STS System: A Sign to Speech and Text, Speech to Sign and Text, and Text to Speech and Sign Converter using DNN

Dr.CHANDRA MOHAN

[1]B.Tech Scholar, Dept. of ECE, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Medchal, Telangana-501301

2Associate Professor, Dept. of ECE, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Medchal, Telangana-501301

[3]Assistant Professor, Dept. of ECE, Vignan's Institute of Management and Technology for Women, Kondapur, Ghatkesar, Medchal, Telangana-501301

Abstract

Communication is the most important skill that everyone needs to express their views to others. If you can speak properly and if the other person can understand properly then no issues, you can communicate easily. But imagine if a person who wants to communicate can't speak and here and the person with whom the first person is communicating can't see. In this case, communication is impossible because deaf and dumb can't hear what a blind person is saying and blind one cannot see what a deaf and dumb person is showing using sign language. The world is always filled with equal opportunities, but in the above case if that kind of people meets and wants to communicate or to mingle with the other category of community it's a quite difficult task. So we need a bridge between these two poles to make them meet and to create equal opportunities to communicate in any kind of scenario. The bridge is nothing but technology. We are going to use a kind of Technology In our proposed system that can convert sign language into speech and text, and text to speech and sign, and speech to sign and text. So that, even if a blind person wants to communicate with a deaf and dumb person they can communicate very easily using this technology. The main objective behind our proposed system is to vanish the communication gap between normal people; hearing and speech impaired, and blind people. The system is a tri- directional communicative solution that can solve all the above problems mentioned and create equal opportunity for communication. We can use a software-based solution in the form of an app or a hardware-based solution using Raspberry Pi in image processing combined with DNN, such that it converts speech to sign and text, sign to speech and text, text to sign, and speech.

Keywords: STS System, Image Processing, DNN, OpenCV, Sign to Text and Speech, Text to Sign and Speech, speech to Text and Sign Conversion, Machine learning, Raspberry Pi, Tri-Directional Communication.

1. Introduction

Nowadays technology is seen everywhere. Very complex problems are made simple using technology. In almost all fields like the biomedical field, in education, in transportation, marketing, banking, and even in cheating, people make use of these human inventions. So, I want to use this technology to give equal opportunity in “view exchange”. That is nothing but communication. If you are speaking to your friends, you can easily express your views to

them and they can easily understand what you are narrating because, each of you can understand the other's language. But imagine if a blind person wants to communicate with a deaf and dumb person, is a bit impossible task because, a blind cannot see actions that are made by a deaf and dumb person in his/her sign language, and a deaf and dumb person cannot hear what blind person is saying. Hence, there will be no communication between them. To overcome this problem a "Tri-directional communication system" need to be implemented, that can convert speech or text to sign, and sign to text or speech. Thus, a Blind person can easily communicate with a deaf and dumb person, and vice versa.

To implement this tri-directional communication system we are going to use the most thriving technology in the world, i.e., Image Processing in combination with Deep Learning. To convert the text to sign we're going to use an algorithm called DNN in combination with an image processing tool kit named OpenCV.

Table1: List of Notations Used

RCNN	Region-based Convolutional Neural Networks
CNN	Convolutional Neural Network
DNN	Deep Neural Network
HMM	Hidden Markov Model
ML	Machine Learning
HOG	Histogram of Oriented Gradients
SVM	Support Vector Machine
LDA	Linear Discriminant Analysis
TTS	Text To Speech
SSL	Spanish Sign Language
ASL	American Sign Language
BSL	British Sign Language
ISL	Indian Sign Language
CSL	Chinese Sign Language
ArSL	Arabic Sign Language
GIF	Graphics interchange Format
RGB	Red, Green, Blue

NLP	Natural Language Processing
API	Application Program Interface
SD	Secure Digital
USB	Universal Serial Bus

2. Literature Review

D. Manoj et al. (2020) proposed a system named Easy talk: a translator for Sri Lankan sign language using machine learning and artificial intelligence. This product comprises an application that uses NLP-based API and natural language Toolkit (nltk) library in NLP in combination with machine learning-based API. They have created different components to perform for tasks namely "hand gesture detector" to capture hand signs using pre-trained models, an "Image classifier" that detects hand signs faster-using RCNN based models. A "text to audio generator" is based on machine learning API and a "Text to sign converter" that uses a reverse translator to convert given text input to GIF images representing sign language [1].

Lance Fernandes et al. (2020) proposed a conventional neural network-based bi- directional sign language translation system that has software consisting of code that captures images to create a versatile custom data set and is preprocessed by converting it to a gray scale image, then to an inverse binary image to recognize the given input sign and gives speech or text as output the system possesses an accuracy of about 99.9 8% it also consists of a speech to sign conversion software to convert the speech to text and a video of image sequences consisting of the gestures [2].

Muttaki Hasan et al. (2016) proposed a machine learning-based approach for the detection and recognition of Bangla sign language that uses OpenCV as an image processing library and Python as a programming language to perform the detection and hand gesture recognition is performed using HOG (Histogram of Oriented Gradients) and SVM (Support Vector Machine) as a classifier and gives output text. This text is taken as a string by the Python compiler and converted into audio through the TTS (Text To Speech) engine using the pyttsx Python library [3].

Mateen Ahmed et al. (2016) proposed a software-based solution named Deaf talk using a 3D animated sign language interpreter using Microsoft's Kinect V2, for people with hearing and speaking disabilities. They have used Kinect V2 to create a 3D humanoid model in real-time that is displayed on the screen as a sign language animation output for the given input text or speech. The system possesses 84% accuracy for sign language to speech and 87% for speech to sign language conversion [4].

Kohsheen Tiku et al. (2020) proposed a Real-time conversion of sign language to text and speech by creating a vision-based application for deaf and dumb people. In this system, 26 ASL alphabets are recognized in real-time from images that are 200x200 pixels in RGB format using a customized SVM model. To achieve this they have used the Oneplus 6

smartphone with Oxygen OS (based on Android Oreo) and the algorithm which is developed using a java-based OpenCV wrapper [5].

Tanuj Bohra et al. (2019) propose a real-time two-way communication system for the speech and hearing impaired using computer vision and deep learning, it is a software- based Paper which can predict 17600 test images in 14 seconds with an average prediction time of 0.000805 seconds with an accuracy of 99% using CNN model. To achieve two-way communication they have used the combination of computer vision, image processing, and machine learning algorithms to get significant results [6].

Mahesh Kumar N et al. (2018) propose an intelligent sign language recognition system using image processing and Mat Lab this system contains four different models to perform the image processing segmentation named pre-processing and hand segmentation in which Eigen values and eigenvectors are extracted, the second one is feature extraction in which RGB to gray scale and gray scale to the binary conversion will take place, next is sign recognition in which dilation and erosion take place, the last one is signed to text conversion in which the pre-processed data is converted into text. The system also uses the LDA algorithm through which dimensionality is reduced and accuracy is increased [7].

Nan Song et al. (2018) proposed a system named "A gesture to emotional speech conversion by combining gesture recognition and facial expression recognition" using DNN (deep neural network) to recognize the sign language and the facial expressions, and SVM (support vector machine) to give the output text of sign and emotional types of facial expressions and HMM based system is used for emotional speech synthesis [8].

Aishwarya V et al. (2018) proposed a system named " Hidden Markov model-based Sign Language to Speech Conversion System in TAMIL" Using an accelerometer- gyroscope sensor-based hand gesture recognition module that converts the hand gesture into Tamil phrases then to audio using HMM-based text to a speech synthesizer. The system makes use of Raspberry Pi and MPU6050 sensor and is attached to gloves to perform the tasks. The overall performance of the system is recorded as 80 - 90% [9]. Tariq Jamil et al. (2020) proposed a text to sign conversion system named "design and implementation of an intelligent system to translate Arabic text into Arabic sign language". The system uses Java programming language bundled with screen builder applications for creating user interfaces. The system also uses a toolkit named "Farias" to achieve quick and accurate processing of Arabic text. After recognition of text, the output is shown in the form of GIF images representing the sign language [10].

The existing system has four main stages there are,

1. Gesture Reading

2. Gesture Processing

3. Speech Conversion

4. Output Modules

In the first stage, the system takes the input as a gesture and sends it to the next stage to process the gesture using some image processing algorithms. In the 2nd stage using image processing that gestures are converted into gray scale, using image recognition algorithms the

processed gesture that is the binary or gray scale gesture is matched with the examples that we have given before to the system, finally validated sign is sent to next stages to convert it into speech or text. In the third stage the process finds all the validated sign is then converted to speech or text using an open CV or TTS text to speech algorithms are conversion methods. In the last phase or stage of the process, the data in the form of speech or text is given to an output module and is received by the end-user in the form of audio, if it is a speech and in the form of text that is displayed on a monitor or a screen.

In the existing system, they have just used image processing to convert the sign to speech or text but in our proposed system we are going to use image processing combined with machine learning and DNN Algorithms, which will help to enhance more user interface and accuracy of the output to the given input.

3. Methodology

Image processing is a method to perform some operations on an image to extract some useful information from it. It is a type of signal processing in which input is an image and output is processed data of the image i.e., binary data. Simply image processing is nothing but the manipulation of the images, by using computer vision we can analyze those manipulated images, and process them to retrieve some data.

Majorly we have 9 steps in image processing. They are,

1. Image Acquisition
2. Image Enhancement
3. Image Restoration
4. Color Image Processing
5. Wavelets
6. Comparison
7. Morphologic Processing
8. Image Segmentation
9. Image Recognition and Interpretation.

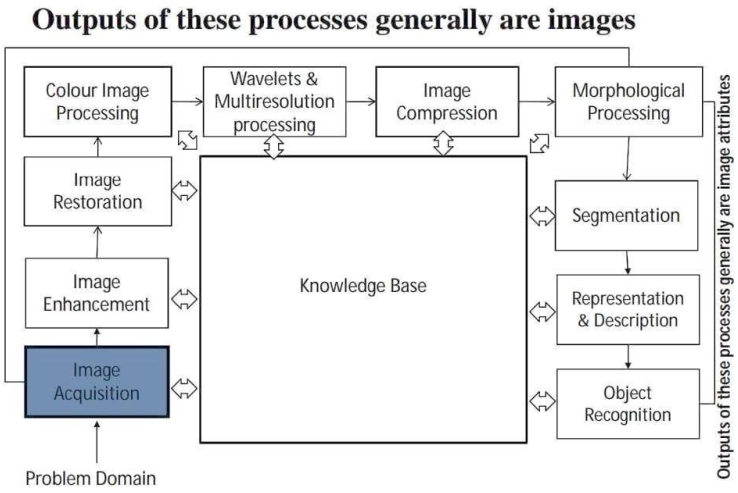

Figure1: Fundamental Steps in Digital Image Processing

Step 1: Image Acquisition

The images captured by a sensor that is a camera, and digitized if the output of the camera or a sensor is not already in digital form, using analog to digital converter.

Step 2: Image Enhancement

Image enhancement is the process of manipulating an image so that the result is more suitable than the original for specific applications. The idea behind Enhancement techniques is to bring out hidden details, or simply to highlight certain features of interest in an image.

Step 3: Image Restoration

Improves the appearance of an image. Tends to be mathematical or probabilistic models. Enhancement, on the other hand, is based on human subjective preferences regarding what constitutes a good enhancement result.

Step 4: Color Image Processing

Use the color of the image to extract features of interest in an image.

Step 5: Wavelets

Wavelets are the foundation of representing images in various degrees of resolution. It is used for image data compression.

Step 6: Compression

Techniques for reducing the storage required to save an image are the bandwidth required to transmit it.

Step 7: Morphological Processing

Tools for extracting image components that are useful in the representation and description of shape. in this step there would be a transition from processes that output images to processes that output image attributes.

Step 8: Image Segmentation

Segmentation producers partition an image into its constituent parts objects The more accurate the segmentation the more rightly recognition is to succeed Step 9: Representation and Description. Representation: decide whether the data should be represented as a boundary or as a complete region. It almost always follows the output of a segmentation stage. Boundary Representation: focus on external shapes characteristics such as corners and inflections. Region Representation: focus on internal properties such as texture or skeleton shape. Description: also called feature selection, deals with extracting attributes that result in some information of interest

Step 10: Recognition and Interpretation, Recognition: the process that assigns a label to an object based on the information provided by its description. Knowledge Base: knowledge about a problem domain is coded into an image processing system in the form of a knowledge database.

4. Experiments and Results

By analysing some papers, we have got an idea of existing systems that have already been proposed by others. We have chosen one model as our existing system from which we have developed our proposed system.

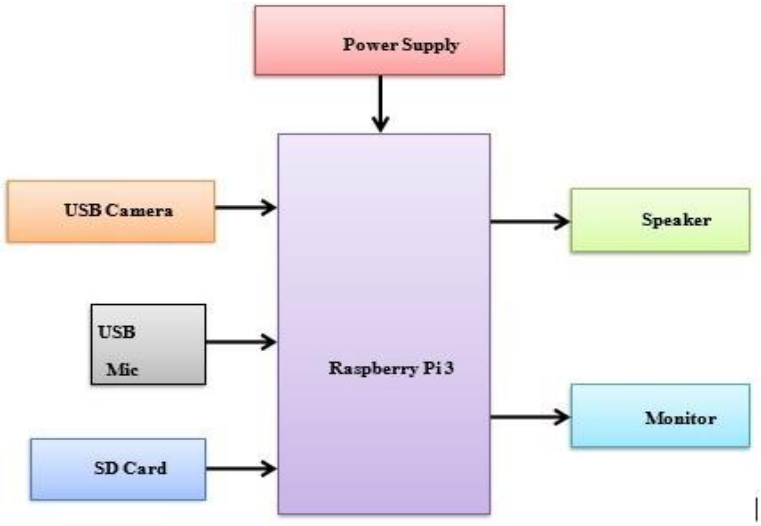

Figure2: Block Diagram of Proposed System

We have chosen Raspberry Pi as our microcontroller so we have connected USB camera, USB mic, and SD card as input devices and output devices are speaker and monitor is also connected to the Raspberry Pi whole system is given a power supply to make it run. The input device; camera is used to capture the signs of hand gestures or movements of the user as an input and store it in an SD card, after processing those gestures it will show text output on the monitor and speech output through the speaker. Another input device USB mic is used

to take or collect the user input in the form of speech and again it is stored in the SD card after processing that's speech signal the output is shown in the form of sign using a monitor and also that text is displayed below the sign in the same monitor.

We will connect a keyboard to the monitor or to the Raspberry Pi, which helps us to give the input as text and that text is processed and then output is shown in the form of a sign on the monitor and a speech output through the speaker.

This is the method that we follow to initiate the hardware process but to run this hardware process, we need software support. We have drawn a design flow chart under the standard software approach in this proposed system. The flowchart is as follows,

Most of the existing systems have only worked on bi- directional Communication algorithms that will only convert sign to speech and text and speech to text but it will not convert speech or text into the sign. In our proposed system we are going to focus on this point of converting speech is text to sign. We are going to convert the bi-directional communication system into a tri-directional communication system.

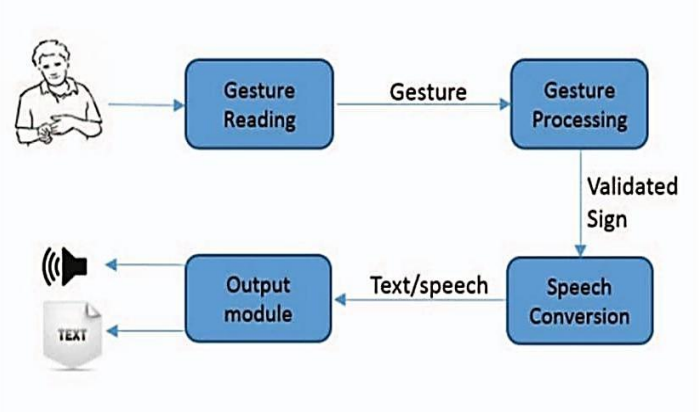

Figure3: Speech to Sign Model of Existing System

Table2: Comparison between CNN and SVM

CNN	SVM
Conventional Neural Network	Support Vector Machine
Increases overall classification performance to 7.7%	Little less performance than CNN
Accuracy is 94%	Accuracy is 80 to 90 %
Used in high precision requirements	SVM combined with CNN to givebetter results

Table3: Comparison between RNN and CNN

RNN	CNN
Ability to process temporal information	Incapable of effectively interpreting temporal information
Process data in sequence like sentence	Can process images and videos
Used in speech recognition	Used in image processing algorithms
Less feature compatibility	More feature compatibility than RNN
RNN is a sequential data algorithm inML	Artificial neural network algorithm indeep learning

Table4: Comparison between ML and Deep Learning

Machine Learning	Deep Learning
ML is computers being able to thinkand act with less human intervention	It is about computers learning to think using structured models of human brain
Can analyze only images but not videos	Analyzes images videos and unstructured data

Table5: Comparison between HOG and SVM

HOG	SVM
Histogram of Oriented Gradients	Support Vector Machine
Used for feature extraction in human detection process	Linear algorithm used for human classification

Table6: Comparison between Glove based and non-Glove based gesture detection methods

Glove Based Sign Recognition	Sign Recognition Without Glove
Signs are detected using sensor attachedto the glove	No external sensors are used or no needto ware any glove to detect the gesture
Not potable	Comparatively portable
Simple algorithm are used	Complex algorithms are used
Less accuracy and performance	More accurate than glove based gesture recognition

Table7: Parameter Wise Analysis Of The Research Work

S. No	Accuracy	Type of Sign Language	Images Taken For Data Set	Images Taken For Testing	Method or Algorithm Used
1.	97%	SSL	250 Images, 26 Classes	38 Samples for each Class	RCNN
2.	99.22%	ASL	34,627	7,172	Glove-Based H/W Device + CNN
3.	86.53%	SSL, BSL	320 Samples, 16 BSL Expressions, 9 Classes	64 Data Sets, 20 Samples for each Class	SVM & HOG with ML
4.	87%	ASL	100 Test Runs with 3 People	-	Kinect V2, 3D Modeling
5.	97%	ASL	100 Images, 27 Classes	20	ML, HOG, SVM, OxygenOS
6.	99%	ASL	70% of Testing Images	17600	Deep Learning & CNN
7.	-	ISL, ASL	10 Samples + 26 Classes = 260 Images	1 Image for each Class	MATLAB & LDA
8.	90.7%	ASL, CSL, Facial expressions	2515 Images of 36 Static Sign Language and 6 Facial Expressions	100 Sets	DNN, SVM &HMM
9.	87.5%	Tamil Sign Language	90 Samples for 16 Gestures	10 Samples for each Gesture	HMM + Glove-Based Device
10.	85%	ArSL	Images of 28 Letters	1 Image for each Class	Word Parsing,Java Programming with Toolkit "Farasa"

5. Conclusions and Future Scope of Work

The proposed system will play a crucial role in helping blind people to equally and normally communicate with deaf and dumb people. If we want to implement a portable system then the Software Solution of an Android application using ml kit and Java libraries are made. f we want to make a device for mass population then an IoT-based embedded system that uses image processing with DNN and OpenCV methods is made.

References

[1] D. Manoj Kumar, K. Bavanraj, S. Thavananthan, G.M.A.S. Bastiansz, S.M.B. Harshanath, J. Alosious. Easy-Talk: A Translator for Sri Lankan Sign Language using Machine Learning and Artificial Intelligence. In 2020, IEEE International Conference on Advancements in Computing (ICAC), pages 506-511.IEEE-2020.

[2] Lance Fernandes, Prathamesh Dalvi, Akash Junnarkar, Professor Manisha Bansode. Convolutional Neural Network based Bidirectional Sign Language Translation System. In 2020, IEEE Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020), pages 769-775.IEEE-2020.

[3] Muttaki Hasan, Tanvir Hossain Sajib, Mrinmoy Dey. A Machine Learning Based Approach for the Detection and Recognition of Bangla Sign Language. In 2016, IEEE 2016 International Conference on Medical Engineering, Health Informatics and Technology (MediTec), pages 1-5. IEEE-2016.

[4] Mateen Ahmed, Mujtaba Idrees, Zain ul Abideen, Rafia Mumtaz, Sana Khalique. Deaf Talk Using 3D Animated Sign Language a Sign Language Interpreter using Microsoft's Kinect v2. In 2016, IEEE SAI Computing Conference (SAI), pages 330-335. IEEE-2016.

[5] Kohsheen Tiku, Jayshree Maloo, Aishwarya Ramesh, Indra R. Real-time Conversion of Sign Language to Text and Speech. In 2020, IEEE Second International Conference on Inventive Research in Computing Applications (ICIRCA-2020), pages 346-351. IEEE-2020.

[6] Tanuj Bohra, Shaunak Sompura, Krish Parekh, Purva Raut. Real-Time Two Way Communication System for Speech and Hearing Impaired Using Computer Vision and Deep Learning. In 2019, IEEE Second International Conference on Smart Systems and Inventive Technology (ICSSIT 2019), pages 734-739. IEEE-2019.

[7] Mahesh Kumar N B. Conversion of Sign Language into Text. In 2018, IEEE International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 9 (2018), pages 7154-7161. IEEE-2018.

[8] Nan Song, Hongwu Yang, Peiwen Wu. A Gesture-to-Emotional Speech Conversion by Combining Gesture Recognition and Facial Expression Recognition. In 2018, IEEE First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia), pages 1-6. IEEE- 2018.

[9] Aiswarya V, Naren Raju N, Johanan Joy Singh S, Nagarajan T, Vijayalakshmi P. Hidden Markov model-based Sign Language to Speech Conversion System in TAMIL. In 2018, IEEE 4th International Conference on Bio-signals, Images and Instrumentation (ICBSII), pages 1-4. IEEE-2018.

[10] Tariq Jamil. Design and Implementation of an Intelligent System to translate Arabic Text into Arabic Sign Language. In 2020, IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pages 1-4. IEEE-2020.