24,698 frames. Feature extraction and Adaptive Deep Convolutional Neural Network (ADCNN) were used for hand classification. The experiment evaluates the result of 100% training data and 99% test data with a runtime of 15,598 s [32]. While other proposed systems used a webcam to track the hand. Then they used the technique and morphology of skin colour (Y-Cb-Cr colour space) to remove the background. Additionally, Kernel Correlation Filters (KCFs) are used to track the ROI. The resulting image goes into a Deep Convolutional Neural Network (CNN). Where the CNN model is used to compare the performance of two mods from Alex Net and VGG Net. Detection rate for both training data and test data is 99.90% and 95.61% in [33]. A new method based on Deep Convolutional Neural Networks, in which the resized image is fed directly into the network, ignoring the segmentation and recognition steps, to classify hand gestures directly. The system works in real time and returns a result with a plain background 97.1% and with a complex background 85.3% in [33]. The depth image generated by the Kinect sensor was used to segment the colour image, then the skin colour was modeled in combination with a convolution neural network, applying the inverse error propagation algorithm to modify the threshold and values. for the neural network. The SVM classification algorithm was added to the network to improve the result in [34]. Other the research study used Gaussian Mixture Modeling (GMM) to filter out non-skin colours from an image, which trained CNN to recognize seven hand gestures, with an average recognition rate of 95.96% in [35]. The following the proposed system uses an action classifier based on a long-term recurrent convolutional network in which multiple sampled frames of the recorded video sequence are fed into the network. To extract the representative frames, representative frames, the deconvolution neural network based on semantic segmentation is used. Image tessellation and binary tessellation are used to train the deconvolution network in [36]. [37] proposes a two-channel convolutional neural network (DC-CNN) where the original image is pre-processed to detect the edge of the hand before it is fed into the network. Each dual channel CNN has a separate softmax and weight classifier used to classify the output results. The proposed system provides a 98.02% detection rate. Finally, a new neural network based on SPD multiple learning for skeletal-based hand gesture recognition proposed by [38].

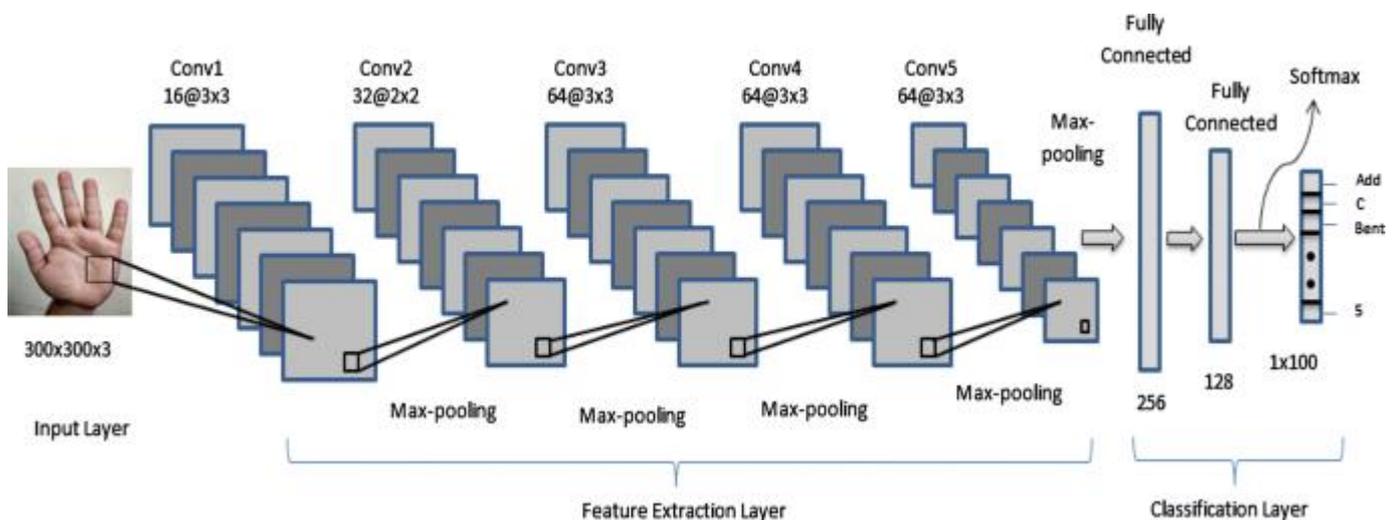


Fig 3. Deep learning convolutional neural network architecture.[39]

V.CONCLUSION

Hand gesture recognition fixes a bug in the interaction systems. Controlling by hand is more natural, simpler, more flexible and cheaper, and there is no need to troubleshoot errors caused by hardware devices because none are required. From the previous sections it was clear that there is a lot to do Trying to develop reliable and robust algorithms using a camera sensor has certain properties to find common problems and get a reliable result. However, each of the above techniques has its pros and cons and can perform well on some challenges while falling short on others.

VI. REFERENCE

- [1] Ahmed Sultan, Walied Makram, Mohammed Kayed, "Sign language identification and recognition: A comparative study", in Open Computer Science 2022; 12: 191–210
- [2] Nakul Nagpal,Dr. Arun Mitra.,Dr. Pankaj Agrawal, "Design Issue and Proposed Implementation of Communication Aid for Deaf & Dumb People", International Journal on Recent and Innovation Trends in Computing and Communication ,Volume: 3 Issue: 5,pp- 147 – 149. published by 2021
- [3] Dr.L.Priya, A.Sathya and S. Kanaga Suba Raja, "Indian and English Language to Sign Language Translator- an Automated Portable Two Way Communicator for Bridging Normal and Deprived Ones ", in 2020 2nd International Conference on Power, Energy, Control and Transmission Systems
- [4] Okan Köpüklü, Ahmet Gunduz, Neslihan Kose , and Gerhard Rigoll, "Online Dynamic Hand Gesture Recognition Including Efficiency Analysis",IEEE TRANSACTIONS ON BIOMETRICS, BEHAVIOR, AND IDENTITY SCIENCE, VOL. 2, NO. 2, APRIL 2020.

- [5] Jestin Joy, Kannan Balakrishnan, Sreeraj M, "SignQuiz: A Quiz based tool for learning fingerspelled signs in Indian Sign Language using ASLR", in IEEE Access (Volume: 7) (2019)
- [6] Frank M. Shipman and Caio D.D. Monteiro, "Crawling and Classification Strategies for Generating a MultiLanguage Corpus of Sign Language Video", in 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)
- [7] Jamal Riffi, Hamid Tairi, Mohamed Adnane Mahraz, "Hand gesture recognition based on convexity approach and background subtraction", in IEEE International Conference on Instrumentation Engineering (ICPCSI-2018)
- [8] Aditya Das, Shantanu Gawde, Khyati Suratwala, and Dhananjay Kalbande. Sign language recognition using deep learning on custom processed static gesture images. In 2018 International Conference on Smart City and Emerging Technology (ICSCET), pages 1–6. IEEE, 2018.
- [9] Kshitij Bantupalli and Ying Xie. American sign language recognition using deep learning and computer vision. In 2018 IEEE International Conference on Big Data (Big Data), pages 4896–4899. IEEE, 2018.
- [10] Akash. Asl alphabet. [Online]. Available: <https://www.kaggle.com/grassknotted/asl-alphabet>, 2018.
- [11] Chandandeep Kaur, Nivit Gill, "An Automated System for Indian Sign Language Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, published by 2017.
- [12] Tu"lay Karayilan and Ozan Kılıc, . Sign language recognition. In 2017 International Conference on Computer Science and Engineering (UBMK), pages 1122–1126. IEEE, 2017.
- [13] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a Convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp.16, doi:10.1109/ICEngTechnol.2017.8308186.
- [14] A. S. Nikam and A. G. Ambekar. Sign language recognition using image-based hand gesture recognition techniques. In 2016 Online International Conference on Green Engineering and Technologies (IC-GET), pages 1– 5, 2016.
- [15] A. Kumar, K. Thankachan, and M. M. Dominic. Sign language recognition. In 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pages 422–428, 2016.
- [16] Mandeep Kaur Ahuja, Amardeep Singh, "Hand Gesture Recognition Using PCA", International Journal of Computer Science Engineering and Technology (IJCSET), Volume 5, Issue 7, pp. 267-27, July 2015.
- [17] Sagar P.More, Prof. Abdul Sattar, "Hand gesture recognition system for dumb people", Published 2015
- [18] Pratibha Pandey, Vinay Jain, "Hand Gesture Recognition for Sign Language Recognition: A Review", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 3, March 2015
- [19] Neelam K. Gilorkar, Manisha M. Ingle, "A Review on Feature Extraction for Indian and American Sign Language", International Journal of Computer Science and Information Technologies, Volume 5 (1), pp 314-318, 2014.
- [20] Kanchan Dabre and Surekha Dholay. Machine learning model for sign language interpretation using webcam images. In 2014
- [21] Ashish Sethi, Kuldeep Kumar, Bhaskara Rao ,Krishnan R, "Sign Pro-An Application Suite for Deaf and Dumb", IJCSET , Volume 2, Issue 5, pp-1203-1206, May 2012.
- [22] International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), pages 317–321. IEEE, 2014.
- [23] Oana Saman & Loredana Stanciu, "Hand Mobility Recovery Through Game Controlled by Gestures", in IEEE 13th International Symposium on Applied Computational Intelligence and Informatics.
- [24] Wan, C.; Probst, T.; Gool, L. Van; Yao, A. Self-supervised 3d hand pose estimation through training by fitting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10853–10862.
- [25] Ge, L.; Ren, Z.; Li, Y.; Xue, Z.; Wang, Y.; Cai, J.; Yuan, J. 3d hand shape and pose estimation from a single rgb image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10833–10842.
- [26] Taylor, J.; Bordeaux, L.; Cashman, T.; Corish, B.; Keskin, C.; Sharp, T.; Soto, E.; Sweeney, D.; Valentin, J.; Luff, B. Efficient and precise interactive hand tracking through joint, continuous optimization of pose and correspondences. ACM Trans. Graph. 2016, 35, 1–12.
- [27] Malik, J.; Elhayek, A.; Stricker, D. Structure-aware 3D hand pose regression from a single depth image. In Proceedings of the International Conference on Virtual Reality and Augmented Reality, London, UK, 22– 23 October 2018; pp. 3–17.
- [28] Tsoli, A.; Argyros, A.A. Joint 3D tracking of a deformable object in interaction with a hand. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 484–500.
- [29] Chen, Y.; Tu, Z.; Ge, L.; Zhang, D.; Chen, R.; Yuan, J. So-handnet: Self-organizing network for 3d hand pose estimation with semi-supervised learning. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, South Korea, 27 October–2 November 2019; pp. 6961–6970. 87. Ge, L.; Ren, Z.; Yuan, J. Point-to-point regression pointnet for 3d hand pose estimation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 475– 491.
- [30] Wu, X.; Finnegan, D.; O'Neill, E.; Yang, Y.-L. Handmap: Robust hand pose estimation via intermediate dense guidance map supervision. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 237–253.

- [31] Cai, Y.; Ge, L.; Cai, J.; Yuan, J. Weakly-supervised 3d hand pose estimation from monocular rgb images. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 666–682.
- [32] Alnaim, N.; Abbod, M.; Albar, A. Hand Gesture Recognition Using Convolutional Neural Network for People Who Have Experienced A Stroke. In Proceedings of the 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 11–13 October 2019; pp. 1–6.
- [33] Chung, H.; Chung, Y.; Tsai, W. An efficient hand gesture recognition system based on deep CNN. In Proceedings of the 2019 IEEE International Conference on Industrial Technology (ICIT), Melbourne, Australia, 13–15 February 2019; pp. 853–858.
- [34] Bao, P.; Maqueda, A.I.; del-Blanco, C.R.; García, N. Tiny hand gesture recognition without localization via a deep convolutional network. *IEEE Trans. Consum. Electron.* 2017, 63, 251–257.
- [35] Li, G.; Tang, H.; Sun, Y.; Kong, J.; Jiang, G.; Jiang, D.; Tao, B.; Xu, S.; Liu, H. Hand gesture recognition based on convolution neural network. *Cluster Comput.* 2019, 22, 2719–2729. *J. Imaging* 2020, 6, 73 29 of 29
- [36] Lin, H.-I.; Hsu, M.-H.; Chen, W.-K. Human hand gesture recognition using a convolution neural network. In Proceedings of the 2014 IEEE International Conference on Automation Science and Engineering (CASE), Taipei, Taiwan, 18–22 August 2014; pp. 1038–1043.
- [37] John, V.; Boyali, A.; Mita, S.; Imanishi, M.; Sanma, N. Deep learning-based fast hand gesture recognition using representative frames. In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, QLD, Australia, 30 November–2 December 2016; pp. 1–8.
- [38] Wu, X.Y. A hand gesture recognition algorithm based on DC-CNN. *Multimed. Tools Appl.* 2019, 1–13.
- [39] Wadhawan, A., Kumar, P. Deep learning-based sign language recognition system for static signs. *Neural Comput & Applic* 32, 7957–7968 (2020)
- [40] Quentin de Smedt, Hazem Wannous, Jean-Philippe Vandeborre. Skeleton-Based Dynamic Hand Gesture Recognition. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016 IEEE Conference on, Jun 2016, Las Vegas, United States. pp.1206 - 1214, [ff10.1109/CVPRW.2016.153ff](https://doi.org/10.1109/CVPRW.2016.153ff). [ffhal01535152f](https://doi.org/10.1109/CVPRW.2016.153ff)
- [41] Mian, Ajmal & Bennamoun, Mohammed & Owens, Robyn. (2004). Automated 3D model-based free-form object recognition. *Sensor Review*. 24. 206-215. [10.1108/02602280410525995](https://doi.org/10.1108/02602280410525995).