Sign Hand Gesture Recognition Using Machine Learning

1st Omkar Chugh
INFT Department
Vivekanand Education Society
Institute of technology
Mumbai,India
d2020.omkar.chugh@ves.ac.in

2nd Devraj Kalyani
INFT Department
Vivekanand Education Society
Institute of technology
Mumbai,India
d2020.devraj.kalyani@ves.ac.in

3rd Viren Khanna
INFT Department
Vivekanand Education Society
Institute of technology
Mumbai,India
2020.viren.khanna@ves.ac.in

4th Deepak Sharma
INFT Department
Vivekanand Education Society
Institute of technology
Mumbai,India
2020.deepak.sharma@ves.ac.in

***Abstract*— Human beings interact with each other either using a natural language channel such as words, writing, or by body language (gestures) e.g. hand gestures, head gestures, facial expression, lip motion and so on. As understanding natural language is important, understanding sign language is also very important. The sign language is the basic communication method within hearing disable people. People with hearing disabilities face problems in communicating with other hearing people without a translator. For this reason, the implementation of a system that recognize the sign language would have a significant benefit impact on deaf people social live.**

In this paper, we have proposed a marker-free, visual Indian Sign Language recognition system using image processing, computer vision and neural network methodologies, to identify the characteristics of the hand in images taken from a video trough web camera. This approach will convert video of daily frequently used full sentences gesture into a text and then convert it into audio. Identification of hand shape from continuous frames will be done by using series of image processing operations. Interpretation of signs and corresponding meaning will be identified by using Haar Cascade Classifier. Finally displayed text will be converted into speech using speech synthesizer.

Keywords— Indian Sign Language (ISL),Computer Vision(CV).

I. INTRODUCTION

A sign is a nonverbal communication method that substitutes nonverbal cues such as body parts, hand forms, hand placements and movements, arm movements, face expressions, and lip movements for spoken words. In order to communicate, most people use both words and signs. A sign language is a language in which communication takes place via actions or signs rather than sounds. As per the concept given above, sign language consists of three main parts [1].

The first crucial element is finger spelling, which implies that there is a sign for every letter in the alphabet. This kind of communication is mostly used to spell names, though occasionally location names are also spelt correctly. This can occasionally be used to emphasise or explain a specific phrase or to indicate terms for which there are no indications [2].Word level sign vocabulary, or the fact that each word in the vocabulary has a matching symbol in the sign language, is the second essential element of any sign language. When combined with a facial expression, the most popular means of communicating for those with hearing impairments is this type. The third crucial element of sign language is non-manual elements. This kind of communication uses body language, lips, tongue, eyebrows, and facial expressions. Word level sign language is really the most often used type of sign language among the deaf community out of all of these components. Hence this study offers more light on frequently used daily words or sentences and their interpretation by Sign Language Interpreter System.

The absence of grammar in Indian Sign Language (ISL) is its most significant feature. Similar to spoken languages and dialects, the structure of sign language varies based on the location and culture. Indian sign language differs from American sign language. A large database of Indian Sign Language (ISL) signs is available on the government website www.indiansignlanguage.org, which was created to empower the deaf [3]. Indian Sign Language has seen very little research done on it thus far. For this reason, having an automated method for translating Indian Sign Language would be very helpful [4].

This technique will undoubtedly benefit Indians with physical disabilities and improve society. With the use of this method, a deafmute person will be able to connect with the general public via computers. Therefore, the primary goal of

this system is to enable everyday communication between deaf individuals and the rest of the world. The teaching and learning process is falling behind in schools for the deaf because there aren't many qualified sign language teachers available. In these situations, anyone can learn or practise sign language using this system, which is intended for sign language teaching [5].

Interfaces for recognising sign language can also be utilised as a natural means of communication between people and machines. This opens up new possibilities for applications like virtual reality games, hardware-free remote controls, and human-computer interaction Additional advantages of this type of gesture-based human-computer interface include the ability to replace a physical keyboard and mouse with a virtual one, point and click, navigate in a virtual environment, interact with a three-dimensional environment, and communicate remotely.

LITERATURE SURVEY

Vision based hand gesture recognition is proposed by Ilan Steinberg and Tomer London [1] using supervised learning algorithm .This system used multiclass classifier having integrated with several binary classifier such as Support Vector Machine Algorithm(SVM) [5] to train and classify hand gestures. Initially acquired images were preprocessed using color normalization, skin detection, blob analysis filtering and feature calculation techniques. Image acquisition, hand segmentation, feature extraction and then classification based on supervised feed forward backpropogation algorithm [2][4]was used by Adithya, Vinod and Usha Gopalkrishnan for hand feature extraction having average recognition rate of 91.11%.

Research on Indian Sign Language made by Pravin Futane and Dr.Rajiv [3] used feature extraction based on shape and geometry feature and lastly learning by General Purpose Fuzzy MinMax(GFMM)neural network. Unsupervised Feature Learning Algorithm is used in designing softmax classifier to understand American Sign Language, an efforts made by Justin Chen, Debabrata, Rukmani Sundaram[6].This approach uses skin modeling using 2D Gaussian curve over 600 training dataset using kinect camera, which is a special camera that supplies dept information was used to identify hand gestures for 40 iteration to obtain estimate weights for classifier. A Hidden Markov Model (HMM) [7] being designed for data occurring over time was used by Nathan and James to correctly classify a feature vector of unknown sign. Feature vector was calculated using Hue Saturation (HSI) model, centering method and grid extraction method. Peter O'Donovan from Canada applied Restricted Boltzmann Machines (RBM) to model gestures and also provided comparison with classical neural network and k-Nearest Neighbors method [8].

Haar-like Algorithm is used for getting the region of interest from hand image preceding preprocessing techniques involving skin detection and size normalization. Fourier transformations is applied to form the feature vectors is then classified by K- Nearest Neighbor (KNN) algorithm to determine Arabic Sign Language (ArSL)[9].

The Other computer vision methods used for hand gesture recognition include specialized mappings architecture principal component analysis, fourier descriptors, neural networks [10], orientation histograms, particle filters etc. The method proposed in this paper used Haar Cascade Classifier that translates word level sign vocabulary in Indian Sign Language to textual as well as audio form.

The rest of the paper is organized as follows. In section III, the system architecture is discussed which follows the preprocessing and classification phase description. In section IV, development environment and experimental setup is explained. In section V, classification results and statistics of our system is presented. In remaining sections conclusion and future scope are presented.

II. SYSTEM ARCHITECTURE

As seen in Figure 1, the Sign Language Interpretation System operates in two stages. First comes the preprocessing phase, also known as image processing, in which various image processing techniques such as background subtraction, blob analysis, filtering and noise removal, grayscale conversion, brightness and contrast normalisation, scaling, and others are used to extract the hand shape and other distinguishable features from the image.

The second step entails classifying an image into one of the several possible gestures using the Haar Cascade Classifier, which has been trained on a training set containing examples of the various motions. These training example photos were shot under various lighting situations and from a variety of angles. Positive, negative, and test sample databases are all included in the training dataset. Positive sample photographs are those that have the ideal hand gesture, whereas negative sample images either lack the necessary gesture or merely show the background elements with no hand moment. These datasets are primarily utilised during the classification phase's training phase. A portion of the classification step may be tested using the test sample dataset.

A. Now that the setup and training have been completed, the system is prepared to interpret the input images from the videos. Next, several indicators are indicated by a database of haar cascade classifiers. Afterwards, the classifier with the highest probability is selected as the most likely interpretation of the sign. The speech synthesis phase comes after the text-to-speech conversion and is referred to as the classification or ANN phase.

B. Preprocessing Phase

TensorFlow combines models and methods from Deep Learning and Machine Learning. It runs well in optimised C++ and has Python as a handy front-end. Developers can design a graph of computations to be performed using TensorFlow. A mathematical operation is represented by each node in the graph, and data is represented by each connection. As a result, the developer may concentrate on the main logic of the application rather than worrying about small details like how

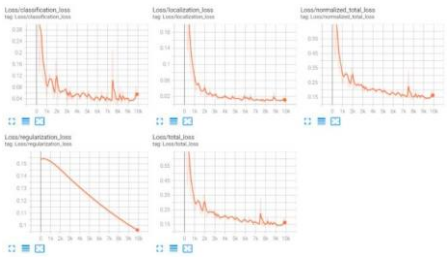
to connect the output of one function to the input of another.

We employ "checkpoints," or save points that a model creates to monitor the extent to which it has self-trained. The training process would restart at the checkpoint if it were to be interrupted. The model can protect itself from system failures thanks to this approach, even though the training process can take a long time. Fig. 5 below illustrates our model's learning rate after 10,000 training steps.

Learning rate of 10000 steps training model

The machine learning algorithm is optimised by the use of a loss function. The model's performance in training and testing sets is used to determine the loss, which is then interpreted. It is the total of all the mistakes produced in training or testing sets for every example. The loss value indicates how well or badly Classification phase is further divided into training and testing stages.

how the model acts following each optimisation cycle Our machine learning model's loss has been declining with each iteration, suggesting that the model's detection accuracy has improved. The loss of our model is depicted below



1) Training stage: Haar Cascade Classifier is trained using 500 positive, 500 negative and 50 test image samples of each gesture. These images are stored in their respective folders. These images especially positive samples are collected from different people with different hand shape, size and color and different lightening condition in various angles. Accuracy of recognition can be improved by locating area of interest in each sample images. This can be accomplished by drawing a box around the region of interest i.e. hand shape. The co-ordinates of region of interest are then analyzed to measure the contrast between each of these images. This stage will enable to build required cascade and find thresholds. Classifier uses Haar like feature like edge, line and center surround features to be trained using simple Haar function given in formula as in (1).

$$H(t)= \begin{matrix} 1 & 0\le t \le 1/2 \\ -1 & 1/2\le t \le 1 \\ 0 & otherwise \end{matrix} \qquad (1)$$

It is advised to use the greatest number of samples possible in order to attain flawless findings. It takes more effort to train a classifier to interpret different signs based on features learned during pre-processing. The training process, which trains the Haar Cascade Classifier for a certain sign, only runs once. The system is prepared to decipher indicators in the video utilising a webcam after training is complete.

2) Testing Phase: Following training, the classifier is now proficient in differentiating between various indicators. Testing is done using a webcam and live footage. This phase's output is textual.

III. DEVELOPMENT ENVIRONMENT

Webcam footage is used as the input for the Sign Language Interpretation System. The video input's frames are taken out. The system will apply a number of image processing methods on the extracted input images. In order to produce segmented output, the processed output is then classed using the Haar Cascade Classifier.

Low-resolution web cameras, mobile integrated cameras, or laptop cameras are required pieces of hardware for the system, while Emgu CV, a cross-platform programme, is needed software.Net wrapper to the OpenCV image processing library.

A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers.

The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes. The output of the convolutional layers is then passed through pooling layers, which are used to down-sample the feature maps, reducing the spatial dimensions while retaining the most important information. The output of the pooling layers is then passed through one or more fully connected layers, which are used to make a prediction or classify the image.

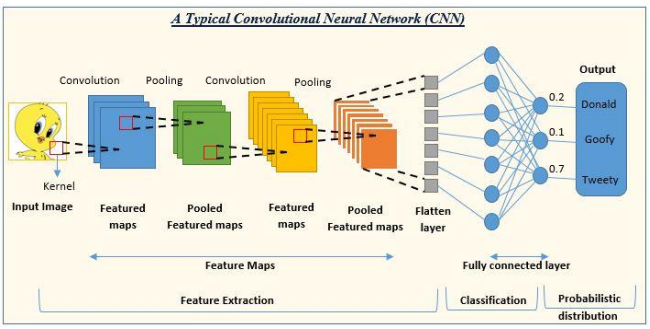


Fig. 1. Sign Language Interpreter Interface reconising .

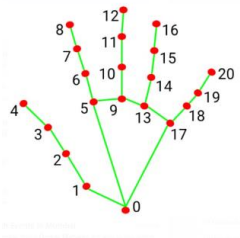
TensorFlow is a free and open-source software library for machine learning and artificial intelligence. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks.

TensorFlow was developed by the Google Brain team for internal Google use in research and production.The initial version was released under the Apache License 2.0 in 2015.Google released the updated version of TensorFlow, named TensorFlow 2.0, in September 2019.

TensorFlow can be used in a wide variety of programming languages, including Python, JavaScript, C++, and Java. This flexibility lends itself to a range of applications in many different sectors.

IV. EXPERIMENTAL RESULTS

Fully segmented words or sentences in both text and audio format are the anticipated outcome of this method. The Sign Language Interpretation will forecast the Indian Sign Language. Each hand motion in the Indian Sign Language result set (Table I) represents a distinct word. It is anticipated that this software will train and classify just a few of the sentences below. The sensor in this system is a computer-mounted video camera, and it can consistently read gestures on a personal computer at a frame rate of 18 to 21 frames per second.

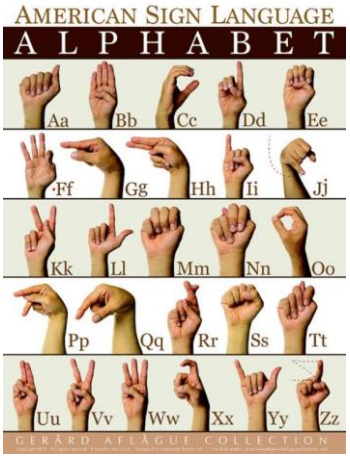


0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

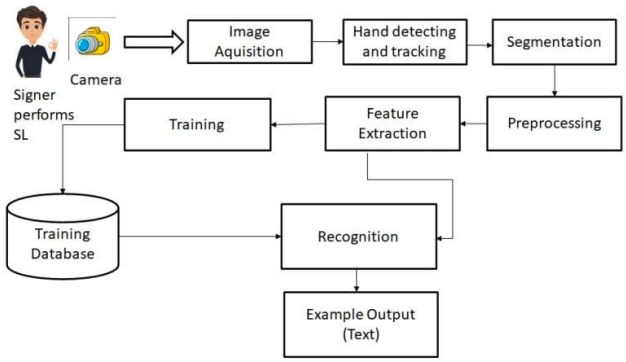
For result analysis testing is performed on two distinct signs for only one second and the results are evaluated to calculate the performance and accuracy of system.

Table displays the findings of the Sign Language Interpretation System experiment. This demonstrates the effectiveness of the Indian Sign Language Interpreter System, as seen by its average accuracy rate of over 92.68%.Even with several hands and in various lighting conditions, the system produces accurate results.Is this result good enough for this system to be used in real time?

V. DATA SET



VI. BLOCK DIAGRAM



CONCLUSION

This work presents the prediction for Indian Sign Language recognition using the Haar Cascade Classifier. Both textual and auditory signs made in front of a computer are interpreted by the system. Compared to all previous ways, this method turns out to be lot faster. Additionally, the convergence rate is faster, which enhances the accuracy and speed of sign language interpretation in the real-time system model.

The ability to complete the training setup prior to actual system use is this system's greatest advantage. As a result, during testing, it interprets data more quickly, reducing processing power and increasing efficiency. Nearly every frame is processed accurately, with an average accuracy speed of 92.68%, according to the statistics system.This system can be developed further by adding complete dataset of ISL gestures.

VII. FUTURE SCOPE

Introduction of face expression and complete body gesture in identifications of sign made by deaf individuals will add considerably to this Sign Language Interpretation System.

It is difficult to design a wide vocabulary for a sign language interpretation system since learning each sign takes time and requires careful consideration of various factors, such as backdrop illumination, hand form, size, and colour.

While generating signs is faster than speaking, the voice synthesiser object takes time to finish the speech, which is why the speech synthesis part of the sign recognition process occasionally results in a delayed answer. This only occurs when signing lengthy sentences, is avoidable by signing more slowly, and may be improved in the future.

- **Real-time Translation**: Develop models for instant translation between sign language and

spoken/written language.

- **Gesture Recognition**: Improve accuracy in recognizing complex sign language gestures,

including subtle nuances.

- **Personalized Learning**: Create adaptive platforms for customized sign language

learning experiences.

- **Multimodal Integration**: Explore combining gestures, facial expressions, and body

movements for better recognition and communication.

REFERENCES

[1] Ilan Steinberg, Tomer M. London, Dotan Di Castro ," Hand Gesture Recognition in Images and Video", Technion-Israel Institute of Technology ,2003.

[2] Adithya V., Vinod P. , Usha Gopalakrishnan, "Artificial Neural Network Based Method for Indian Sign Language Recognition", IEEE Conference on Information and Communication Technologies (ICT 2013), JeJu Island ,pp.1080-1085,April 2013.

[3] P. R.Futane , Dr. R. V. Dharaskar ,"Video Gestures Identification And Recognition Using Fourier Descriptor And General Fuzzy Minmax Neural Network For Subset Of Indian Sign Language", IEEE Conference on Hybrid Intelligent Systems (HIS), 12th International Conference, Pune,pp 525-530,Dec 2012.

[4] Corneliu Lungociu "Real Time Sign Language Recognition Using Artificial Neural Networks" , Studia Univ. Babes_Bolyai, Informatica, Volume LVI, Number 2011.

[5] J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: http://www.atm.com R. Kurdyumov,P. Ho,J. Ng. (2011, December16) *Sign Language Classification Using Webcam Images* [Online]. Available:http://cs229.stanford.edu/.../KurdyumovHoNg-SignLanguageClassificationUsi...

[6] J. K. Chen, D. Sengupta, R. R. Sundaram, IT University of Copenhagen *Sign Language Gesture Recognition with Unsupervised Feature Learning*[Online].Available:http://cs229.stanford.edu/.../ChenSenguptaSundaram-SignLanguageGestureRe...

[7] Nathan L. Naidoo, James Connan(2009),*Gesture Recognition Using Feature Vectors*, University of Western Cape, Available:http://connan.co.za/jconnan/publications.

[8] Peter O'Donovan ,"Static Gesture Recognition with RestrictedBoltzmann Machines" , ,University of Toronto,Canada, 2007

[9] Nadia R. Albelwi, Yasser M. Alginahi,"Real-Time Arabic Sign Language (ArSL) Recognition" , , Taibah University, © ICCIT 2012,pp 497-501.

[10] Jonathan C. Rupe,"Vision-Based Hand Shape Identification for Sign Language Recognition", MS in CE Thesis. RIT. 2005.