**CHAPTER 1**

**ABSTRACT**

**Gesture recognition** is a topic in computer science with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Gesture recognition enables humans to communicate with the machine and interact naturally without any mechanical devices. Using the concept of gesture recognition, it is possible to point a finger at the computer screen so that the cursor will move accordingly. This could make conventional input devices such as mouse, keyboards and even touch screens redundant.

In this project, we design a real-time human computer interaction system based on hand gesture. Hand gesture recognition system provides us a novel, natural, innovative user friendly way of communication with the computers. The whole system consists of three components: **Hand Detection, Gesture Recognition and Human-Computer Interaction (HCI)** based on recognition, and realizes the robust control of mouse and keyboard events with a higher accuracy of gesture recognition.

**CHAPTER 2**

**INTRODUCTION**

Communication in daily life is performed via the help of vocal sounds and body language. However vocal sounds are the main tool for interaction, body language and facial expressions have a serious support in the meanwhile. Even in some cases, interacting with the physical world by using those expressive movements instead of speaking is much easier. Body language has wide range of activities namely eye expressions, slight change in skin color, variation of the vibrations in vocal sounds etc. But the most important body language expressions are performed using hands.

Hand gesture recognition is identifying expressions which are meaningful of the hand or hand in motion. Hand gestures are used in our daily life as a nonverbal method of communication. Gestures are actions that contain some meaningful messages. There are number of systems used in input system for computer interaction such as mouse, keyboard. But in early years there were number of techniques for gesture recognition on tracking such as instrumental glows and optical markers etc. But the most efficient way is to identify hand gestures via webcam. This paper introduces hand gestures recognition system which uses hand gestures for computer interaction. We are using OpenCV image processing and C++ for pre-processing, hand detection, hand tracking and event triggering.

**OPEN CV**: Open CV (open source computer version library) is a library which explicitly focuses on computer version and real time CV. It is available free for everyone. It is compatible with C++, Python, Java and supports Windows, Linux, Mac OS and Android. Open CV was designed for relative applications. It has above 2000 optimized algorithms and comprehensive set of art CV and ML algorithms.

**CHAPTER 3**

**PROBLEM STATEMENT**

* A hand gesture recognition system was developed to capture the hand gesture being performed by the user and to control a computer system by that incoming information.
* Many of such systems in literature have strict constraints like wearing special gloves, having uniform background, long-sleeved user arm, being in certain lightning conditions, etc.
* This system eliminates some of the above constraints by just detecting the natural movement of the hand using a web camera and enabling the system to correctly determine the gesture that is being made and to perform the desired action.
* The scenario that will be used in this project will be the reading of a book and performing certain actions such as switching of pages, scrolling in a page, zooming in or zooming out of pages by recognizing hand gestures.

**CHAPTER 4**

**OBJECTIVES OF THE PROPOSED PROJECT**

Coming across the various spheres of life where this wearable technology can be applied, our major motivation for this project came from one such need. Among us, there are some people who do not have the voice to express, and sometimes, the capability to hear. The deaf and dumb sign language was developed to aid the communication for such people. Our idea was to create a device which could help them communicate, to give them a voice.

The proposed system is used to detect the hand movements of a person and performing certain actions such as switching of pages, scrolling up or down in a page, zooming in or zooming out of pages. This is the main objective of our system.

The hand detection is done using Machine Learning techniques. A dataset with various images of the human hand is used. A webcam is used to capture the images of the hand and compare it with the dataset in order to recognize the valid hand movements that are needed to perform the required actions.

**CHAPTER 5**

**LITERATURE REVIEW**

The survey throws light on the key aspects of the papers studied and also highlights their positive and negative points:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SL.NO** | **AUTHOR, TITLE AND YEAR OF PUBLICATION** | **METHODOLOGY** | **ADVANTAGES** | **LIMITATIONS** |
| **1.** | Joseph Redmon, Santosh Divvala, Ross Girshick Ali Farhadi  **You Only Look Once:**  **Unified, Real-Time Object Detection** –  May 2016. | i) Resizes  the input image to 448\*448.  ii) A single neural network predicts  bounding boxes and class probabilities directly from  full images in one evaluation using S x S grid on input.  iii)Processes images in real time at 45 frames per second (fps). | i)YOLO is extremely fast. Since, frame detection is a regression problem it does not require a complex pipeline.  ii) YOLO achieves more than twice the mean average precision of  other real-time systems.  iii) YOLO learns generalizable representations of objects.  When trained on natural images and tested on artwork,  YOLO outperforms top detection methods like DPM  and R-CNN by a wide margin. | YOLO imposes strong spatial constraints on bounding box predictions since each grid cell only predicts two boxes  and can only have one class. Also YOLO makes arbitrary guesses on boundary boxes. |
| **2.** | Joseph Redmon, Ali Farhadi  **YOLO9000:**  **Better, Faster, Stronger** – December 2016. | i)YOLO9000 also known as YOLOv2 is an improvised version of YOLO algorithm and is capable of detecting over 9000 object categories.  ii)Each prediction includes 4 parameters for the boundary box, 1 box confidence score (objectness) and 20 class probabilities. i.e. 5 boundary boxes with 25 parameters: 125 parameters per grid cell. | i)Improves accuracy by moving the class prediction from the cell level to the boundary box level.  ii)YOLOv2 has added batch normalization in convolution layers. This removes the need for dropouts and pushes mAP up 2%. | Difficult to detect objects when too many objects are overlapped. |
| **3.** | Hyeok-June Jeong, Kyeong-Sik Park, Young-Guk Ha  **Image Preprocessing for Efficient Training of YOLO**  **Deep Learning Networks**. **–**  January 2018. | This paper implements a pre-processor for YOLO object detector.  It uses the following 4 steps:  i) Image picker: This subsystem randomly picks out  objects from a prepared crawled image set.  ii) Scale Modifier: This subsystem reduces the size of  objects cropped in the Image picker to an appropriate  size.  iii) Image Maker: In this step, the modified object image is  pasted into base images. Base images with similar  backgrounds and sizes.  iv)Annotation Creator: This subsystem annotates the  position and size of newly placed objects in the base  image. | Since the images are pre-processed the performance of YOLO is better when compared to the performance without pre-processing. | The response / detection time is more due to the pre-processing of images.  Also this methodology cannot be used for real time systems. |
| **4.** | Wenbo Lan, Jianwu Dang, Yangping Wang and Song Wang  **Pedestrian Detection Based on YOLO Network Model –** August 2018. | This paper proposes the YOLO-R network structure, an improvised version of the YOLO network. | This system captures the gestured image and compares it with the predefined images in the database. On finding a match, it outputs the number which the gesture signifying. For example, if the gesture is just a forefinger, then the generated output is 1, etc. | The edge detection and segmentation algorithms used here are not very efficient when compared to neural networks. The dataset being considered here is very small and can be used to detect very few sign gestures. The histogram matching method is not very reliable and can give false output at times. |
| **5.** | Dnyanada R Jadhav, L. M. R. J Lobo,  **Navigation of PowerPoint Using Hand Gestures .**  (International Journal of Science and Research (IJSR) -  2015) | The System architecture consists of :  1.Image acquisition  2.Segmentation of hand region.  3.Distance transform method for gesture recognition. | 1.Distance transform method is better in performance than circular profiling method.  2. The presented gesture recognition system recognizes both static and dynamic gestures. | The limitations are:  1.The number of gestures that are recognized are less.  2.The gestures where not used to control any applications. |
| **6.** | Ruchi Manish Gurav  Premanand K. Kadbe ,  **Real time Finger Tracking and Contour Detection for Gesture Recognition using OpenCV.**  **(**International Conference on Industrial Instrumentation and Control (ICIC) – 2015) | There are three main algorithms that are used:  1.Viola and jones Algorithm.  2.Convex Hull Algorithm.  3. The AdaBoost based learning Algorithm. | The work was  accomplished by training a set of feature set which is local contour sequence.  The main advantage of Local contour sequence is that it is invariant to rotation, translation and  scaling so it is a good feature to train the learning machine.  We have achieved 92% accuracy with Convex Hull and 70% accuracy with AdaBoost. | The limitations of this system are that it requires two sets of images for classification.  One is the positive set that contains the required images.  The other is the negative set that contains contradicting images. |
| **7.** | Pei Xu,  **A Real-Time**  **Hand Gesture Recognition And Human Computer Interaction System.**  (Arxiv Journal  Repository-  Cornell  University  Library – April  2017) | The system consists  of three components:  1.Hand detection  2.Gesture  recognition  3. Human-Computer  Interaction (HCI)  It has implemented  the following  methodology:  1.The input image is  preprocessed and the  hand detector tries to  filter out the hand  from the input image  2.A CNN classifier is  employed to  recognize gestures  from the processed  image, while a  Kalman Filter is used  to estimate the  position of the  mouse cursor.  3.The recognition  and estimation  results are submitted  to a control centre  which decides the  action to be taken. | 1.The system uses  Convolutional  Neural  Networks(CNN)  which reach an  accuracy rate of  99% rather than  other approaches  such as Hidden  Markov  Model(HMM),  Orientation  Histogram which  are less accurate.  2.It uses only one  monocular  camera to capture  the image. | 1.The system recognizes only  static images.  2.The CNN used is not  robust and reliable since the  number of images and used  to train and test the classifier  is less.  3.It only recognizes a gesture  and not a motion of gestures  to control the mouse actions  (Gesture one for dragging the  mouse, Gesture two for  clicking the mouse, etc). |
| **8.** | P.Suganya, R.Sathya, K.Vijayalakshmi  **Detection And Recognition Of Hand Gestures To Control The System Applications By Neural Networks.**  (International Journal of Pure and Applied Mathematics, Volume 118 No. 10 -2018) | This project focuses on detection of hand gestures using java and neural networks. It is divided into two phases:-  1.Detection module using java where in the hand is detected using background subtraction and conversion of video feed into HSB video feed thus detecting skin pixels,  2.The second module is the prediction module, a convolutional neural network is used. The neural network is written with the Theano library. A NumPy array is used to collect the dataset images. The input feed image is gained from Java. The input image is fed into the neural network and is analyzed with respect to the dataset images. | 1.Optimised Recognition of Gestures Aided by Neural Networks (ORGAN) is used to control the mouse actions such as mouse press, mouse release, cursor movement etc. using hand gestures.  2.ORGAN had proved to predict hand gestures using an optimized detection and prediction mechanism.  It had shown its effectiveness without the use of any specialized hardware for boosting its performance and efficiency. | There a few limitations regarding this project:  1.Socket programming is required in order to connect the java and python modules.  2.This socket programming is unreliable sometimes and requires constant internet connection.  3.The training dataset used is small.  4.In order to control the cursor movement robot class in java is being used and that again increases the complexity of the system.  5.The number of gestures that are trained and classified are less since the dataset used is small. |

**CHAPTER 6**

**LIMITATIONS OF EXISTING SYSTEM**

A hand gesture recognition system was developed to capture the hand gesture being performed by the user and to control a computer system by that incoming information. Many of such systems in literature have strict constraints like wearing special gloves, having uniform background, long-sleeved user arm, being in certain lightning conditions, using specified camera parameters etc.

Many of the existing systems have implemented gesture recognition using only spatial modelling, i.e. recognition of a single gesture and not temporal modelling i.e. recognition of motion of gestures. Also the existing systems have not been implemented in real time, they use a pre captured image as an input for gesture recognition.

Some of the limitations of the existing systems are that they ruin the naturalness of a hand gesture recognition system and also correct detection rates and the performances of those systems are not well enough to work on a real time HCI system.

**CHAPTER 7**

**RESEARCH GAPS AND CHALLENGES**

The following are the research gaps and challenges faced:

* **Variation of illumination conditions :** any change in the lighting condition affects badly on the extracted hand skin region.
* **Background problem :** refers to the complex background where there is other objects in the scene with the hand objects and these objects might contain skin like color which would contradict and produce misclassification problem.
* **Translation problem :** the variation of hand positions in different images also leads to erroneous representation and extraction of the features.
* **Real time motion detection :** capturing the motion of gestures in real time and classification of the right action.
* **Training the Neural Network :** building, modelling and training the neural network to produce accurate results is a major challenge as it has to built and trained using a very large dataset and evaluate its results.
* **Datasets :** The quality of the available datasets and size of the images / video frames have a great impact on the gesture recognition.
* **Computational Complexity :** Training and implementation of the real time gesture recognition system requires a lot of computational capabilities such as use of high end Graphical Processing Units (GPU’s) .

**CHAPTER 8**

**PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY**

The system aims to design a vision-based hand gesture recognition system with a high correct detection rate along with a high-performance criterion, which can work in a real time HCI system without having any of the mentioned strict limitations (gloves, uniform background etc.) on the user environment.

It is composed of a human computer interaction system which uses hand gestures as input for communication. System is initiated with acquiring an image from a web-cam or a pre-recorded video sequence. Skin color is determined by an adaptive algorithm in the first few frames. Once the skin color is fixed for the current user, lightning and camera parameter conditions, hand is localized with a histogram clustering method. Then a hand gesture recognition algorithm is applied in consecutive frames to distinguish the current gesture. Finally, the gesture is used as an input for a computer application.

In the system that we are going to develop we are making use of Open Source Computer Vision, also known as OpenCV, a real time computer vision library with many image processing functions developed by Intel for the C++ or Java programming platform. This API provides many functions to perform calculation on captured video sources and special image processing on each frame of the video to support the HGR software. The video capture and the face detection components of this project are heavily supported by some of the functions built into the OpenCV library. The steps to be followed in hand gesture recognition are:

**1.HAND DETECTION :**

We need to have an input image from webcam which is fetching at the speed of 20 frames per second. The distance between hand and camera should be in the range of 30 to 100 cm. The video input is stored frame by frame into a matrix after preprocessing. For faster processing we use skin color to detect the variants and hand area. Hand gestures are meaningful continuous hand action. The image sequences captured by webcam contains garbage gestures which need to be removed by counter tracing algorithm.

**2.FEATURE EXTRACTION :**

a) **Centroid:** On basis of image intensities we divide image into two parts one with hand and other with non-hand region part. The centroid divides these two halves at its geometric center. Centroid is the center of mass of an object which is calculated using image movement. Image moment, which is the weighted average of pixel’s intensities of the image.

b) **Thumb detection:** Thumb is important in hand detection and a finger counting. The presence of thumb defines a set of gestures while absence defines other set. The thumb is on the corner of the box so we can identify which hand is there right or left-hand using thumb detection.

c)**Euclidean distance:** Calculate the distance between all tip of the fingers and centroid using Euclidean distance formula that is (a, b) = {(x(a) –x(b)) 2 +(y(a) –y(b)) 2 } 1/2 ,Where a and b are the are figure tips.

This distance is used to remove redundant fingers. If there are two fingers at a closer distance they shouldn’t be considered in figure count.

**3. GESTURE RECOGNITION :**

First, the number of fingers present in the hand gesture is determined by making use of defect points present in the hand gesture. The resultant gesture obtained is fed through a 3Dimensional Convolutional Neural Network consecutively to recognize the current gesture.

The recognized gesture is used an input for switching of pages, zooming in or zooming out, scrolling up or down, etc as each gesture would do a different task.

**SYSTEM ARCHITECTURE**



Flow diagram of hand gesture recognition



Gesture fed through a 3D CNN to recognize it

**CHAPTER 9**

**SYSTEM REQUIREMENT SPECIFICATIONS**

**HARDWARE REQUIREMENTS :**

1. **Web Camera :** 10 Megapixels or higher
2. **Processor:** Intel Core i5 or higher
3. **Graphical Processing Unit (GPU) :** 4 GB
4. **Memory :** 25 GB Disk Space

**SOFTWARE REQUIREMENTS :**

1. **Operating System :** Windows / Linux.
2. **Dataset :** of different hand gestures to train the neural network.
3. **OpenCV :** A real time computer vision library.
4. **NumPy :** Python package for scientific computing.
5. **Pandas :** Data analysis and manipulation library.
6. **Scikit - learn** : Python library for machine learning.

**CHAPTER 10**

**EXPECTED OUTCOMES**

The expected outcomes of the proposed project are :

1. A vision based hand gesture recognition system with a high correct detection rate along with a high performance criterion, which can work in real time HCI system without having any of the limitations such as gloves, uniform background, etc on the user environment.
2. The application to which the system will be put to use is a scenario where the user is reading a book on his/her laptop and he/she can perform various viewing actions such as switching of pages, scrolling the pages, zooming in and zooming out of pages.
3. Performing these actions even with movement of the eye of the user will also be implemented depending on the requirement.

**CHAPTER 11**

**REFERENCES**

[1] Meenakshi Panwar, Pawan Singh Mehra**, Hand Gesture Recognition for Human Computer Interaction**, International Conference on Image Information Processing India (2011).

[2]Rafiqul Zaman Khan and Noor Adnan Ibraheem, **Comparitive Study Of Hand Gesture Recognition System**, AIRCC Digital Library – 2012.

[3]Arpita Ray Sarkar ,G. Sanyal , S. Majumder ,**Hand Gesture Recognition Systems: A Survey** ,International Journal of Computer Applications Volume 71– No.15, May 2013.

[4]Manjunath A E ,Vijaya Kumar B P , Rajesh H, **Comparative Study of Hand Gesture Recognition Algorithms**, International Journal of Research in Computer and Communication Technology, Vol 3, Issue 4, April- 2014.

[5] Dnyanada R Jadhav, L. M. R. J Lobo, **Navigation of PowerPoint Using Hand Gestures** ,International Journal of Science and Research (IJSR) 2015.

[6] Ruchi Manish Gurav, Premanand K. Kadbe, **Real time finger tracking and contour detection for gesture recognition using OpenCV**, IEEE Conference May 2015, Pune India.

[7] Pei Xu, Department of Electrical and Computer Engineering, University of Minnesota, **A Real-time Hand Gesture Recognition and Human-Computer Interaction System** ,Research Paper April 2017.

[8] P.Suganya, R.Sathya, K.Vijayalakshmi , **Detection and Recognition of Gestures To Control The System Applications by Neural Networks**, International Journal of Pure and Applied Mathematics 2018.