GESTURE RECOGNITION USING TUNED CONVOLUTION NEURAL NETWORK

A Thesis submitted in partial fulfillment of the requirement for the Award of the Degree of

MASTER OF ENGINEERING

In

Electronics and Communication

Submitted by

Maninder Singh

Roll No:801761010

Under the Supervision

of

Dr. Sunil Kumar

Assistant Professor, ECED

Mr. Sukhwinder Kumar

Lecturer, ECED

ELECTRONICS AND COMMUNICATION ENGINEERING DEPARTMENT

THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY), PATIALA, PUNJAB

AUGUST, 2019

DECLARATION

I, **Maninder Singh** hereby declare that the work presented in this Project **"GESTURE RECOGNITION USING TUNED CONVOLUTION NEURAL NETWORK "** in partial fulfillment of the requirement for the award of degree of **Master of Engineering (Electronics and Communication)** submitted at **Electronics and Communication department,** Thapar Institute of Engineering and Technology (Deemed to be University), Patiala is an authentic record of work carried out under supervision of **Dr. Sunil Kumar (Assistant Professor, ECED, IIITM, Gwalior) and Mr. Sukhwinder Kumar (Lecturer, ECED, T.I.E.T, Patiala)** from September 2018 to August 2019.

The matter presented in this seminar has not been submitted either in part or full to any other university or institute for the award of any other degree.

Date: 30/08/2019

Maninder Singh
Roll no:801761010

It is certified that the above statement made by the candidate is correct to the best of my knowledge and belief.

30/08/2019

Dr. Sunil Kumar
Assistant Professor
Electronics and Communication Department
IIITM, Gwalior

Date: 30/08/2019

Mr. Sukhwinder Kumar
Lecturer
Electronics and Communication Department
Thapar Institute of Engineering and Technology
(Deemed To Be University), Patiala, Punjab

Date: 30/08/2019

ACKNOWLEDGEMENT

First of all, I express my sincere thanks to the all almighty GOD for honoring me with knowledge, intelligence, well-being, cognizance and the brainpower to conduct this research successfully. I would like to thank my **Parents, other Family members and friends** who have been always supportive directly or indirectly to make my educational career superior.

I wish to express my deep gratitude and sincere thanks to my Supervisor, **Dr. Sunil Kumar,** Assistant Professor in Electronics and Communication Department, ABV-Indian Institute of Information Technology and Management, Gwalior, Madhya Pradesh and **Mr. Sukhwinder Kumar,** Lecturer in Electronics and Communication Department, Thapar Institute of Engineering and Technology, Patiala, Punjab for his invaluable guidance, constant encouragement, constructive comments, sympathetic attitude, and immense motivation, which has sustained my efforts at all stages of this work. Their valuable advice and suggestions for the corrections, modifications and improvement did enhance my work.

I would like to express my gratitude to **Dr. Alpana Agarwal,** the Head of Electronics and Communication Department, Thapar Institute of Engineering and Technology, Patiala, Punjab for providing me with adequate environment in carrying out the work.

I am also thankful to my friend, **Ipsita Kumar,** a student of Electronics and Communication Department, Thapar Institute of Engineering and Technology, Patiala, Punjab for her support and help in the completion of my thesis.

Finally, I want to extend my gratitude to all those people who directly or indirectly helped me in carrying out this work in right direction.

Maninder Singh

ABSTRACT

Skin color segmentation is used to discriminate between skin and non-skin pixels of an image. But when we are talking about robust techniques for detection of skin pixels, there are always some difficulties as skin segmentation is still an ongoing hard problem to be sorted out by the researchers. In order to segment human skin regions from non-skin regions, a reliable skin model is needed who is adaptable to different colors and light conditions. In this paper, implementation and extraction of skin pixels in RGB color model is being presented and depicted that there is a requirement of switching color models by observing the effect of noise, light etc. The color spaces that are frequently used in studies are RGB, HSV, and YCbCr. The presence of light, shadows, noise can affect the appearance of the skin color. However, an effective skin segmentation algorithm should be capable to detect skin pixels efficiently by overriding these effects. In this research study, RGB based skin segmentation technique is being presented for extraction of skin pixels. Therefore, for robust skin pixel detection, a dynamic skin color model that can cope with the changes must be employed.

We present the automated system for switching of color models automatically in different color space such RGB into YCbCr or vice versa to get the better visible image pixels. The experiment result shows that, the algorithm gives hopeful results. Followed by skin segmentation of hand, gesture recognition has been taken under consideration. The training model used for hand gesture recognition is CNN (convolutional neural network) due to its advantages of adaptability and self-training. Finally, the accuracy of hang gesture recognition has been improved and made more precise so that it can effectively be used in various applications and for interface.

Keywords Skin Segmentation, Hand Gesture Recognition, Convolution Neural Network (CNN)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

HGR	Hand Gesture Recognition
RGB	Red Green Blue
HSI	Hue Saturation Intensity
HSV	Hue Saturation Value
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
CNN	Convolutional Neural Network
HCI	Human Computer Interaction
SVM	Support Vector Machine
HMM	Hidden Markov Model
CDP	Conditional Density Propagation
FSM	Finite State Machine
CAD	computer aided design
ANN	Artificial Neural Network
InSL	Indian Sign Language
LSDM	Local Skin Distribution Model
SPM	Skin Probability Maps
GSDM	Global Skin color Distribution Model
IDM	Image Distribution Model
SDM	Skin color Distribution Model
FSPM	Fusion-Based Skin Probability Maps
CFS	Correlation based Feature Selection
RFE	Recursive Feature Elimination
ASL	American Sign Language
SIFT	Scale Invariant Feature Transform
HOG	Histogram of Oriented Gradients
GPU	Graphics Processing unit
CPU	Central Processing unit
ABF	Adaptive Bilateral Filter

PCA	Principle Component Analysis
RAM	Random Access Memory
IDLE	Integrated Development Environment

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

In past few years there has been a growing interest in the area of skin segmentation and gesture recognition. Skin segmentation is a separation of skin pixel and non-skin pixel [27]. A basic skin segmentation is shown in figure 1.1. Skin Segmentation is an important Step for Human Computer Interaction (HCI). Skin Segmentation includes Face Recognition, Face Expression Recognition, Eyes Tracking, Gesture Recognition. Applications based on Skin segmentation is dealing with the problems such as detection and recognition of humans and their activities that require the segmentation of skin regions as first step. If we talk about the skin color, it varies widely in range which makes it difficult to set a boundary around all regions. But it is observed that in small regions there is a limited range of skin color. The utilization of color information is a challenging task, as in images the appearance of the skin color is dependent on different factors, such as Aging, Ethnicity, illumination condition, Camera characteristics etc. Skin color varies due to aging and different lighting conditions, also the accuracy of skin detection depends upon the background. In other words, the detection of skin is very much prone to get affected by the background color which is nearly similar to the skin color or whose value is close to the actual skin pixel value. First step in some of the applications of skin segmentation is finding faces in an image which is called face detection. Followed by face detection, comes face recognition, facial expression recognition, eye tracking and gesture recognition. Color images can be represented in several color models that includes red-green-blue (RGB), hue-saturation-intensity (HSV) and YCbCr.

Detection of skin is the approach to finding pixels and areas of skin color in an image or video. Typically, this phase is used as a pre-processing step to find areas where pictures may contain human faces and limbs. Several approaches to computer vision for skin detection have been created. Rarely a skin detector transforms a designated pixel into a relevant color space and then use a skin classifier to label the pixel the appearance of skin in an image focuses on the circumstances of illumination where the picture was caught. A major obstacle in skin authentication is therefore to portray the color in a manner that is invariant or at least insensitive to modifications in lighting [35].

Skin video segmentation (Color based) is an efficient approach which enables very reliable skin segmentation irrespective of the illumination variation which occurs during tracking. The very basic of face recognition and tracking of gestures involve locating followed by tracking of the skin colored pixels. The very useful aspect of this technique is the orientation and the non-varying size for which it is used in early stage localization of the high level systems. The challenge however is to incorporate variations and later adjust with the variation in the illumination conditions which may or may not occur within the image sequence.

Figure 1.1 Skin segmentation

1.1.1 COLOR REPERSENTATION OF AN IMAGE

Previously, extraordinary color spaces have been utilized in skin division as appeared in figure 1.2. At times, shading characterization is finished utilizing just pixel chrominance since it is normal that if pixel luminance is disposed of skin division may turn out to be progressively powerful to lighting varieties. A picture that incorporates shading and picture data for every pixel or outwardly worthy outcomes, it gave (shading channels) for every pixel, which are deciphered as the directions in shading space. RGB shading space is most usually utilized in Image handling, however another spaces, for example, YCbCr, HSV are regularly utilized in different settings in picture preparing. RGB picture has three qualities every pixel and they measure the force worth and chrominance estimation of light. The genuine data of a picture store in the advanced picture information that is the splendor data in RGB Image

Colors are indicated in the three primary colors: red (R), green (G), and (B) blue.

Colors are specified in terms of hue(H), saturation(S) and intensity (V) which are the three color characteristics. There is really no linear transformation in both HSV and RGB.

YCbCr: luminance (Y-channel) as well as chrominance (Cb and Cb) colors are indicated. Fig 1.2 is the representation of RGB and YCbCr color space transformation.

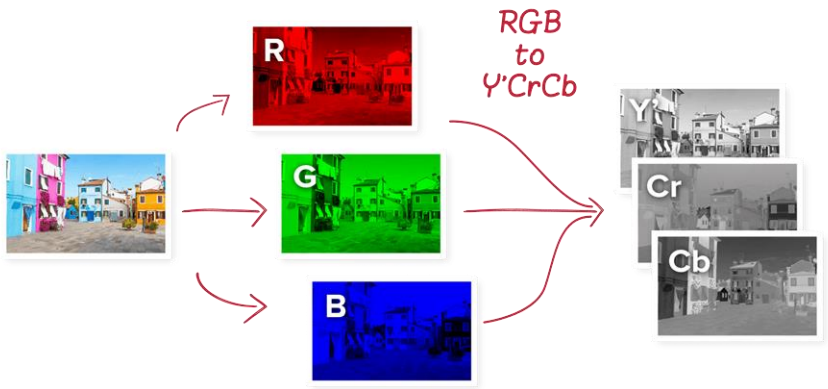
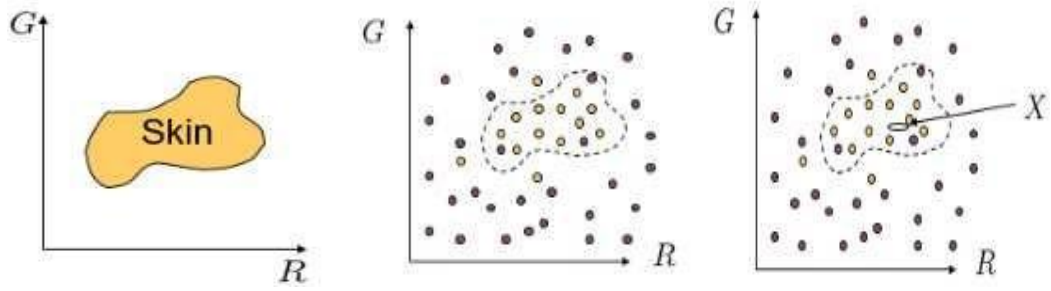


Figure 1.2 Red-green-blue (RGB) to YCbCr colour domain transformation

In different utilizations of programmed controlling utilizing picture preparing the initial step is to find the essence of the individual which is named as face restriction. Face restriction is finished by division through skin shading utilizing a versatile thresholding technique. The upside of versatile technique is that it adjusts naturally to the skin shade of the client. Finding a face is in a picture a troublesome issue in light of the fact that there are changes are available in the size of a picture, focuses, point and direction. Additionally, the demeanor of face, lighting condition changes and in part changes in the estimation of countenances called geometry trademark. Limitation of the face is a disentangled issue is face recognition. Finding a face in picture, when you watch and perform there is one face present in the picture. Along these lines, strategy is consequently attempting to find just confront. In a few applications the situation is normal or same, for example, following of an eye, face acknowledgment, and Gesture Recognition.

1.1.2 SKIN CLASIFICATION

Skin recognition is the procedure of skin-shaded pixels and districts in a picture or a video. This procedure is commonly utilized as a pre-handling venture to discover areas that possibly have human appearances and appendages in pictures. A few PC vision methodologies have been produced for skin recognition. A skin indicator normally changes a given pixel into a proper shading space and afterward utilize a skin classifier to name the pixel whether it is a skin or a non-skin pixel. Fig. 1.3 shows skin pixel limitation and characterization system. The presence of skin in a picture relies upon the enlightenment conditions, where the picture was caught. Hence, a significant test in skin identification is to speak to the shading in a manner that is invariant or possibly harsh toward changes in light. Figure 1.3 Skin pixel localization and classifications strategy.



(a) (b) (c)

Figure 1.3 Skin pixel localization and classifications.

1.1.3 APPLICATIONS OF SKIN SEGMENTATION

Some of the significant skin segmentation applications are explored in this chapter. It includes the various applications for the human computer interface and source to communicate with the high tech device.

- **Human Computer Interface:** This involves further more applications such as in gaming, in which gaming control can be done by the movements of hands. It can also be a source of communication for deaf and dumb people.

- **Robot Control:** Utilizing HCI you can undoubtedly control the robot's administration robots straightforwardly associate with individuals, so finding a progressively normal and simpler UI is of essential significance. While prior works have concentrated principally on issues, for example, control and route in the earth, couple of mechanical frameworks are utilized with easy to use interfaces that have the capacity to control the robot by regular methods. To encourage a plausible answer for this necessity, we have executed a framework through which the client can offer directions to a remote robot utilizing signals. Through this technique, the client can control or explore the robot by utilizing motions of his/her palm, along these lines communicating with the mechanical framework. The direction sign is produced from these motions utilizing picture preparing. These signs are then passed to the robot to explore it in the predefined bearing.

- **Home automation:** Hand signals are utilized to control the home machines, for example, fans, lights, and so forth. The stage utilized for the acknowledgment of the motion is the MATLAB recreation instrument. When the main signal is caught and prepared, one apparatus is controlled, at that point it is required to rerun the program. Research is going on procedure to control all the home apparatuses by the motions once the program is on run. The future headway will be founded on the IoT premise, we can control the home machines in and around the globe by the assistance of Internet of things. The apparatuses as well as these signals are utilized to control volumes tuning, TV channels, speed controls, the controller of a fan can be controlled by the motion.

- **Medical applications:** Hand motion is one of the most significant methods correspondences without contact among human and machines. There is an incredible enthusiasm for instructing electronic hardware in medical procedure rooms by hand motion for lessening the hour of medical procedure and the potential for contamination. There are difficulties in usage of a hand signal acknowledgment framework. It needs to satisfy prerequisites, for example, high exactness and quick reaction.

1.2 GESTURE RECOGNITION

One of the most normal and significant correspondence modalities is motions. Collaborations like human correspondence are performed once the human arranged intuitive frameworks are acknowledged, for which motion acknowledgment is significantly required. The most alluring picture succession for intuitive frameworks is the non-contact and constant motion acknowledgment. A lot of systems have been proposed for sign affirmation from picture game plan. Several these procedures used for movement affirmation from picture gathering are DP organizing, constrained state machine, HMM, neural frameworks. In the strategies referenced above time changing features are removed from sign picture gathering. With these techniques, the degree data of signal, for example, greatness, speed and others can't be acquired These strategies likewise put accentuation on the grouping of a sort of motion. The degree data assumes a significant job in correspondence which essentially speak to client's feelings, disposition, articulations, etc. Along these lines, both sort of motion and level of data are critical to be perceived by intelligent frameworks. What's more, these techniques acknowledge just one picture grouping. As the motion acknowledgment utilizes just one picture grouping, it is hard for it to perceive a confounded motion accurately uniquely single input picture sequence auto-occlusion and confusion such that gesture is usually performed in 3D space. sequence. Due to the large range of signs and a variety of different features defining each sign, gesture recognition is a complex problem for understanding sign languages. Many methods for gesture recognition of sign language looks at the subsets of the parameters that focuses on the issues of tracking hands and head position of the signer successfully. It also classifies the hand gesture based on neural network, motion analysis 3D analysis etc. Another problem, that is a bit more difficult than the previous one is to recognize the sign from a continuous image stream which is popularly known as Motion Epenthesis or Continuous Sign Language Recognition. Gestures from A to Z are shown in figure 1.4 as sign language.

Considering the historical backdrop of hand motion acknowledgment, it began with the innovation of glove based control interface for PC control. It was acknowledged by the scientists that the signals that are enlivened by the gesture based communication can be utilized to offer straightforward directions for any PC interface. The glove based system step by step improved and advanced with the improvement of different gadgets and instruments, for example, infrared cameras, fiberoptic curve sensors and much precise accelerometers. In the long run, a portion of the advancements in glove based framework, offered the capacity to acknowledge PC vision based calculation with no sensors joined to the glove [30]. There has been a little more advancement made to gloves, that is, colored glove consisting of unique colors for finger tracking. After the concept of gesture recognition using gloves technique, came to the path, gesture recognition with the help of image processing using skin segmentation followed by various methods in between [10],[24].

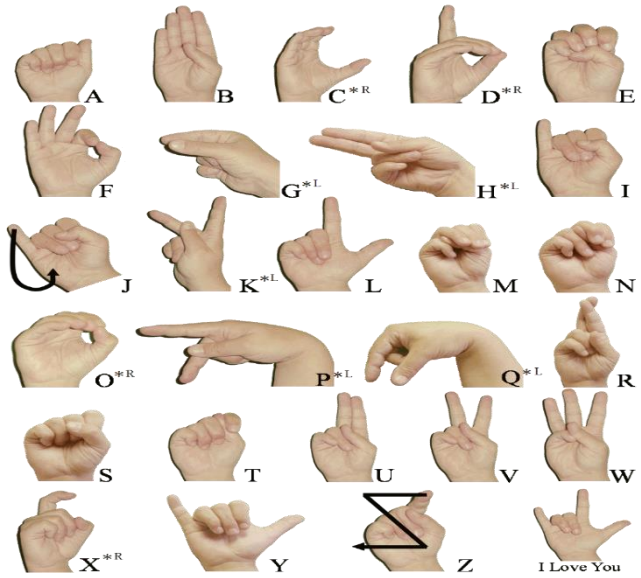


Figure 1.4 Hand Gesture for Alphabets

1.2.1 HCI

Computer has been a very important part of human life now-a- days. Machine (CPU) are used by people both at their home or offices. For making the communication easy between humans and computers, various but special i/o devices are designed with the time. one of the most common devices are keypad and optical mouse. By the introduction of every new device, an attempt is made for the humans to be able to perform more difficult and complicated communication with the computer thus, making the computer more intelligent. It has been possible due to the creation of successful human computer interface, by the computer professionals and their result-oriented efforts. (CP) and intuitiveness are the crucial attributes of (CP) to survive in a competitive environment, as the complexities of human requirements and needs have turned into huge demands and they continue to grow. The (CP) have been incredibly successful at making the communication much easier between the humans and computers. With the arrival of every new product in the market, it makes it easier to perform the job, thus reducing the complexity. For instance, it has taken better human control over complex work systems such as cars, planes, monitoring systems etc. it has also helped in facilitating tele operating, robotic use and much more. Initially, the focus was on speed rather than the modifiable features and so the (CP) are pretend that kind of CP. Provide The focus area thus has been revisited and shifted towards a more human friendly environment. The main idea is to develop a user-friendly human computer interface (HCI) and make the computers understand human language. A few steps towards making it human compatible is making it understand speech and voice, facial expressions and human gestures. In which, gesture comes under the non-verbal exchange of information. Human being can perform innumerable gestures at a particular place and time. It is a subject for computer vision researchers as human gestures are obtained and perceived through vision. The aim is to determine these human gestures by creating human

computer interface (HCI). A complex programming algorithm is required to perform the coding of these gestures into machine language [5].

There is no existing specific identification of hand gesture due to its vast range of applications. This statement can only explain the little amount of gestures. Even though the researchers have tried to define gestures but its actual mean is still arbitrary. Research scholar defined gestures as the motion of the body and body parts that is intended to communicate for a specific gesture, both the sending and the receiving end must have the same set of information, thus leading it to a successful communication. Gesture can be explain as an expressive body part which shows a specific massage that is, precisely communicated b/w the sender and the receiver at the receive end. A gesture can be further categorized into two major categories i.e. dynamic and static. In the case of dynamic gesture, intended to change over the period of time whereas static gesture is observed at a spurt of time. It is important to interpret all the static and dynamic gestures over a period of time to understand the complete message. This complex process is called gesture recognition. Gesture recognition is the process of recognizing and interpreting a stream continuous sequential gesture from the given set of input data. HCI (human-PC collaboration) is the investigation of how individuals interface with PCs and to what degree PCs are or are not created for fruitful cooperation with people. A noteworthy number of real organizations and scholastic establishments presently study HCI. Truly and with certain special cases, PC framework engineers have not given much consideration to PC usability. Numerous PC clients today would contend that PC creators are as yet not giving enough consideration to making their items "easy to understand." However, PC framework engineers may contend that PCs are amazingly mind boggling items to plan and make and that the interest for the administrations that PCs can give has consistently outdriven the interest for usability. Figure 1.5 show basics structure of HCI.

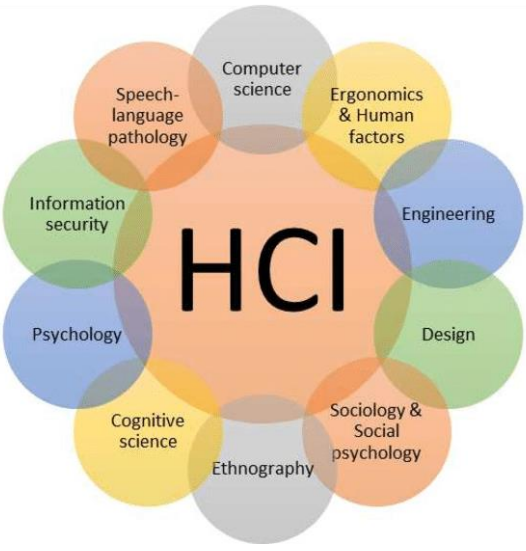


Figure 1.5 Human Computer Interface (HCI)

1.3 HAND GESTURE RECOGNITION TECHNIQUES

- Template Matching
- Feature Extraction Analysis
- Active Shapes Model
- Principal Component Analysis
- Linear Fingertip Models
- Casual Analysis

1.4 TOOLS FOR GESTURE RECOGNITION

There are different types of tools available for data mining classification technique which are used by the researchers these tools are basically used for segmentation and feature extraction of data.

1.4.1 HIDDEN MARKOV MODELS (HMM)

Mainly in HMM model double processes are performed as shown in figure 1.6. Initial first one is to create Markov chain contain limited number of states and the second is arbitrary capacity arrangement of related with First one state. Corresponding to present state observation symbols are generated. According to random functions that symbol is generated for first one process. Probabilities pair defined as follows for the all transition between the all these states.

- Probability identified with the change which make accessible likelihood for experiencing the progress.
- Probability identify with yield which depicts confined likelihood of delivering a yield image from limited letter set when given a state.

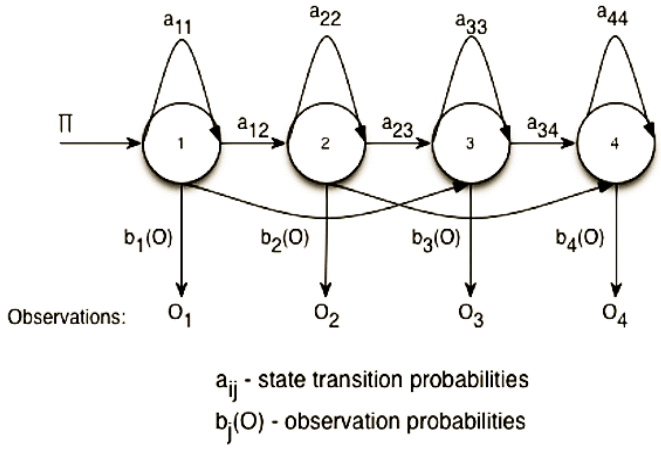


Figure 1.6 Hidden Markov Model

Spatio and temporal information can effectively and naturally handle by HMM just because it has numerous mathematical structures. Here one is only able to see that sequences of observations so its termed as “hidden”. For evaluation, learning and decoding process the well- designed and well organized algorithm developed by Baum–Welch and Viterbi is also involved here. Expressed in HMM π.= (A B) an observation which contain strings set O = { O1,.., OT }, where t varies from 1 to T. A set with N, states can be given as {s1, . . ., sN}; for k discrete observation, a set of symbols are given as {v1, . . ., vk} N is the set of HMM states, K is the number of code book nodes, and the percentage of correctly labelled examples is verification and test score.

1.4.2 PARTICLE FILTERING AND ALGORITHM CONDENSATION

Condensation algorithm (Conditional Density Propagation) This is a calculation of PC vision. The fundamental head use of build up calculation is to distinguish and furthermore track the form of article moving in a jumbled foundation or condition. Following of an article is one of the essential and furthermore troublesome parts of PC vision and is commonly an earlier condition to question acknowledgment. Having the option to recognizing which pixels in a picture to make up the shape of item is a non-paltry issue. To take care of this issue Condensation is a probabilistic calculation is required. The creation of the algorithm was inspired by Kalman's failing to properly track elements in the presence of interesting background clutter. The existence of clutter tends to generate multi-modal and thus poorly modelled probability distributions for the object state. The most general type of the condensation algorithm does not require any assumptions.

1.4.3 FSM APPROACH

An ordered sequence of states is modelled as gestures for spatio-temporal configuration space, FSM method. The quantity of states in FSM are changed according to the applications. Trajectory is recognized as gesture from unsegmented sample. The trajectories group is created with the incessant sensor information stream. The trajectories of gestures are signified by the point cluster in 3D space. The chronological Hand signature action is extracted from the partial Vibrant set of gestures. Energy perception is used for this motion which is then understood and examined By a FSM of determinism. The comparative divert is measured for temporal signature determination such that the motion is slow or quick. For the accomplishment of additional cross-culture gestures the FSM is redefined. The redefinition is done according to the applicable rules of society.

1.4.4 YUV COLOR SPACE AND CAMSHIFT ALGORITHM

For the recognition of hand gestures by this method, following five steps are taken under consideration. (1) At first, A video stream of hand gestures is recorded using a digital camera. The frames will then be discussed.

(2) By using YUV color space, skin color-based segmentation is achieved. In this the chrominance and intensity are isolated from each other which is the main factor of using YUV color system. Y indicates intensity whereas UV signifies the chrominance part.

(3) For the separation of hands from the rest of the image, CAMSHIFT algorithm shown in figure 1.7 is used. Hand being the largest connected region, the segmentation of the hand from the body is of a great need.

(4) Followed by this, the centroid of the location of the hand is evaluated in each frame. For instance, zeroth and first moments are calculated which are then used for the calculation of other centroids.

(5) The different centroids are joined to form a shape telling a direction.

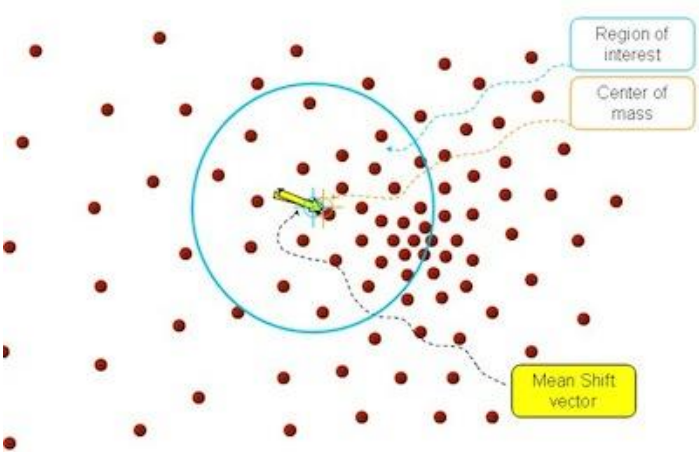


Figure 1.7 Camshift Algorithm

1.4.5 ARTIFICIAL NEURAL NETWORKS (ANN)

ANN is used by many researchers for gesture recognition. Layers of ANN are shown in figure 1.8. ANN has been classified as one of the most usable technique by most of the researchers while on the other hand it has also been used for extracting shape of hands by other researchers.

Adobe photoshop filter is used for finding the edges of the input images. To obtain the feature of an image, the histogram of local orientation is calculated. These vectors also contribute to manage neural network systems. For the recognition of Indian sign language (InSL), two repetitive models are used both partially and fully separately. For the data of input image, a colored glove model is used and for segmentation, HSI model is used. Later, for the recognition of hand area, YCbCr color space is used. For the recognition of skin color, threshold value is also used in the case of YCbCr color space.

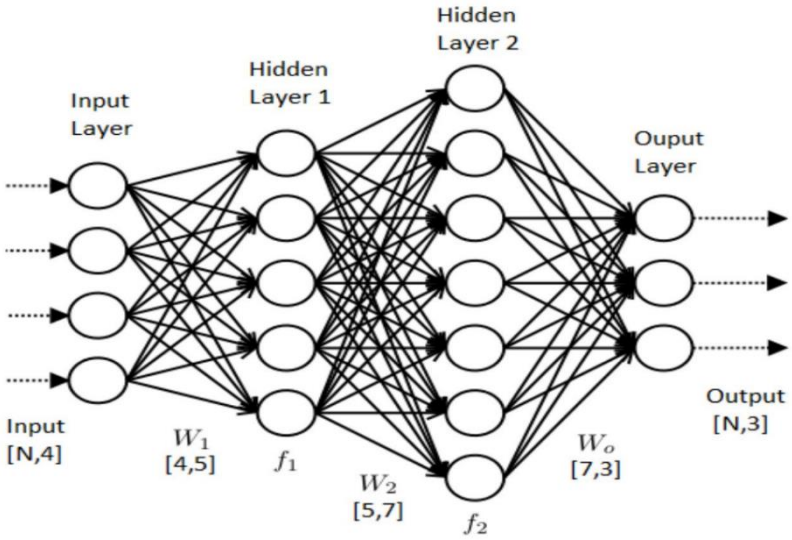


Figure 1.8 Artificial Neural Network

1.4.6 CONVULUTION NEURAL NETWORK (CNN)

Convolutional neural systems have a significant job for versatile picture preparing. CNNs additionally structure a connection between versatile channels and general feed-forward neural systems. With the assistance of at least one layers of two dimensional channels and conceivable actuation capacities or/and down examining, the two dimensional CNNs are framed. Interpretation invariance and spatially neighborhood associations are the key properties that are controlled by CNNs. The information based versatile strategies, for example, CNNs are valuable for picture preparing. They can likewise be utilized in different applications where the information exhibits are huge and incidentally/spatially appropriated. Further refinements of CNNs are proposed, for example, augmentations to three dimensional for example video preparing or execution of distinguishable channels. Structure of CNN is shown in figure 1.9

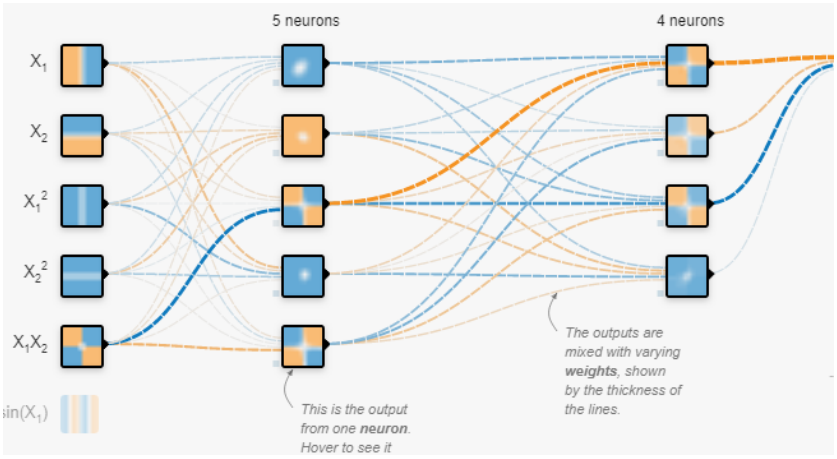


Figure 1.9 Convolutional neural network

Table 1.1 Comparison Between SVM, FSM and CNN

Accuracy %	Average	Max	Min	Std
Particle Filters	**88.31**	**90.80**	**72.73**	**7.25**
SVM	**75.53**	**88.96**	**14.56**	**19.79**
FSM	**79.22**	**90.57**	**61.55**	**7.41**
CNN	**80.56**	**91.56**	**73.86**	**7.98**

1.5 GESTURE BASED APPLICATION

Applications based on gesture are broadly classified into two categories depending upon their purpose Multidirectional client control and language of symbolism.

a) 3D Design

Computer aided design (PC supported plan) goes under one of the 3D structures as it is a HCI that gives a stage to control and understanding of 3-Dimensional information sources which can be resolved as signals. Control of 3D data sources utilizing mouse is a very tedious assignment as it includes entangled procedure of deteriorating a six degree opportunity task into three successive two degree undertakings. Another innovation has been taken into way for example 3D RAW innovation. In this innovation a pen is inserted in a gadget to follow the situation of the pen and the direction is in 3D. The plane where the article rests is a 3D space sensor implanted in level palette. The items can be pivoted and interpreted by the synchronous CAD model development and the client signal development so as to see them from every one of the sides as they are being adjusted and made.

b) Tele presence

In some of the cases such as emergency hostile conditions or inaccessible remote areas or system failure there may raise a need of manual operations. For human operators, it is regularly difficult to be physically present close machines. Tele nearness is the zone of specialized insight whose fundamental rationale is to give a situation of physical activity bolster that through the administrator that will convey the entire procedure and convey to the significant errand. speak to the common habitat a characteristic marvel of controlling outside vehicle by use hand motion language. Tele nearness likewise incorporate

under ocean mission prescription assembling, space and support of atomic reactors as its different prospects.

c) Virtual reality

In real as well as in imaginary world, physical presence can be simulated by applying virtual reality to computer-simulated environments. Most presented virtual reality environments are the primary visual experience displayed either by stereoscopic displays or on a PC monitor. There are also simulations that include additional sensory information, such that the sound through speakers or headphones. Followed by this there are advanced systems that include tactile information which is known as force feedback, in gaming and medical applications.

d) Sign language

It is the most fundamental, characteristic and crude type of language that could be absolutely gone back to as right on time as the coming of human progress, when the underlying hypotheses of gesture based communication showed up in the history. The time of gesture based communication give an impart before the talking. From that very time till date gesture based communication at significantly include and has been versatile in our day by day schedule life as the reason for correspondence process. In the present life gesture based communications are by and large broadly utilized worldwide as the global language for hard of hearing and unable to speak individuals, at work spots or family zones as well as in the realm of games, religious practices and considerably more. Signals one of the most introductory types of data learnt by a tyke when he/she needs to express its solace, warmth, need of nourishment and different angles. It helps in communicating the sentiments and considerations viably and furthermore upgrades the accentuation of communicated in language. A straightforward motion saying 'hello there' or 'farewell' utilizing one hand have same significance everywhere throughout the world making it a universal motion. Numerous individuals making a trip the world over to different abroad nations without the information of the official language of that spot are as yet ready to impart utilizing communication via gestures and signals. These are a couple of models that demonstrates motions to be universal and utilized practically everywhere throughout the world. At numerous associations and different employments all around the globe are signals are the fundamental methods for data. A couple of more instances of motions and communication via gestures are to such an extent that at air terminals and in planes there is a predefined set of signs and signals that causes individuals on the ground to have the option to speak with the pilots in this way offering ways to the pilot and furthermore how to take off and arrive on the runway. Another model is thinking about an arbitrator in games, in practically any and each game the ref uses motions to impart his choices subsequently making signals basic in the realm of games. A progression of motion and sign is gotten by the pitcher in baseball. Individuals with hearing issues have built up a sign and motion language where all the characterized signalled have been appointed an importance. With the assistance of this language they can speak with all sort of words and the live in ordinary world. The acknowledgment of the motions in the dialects, for example, communication via gestures is without a doubt spoken to as the most hard acknowledgment issue. In

the event that we consider a working gesture based communication acknowledgment framework, it gives an extraordinary open door for the hard of hearing to impart effectively without the need of translator. In this way, taking a shot at such models is still under thought [7].

e) PC and Gaming

Gesture Recognition has been acquainted with the human-PC association and gaming industry which currently empowers for example game architects to show signs of improvement comprehension of the game understanding so it is to some degree conceivable to control the experience and make includes that push the limits of reality considerably more. Figure 1.10 shows motion controller that can be utilized in this application.

f) Human Factors and Simulation

Eye tracking has taken an important place in the area of automotive research. There is a going on research in automobiles where the vehicles might start responding towards the driver's eye movements and eye gaze. Furthermore, eye tracking can make biometric sensor technology more successful and powerful.

Figure 1.10 Gesture Controller

1.6 OUTLINE OF THESIS

The thesis is divided into five sections. Each section plays an important role in the completion of thesis. A brief about each section is provided in this section.

Chapter 1: In this chapter a brief introduction is given about basics of skin segmentation and Gesture Recognition. Various methods and applications of skin segmentation and Gesture Recognition. Followed by this, a brief about extraction of features of HGR is given. the main focus is on one of the features that is Hand Gestures. In the end a small information about HGR technique and tools for HGR is provided with the general applications of Hand Gestures.

Chapter 2: The chapter is provided with a literature survey of various paper on hand gesture recognition. Different methods and algorithms such as template matching, and models like SVM, ANN, CNN etc. proposed by researchers is explained with the results in the form of accuracy of the algorithms has been explained. With the help of this chapter further methodology has been proposed.

Chapter 3: In this chapter the criteria of proposed methodology has been explained with the help of a proper flow diagram. CNN and its layers are briefly explained. The light has been put on each and every step included in the proposed method. The method has been proposed on the basis of HGR leading to the comparison of the experimental values to actual values.

Chapter 4: The chapter explains about the results of accuracy of HGR obtained by the experimental values and its comparison with the actual values. The results are obtained in the form of TP (true positive), TN (true negative), FP (false positive) and FN (false negative) by finding the error between both the values. CNN model accuracy and Confusion Metrix

Chapter 5: This chapter consists of the conclusion and future scope sections of the thesis on the basis of hand gesture recognition.

CHAPTER 2
LITRETURE REVIEW

2.1 INTRODUCTION

Literature survey started from simple threshold-based skin segmentation. Follow through the color-based skin segmentation methods and probabilistic skin segmentation. We started from the basic problems what are the colors model that used in skin segmentation. First step in some of the applications of skin segmentation is finding faces in an image. In this literature review we found that the best color models for skin segmentation and gesture recognition

Manuel *et al.*, the creators in 2013 displayed an exhibition examination among a few shading models including RGB, HSV and YCbCr based on shading skin division [13]. The procedure of division starts with the change of the picture into the ideal shading model. The detecting procedure gives the picture in the RGB shading model. Thusly, this shading model does not expect of this progression. HSV, YCbCr model requires this past advance. The least complex change is the YCbCr in light of the fact that just suggests a network increase. The least difficult calculation for shading division is finished by choosing an edge for each channel to isolates the skin shading. The edge is chosen by acquiring a few skin tests, and after that getting the histogram of each channel of those examples. The division is finished by contrasting every pixel and edge. The portrayed procedure was mimicked in MATLAB to display the proposed execution. So as to think about the result of the division under each shading model, 31 pictures containing just one face where physically fragmented. At that point, the subsequent veil of each picture was contrasted and the ground truth [26]. The best execution is accomplished by the YCbCr and HSV shading models. Analysts utilized YCbCr tint model. These outcomes are arrived at utilizing fixed edge. YCbCr takes less time because of the way that transformation among RGB and YCbCr is less complex.

Phung *et al.*, the author in 2005 introduced: shading portrayal, shading quantization, and characterization calculation [33]. Investigation of a few delegate shading spaces utilizing the Bayesian classifier with the histogram procedure demonstrated that skin division dependent on shading pixel grouping is to a great extent unaffected by the decision of the shading space. Division execution debases when just chrominance diverts are utilized in order. Moreover, it was discovered that shading quantization can be as low as 64 containers for every channel, albeit higher histogram sizes give better division execution. Most existing skin division procedures include the order of individual picture pixels into skin and non-skin classes based on pixel shading. The method of reasoning behind this methodology is that the human skin has reliable hues which are unmistakable from the shades of numerous different articles. Order Algorithm Several calculations have been proposed for skin shading pixel characterization. They incorporate piecewise direct classifiers, The Bayesian classifier with the histogram method, Gaussian classifiers [27],[34].

- **Linear Classifier:** A linear decision boundary is used to separate this category of classifiers, skin and non-skin colors.
- **Bayesian Classifier:** The Bayesian minimum cost decision rule is a well-established method in the classification of statistical patterns. Use this decision rule to consider a color pixel x as a skin pixel if

$$\frac{p(x|skin)}{p(x|non-skin)} \geq \tau$$

Gaussian Classifiers

The class-conditional pdf of skin colors is approximated by a functional parametric shape which is generally selected as a Gaussian unimodal or a Gaussian combination. The skin class-conditional pdf has the shape in the situation of the unimodal Gaussian model.

$$p(x|skin) = g\,(x;m_s,C_s)$$

$$= (2\pi)^{-d/2}|C_s|^{-1/2}exp{-}\frac{1}{2}(x-m_s)'\,C_s{}^{-1}\,(x-m_s$$

Where x is a d-dimensional vector, Ms is the middle vector and Cs is the skin class covariance matrix. If we expect a uniform distribution of the non-skin class, the Bayesian rule in Eqn. Reduces to the following: a pixel x color is regarded as a pixel of the skin if:

$$(x-m_s)^T\,C_s{}^{-1}\,(x-m_s) \leq \tau'$$

where τ is a threshold and the left-hand side of Eqn. (2.4) is the squared Mahala Nobis distance. The resulting decision boundary is an ellipse in 2D space and an ellipsoid in 3D space. In this study, we also investigate the approach of modeling both skin and non-skin distributions as unimodal Gaussian. In this case, it can easily be shown that *x* is a skin
pixel if:

$$(x-m_s)'\,C_s{}^{-1}\,(x-m_s) - (x-m_{ns})'\,C_{ns}{}^{-1}\,(x-m_{ns}) \leq \tau'$$

In case, skin pixel assumes to be multi-modal Gaussian distribution, decision rule is decided on the basis of probability given by the following equation.

$$p\left(x|skin\right) = \sum_{i=1}^{Ns} \omega_{s,i} g\left(x; m_{s,i,} C_{s,i}\right.$$

$$p\left(x|nonskin\right) = \sum_{i=1}^{Ns} \omega_{ns,i} g\left(x; m_{s,i,} C_{s,i}\right.$$

Chakraborty et al, authors in 2016 researched division of skin locales from shading pictures. Be that as it may, exactness of existing skin location strategies are seriously influenced by the shading comparability between the foundation and genuine skin areas. Probabilistic methodologies utilizing skin likelihood maps (SPMs) can take care of this issue to a degree. In this paper, a novel strategy has been proposed which uses seeded area developing technique. District developing is executed by a versatile cost proliferation and neighborhood examination plot. The underlying seeds are acquired from the SPM [25]. It saw from the exploratory outcomes that the proposed strategy can perform better contrasted with the current skin division strategies for various enlightenment and foundation conditions. The creators, proposed an approach to identify a novel skin data of pixel dissemination in a picture for a specific shading space. In this technique, a Local Skin Distribution Model (LSDM) has been gotten from the picture pixels dissemination model utilizing pixels from the facial district. At long last, the joint skin conveyance model is gotten by combining LSDM with Global skin shading circulation model. After this, a powerful district developing (was technique proposed to permit skin zones to develop progressively. Another skin recognition calculation was proposed by using the data of dispersion of picture pixels. The rule that a picture can be grouped into various Gaussian circulated bunches was considered. In this way, the pixel dispersion can be displayed as a blend of Gaussian capacities. The picture pixel appropriation model is named as Image Distribution Model. In this way, the Gaussian dispersed bunches of skin pixels ought to be nearer to the group of reference skin pixels in a given shading space.

Smita Tripathi et al, the creators in 2011 showed a technique for extraction of face. The strategy was spoken to with the mix of skin shading extractor and layout coordinating plan [16]. Initial step was to concentrate faces with the assistance of skin shading extractor. The shading space model utilized for the extraction was YCbCr shading model. The motivation behind utilizing this shading model was that it effectively secludes both non-skin and skin pixels. On account of skin shading extractor, it can undoubtedly separate among appearances and non-faces. In mix to this layout coordinating strategy was utilized so as to get faces all the more precisely and to evade non faces. By the down to earth improvement of the examination it has been demonstrated that the proposed criteria is superior to anything the skin shading extractor. For the execution of the test Windows XP was utilized as the working framework, a 2.2 GHz processor and the RAM was of 1GB. The usage was finished by MATLAB form R2007b. the dataset worked for the achievement to this plan was RGPV which contained 30 pictures. The structure of the pictures was to such an extent that it was having different

various pictures of recognized shapes, sizes, shading, position and articulations. The underlying picture taken for the execution of the technique was in RGB structure. The RGB picture was changed over into a grayscale variant. The grayscale picture was pursued and changed over to a YCbCr picture. At that point, to YCbCr picture, a limit is connected. On applying the limit, the edge picture is changed to the paired picture. The qualities in this picture are disseminated in two sections for example skin area changes into white shading with worth 1 and the non-skin district changes into dark shading with worth 0. After this, skin extractor was connected to the picture that detached skin pixels and non-skin pixels, not simply this it additionally expels non skin pixels from the picture. Indeed, even after this, a couple non-skin pixels are probably going to be removed to stay away from this error sobel channel was taken into the technique for appropriate face extraction. The most precise technique among the two was layout coordinating plan. Skin shading finder likewise recognized a couple non faces alongside appearances. For upgrading and ad libbing the idea of face extraction with this strategy was by diminishing the outcomes for the sort false positive (FP). To improve the idea of skin extractor, it was cultivated alongside format coordinating. In future research it was proposed to actualize utilizing different shading space and take a gander at their comparison. The reference skin pixel group was acquired from the skin model got from facial pixels. Thusly, another Local Skin Distribution Model (LSDM) was inferred by estimating the comparability among IDM and the reference skin pixel circulation model. Toward the end, a combination based skin model was determined by joining GSDM with the proposed SDM. In this way, a unique district developing strategy was connected so as to improve the division precision [10].

Biplab Ketan Chakraborty, et al [13], the authors proposed the manner in which the capacity of computer to perceive hand signal outwardly is basic for advancement in human–computer cooperation. Motion acknowledgment has applications going from communication via gestures to therapeutic help to computer generated reality. Nonetheless, signal acknowledgment [21] is very testing not just due to its assorted settings, various elucidations, and spatio-transient varieties yet in addition as a result of the complex non-inflexible properties of the hand. This reviews significant imperatives on vision-based motion acknowledgment happening in location and pre-handling, portrayal and highlight extraction, and acknowledgment. Current difficulties are investigated in detail. Signal portrayal Static motions can be spoken to utilizing a 2D model and 3D model. The least difficult approach to speak to a static signal or posture is to utilize an appearance-based model, which attempts to recognize motions straightforwardly from visual pictures. Single picture signal is anything but difficult to distinguish where as in video or constant applications, motions of hand is hard to recognize. Since video has contained 30 edges in 1/sec [22].

Youssef Filali et al [6], Among the most hazardous malignancy on the planet is skin disease. If not analyzed in beginning times it may be difficult to fix. The point of this work is to show an investigation

of skin division, highlights choice and characterization approaches [1]. In the division arrange, we will display the consequence of the utilization of a pre-preparing dependent on a multiscale decay model where geometrical segment is utilized to get a decent division. The highlights are right off the bat removed utilizing the surface segment and shade of the injury, and afterward we will introduce a relative investigation of certain highlights choice methodologies that select the applicable ones. In highlight arrangement we will think about between the most and great classifiers utilized in writing. Otsu for division, surface and shading for highlights extraction, ReliefF, Correlation based Feature Selection (CFS), Recursive Feature Elimination (RFE), chi2 technique as a component's determination and SVM with quadratic part as a classifier. The proposed methodology has been executed and tried on therapeutic pictures. As future work, concerning the division and order of skin injury, utilized plan to utilize new calculations and strategies to grow more our frameworks of a higher quality finding on dermatology pictures like profound learning calculations [8],[11].

Alijanpour N et al, [10] As computer become progressively unavoidable in the public arena, encouraging regular human PC communication (HCI) will positively affect their utilization. Subsequently, there has been developing enthusiasm for the advancement of new methodologies and advances for spanning the human–PC obstruction. A definitive point is to bring HCI to a system where collaborations with PCs will be as normal as an association among people, and to this end, joining motions in HCI shown in figure 2.2 is a significant research territory. Motions [32] have for some time been considered as a collaboration strategy that can possibly convey progressively characteristic, inventive and natural strategies for speaking with our PCs. This paper gives an examination of near overviews done around there. The utilization of hand signals as a characteristic interface fills in as a propelling power for research in motion scientific categorizations, its portrayals and acknowledgment systems

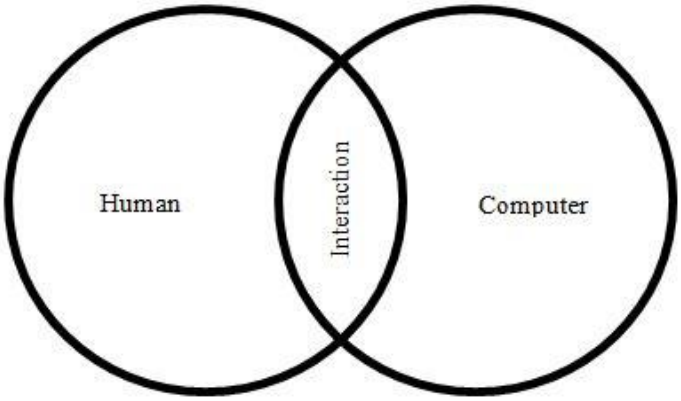


Figure 2.1 Natural human computer interaction (HCI)

Roziati Zainuddin et al., in 2012 proposed a color pixel bunching model for skin division under unconstrained scene conditions [14]. The proposed model can beat affectability to varieties in lighting conditions and complex foundations. Our methodology depends on structure multi-skin shading bunching models utilizing the Hue, Saturation, and Value shading space and staggered division. Skin locales are separated utilizing four skin shading grouping models, to be specific, the standard-skin, shadow-skin, light-skin, and high-redskin models. In addition, skin shading adjustment (skin lighting) at the shadow-skin layer is utilized to improve the discovery rate. The exploratory outcomes from an enormous picture informational index show that the proposed grouping models could accomplish a genuine positive pace of 96.5% and a bogus positive pace of roughly 0.765%. The trial results demonstrate that the shading pixel bunching model is more effective than different methodologies. The point of the creators is to exhibit an investigation of skin division, highlights determination and arrangement draws near. In the division organize, we will introduce the aftereffect of the utilization of a pre-preparing dependent on a multi-scale disintegration model where geometrical segment is utilized to get a decent division. The highlights are right off the bat removed utilizing the surface segment and shade of the sore, and afterward we will exhibit a near investigation of certain highlights determination approaches that select the applicable ones. In highlight characterization we will look at between the most and great classifiers utilized in writing. Otsu for division, surface and shading for highlights extraction, Relief, Correlation based Feature Selection (CFS), Recursive Feature Elimination (RFE), chi2 strategy as a component choice and SVM with quadratic part as a classifier. The proposed methodology has been actualized and tried on medicinal pictures. As future work, concerning the division and arrangement of skin injury, utilized mean to utilize new calculations and strategies to grow more our frameworks of a higher quality analysis on dermatology pictures like profound learning calculations [8].

Meghdad Kurmanji et al, the authors in 2019 i.e. recently found that as compared to hand gesture recognition in still images, the same is more challenging in videos. The challenges came into the way as it is difficult to represent the temporal features and they also have longer training time mostly in real-time applications [2][6]. Followed by studying a brief about this challenge, the authors in this article, the potential of 2D over 3D CNNs to represent temporal characteristics and also to classify hand gestures in videos was explored. Further to capture the dynamics of the hand movements in a single frame, the frame sequence of the hand gesture was mapped to a chronological tiled pattern. Followed by this, a feature vector was generated by using 2D CNNs. The feature vector consisted of both temporal and special features. In addition to this, a new approach has been proposed that was basically used for Predictions by means of a two-stream architecture and data fusion to exploit information on depth. The authors also undertook an inquiry into various kinds of methods of augmentation. Research findings mean that the suitable use of 2D CNNs can readily outperform a 3D CNN application in terms of time, recall and precision in this assignment.

Table 2.1 Configuration of 2D CNN

MAX EPOCHS	OPTIMIZER	EPOCH STEP	MOMENT.	BATCH ACC.	BATCH SIZE	LR.
200	NESTEROV	10	0.9	2	25	10^{-x}

Table 2.2 Configuration of 3D CNN

MAX EPOCHS	OPTIMIZER	EPOCH STEP	MOMENT.	BATCH ACC.	BATCH SIZE	LR.
200	SGD	10	0.9	2	25	10^{-x}

Table 2.3 Mean and Standard Deviation of Accuracy and Recall

CNN type	VIVA Dataset					
	Accuracy	std	Recall	std	T1 (min)	T2 (sec)
GoogleNet	72.5	6.1	66.5	6.5	480	5
Alexnet	70.5	7	63.5	7	155	3.5
	Cambridge Dataset					
GoogleNet	77	2.3	69	2.4	365	5
Alexnet	75	2.5	65	2.5	102	3.5

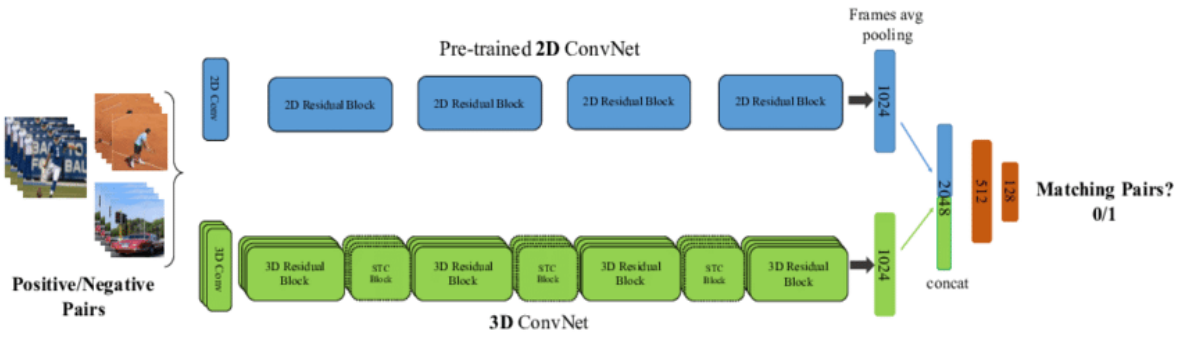


Figure 2.2 Block diagram for the potential of 2D and 3D CNN

Jing-Hao Sun et al, the authors in 2018, through this research explain that the demand of interface, interaction and communication between humans and machines is increasing more and more extensively due to the rapid ongoing developments in computer vision [3]. The applications of hand gesture recognition are in huge terms due to its capability of expressing enriched information. The applications are such that Recognition of hand movements is commonly used in intelligent furniture, robot control and different viewpoints. In this paper, the division of hand motion has been acknowledged by the foundation of skin shading model and furthermore by AdaBoost classifier. As per the disposition of skin shading for hand motion, the AdaBoost classifier depends on Haar. At last, the area of hand that has been distinguished continuously has been perceived by convolutional neural system. After the total acknowledgment utilizing neural system Also, the denaturation of hand motion comprising one casing of video is being cut for examination. With this entire strategy, the human hand can without much of a stretch be portioned from jumbled and complex foundation. The constant hand motion following can likewise be acknowledged by CamShift. It demonstrates the precision of 98.3%

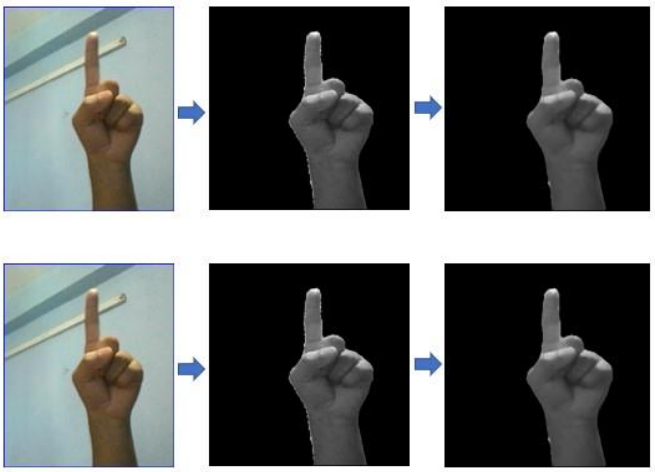


Figure 2.3 Skin color segmentation of hand.

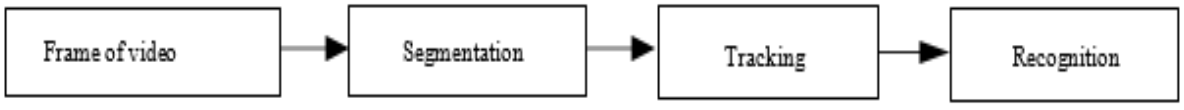


Figure 2.4 Block diagram for gesture recognition

Jaya Shukla et al, the authors in 2014 Represented a technique for hand recognition of signs with help of Microsoft Kinect sensor [24]. Kinect shows in figure 2.5, can be used for dense capturing and also for 3-dimensional real time images of an element. A combination of both learning and modelling Recognition of the hand gesture strategy has been proposed by the authors in this paper. A major feature of Kinect, i.e. Kinect depth feature was used for the proper segmentation of hand gesture images that were taken with the Kinect. Further, to find a contour of the hand pictures segmented, a few image handling techniques are employed. Then, the convexity defects and convex hull for the contour have been calculated. The convexity defects and contour area are used as a classification feature. In this proposal the gestures are classified using Bayes classifier. In this research, five hand gesture classes have been considered i.e. it shows one, two, three, four and five fingers one at a time. The algorithm was tested and implemented for 15 images of each class. The classification rate came out to be of 100%.

Figure 2.5 Kinect Sensor

Later, after applying the depth thresholding, binary images are being obtained. The tool used to classify was a machine learning tool termed as Weka. Followed by this, a mapping from attributed to the class variable was found by the Bayes classifiers that was based on a few data sets such as unknown real world attribute example that could predict high precision class. The feature attributes came out to be convexity defect and contour area. The authors also gave a brief on the research that can be done in the future. The future research on the basis of this work are as follows: (1) hand gesture recognition using two hands. (2) possibility of gesture recognition from different rotations and orientations. (3) recognition of not only static but also dynamic gestures.

Chung-Ju Liao et al, the authors in 2015 Introduced an experience based model of hand motion recognition. Unlike other studies about hand gesture recognition, possible dynamic motions of hand and complicated background is considered in this study. Thus, there are many problems considered in this proposal except for the simple background subtraction. The problem considered in this study are such that the Skin detection of color picture, hand detector reaching in the perspective of the camera and also full palm detection. The system proposed in this paper consisted of four stages i.e. (1) detection of hand

appearance in the camera view, (2) hand region segmentation, (3) full palm detection and (4) recognition of hand gesture. Detecting the appearance of the hand is to determine when the hand enters the camera view field. A few of morphological double-stage skin color detection approaches also applied to the proposed system to remove the sound impact. The suggested technique of two stage skin color detection is obtained from the idea of handling outliers of extracting the hand from a complex and cluttered back story followed by this, the Full palm detection is performed to understand if the hand gets beyond the camera view area. The design used to determine of hand going beyond the camera view is known as ergonomics. The system is run by keeping mini dv at the center of the user's color, followed by which individuals are authorized to bring the fingertips in the camera area regardless what they seem to be wearing whether full sleeves or half sleeves. Also, whether the hand is moving or in a still position was not the matter of consideration. In the end, the frame's corners have been cut in the four sections of the row. Including this frame splitting has been done in order to judge if the image included palm. lastly, the outcomes of testing demonstrated the efficiency of the Introduced algorithm for the recognitions such that the results were quite promising [23], [22].

Marco E. Benalcazar et al, there are a number of applications of gesture recognition specially in the fields of medical and engineering. The basic problem of hand gesture recognition is to identify a given gesture performed by hand at any moment. The authors in 2017 proposed a real time hand gesture model. The input of this method was to measure surface electromyography by a commercial sensor known as the myo armband that is placed on the forearm [16]. The label of the gesture executed by the user at any time comes out to be the output. The proposed model is based on the dynamic time warping algorithms and k-nearest neighbor. The model could easily be made to learn to recognize any hand gesture. For the evaluation of the model, there was a measurement and a comparison of the accuracy was made between the proprietary algorithm i.e. the myo armband model and the model presented by the authors. The comparison was made at 5 classes of gesture recognition which were such that, pinch, fist, open, wave in and wave out. The data in the model was transmitted via Bluetooth to the computer. The proposed method basically composed of 5 model: (1) signal acquisition, (2) preprocessing of the signal (3) extraction of features, (4) classification and (5) post-processing. As a result, it was determined that the model presented had 3% more accuracy i.e. 86% than that of myo method which had 83% accuracy for instance [17],[18].

Sarang Suresh Kakkoth et al, gestures language has been the basic mode of information since the ancient times until human developed vocal language to communicate. Even after the development of vocal language, gesture kept its path in huge terms and still is equally significant as vocal language. The hand gesture recognition not only have applications in entertainment world such as for human computer interaction or gaming but also its applications lies in human life improvement section such as sign language for deaf and dumb people, traffic signal control system and the list go on. In this paper in 2017, the authors compiled a research on Some of the steps associated in this implementation of hand

gesture recognition system that included data acquisition, segmentation, tracking, feature extraction and gesture recognition. According to the researchers, hand gesture technique can basically be divided into 3 techniques that involves sensor based technique, vision based technique and (3) depth based technique. Each of the technique has its own advantages and disadvantages depending upon the applications they are used in. the research work here involves the modification of the techniques in various steps according to the improvements in hardware and development in software framework. Finally, in this paper the amalgamation of various techniques and their sub steps is presented. It can come handy in the case of real time hand gesture recognition [14],[22].

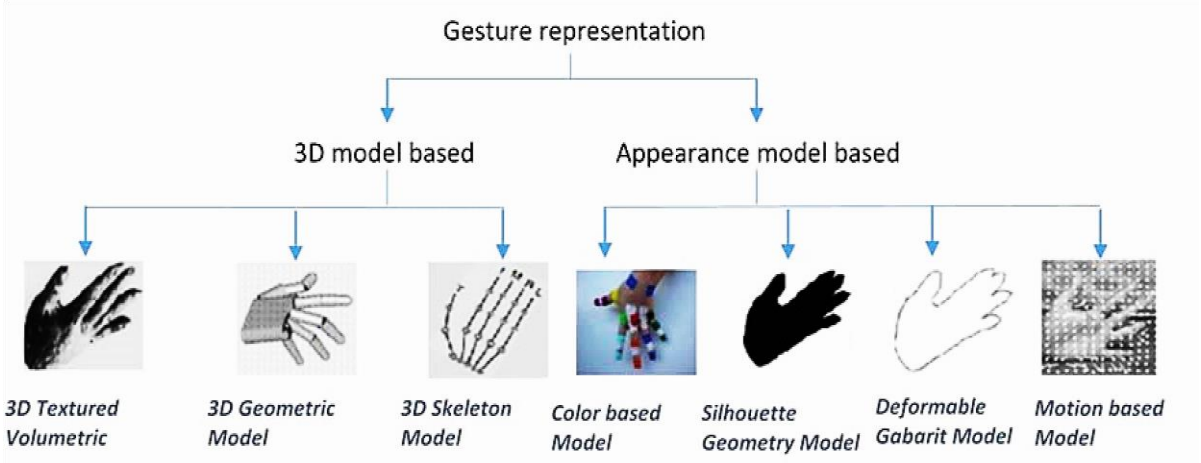


Figure 2.6 Representations of hand gestures based on vision

Juhi Ekbote et al, all over world sign language is a standard for communication between deaf and dumb people community. It is the most basic and significant way of interaction and communication between hearing and speech impaired people and normal people without the need of any interpretation. there is a standard and specific sign language for every country of its own. In India the dialect is termed as Indian sign language. In 2017, in this research work an automatic recognition system for Indian sign language numerals starting from 0 to 9. There data base used by the researchers was self-created that consisted of 1000 images divided into 100 image per numeral sign has been developed. There were three techniques used for extraction of desired features i.e. shape descriptors, Scale Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG). For the classification of sign two basic techniques were used i.e. support vector machine (SVM) and Artificial Neural Networks (ANN). As the experimental results, the combined approach of SIFT and HOG with SVM classifier provided an accuracy of 93% while the combination of HOG and ANN provided the accuracy as high as 99%. Figure 2.7 shows the accuracy obtained by applying different combination of feature extraction techniques and classifiers [19].

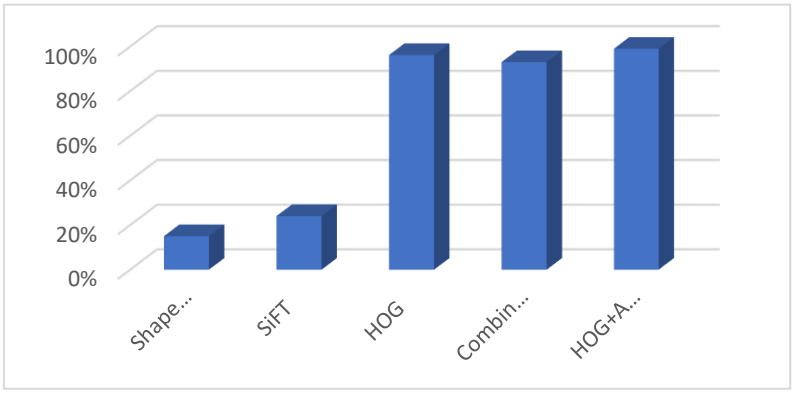


Figure 2.7 Accuracy obtained by combination of feature extraction techniques and classifiers

Saha Himadri Nath et al, the world consists of a total population of nearly 7 billion individuals, over 500 million of whom actually suffer mostly from one or the other type of disability be it sensory, mental or physical. Such type of deformities Does not allow them to participate fully in society and also makes a limitation for them from the entertainment of equality, equal possibilities. Language of signs is a common source of interaction or communication to the silly and stupid sign language a very efficient way of talking or interaction where the hand gesture replaces the former. Hand manipulations are used to convey the conveyor's ideas fluently with a combination of various shapes of hand, orientation and movement of hands, positioning of the palm and alignment of fingers. Signs are used to convey words and phrases to the crowd. In 2018, in this paper the authors tend to optimize an algorithm for hand gesture recognition with a reasonable accuracy [12]. In this case the pattern recognition input is provided by the hands. Previous models of reference such as American sign language (ASL) are already available. Data of the images is collected through webcam followed by the preprocessing part. Next step in the algorithm is to do the segmentation of the figure which is done by approximate convex decomposition and polygon approximation. Followed by this come the part of extraction feature that is accomplished by recording the distinctive function between the different convex hand sections. The extracted characteristics are regarded to be the resulting singularities. This includes training with the characteristics acquired. The characteristics acquired are roughly unique for various hand gestures. This algorithm successfully made disabled individuals socially acceptable as it became easy to identify the sign language.

Figure 2.8 Gestures Sign Language

2.2 RESEARCH GAPS

In the past years, A series of processes in the area of skin segmentation to Gesture recognition and other related process has been widely explored and worked upon. A few of these researches can be taken a look from Literature survey in the above section. As we have discussed before, skin segmentation is the first step of most of the applications based on this area. The first challenge in this area comes out as we can easily find the skin pixel from an image in normal environment but where the background is similar to skin, there we face some difficulties to separate actual skin pixels and non-skin pixel from an image. Further, in the case of cluttered background it is even harder to set a boundary between actual skin pixels and non-skin pixels. One of the research gaps that arise here is to get more accuracy in skin segmentation from cluttered background. Recognition of hand gesture in a video or real time applications leads to some delay in the system as it takes time for each frame to be processed. Gesture recognition in different lighting conditions is another challenge. Researchers defines that the detection and monitoring of skin colored areas in the videos is a significant study direction under different lighting circumstances. Observe CNN model and increase accuracy of the modified model

2.3 OBJECTIVE OF THESIS

On reviewing the previous papers and work done on hand gesture recognition, the first objective of this paper is to determine the accuracy of the skin segmentation for gesture recognition. Followed by this, gestures are to be recognized for sign language. Further, the features and advantages of using CNN model for analyzing and training are to be shown. Once, the models are train with the preprocessed data set it allows to recognize new gesture by itself. This model is adaptive in nature. One of the main features of CNN is of great advantage for gesture recognition is that with the help of new real time input data it trained itself The main motive of the research is to increase the accuracy of gesture recognition in real time images and thus making it more precise for the interface, interaction and various other applications.

CHAPTER 3
METHODOLOGY

3.1 INTRODUCTION

In this chapter, methodology and our proposed work has been discussed for Hand Gesture Recognition (HGR) Using image processing computer vision and classification techniques of an image it is possible to recognize hand gestures. That process of hand gesture recognition is mainly divided into three parts hand location, hand segmentation and its classification. My method is locating hand in an image and extracts the hand and generate new images for postprocessing and uses in Convolutional Neural Network (CNN) for classification and then recognize the gestures of hand [15].

We can easily adapt to different environments, recognize the patterns and generalize from our knowledge about determining what is it in the image, but computers do not have these skills. When a computer takes an image as an input it sees a matrix of pixel values, as shown in Figure 3.1. Depending on the number of image channels (RGB = 3 or Grayscale = 1) pixels are unique values of channel length arrays, that describe the intensity of the pixels at each point. If the image is Grayscale, pixels have an integer value in the range [0,255]. For RGB pixels having an array of length 3, one for each color with same value range as Grayscale [0,255]. The idea of image classification using machines is to use these values to detect patterns in the image and provide a resulting label or a probability of the input belonging to a class.

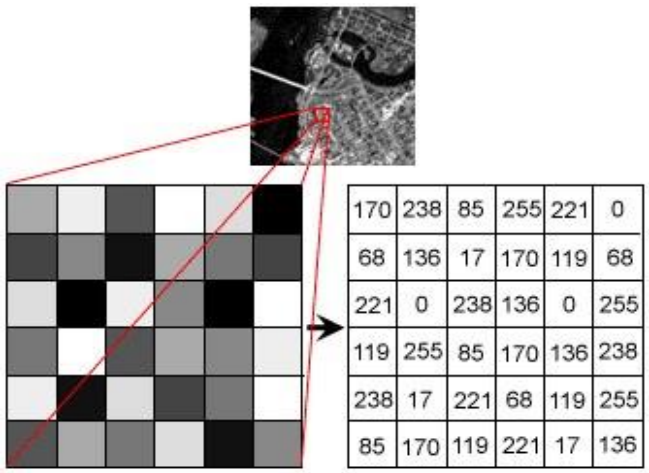


Figure 3.1 How Computer See an Image

3.2 TOOLS USED

3.2.1 Python 3

Python is a most commonly used for powerful programming language ideal for scripting and dynamic application development. It is also used in web development (like: Django), used for scientific and mathematical computing (Orange, NumPy) to the desktop graphical user Interface (Pygame, Panda3D). Python is easy to use and learn. It is high level programming and developer friendly language.

3.2.2 OpenCV

OpenCV is the library of programming functions mainly used in real time computer vision. OpenCV support a lot of algorithms which is related to Open Computer Vision and Machine Learning. It supports the Numpy Tensorflow and keras which is required for our proposed method. It provides the support to programmer to express the ideas in fewer lines of code and instructions without reducing any readability.

3.2.3 TensorFlow

While we are as yet dealing with the early use of AI, further developed calculations and their branches, presenting, for example, Deep Learning which is known as Neural Networks, TensorFlow is the essential programming device of profound learning. TensorFlow is an open source computerized reasoning library, to fabricate models utilizing information stream charts. It enables designers to make or make enormous scale neural systems with various layers. TensorFlow is chiefly utilized for: Perception, Classification, Discovering, Prediction and Creation. TensorFlow is for the most part utilized by Social Media, Handset Manufacturers and Telecom, Image Search, Face Recognition, Motion Detection, Machine Vision and Photo Clustering and furthermore in Automotive. Motion Recognition intends to perceive and distinguish signs and articles in pictures just as for understanding the primary substance and setting. TensorFlow item acknowledgment calculations used to group and recognize subjective articles, inside high goals pictures. This is for the most part utilized as in designing applications to recognize the shapes for displaying purposes by breaking down various photographs of trees for instance, this innovation can figure out how to distinguish the tree it has never observed at this point.

3.2.4 Keras

Keras is API of high level neural networks, written in Python, capable of running on top of TensorFlow, Theano or CNTK. Keras used if you need a deep learning library it allows for fast and easy prototype supports both recurrent networks, convolutional networks and, as well as dual combinations. Runs seamlessly on CPU and GPU.

IDLE PYTHON 3.7

Python 3.7 is open source application integrated development environment for Python which is used for statistical computing and graphics. IDLE can be used to execute single statement like Python Shell and also to modify, create and execute Python scripts. **IDLE Python 3.7** provides a fully featured text editor environment to create Python scripts that includes the features, syntax highlighting, smart indent and autocompletion. It also has a debugger with stepping and breakpoints features.

Figure 3.2 Python 3.7

3.3 PROPOSED METHOD

In Proposed Method Using an image processing as well as image classification techniques it is possible to recognize hand gestures in images. The process of hand gesture recognition is divided into hand location, hand segmentation and its classification. My method locates and extracts the hand to generate new images, uses them in a Convolutional Neural Network (CNN) to classify and then recognize hand gestures. To build a classification model it is required training data, by segmenting a user's hand, from the webcam, using previous knowledge and tools, I conducted a preliminary study to detect the constraints of the problem. Results provided information about how to select better training data and a more suitable classification model. which I used this information to improve the system's recognition. To determine if a dataset has enough variance to generalize the problem. The complexity of the scene (colors, reflective materials, and other objects) takes an important role in the capture. The capture of images may not be good enough to detect, recognize or classify signs. I created an "ideal room with no elements that affects the image capture using the webcam. When capturing each image, I used the camera's depth data. The user's position was crucial because the depth sensor had a limited optimal range of capture. This range was between 1.5m and 3.5m. I set the user seated at a distance of 2m in the center of the field of view of the camera. I used the captured images to extract only the hands for processing using a classification model. Image classification is the task of taking an input image, processing it and as signing a unique label (or a class) from a set of labels. The result may be a label or a probability of belonging to diverse classes. Humans, are able to easily recognize objects in images. As we know earlier, we used high resolution webcam to capturing the images. We go for a normal environment to capture the real time image. Block diagram of proposed methodology is shown in figure (3.3)

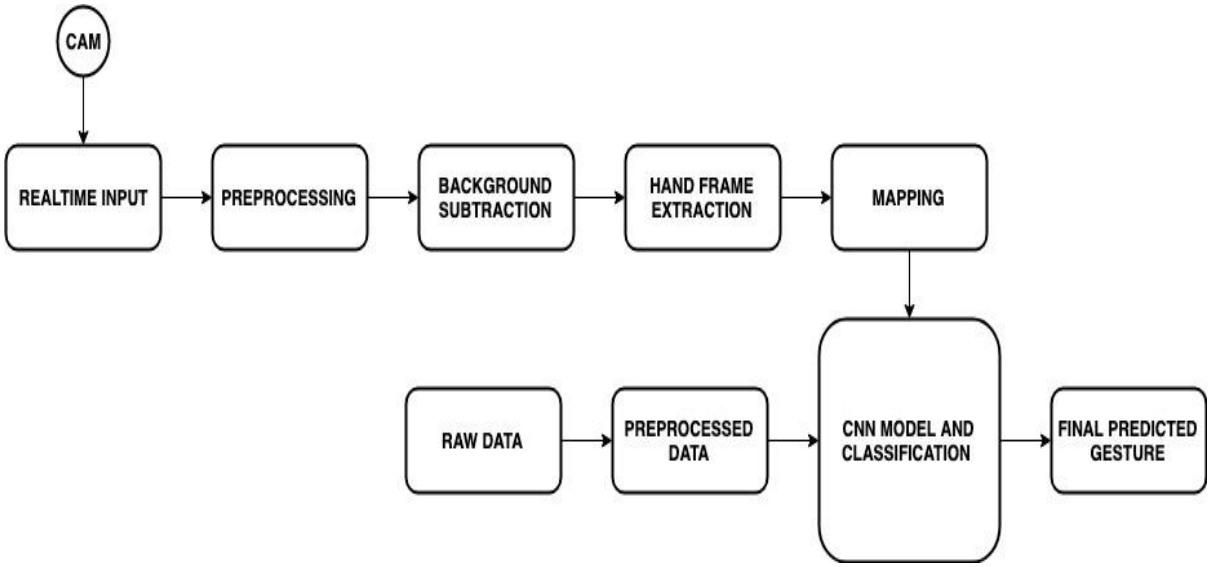
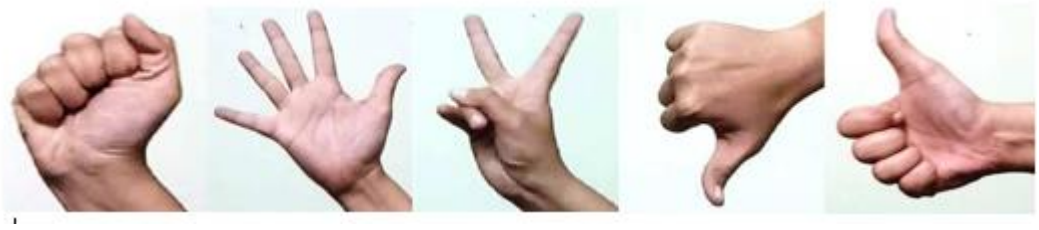


Figure 3.3 Block Diagram of Proposed Methodology

3.3.1 REAL TIME INPUT

The first step required to establish a research in this area is to acquire a real time data of various images of different positions of different hand in different illumination conditions, different positions and of different skin. The input data of images that has to be taken under consideration are of various skin types. A part of data used for testing is shown in figure 3.4.



(A) **(E)** **(K)** **(T)** **(Y)**

Figure 3.4 Real Time Input Gesture for Alphabets

The training of data base has to be accomplished so as to make the project universal for any skin type. Such that any faces, hand with any skin type can be determined easily When a computer takes an image as an input it sees a matrix of pixel values, Shown in Figure 3.5.

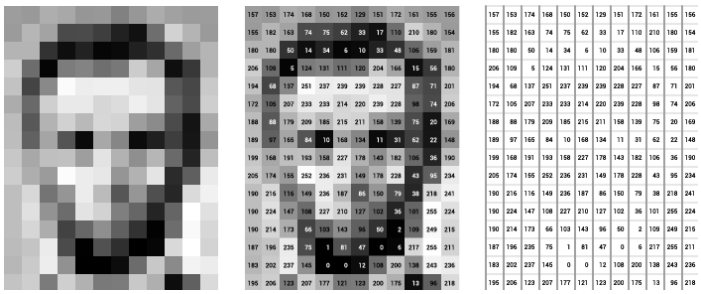


Figure 3.5. Image as Matrix Pixel Value

3.3.2 PREPROCESSING

In our Proposed Method After acquiring an input from webcam second stage preprocessing includes conversion of images such as from RGB to grayscale and YCbCr shows in fig 3.6. Furthermore, at the preprocessing stage itself filtering of noise is also established. as it is known that digital images are mostly and usually prone to noise during their transmission and acquisition. It has been a challenge in the enhancement of images to remove the noise and on the very same verge keep the features of the images in an active manner. The feature that are to be maintained during the processing of noise removal are such that, edges, texture fine details and many more. Majorly, there are two types of noises that come under consideration which have to be removed or at least have to be reduced to keep the quality of the image. The main two types of noises that occur in images are Gaussian noise and impulse noise which are introduced during the transmission and acquisition of image. The main focus always comes on the removal of Gaussian noise. Not only that, noise in the digital images can also occur if the damaged image is scanned. Also, sensors in camera can also cause noise in the image due to any kind of malfunctioning in sensors. That's not it, noise can get to an image even if there is any flaw in data transmission or even if there is any type of electronic interference. There have been a lot of methods introduced for the removal of Gaussian and impulse noise. A few of these methods can be seen as follows, a noise removal method was introduced by Buyue Zhang known as Adaptive Bilateral Filter (ABF) [28],[31]. The same method was also meant for the sharpness enhancement of the image. The sharpness of the image was managed by the increment in the slope of the edges of the image without the production of undershoot or overshoot by the ABF method. The ABF method provides a robust and more reliable and is efficient to implement. The ABF works very well for both text images and natural images.

$$
\begin{aligned}
Y' &= 16 + \frac{65.738 \cdot R'_D}{256} + \frac{129.057 \cdot G'_D}{256} + \frac{25.064 \cdot B'_D}{256} \\
C_B &= 128 - \frac{37.945 \cdot R'_D}{256} - \frac{74.494 \cdot G'_D}{256} + \frac{112.439 \cdot B'_D}{256} \\
C_R &= 128 + \frac{112.439 \cdot R'_D}{256} - \frac{94.154 \cdot G'_D}{256} - \frac{18.285 \cdot B'_D}{256}
\end{aligned}
$$

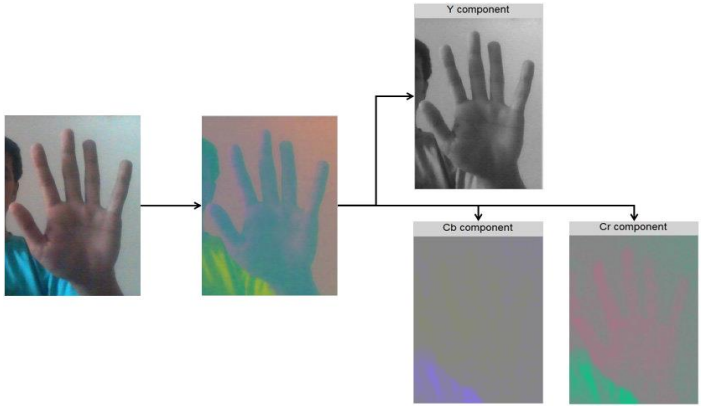


Figure 3.6 RGB to grayscale and YCbCr

It is basically a fuzzy set that takes a peer group as its support set [29]. This method is capable of reducing both impulse and Gaussian noise separately as well as mixed. With an upgradation. This method was meant to detect noise and act as a universal noise removal this method is very much capable of removing both impulse and additive Gaussian noise. This filter simultaneously removes both the noises that is Gaussian and impulse. Figure 3.7 shows (a) RGB to Greyscale, (b) Filtered image

$$g(x,y)=\frac{\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)f(x+s,y+t)}{\sum_{s=-a}^{a}\sum_{t=-b}^{b}w(s,t)}$$

(a)

(b)

Figure 3.7 (a) RGB to Greyscale (b) Filtered images

3.3.3 BACKGROUND SUBTRACTION

For the extraction of moving foreground from the static background. This is one of the most important steps of pre-processing. Background subtraction can easily be done if the image of static background is available and foreground image can be obtained. In most of the cases, such images with segregated static back ground are not available for which it is required to be adjust with available images. There are also cases where shadows can also complicate the process. In this methodology, OpenCV is used for background subtracting. Background subtracting is a technique that is widely and commonly used for the generation of a foreground mask with the help of static cameras. It basically performs a subtraction between the background model and current frame for calculating BS. It basically consists of the static part in the scene or in general everything that is present in the background. Background modelling can be divided into two steps: (1) background initialization and (2) background update. In the first step, computation of the initial model of the background is occurred while in the second case model is updated so as to adapt to the changes occurred in the scene. In this methodology, background subtractor MOG2 gaussian mixture-based background/foreground segmentation OpenCV algorithm is used. The most important feature of this algorithm is that it selects for each pixel the appropriate number of Gaussian distributions. Figure 3.8, Gaussian Mixture-Based Background Subtraction

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

$$\phi(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

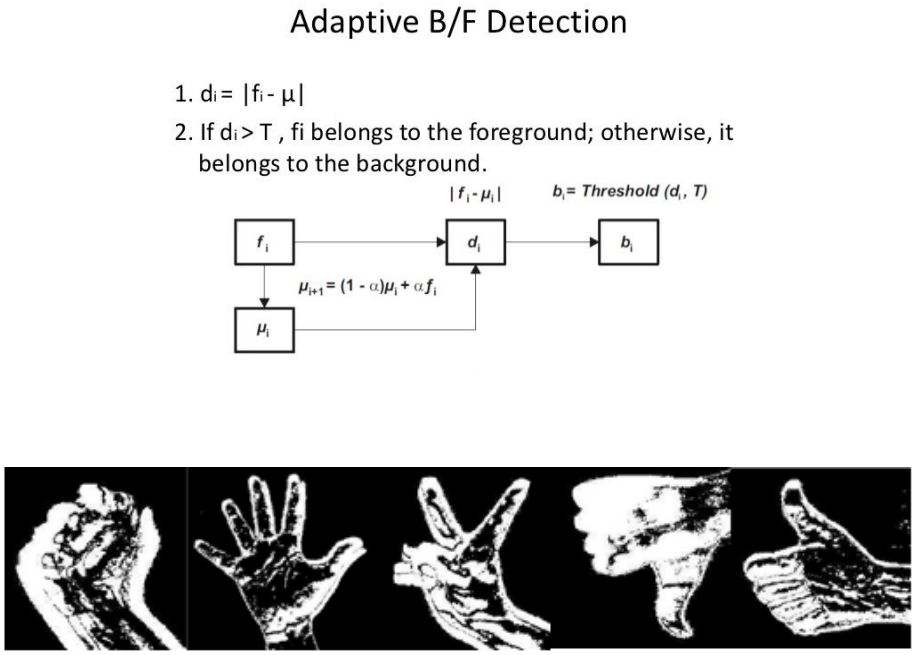


Figure 3.8 Gaussian Mixture Based Background Subtraction

3.3.4 FRAME EXTRACTION

Once the foreground is obtained from the static background, the region of interest (ROI) i.e. hand gesture is obtained. Further, the frames from images are extracted which is basically a part of dataset followed by the application of some image processing techniques. After the extraction of frames, smoothening of the images are to be done for which median blur and bilateral filters are applied on the frames. Low pass filter kernel is also applied for the smoothening of images which involves convolving. It also helps in the removal of high frequency content like noise. Edges of the images that tend to blur are also a part of the process. During the removal of noise, bilateral filter helps in keeping the edges sharp effectively [28]. Bilateral filter consists of two Gaussian filters one is the function of space making sure that only neighbor pixels are considered. The other one is a function of pixel difference that assures the blurring of only the pixels that have similar intensity to central pixel are considered. Salt and pepper noise is removed. Median blur function also replaces the central elements with some of the image pixels shows in figure 3.9.

$$B(x,y,t) = median\{I(x,y,t-i)\}$$
$$|I(x,y,t) - median\{Ix,y,t-i)\}|\,Th$$
$$where\ \_i \in \{0,...,n-1\}$$

Figure 3.9 Hand Gesture After Applying Low Pass Filter

3.3.5 RAW DATA

Raw data is basically the data on which the whole method is proposed. The data mainly consist of the images of hand gesture. The images of hand gestures are of different positions and gestures. The images in this data is consists of various cluttered back ground which are later refined according to the method. The data set is a combination of 52,000 images of hand gestures. Out of these 52,000 images 41,600 are the training images while the rest 10,400 are the testing images shown in table 3.1. The data set also consists of the skin mask and visualized skin mask. The images enclosed only hands with no other object or background. The data base consists enough number of images to generalize the gestures. The images were in the .png format which made them easy to compare to other data sets and also the content was easy to see. Figure 3.10, show the database images samples

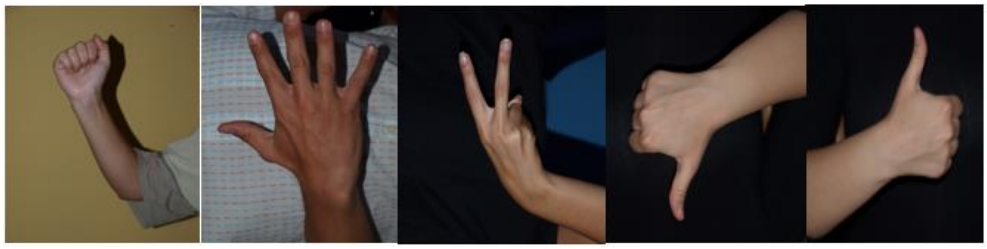


Figure 3.10 Raw Database Images

3.3.6 PREPROCESSED DATA

The preprocessed data consists gestures of the particular signs i.e. from A to Z. This data is already segmented such that the noise has already been filtered from the images. The preprocessed data is divided into two parts which includes 80% of the training data and 20% of the testing data. Preprocessing of data is basically a data mining technique that is used to transform raw data into an understandable format. The output of the transformation of the raw data is mainly termed as preprocessed data. In real-world applications the data is mostly inconsistent, incomplete, lack in some behaviors or trends and are more prone to the errors. To resolve such issues, data preprocessing is required making it suitable for further processing. Preprocessed data for training and testing model is shown in figure 3.11.

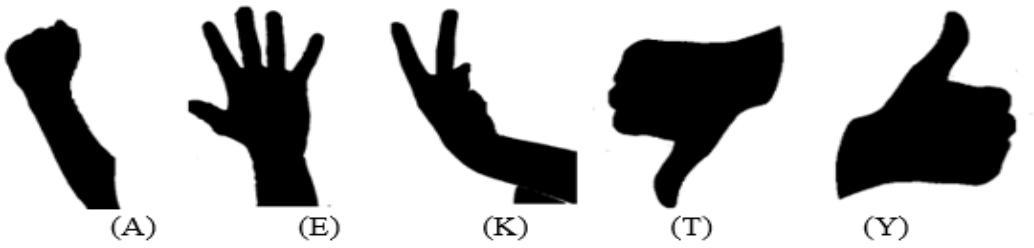


Figure 3.11 Preprocessed Data for Training and Testing Model

Table 3.1 Dataset Information and Characteristic

Dataset	**Classes**	**Input Dimension**	**No. of samples**	
			Training	***Testing***
HGR1	26	52000	41600	10400

3.3.7 CNN MODEL

Convolutional neural systems have a significant role for adaptive image processing. CNNs likewise structure a connection between versatile channels and general feed-forward neural systems. With the assistance of at least one layers of two-dimensional channels and conceivable actuation capacities or/and down examining, the two-dimensional CNNs are framed. Interpretation invariance and spatially neighborhood associations are the key properties that are controlled by CNNs. The information based versatile techniques, for example, CNNs are helpful for picture handling. They can likewise be utilized in different applications where the information exhibits are huge and briefly/spatially circulated. Further refinements of CNNs are proposed, for example, expansions to three dimensional for example video handling or usage of divisible channels. Convolutional Neural Networks (CNN) have accomplished achievement in pictures order issues.

The Convolutional Neural Networks contains three sorts of layers to be specific, convolutional layer, subsampling or pooling layer, and last completely associated layer. Regularly, whole CNN design is acquired by stacking a few layers of the previously mentioned layers. The highlights extricated from CNN utilizing three kinds of layers they are various leveled. The base layers in a Convolutional Neural Networks gather the low-level highlights. Abnormal state layers learn highlights with increasingly unique data. This is valuable for the assignment named order. Convolution layer that researches the spatially nearby and relationship of its sources of info at that point maps it into the following layer, called an element map. Distinctive component maps are developed from various pieces. Movement of the element map in the layer is determined by (1) First convolutional layer in fig can be trailed by the pooling layer. The capacity utilized in pooling layer is max work or can say normal capacity, in this layer the most well-known capacity utilized is the maximum pooling capacity. This maximum capacity is registering the abnormal state includes in a nearby window. Pooling layer is utilized to diminish size of the highlights and to diminish the required calculation time. In this segment the fundamental calibrating and adjustment procedures is connected to the essential Convolutional Neural Networks (CNN) design are talked about. This adjusted and model, CNN, is appeared in Figure 3.10. The exhibition of the CNN improved by tuning. The proposed strategy is likewise including information increase.

Machine learning algorithm works on the basis of the concept that it works more effectively if more data is accessed by it. Even if the quality of data is not that good, algorithm can perform very well till the moment it is possible for the model to extract the useful data from the original data set. The basic concept of data augmentation is to transform the base data and increase the number of training data. The transformed data is retrieved from the original data during the training process with a precisely little computation. This means that before the data is fed to the model for the training process, it is transformed. These conversions make it possible for the network to observe the impacts of introducing

it to information and also to train the network to act better. The main objective is not only to reduce overfitting by means of an increase, but also to increase the data to improve the classifier. In the case of tuned CNN in proposed method, to derive new images, random vertical and horizontal shifts are applied. For numerical increment in the size of the data base and to make it robust, deep learning approach is required. Generally, the images are shifted by 20% of the original dimensions both vertically and horizontally using data augmentation to improve the efficiency of CNN, the transformed images are also included in the data set. It also prevents over convenient and enables the function for best learning from the various pictures of a particular gesture. An overview of the proposed hand gesture recognition model is considered which is shown in figure 3.12.

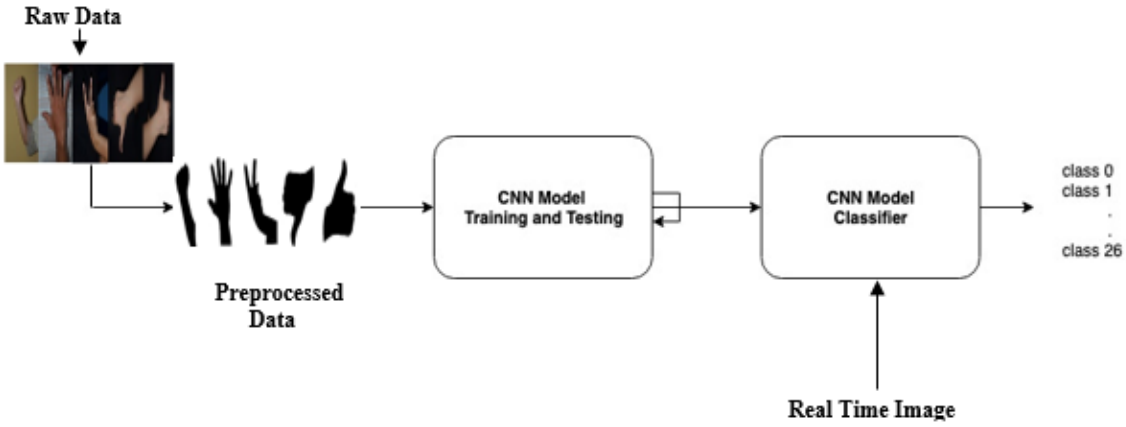


Figure 3.12 Overview of the proposed hand gesture recognition model

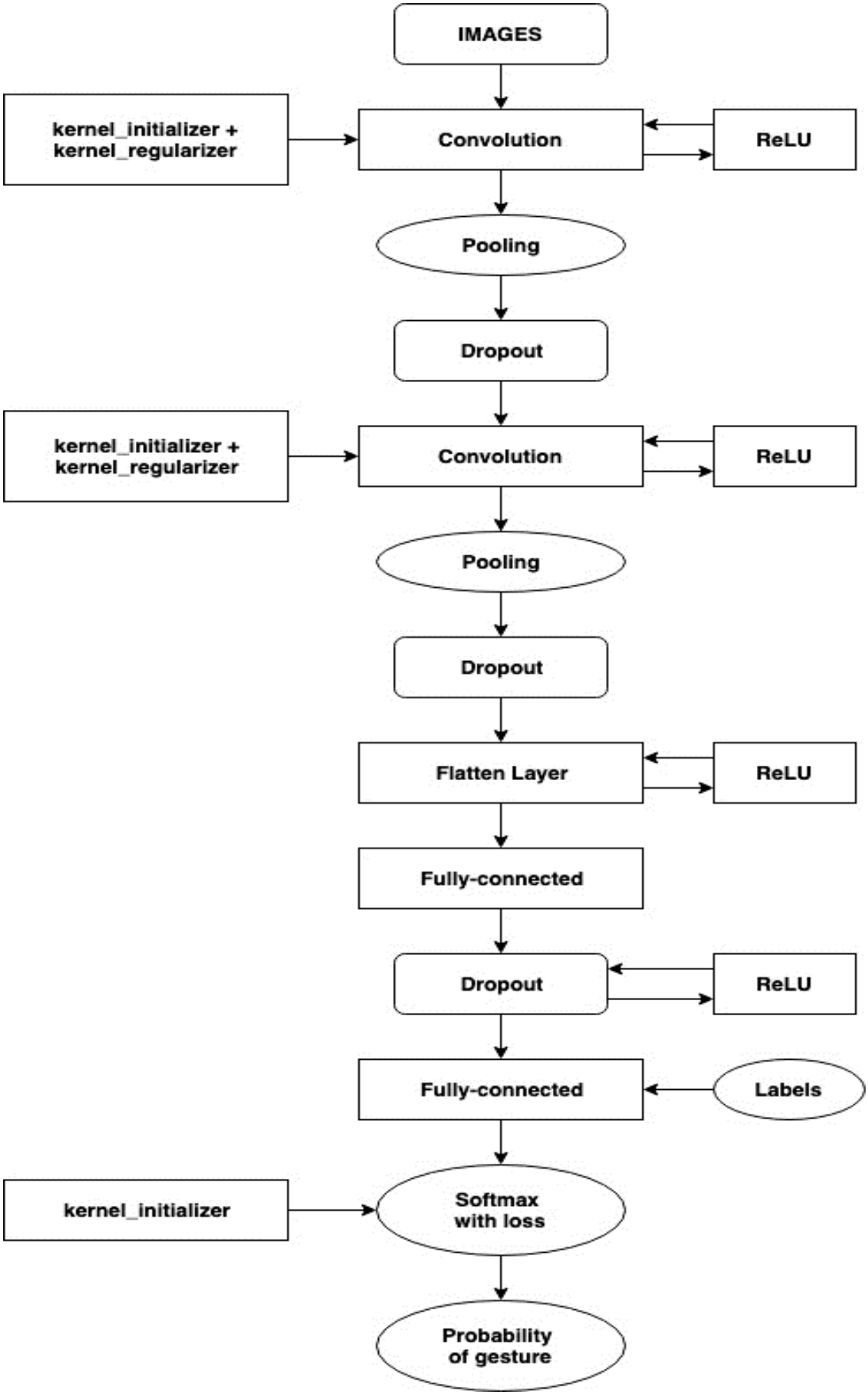


Figure 3.13 Architecture of the Proposed CNN for HGR

Table 3.2 Configuration of CNN

Layers Operation	Layers Configuration
Convolution	32 filters, 5x5 kernel and ReLU
Max-Pooling	2x2 kernel
Dropout	20%
Convolution	32 filters, 3x3 kernel and ReLU
Max-Pooling	2x2 kernel
Dropout	20%
Flatten layer	800 Neurons
Fully connected	128 Neurons
Dropout	20%
Fully connected	64 Neurons
Output layer	Softmax 6 classes

3.3.8 FINAL PRIDICTED GESTURE

Final predicted gesture is the final result of the gesture that is obtained. The gesture in this is assigned to a particular sign once the CNN model is trained, due to its feature of adaptability, at first it adapts the sign performed by the gesture. Followed by this, it classifies the gesture by the classifier into the class consisting of the images of that particular gesture. And in the end, it shows the output gesture of that particular sign. In layman's terms, the input gesture is firstly grouped with the images of same gesture which is assigned to a particular sign in the form of a class of that particular sign. Figure 3.14 shows the classified output at the end and recognize the gesture hand.

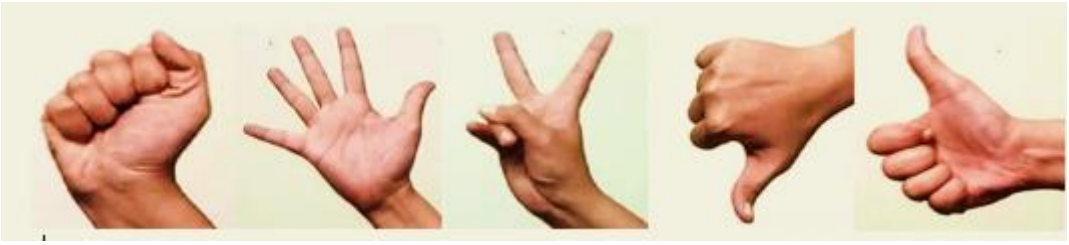
Figure 3.14 Classified output Recognized Gesture

CHAPTER 4
RESULTS AND OBSERVATION

4.1 INTRODUCTION

This Section analyze the classified results of Convolution Neural Network (CNN) model of machine learning classification with different type of testing data and kernel. The total dataset contains 52,000 preprocessed images which has 26 different classes of hand gestures. This section also includes real time data which is taken from webcam. Followed by Preprocessing and segmentation background subtraction and filtering. Figure 4.1, Show real time of gestures for selected alphabets.

4.2 REAL TIME IMAGE DATA AFTER HAND LOCALIZATION



(A) **(E)** **(K)** **(T)** **(Y)**

Figure 4.1 Show Real Time of Gestures for Selected Alphabets

Figure shows the real time data for gesture recognition. This data is the basic input for further processing. There is a whole set of gestures captured by the camera but a few of them are selected to show the results. The gestures taken as input are for the representation of alphabets A, E, K, T and Y.

4.3 RESULTS OF PREPROCESSING AND FILTERING

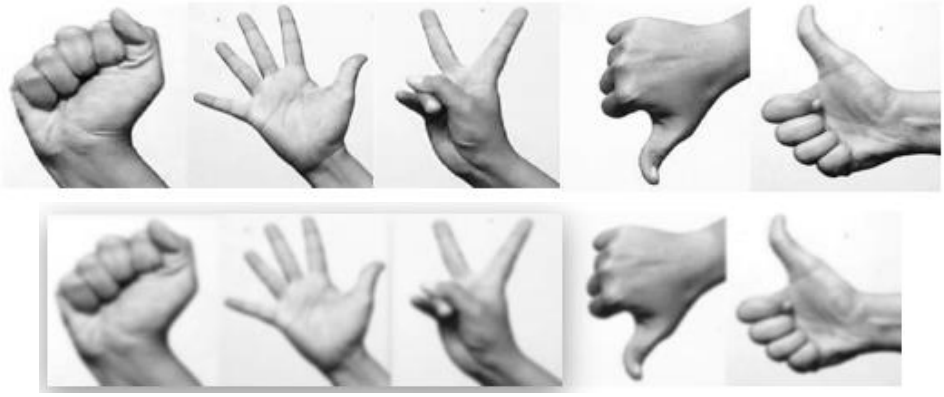


Figure 4.2 Gaussian Filtered Images

After the input image is acquired, the image is converted from RGB color space to gray scale color space as can be seen in figure the filtered images of the input. The filter used for filtration in this is gaussian filter shown in fig 4.2.

4.4 RESULT OF BACKGROUND SUBTRACTION

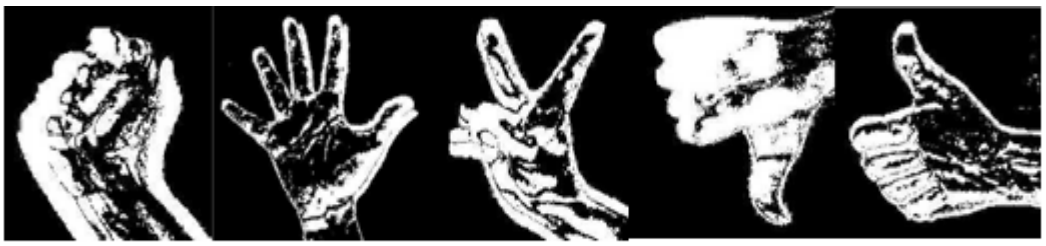


Figure 4.3 Images with Subtracted Background.

This result obtains after Background Subtraction From the real time from the real time data shown in figure 4.3 shown gaussian mixture-based background/foreground segmentation. The most important feature of this algorithm is that it selects for each pixel the appropriate number of Gaussian distributions

4.5 RESULTS OF FRAME EXTRACTION

Figure 4.4 Image Using Bilateral Filter.

Followed by the preprocessing of the image. The next step that comes along is the frame extraction. As the name says, in this the frames of the images are extracted using. In this the pixel and non pixel area Figure shows the improvised image after frame extraction. In this image the edges are highlighted and set a boundry for input image of hand gesture for further processing. Figure 4.4 shown image using bilateral filter. Figure 4.5 shown edge detection of hand gesture images.

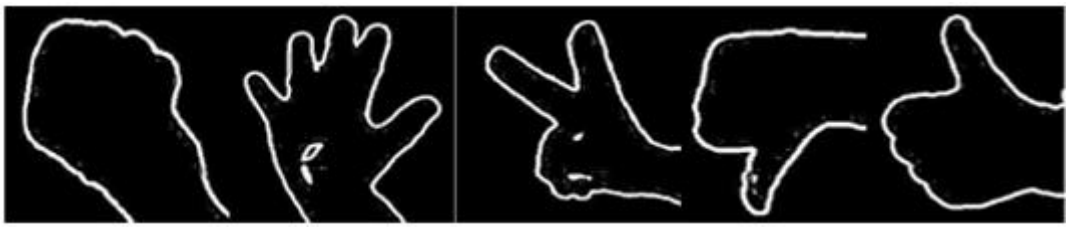


Figure 4.5 Edge Detection of Hand Gesture Images.

4.6 RESULT FOR FINAL PREDECTED GESTURES

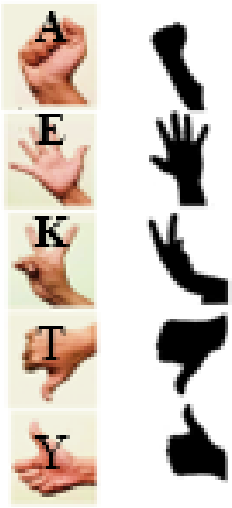


Figure 4.6 Final Predicted Result.

This show the final out put of our model recognized the gestures with their catogary.Result for final predicted gestures are shown in figure 4.6, contain the final stage result where the signing gesture to the alphabets are precisely predicted with great accuracy of our proposed model.

4.7 RESULT OF CONVULUTION NEURAL NETWORK (CNN)

Configuration

CNN has multiple layers. Operation of the layers and configuration for convolution, filters kernal and ReLU. Max pooling kernel, dropout, Flatten layes. Neurons, and Fully connected neurons. Proposed Convolution Neural Network (CNN). Configuration settings for proposed CNN is shown in Table 4.1.

Table 4.1 CNN Configuration

Layers Operation	**Layers Configuration**
Convolution	32 filters, 5x5 kernel and ReLU
Max-Pooling	2x2 kernel
Dropout	20%
Convolution	32 filters, 3x3 kernel and ReLU
Max-Pooling	2x2 kernel
Dropout	20%
Flatten layer	800 Neurons
Fully connected	128 Neurons
Dropout	20%
Fully connected	64 Neurons
Output layer	Softmax 6 classes

4.8 PARAMETERS

This is the Parameters of our proposed CNN are comparing with normal CNN show in Table 4.2.

Table 4.2 Parameters of normal CNN and proposed CNN

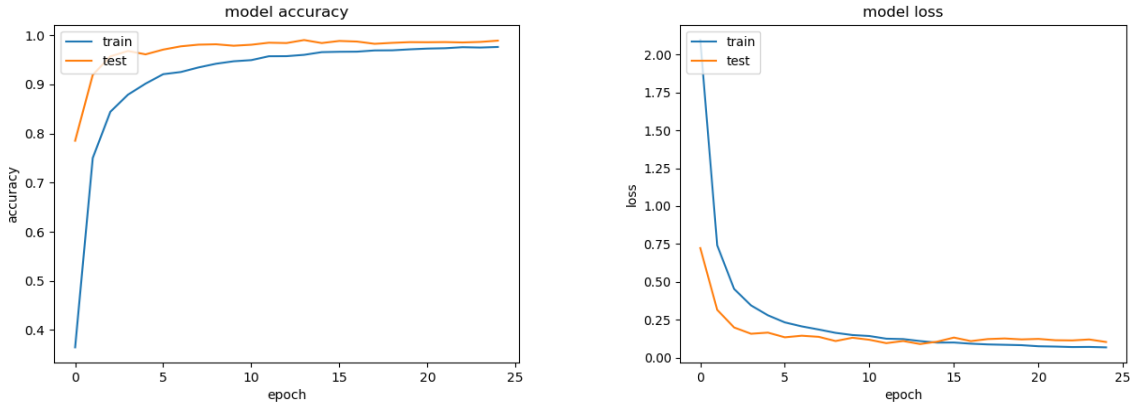
Network	(CNN)	Proposed CNN
Training samples	41600	41600
Activation function	ReLU SoftMax	ReLU-SoftMax
Learning Rate	.01	.01
Iterations.	10	10
Cost function	categorical cross entropy	categorical cross entropy
Optimization.	Adam	Adam
Data augmentation.	None	yes
Network initialization.	None	ReUL Soft max (glorotuniform)
l2 Regularization.	None	l2 Regularization

4.9 ACCURACY

Our proposed CNN model train with the specified dataset the training and testing is performed on provided data set and it has an option to train itself from the new input gesture from input camera so that accuracy of our proposed CNN or Tuned CNN is very high. The accuracy of our pre trained model is calculated for both training and testing partition 80-20 is shown in Table 4.3, figure 4.7. Training accuracy and losses. figure 4.8 shows Confusion Matrix (a) CNN, (b) proposed CNN.

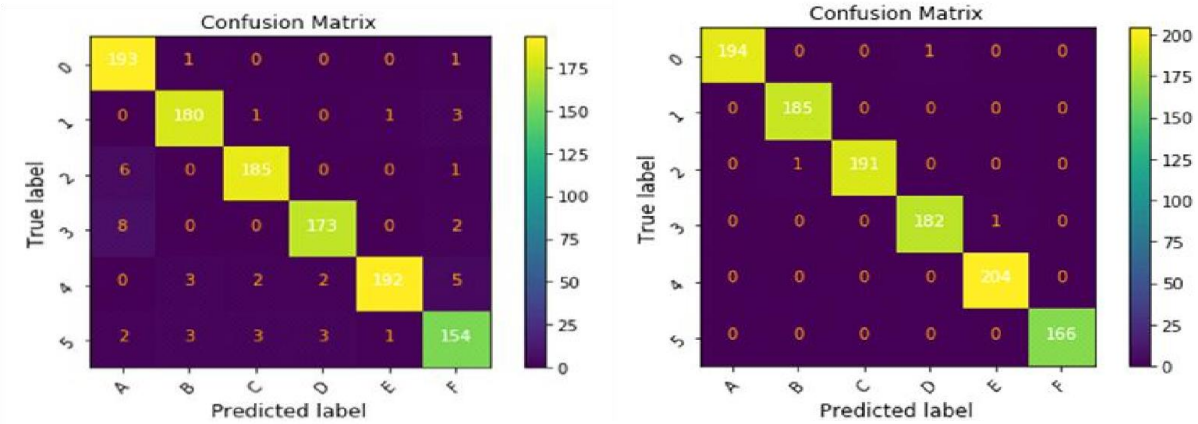
Table 4.3 Accuracy of CNN and Proposed CNN

Methods	**No. Epoch**	**Precision**	**Recall**	**F1 Score**	**Accuracy (%)**
CNN	26	0.95	0.95	0.95	95.10
Proposed CNN	26	1	1	1	97.73



(a) (b)

Figure 4.7 (a) Model accuracy, (b) Model Losses



(a) (b)

Figure 4.8 Confusion matrix (a) CNN, (b) Proposed CNN

CHAPTER 5
CONCLUSION AND FUTURE SCOPE

5.1 CONCLUSION

Various research or studies have various characteristics, advantages, disadvantages and most importantly future scope. By looking at the review it can be concluded that in the area of skin segmentation in image processing, it can be implemented on various parts such as face. While through the study it is observed that skin segmentation of hands can make out to be in great applications such as communication between deaf and dumb people. This study is yet to be fully justified. As in the case of future research hand is to be detected from a cluttered background in real time. The performance of the proposed method highly depends on the result of hand detection. If there are moving objects with the color similar to that of the skin, the objects exist in the result and then degrade the performance of the hand gesture recognition. This gives us another case to be taken care of, that is, the accuracy in skin segmentation for hand gestures followed by the artificial neural network we have test the basic algorithms for data mining and classification like SVM, ANN, CNN. We conclude that these tools and algorithms are very useful for classification. In Convolution Neural Network we were using the CNN model and perform the training and testing on preprocessed data. Once the model trained it will give us the better performance to recognize the Hand Gestures. I used Tuned CNN basically I used Convolution Neural Network on tuned or manual configuration settings. This method increases the performance and accuracy of the Hand Gesture Recognition model.

5.2 FUTURE SCOPE

Future scope of Hand Gesture Recognition technology life has become a science fiction movie. This technology became more faster reliable with the human computer interaction. This emerging trend will give better performance and be better define the gesture for Hand Gesture Recognition. It can be used in medical research, gaming etc. that can be controlled with the hands in the future such as:

1. 3D Designing – interaction with 3D objects similar to real life
2. Hardware Control - Gesture controllers can be used to control hardware like robots, cars or even drones
3. Smart Display - Interaction with Displays at public spaces
4. Smart Helmets - Interaction with helmets with driving like speed or navigation.
5. We want to make it into a complete product that makes the deaf and dumb people to communicate like normal people.

REFERENCES

[1] Filali, Youssef, et al. "Texture Classification of skin lesion using convolutional neural network." *2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*. IEEE, 2019.

[2] Kurmanji, Meghdad, and Foad Ghaderi. "A Comparison of 2D and 3D Convolutional Neural Networks for Hand Gesture Recognition from RGB-D Data." *2019, 27th Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2019.

[3] Sun, Jing-Hao, et al. "Research on the Hand Gesture Recognition Based on Deep Learning." *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*. IEEE, 2018.

[4] Codella, Noel CF, et al. "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)." *Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on*. IEEE, 2018.

[5] Singha, Joyeeta, Amarjit Roy, and Rabul Hussain Laskar. "Dynamic hand gesture recognition using vision-based approach for human–computer interaction." *Neural Computing and Applications* 29.4 (2018): pp.1129-1141.

[6] Alom, Md Zahangir, et al. "Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation." *arXiv preprint arXiv:1802.06955* (2018).

[7] Bhujbal, Mr Vishal P., and K. K. Warhade. "Hand Sign Recognition Based Communication System for Speech Disable People." *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2018.

[8] Li, Yuexiang, and Linlin Shen. "Skin lesion analysis towards melanoma detection using deep learning network." *Sensors*18.2 (2018): 556. 17)

[9] Panchal, Twinkal H., and Pradip R. Patel. "A Novel Approach of Sign Recognition for Indian Sign." (2018).

[10] Pigou, Lionel, et al. "Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video." *International Journal of Computer Vision* 126.2-4 (2018): 430-439.

[11] Y. Filali, A. Ennouni, M. A. Sabri and A. Aarab, "A study of lesion skin segmentation, features selection and classification approaches," *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, 2018, pp. 1-7.

[12] Saha, Himadri Nath, et al. "A machine learning based approach for hand gesture recognition using distinctive feature extraction." *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2018.

[13] Chakraborty, Biplab Ketan, et al. "Review of constraints on vision-based gesture recognition for human–computer interaction." *IET Computer Vision* 12.1 (2017): 3-15.Sanchez-Cuevas,

[14] Kakkoth, Sarang Suresh, and Saylee Gharge. "Survey on real time hand gesture recognition." *2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*. IEEE, 2017.

[15] Nishihara, H. Keith, et al. "Hand-gesture recognition method." U.S. Patent No. 9,696,808. 4 Jul. 2017.

[16] Benalcázar, Marco E., et al. "Real-time hand gesture recognition using the myo armband and muscle activity detection." *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*. IEEE, 2017.

[17] Benalcázar, Marco E., et al. "Hand gesture recognition using machine learning and the Myo armband." *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017.

[18] Jiménez, Luis A. Estrada, Marco E. Benalcázar, and Nelson Sotomayor. "Gesture recognition and machine learning applied to sign language translation." *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th-28th, 2016*. Springer, Singapore, 2017.

[19] Ekbote, Juhi, and Mahasweta Joshi. "Indian sign language recognition using ANN and SVM classifiers." *2017 International conference on innovations in information, embedded and communication systems (ICIIECS)*. IEEE, 2017.

[20] Chakraborty, Biplab Ketan, Manas Kamal Bhuyan, and Sunil Kumar. "Combining image and global pixel distribution model for skin colour segmentation." *Pattern Recognition Letters* 88 (2017): 33-40.

[21] Chakraborty, Biplab Ketan, M. K. Bhuyan, and Sunil Kumar. "Adaptive propagation-based skin segmentation method for color images." *Communication (NCC), 2016 Twenty Second National Conference on*. IEEE, 2016.

[22] Rautaray, Siddharth S., and Anupam Agrawal. "Vision based hand gesture recognition for human computer interaction: a survey." *Artificial Intelligence Review* 43.1 (2015): 1-54.

[23] Liao, Chung-Ju, Shun-Feng Su, and Ming-Chang Chen. "Vision-based hand gesture recognition system for a dynamic and complicated environment." *2015 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2015.

[24] Shukla, Jaya, and Ashutosh Dwivedi. "A method for hand gesture recognition." *2014 Fourth International Conference on Communication Systems and Network Technologies*. IEEE, 2014.

[25] M. Kawulok, “Fast propagation-based skin regions segmentation in color images,” in Proc. 10th IEEE Int. Conf. and Workshops Autom. Face and Gesture Recognition (FG), April 2013, pp. 1–7.

[26] Manuel C., Ruth M. Aguilar-Ponce, and J. Luis Tecpanecatl-Xihuitl. "A comparison of color models for color face segmentation." *Procedia Technology* 7 (2013): 134-141.

[27] Naji, Sinan A., Roziati Zainuddin, and Hamid A. Jalab. "Skin segmentation based on multi pixel color clustering models." *Digital Signal Processing* 22.6 (2012): 933-940.

[28] Lin, Chih-Hsing, Jia-Shiuan Tsai, and Ching-Te Chiu. "Switching bilateral filter with a texture/noise detector for universal noise removal." *IEEE Transactions on Image Processing* 19.9 (2010): 2307-2320.

[29] Morillas, Samuel, Valentín Gregori, and Antonio Hervás. "Fuzzy peer groups for reducing mixed Gaussian-impulse noise from color images." *IEEE Transactions on Image Processing* 18.7 (2009): 1452-1466.

[30] Garg, Pragati, Naveen Aggarwal, and Sanjeev Sofat. "Vision based hand gesture recognition." *World Academy of Science, Engineering and Technology* 49.1 (2009): 972-977.

[31] Zhang, Buyue, and Jan P. Allebach. "Adaptive bilateral filter for sharpness enhancement and noise removal." *IEEE transactions on Image Processing* 17.5 (2008): 664-678.

[32] Alijanpour, Nasibeh, Hossein Ebrahimnezhad, and Afshin Ebrahimi. "Inner distance based hand gesture recognition for devices control." *2008 International Conference on Innovations in Information Technology*. IEEE, 2008.

[33] Phung, Son Lam, Abdesselam Bouzerdoum, and Douglas Chai. "Skin segmentation using color pixel classification: analysis and comparison." *IEEE transactions on pattern analysis and machine intelligence* 27.1 (2005): 148-154.

[34] Wolfel, M., & Ekenel, H. K. (2005, September). Feature weighted Mahalanobis distance: improved robustness for Gaussian classifiers. In *Signal Processing Conference, 2005 13th European* (pp. 1-4). IEEE.

[35] M. J. Jones and J. Rehg, “Statistical color models with application to skin detection,” Int. J. Comput. Vis., vol. 46, no. 1, pp. 81–96, 2002.