CONVERSION OF SIGN LANGUAGE TO TEXT

V.Suneetha, P.Hema Ramya Naga Sri Sai, S.Priyanka, S.Sameera Chowdary, S.Renuka, P.Joshna

Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women, Visakhapatnam, Andhra Pradesh, India.

ABSTRACT

Speech-impaired people cannot communicate through hearing or speech; instead, they use sign language. All people with speech impairments utilise sign language, but deep learning technology is now making strides in the fields of gesture and motion detection. The suggested approach aims to translate hand motions into equivalent English text in real time. It's a good idea to have a backup plan in place, especially if you're planning to travel. Similar studies have been conducted in the past, the majority of which concentrated solely on the translation of English alphabets into signs or just on numerals. For the classification of hand motions, CNN will be used. The communication gap between signers and non-signers is closed by using this approach. This will make it easier for folks who have trouble speaking to communicate. Many studies assisted us in developing the concept of artificial neural networks for our project.

Keywords; Differently abled, Hand Gesture Recognition sign Language,Network- Security, Machine Learning.

INTRODUCTION

Sign language is a language for the deaf and dumb that uses simultaneous orientation and motion of hand shapes rather than acoustically conveyed different sounds. Deaf and dumb people depend on sign language interpreters to interact. On the other hand, it takes a lot of time and money to locate qualified and experienced translators for their ongoing daily needs. The most basic form of communication for those who are deaf or hard of hearing is sign translation. Less fortunate people face challenges in their daily lives. Our goal is to create a system that makes interactions possible. The use of your hands to create shapes or motions by connecting them to your head or other bodily traits, together with distinctive facial features, is referred to as sign communication. For this reason, a classification system needs to be able to recognise different hand movements and orientations, as well as facial expressions and even hand position.

A concept for a straightforward yet expandable system that can tell apart between static and dynamic ISL motions was put out, with an emphasis on the letters a through z. Since American Sign Language is used by the majority of impaired people, it was chosen. Hand-talking, often known as gesturing, has become a popular method of transmitting information to those who are unable to speak vocally. A language that communicates words and letter combinations through gestures is known as gesture-based communication. generally find it difficult to communicate verbally with others who lack the ability to use gestures. Because of its popularity and widespread use, we used American Sign Language (ASL) for the purposes of this investigation and the use of the English language for trial and error inside this work.

LITERATURE SURVEY

Sakshi Goyal and Ishita Sharma (2015) create a Real - time system Identification System that collects the data, which is subsequently divided into numerous frames and characteristics such as Guassian difference in Centroids a Feature Extractor. Iker Vazquez Lopez (2017) created Language Transcriptor,a software which interprets hand movements in photographs by applying image identification and analysis algorithms. The identification of gestures is divided into three phases: hand location, hand segmentation, and categorization.

[2] Prof. Radha S. Shirbhate and Mr. Vedant D. Shinde (2020) built a system to develop a Sign Language Classification using various Computer Vision Algorithms such as SVM and KNN to produce an automated sign language sign language recognition implemented in real using multiple tools. Mohammad Elham Walizad and Mehreen Hurroo (2015) developed a Signs Language Recognition in that Convolutional Networks and machine vision are used in this system. Splitting is used to assess the whole skin tone region. Segmentation strategies are employed. Because the images generated by OpenCV are all shrunk to the same size, there is no visible difference between shots of different movements. Abhijith Bhaskaran, et.al. proposed the method of hand gesture recognition using smart gloves. The proposed idea is the smart gloves technique for recognition of hand gesture. The proposed idea in this system is a smart glove technique which converts sign language to speech or voice output. The smart glove consists of an Inertial Measurement Unit (IMU) and flex sensors to recognize the gesture. A state estimation method has been developed to track the motion of hands in three dimensional space. Flex sensor is used to track the fingers orientation. This prototype has been tested for its feasibility in conversion of Indian sign language to speech output. The glove is a multipurpose glove that can be used in various fields like gaming, robotics and medical field even though it is intended for sign language to speech conversion. The drawback of this system is life of sensor like gadgets is limited. Nearly 72 million of the world's population is of deaf/mute persons as per the report provided by the World Federation of the Deaf. For deaf people, learning sign language is a fundamental human right from which all other human rights can be attained. Almost 240 spoken languages are represented by sign language worldwide. For a variety of uses, there are many glove-based gadgets on the market. Glove-based technology's ability to identify movements and translate them into real-time communication for mute persons is one of its The most crucial characteristics.

A study field with significant potential for impact is sign language translation, generation, and recognition. Processing sign language would lower the barrier to communication for the silent persons. Google has developed a wristband that recognises the user's gesture, which is later determined to be only hypothetically conceivable. Moreover, TOSHIBA created an android that aids in deciphering sign language. The technology in the current systems simply translates sign language into the corresponding text or speech.

PROPOSED SYSTEM

In this strategy, the camera takes photos and stores them in the database under unique folders for each letter and number. Since grayscale only stores information about intensity, it is much simpler to apply a threshold to convert it to a binary image. The image will first be captured in RGB format and then converted to Grayscale. The images are then quickly converted to binary images via a process called grayscale thresholding.

Gaussian Filter will be applied for this because it is faster than others and has a median filter. Thresholding is essential for reducing background noise and retaining only the hand in the image. After that, the CNN layer enters the image, matches the sequence, and changes that to texts. There are two CNN layers in total. The very first CNN layer will categorize 26 symbols, while the second layer will classify similar-looking symbols.

The goal of our project is to communicate with deaf and dumb people by using a interface with high accuracy. The following conditions must be met:

- Data confidentiality: Our system must guard the confidentiality of the contents of outsourced data against both nosy providers and nefarious users.
- Flexible access control should ensure that users belonging to various groups and with varying rights are subject to flexible security policies. The confidentiality of the contents of the outsourced data should be guaranteed by these access control rules.

- Availability, scalability, and capacity. The system must consistently achieve 100% availability.
- More users must be supported by the system.
- Maintenance: The system should be optimised as much as feasible for supportability, or case of maintenance.

Base Paper Explanation

The process of communicating involves exchanging ideas and messages via a variety of means, including speech, signalling, behaviour, and pictures. Those who are Deaf and Mute (Dumb) (D&M) use their hands to make a variety of gestures to communicate with others. The nonverbal communication that takes place through gestures is comprehended visually. Sign language is the nonverbal form of communication used by the deaf and the dumb.

A visual language, sign language has three main elements. These are non-manual characteristics, word level sign vocabulary, and fingerspelling. In this project, my main goal is to create a model that can recognise hand motions based on finger spelling and combine them to create words.

The first phase of establishing this project consisted of creating the folders for storing the training and testing data. The construction of the training and testing dataset comes after the folder has been created in the second phase. Our machine's webcam will record the images in this. A blue boundary designates the Area Of Interest (ROI) in each frame.

Guassian blur is applied to the image after the ROI image has been captured, aiding in the extraction of the image's numerous elements. The next step is to create a model for training after the creation of training and testing data. In this case, automated sign language is produced using SVM and KNN algorithms, while real sign language is realised utilising a variety of tools.

Proposed System Architecture

CNNs are a type of neural network that are very effective at addressing computer vision issues. They took their cues from how our brains actually perceive vision, which occurs in the visual cortex. They utilise a filter or kernel to iteratively calculate the values of each pixel in the image and assign the proper weights to enable the detection of a particular feature. The CNN has layers like convolution, max pooling, flatten, dense, dropout, and a fully connected neural network layer, among others. Together, these layers create a very effective tool for locating features in images. Beginning with low level features, higher level features that are more sophisticated are gradually picked up by the starting layers.

The CNN Architecture functioning

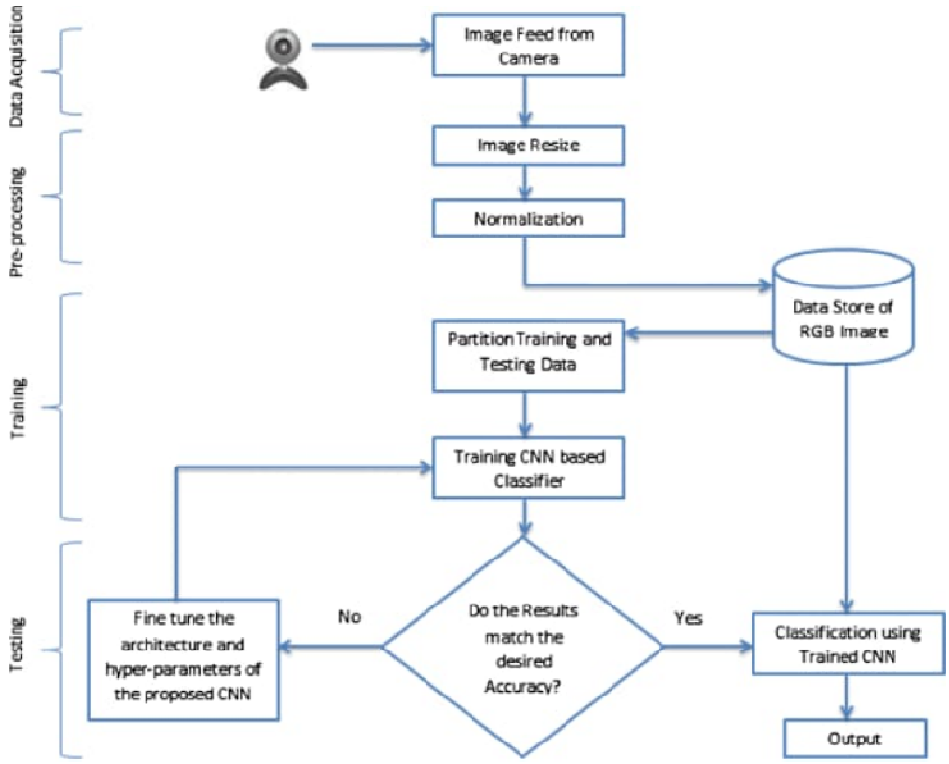
There are 11 layers in the CNN model used for this project. Three convolutional layers are present. The first convolutional layer, which is in charge of locating low level features like lines, will accept a grayscale image with a size of 50*50 pixels. With 16 filters of size 2*2, this layer generates an activation map of size 49*49 for each filter, translating to an output of size 49*49*16. The addition of a rectifier linear unit layer removes all negative values from the map and replaces them with 0.

Using a maxpooling layer, which only takes into account maximum values in 2*2 sections of the map, decreases the activation to 25*25. This process improves the likelihood of finding the target characteristic. A second convolutional layer comes next. It is in charge of recognising details like angles and curves. With 32 filters of size 3*3, this layer produces an activation map of size 23*23, meaning the output is equal to 23*23*32. By locating the highest values in 3x3 sections of the map, a maxpooling layer further shrinks the

activation map to 8x8x32. In order to recognise high level features like movements and shapes, a third convolutional layer is employed. the input is reduced by 64 filters of size 5*5 to an output of 4*4*64.

The map shrinks to 1*1*64 after the maxpooling layer. The map is compressed into a 64-element 1D array. The map is expanded to a 128 element array using a dense layer. To prevent overfitting, a dropout layer eliminates random map pieces. A dense layer ultimately lowers the map to a 44-element array, representing the number of classes. There is a matching chance of prediction assigned to each class. The predicted gesture is given as the class with the highest probability.

Block Diagram



Advantages of Proposed System

1. Sign language translation (SLT) is an important application to bridge the communication gap between deaf and hearing people.
2. Being proficient in ISL allows you to communicate with a wide range of hearing, hard of hearing, and deaf individuals.
3. The gestures or symbols in sign language are organized in a linguistic way.

System Methodology

Data acquisition: The many methods for gathering information about the hand gesture include the following:

1. Use of sensory devices: It precisely determines the configuration and position of the hands using electromechanical devices. Information can be extracted using a variety of glove-based methods. However it is pricey and difficult to use.

2. Vision-based approach: In vision-based approaches, the webcam on the computer is used as an input device to view the information of hands and/or fingers. The cost of the Vision Based techniques is reduced because they only require a camera to realise a natural interaction between people and computers. By describing artificial vision systems that are implemented in software and/or hardware, these systems frequently supplement biological vision.

The main difficulty in vision-based hand detection is dealing with the enormous variability in the appearance of the human hand caused by a great number of hand movements, the potential for different skin tones, as well as the various viewpoints, scales, and camera shutter speeds used to capture the scene.

Data Pre-Processing and Feature extraction for vision-based approach:

The method for hand detection in [1] incorporates background subtraction and threshold-based colour detection. Because hands and faces both entail comparable skin tones, we may utilise the AdaBoost face detector to distinguish between them.

By using a filter called Gaussian Blur, we may also extract the relevant image that needs to be trained (also known as Gaussian smoothing). The filter is described in [3] and is simply deployed using open computer vision (commonly known as OpenCV).

Instrumented gloves can be used, as described in [4, to extract the necessary image that will be used for training. When compared to adding filters to the data obtained through video extraction, this speeds up calculation for pre-processing and provides us with data that is clearer and more precise.

We attempted to manually separate an image using colour segmentation techniques, but because skin tone and colour are so strongly influenced by lighting, the results of our efforts were not very successful.

Furthermore, because we have a large number of symbols to train for our project, many of which resemble one another, such as the gesture for the symbol "V" and the digit "2," we decided that rather than segmenting the hand out of a random background, we should keep the background of the hand a stable single colour so that we don't need to segment it on the basis of skin colour. This would enable us to get better outcomes.

Gesture Classification

Hidden Markov Models (HMM) are employed in [1] to categorise the gestures. The dynamic elements of gestures are covered by this model. By following the skin-color blobs that correspond to the hand into a body-facial space that is centred on the user's face, gestures can be recovered from a series of video shots.

Recognizing the two categories of gestures—deictic and symbolic—is the aim. An indexing table with quick look-ups is used to filter the image. Skin colour pixels are collected into blobs after filtering. Blobs are statistical objects used to identify homogenous areas based on the location (x, y) and colorimetry (Y, U, V) of the skin colour pixels.

Naive Bayes Classifier, an efficient and quick technique for recognising static hand gestures, is utilised in [2]. It is based on categorising various gestures using geometric invariants that are obtained from segmented visual data.

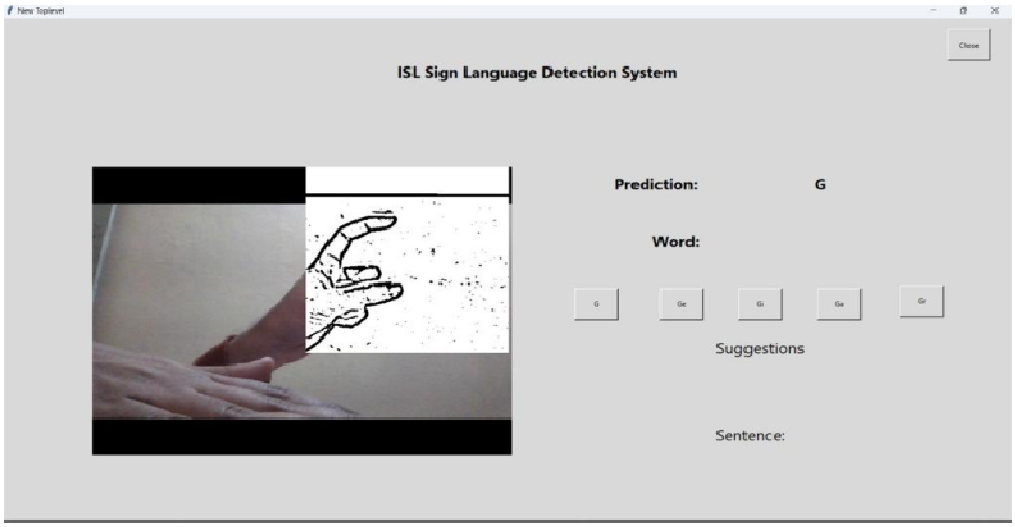
As a result, unlike many other recognition techniques, this technique is not based on skin tone. With a static background, the movements are taken from each frame of the video. The initial stage is to segment, label, and extract geometric invariants from the objects of interest. The classification of gestures is done next using

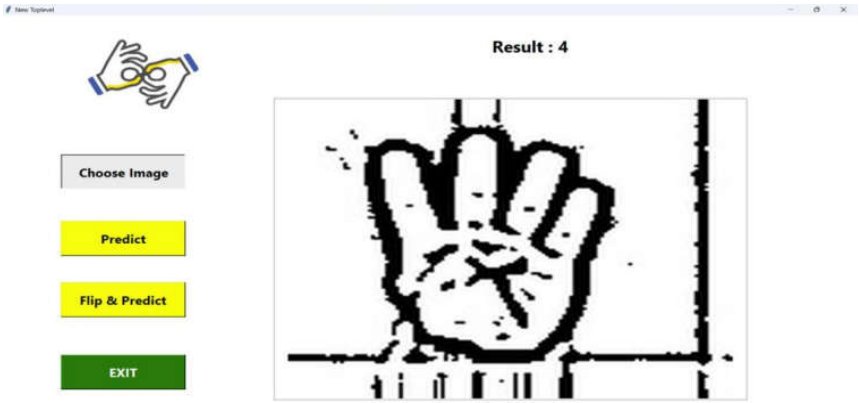
the K closest neighbour method with distance weighting (KNNDW) to give relevant information for a locally weighted Naive Bayes classifier.

Hsien-I Lin, Ming-Hsiang Hsu, and Wei-Kai Chen, graduates of the Institute of Automation Technology National Taipei University of Technology Taipei, Taiwan, report in their paper titled "Human Hand Gesture Recognition Using a Convolution Neural Network" that they have created a skin model to extract the hands from an image before applying a binary threshold to the entire image.

They calibrate the threshold picture after acquiring it in order to centre the image around the principal axis. To train and anticipate outputs, they feed this image into a convolutional neural network model. They used this model to obtain an accuracy of about 95% for the seven hand motions after training it on those.

RESULTS AND DISCUSSIONS:





CONCLUSION & FUTURE SCOPE

Project is a functional system for deaf and mute persons to convert Indian Sign Language to text in real-time. The single spelling-based signs, where each letter of the alphabet is represented by a distinct hand sign, are the main emphasis of this project at the moment. On our testing dataset, a 97.3% accuracy was attained. According to our method, the training data include background noise in the form of patterned clothing and other things that may be present in real-world circumstances.

The distance dataset, however, has minimal background noise and good lighting. By teaching the model to disregard background noise and concentrate solely on hand signals and gestures, we can improve the results.

REFERENCES

[1] T. Yang, Y. Xu, and “A., Hidden Markov Model for Gesture Recognition”, CMU-RI-TR-94 10, Robotics Institute, Carnegie Mellon Univ., Pittsburgh, PA, May 1994.

[2] Pujan Ziaie, Thomas M uller, Mary Ellen Foster, and Alois Knoll “A Na¨ive Bayes Munich, Dept. of Informatics VI, Robotics and Embedded Systems, Boltzmannstr. 3, DE-85748 Garching, Germany.

[3] Mohammed Waleed Kalous, Machine recognition of Auslan signs using PowerGloves: Towards large-lexicon recognition of sign language.

[4] Pigou L., Dieleman S., Kindermans PJ., Schrauwen B. (2015) Sign Language Recognition Using Convolutional Neural Networks. In: Agapito L., Bronstein M., Rother C. (eds) Computer Vision - ECCV 2014 Workshops. ECCV 2014. Lecture Notes in Computer Science, vol 8925. Springer, Cham

[5] Zaki, M.M., Shaheen, S.I.: Sign language recognition using a combination of new vision-based features. Pattern Recognition Letters 32(4), 572–577 (2011).

[6] N. Mukai, N. Harada andY. Chang, "Japanese Fingerspelling Recognition Based on Classification Tree and Machine Learning," 2017 Nicograph International (NicoInt), Kyoto, Japan, 2017, pp. 19-24. doi:10.1109/NICOInt.2017.9

[7] Byeongkeun Kang, Subarna Tripathi, Truong Q. Nguyen” Real-time sign language fingerspelling recognition using convolutional neural networks from depth map” 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)

[8] Byeongkeun Kang, Subarna Tripathi, T. Q. N. (2015). “Real-time sign language fingerspelling recognition using convolutional neural networks from the depth map.” 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). 10.1109/ACPR.2015.7486481

[9] Lin, H.-I., Hsu, M.-H., and Chen, W.-K. (2014). “Human hand gesture recognition using a convolution neural network.” IEEE International Conference on Automation Science and Engineering,10.1109/CoASE.2014.6899454.

[10] Pigou, L., Dieleman, S., Kindermans, P.-J., and Schrauwen, B. (2015). “Sign language recognition using convolutional neural networks.” 8925, 10.1007/978-3-319-16178-5_40, 572–578.

[11] Tie, Y. and Xu, Y. (1994). “Hidden Markov model for gesture recognition.” CARNEGIE MELLON UNIV PITTSBURGH PA ROBOTICS INST, 10.21236/ada282845