**Real Time Hand Gesture Recognition System for Dynamic Applications.**

Submitted in partial fulfilment of the requirements for the award of degree of

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE & ENGINEERING**



**Submitted To:**

**Prof. Dr. Alok Kumar**

**Submitted By:**

**NAME (Enrolment No.)**

**1)Kondra Sharath Reddy 21CS002388**

**2)Gaddam Kalyan Teja 21CS002373**

**3)M. Shiva Keshava Reddy 21CS002396**

**Sir Padampat Singhania University, Udaipur.**

**January 2024.**

**DECLARATION**

I **Kondra. Sharath Reddy(21CS002388)** hereby declare that the partial submission of the project work entitled **Real Time Hand Gesture Recognition System for Dynamic Applications** is an authentic record of our own team work carried out at Sir Padampat Singhania University as requirement of Minor Project for the award of degree of B. Tech CSE(AIML), under the guidance and supervision of **Prof. Dr. Alok Kumar** (Faculty Coordinator), during January to March 2024.

Word Count:4582

Name: Sharath Reddy, Kalyan Teja, Shiva Keshava

Enrolment. No: 21CS002388, 21CS002373, 21CS002396.

Date of submission: \_\_\_\_\_\_\_\_\_\_\_\_

Signature of Student:

Certified that the above statement made by the student is correct to the best of our knowledge and belief

Prof. Dr. Alok Kumar

Faculty Coordinator

**Acknowledgement**

I **K.SharathReddy(21CS002388),** Besides of my team would like to express my heartfelt appreciation to all those who have played a role in the advancement and realization of this project focusing on **Real Time Hand Gesture Recognition System for Dynamic Applications.**

Your continuous support and guidance have been instrumental in the successful completion of this academic endeavour.

I am deeply grateful to my esteemed faculty advisor, **Prof. Dr. Alok Kumar**, whose unwavering support, insightful feedback, and encouragement have been invaluable throughout this journey. Your mentorship has profoundly influenced the direction and progress of this research endeavour. Thank you for your invaluable insights, thought-provoking discussions, and for providing clarity to my ideas.

I extend my sincere thanks to the faculty members of the Department of Computer Science & Engineering at **Sir Padampat Singhania University**, Udaipur, for their expertise and guidance, which have significantly enriched the methodologies and approaches employed in this project.

Special appreciation is also extended to my peers for their collaborative efforts, brainstorming sessions, and constructive critiques, all of which have contributed to refining and improving this project.

I am grateful for the unwavering support and dedication of everyone involved. Together, we are making significant contributions towards the advancement of autonomous vehicle technology and the enhancement of road safety.

**Abstract**

The project introduces an application using computer vision for Hand gesture recognition. A camera records a live video stream, from which a snapshot is taken with the help of interface. The system is trained for each type of count **hand gestures** (one, two, three, four, and five) at least once. After that a test gesture is given to it and the system tries to recognize it.

Research was carried out on a number of algorithms that could best differentiate a hand gesture. It was found that the diagonal sum algorithm gave the highest accuracy rate. In the preprocessing phase, a self-developed algorithm removes the background of each training gesture. After that the image is converted into a binary image and the sums of all diagonal elements of the picture are taken. This sum helps us in differentiating and classifying different hand gestures.

Previous systems have used data gloves or markers for input in the system. I have no such constraints for using the system. The user can give hand gestures in view of the camera naturally. A completely robust hand gesture recognition system is still under heavy research and development; the implemented system serves as an extendible foundation for future work.

**Table of Content**

CHAPTER 1 INTRODUCTION……………………………………………7

1.1 DIGITAL IMAGE PROCESSING……………………………………7

1.2 BIOMETERICS………………………………………………………8

1.3 HAND GESTURE DETECTIOIN AND RECOGNITION………….8

1.3.1 DETECTION…………………………………………………………8

1.3.2 RECOGNITION……………………………………………………...10

1.4 SOFTWARE TOOLS………………………………………………...11

1.5 OBJECTIVES………………………………………………………..11

CHAPTER 2: LITERATURE REVIEW…………………………………....12

2.1 CAMERA ORIENTATIONS AND DISTANCE……………………..12

2.2 DIFFERENT RECOGNITION APPROACHES……………………..13

2.2.1 Tracker-Based Gesture Recognition………………………………….13

2.2.2 Body Suits…………………………………………………………….13

2.2.3 Head and Face Gesture………………………………………………..13

2.2.4 Hand and Arm Gestures………………………………………………14

2.2.5 Vision-Based Gesture Recognition…………………………………...14

CHAPTER 3: METHODOLGY……………………………………………..16

3.1 PROJECT CONSTRAINTS…………………………………………..16

3.2 THE WEBCAM SYSTEM (USB PORT)……………………………..17

3.3 BRIEF OUTLINE OF THE IMPLEMENTED SYSTEM……………...17

3.4 Real Time Classification………………………………………………...18

CHAPTER 4: FEATURE EXTRACTIONS…………………………………...19

4.1 NEURAL NETWORKS………………………………………………...19

CHAPTER 5: RESULTS AND DISCUSSION………………………………..22

5.1 Experiments & Analysis…………………………………………………..22

5.2 Gesture Recognition………………………………………………………23

CHAPTER 6: CONCLUSION AND FUTURE WORK………………………24

6.1 CONCLUSION…………………………………………………………24

6.2 FUTURE WORK……………………………………………………….25

REFERENCES………………………………………………………………...26

GITHUB REPOSITORY………………………………………………………27

**LIST OF FIGURES**

Fig1.1 Lighting condition and Background

Fig1.2 Hand Gesture recognition flow chart

Fig3.1 System Implementation

Fig3.5 Real time Classification

Fig4.1 Neural Network Block diagram

Fig4.3 NN For Mean and Standard Deviation

Fig5.2 Classification Percentage

**CHAPTER 01**

1. **Introduction**

Recent developments in computer software and related hardware technology have provided a value added service to the users. In everyday life, physical gestures are a powerful means of communication. They can economically convey a rich set of facts and feelings. For example, waving one's hand from side to side can mean anything from a "happy goodbye" to "caution". Use of the full potential of physical gesture is also something that most human computer dialogues lack.

The task of hand gesture recognition is one the important and elemental problem in computer vision. With recent advances in information technology and media, automated human interactions systems are built which involve hand processing task like hand detection, hand recognition and hand tracking.

This prompted my interest so I planned to make a software system that could recognize human gestures through computer vision, which is a sub field of artificial intelligence. The purpose of my software through computer vision was to program a computer to "understand" a scene or features in an image.

A first step in any hand processing system is to detect and localize hand in an image. The hand detection task was however challenging because of variability in the pose, orientation, location and scale. Also, different lighting conditions add further variability.

* 1. **Digital Image Processing**

Image processing is reckoned as one of the most rapidly involving fields of the software industry with growing applications in all areas of work. It holds the possibility of developing the ultimate machines in future, which would be able to perform the visual function of living beings. As such, it forms the basis of all kinds of visual automation.

* 1. **Biometrics**

Biometric systems are systems that recognize or verify human beings. Some of the most important biometric features are based physical features like hand, finger, face and eye. For instance, finger print recognition utilizes of ridges and furrows on skin surface of the palm and fingertips. Hand gesture detection is related to the location of the presence of a hand in still image or in sequence of images i.e. moving images. Other biometric features are determined by human behavior like voice, signature and walk. The way humans generate sound for mouth, nasal cavities and lips is used for voice recognition. Signature recognition looks at the pattern, speed of the pen when writing one’s signature.

1. **Hand Gesture Recognition and Detection**
2. **Detection**

Hand detection is related to the location of the presence of a hand in a still image or sequence of images i.e. moving images. In case of moving sequences, it can be followed by tracking of the hand in the scene but this is more relevant to the applications such as sign language. The underlying concept of hand detection is that human eyes can detect objects which machines cannot with that much accuracy as that of a human. From a machine point of view, it is just like a man fumble around with his senses to find an object.

**Variations in image plane and pose**

The hands in the image vary due to rotation, translation and scaling of the camera pose or the hand itself. The rotation can be both in and out of the plane.

**Lighting Condition and Background**

As shown in Figure 1.1 light source properties affect the appearance of the hand. Also, the background, which defines the profile of the hand, is important and cannot be ignored.

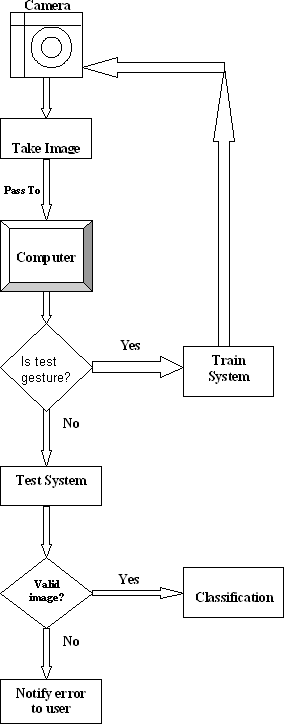






**Figure: 1.1: Lighting Condition and Background**

**1.3.2 Recognition**

Hand detection and recognition have been significant subjects in the field of computer vision and image processing during the past 30 years. There have been considerable achievements in these fields and numerous approaches have been proposed. However, the typical procedure of a fully automated hand gesture recognition system can be illustrated in the **Figure 1.2** below:

**Figure 1.2: Hand Gesture Recognition Flow Chart**

**1.4 Software Tools**

Due to the time constraint and complexity of implementing system in C++, the aim was to design a prototype under MATLAB that was optimized for detection performance. A system that accepted varying inputs of different sizes and image resolutions was implemented; constructing a well coded and documented system for easier future development.

**1.5 Objectives**

Due to the time constraint and complexity of implementing system in C++, the aim was to design a prototype under MATLAB that was optimized for detection performance. A system that accepted varying inputs of different sizes and image resolutions was implemented; constructing a well coded and documented system for easier future development.

**CHAPTER 02**

1. **Literature Review**

Hand gesture recognition research is classified in three categories. First “**Glove based Analysis**” attaching sensor with gloves mechanical or optical to transduces flexion of fingers into electrical signals for hand posture determination and additional sensor for position of the hand. This sensor is usually an acoustic or a magnetic that attached to the glove. Look-up table software toolkit provided for some applications to recognize hand posture.

The second approach is “**Vision based Analysis**” that human beings get information from their surroundings, and this is probably most difficult approach to employ in satisfactory way. Many different implementations have been tested so far. One is to deploy 3-D model for the human hand. Several cameras attached to this model to determine parameters corresponding for matching images of the hand, palm orientation and joint angles to perform hand gesture classification. Lee and Kunii developed a hand gesture analysis system based on a three-dimensional hand skeleton model with 27 degrees of freedom. They incorporated five major constraints based on the human hand kinematics to reduce the model parameter space search. To simplify the model matching, specially marked gloves were used.

The Third implementation is “**Analysis of drawing gesture**” use stylus as an input device. This drawing analysis led to recognition of written text. Mechanical sensing work has used for hand gesture recognition at vast level for direct and virtual environment manipulation. Mechanically sensing hand posture has many problems like electromagnetic noise, reliability and accuracy. By visual sensing gesture interaction can be made potentially practical but it is most difficult problem for machines.

* 1. **Camera Orientation and Distance**

It is very important to careful about direction of camera to permit easy choice of background. Two good and more effective approaches are to point the camera towards wall or floor. Lighting was standard room; intensity of light would be higher and shadowing effects lower because camera was pointed downwards. The distance of the camera from the hand should be such that it covers the entire gesture mainly. There is no effect found on the accuracy of the system if the image is a zoomed one or not; the principle is to cover the entire hand area majorly.

* 1. **Different Recognition Approaches**
     1. **Tracker-Based Gesture Recognition**

There are many tracking systems available commercially which can used for gesture recognition, primarily tracking eye gaze, hand gesture, and overall body and its position. In virtual environment interaction each sensor has its own strengths and weaknesses. Gestural interface eye gaze can be useful, so I focus here on gesture based input from tracking the hand and the body.

* + 1. **Body Suits**

Process of small place of strategically dots placed on human body, people can perceive patterns such as gestures, activities, identities and other aspects of body. One way of approach is recognition of postures and human movements is optically measure of 3D position such as markers attached to body and then recovers time varying articulate structure of body. This articulated sensing by position and joint angles using electromechanically sensors. Although some of system require small ball or dot placed top user clothing. I prefer body motion capture by “body suits” generically.

* + 1. **Head and Face Gestures**

When people interact with one another, they use an assortment of cues from the head and face to convey information. These gestures may be intentional or unintentional, they may be the primary communication mode or back channels, and they can span the range from extremely subtle to highly exaggerate. Some examples of head and face gestures include: nodding or shaking the head, direction of eye gaze, raising the eyebrows, opening the mouth to speak, winking, flaring the nostrils and looks of surprise, happiness, disgust, anger, sadness, etc.

People display a wide range of facial expressions. Ekman and Friesen developed a system called FACS for measuring facial movement and coding expression; this description forms the core representation for many facial expression analysis systems.

A real-time system to recognize actions of the head and facial features was developed by Zelinsky and Heinzmann, who used feature template tracking in a Kalman filter framework to recognize thirteen head/face gestures.

* + 1. **Hand and Arm Gestures**

These two parts of body (Hand & Arm) have most attention among those people who study gestures in fact much reference only consider these two for gesture recognition. The majority of automatic recognition systems are for deictic gestures (pointing), emblematic gestures (isolated signs) and sign languages (with a limited vocabulary and syntax). Some are components of bimodal systems, integrated with speech recognition. Some produce precise hand and arm configuration while others only coarse motion.

Stark and Kohler developed the ZYKLOP system for recognizing hand poses and gestures in real-time. After segmenting the hand from the background and extracting features such as shape moments and fingertip positions, the hand posture is classified. Temporal gesture recognition is then performed on the sequence of hand poses and their motion trajectory. A small number of hand poses comprises the gesture catalog, while a sequence of these makes a gesture.

Freeman developed a real-time system to recognize hand poses using image moments and orientations histograms, and applied it to interactive video games. Cutler and Turk described a system for children to play virtual instruments and interact with life like characters by classifying measurements based on optical flow.

* + 1. **Vision-Based Gesture Recognition**

The most significant disadvantage of the tracker-based systems is that they are cumbersome. This detracts from the immerse nature of a virtual environment by requiring the user to put on an unnatural device that cannot easily be ignored, and which often requires significant effort to put on and calibrate. Even optical systems with markers applied to the body suffer from these shortcomings, albeit not as severely. What many have wished for is a technology that provides real-time data useful for analyzing and recognizing human motion that is passive and non-obtrusive. Computer vision techniques have the potential to meet these requirements.

Typically, the camera locations are fixed in the environment, although they may also be mounted on moving platforms or on other people. For the past decade, there has been a significant amount of research in the computer vision community on detecting and recognizing faces, analyzing facial expression, extracting lip and facial motion to aid speech recognition, interpreting human activity, and recognizing particular gestures.

Unlike sensors worn on the body, vision approaches to body tracking have to contend with occlusions. From the point of view of a given camera, there are always parts of the user’s body that are occluded and therefore not visible – e.g., the backside of the user is not visible when the camera is in front. More significantly, self-occlusion often prevents a full view of the fingers, hands, arms, and body from a single view. Multiple cameras can be used, but this adds correspondence and integration problems.

Unlike special devices, which measure human position and motion, vision uses a multipurpose sensor; the same device used to recognize gestures can be used to recognize other objects in the environment and also to transmit video for teleconferencing, surveillance, and other purposes. There is a growing interest in CMOS-based cameras, which promise miniaturized, low cost, low power cameras integrated with processing circuitry on a single chip.

This technique was also used by us for recognizing hand gestures in real time. With the help of a web camera, I took pictures of hand on a prescribed background and then applied the classification algorithm for recognition.

**CHAPTER 03**

1. **Methodology**

There have been numerous researches in this field and several methodologies were proposed like Principal Component Analysis (PCA) method, gradient method, subtraction method etc. PCA relates to Linear transformation consist on statistical approach. This gives us powerful tool for pattern recognition and data analysis which mostly used in image processing techniques for data (compression, dimension and correlation). Gradient method is also another image processing technique that detect colour patches applying low pass filters is also known as edge detection method. Subtraction method is very simple that subtract input image pixel to another image or constant value to provide output. I have also studied different approaches to hand gesture recognition and came to know that implementation of such techniques like PCA and Gradient method is complicated, we can produce same output as these techniques gives us by simple and easy implementation. So, I have tried four different algorithms and finally selected the one, which was most efficient i.e. diagonal sum algorithm. This algorithm is able to recognize maximum gestures correctly.

* 1. **Project Constraints**

I propose a vision-based approach to accomplish the task of hand gesture detection. As discussed above, the task of hand gesture recognition with any machine learning technique suffers from the variability problem. To reduce the variability in hand recognition task we assume the following assumptions:

* Single coloured camera mounted above a neutral-coloured desk.
* User will interact by gesturing in the view of the camera.
* Training is must.
* Hand will not be rotated while image is capturing.

The real time gesture classification system depends on the hardware and software.

**Hardware**

* + - Minimum 2.8 GHz processor Computer System or latest
    - 52X CD-ROM drive
    - Web cam (For real-time hand Detection)

**Software**

* + - Windows 2000(Service Pack 4),XP, Vista or Windows 7
    - MATLAB 8.0 or latest (installed with image processing toolbox)
    - Vcapg2.dll (Video Capture Program Generation 2)
    - DirectX 9.0 (for supporting Vcapg2)
  1. **The Webcam System**

Below is the summary of the specifications of the camera which this system required:

**Resolution:** 640x480

**Video frame rate:** 30fps @640x480

**Pixel depth:** Minimum 1.3-mega pixels

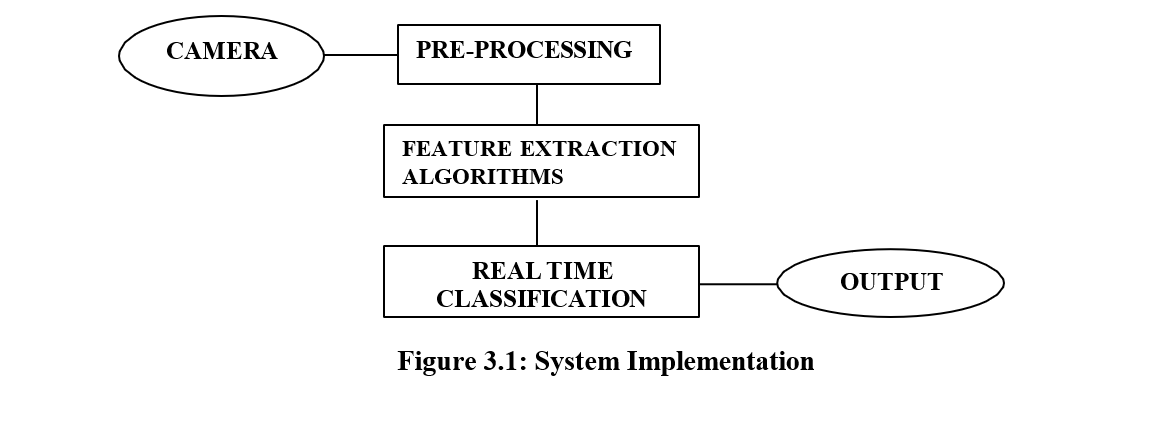
**Connection Port:** USB

In my project web cam was attached via USB port of the computer. The web cam worked by continually capturing the frames. In order to capture a particular frame, the user just need to select the particular Algorithm METHOD button on the interface and the hand was detected in the particular frame. The web cam took color pictures, which were then converted into grayscale format. The main reason of sticking to grayscale was the extra amount of processing required to deal with color images.

* 1. **Brief Outline of the Implemented System**

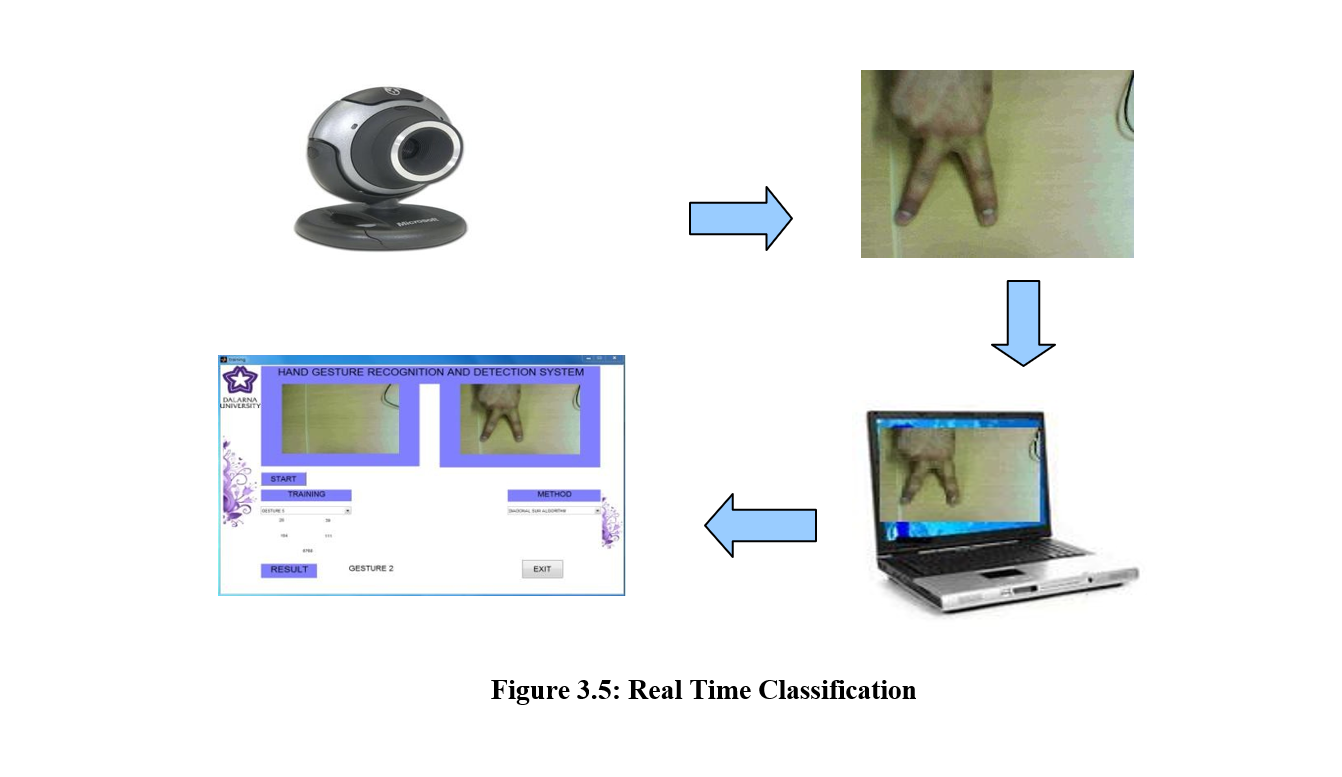
Hand gesture recognition system can be divided into following modules:

* Preprocessing
* Feature extraction of the processed data
* Real time classification



* 1. **Real Time Classification**

**Figure 3.5** shows the concept for real time classification system. A hand gesture image will be passed to the computer after being captured through camera at run time and the computer will try to recognize and classify the gesture through computer vision.



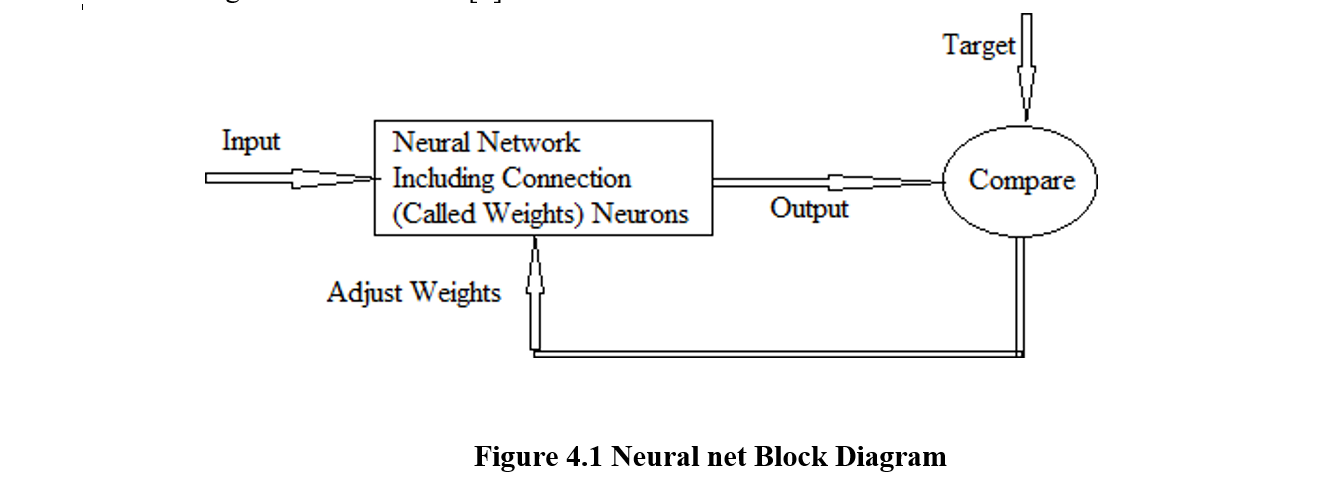
**CHAPTER 04**

**Feature Extractions**

**4.1 Neural Networks**

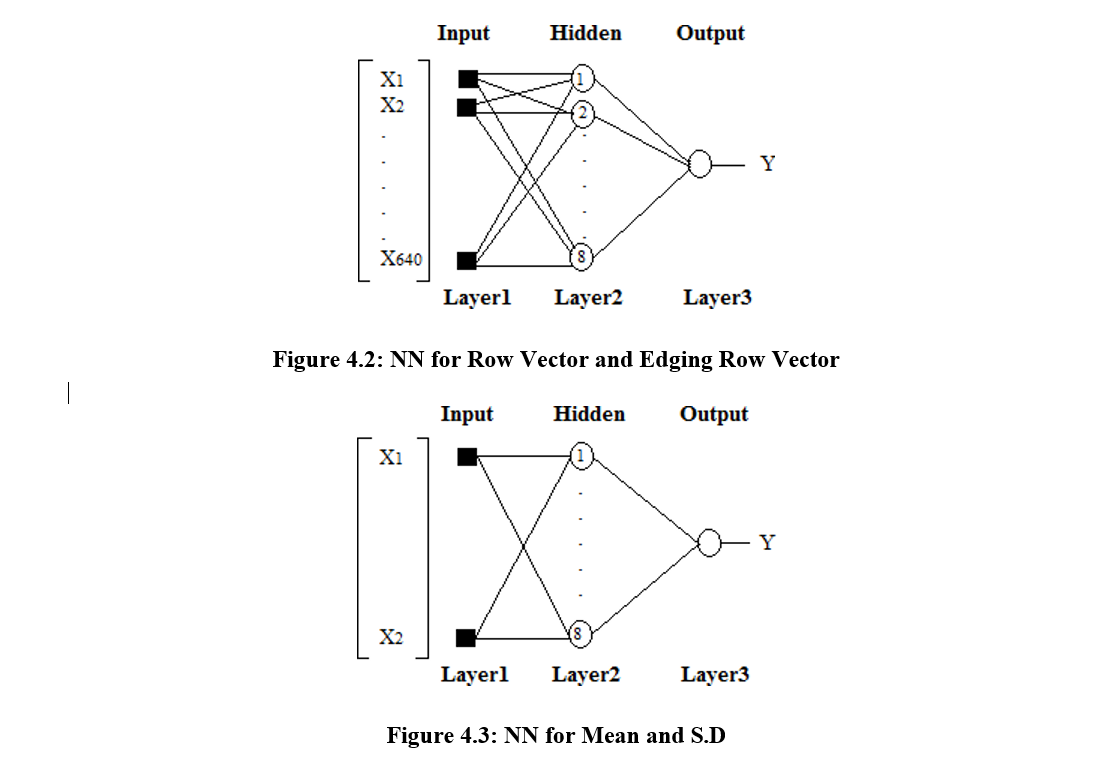
Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements.

Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. Such a situation is shown in Figure 4.1 below. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target. Typically, many such input/target pairs are used, in this supervised learning to train a network.

****

Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, and vision and control systems .Today neural networks can be trained to solve problems that are difficult for conventional computers or human beings.

Once data ready for representation then next step is to design NN for training and testing data. In first two algorithms Row Vector and Edging and Row Vector passing algorithm have three layers feed forward network: Input, Hidden and Output. Number of neurons in Input is 640 which are equal to number of features extracted from each of algorithm and one neuron for Output layer for skin class to be recognized. But for Mean and standard deviation there are only two input which is also equal to extracted features from this algorithm. Neural network Architecture has number of parameters such as learning rate (or), number of epochs and stopping criteria which is based on validation of data. Training of Mean Square Error at output layer which is set trial values and which is set by several experiments.

****

**CHAPTER 05**

**Results and Discussion**

The hand gesture recognition system has been tested with hand images under various conditions. The performance of the overall system with different algorithms is detailed in this chapter. Examples of accurate detection and cases that highlight limitations to the system are both presented, allowing an insight into the strengths and weaknesses of the designed system. Such insight into the limitations of the system is an indication of the direction and focus for future work.

System testing is actually a series of different tests whose primary purpose is to fully exercise the computer-based system. It helps us in uncovering errors that were made inadvertently as the system was designed and constructed. We began testing in the „small‟ and progressed to the „large‟. This means that early testing focused on algorithms with very small gesture set and we ultimately moved to a larger one with improved classification accuracy and larger gesture set.

**5.1 Experiments & Analysis**

Performed experiment shows the achieved result and estimate gesture recognition system projected in chapter 4. The experiment divided into two categories to better analyze system performance and capabilities. The more general approach to work with differently user independent system developed to interact with multi users with different kind of skin colors and hands shapes. It is very important approach to attempt for independent multi- user system. The system can be used by various users.

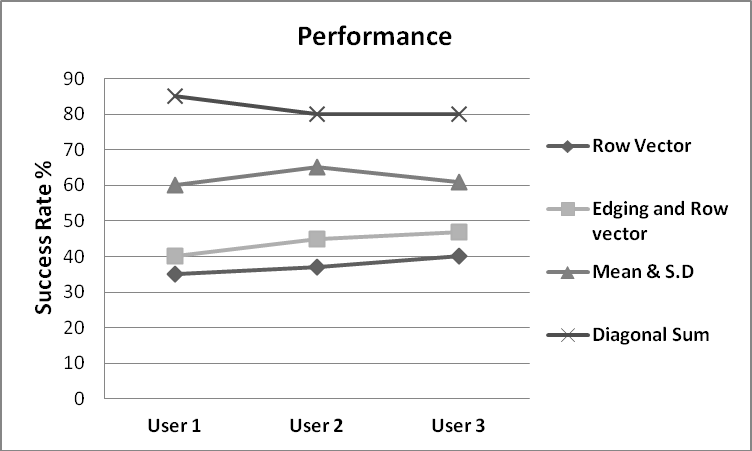
Two main aims for this work to detect hand and recognition of hand gesture with neural network and real classification. The first aim to detect hand with different skin tones, using explicitly defined skin region. Secondly gesture recognition with neural network and real classification by different algorithms. This system designed to test the hypothesis that detection and recognition rate would increase as:

* Hand detection with different skin tones
* More training pattern are used to train neural network
* Gesture recognition

**5.2 Gesture Recognition**

This experimental hypothesis was to recognize of gestures that user gave to system either with training with neural network and real time classification. The database to test hand gesture recognition system created 75 static images with different background. The static images size 640x480 pixels collected from camera. I grabbed gestures from different people and implement different methods of recognition with more than one user ultimately system has ability to recognize gestures of different peoples.

The accuracy percentage measured per hand gesture from different users. In **Figure 5.2** prove the effectiveness of all method implemented. Classification of each gesture can be seen from following results. In following results comparison that classification percentage varies with different methods.



**Figure 5.2: Classification Percentage**

**CHAPTER 06**

**Conclusion and Future Work**

**6.1 Conclusion**

This chapter summarizes my work at every stage of the project. At the time I started my thesis, I had a brief idea of how I will bring it from a topic on the paper to a real product. Due to knowledge of Computer Vision and Biometric subjects I had background in the image-processing field but not at expert level but my constant effort helped me to go through and succeed eventually.

As required in every project, research is of utmost importance. So, I spent the pretty much time in going through the background literature. I looked at various approaches of doing my thesis and developed four different methods: Row vector algorithm, Edging and row vector passing algorithm, Mean and standard deviation of edged image and Diagonal sum algorithm.

The first limitation that was discovered in all the algorithms used with neural networks was that their performance depended on the amount of training dataset provided. The system worked efficiently after being trained by a larger dataset as compared to a smaller dataset.

The Row vector algorithm used initially was a very vague approach adopted for classification as it was found through experiments that the row vectors of two different images could happen to be the same.

In the edging and row vector-passing algorithm, the edging parameter was introduced in addition to the row vector to improve the gesture classification accuracy but it was found that due to self-shadowing effect found in edges, the detection rate was not sufficiently improved.

The next parameters tried for classification were mean and standard deviation. They also failed to give satisfactory results (i.e. above 60%) but still they were among the best parameters used for detection with neural networks.

**6.2 Future Work**

The system could also be made smart to be trained for only one or two gestures rather than all and then made ready for testing. This will require only a few changes in the current interface code, which were not performed due to the shortage of time.

One time training constraint for real time system can be removed if the algorithm is made efficient to work with all skin types and light conditions which seems impossible by now altogether. Framing with COG (Centre of gravity) to control orientation factor could make this system more perfect for real application.

**References**

1. Robyn Ball and Philippe Tissot, Texas A&M University, **Demonstration of Artificial Neural Network in Matlab.**
2. Howard Demuth, Mark Beale, **Neural Network Toolbox**
3. Steven Daniel Lovell, February 2005, **A System for Real-Time Gesture Recognition and Classification of Coordinated Motion**
4. Ray Lockton, Balliol College, Oxford University, **Hand Gesture Recognition Using Computer Vision.**
5. Hatice Gunes, Massimo Piccardi, Tony Ja, 2007, Research School of Information Sciences and Engineering Australian National University **Face and Body Gesture Recognition for a Vision-Based Multimodal Analyzer**
6. J. Heinzm ann and A. Zelinsky **Robust Real - Time Face Tracking and Gesture Recognition**
7. Vladimir Vezhnevets, Vassili Sazonov, Alla Andreeva, Moscow State University **A Survey on Pixel-Based Skin Color Detection Techniques,**
8. TEY YI CHI, Universiti Teknologi Malaysi, **FUZZY SKIN DETECTION**
9. Robyn Ball and Philippe Tissot, Texas A&M University, **Demonstration of Artificial Neural Network in Matlab**
10. Howard Demuth, Mark Beale, **Neural Network Toolbox**
11. Ari Y. Benbasat and Joseph A. Paradiso in MIT Media Laboratory, Cambridge **An Inertial Measurement Framework for Gesture Recognition and Applications**
12. Peggy Gerardin, Color Imaging Course Winter Semester 2002-2003, **Color Image Segmentation**
13. Michael Nielsen, Moritz Störring, Thomas B. Moeslund, and Erik Granum*,* March 2003, **A procedure for developing intuitive and ergonomic gesture interfaces for man- machine interaction.**

**GITHUB REPOSITORY**

https://github.com/kalyanteja18/Hand-Gesture-Recognition-for-Dynamic-Applications