Hand Gesture Recognition A Literature Review

Aditya Verma[1] Aleena Khan[2] Aman Dubey[3] Kavita Namdev[4] Ritesh Khedkar[5]
[1,2,3]Final year CSE Student [4,5]Sr. Associate Professor
[1,2,3,4,5]Department of Computer Science Engineering
[1,2,3,4,5]Acropolis Institute of Technology and Research, Indore-452007, Madhya Pradesh, India

Abstract— This paper aims to discuss a gesture recognition approach that focuses on hand gestures. We propose a novel deep learning architecture that uses a Convolutional Neural Network (CNN) and OpenCV library. This HGR based project will be useful for mute people in communicating or expressing themselves. By using our project they'll be able to give voice to their hand gestures so that other people who may or may not have knowledge about hand gesture language can understand them and communication becomes possible. By using OpenCV to capture the hand gestures and CNN the project will be able to read the hand gestures of the person and produce audio messages as per the recognized hand gestures. We evaluate our approach using a dataset of hand gestures involving either one or both hands simultaneously and compare the proposed approach to another that uses hand-crafted features.

Keywords: OpenCV, HGR (Hand Gesture Recognition), VE (Virtual Environments), DL (Deep Learning), ER (Entity Relationship), DB (Database), ML (Machine Learning), CNN (Convolutional Neural Networks), ASL (American Sign Language), MLP (Multi-layer Perceptron), HOG (Histogram of Gradients)

I. INTRODUCTION

Gestures are the physical movements of fingers, hands, arms, or body that carry special meaning that can be translated to interaction with the environment. There are many devices that can sense body positions, hand gestures, voice recognizers, facial expressions recognizer, and many other aspects of human actions that can be used as powerful HCI. Gesture recognition has many applications such as communication tools between hearing impaired and virtual reality applications and medical applications.

Hand gesture recognition techniques can be divided into two main categories: appearance-based approaches and three-dimensional hand model-based approaches. Appearance based approaches depend on features extracted from the model image to model the hand appearance. After that, all input frames from video streaming are compared with the extracted features to detect the correct gesture. Three-dimensional hand model-based approaches convert the 3D image to 2D image by projection. Then, the hand features are estimated by comparing the estimated 2D images with the input images to detect the current 2D hand gesture. Generally, appearance-based approaches performance is better than 3D hand model's performance in real-time detection but three-dimensional hand model-based approaches offer a rich description that potentially allows a wide class of hand gestures.

Hand gesture detection project based on Object detection and CNN. It also uses open source computer vision which is a highly optimized library focusing on real-time applications. The purpose behind using OpenCV is its availability. It means it is free for commercial use.

This project will be useful for mute people in communicating or expressing themselves. By using this, they will be able to give voice to their hand gestures so that other people who may or may not have knowledge about hand gesture language can understand them and communication becomes possible. This project will read the hand gestures of the person and interpret the message by data available in its database and then produce audio message as the output. This project proposes a system for hand gesture detection by using the Background Subtraction algorithm for background elimination and Convolution Neural Network for hand gesture recognition. Single hand gesture is detected and translated into commands that can be used to produce audio signals.

II. LITERATURE SURVEY

The HGR deals with the detection and recognition of hand gestures. Images of the hand gestures are taken using a camera input device or webcam and matched with the images in the database and the best match is returned. Gesture recognition is one of the essential techniques to build user-friendly interfaces. For example, a robot that can recognize hand gestures can take commands from humans, and for those who are unable to speak or hear, having a robot that can recognize sign language would allow them to communicate with it. Hand gesture recognition could help in video gaming by allowing players to interact with the game using gestures instead of using a controller. However, such an algorithm needs to be more robust to account for the myriad of possible hand positions in three-dimensional space. It also needs to work with video rather than static images. That is beyond the scope of our project. Computer vision and Digital Image Processing The sense of sight is arguably the most important of man's five senses. It provides a huge amount of information about the world that is rich in detail and delivered at the speed of light. However, human vision is not without its limitations, both physical and psychological. Through digital imaging technology and computers, man has transcended many visual limitations. He can see into far galaxies, the microscopic world, the subatomic world, and even "observe" infra-red, x-ray, ultraviolet and other spectra for medical diagnosis, meteorology, surveillance, and military uses, all with great success. While computers have been central to this success, for the most part man is the sole interpreter of all the digital data. For a long time, the central question has been whether computers can be designed to analyze and acquire information from images autonomously in the same natural way humans can. This is the province of computer vision, which is that branch of artificial intelligence that ultimately aims to use computers to emulate human vision, including learning and being able to make inferences and taking actions based on visual inputs. The main difficulty for computer vision as a relatively young discipline is the

current lack of a final scientific paradigm or model for human intelligence and human vision itself on which to build infrastructure for computer or machine learning. The use of images has an obvious drawback. Humans perceive the world in 3D, but current visual sensors like cameras capture the world in 2D images. The result is the natural loss of a good deal of information in the captured images. Without a proper paradigm to explain the mystery of human vision and perception, the recovery of lost information from 2D images represents a difficult hurdle for machine vision. However, despite this limitation, computer vision has progressed, riding mainly on the remarkable advancement of decades-old digital image processing techniques, using the science and methods contributed by other disciplines such as optics, neurobiology, psychology, physics, mathematics, electronics, computer science, artificial intelligence and others. Computer vision techniques and digital image processing methods both draw the proverbial water Real-Time Hand Gesture Detection and Recognition Using Simple Heuristic Rules from the same pool, which is the digital image, and therefore necessarily overlap. Image processing takes a digital image and subjects it to processes, such as noise reduction, detail enhancement, or filtering, for the purpose of producing another desired image as the end result. For example, the blurred image of a car registration plate might be enhanced by imaging techniques to produce a clear photo of the same so the police might identify the owner of the car. On the other hand, computer vision takes a digital image and subjects it to the same digital imaging techniques but for the purpose of analyzing and understanding what the image depicts. For example, the image of a building can be fed to a computer and thereafter be identified by the computer as a residential house, a stadium, high-rise office tower, shopping mall, or a farm barn. The first is the feature extraction approach, which focuses on simple computations applied directly to digital images to measure some usable characteristic, such as size. This relies on generally known image processing algorithms for noise reduction, filtering, object detection, edge detection, texture analysis, computation of optical flow, and segmentation, which techniques are commonly used to pre-process images for subsequent image analysis.

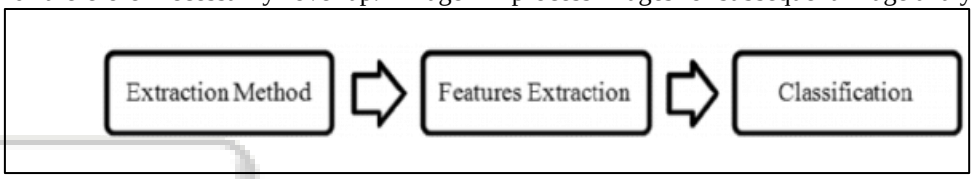

Fig. 1: Gesture Recognition Steps

The idea is to use the features as a unique representation of the image. Since a digital image is a two-dimensional matrix of pixels values, region-based object descriptions are affected by geometric transformations, such as scaling, translation, and rotation. For example, the numerical features describing the shape of a 2D object would change if the shape of the same object changes as seen from a different angle or perspective. However, to be useful in computer vision applications, object descriptions must be able to identify the same object irrespective of its position, orientation, or distortion. One of the most popular quantitative object descriptors are moments. This focuses on research, covering relevant literature relating to this project on touchless hand gesture based human computer interaction. Appropriate journals, books, Internet sites will be used to gather the relevant literature. This chapter describes the literature associated with gesture recognition and image processing. It discusses the different steps that comprise image processing and the techniques used to accomplish the stages. The literature also involves previous studies that aimed at the development of similar systems. The techniques used previously are analyzed and appropriate ones chosen to devise a feasible solution. Since they are invariant to geometric transformations, a set of moment invariants computed for an image may be considered as a feature vector. A set of feature vectors might constitute a class for object detection and recognition. The feature vectors of a class of reference images can be compared with the feature vectors of the image of an unknown object, and if their feature vectors do not match, then they may be considered as different objects.

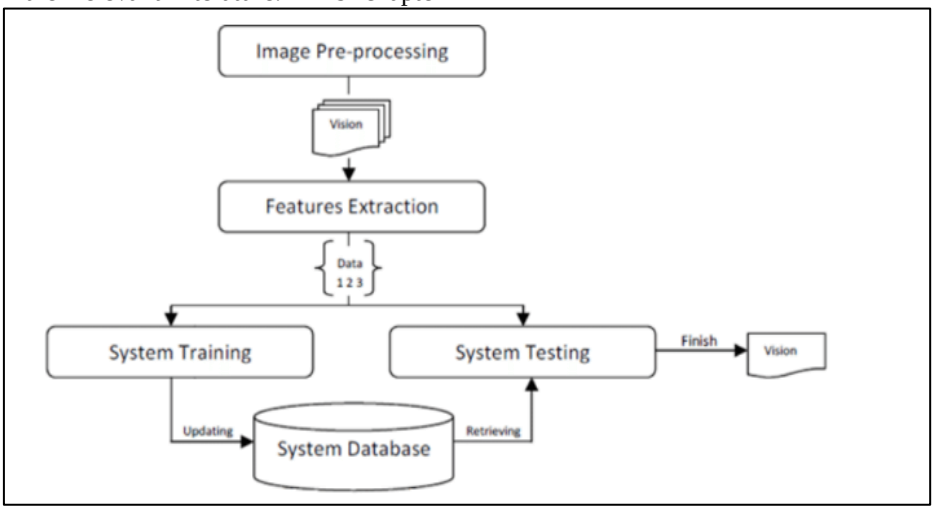

Fig. 2: Architecture of HGR Project

A. Hand Gesture Recognition Using Neural Network

Research on hand gestures can be classified into three categories. The first category, glove-based analysis, employs sensors attached to a glove that transduces finger flexions into electrical signals for determining the hand posture. The relative position of the hand is determined by an additional sensor. This sensor is normally a magnetic or an acoustic sensor attached to the glove. The second category, vision-based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way. Several different approaches have been tested so far. One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras, and parameters corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification.

B. Related Work

Hand gesture recognition is important for designing touchless interfaces in cars. Such interfaces allow drivers to focus on driving while interacting with other controls, e.g., audio and air conditioning, and thus improve drivers' safety and comfort. In the last decade, many vision-based dynamic gestures recognition algorithms were introduced. To recognize gestures, different features such as hand-crafted descriptors and articulated models, were used. As gesture classifiers, hidden Markov models, conditional random support vector machines (SVM) have been widely used. However, robust classification of gestures under widely varying lighting conditions, and from different subjects is still a challenging problem. To improve classification accuracy, gesture recognition methods with multi-modal sensors were introduced. Never ova et al. successfully combined RGBD data from the hand region with upper-body skeletal motion data using convolutional neural networks (CNNs) for recognizing 20 Italian sign language gestures. However, their technique was intended for gestures performed indoors only. They reported the best performance with a combination of histogram of gradient (HOG) features and an SVM classifier. Fused information of hand gestures from depth, color and radar sensors and jointly trained a convolutional neural network with it. They demonstrated successful classification results for widely varying lighting conditions, which motivated our work.

C. Application Areas of Hand Gestures System

Hand gestures recognition systems have been applied for the different applications on different domains, as mentioned in including; sign language translation, virtual environments, smart surveillance, robot control, medical systems etc. overview of some hand gesture application areas are listed below.

1) Sign Language:

Since sign language is used for interpreting and explaining a certain subject during the conversation. A lot of systems have been proposed to recognize gestures using different types of sign languages. For example recognized American Sign Language ASL using boundary histogram, MLP neural network and dynamic programming matching. recognized Japanese sign language JSL using Recurrent Neural Network. Shapes for drawing are; triangle, rectangular, circle, arc, horizontal and vertical line for drawing, and commands for editing graphic systems are; copy, delete, move, swap, undo, and close.

2) Virtual Environments (VEs):

One of the popular applications in gesture recognition systems is virtual environments VEs ,especially for communication media systems provided 3D pointing gesture recognition for natural human computer Interaction HCI in a real-time from binocular views. The proposed system is accurate and independent of user characteristics and environmental changes. A set of hand gestures are used to control the TV activities, such as turning the TV on and off, increasing and decreasing the volume, muting the sound, and changing the channel using open and close hands. 3D Modeling to build 3D modeling, a determination of hand shapes are needed to create, build and view the shape of the hand.

3) Graphic Editor Control:

Graphic editor control system requires the hand gesture to be tracked and located as a preprocessing operation used 12 dynamic gestures for drawing and editing graphic system. Shapes for drawing are; triangle, rectangular, circle, arc, horizontal and vertical line for drawing, and commands for editing graphic systems are; copy, delete, move, swap, undo, and close.

To build 3D modeling, a determination of hand shapes is needed to create, built and view 3D shape of the hand. Some systems built 2D and 3D objects using hand silhouettes. 3D hand modeling can be used for this purpose also which is still a promising field of research.

III. RESEARCH METHODOLOGY

We use a convolutional neural network classifier for dynamic hand gesture recognition. It is the algorithm for hand gesture recognition used in this paper.

Initially we took the convolutional neural network classifier algorithm as a basis, tested it, and made some changes to improve its results. First, we present the original algorithm from the paper. A live video stream is segmented into individual images from which a region of interest is extracted and converted to a 3D intensity matrices (height * width * intensity value of each pixel). These matrices are then fed as input to the Background Subtraction module. Color properties are not used, the images are converted to grayscale. The algorithm to separate background and foreground in the incoming image is based on the Approximate Median Model'. However, this technique requires two reference background images, namely, 'Current Background' and 'Buffered Background'. This technique of storing two backgrounds can be considered as a dual background method. One of the interesting features of this technique is that both the backgrounds are updated dynamically. The first one is updated frequently while the second one has a slower update rate. The first frame of the incoming video is initialized as 'Current Background'. Subsequently, the intensity of each pixel of this current background is compared with the corresponding pixel of the next frame. If it is less, then the intensity of that pixel of the current background is incremented by one unit, otherwise it

is decremented by one unit. In case of equality, the pixel intensities remain unchanged. This way, even if the foreground is changing at a fast pace, it will not affect the background but if the foreground is stationary, it gradually merges into the background. Since we are interested in all those objects which are stationary for a long period of time, we maintain another set of background image called 'Buffered Background'. Here, all those pixels which do not belong to the prospective abandoned objects set are made equal to that of 'Current Background'. This is done at an interval of every 20 seconds. Difference of the two backgrounds is represented as a binary image with the white portion representing foreground. The algorithm was implemented in Python with the help of OpenCV.

IV. PROCESS DIAGRAMS

A. Process Flow Diagram:

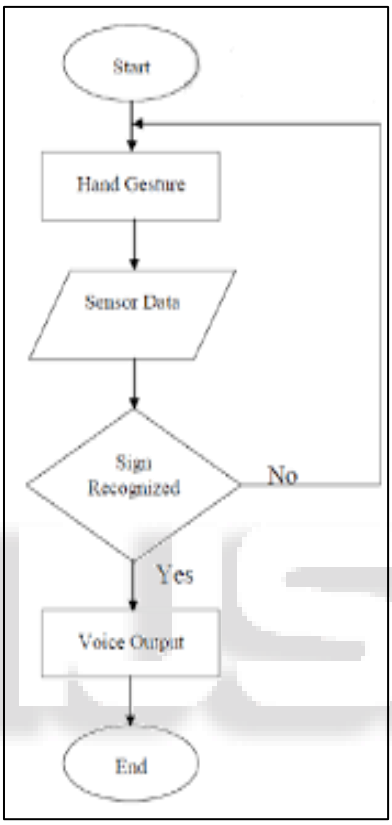

Fig. 3: Process Flow Diagram

B. Use Case Diagram:

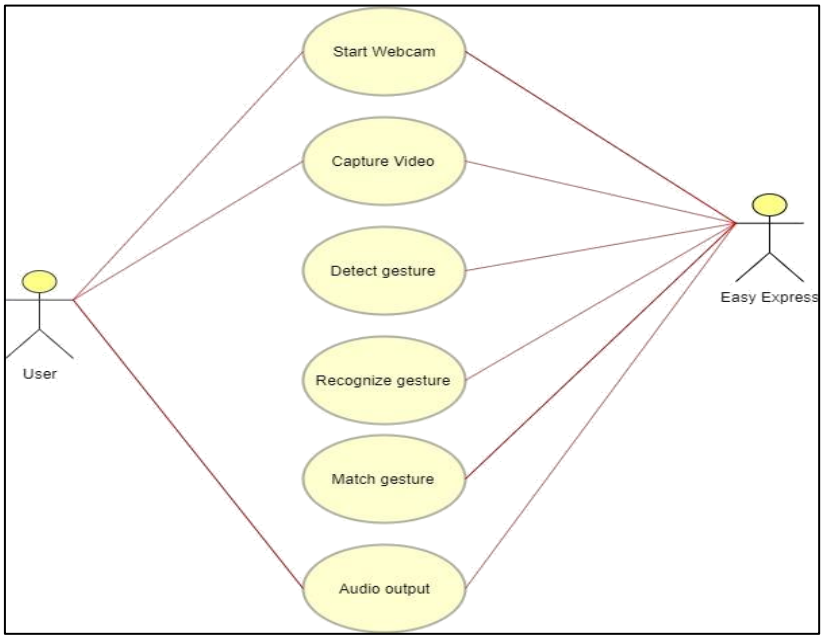

Fig. 4: Used Case Diagram

C. Activity Diagram:

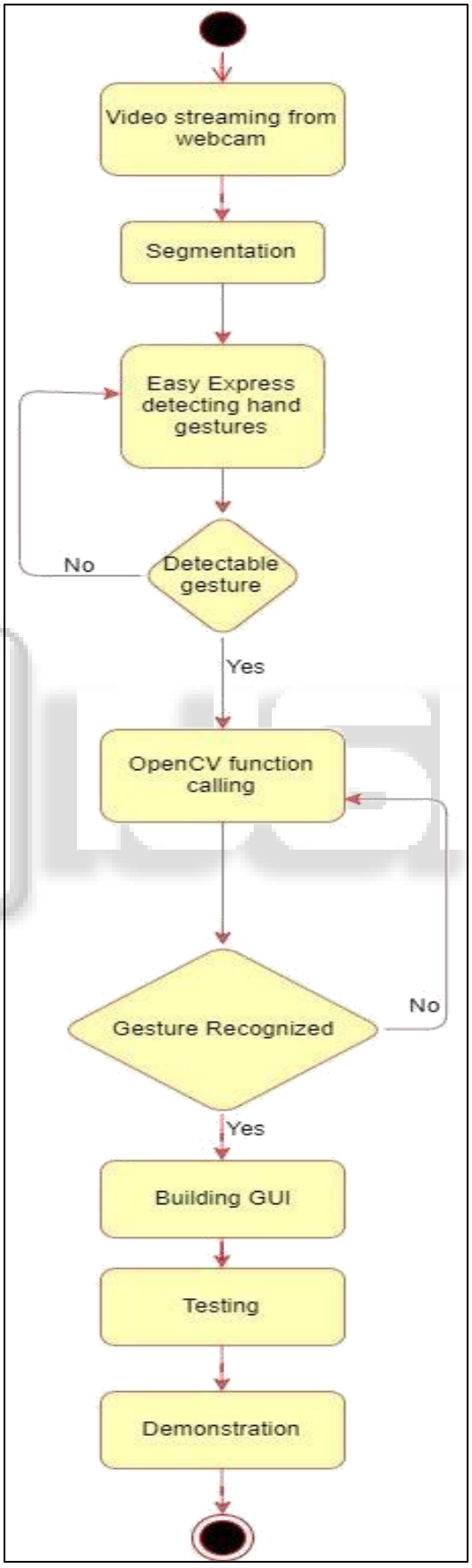

Fig. 5: Activity Diagram

D. Sequence Diagram:

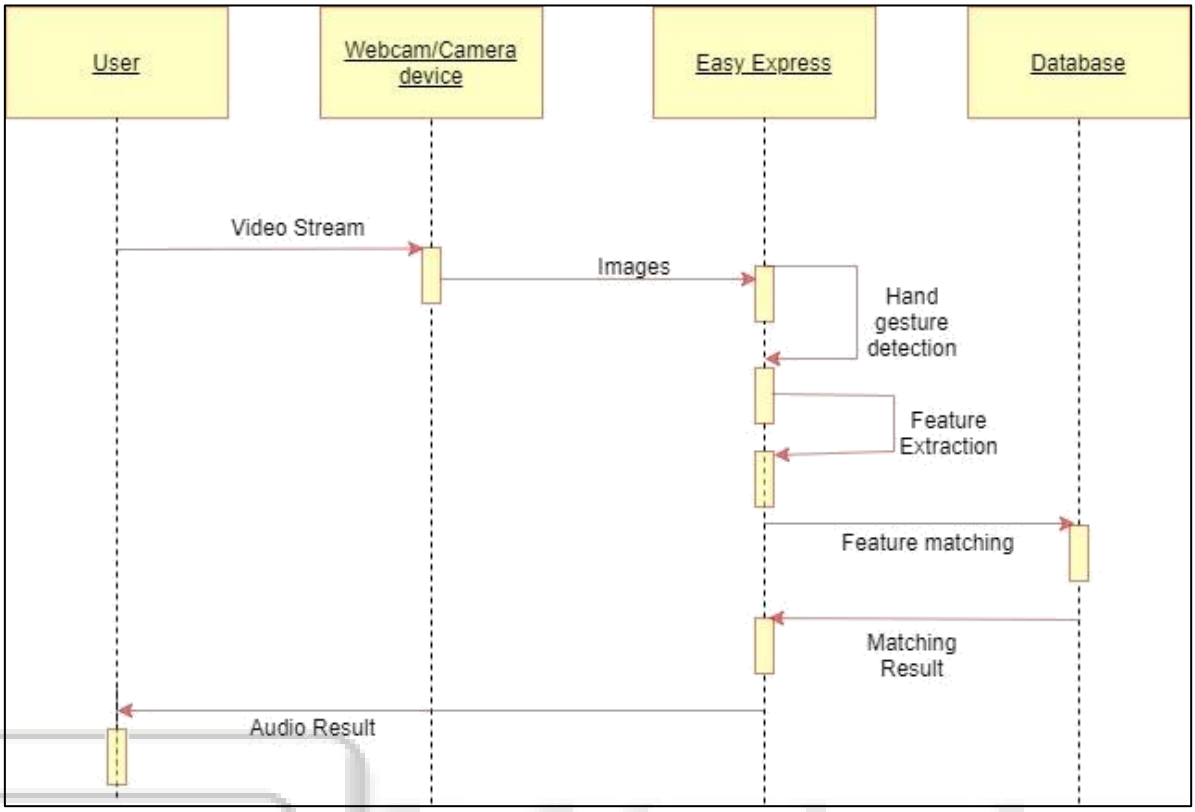

Fig. 6: Sequence Diagram

V. RESULTS

HGR resulted as a beneficial tool for solving the problem of communication especially in case of mute people. It is a project which can be useful for mute people in communicating or expressing themselves. It will help people in giving voice to their hand gestures so that other people can understand them easily. It will read the hand gestures of the person and then it will give the audio message for the recognized hand gestures. This project has a positive scope in future when it will be implemented on people in real life. Using advanced hardware and tools it can be extended to a very advanced extent.

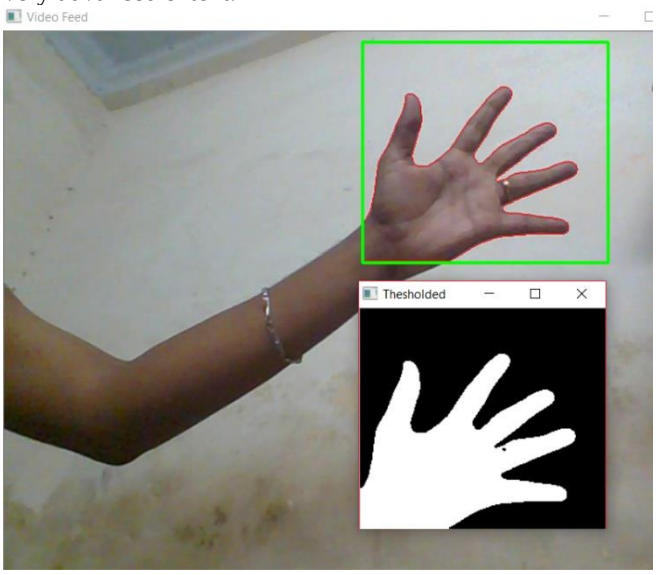

Fig. 7: Palm Detection using Thresholding

Having two backgrounds has an added advantage that the user can adjust the time interval between the update of reference background frames to suit different needs and environments. The Buffered Background is updated intelligently, only when it is necessary. Moving objects and unnecessary details like shadows, reflections on floors and walls are filtered off efficiently.

However, the algorithm requires an initial training period where the video feed must not contain any moving people. Otherwise it classifies the people present initially as missing objects.

We tried to make it adapt to rapid lighting fluctuations with a small extent of success, but with a small extent of success.

VI. CONCLUSION

The OpenCV library, which is an open source library which contains functions that specializes in image processing and gesture recognition, used in this project is compatible with programming language. The main problem that arises at the implementation phase was familiarizing with the programming language as well as the OpenCV. Having to learn a whole new programming language leads to a major amount of time loss. Also, understanding how the OpenCV functions work and how we can use them was a hectic task. After being able to successfully implement some simple functions we finally started with the coding of the software while continuing my research to get more in depth knowledge about the technology.

HGR project is very beneficial for society as it solves the problem of communication especially in case of mute people. It is a project which can be useful for mute people in communicating or expressing themselves. It will help people in giving voice to their hand gestures so that other people can understand them easily. HGR will read the hand gestures of the person and then it will give the audio message for the recognized hand gestures. This project has a positive scope in future when it will be implemented on people in real life. Using advanced hardware and tools it can be extended to a very advanced extent. Using our project mute people will be able to give voice to their hand gestures so that other people who may or may not have knowledge about hand gesture language can understand them and communication becomes possible.

A hand gesture gives us a sequence of hand postures connected by continuous hand or finger movements over a short period of time. Hand gestures provide a separate complementary modality to speech for expressing one's ideas. So, hand gestures recognizing systems can be a natural way of communicating between the computer and humans.

OpenCV is a highly optimized library focusing on real-time applications. We used OpenCV because it is open. It means it is free for commercial use. It is a library of programming functions mainly aimed at real-time computer vision.

REFERENCES

[1] International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012 [Online] [Accessed: July 04, 2012] Available: https://www.researchgate.net/publication/284626785_Hand_Gesture_Recognition_A_Literature_Review
[2] S. D. Badgujar research paper on Hand Gesture Recognition System [Online] Available: www.ijsrp.org/research-paper-0214/ijsrp-p2667.pdf [Accessed: Feb. 02, 2014]
[3] Official OpenCV website. Available: www.opencv.org
[4] Processing sketchbook tool website. Available: www.processing.org
[5] Ghotkar International Journal on sign-language gesture recognition [Online] Available: www.academia.edu/40445563/Vision_Based_Hand_Gesture_Recognition_Using_Fourier_Descriptor_for_Indian_Sign_Language [Accessed: Dec. 06, 2016]

