Indian Sign Language Recognition System

Atharva Shinde[1], Anushri Shivale[2], Siddhesh Phapale[3], Assistant Prof.Renuka Kajale[4]

[1,2,3]Students, CS Department, NMIET, Talegaon, Pune, India
[4]Assistant Professor, CS Department, NMIET, Talegaon, Pune, India

Abstract **:** People can interact and exchange ideas and emotions through communication. The social contacts of the deaf community are hindered by multiple factors. The people converse with each other using sign language. A technology can translate sign languages into a form that is comprehensible in order to communicate with ordinary people. Developing a real-time text-to-Indian Sign Language (ISL) translation system is the aim of this project. For the most part, manual labor is used. In this paper, we describe a convolutional neural network-based deep learning method for classifying signs. We initially construct a classifier model using the numerical signs and the Python-based Keras convolutional neural network implementation. In phase two, a second real-time system was used to use skin segmentation to detect the Region of Interest in the frame that displays the bounding box. To forecast the sign, the segmented region is fed into the classifier model. The accuracy rating of the system for the same subject is 99.56% in poor light and 97.26% in high light. It was observed that the classifier improved with varying image capture angles and backgrounds. The RGB camera system is the main emphasis of our strategy.

Keywords— Real-time systems, areas of interest, convolutional neural networks, and deep learning.

I. INTRODUCTION

Sign language is used by both the deaf and dump communities to refer to individuals who are physically impaired. People use multiple languages to communicate with each other around the world. Sign languages include American Sign Language, Chinese Sign Language, Indian Sign Language, and others. Depending on whether motion, single-handed, or double-handed representations are present, the symbols change in each case. Sometimes, instead of using static symbols to represent letters, words like "hello," "Hai," etc. are represented using dynamic symbols. A real-time technology is going to facilitate communication between these communities. Once it has been altered using the Computer Vison technique, it can be translated into any language. Many studies in this field have been carried out in an effort to develop an accurate and efficient system. The original method employed by the researchers made use of a handcrafted feature, but it was limited and only applied in specific situations. Most studies rely on pattern recognition and feature extraction using techniques like HOG, SIFT, LBP, etc. But frequently, a single feature is insufficient for a system; in order to solve this problem, the hybrid technique was created. But we have to fix issues faster in a real-time system. These days, we employ parallel implementation to speed up our computers' processing. Most of the time, our system solves problems with a single core. The GPU system, which has more cores than a CPU system, can be used to solve issues using parallel computing [12]. Using the deep learning paradigm, we can model a self-learning system that meets our requirements. Convolutional neural networks are among the most widely used deep learning systems that are capable of solving any computer vision problem. We employed a region of interest convolutional neural network to accomplish our method in real time.

1.1 Project Domain Description

Hand signs are a communication tool employed by the dumb, and normal people have a hard time understanding what they are attempting to convey. Thus, systems that are able to discriminate between different indicators and notify the public are needed.

1.2 Application

The main purpose of the application is to give those with hearing difficulties a way to communicate. They can communicate with people who might not be familiar with sign language thanks to this technology, which translates their movements into spoken words. This technology can help kids who use sign language communicate with peers or teachers who do not in an educational setting. Additionally, it can be used into e-learning systems to offer content that is accessible for online courses. By using this technology, companies can improve inclusivity and accessibility by offering help to clients who use sign language in customer service settings.

II. LITERATURE SURVEY

Computer vision is one of the hottest technological advancements right now. It's utilized in a lot of AI-based systems, such markets, automobiles, robots, and so on. More system-related concerns are related to object detection and image classification. The sign language system can be implanted using this method. Many more methods were used in the previous systems. used literary works as the foundation for the ISL Recognition system [1]. Color segmentation is done with a glove, and recognition is done with Principal Component Analysis, or PCA. Every 20th frame, real-time data frames are used as the input for recognition. The overlapping and movement of the sign are issues with this method. For recognition, both the fingertip algorithm and PCA are used. Current research has focused on static markers of ISL [2] from still or moving images that were taken with data gloves or colored gloves in controlled environments, like with a single background and specialized equipment. The position and light are more important in the system. The signer must be familiar with the system in order for it to function in these conditions. There are several methods for preprocessing Otsu's thresholding, such as taking backdrop subtraction, motion-based segmentation, and skin tone into account [3-5]. The feature extraction stage makes use of wavelet decomposition, Fourier descriptors, and scale-invariant feature transform. Sign classification techniques include K Nearest Neighbor (KNN), Hidden Markov Models (HMM), Multiclass Support Vector Machines (SVM)[6]. Fuzzy systems, Artificial Neural Networks (ANN), and numerous others. used an edge detection method in a different investigation to identify hand gestures [7]. The frame features are retrieved using the database's edge detection and sorting features. Apply template matching to forecast the gesture using the newly created database. The least distance is used to match templates in this case. Both static symbols and dynamic gestures can be recognized by the system. employing a fuzzy and a fuzzy membership function-based approach, the system extracts the spatial aspects of signs [8]. An acceptable symbolic similarity measure is combined with the Nearest Neighbor classifier. Reheja [9] and colleagues used the Microsoft Kinect sensor device to construct an Indian sign language gesture detection system. For their research, they employed RGB and Depth Kinect images. 2. The study shows that the system's accuracy is improved when RGB-D images are used. The model extracts the HU-Moments—that is, moments that are invariant to angle, position, and shape—and feeds them as features to the SVM classifier. Pranali Loke et al. built an Android app-based system for Indian sign language [10]. The Android system gathers images and sends them to the server. These images are sent by the server system to the Matlab program, where a neural network is used to train the system and features are retrieved using the operator Sobel. Text is generated by the system through pattern recognition and classification analysis of the images. A method for recognizing American Sign Language (ASL) from depth images captured by the Kinect sensor was created by Beena M.V. et al.To train the system, a total of 1000 images representing each numerical sign were employed [11].After extracting features from the block-processed photos and training an artificial neural network (ANN), the method generated a 99.46% accuracy for the depth pictures.The system has been trained to execute more quickly on the.Convolutional Neural Network (CNN) is used in a continuation of the study for static Kinect depth picture symbols, utilizing softmax classification. The way it's implemented demonstrates that as the number of classes increases, the custom feature becomes insufficient for categorization purposes. Because it can learn from the given training data, the CNN structure will outperform other conventional approaches in terms of accuracy.

III. PROPOSED WORK

Indian sign language is a sophisticated technique that uses both hands. Convolutional neural networks are effectively applied to the image for practical usage and to increase the accuracy of categorization. The suggested system's fundamental processes are illustrated below.

Steps

1. Enter the video frame or picture.
2. Track down the handicraft.
3. Take the feature out.
4. Sorting and forecasting

The majority of the object detection problems require an image data collection and bounding box mapping to train the model. Labeling each image's bounding box is an expensive process. Additionally, using skin segmentation, we suggested a region of interest predictor. We crop the image from the segmented, constrained region and feed it to the classifier for prediction. In the first section, we input the video frame and use CLACHE (Contrast Limited Adaptive Histogram Equalization) to adjust the image's lightness using the LAB color system. Apply blurring using Gaussian blurring to the original image in the following step. We do a thresholding operation using the HSV color space to generate the skin. We can change the threshold values while running in some circumstances where the light variation is high. Finding the biggest contours in the segmented photos and drawing a rectangular box around the area that displays the output categorized result as text are the final steps. The convolutional neural network model is fed the bounding box in order to produce the sign as text.

4.1 System Design:

The project passes through a number of phases in this suggested method, including:

1) Data acquisition
2) Data Preprocessing
3) Feature Extraction
4) Gesture classification

The following methods can be used to obtain data regarding the hand gesture:

1) Making use of sensory apparatus: Electromechanical devices are utilized to precisely offer the location and configuration of the hand. Different glove-based approaches can be used to extract information. But it is expensive and not user friendly.
2) Vision based approach:In vision-based methods, the computer webcam is the input device for observing the information of hands and/or fingers. The Vision Based methods simply a camera, enabling a seamless and cost-effective contact between people and computers without the need for additional hardware. These systems often describe artificial vision systems that are implemented in hardware or software, which serves to supplement biological vision. [5].The primary difficulties in vision-based hand identification include adjusting to the great variety in the look of the human hand resulting from an enormous number of hand movements, various skin-color options, and variances in viewpoints, scalesand the speed with which the camera captures the scene.

B) Data Pre-processing:Resize all images to a uniform size to ensure consistency in input dimensions.Normalize pixel values to a common scale (e.g., [0, 1]) to mitigate variations in lighting conditions.Apply image enhancement techniques like contrast adjustment, histogram equalization, or noise reduction to improve image quality and enhance gesture features.

C) Feature Extraction: Once the hands or relevant regions are isolated, extract discriminative features from the image patches or frames to represent the sign language gestures effectively.Features are:
Shape Features: Extract geometric properties of the hand or hand contours, such as centroid, area, aspect ratio, convex hull,etc.
Appearance Features: Utilize texture descriptors (e.g., Local Binary Patterns), color histograms, or gradient-based features to capture visual appearance characteristics.
Motion Features: Compute optical flow or frame differences to capture dynamic motion information in sign language gestures.
Spatial-Temporal Features: If working with sequences of frames, extract spatio-temporal features to capture both spatial hand configurations and temporal dynamics of gestures.

D) Gesture Classification: The primary The assignment is to identify and classify different sign language gestures accurately. This involves training a deep learning model, such as a convolutional neural network (CNN), on a dataset containing images or videos of sign language gestures annotated with their corresponding meanings.

IV. ALGORITHM

Convolutional Neural Networks (CNNs) are deep learning algorithms widely used for analyzing visual data. CNNs consist of layers that extract features from input images through convolutional operations, downsample the feature maps through pooling, and perform high-level reasoning via fully connected layers. Through training with labeled data, CNNs learn to recognize patterns and objects in images, making them suitable for tasks like image classification, object detection, and facial recognition. Transfer learning, where pre-trained models are fine-tuned for specific tasks, accelerates training and improves performance. CNNs have applications in diverse fields such as image recognition, medical imaging, autonomous vehicles, and more, playing a pivotal role in advancing computer vision technology.

4.1 Convolutional Layer:

CNNs use convolutional layers to extract features from input data. These layers consist of filters (also called kernels) that slide over the input image, performing convolution operations to produce feature maps. Each filter recognizes specific patterns or features, like edges, textures, or forms, by computing dot products between the filter weights and the input data.

4.2 Pooling Layer:

Pooling layers are typically used after convolutional layers to minimize the spatial dimensions of the feature maps while retaining important information. Common pooling operations include max pooling, where the maximum value within each pooling window is retained, and average pooling, where the average value is computed. Pooling makes the learned features more resilient to tiny translations and distortions in the input data.

4.3 Fully Connected Layer:

After extracting characteristics from the input data using convolutional and pooling layers, CNNs often include one or more fully connected layers to perform high-level reasoning and classification. These layers connect every neuron in one layer to every neuron in the next layer, allowing the network to learn complex relationships between features and make predictions based on the learned representations.

4.4 Final Output Layer:

The final output layer in a Convolutional Neural Network (CNN) is responsible for producing the network's predictions or classifications based on the features extracted from the input data.

V. ADVANTAGES

1) Enhanced Communication: Sign language to speech conversion technology enables individuals who are deaf or hard of hearing to communicate effectively with those who do not understand sign language, thereby enhancing their ability to interact and exchange ideas with the broader community.

2) Real-time Translation: The development of a real-time system for sign language translation allows for instantaneous communication between sign language users and non-signers, facilitating natural and fluid interactions without significant delays or barriers.

3) Accessibility: Sign language to speech conversion technology makes information more accessible to individuals with hearing impairments by providing them with alternative means of communication, thereby reducing dependence on interpreters or written communication.

4) Adaptability: The system's ability to improve classification accuracy even with varied backgrounds and image capture angles demonstrates its adaptability to different environmental conditions, making it suitable for diverse real-world scenarios.

VI. APPLICATIONS

1) Accessibility Technology: Its principal use is to provide a way of communication for people who have hearing problems. This technology allows people to communicate with others who do not understand sign language by translating their motions into spoken words.

2) Education: This technology can help children who use sign language communicate with teachers and peers who do not. It can also be linked into e-learning platforms to make online course information more accessible.

3) Healthcare: It personnel can utilize this technology to successfully communicate with deaf or hard-of-hearing patients, ensuring that they receive appropriate treatment and comprehend medical instructions.

4) Customer Service: Businesses can use this technology in customer service contexts to assist consumers who communicate with sign language, thereby increasing inclusion and accessibility.

5) Emergency Responders: This can utilize this technology to communicate with deaf or hard-of-hearing people during an emergency, ensuring they receive prompt assistance and information.

VII. FUTURE SCOPE

1) Expansion to Other Sign Languages: While the project focuses on translating Indian Sign Language (ISL) into text, there is a scope to extend the system to support other sign languages used globally. This would involve collecting data for different sign languages, training models, and adapting the system to accommodate linguistic and cultural variations.

2) Gesture Recognition and Translation: In addition to classifying individual signs, future developments could focus on recognizing and translating complex gestures, facial expressions, and non-manual markers used in sign languages to convey meaning. This would require advancements in deep learning techniques and multimodal integration to capture the nuances of sign language communication.

3) Mobile and Wearable Applications: Developing mobile applications or wearable devices that provide on-the-go access to sign language translation services can enhance accessibility and convenience for individuals with hearing impairments.

VIII. CONCLUSION

The real-time system is designed to process numbers ranging from 0 to 9. The first step towards Indian Sign Language being recognized is this. This is the first step toward the recognition of Indian Sign Language. The 3000 static symbols of RGB. The system employed 100 images for every symbol during testing. The model was created by utilizing a region-based convolutional neural network in a deep learning system in an efficient manner. During testing, the system's accuracy for the same subject was 99.56%; however, in low light, it dipped to 97.26%. In the future, add more symbols, such as the double hand notation, from the alphabets of the static symbols used in Indian sign language. The low light difficulties need to be addressed by expanding the dataset.

REFERENCES

[1] Divya Deora, Nikesh Bajaj, Indian Sign Language Recognition, 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking, IEEE 2012-978-1-4673-1627-9/12.

[2] Anuja V. Nair, Bindu V., A Review on Indian Sign Language Recognition, International Journal of Computer Applications (0975 – 8887) July 2013, Volume 73– No.22

[3] Jorge Badenas, Josee Miguel Sanchiz, Filiberto Pla, Motion-based Segmentation and Region Tracking in Image Sequences, Pattern recognition 2001, 34, pp. 661-670.

[4] Ping-Sung Liao, Tse-Sheng Chen, Pau-Choo Chung, 2001, A Fast Algorithm for Multilevel Thresholding, Journal of Information Science and Engineering 17, pp. 713-727.

[5] Dr. Alan M McIvor, Background subtraction techniques, Image and Vision Computing, Newz Zealand 2000 (IVCNZ00).

[6] Aseema Sultana, T. Rajapushpa, Vision Based Gesture Recognition for Alphabetical Hand gestures Using the SVM Classifier, International Journal of Computer Science and Engineering Technology, Volume 3, No. 7, 2012.

[7] Purva A. Nanivadekar, Dr. Vaishali Kulkarni, Indian Sign Language Recognition: Database Creation, Hand Tracking and Segmentation, International Conference on Circuits, Systems, Communication and Information Technology Applications, IEEE 2014,978-1- 4799-2494-3/14.

[8] Nagendraswamy H S, Chethana Kumara B M, Lekha Chinmayi R, Indian Sign Language Recognition: An Approach Based on Fuzzy-Symbolic Data, 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, 978-1-5090-2029-4/16.

[9] J. L. Raheja , A. Mishra, A. Chaudhary, Indian Sign Language Recognition Using SVM, Pattern Recognition and Image Analysis, 2016, Vol. 26, No. 2, pp. 434–441.

[10] Pranali Loke, Juilee Paranjpe, Sayli Bhabal, Ketan Kanere, Indian Sign Language Converter System Using An Android App. ,International Conference on Electronics, Communication and Aerospace Technology, 2017 IEEE ,978- 1-5090-5686- 6/17.

[11] M.V. Beena and M.N. Agnisarman Namboodiri, ASL Numerals Recognition from Depth Maps Using Artificial Neural Networks, Middle-East Journal of Scientific Research 25 (7): 1407-1413, 2017,ISSN 1990-9233.

[12] Beena M.V., Dr. M.N. Agnisarman Namboodiri, Automatic Sign Language Finger Spelling Using Convolution Neural Network: Analysis, International Journal of Pure and Applied Mathematics, Volume 117 No. 20 2017, 9-15.