Hand and Head Gesture Tracking and Recognition

Altamash Kalania[1] Kamal Siddiqui[2] Munishwaran Chelliyah[3]
[1,2,3]Department of Computer Engineering
[1,2,3]Atharva College of Engineering, Mumbai University, Mumbai, India

Abstract— Hand & Head gesture recognition system received great attention in the recent few years because of its manifoldness applications and the ability to interact with machine efficiently through human computer interaction. In this paper a survey of recent hand & Head gesture recognition systems is presented. Key issues of hand & Head gesture recognition system are presented with challenges of gesture system. Review methods of recent postures and gestures recognition system presented as well. Summary of research results of hand & Head gesture methods, databases, and comparison between main gesture recognition phases are also given. Advantages and drawbacks of the discussed systems are explained finally.

Key words: Head Gesture, Recognition

I. INTRODUCTION

The essential aim of building hand and Head gesture recognition system is to create a natural interaction between human and computer where the recognized gestures can be used for controlling a robot or conveying meaningful information . How to form the resulted hand gestures to be understood and well interpreted by the computer considered as the problem of gesture interaction .

Human computer interaction (HCI) also named Man-Machine Interaction (MMI) refers to the relation between the human and the computer or more precisely the machine, and since the machine is insignificant without suitable utilize by the human . There are two main characteristics should be deemed when designing a HCI system as mentioned in : functionality and usability. System functionality referred to the set of functions or services that the system equips to the users , while system usability referred to the level and scope that the system can operate and perform specific user purposes efficiently .

Gestures can be static (posture or certain pose) which require less computational complexity or dynamic (sequence of postures) which are more complex but suitable for real time environments . Different methods have been proposed for acquiring information necessary for recognition gestures system . Some methods used additional hardware devices such as data glove devices and color markers to easily extract comprehensive description of gesture features . Other methods based on the appearance of the hand using the skin color to segment the hand and extract necessary features , these methods considered easy, natural and less cost comparing with methods mentioned before.

II. LITERATURE REVIEW

A. Hand Gesture Detection:

A Hand Gesture Detection is a system that Detect the Hand Gesture with Meaningful Gesture, meaning that it does not detect anything it only detects some pre-defined Gestures.

B. Head Gesture Detection:

A system that Detect the Head Gesture with only few pre-defined Motion and taking Necessary action of that motion. This System contains further 2 parts which are :

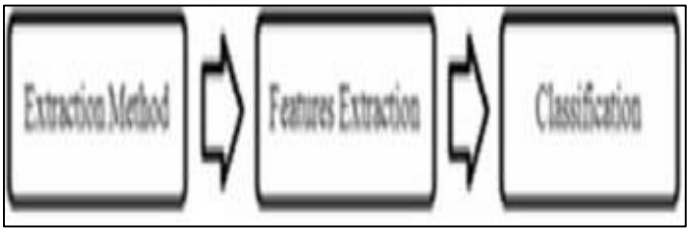

Fig. 1: Gesture Recognition System Steps

C. Features Extraction:

Good segmentation process leads to perfect features extraction process and the latter play an important role in a successful recognition process [6]. Features vector of the segmented image can be extracted in different ways according to particular application. Various methods have been applied for representing the features can be extracted. Some methods used the shape of the hand such as hand contour and silhouette while others utilized fingertips position, palm center, etc. created 13 parameters as a feature vector, the first parameters represents the ratio aspect of the bounding box of the hand and the rest 12 parameters are mean values of brightness pixels in the image. used Self-Growing and Self-Organized Neural Gas (SGONG) neural algorithm to capture the shape of the hand, then three features are obtained; Palm region, Palm center, and Hand slope.

calculated the Center Of Gravity (COG) of the segmented hand and the distance from the COG to the farthest point in the fingers, and extracted one binary signal (1D) to estimate the number of fingers in the hand region. divided the segmented image into different blocks size and each block represents the brightness measurements in the image.. [1]

D. Gestures Classification:

After modeling and analysis of the input hand image, gesture classification method is used to recognize the gesture. Recognition process affected with the proper selection of features parameters and suitable classification algorithm.[7] For example edge detection or contour operators cannot be used for gesture recognition since many hand postures are generated and could produce misclassification. Euclidean distance metric used to classify the gestures. [2]

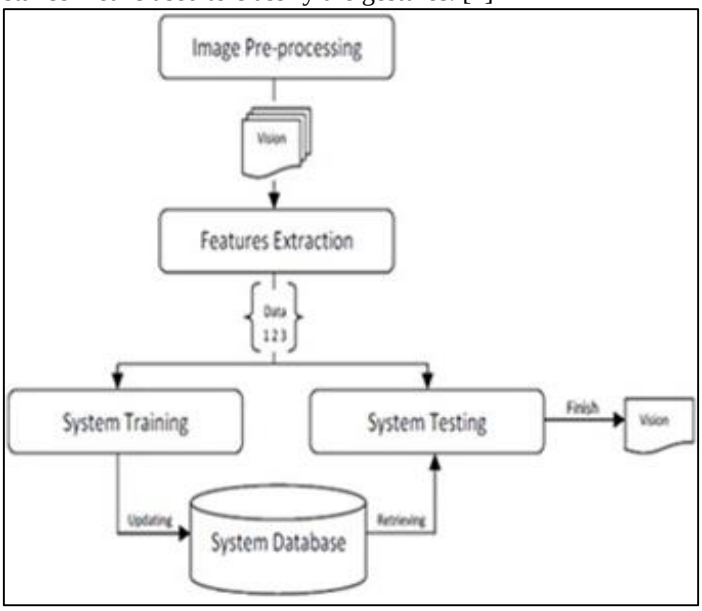

Fig. 2: Architecture of Gesture Recognition System

Neural network has been widely applied in the field of extracted the hand shape, and for hand gesture recognition. Other soft computing tools are effective in this field as well, such as Fuzzy C.

III. RELATED WORK

A. Face Detection:

Facial recognition technologies are used in a wide array of contexts, reflecting a spectrum of increasing technological sophistication in today's world. At the simplest level, the technology can be used for facial detection; that is, merely to detect and locate a face in a photo. Current uses of facial detection include refining search engine results to include only those results that contain a face; locating faces in images in order to blur them. A more refined version of facial recognition technology allows assessing characteristics of facial images. Face recognition has always been a very challenging task for the researches. [3]

On the other hand, it has always been very difficult to implement due to all different situation that a human face can be found. Due to the difficulty of the face recognition task, the number of techniques is large and diverse. It is not think that images are always capture in ideal conditions, there may be illumination, pose, and expression variation. Such challenges are more prominent in heterogeneous face recognition. In last decades there were many method developed to tackle such problem. From face recognition surveys it implies that they have face recognition of face images which are of same type. This restricts the face recognition for specific datatype. Such situation can be tackle by using face images of different modality, it refer as heterogeneous faces.[5] When designing a face detection and face recognition system, in addition to considering the aspects from psychophysics and neuroscience and the factors of human appearance variations, there are still some design issues to be taken into account. First, the execution speed of the system reveals the possibility of on-line service and the ability to handle large amounts of data. Heterogeneous face recognition using kernel methods concentrate their works on how to speed-up the existing algorithms and how to handle large amounts of data simultaneously. Second, the training data size is another important issue in algorithm design. It is trivial that more data are included, more information we can exploit and better performance we can achieve.[4]

B. Facial Expression Detection:

Most Facial Expression Recognition and Analysis systems proposed in the literature focus on the binary occurrence of expressions, often either basic emotions or FACS Action Units (AUs). In reality, expressions can vary greatly in intensity, and this intensity is often a strong cue for the interpretation of the meaning of expressions. In addition, despite efforts towards evaluations standards (e.g. FERA 2011), there still is a need for more standardised evaluation procedures. They therefore suffer from low comparability. This is in stark contrast with more established problems in human behaviour analysis from video such

as face detection and face recognition. Yet at the same time, this is a rapidly growing field of research, due to the constantly increasing interest in applications for human behaviour analysis, and technologies for human-machine communication and multimedia retrieval.[3]

In these respects, the FG 2015 Facial Expression Recognition and Analysis challenge (FERA2015) shall help raise the bar for expression recognition by challenging participants to estimate AU intensity, and it will continue to bridge the gap between excellent research on facial expression recognition and low comparability of results. We do this by means of three selected tasks: the detection of FACS Action Unit (AU) occurrence, the estimation of AU intensity for pre-segmented data (i.e. when AU occurrence is known), and fully automatic AU intensity estimation (i.e. for every frame when AU occurrence is not known).[2]

IV. AIM AND OBJECTIVE

A. Aim:

Even after So many Advancement in the Technology user can only interact with keyboard and mouse with PC. Hence this led to the development of a system that provides much more convenient Human Computer Interface System.

B. Objective:

The main objective of this project is to provide flawless and seamless interaction with Computers with as much Friendly Interface as Possible. Allowing the user to do Daily Task!

User can interact with Computer using Head and Hand Gesture to perform certain common task and also Voice as Additional support for better Interaction and also as an alternative.

V. PROPOSED SYSTEM

The main Objective of our project is to save a ton of time for user by providing easier and faster interface with computer. The user can use microphone and webcam as input method instead of keyboard and mouse.

By using Computer Vision and Image Processing together we can build a smart system that does not exist. The system can understand the gesture of the user and does different operations accordingly.

Firstly, Using Computer Vision Techniques the system will take multiple images of the actions made by the user and then using Image Processing techniques the system will then perform "Addition and Subtraction Algorithm" and then analyze the Gesture Meaning and Finally Executing the Job the user want him to do. [5]

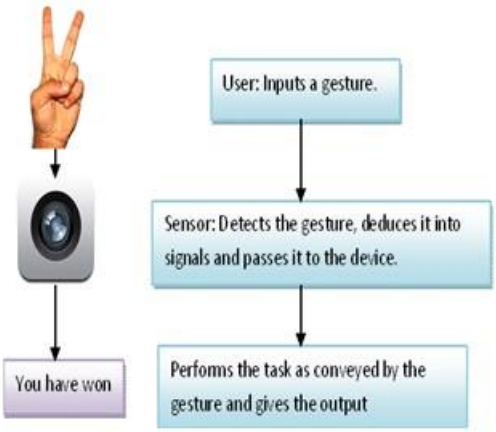

Fig. 3: Basic Overview

A. Data Collection & Analysis:

In order to understand the Gestures the first and foremost thing is the Data that the System need to perform operation. So the System saves the images taken from the webcam every frame and stores them in a List.
System will save Images in either Data Structure:

1) Linked List.
2) Arrays.

In First stage the system will store images using any Data Structure stated above. Then these stored images are analyzed using Algorithms.[2]

B. Binary Image Extraction:

The Images that are stored are then subtracted with all the other images it has in the List and then it converts to Grayscale Image. A Grayscale image is the image that contains grey, black and white pixels. So the System converts the Image to Grayscale image.Binary Image Extractor is the process in the system that converts this Grayscale image to Binary Image. It

only takes Grayscale image as the input so this is the reason we convert the normal image to Grayscale image so that finally we can convert to Binary Image.[1]

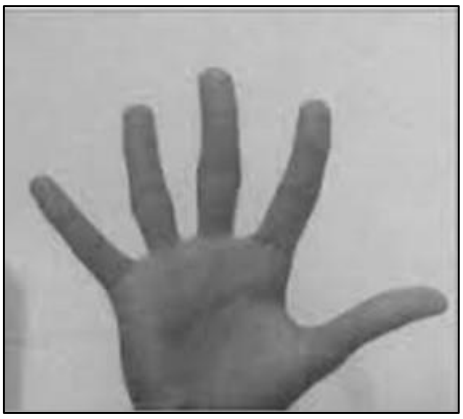

Fig. 4: Grayscale Image

Fig. 5: Binary Image

C. Recognition Engine:

Recognition Engine is the Heart of the System that Recognizes the Gestures. This engine has to be efficient and smart in order to process Detect the Gestures. The Recognition Engine take Arrays of Binary Image and then it Analyze them using "Subtraction Algorithm".[3]

Finally after applying the algorithm it gets the actual meaning of the Gesture and then it checks the Database for any Pre-Defined Gesture. If found then it Acknowledges the Expert System which is Responsible for Carrying out Actual operation.

D. Haar Classifier:

First, a classifier is trained with a few hundred sample views of a particular object (i.e., a face or a car), called positive examples, that are scaled to the same size (say, 20x20), and negative examples - arbitrary images of the same size.

After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a "1" if the region is likely to show the object (i.e., face/car), and "0" otherwise. To search for the object in the whole image one can move the search window across the image and check every location using the classifier.[7]

E. Methodology:

This section describes the overall methodology adopted in the project. The goal of our project is to use Various Algorithms for Detecting Hand and Head of the User even in the Dark. This Detection is done suing Webcam of at least 1.0 Mega Pixel. Finally the Algorithm uses the best possible solution to find and track the hand/head of the user even in the dark room.[1]

We proposed a robust real-time hand gesture recognition method. In our method, firstly, a specific gesture is required to trigger the hand detection followed by tracking; then hand is segmented using motion and color cues; finally, in order to break the limitation of aspect ratio encountered in most of learning based hand gesture methods, the scale-space feature detection is integrated into gesture recognition. Applying the proposed method to navigation of image browsing, experimental results show that our method achieves satisfactory performance.

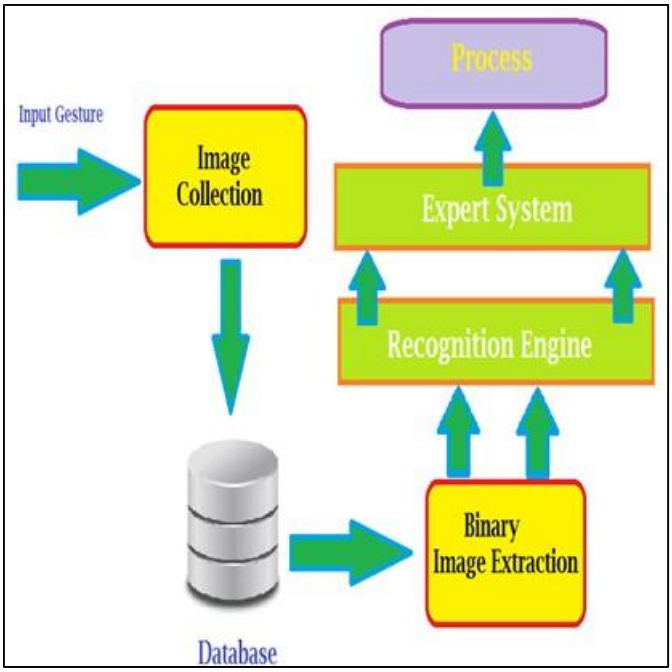

Fig. 6: System Overview of Gesture Detection Methodology

VI. SCOPE

Gestures have been an integral part of human interaction since old times. Most gestures originate from either face or hand, but they can actually be done by any part of the body. Automobile drivers have depended on gestures to maneuver through traffic previously – from waving hands during turns to shaking fists after overtaking.

New technology is enabling automobile manufacturers to integrate gesture recognition features in their cars to let drivers manage the control systems of the car. For example, an approaching hand can activate the in-car infotainment system, or tilting your head can switch on the turning indicator.

Driver distraction is a major cause of concern for maintaining safety on the road. Taking your hands off the steering wheel and eyes off the road to adjust air conditioning or making a call can have dire consequences. Gesture based car controlling systems would enable drivers to do all that and even more without even looking at the dashboard.

Gesture and motion controls are not a new phenomenon. Consumers have tasted it in other aspects of their lives, whether it is motion-activated burglary systems or playing games on their favourite video gaming console such as Nintendo Wii or Xbox 360.[4]

VII. IMPLEMENTATION PLAN

A. Software Requirements:

1) Visual Studio 2015:

Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for Microsoft Windows, as well as Websites, Web Applications and Web Services. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silver light. It can produce both native code and managed code.

2) Windows OS:

Microsoft Windows (or simply Windows) is a metafamily of graphical operating systems developed, marketed, and sold by Microsoft. It consists of several families of operating systems, each of which cater to a certain sector of the computing industry. Active Windows families include Windows NT, Windows Embedded and Windows Phone; these may encompass subfamilies, e.g. Windows Embedded Compact (Windows CE) or Windows Server. Defunct Windows families include Windows9x and Windows Mobile.

3) .NET Framework

.NET Framework (pronounced dot net) is a software framework developed by Microsoft that runs primarily on Microsoft Windows. It includes a large class library known as Framework Class Library (FCL) and provides language interoperability (each language can use code written in other languages) across several programming languages. Programs written for .NET Framework execute in a software environment (as contrasted to hardware environment), known as Common Language Runtime (CLR),an application virtual machine that provides services such as security, memory management, and exception handling. FCL and CLR together constitute .NET Framework. .Net Framework 4.0

4) OpenCV Library

OpenCV (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision, originally developed by Intel research center in Nizhny Novgorod (Russia), later supported by Willow Garage and now maintained by Itseez. The library is cross-platform and free for use under the open-source BSD license.

B. Hardware Requirements:

Processor	:	Intel Processor IV or above.
RAM	:	512 MB or more.
Hard disk	:	40 GB or more.
Microphone	:	Internal or External.
WebCam	:	1.0 Mega Pixel or More.
Speaker (Optional)	:	Internal or External.

ACKNOWLEDGEMENT

We would like to thank our project guide Prof. Rajesh Gaikwad for his enormous co-operation and guidance. We have no words to express our gratitude for a person who wholeheartedly supported the project and gave freely of his valuable time while making this project. All the inputs given by him have found a place in the project. The technical guidance provided by him was more than useful and made the project successful. He has always been a source of inspiration for us. It was a memorable experience learning under such a highly innovative, enthusiastic and hard working teacher.

We are also thankful to our Principal Dr. S.P. Kallurkar,our HOD Prof. Mahendra Patil, our Project co-ordinator Prof. Deepali Maste and all the staff members of the Computers department who have provided us various facilities and guided us to develop a very good project idea.

REFERENCES

[1] G. R. S. Murthy, R. S. Jadon. (2009). "A Review of Vision Based Hand Gestures Recognition," International Journal of Information Technology and Knowledge Management, vol. 2(2), pp. 405410.
[2] P. Garg, N. Aggarwal and S. Sofat. (2009). "Vision Based Hand Gesture Recognition," World Academy of Science, Engineering and Technology, Vol. 49, pp. 972-977.
[3] Fakhreddine Karray, Milad Alemzadeh, Jamil Abou Saleh, Mo Nours Arab, (2008) ."HumanComputer Interaction: Overview on State of the Art", International Journal on Smart Sensing and Intelligent Systems, Vol. 1(1).
[4] Wikipedia Website.
[5] Mokhtar M. Hasan, Pramoud K. Misra, (2011). "Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization", International Journal of Computer Science & Information Technology (IJCSIT), Vol. 3(2).
[6] Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.
[7] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems, Man and Cybernetics.