International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 6, 2025
https://theaspd.com/index.php

Image Enhancement, Restoration and Segmentation Methods for Vision Based Sign Language Recognition

Shivani Chauhan[1], Second Sonika Kandari[2], Kamal Kant Verma[3]
[1,2,3]School of Engineering and Technology, Shri Guru Ram Rai University Dehradun, India,
School of Engineering and Technology, Shri Guru Ram Rai University Dehradun, India,
School of Computer Science and Engineering, IILM University Greater Noida, India
shivichauhan687@gmail.com, dean.set@sgrru.ac.in, dr.kamalverma83@gmail.com

Abstract
SLR (Sign Language Recognition) plays an essential part in increasing the communication for the deaf and speechless people. Because Sign Language is the sole language used by the deaf and speechless to exchange messages with one another. Sign Language is entirely distinct from generally spoken languages because It possesses its own grammar. But still SLR, or sign language recognition, is the most difficult area for practitioners and researchers. The SLR is mostly required because of its ability to remove the communication barrier for deaf and dumb people. To overcome this problem this research presents the comprehensive review of different studies based on sign language which includes various techniques of Artificial Intelligence (AI) used for vision based SLR for human understanding that have employed in recent past which will help to develop a model to fill the communication gap. The main objective of this work is comparison of image enhancement techniques (Histogram Equalization, Contrast Limited Adaptive Histogram Equalization), image restoration techniques (Mean Filter, Gaussian Filter) and image segmentation techniques (Edge Based, Region Based and Cluster Based) by applying on images and also discuss their scope of area and limitations. Important SLR challenges like database limitations and regional sign languages have also been covered. This paper aim is to help researcher to know about presented methods of sign language translations along with new findings, so that a new system can be developed with more accuracy, easy to use and comfortable for deaf people. Remaining part of this paper will contain the literature review and different approaches of sign language recognition and different techniques used for vision-based approach of sign language recognition.
Keyword: AI, SLR, vision based, Image Enhancement, Restoration and Segmentations.

1. INTRODUCTION

AI is the area of computer science in which computer system is able to perform tasks like object detection, speech recognition, face recognition etc. Machines also have the capability of thinking and learning. An important branch of AI is NLP (Natural Language Processing), used to collaborate between human and machine. The objective of NLP is to understand the human languages. AI is also used for Sign Recognition. Signs are a medium of communication used to share information among people[1]. Specially for the people with disability. According to the survey 15-20% population of the world are deaf and dumb. Like normal people these people cannot communicate with others. People that are deaf or dumb utilise sign language as a sharing the meaningful information. But normal people cannot understand the sign language, because sign language is not an international language. Sign Language is a visual based language which is based on three main parameters hand shape, facial expression as well as physical motion. Recognition of Sign Language is a process of understanding the gesture done by signer and converting in the form of speech or text which understood by normal people too[2]. SLR bridges the communication gap between traditional and people with disability (deaf and dumb). In other words, SLR system works as bridge between normal people and disable people. Sign Language Recognition Technique is mainly divided into five categories. First, acquiring 2) The pre-processing stage 3. Division Feature extraction is step four, while classification is step five. The two most popular approaches of SLR are glove- and vision-based[3].

In Vision based approach gesture is captured by camera in the form of video. Then the video goes for pre-processing to eliminate noise. After that segmentation is done[4]. In glove-based approach tracking of the signs given by signer is done with sensor-based gloves. In glove-based approach signer should wear the gloves. The problem with glove-based approach is that it cannot track the non-manual signs like face expression. In this approach segmentation is not required[5].

According to the survey 2021 conducted by Over 70 million persons are members of the "World Federation of Deaf" in the world. There are 300 plus sign languages are available in the world for e.g., ASL, Indian Sign Language, Deaf and dumb persons use Chinese Sign Language, among other sign languages, to communicate with others[2]. Sign language is a natural language different from spoken language. It has the same properties as spoken language but with grammar is different from English. The important parameters of sign language are hand shape, face expression, movement and non-manual signals. Deaf and Dumb people use the sign language to share the information. But normal people cannot understand the signs used by signer[4]. To understand the meaning of signs and convert into its equivalent text or speech "Sign Language Recognition "Technique is used. Sign Language varies country to country: In America American Sign Language used, In China Chinese Sign Language used, In India Indian Sign Language used[4]. ISL is in Primary phase of development. Unlike American Sign Language and Chinese Sign Language, ISL uses two hands to make gesture.

Recently AI (Artificial Intelligence) has achieved many advancements on Sign Language Recognition for interpreting operations[4]. Now AI based intelligent SLR system is using in many applications like Robotics, Virtual Reality, Online hand tracking, computer human interaction etc[6]. Following Figure shows the alphabets used in Indian Sign Language and American Sign Language.

Fig.1 ISL (Indian Sign Language)

Fig.2 American Sign Language alphabets (A-Z & a-z)

Motivation: This paper is motivated by increasing the need of recognition system for communication for those who are dumb or deaf as communication is only the way which is used by human to express their needs, feelings and thoughts to people mostly by speech. But this world also has the people with some disability like deaf and dumb people. For these people communication is not possible by speech. Instead of many advancements in technology there are some issues which lack the integration of SLR systems into daily life such as lack of databases, different sign languages. The goal of this research is to create a model that can be applied in the actual world to close the communication gap quickly and accurately.

1.A Need of Sign Language Recognition

Because 15–25% of people can't have normal conversations. SLR is essential for the deaf and hard-of-hearing population to communicate more successfully and be more inclusive. The SLR is required to overcome the following problem:

Communication for Deaf and Hard of Hearing People: Many people who are Deaf or Hard of Hearing, sign language is their main form of communication. By bridging the communication gap and encouraging inclusivity, sign language recognition technology enables them to speak more effectively with others who might not understand sign language.

Accessibility: The recognition of sign language improves the lives of the deaf and hard of hearing in many areas. It can be used to improve communication in a variety of settings, including public announcements, customer service, healthcare, and education. It helps the deaf population access information and services by delivering real-time translation between sign language and spoken or written language.

Education: Deaf and hard-of-hearing people can benefit significantly from learning sign language. It can be incorporated into instructional technologies and resources to support learning, enabling deaf students to access educational materials, take part in class discussions, and interact with peers and teachers.

The development of assistive technology and equipment that improve the daily lives of the deaf and hard of hearing can make use of sign language recognition. In order to communicate with others who do not understand sign language, it can be utilised to develop sign language interpreters or virtual avatars that can translate signs into spoken or written language.

Research and Development: In the fields of sign language linguistics, cognitive science, and human-computer interaction, sign language recognition has a considerable impact. Researchers can progress linguistic research and create more complex sign language recognition systems by analysing sign language data to acquire insights into the structure, syntax, and semantics of sign languages.

In general, sign language recognition technology has the ability to remove obstacles to communication, empower people who are deaf or hard of hearing, and advance inclusivity and accessibility in a variety of spheres of life.

1.B CHALLENGES

Because SLR incorporates so many distinct gestures and facial expressions, it is challenging and complex. In addition to the manual elements, lip shapes and brow positions set similar signs apart; several manual signs, for example, seem to be in the same position. But with the help of facial expression and lip movement, these can be identified. Face emotions, body movements, palm posture, hand gestures, shape, position, and orientation all contribute to sign language. SLR performance is significantly impacted by following parameters[7].

Variability: Across areas, cultures, and individuals, sign languages differ significantly. The sign languages of many nations each have their own vocabulary, grammar, and syntax, such as the American Sign Language (ASL) and British Sign Language (BSL). Creating a universal recognition system that takes into account all these variables is difficult since sign language users may have their own quirks and variants.

Data collection can be difficult since there aren't many people who can sign fluently and because annotating big amounts of data takes a lot of effort. Large amounts of labelled data are often needed for training robust models in sign language recognition systems. The limited datasets for sign languages compared to spoken languages make it more challenging to develop precise and dependable models.

Background Clutter and Lighting Conditions: In real-world settings, background clutter and a range of lighting conditions must be taken into account by sign language identification systems. It can be difficult to tell the signer's hands apart from the background, especially if the lighting is bad or the background has intricate patterns. The accuracy of the recognition system may suffer as a result of these circumstances

Real-Time Processing: In order to enable communication between sign language users and non-signers, sign language recognition systems are frequently necessary to function in real-time. For applications like

sign language interpretation or communication with hearing-impaired people, achieving low-latency performance is essential. To provide prompt and precise recognition, real-time processing needs effective algorithms and hardware designs that are optimised.
Limited Resources: The lack of resources, such as annotated datasets, pre-trained models, and standardised evaluation criteria, hinders the development of sign language recognition systems. Sign languages have gotten considerably less attention than spoken languages, which benefit from substantial study and resources, leading to a dearth of tools and resources.

2. LITERATURE REVIEW

In last years various machine learning and deep learning models have been proposed by researchers to recognise human gesture. As [42] proposed the approach to recognise human activity by using 3DCNN, LSTM and SVM. The 2D-ConvNet used in [43] to recognise the vision based human acitivity. Therefore, in a review paper by H. Housni et al. Three modules make up the Automatic Sign Language Recognition System i.e SL database, feature extraction, and classification. The three primary channels of sign language are hands, facial expression, and body posture. The feature extraction serves as an input to the classification module, which is the second module. The training data sets are used in classification to categorise the signs. The most popular classifiers are neural networks and HMMs. The SL database, which is the third module, is crucial to an effective ASLR system[8]. In a research by Wenhui Zhang et al. a 3D somatosensory camera is the foundation of the suggested approach. The Kinect is utilised to capture the colour, depth, and skeleton frame in the first stage. The issue of background and other environmental elements can be resolved with Kinect. For the extraction of hand position features, Kinect uses the Window SDK. The HOG (Histogram of Oriented Gradient) is then used in image processing to detect objects. The SVM (Support Vector Algorithm) is employed as the final feature for the hand form[9]. In addition, a study by Dr Gomathi V. et al. camera is used by the suggested system to record the gestures. The system is provided the raw video that was captured against a moving background. Smoothing and blurring will be used in the first phase, preprocessing, to remove high intensity noises from video. and BGR to HSV conversion of the photos. Images in binary format are obtained after preprocessing. Morphological Transformations come next. Two fundamental morphological operators for eliminating all noises are erosion and dilation. Noises from the fare ground are eliminated using this method. After that, background noise is removed using median blurring, and the final image is obtained using a straightforward threshold function. Finding the contours that will be used to detect the object is the last step. A contour is a curve that connects all the places along an object's edges that are the same colour or intensity. Finally, feature extraction and selection are completed. For the purpose of grouping comparable data items, fuzzy clustering is used[10]. A deep transfer learning and SVM approach is used to recognise vision based human activity [44]. Mevin M. Dominic et al. proposed the initial step in the suggested procedure is to sample skin tones to decrease noise. Video is used as the input medium. The following step involves image preprocessing and manual segmentation. Face removal and face detection are done using the Viola and Jones method. AdaBoost classifier is used in this approach. A segmentation method is the HSV algorithm. The next phase is post processing, which is used to find and monitor the centre of gravity and finger tips to determine whether they are moving. The following step involves comparing thresholds. There is a static vs. dynamic classification. Then, dynamic gestures are further separated into those that include major hand movement or those that involve significant fingertip movement. The feature extraction process is now complete. We employ Zernike moments to determine the hand's position for static gestures, while COG is utilised for expressive motions[11]. Priyanka C Pankajakshan et al. proposed the methodology in which different steps of image processing are included in the suggested system. The webcam image is taken in the initial step. The collected image must be processed beforehand. After that, noise is removed using image processing. Grayscale conversion and thresholding are both a part of image processing. Additionally, erosion and dilation are done. The following stage, known as skin segmentation, uses the image's skin areas to recognise hand gestures. Segmentation makes use of the YCbCr domain. The third phase, hand tracking, is utilised to track the hand for an efficient hand recognition system to provide quick and error-free recognition. because the

user's hand may move accidentally or on purpose. The next phase is the clever edge detector algorithm's feature extraction. The ANN is then employed for recognition[12]. In addition Su-Jing Wang et al. uses tensor to perform SLR-based modelling on the basis of colour segmentation[13]. In addition, a research done by Archana S. Ghotkar et al. a system is proposed in which Four modules make up the suggested system in this paper: 1. Hand tracking - Webcam built-in captures input. The Camshift algorithm is used for hand tracking. The second step is to segment the hand using the HSV colour model.
3. Feature Extraction: Classifiers are used to extract features from data. Features with general applications and their relationships are extracted. 4. Gesture Recognition: The following stage after extraction is recognition, which involves looking up features in a database. For recognition, a genetic algorithm is employed[14]. In an approach proposed by Shangeetha. R.K. et al. the input frames are transformed from RGB to HIS model during segmentation. Following segmentation, the hand region is given the colour white, and the remaining portions are given the colour black. The centroid of the hand is then determined using the distance transform method. The Euclidean distance is determined from each white pixel. The split image has the palm region deleted. Fingertip detection comes next. The fingertip is defined as the location along the major axis that is furthest away from and closest to the palm area for each of the extracted finger regions. The distance between each pair of fingers is computed during the finger identification process using the base point of each finger. The next step is the splitting of the fingers into three halves, with the exception of the thumb, to account for the semi-closed fingers. Since the distance between each finger's tip and the point where the palm and finger meet is known. Detecting angles is the last stage[15].
A research by Paulraj M P et al. system contains three stages of processing: 1. Preprocessing, Second, Feature Extraction and 3. Classification of Gestures.40 frames are needed to fully record a gesture, which is recorded at a resolution of 640*480 pixels and 30 frames per second. But to minimise gathering unnecessary data, just 9–26 frames were captured for processing. The method of skin colour detection in RGB is employed. The moment invariant is discovered during feature extraction from the single blob in the collection of picture frames. The background is 0, and the blob has a high intensity value of 1. An artificial neural network is used to classify gestures[16]. Yonas Fantahun Admasu et al. divide the system into six phases in this paper: The gathering of data is the initial step. During this ESL project, students provided 640*480 photographs in jpg format, totaling 170 images. Images of the blackboard are taken. White gloves were worn in an effort to lessen the wide range of skin tones. Image preprocessing is the subsequent stage. Operations including picture size normalisation, image background subtraction, image correction, image segmentation, and image filtering, among others, are included in image preprocessing. Before applying a neural network, the third stage, feature extraction, transforms data into a form that can be represented. Gabor filtering with PCA (Principal Component Analysis) is used for feature extraction. The data input for ANNs is genuine, and the following step is data representation. The next stage is data representation; PCA's real number representation of the input data for ANN is used in this step. The most typical encoding for ANN output data is a one-of-N code. The training, testing, and validation data set phase comes next. This method separates the source data into two or more data sets at random for training and testing. Recognition is carried out using an artificial neural network at the final stage[17]. A methodology proposed by P. Subha Rajam et al. comprises four key components. 1) Data gathering 2) Extraction of Palm Images 3) Recognising signs 4) The binary to text conversion phase. In total, 320 photos were taken, with 10 images of each 32-sign being used for training. 160 photos, or 5 of each, are utilised for testing. Using an LG smart Cam, pictures are taken in a busy atmosphere in front of a chalkboard. The photos are downsized to 128*128 pixel resolution. Images are transformed into grayscale images, which are then transformed into monochrome images. The palm's contour is then extracted using the Canny Edge Detection algorithm. The scan and testing step is applied to the images. I) Marking of feature points is a step in the scanning process. ii) Evaluation of finger heights with the 'UP' position iii) calculating the angle between the horizontal line passing through the reference point and the line connecting the feature point of the 'UP' fingers, iv) counting the instances in the number 'd'. The following stage is the training step, which establishes the image threshold height. A palm held at a fixed size distance from the camera is used to take pictures. For each figure, the minimum and maximum thresholds are specified. Each category in which all photos fall comprises a

subset of signs. The output is in binary and has a length of 5. The most important portion is the "LITTLE" finger. The least significant bit is used to symbolise "THUMB". In addition, a research done by Priyanka Mekala et al. proposes a neural network-based architecture for real-time sign language recognition. The image acquisition module is the first one. The camera records the hand motions. Make sure the front of the gesture is captured by the camera. Place the camera in front of the signer. Preprocessing, which removes noise and digitization error, is the following module. Filtering the image is the initial stage of preprocessing, which removes the image's undesired noise. The median or moving average filter is applied. Background subtraction comes next. For background subtraction, the running Gaussian average method is utilised since it is quick and uses little memory. Feature extraction is the following stage. For the recognition of objects, feature vectors are required. Two feature subcategories: hand movement and form. The point of interest (POI) attribute is used to describe the figure tip. The shape and movement of the hand are thought to be represented by two POI. The word "TP" (Track Point) refers to the hand's midpoint. For sign recognition in the gesture categorization module, neural networks are employed. The suggested approach makes use of a combinational neural network[18]. Dominique Uebersax et al. proposed a method in which palm detection, orientation estimation, and classification are the three main phases in letter recognition. The computation of gravity follows normalisation. having the palm's estimated centre radius. It is thought that pixels further from the centre do not belong to fingers. Principal component analysis is used to determine the hand orientation. Since PCA is not very accurate, fingertip detection is used to further refine the orientation. Now that the data has been split and normalised, classification may begin. There are three classifiers used: The first codebook examines the pixel-by-pixel depth separation between the seen hand and the code entry. The second one is dependent on hand orientation. The third is substantially more potent and is based on ANMM[19]. In a review paper by M. Ebrahim Al-Ahdal et al. discusses various capture methods and recognition strategies. The authors proposed a system with three levels: data collecting, gesture recognition, and language modelling level after thoroughly examining the advantages and disadvantages of each method. Data gathering with data gloves is part of the first stage. Multi-channel EMG sensors are introduced to get over the segmentation restriction. The second level includes gesture recognition, where four channel HMMs are used to categorise the gestures. Each channel represents a movement channel, hand form channel, rotation channel, and orientation channel based on a modelled lexicon. The third level of phrase recognition uses a single channel HMM to match recognised gestures with modelled sentences based on language grammar and semantics[20].

A survey done by Lalit Kaneet et al. in this survey-based research, various sign recognition system methodologies and approaches (vision based, glove based) are discussed. The signer needs to wear sensor-based gloves while using the glove-based technique. The drawback of this method is its inability to detect non-manual gestures, such as facial expressions and eye movements. This study discusses an SLR system that is vision-based. Image processing, segmentation, descriptor specification, and pattern classification are the system's other four components. Skin tone is used for segmentation. While some scientists utilise the RGB paradigm, others favour the HSV model. YCbCr colour space is used by some people. Principle Curvature Based Region Detector and 2-D Wavelet Packet Decomposition are used in Indian Sign Language Recognition for extracting hand descriptors (PCBR), while some employ Histogram of Orientation Gradients (HOG) and Histogram of Edge Frequencies (HOEF) for feature extraction. Additionally, SVM or a multi-layer sensory neural network is employed for the classification KNN. Principle Component Analysis is utilised in American Sign Language (ASL) for hand configuration and orientation, while HMM is employed as a classifier for pattern classification[21]. Furthermore [47] developed a deep learning based hybrid model for ASL with 99.57% accuracy.

A glove based approach is proposed by Kunal Kadam et al. in which gloves are utilised in the proposed system to collect input. The gloves are on the signer. Sensors for measuring finger flex are installed to gloves. A correctly made sign is made using an LCD display. Two ways exist for teaching and learning. By entering various motions, a database is formed in teach mode. The existing database is compared with the gesture in learn mode. Five times the worth of a gesture are matched. The notification indicating the gesture is invalid is printed if it does not match the database already in use[5]. I. Bulugu[48] developed glove based SLR system for Tanzanian Sign Language(TSL) with 96.5 % accuracy.

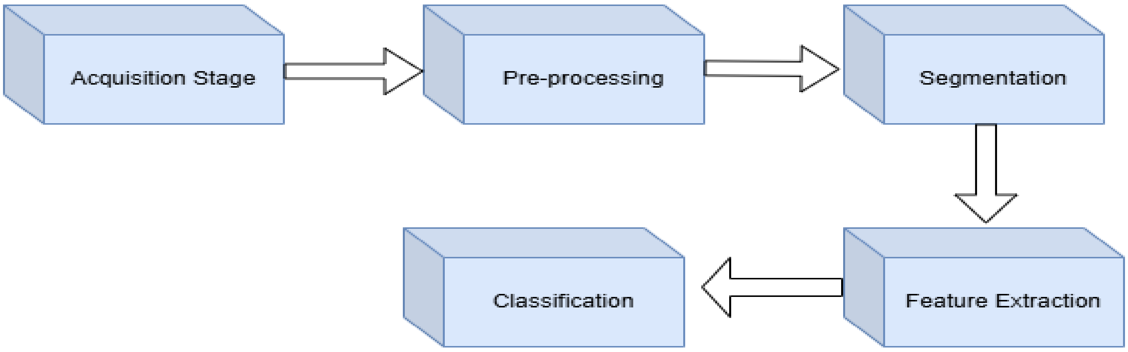


Fig.3 Frequency of review papers

3. APPROACHES

Sign Language Recognition approaches depends on the device used to take the input. If Camera is used to capture the gesture, it is Vision Based Approach else if Sensor based gloves are used to track the gesture, then it is Glove Based Approach[22].

A. VISION BASED

In vision-based approach webcams are employed to record gestures. The accuracy of output is dependent on the angle of camera[4][3]. Because image of a single sign from different angle will be different this will cause the different meaning of a single sign. In some researches more than one camera is used in different angle and more than one view of gesture is captured. Sign Language recognition in vision-based approach is divided into five steps: 1) Acquisition Stage 2) Pre-processing 3) Segmentation 4) Feature Extraction and 5) Classification as shown in fig 3.

fig.4 Basic vision based approach model

A.1 ACQUISITION STAGE

This is the initial stage of recognizing sign language, where input is gathered using a camera, gloves, a Kinect, and a jump motion controller, among other input devices. After comparing input devices, a web camera is the best option because it is affordable, convenient, and simple to use[4].

Table 1. Comparison of existing acquisition devices

Device	Area of scope	Loopholes
Web Cam[12][20]	Priyanka C Pankajakshan et al. and Archana S. Ghotkar et al. used web cam in their research.Users didn't need to wear any additional external devices; they could simply use their hands inside the camera collection range.	Preprocessing is important to remove extraneous background elements that could compromise accuracy.

Gloves[5]	Kunal Kadam et al. used gloves for gathering data, these gadgets are not impacted by the environment outside. Improved recognition accuracy is offered.	It lessens how organic interactions are. User finds it difficult to utilize, really pricey.
Kinect[8]	Wenhui Zhang et al. used a 3D somatosensory Kinect camera to capture the colour, depth, and skeleton frame for gesture recognition. It is helpful for many applications involving human-computer interaction.	Noise, complicated backgrounds, hand and facial segmentation, and illumination can all have an impact.

A.2 PRE-PROCESSING

When camera capture the gesture, it is possible background noises are also captured. Because the accuracy of result depends on the input, it is necessary to eliminate the unnecessary background noises from image[23]. During pre-processing stage high intensity noises are eliminated from image by using smoothing and blurring techniques. Some researcher uses the morphological operation during pre-processing[24]. It is the part of image processing contains the set of operations used to process the image which is based on shape. The two main operations of morphological operations are dilation and erosion used to eliminate small object from image[25].

At this stage, image improvement and restoration are carried out to boost image quality so that it will be better for either human or machine analysis. Both image repair and enhancement techniques are widely available[23]. There are several image enhancement techniques available, including Histogram Equalisation, Adaptive Histogram Equalisation, Contrast Limit Adaptive Histogram Equalisation (CLAHE), and others. And methods for restoring images include Gaussian, Mean, and Median filters, among others[25].

In this work methods of pre-processing were applied to a supplementary dataset to identify the gap.

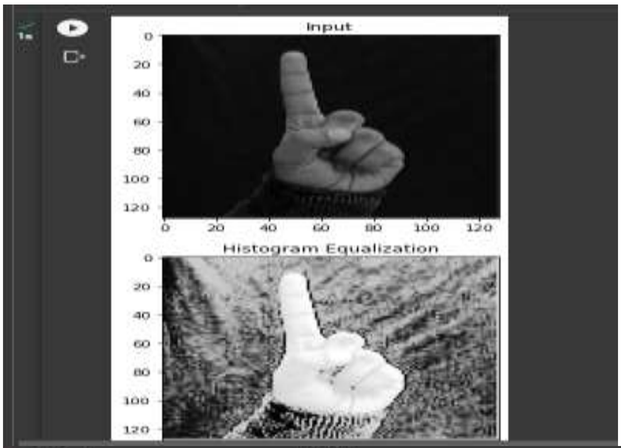
A.2.i Image Enhancement Techniques

A collection of techniques and algorithms are referred to as image enhancement techniques and are used to increase the visual appeal and quality of digital photos. These methods try to improve the brightness, contrast, sharpness, and colour of an image, among other qualities.

In this work some methods applied on images for improving images and the result is shown in following figures (5 & 6):

1.By redistributing an image's intensity levels, the histogram equalisation approach improves contrast. In the Fig.5 Top grayscale image named as “Input” which is the original image and the second image is the result of HE (histogram equalization) technique shows that it operates by extending the image's histogram to encompass the entire range of intensity values. The original image contains the uneven lighting which cause to poor visibility. Whereas the bottom image shows improved contrast used to distinct the hand and the background.

2.In addition Fig.6 contains the two grayscale images in this top image named as “Input” is the original image and the bottom image is the result of CLAHE (Contrast Limit Adaptive Histogram Equalization). The original image has darker background which lead to some details less visible. The image after applying CLAHE has much better contrast enhancement than histogram equalization.



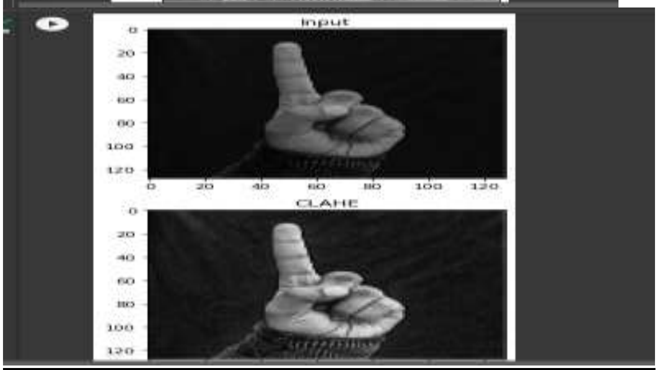


fig.5 Histogram Equalization

fig.6 Contrast Limit Adaptive Histogram Equalization(CLAHE)

By dispersing pixel intensity values around the histogram, HE improves the image's contrast and increased brightness is seen in the final image, particularly in the hand area, but at the expense of some overexposure (decreased detail in bright regions). While CLAHE prevents overexposure by enhancing contrast while reducing amplification in bright spots. With improved texture and detail visibility, the treated image keeps its more natural appearance and in contrast to conventional histogram equalization, the background stays smooth and noise-free. Although Histogram Equalization works well, it can cause excessive brightness in some places. But better local contrast control is offered by CLAHE, which preserves finer details without overexposing bright areas.

A.2.ii Image Restoration Techniques-

Image restoration techniques are a collection of procedures and algorithms used to improve or repair damaged or degraded images. By lowering noise, eradicating artefacts, increasing features, and sharpening an image, these techniques seek to restore its original content and quality. In this work here are several methods for image restoration that are applied on image and results are shown in figures. Fig.7 contains the original image before applying any restoration techniques. A 5x5 mean kernel filter was used to smooth this image. The result is a minor blurring effect that preserves certain details while lowering some noise (as shown in Fig. 8). In Fig. 9 to smooth this image, a bigger 10x10 mean kernel was used. Blurring has a more noticeable effect, lowering noise levels but also obliterating details. Next The original image now seems grainy due to the addition of Gaussian noise. With tiny, erratic changes in intensity, this kind of noise impacts the entire image (as shown in Fig. 10). And finally, Salt-and-pepper noise has distorted this image, randomly adding black and white pixels.

Salt-and-pepper noise produces definite black and white spots, in contrast to Gaussian noise, which gradually alters pixel brightness (Fig.11).

Fig.7 Original Image **Fig.8 Mean Filter with 5*5 kernel**

Fig.9 Mean Filter with 10*10 kernel

Fig.10 Image with Gaussian noise

Fig.11 Image with S&P noise

High-frequency noise is reduced by the mean 5*5 kernel while a respectable degree of detail is preserved in the 10* 10 The image appears considerably more blurred than in the 5x5 casing because the edges and finer details are further lost.
In natural imaging systems, low-light cameras may introduce some noise (Gaussian noise) . because of Such noise Pixel intensity varies at random; this type of noise is common. Usually, sensor defects or issues with image transmission result in another type of noise. There is salt and pepper noise in the picture (fig.11), which shows up as intermittent black and white pixels. Although mean filtering efficiently decreases noise, it causes blurring and a loss of clarity, especially when using a bigger kernel.

Because Gaussian noise causes random intensity variations throughout the image, it must be reduced using techniques like bilateral filtering or Gaussian blur. Certain pixels are disturbed by salt and pepper noise, which is better removed using techniques like median filtering.

A.3 SEGMENTATION

Skin segmentation is utilised in the second stage to identify the skin regions from the image to identify the gesture. In this step skin color will be separated from non-skin color.

The aim of segmentation technique is to divide the image into segment based on similar properties like color, shape or texture. These partitioned segments are also known as regions or objects[22].

At this point, the image has been divided into segments, which are useful geographic units. To identify the region of interest in an image, segmentation is used. Edge-based, thresholding-based, cluster-based, region-based, or based on artificial neural networks are some examples of image segmentation techniques[26][41]. In this work experiment with edge-based, region-based, and cluster-based techniques have been done and discovered the following gap.

A.3.i Edge Based Technique

A group of image processing techniques known as "edge-based techniques" are centred on identifying and improving the edges in an image. Edges in an image reflect rapid changes in brightness or colour and frequently correspond to object boundaries or important details. Edge-based methods can be used for a range of assignments, including object detection, feature extraction, and picture segmentation, by emphasising these edges. The original, unedited image of a hand gesture on a black background is seen in Fig.12. Second picture emphasizes textures and edges in a horizontal orientation. Fingers, creases, and sleeve patterns are examples of horizontal details that are improved (as shown in Fig.13). Fig.14 shows Strong enhancement is applied to vertical structures such as the wrist and finger outline.

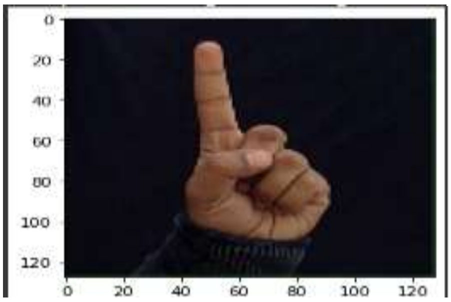


Fig.12 Original Image

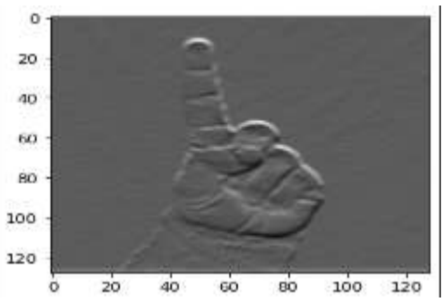


Fig.13 Horizontal Segmentation

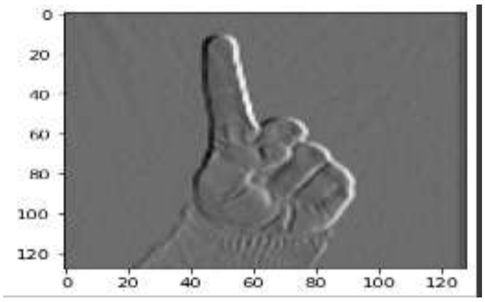


Fig.14 Vertical Segmentation

A.3.ii Region Based Technique

A region-based methodology is a technique or strategy that is used to analyse or alter data within certain regions or areas of interest across numerous sectors, such as computer vision, image processing, and data analysis. Region-based techniques concentrate on specific regions or subsets of data to execute targeted operations or extract pertinent information, as opposed to considering the full dataset as a whole.

Region-based algorithms are frequently used in computer vision and image processing for tasks like object detection, segmentation, and tracking. These methods entail segmenting an image or scene into smaller areas or regions of interest (ROIs), which are then analysed or processed independently. This strategy enables more localised analysis and can enhance the algorithms' effectiveness and precision.

Data analysis and machine learning both use region-based strategies to manage spatially or geographically related data. For instance, region-based techniques are used in geographical information systems (GIS) to analyse spatial data within specified regions, such as discovering patterns or trends in a given area.

Whether it be for computer vision, image processing, or geographic data analysis, region-based techniques generally offer a mechanism to concentrate on particular regions of interest while analysing data. These methods can enhance the efficacy, efficiency, and accuracy of various algorithms and models by focusing on particular locations. These methods entail segmenting an image or scene into smaller areas or regions of interest (ROIs), which are then analysed or processed independently. The result of region-based segmentation is shown in Fig.17. The first image is the original image with black background as shown in Fig.15 and converted into grayscale before applying segmentation technique to reduce computational time as shown in Fig.16.

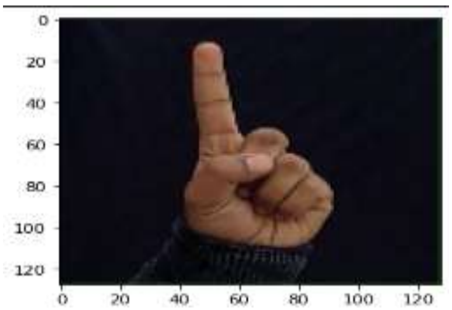


Fig.15 Original Image

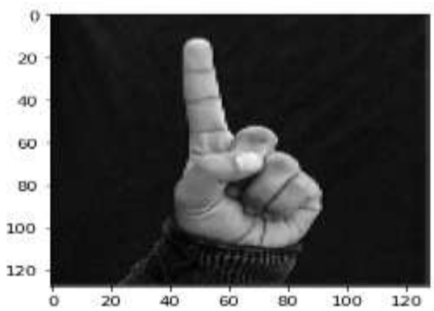


Fig.16 Grayscale Image

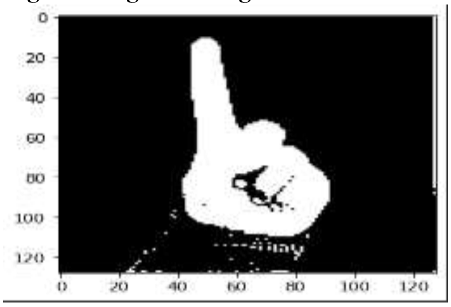


Fig.17 Output

A.3.iii Cluster Based Technique

A class of machine learning algorithms known as cluster-based techniques is used to cluster data, which is the process of putting related objects or data points together. These methods divide a dataset into discrete clusters based on the degree of similarity or dissimilarity between the data points to find innate patterns or structures within the data.

In several fields, including image processing, market segmentation, consumer profiling, anomaly detection, and recommendation systems, cluster-based techniques are frequently utilised. Each method has its own advantages, disadvantages, and applicability for various data kinds and clustering circumstances. In this work cluster based technique (K-means) is applied on image as shown in Fig.18 and 19. Where Fig.18 is original image and 19 is the result of K-means clustering. The image is segmented with 3 colors based on the similarity of colors.

Fig.18

Fig.19

In the segmentation based on edges The smooth areas of the hand and finger are suppressed, and the edges are accentuated. In contrast, region-based separating the foreground (hand) from the background, a binary (black and white) effect is produced. There may be considerable noise in the segmentation, especially in smaller areas. Finally, by reducing color variations, the cluster-based segmented image highlights discrete sections while maintaining the basic structure. Although edge-based techniques can be susceptible to noise, they are effective in identifying object boundaries.
While region-based segmentation is good at isolating objects, it might not work well in areas with a lot of texture. K-means cluster-based segmentation is helpful for object separation and color simplification.

Table 2. Comparison of Image Segmentation Techniques-

Segmentation Technique	Area of scope	Loopholes
Edge Based Technique	Suitable for photos with enhanced object contrast.	It's not appropriate for photographs with excessive edges or noise.
Cluster Based Technique	Due to the usage of fuzzy partial membership, it is more applicable to problems in the real world.	It is difficult to determine how membership functions.
Region Based Technique	It is more useful and less sensitive to noise when the similarity criteria are simple to define.	In terms of processing time and memory usage, it is quite expensive.

A.3 FEATURE EXTRACTION

Feature extraction is employed in order to find smaller set of new variables by eliminating redundant variables. In sign language recognition prediction is based on features. If number of features will be high, it will be difficult to work on it. Sometimes many features are redundant[22]. Feature Extraction is used to eliminate these unnecessary repeated features so that result will be accurate. There are various methods used to extract features from segmented image.
The most pertinent features from the image are extracted at this step of recognition using an extraction approach. Fourier Descriptor(FD), Shift-Invariant Feature Transform(SIFT), Oriented Gradient Histogram, Principal Component Analysis(PCA), and Speed Up Robust Feature(SURF) are just a few of the significant extraction approaches employed to obtain a good performance[27].

Table 3. Comparison of existing Feature Extraction techniques

Techniques	Area of scope	Loopholes
Principal Component Analysis (PCA)[15]	Minimal sensitivity to noise. Lower the data's excessive dimensionality and redundant features.	Ability to interpret lost data Only pair-wise correlations between the brain image voxels are taken into consideration by PCA.

Histogram of Oriented Gradient (HOG)[8] [21]	The gradient structure or edge that is a key component of the local perception is captured.	The rotation of images can affect it. Because it uses an extensive scanning strategy across the entire region of interest, it is computationally intensive.
Fourier Descriptor (FD)[28]	Rotation, scale, and translation are all invariant.	Noise resistance and affine transform are terrible. Since the locations of the frequencies are unknown, only their magnitudes can be used to locate local features.
Scale Invariant Feature Transform (SIFT)	Comparable to no other adjectives, it is more accurate. It is scale- and rotation-invariant.	Compared to SURF, SIFT is comparatively slow. It performs poorly when the lighting changes.

A.4 FEATURE CLASSIFICATION

Feature classification technique is used to recognize the pattern. It classifies the amount of data into different classes. There are various methods used to classify the data.

This is the final stage of sign language recognition where classification of image is done with the help of classifier. Many classifiers are available used by researchers in the understanding of sign language.

i)HMM (Hidden Markov Model) is a mathematical model used to extract the hidden information from observable symbols[29][40].

ii)ANN (Artificial Neural Network) is a classification algorithm employed to identify the data. It is similar to the human brain's structure. Like human brain ANN al/so have neurons. The architecture of ANN is classified into three levels.1) Layer of input; 2) Hidden layer 3) The layer of output. An Artificial Neural Network can have multiple hidden layers[30][40].

iii)KNN (K-Nearest Neighbor) is a learning algorithm use dataset to classify the category of new data. It matches the new data with available data and compare the fresh data with the existing data for similarity and classify the category based on the similarity[30].

iv) CNN (Convolutional Neural Network) is a cutting-edge deep learning technique that uses an image as input to rank the objects in the image and distinguish them from one another. Compared to other classifiers, CNN requires less pre-processing. The CNN's architecture is comparable to the neuronal connection network in the human brain[29].

v) SVM (Support Vector machine) is employed in categorization and outlier detection. It finds hyperplane in N-D (N Dimensional) space that differentially classifies data points. The dimension depends on the number of features. If the features are two, then it is just a line else if the features are three it is 2D. If the features are more than three it is difficult to found hyperplane i.e., SVM can perform classification only between two classes. Hyperplane is a boundary between two class[22][40].

Table 4. Comparison of existing classifiers

Classifier	Area of scope	Loopholes
K-Nearest Neighbors[21]	Implementation is simple. It is easy to understand the algorithm.	Extremely perceptive to unimportant details. Large datasets don't work well with it.
Artificial Neural Networks (ANN)[12][15]	When the learnt target function needs to be rapidly assessed, it is useful.	Selecting an appropriate network architecture is a difficult and costly computing task.

Support Vector Machine (SVM)[8]	When interacting with continuous features and several dimensions, it performs better.	To reach its highest level of prediction accuracy, the dataset must have a substantial sample size.
Hidden Markov Model (HMM)[11][18][21]	In terms of recognition, it does reasonably well. The implementation and analysis are simpler.	HMMs usually have a large number of unstructured parameters. Comprehensive training is required to achieve better results.
Convolution Neural Network (CNN)[31]	It detects important traits automatically, without human intervention.	High computation costs. A significant amount of training data is required to achieve good accuracy.

Table 5. Various classifiers used for different sign language recognition datasets with accuracy

Classifiers	**Result**	**Dataset**
CNN[31]	-	Arabic Sign Language (ArSL)
Gabor Filter and PCA for feature extraction. Artificial Neural Network for Ethiopian Sign Language recognition[15]	98.53%	Ethiopian Sign Language
HOG and SVM algorithms[8]	89.8%	3D sign language dataset with 72 words collected using Kinect
Coupled HMM[32]	90.80%	Indian Sign Language dataset with 25 dynamic sign words
Genetic algorithm[28]	-	Indian Sign Language
Temporal Graph Convolutional Networks(TGCN)[33]	-	Word Level American Sign Language
Random Forest classifier with inertial measurement unit based gloves[34]	92%	French Sign Language
ANN[35]	79.58%	American Sign Language
Multi headed CNN[36]	98.981%	American Sign Language Finger Spelling dataset
Multimodality model (Late Fusion of Computer Vision and Leap Motion)[37]	82.55%	British Sign Language
Combinational Neural Network[38]	48% with noise data	26(A-Z) English alphabets
BiLSTM[39]	94.46%	Saudi Sign Language
BiGRU[47]	99.57%	ASL

B. **GLOVE-BASED**

In glove-based approach sensors are used to track the sign. A glove has five sensors for each finger. Some researchers have used LCD for blending the accurate sign so that the result will be accurate. The problem

with this approach is that user have to wear the gloves. It is difficult to carry the gloves all time and if user move the finger un intentionally will also be consider as sign [5] [18] [34].
A glove with multiple sensors is used in a glove-based method to sign language recognition to record the hand motions and movements used in sign language. To identify and interpret the related indicators, the sensor data is next processed and analysed.

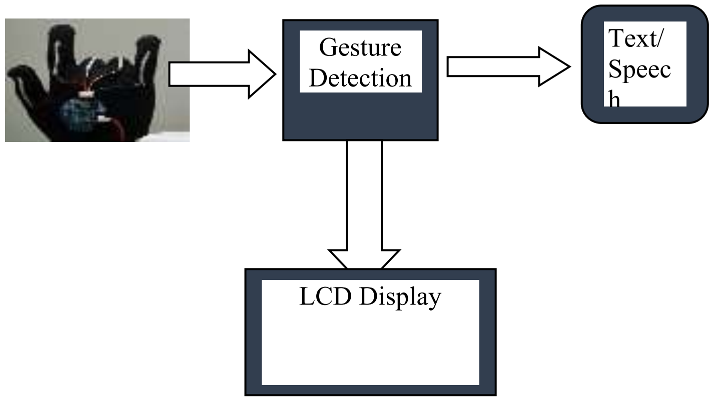


Fig.20 Diagram of glove-based approach for converting sign to text/speech

The main phases in a glove-based sign language recognition system are as follows [45] [46]:
1.Sensor Integration: To reliably record hand movements, the glove is equipped with sensors like flex sensors, accelerometers, gyroscopes, or even camera-based systems. These sensors take measurements of things like acceleration, finger flexion, and orientation.
2. Data collection: While the user wears the glove and makes sign language gestures, sensor data is being gathered. The sensor readings are continuously recorded by the glove, which also captures finger and hand movements.
3.Preprocessing: To reduce noise, weed out unnecessary data, and normalise the values, the raw sensor data may need to be processed. This action aids in enhancing the precision of later processing operations.
4.Feature Extraction: From the pre-processed sensor data, pertinent features are extracted. Finger placements, joint angles, or other pertinent metrics that capture the distinctive qualities of each sign gesture may be included in these elements.
5.Gesture Recognition: To categorise the retrieved information and identify the relevant sign gestures, machine learning or pattern recognition algorithms are used. There are other strategies that can be applied, such as:

- Rule-based Systems: To map the collected features to sign motions, these systems use hand-crafted heuristics and established rules. They might not be able to handle sign language's nuances and complexity, though.
- Machine learning: To recognise sign motions, labelled data can be used to teach supervised learning methods like as neural networks, KNN, Random Forests, and Support Vector Machines (SVM). During training, the sensor data features are used as the input, and the matching sign motions are used as the output labels [48].
- Deep Learning: To learn complicated spatial and temporal patterns in the sensor data for sign language recognition, Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) can be utilised. In a variety of gesture recognition challenges, deep learning models have demonstrated encouraging results.

6.Output and Interpretation: Once the sign gesture has been identified, the system can interpret it and deliver the desired output. This output could be speech, text, or any other format that makes communication easier.

4. CONCLUSION AND FUTURE WORK

This paper reviews and analyzes techniques for image enhancement (Histogram Equalization, Contrast Limited Adaptive Histogram Equalization), image restoration (Mean Filter, Gaussian Filter), and image segmentation (Edge Based, Region Based, and Cluster Based) in the context of Vision-Based SLR by applying them to images. It also discusses the limitations and scope of these techniques. Taken together, these techniques greatly improve the precision, speed, and resilience of vision-based SLR systems. By making the system more functional in real-world settings, they facilitate improved communication between the deaf population and the general public.

In future the accuracy, efficiency, and real-time adaptability of visual processing systems will be improved in on picture enhancement, restoration, and segmentation techniques for vision-based sign language recognition. Future work must concentrate on developing more resilient algorithms that can manage difficult environmental conditions like dim illumination, occlusion, motion blur, and complicated backgrounds since accurate interpretation of hand, facial, and body movements is crucial to sign language identification.

5. REFRENCES

[1] S. G. Moreira Almeida, F. G. Guimarães, and J. Arturo Ramírez, "Feature extraction in Brazilian Sign Language Recognition based on phonological structure and using RGB-D sensors," *Expert Systems with Applications*, vol. 41, no. 16. Elsevier Ltd, pp. 7259–7271, Nov. 15, 2014. doi: 10.1016/j.eswa.2014.05.024.

[2] Suharjito, R. Anderson, F. Wiryana, M. C. Ariesta, and G. P. Kusuma, "Sign Language Recognition Application Systems for Deaf-Mute People: A Review Based on Input-Process-Output," in *Procedia Computer Science*, Elsevier B.V., 2017, pp. 441–448. doi: 10.1016/j.procs.2017.10.028.

[3] I. A. Adeyanju, O. O. Bello, and M. A. Adegboye, "Machine learning methods for sign language recognition: A critical review and analysis," *Intelligent Systems with Applications*, vol. 12. Elsevier B.V., Nov. 01, 2021. doi: 10.1016/j.iswa.2021.200056.

[4] M. J. Cheok, Z. Omar, and M. H. Jaward, "A review of hand gesture and sign language recognition techniques," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 1, pp. 131–153, Jan. 2019, doi: 10.1007/s13042-017-0705-5.

[5] K. Kadam, R. Ganu, A. Bhosekar, and S. D. Joshi, "American sign language interpreter," in *Proceedings - 2012 IEEE 4th International Conference on Technology for Education, T4E 2012*, 2012, pp. 157–159. doi: 10.1109/T4E.2012.45.

[6] S. Kausar and M. Y. Javed, "A survey on Sign Language recognition," in *Proceedings - 2011 9th International Conference on Frontiers of Information Technology, FIT 2011*, 2011, pp. 95–98. doi: 10.1109/FIT.2011.25.

[7] D. M. Madhiarasan and P. P. P. Roy, "A Comprehensive Review of Sign Language Recognition: Different Types, Modalities, and Datasets," vol. XX, no. X, pp. 1–30, 2022, [Online]. Available: http://arxiv.org/abs/2204.03328

[8] A. Er-Rady, R. Faizi, R. O. H. Thami, and H. Housni, "Automatic sign language recognition: A survey," *Proc. - 3rd Int. Conf. Adv. Technol. Signal Image Process. ATSIP 2017*, vol. 9, pp. 1–7, 2017, doi: 10.1109/ATSIP.2017.8075561.

[9] Y. Chen and W. Zhang, "Jf," pp. 1947–1951, 2016.

[10] H. Muthu Mariappan and V. Gomathi, "Real-time recognition of Indian sign language," *ICCIDS 2019 - 2nd Int. Conf. Comput. Intell. Data Sci. Proc.*, pp. 1–6, 2019, doi: 10.1109/ICCIDS.2019.8862125.

[11] A. Kumar, K. Thankachan, and M. M. Dominic, "Sign language recognition," *2016 3rd Int. Conf. Recent Adv. Inf. Technol. RAIT 2016*, pp. 422–428, 2016, doi: 10.1109/RAIT.2016.7507939.

[12] P. C. Pankajakshan and B. Thilagavathi, "Sign language recognition system," *ICIIECS 2015 - 2015 IEEE Int. Conf. Innov. Information, Embed. Commun. Syst.*, pp. 2–5, 2015, doi: 10.1109/ICIIECS.2015.7192910.

[13] S. J. Wang, L. B. Zhang, D. C. Zhang, C. C. Jia, N. Zhang, and C. G. Zhou, "A sign language recognition based on tensor," *2010 Int. Conf. Multimed. Inf. Technol. MMIT 2010*, vol. 2, pp. 192–195, 2010, doi: 10.1109/MMIT.2010.21.

[14] A. Susitha, N. Geetha, R. Suhirtha, and A. Swetha, "Static and Dynamic Hand Gesture Recognition for Indian Sign Language," *Lect. Notes Networks Syst.*, vol. 256, pp. 48–66, 2022, doi: 10.1007/978-3-030-82469-3_5.

[15] R. K. Shangeetha and S. Padmavathi, "COMPUTER VISION BASED APPROACH FOR INDIAN SIGN LANGUAGE CHARACTER RECOGNITION Dept . of Information Technology Amrita VishwaVidyapeetham Valliammai . V Dept . of Information Technology Amrita VishwaVidyapeetham Asst . Professor Technology Amrita VishwaVid," pp. 181–184.

[16] M. P. Paulraj, S. Yaacob, M. S. Bin Zanar Azalan, and R. Palaniappan, "A phoneme based sign language recognition system using skin color segmentation," *Proc. - CSPA 2010 2010 6th Int. Colloq. Signal Process. Its Appl.*, pp. 86–90, 2010, doi: 10.1109/CSPA.2010.5545253.

[17] Y. F. Admasu and K. Raimond, "Ethiopian sign language recognition using Artificial Neural Network," *Proc. 2010 10th Int. Conf. Intell. Syst. Des. Appl. ISDA'10*, pp. 995–1000, 2010, doi: 10.1109/ISDA.2010.5687057.

[18] P. Mekala, Y. Gao, J. Fan, and A. Davari, "Real-time Sign Language Recognition based on Neural Network Architecture."

[19] D. Uebersax, J. Gall, M. Van Den Bergh, and L. Van Gool, "Real-time sign language letter and word recognition from depth data," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 383–390, 2011, doi: 10.1109/ICCVW.2011.6130267.

[20] M. Ebrahim Al-Ahdal and M. T. Nooritawati, "Review in sign language recognition systems," *2012 IEEE Symp. Comput. Informatics, Isc. 2012*, pp. 52–57, 2012, doi: 10.1109/ISCI.2012.6222666.

[21] L. Kane and P. Khanna, "Towards establishing a mute communication: An Indian sign language perspective," *4th Int. Conf. Intell. Hum. Comput. Interact. Adv. Technol. Humanit. IHCI 2012*, 2012, doi: 10.1109/IHCI.2012.6481787.

[22] M. Safeel, T. Sukumar, K. S. Shashank, M. D. Arman, R. Shashidhar, and S. B. Puneeth, "Sign Language Recognition Techniques- A Review," in *2020 IEEE International Conference for Innovation in Technology, INOCON 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/INOCON50539.2020.9298376.

[23] W. Zhou, X. Ma, and Y. Zhang, "Research on Image Preprocessing Algorithm and Deep Learning of Iris Recognition," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Sep. 2020. doi: 10.1088/1742-6596/1621/1/012008.

[24] Y. Zhu, H. Yuan, C. Zhang, and C. Lee, "Image-preprocessing method for near-wall particle image velocimetry (PIV) image interrogation with very large in-plane displacement," *Meas. Sci. Technol.*, vol. 24, no. 12, 2013, doi: 10.1088/0957-0233/24/12/125302.

[25] I. Patel, S. Patel, and A. Patel, "Analysis of Various Image Preprocessing Techniques for Denoising of Flower Images Digital ScareCrow using Iot View project CHARUSAT Apps (Mobile Application) View project Analysis of Various Image Preprocessing Techniques for Denoising of Flower Images," *Artic. Int. J. Comput. Sci. Eng.*, 2018, [Online]. Available: www.ijcseonline.org

[26] S. Joudaki, D. bin Mohamad, T. Saba, A. Rehman, M. Al-Rodhaan, and A. Al-Dhelaan, "Vision-based sign language classification: A directional review," *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, vol. 31, no. 5. Taylor and Francis Ltd., pp. 383–391, Nov. 03, 2014. doi: 10.1080/02564602.2014.961576.

[27] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *Proceedings of 2014 Science and Information Conference, SAI 2014*, Institute of Electrical and Electronics Engineers Inc., Oct. 2014, pp. 372–378. doi: 10.1109/SAI.2014.6918213.

[28] Institute of Electrical and Electronics Engineers Madras Section, Sri Shakthi Institute of Engineering and Technology, International Conference on Computer Communication and Informatics 2012.01.10-12 Coimbatore, and ICCCI 2012.01.10-12 Coimbatore, *International Conference on Computer Communication and Informatics (ICCCI), 2012 10-12 Jan. 2012, Coimbatore, India*.

[29] R. Rumana, R. Sandhya Rani, and R. Prema, "A Review Paper on Sign Language Recognition for The Deaf and Dumb." [Online]. Available: www.ijtrs.com

[30] K. Amrutha and P. Prabu, "ML based sign language recognition system," in *2021 International Conference on Innovative Trends in Information Technology, ICITIIT 2021*, Institute of Electrical and Electronics Engineers Inc., Feb. 2021. doi: 10.1109/ICITIIT51526.2021.9399594.

[31] M. Mustafa, "TI TI," *J. Ambient Intell. Humaniz. Comput.*, vol. 12, no. 3, pp. 4101–4115, 2021, doi: 10.1007/s12652-020-01790-w.

[32] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "Coupled HMM-based multi-sensor data fusion for sign language recognition," *Pattern Recognit. Lett.*, vol. 86, pp. 1–8, Jan. 2017, doi: 10.1016/j.patrec.2016.12.004.

[33] D. Li, C. Rodriguez Opazo, X. Yu, and H. Li, "Word-level Deep Sign Language Recognition from Video: A New Large-scale Dataset and Methods Comparison." [Online]. Available: https://dxli94.github.io/

[34] S. Kasireddy, P. M. Scholl, J. Kempfle, and K. Van Laerhoven, "Gestures with an IMU-Based Glove †," pp. 1–18, 2018, doi: 10.3390/informatics5020028.

[35] M. Rivera-acosta, S. Ortega-cisneros, and F. Sandoval-ibarra, "Neural Network," 2017, doi: 10.3390/s17102176.

[36] R. K. Pathan, M. Biswas, S. Yasmin, M. U. Khandaker, M. Salman, and A. A. F. Youssef, "Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network," *Sci. Rep.*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-43852-x.

[37] J. J. Bird, A. Ekárt, and D. R. Faria, "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language," *Sensors (Switzerland)*, vol. 20, no. 18, pp. 1–19, Sep. 2020, doi: 10.3390/s20185151.

[38] M. Palmeri, F. Vella, I. Infantino, and S. Gaglio, "Sign languages recognition based on neural network architecture," *Smart Innov. Syst. Technol.*, vol. 76, pp. 109–118, 2018, doi: 10.1007/978-3-319-59480-4_12.

[39] L. Al Khuzayem, S. Shafi, S. Aljahdali, R. Alkhamesie, and O. Alzamzami, "Efhamni: A Deep Learning-Based Saudi Sign Language Recognition Application," *Sensors*, vol. 24, no. 10, 2024, doi: 10.3390/s24103112.

[40] K. K. Verma, B. M. Singh, and A. Dixit, "ORIGINAL RESEARCH A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system," *Int. J. Inf. Technol.*, 2019, doi: 10.1007/s41870-019-00364-0.

[41] "n t e r n a t i o n a l o u r n a l o f".

[42] K. K. Verma and B. M. Singh, "Deep Multi-Model Fusion for Human Activity Recognition Using Evolutionary Algorithms," vol. 7, pp. 44–58, 2021, doi: 10.9781/ijimai.2021.08.008.

[43] K. K. Verma, B. M. Singh, H. L. Mandoria, and P. Chauhan, "Two-Stage Human Activity Recognition Using," pp. 125–135, 2020, doi: 10.9781/ijimai.2020.04.002.

[44] K. K. Verma and B. M. Singh, "Vision based Human Activity Recognition using Deep Transfer Learning and Support Vector Machine," *2021 IEEE 8th Uttar Pradesh Sect. Int. Conf. Electr. Electron. Comput. Eng.*, pp. 1–9, doi: 10.1109/UPCON52273.2021.9667661.

[45] S. M. Biju, O. Al-khatib, H. Z. Sheikh, and F. Oroumchian, "Glove based wearable devices for sign language-GloSign," vol. 12, no. 4, pp. 1666–1676, 2023, doi: 10.11591/ijai.v12.i4.pp1666-1676.

[46] S. J. Fernando and D. J. David, "SMART GLOVES FOR SIGN LANGUAGE RECOGNITION : BRIDGING COMMUNICATION GAP," vol. 11, no. 7, pp. 426–435, 2024.
[47] B. Alabduallah, R. Al Dayil, A. Alkharashi, and A. A. Alneil, "Innovative hand pose based sign language recognition using hybrid metaheuristic optimization algorithms with deep learning model for hearing impaired persons," pp. 1–25, 2025.
[48] I. Bulugu, "Tanzanian sign language recognition system for an assistive communication glove sign tutor based on the inertial sensor fusion control algorithm," *J. Electr. Syst. Inf. Technol.*, vol. 12, no. 1, 2025, doi: 10.1186/s43067-025-00199-9.