Department of Computer Science and Engineering,
NMAMIT, Nitte.

Project Report

on

Facial Emotion and Hand Gesture Recognition

Submitted to

NMAM INSTITUTE OF TECHNOLOGY, NITTE
(An Autonomous Institution under VTU, Belagavi)

In partial fulfillment of the requirements for the award of the

Degree of Bachelor of Engineering
in
Computer Science and Engineering

by

SHASHANK	**4NM17CS163**
SHRIKESH I	**4NM17CS177**
MADAN A R	**4NM17CS095**
KIRAN MAHADEV GIRADDI	**4NM17CS088**

Under the guidance of

Mr. RAJU K,
Assoc. Professor
Dept. of CSE, NMAMIT, NITTE

NITTE EDUCATION TRUST | NMAM INSTITUTE OF TECHNOLOGY

(ISO 9001:2015 Certified), Accredited with 'A' Grade by NAAC

CERTIFICATE

Certified that the project work entitled ***Facial Emotion and Hand Gesture Recognition*** *is a bonafide work carried out by* ***Shashank (4NM17CS163), Shrikesh I (4NM17CS177), Madan A R (4NM17CS095), Kiran Mahadev Giraddi(4NM17Cs088),*** *in partial fulfillment for the award of Degree of Bachelor of Engineering in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi during the year 2020-21. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project Phase- 1 prescribed for the said Degree.*

Name & Signature of Guide **Signature of HOD** **Signature of the Principal**

Mr Raju K, Assoc.Professor

External Viva

Name of the Examiners	**Signature with Date**
1. _____	_____
2. _____	_____

ACKNOWLEDGEMENT

The satisfaction that accompanies the completion of any task would be incomplete without the mention of all the people, without whom this endeavour would have been a difficult one to achieve. Their constant blessings, encouragement, guidance and suggestions have been a constant source of inspiration

First and foremost, my gratitude to my project guide ***Mr Raju K*** for the constant guidance throughout the course of this Project Phase-1 and for the valuable suggestions.

I also take this opportunity to express a deep sense of gratitude to the project coordinators for their valuable guidance and support.

My thanks to our beloved HOD, **Dr. Uday Kumar Reddy,** for extending support in carrying out this project in the department and providing us with all necessary facilities.

My sincere thanks to our beloved principal, **Dr. Niranjan N Chiplunkar** for permitting us to carry out this project at our college and providing us with all needed facilities.

Finally, thanks to staff members of the Department of Computer Science and Engineering and our friends for their honest opinions and suggestions throughout the course of our project Phase-1.

Shashank (4NM17CS163)
Shrikesh I (4NM17CS177)
Madan A R (4NM17CS095)
Kiran Mahadev Giraddi (4NM17CS088)

ABSTRACT

Modern world is changing in each pulse. New technologies are taking place in every sector of our day to day life. Image processing is one of the major pioneers in this changing world. With a single click many things are taking place. Many things are possible with the help of an image. Object detection and identification is a popular task in computer vision. Detecting emotions of a person from images or streams of images and identifying their hand gestures are an important application of object detection. These applications have huge potential in various fields.

In this project, we have implemented a method to identify a person's emotional state and recognise their hand gestures. We use Deep Conventional Neural Networks to implement this.

TABLE OF CONTENTS

Contents	Page
Title Page	i
Certificate	ii
Acknowledgment	iii
Abstract	iv
Table of Contents	v-vi
Table of Figures	vii

CHAPTER NO.	CHAPTER NAME	PAGE NO.
1	**INTRODUCTION**	**1-4**
	Overview	2
	Problem Statement	2
	Study Area	2-3
	Objective	3
	Motivation	4
	Organization of the Report	4
2	**LITERATURE SURVEY**	**5-9**
	2.1 Existing System	5-8
	2.2 Proposed System	9
3	**SYSTEM ANALYSIS AND REQUIREMENTS**	**10-12**
	3.1 System Analysis	10
	3.1.1 Relevance of platform	10
	3.1.2 Relevance of programming language	10
	3.2 Requirement Analysis	11
	3.2.1 Scope and Boundary	11
	3.3 Functional requirements	11
	3.3.1 Software Requirements	11

3.3.2 Hardware Requirements 11
3.4 Non Functional requirements 11-12

4 SOFTWARE APPROACH 13-14

4.1 About Python 13
4.2 About Anaconda 13
4.3 About OpenCV 13-14
4.4 About Tensorflow 14

5 SYSTEM DESIGN 15-18

5.1 High Level Design Architecture 15-16
5.2 Low Level Design Architecture 16-18
5.2.1 Sequence Diagram 16-17
5.2.2 Use Case Diagram 17-18

*

LIST OF FIGURES

Figure no.	Description	Page No.
1	Face Registration	5
2	Seven Basic Emotions	6
3	Background Subtraction	7
4	CNN Architecture	16
5	Sequence Diagram	17
6	Use Case Diagram	18

CHAPTER 1

INTRODUCTION

Object recognition is to describe a collection of related computer vision tasks that involve activities like identifying objects in digital photographs. Image classification involves activities such as predicting the class of one object in an image. Object localization refers to identifying the location of one or more objects in an image and drawing a bounding box around their extent. Object detection does the work of combining these two tasks and localizes and classifies one or more objects in an image. When a user or practitioner refers to the term “object recognition“, they often mean “object detection“. It may be challenging for beginners to distinguish between different related computer vision tasks.

Facial expressions play a key role for understanding and detecting emotion. Even the term “interface” suggests how important face plays in communication between two entities. Studies have shown that reading of facial expressions can significantly alter the interpretation of what is spoken as well as control the flow of a conversation. The ability for humans to interpret emotions is very important to effective communication; accounting for up to 93% of communication used in a normal conversation depends on emotion of an entity. For ideal human-computer interfaces (HCI), we would desire that machines have the capability to read human emotion.

Gesture recognition has been a very interesting problem in the Computer Vision community for a long time. This is particularly due to the fact that segmentation of foreground objects from a cluttered background is a challenging problem in real-time. The most obvious reason is because of the semantic gap involved when a human looks at an image and a computer looking at the same image. Humans can easily figure out what’s in an image but for a computer, images are just 3-dimensional matrices.

1.1 OVERVIEW

Face detection and gesture recognition has been around for ages. Taking a step forward, human emotion displayed by face and felt by brain, captured in either video, electric signal (EEG) or image form can be approximated. Human emotion detection is the need of the hour so that modern artificial intelligent systems can emulate and gauge reactions from face. This can be helpful to make informed decisions be it regarding identification of intent, promotion of offers or security related threats. Recognizing emotions from images or video is a trivial task for the human eye, but proves to be very challenging for machines and requires many image processing techniques for feature extraction. Several machine learning algorithms are suitable for this job.

1.2 PROBLEM STATEMENT

We are going to recognize facial emotion and hand gestures from a video sequence. To recognize hand gestures from a live video sequence, we first need to take out the hand region alone removing all the unwanted portions in the video sequence.There are six basic universal emotions for human beings. These are happy, sad, angry, fear, disgust and surprise. From a human's facial expression we can easily detect this emotion. In this project we implement a useful way to detect these basic universal emotions along with neutral expression from frontal facial emotion. Our aim is to develop a method of face mood and hand gesture detection that is fast, robust, reasonably simple and accurate with a relatively simple and easy to understand algorithm.

1.3 STUDY AREA

Image Processing is a useful method for performing different operations on an image to get a better image or to get some useful information from it. Normally image processing methods consider an image as a two dimensional

signal. Because of this usefulness of image processing, in our research we are dealing with this method. Mainly the project aim is to detect a human's facial expression by applying image processing techniques. Gesture recognition can be used to design a simple computer vision application that can detect and recognize simple hand gestures for robot navigation control based on simple heuristic rules.

1.4 OBJECTIVE

In this project, we have developed convolutional neural networks for a facial expression recognition and gesture detection task. The goal is to classify each facial image into one of the seven facial emotion categories and recognise most common hand gestures such as counting of fingers,thumbs up , thumbs down etc. We trained CNN models with input data consisting of various images. We developed our models and exploited Graphics Processing Unit computation in order to expedite the training process. The input is taken directly from the webcam of the computer, and the person is detected, classified based on their action. The Objectives of the project are

- To develop convolutional neural networks for facial expression recognition and hand gesture detection.
- To train convolutional neural networks on the input data.
- To classify each facial image into one of the seven facial emotion categories.
- To recognise most common hand gestures such as counting of fingers, thumbs up , thumbs down etc

1.5 MOTIVATION

Facial expression recognition or computer-based facial expression recognition system is important because of its ability to mimic human coding skills. Facial expressions and other gestures convey nonverbal communication cues that play an important role in interpersonal relations. Similarly A hand gesture recognition system provides a natural, innovative and modern way of non verbal communication. It has a wide area of application in human computer interaction and sign language.

1.6 ORGANISATION OF THE CHAPTERS

The project report has been organized under nine chapters, which are as follows:

Chapter I: Introduces to the main idea of the project. It gives a brief knowledge about the aim and methodology of the same.

Chapter II: It includes literature survey of related works.

Chapter III: Discusses the system requirements that are needed for the project. These include functional requirements, non-functional requirements, user requirements and hardware requirements.

Chapter IV: Includes the system design details which includes flowchart, sequence diagram.

Chapter V: Includes the implementation details of the project, application is explained in detail. It also deals with software approaches.

CHAPTER 2

LITERATURE SURVEY

2.1 EXISTING SYSTEM:

As per various literature surveys it is found that for implementing Facial Emotion recognition four basic steps are required to be performed.

1. Preprocessing
2. Face registration
3. Facial feature extraction
4. Emotion classification

Preprocessing: Preprocessing is a common name for operations with images at the lowest level of abstraction both input and output are intensity images. Most preprocessing steps that are implemented are – reducing the noise, converting the image To binary/grayscale and geometric transformation

Face Registration: Face Registration is a computer technology being used in a variety of applications that identifies human faces in digital images. In this face registration step, faces are first located in the image using some set of landmark points called “face localization” or “face detection”.

Fig 1.Face Registration

Facial Feature Extraction: Facial Features extraction is an important step in face recognition and is defined as the process of locating specific regions, points, landmarks, or curves/contours in a given 2-D image or a 3D range image. In this feature extraction step, a numerical feature vector is generated from the resulting registered image. Common features that can be extracted are lips ,eyes , eyebrows and nose tip.

Emotion Classification: In the final step of classification, the algorithm attempts to classify the given faces portraying one of the seven basic emotions.

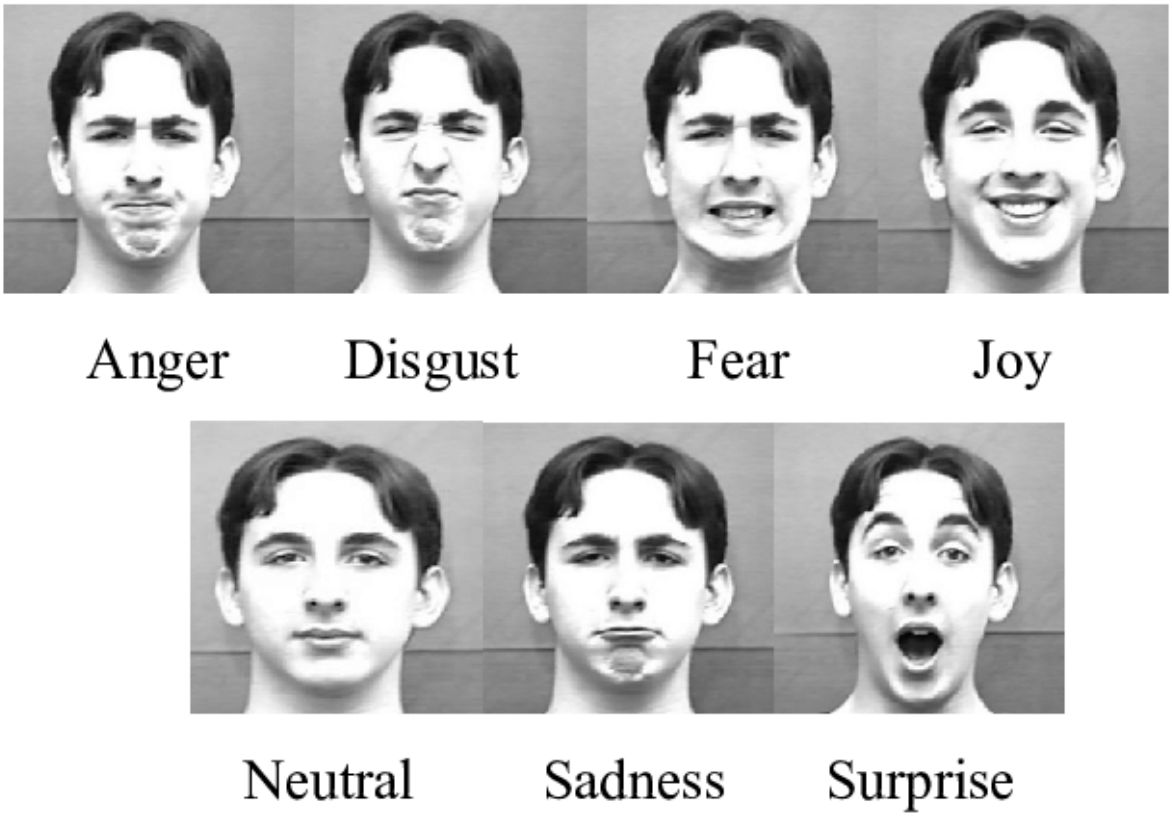

Fig 2.Seven Basic Emotions

Implementing hand gesture detection and recognition from a video sequence involves three simple steps.

1. Background Subtraction
2. Motion Detection and Thresholding
3. Contour Extraction

Background Subtraction: Background Subtraction is an efficient method to separate foreground from background. To do this, we use the concept of running averages. We make our system look over a particular scene for 30 frames. During this period, we compute the running average over the current frame and the previous frames.

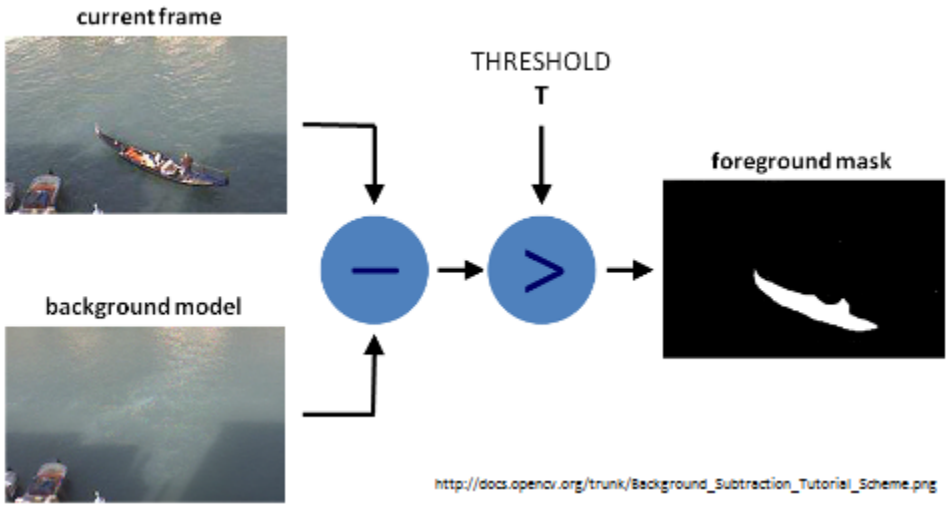

Fig 3.Background Subtraction

Motion Detection and Thresholding: To detect the hand region from this difference image, we need to threshold the difference image, so that only our hand region becomes visible and all the other unwanted regions are painted as black. Thresholding is the assignment of pixel intensities to 0's and 1's based on a particular threshold level so that our object of interest alone is captured from an image.

Contour Extraction: After thresholding the difference image, we find contours in the resulting image. The contour with the largest area is assumed to be our hand. Contour is the outline or boundary of an object located in an image.

Different approaches which are followed for Facial Expression and Hand Gesture Recognition:

Neural Network Approach: The neural network contained a hidden layer with neurons. The approach is based on the assumption that a neutral image corresponding to each image is available to the system. Each neural network is trained independently with the use of on-line back propagation.

Principal of Component Analysis : Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called Principal Component.

Support Vector Machine: In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary model (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a presentation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

2.2 PROPOSED SYSTEM:

This project aims to recognise the facial expression and classify into one of seven categories and identify commonly used hand gestures using convolutional neural networks. Facial expressions and other gestures convey nonverbal communication cues that play an important role in interpersonal relations. The model is trained on the FER-2013 dataset which was .This dataset consists of 35887 grayscale, 48x48 sized face images with seven emotions - angry, disgusted, fearful, happy, neutral, sad and surprised.

Our aim is to develop a method of facial expression and hand gesture detection that is reasonably simple and accurate with a relatively simple and easy to understand algorithm. User is able to provide the input from the webcam and the system based on its trained data is able to correctly predict the facial emotion and hand gesture.

CHAPTER 3

SYSTEM ANALYSIS AND REQUIREMENTS

3.1 SYSTEM ANALYSIS

3.1.1 Relevance of Platform

OpenCV is a cross-platform library using which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection. The library is cross-platform and free for use under the open-source Apache 2 License. Starting with 2011, OpenCV features GPU acceleration for real-time operations

3.1.2 Relevance of Programming Language

Python is a powerful and object-oriented high-level programming language. It has very simple easy-to-use syntax. It works on cross-platform operating systems and can be used across to develop a wide range of applications including those intended for image processing, text processing, web, and enterprise level using scientific, numeric and data from the network.

3.2 REQUIREMENT ANALYSIS

3.2.1 Scope and Boundary

In this project facial expression and hand gesture recognition system is implemented using convolution neural networks. Facial images are classified into seven facial expression categories namely Anger, Disgust, Fear, Happy, Sad, Surprise and Neutral. FER-2013 Dataset is used to train and test the classifier.

3.3 FUNCTIONAL REQUIREMENTS

3.3.1 Software Requirements:

- Software: 1. Python Environment
  2. Notepad++
  3. Anaconda
  4. OpenCV
  5. Tensorflow

3.3.2 Hardware Requirements:

- Operating system: Windows 7 and above.
- RAM: 4GB and above
- Processor: Intel® Core(TM)2 duo CPU T6500
- Processor speed: 2.67 GHz.
- CPU: 64-bit operating system.

3.4 NON FUNCTIONAL REQUIREMENTS

In systems engineering and requirements engineering, a non-functional requirement (NFR) is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviours. Non-functional requirements are conditions under which the system must be able to function and the quality the system must have. It defines how a system is supposed to be

- ➢ Performance
  - With ideal condition output of code should be fast and error free.
  - Code performance shall not decrease with time or by usage

- ➢ Flexibility
  - Project code will be easy to learn and use.
  - Is able to analyze and give the output as quickly as possible
- ➢ User-Friendly
  - The users should be able to navigate the project code easily.
  - The project code should be self explanatory.
- ➢ Response Time
  - The output should be quick without consuming much buffer time.
- ➢ Understandability
  - All users can access the output because of its simplicity.

.

CHAPTER 4

SOFTWARE APPROACH

4.1 ABOUT PYTHON

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python was designed to be highly readable which uses English keywords frequently whereas other languages use punctuation and it has fewer syntactic constructions than other languages. Python is a general-purpose language, which means it can be used to build just about anything, which will be made easy with the right tools/libraries. Professionally, Python is great for backend web development, data analysis, artificial intelligence, and scientific computing. Many developers have also used Python to build productivity tools, games, and desktop apps, so there are plenty of resources to help you learn how to do those as well.

4.2 ANACONDA

It is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications (large-scale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 6 million users, and it includes more than 250 popular data science packages suitable for Windows, Linux, and MacOS.

4.3 OPENCV

OpenCV (Open Source Computer Vision Library) is an open source computer vision and machine learning software library. OpenCV was built to provide a

common infrastructure for computer vision applications and to accelerate the use of machine perception in commercial products. Being a BSD-licensed product, OpenCV makes it easy for businesses to utilize and modify the code. The library has more than 2500 optimized algorithms, which includes a comprehensive set of both classic and state-of-the-art computer vision and machine learning algorithms. OpenCV has more than 47 thousand people in the user community and an estimated number of downloads exceeding 18 million. The library is used extensively in companies, research groups and by governmental bodies.

4.4 TENSORFLOW

TensorFlow is a free and open-source software library for machine learning. It can be used across a range of tasks but has a particular focus on training and inference of deep neural networks. Tensorflow is a symbolic math library based on dataflow and differentiable programming. It is used for both research and production at Google. TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache License 2.0 in 2015.It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

CHAPTER 5

SYSTEM DESIGN

5.1 HIGH LEVEL DESIGN ARCHITECTURE

General CNN architecture With preprocessing, the input layer is predefined to a fixed size, which can then be entered into the next layer. For this step, to detect the faces in each image, the project used OpenCV, a popular computer vision library that includes pre-trained filters. This step helps to significantly reduce the size. The input layer can then be transferred to the Convolution2D layer, where the number of filters is specified as a super-parameter. A set of filters, such as the kernel, is randomly generated weights. Each filter, such as a sliding window, traverses the entire image to construct a feature graph with shared weights. The convolution layer generates a feature map showing how the pixel values are raised, such as edge, light, and pattern detection. Applying pooling to help reduce the dimension behind the convolution layer is a relatively important process in building a generic CNN architecture, as adding more convolution layers can increase computational costs. The project uses the MaxPooling2D method, a popular pooling technique that uses a 2x2 window that traverses feature maps, retaining only the maximum value of pixels. When pixels are merged, the size of the image is reduced by 4. For the output layer, the project uses softmax instead of sigmoid as the activation function. The layer outputs the probability of each facial expression class. In this way, the CNN model can provide the probability of each emotion, and select the emotion with the highest predictive score as the recognition result. Figure below shows the CNN architecture that was ultimately designed to detect faces and recognize specific facial expressions for each face.

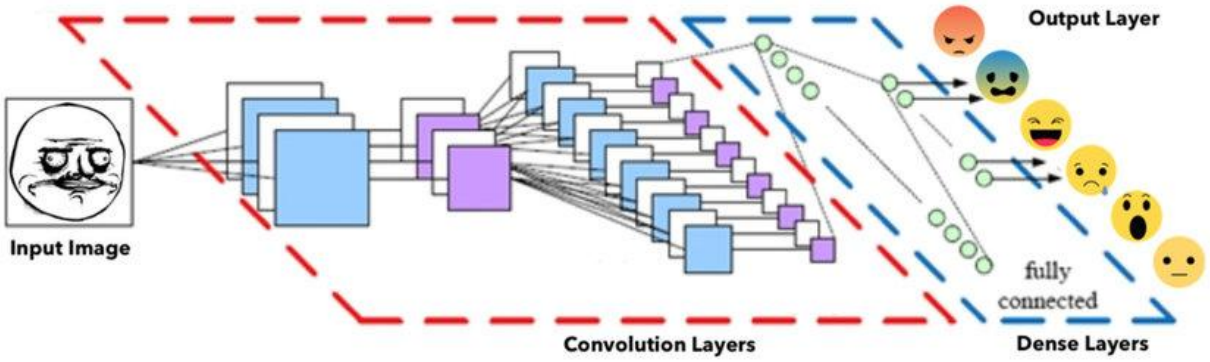

Fig 4.CNN Architecture

5.2 LOW LEVEL DESIGN ARCHITECTURE

5.2.1 Sequence Diagram

A sequence diagram shows object interaction arranged in time sequence. It describes interactions among classes in terms of an exchange of messages over time. It is also called an event diagram. A sequence diagram is a good way to visualize and validate various run time scenarios. These can help to predict how a system will behave and to discover responsibilities a class may need to have in the process of modelling the new system. Messages are arrows that represent communication between the objects. Lifelines are vertical dashed lines that indicate the object presence over time.

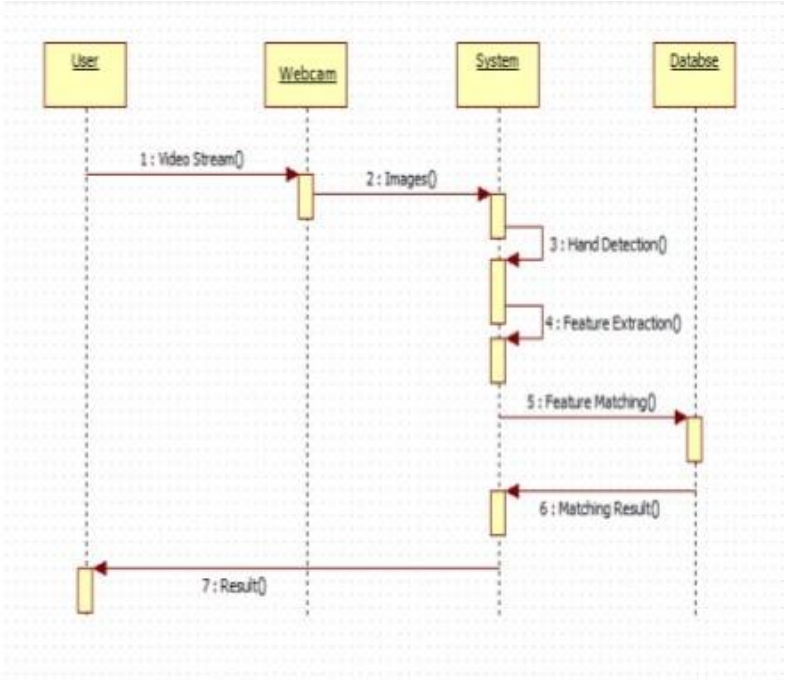

Fig 5.Sequence diagram

5.2.2 Use Case Diagram

A use case diagram is a dynamic or behavior diagram in UML. Use case diagrams model the functionality of a system using actors and use cases. Use cases are a set of actions, services, and functions that the system needs to perform.

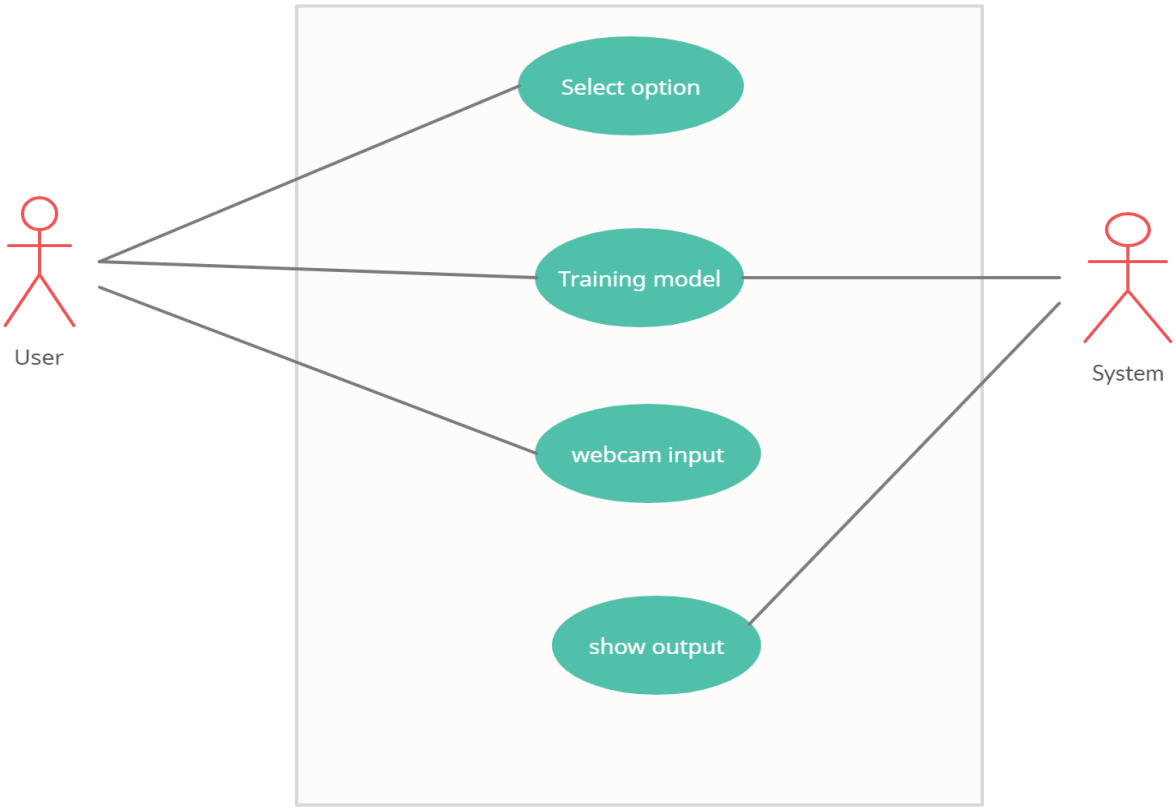

Fig 1.Face Registration

CHAPTER 6

SYSTEM IMPLEMENTATION

CHAPTER 7

SYSTEM TESTING

CHAPTER 8

RESULTS AND DISCUSSION

CHAPTER 9

CONCLUSION AND FUTURE WORK