HAND GESTURE RECOGNITION
&
VOICE CONVERSION

Signal and Systems

Submitted in partial fulfillment for the award of the course
Of
BACHELOR OF ENGINEERING
In
COMPUTER SYSTEMS ENGINEERING

Submitted By:
ZAIN UL ABIDIN ,
ZAGHUM ABBASS,
WAQAR AHMED

Guided by:
Dr: Junaid Bhatti

Department of Computer Systems Engineering
Sukkur IBA University

ABSTRACT: Communication is main method about inter-person communication. Number about deaf & dumb persons has dramatically increased in recent years due to birth defects, mishaps, & mouth diseases. Individuals who are deaf or dumb must use a visual medium to communicate among others because they are unable to do so among hearing people. Many different languages are used & translated all around world. People who have trouble hearing & speaking are referred to as "Special Persons." other person has trouble understanding what "The Dumb" & "The Deaf" persons are trying to say, respectively. In order for dumplings to communicate among regular people, sign language is essential. Speaking among others who aren't silent is really difficult for silent people. because public is not taught hand sign language. They find it quite challenging to communicate in an emergency. solution to this problem is to translate sign language into audible speech. There are numerous efficient techniques for spotting hand motions or gestures, such as voice-to-text conversion using CNN & SVM algorithms. Here proposed study utilised SVM technique, however Python SVM is not accurate in identifying hand gesture, therefore we are using deep learning Convolution Neural Network to train hand gesture photographs, & then this trained model can be used to predict those taught hand gesture from webcam.

1. INTRODUCTION

In today's rapidly evolving technological landscape, advancements have revolutionized how we navigate daily life. However, amidst this progress, certain groups, notably individuals who are deaf or mute, face significant communication challenges that have often been overlooked.

The inability to easily communicate with others can lead to feelings of isolation and hinder opportunities for meaningful interaction. Recognizing this, researchers have turned their attention to developing technologies aimed at bridging this communication gap.

One such innovation is HGRVC (Hand Gesture Recognition & Voice Conversion) technology, which utilizes web cameras and advanced algorithms to track and interpret hand gestures. By converting these gestures into text or speech, HGRVC facilitates communication between individuals who are deaf or mute and the broader community.

Additionally, efforts are underway to improve the recognition of sign language through visual approaches, leveraging camera-based systems to interpret hand movements. While these advancements hold promise, challenges such as the need for constant hand visibility persist.

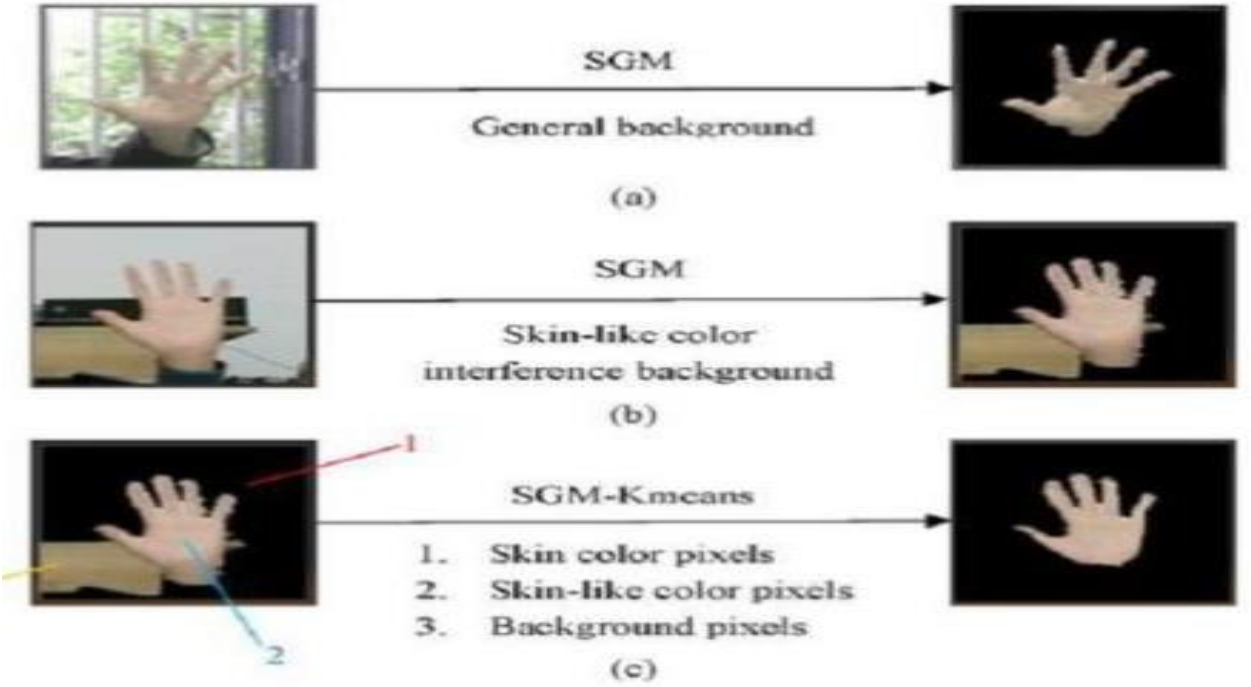


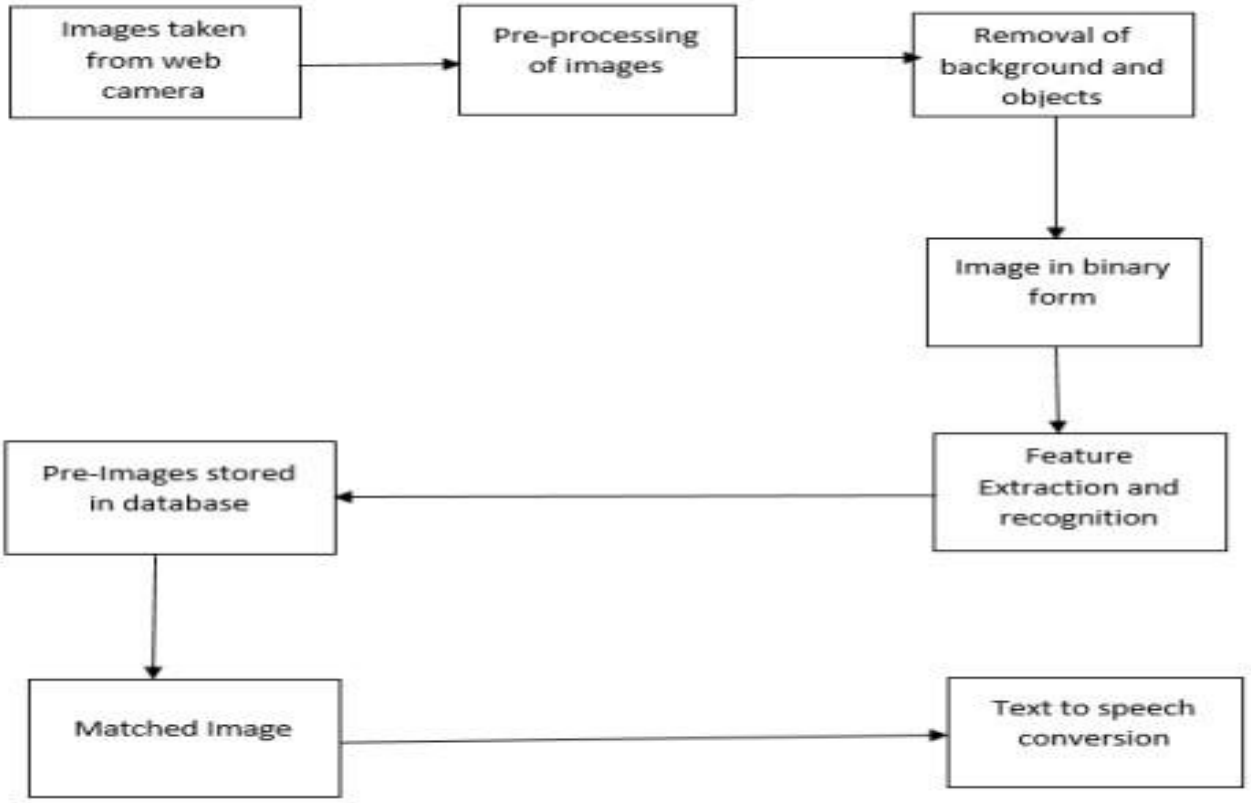
Fig.1: (a) SGM segmented image in general background. (b) SGM segmented image in skin-like background. (c) SGM-Kmeans segmented image in skin-like background.

2. PROPOSED SYSTEM

Here about this research is constructing a machine learning model that can predict hand gesture from a camera & then turn recognised gesture into voice so that non-Deaf & non-Dumb people may understand what Deaf & Dumb people are saying. We are using a deep learning Convolution Neural Network to train hand gesture images, & we are using that trained model to predict those learnt hand motions from webcam. we used SVM technique in suggested investigation, although Python SVM is not reliable for distinguishing hand motion.

This project aims to develop a system that can convert hand gestures into text. project's objective is to add photographs to database, which will match them & convert them to text. As part about detection process, hands are observed in motion.

method generates text output, reducing communication gap between humans & deafmutes.



2.1 Training about System:

The user must enter desired number in order to store samples in database. More samples should be used to get greater accuracy than 5. user must select folder where photos will be saved. Click start video to start web camera & start database creation process. Click capture image to add specified number about photos to training folder for each sample. When number about images displayed matches number about successfully taken images, database construction is complete.

2.2 Image Pre-Processing:

The acquired images are pre-processed to enhance their intrinsic features. Pre-processing basically involves removing foreground & backdrop about an image to focus solely on hand gestures. preprocessed image is then shown as a series about binaryized (all-black & all-white) pixels.

2.3 Feature Extraction & Recognition:

Feature extraction is part about dimensionality reduction process, which reduces size & complexity about a starting collection about raw data into manageable chunks. As a result, processing will be less complicated for you.

MODULES:

1. Upload Hand Gesture Dataset
2. Preprocess dataset
3. Model Generation
4. Train CNN Gesture Images
5. Sign Language Recognition from Webcam
6. Extract image from webcam
7. Convert image to binary or grey format & back ground removal
8. Extract features from image
9. Recognition & play audio

3. IMPLEMENTATION

• Here proposed study utilised SVM technique, however Python SVM is not accurate in identifying hand gesture, therefore we are using deep learning Convolution Neural Network to train hand gesture photographs, & then this trained model can be used to predict those taught hand gesture from webcam.

CNN ALGORITHM:

Convolutional neural networks, often known as CNNs or ConvNets, are particularly skilled at processing input among a grid-like design, such as images. A digital image is a binary representation about visual data.

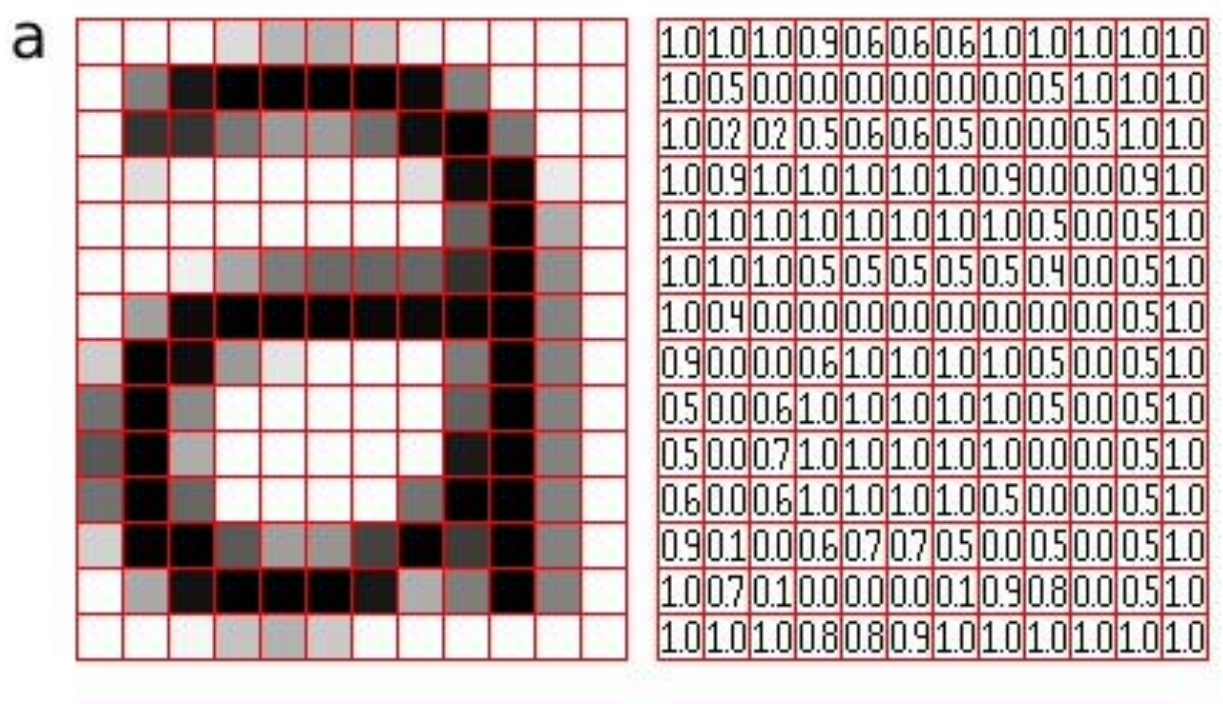


Fig.3: Representation about image as a grid about pixels

The human brain starts processing a huge amount about data as soon as we perceive a picture. Each neuron has a unique receptive field, & because they are connected to one another, they collectively encompass whole visual field. Similar to how each neuron in biological vision system responds to stimuli only in confined area about visual field known as receptive field, each neuron in a CNN processes data only in its receptive field. layers first pick up on lines, curves, & other simpler patterns before moving on to more complex patterns like faces & objects. Using a CNN, one can give computers sight.

Convolutional Neural Network Architecture:

The conventional three layers about a CNN are fully connected, pooling, & convolutional layers.

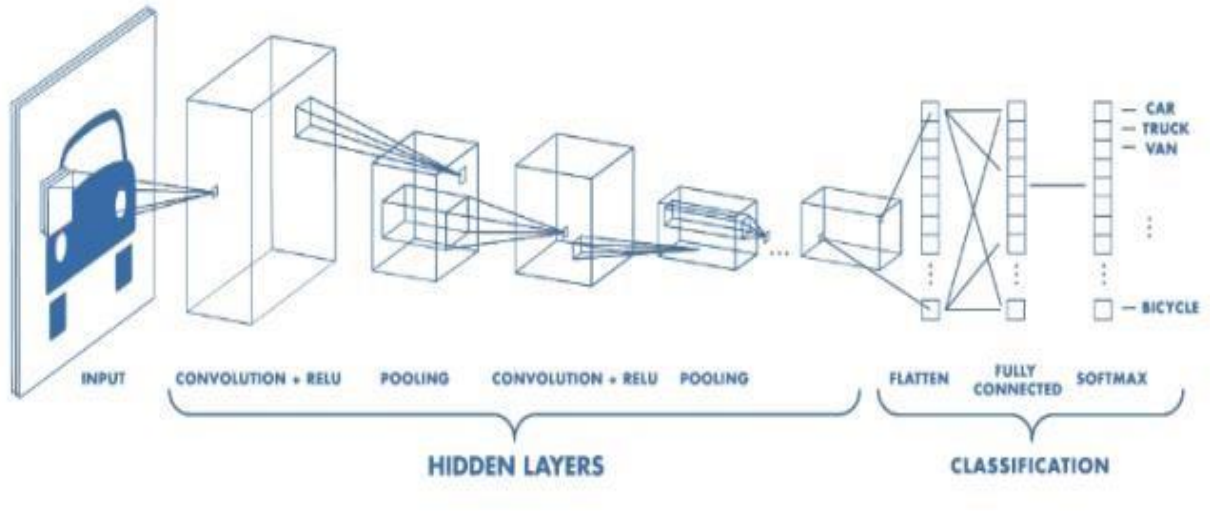


Fig.4: CNN architecture

Convolution Layer: convolution layer is a key component about CNN. It carries majority about network's computational load. kernel—a group about learnable parameters—and limited region about receptive field are two matrices that are combined in this layer to form a dot product.

Pooling Layer: At specific points, pooling layer takes place about network's output through getting a summary statistic from nearby outputs. This helps to reduce spatial size about representation, which minimises amount about computation & weights required. pooling procedure is applied to each slice about representation independently.

Similar to a regular FCNN, fully connected layer has complete connections between every neuron in it & every neuron in layer above & below it. As a result, it can be estimated using standard matrix multiplication & bias effect. among help about FC layer, representation between input & output is mapped.

5. EXPERIMENTAL RESULTS

Step 1:Collect Images and Create Data Sets

Step 2: Train Classifier

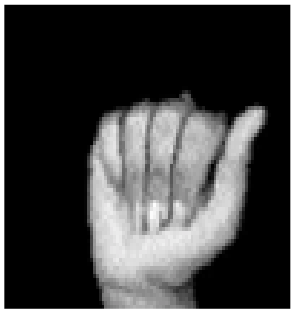
Train image 1

Train image 2

Train image 3

Test Image 1

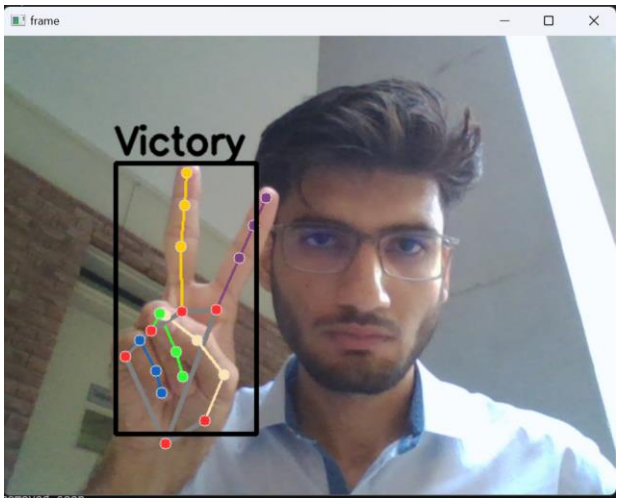
Test Image 2

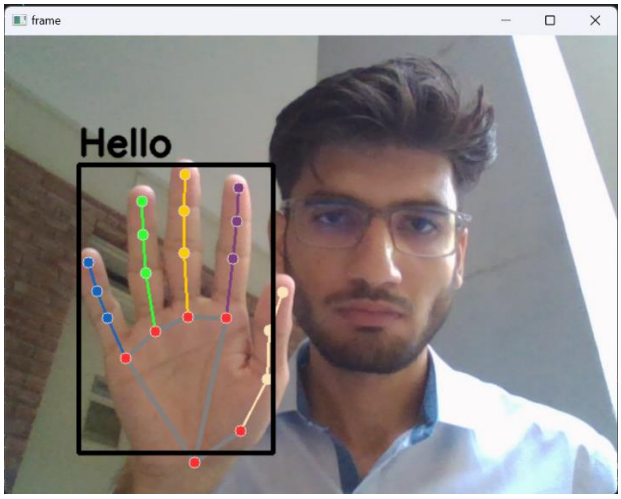


Test Image 3

```
100.0% of samples were classified correctly !
```

Step 3: Interface Classifier





Here, all you have to do is act out motion as it appears on screen above. If your hands are adjusted, forecast might be off; however, if your gesture is fixed, it will be on money. When project is run, modules listed below are executed for each forecast.

- Webcam image extraction, binary or grayscale conversion, & background removal
- Play music & recognise & extract visual features

6. CONCLUSION

For dumb & deaf persons, image processing has been used to translate voices & recognise hand movements. method takes an image as input & outputs text & audio. implementation about this system offers accuracy about up to 90% & performs well in most test situations. objective about this project is to build a machine learning model that can forecast hand motions from webcam footage & then convert recognised hand gestures into voice, enabling hearing & hearingimpaired people to converse among regular people. finished product is shown as static text after

captured image has been retrieved from image dataset. A hand motion is transformed into a picture through feature extraction & categorization.

Reference;

- **Github**
- **Stackoverflow**
- **Research Gate**