Hands free HCI based on head tracking using feature extraction

Mrs. Nitty Sarah Alex[1]
Senior Assistant Professor, New Horizon College of engineering, Bangalore, India

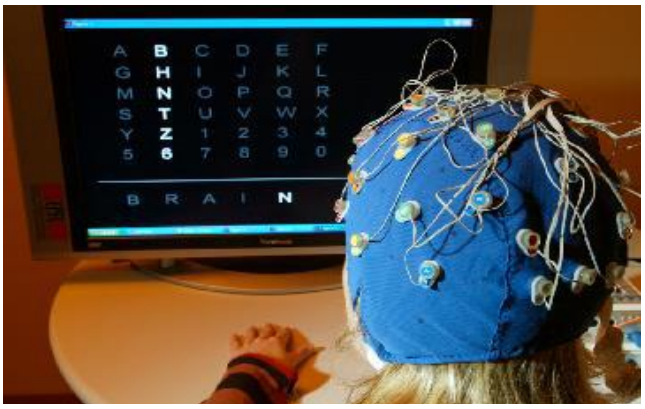
***Abstract*— The proposed system is an efficient way for Human Computer Interaction which consists of four cameras attached to a computer screen. The intention of this project is to ease the amount of effort required to instruct the computer. The component based method is considered for feature extraction since it out performs the other methods. Earlier interaction technique with keyboard and mouse is reduced and a system is proposed which uses a technique in which the eyes, nose and lips are used as medium of interaction. The aim of the project is to monitor the head movements to control the cursor movements.**

***Index Terms*— Human Computer Interaction (HCI), feature extraction.**

I. INTRODUCTION

Head movement based interaction is an upcoming user interface technique in the scenario. Many of the current systems have been developed to help people who are paralysed. The attention has been focused on to detect the user movements with a video camera and translate the motion into symbols to manipulate the operation of the system. It mainly helps the disabled and aged people who require a limited action to communicate.

A real time tracking system on a common PC is made possible due to the great progress in the field of Face detection/tracking and recent advances in hardware have made it possible. The disadvantages of this technique led to the invention of other HCI or the need for a better way of interaction led to the inventions of some HCI.



II. RELATED WORK

There is a lot amount of work and research going on related to head tracking based technique of interaction. This technique is being preferred to some of the other techniques because it does not require any physical interaction with the system.

The existing technologies that are in use and recent advances as mentioned in [1] on HCI gives common architectures or architectural patterns used in the design of HCI systems which includes unimodal and multimodal configurations.

The various new interaction methods which ease the use of computers to increase interaction efficiency in [2] specifies that eye tracking is a promising upcoming technology which enables it through eye gazing techniques to allow a user to move the mouse pointer using his eyes.

The various feature extraction techniques [3] that are existent and are in use effectively reveals the advantages and disadvantages of each of the corresponding feature extraction techniques like template based, color based method.

The technique for human computer interaction (HCI) using open source like python and OpenCV which proposes an algorithm [4] consisting of pre-processing, segmentation and feature extraction. Here features like moments of the image, centroid of the image and Euclidean distance is calculated were the input hand gesture images are taken by a camera. The algorithm is able to recognize the number of fingers present in the hand gesture.

A major advantage of this technology is that it offers the privilege of helping the technical system developers develop a system which can offer the existence of virtual reality (VR) which makes the user feel that there is another world on the other side of the screen. This technique of interaction is been preferred by many of the gamers to have an enriched experience during their gaming sessions.

There are many ways to implement this technology out of which some perform well whereas some of them work very slow making it quite unpleasant to use this technique. So as such a system need to be build which offers this technique with a greater potential to perform well.

The various advantages and disadvantages interaction techniques need to highlighted to understand the importance of the proposed technique of interaction. There are many techniques existing to interact with the computer and each of them has their own motives and needs but along with that they also have their own disadvantages. Below given are the few interaction techniques with their explanation with a brief related explanation about them.

A. Human Computer Interaction (HCI) through gestures.

Gesture recognition [4] enables humans to communicate with the machine (HMI) and interact naturally without any

ISSN: 2278 – 7798

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 potentially make conventional input devices such as mouse, keyboards and even touch-screens redundant.

This technology can be conducted with techniques from computer vision and image processing which is suitable in a controlled lab setting but does not generalize to arbitrary settings. If there are no high contrast stationary backgrounds and ambient lighting conditions, the recognition is very likely to make mistakes and the machine cannot recognize the start and end points of meaningful gestures from continuous motion of the hands.

B. Human Computer Interaction (HCI) through brain computer interface.

Advances in cognitive neuroscience and brain imaging technologies have started to provide us with the ability to interface directly with the human brain. This ability of human brain interaction is made possible through the use of sensors that can monitor some of the physical processes that occur within the brain that correspond with certain forms of thought.

Researchers have used these technologies to build brain computer interfaces (BCIs), communication systems that do not depend on the brain's normal output pathways of peripheral nerves and muscles. In these systems, users explicitly manipulate their brain activity instead of using motor movements to produce signals that can be used to control computers or communication devices. The major advantage of this technique is that it is majorly beneficial to the physically paralysed.

C. Human Computer Interaction (HCI) based on voice user interface (VUI).

This technique uses voice user interface to interact with the system where the words spoken by the user are recorded analyzed after which the required action is performed. However, VUIs are not without their challenges. People have very little patience for a "machine that doesn't understand". In VUIs need to respond to input reliably, or they will be rejected and often ridiculed by their users.

D. Human Computer Interaction (HCI) based on eye movement.

In this technique the user interacts with the computer through eye movements where the movement of the eye causes the cursor to move on the screen along with which aids the icons to open due to continuous focusing on it.

But this technique to be implemented requires infrared cameras which help in detecting the eye precisely. Another disadvantage is that stray movements of the eye cause the cursor also to move along.

E. Human Computer Interaction (HCI) through touch.

It is the most common technique that is used by users to interact with any possible system. The sensation of touch can be used to implement a security system which makes this technique very useful. The touch screen panels used are usually thick. The touch causes smears on the screen. Sometimes continuous usage of your hands to interact may cause it to pain which is another disadvantage of this technique.

III. Proposed Approach

This paper introduces a novel Hands-free HCI system that enables to navigate mouse cursor through head movements which is detected with the help of four cameras. As the basic idea of how our system actually works and other techniques is known which are under progress in the market we move ahead with learning how our system is going to implement this technique in the most efficient and effective possible manner.

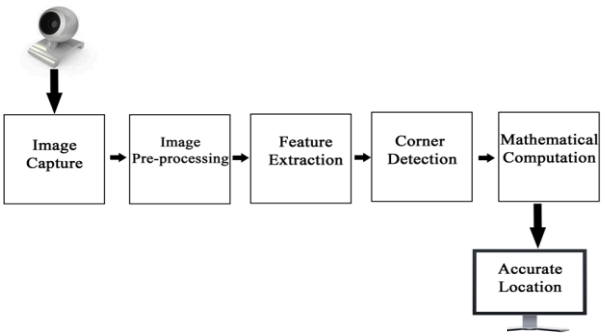


Fig. 1 HCI System Architecture

A. Capture of image and Image processing

A video camera which captures images in real time to a computer system on a computer network is placed on the four corners of the screen to capture the image of the user from all possible angles. When an image is captured using the web camera a few parameters like brightness, exposure, gamma, shutter and gain are kept in mind. All of them are set to zero in order to manage any manual image process.

A webcam is a video camera that feeds its image in real time to a computer or computer network. Unlike an IP camera (which uses a direct connection using Ethernet or Wi-Fi), a webcam is generally connected by a USB cable, FireWire cable, or similar cable. Here with the help of four USB cabled web cameras with 5 MP (VGA 1280x720) sensor, which has all the camera drivers built into the Operating System (Ubuntu). OpenCV claim to fame is its ability to effortlessly work with video feed, either live or recorded with the help of a class for video capturing from video files or cameras.

The video capture object is created with the web camera by object creation syntax: **VideoCapture (0)**; where 0 is the ID of the camera used.

Image processing is easiest but most important phase of a head tracking system. Once the image is captured it needs to be processed in order to extract the features from the image. The procedure is as follows:

1. Detect the projection of a pattern on the camera.
2. Then track the position of these projections to improve the quality of the image processing and pattern detection from the captured image.
3. The head tracking system uses the predicted position of projection to decrease the searching area and also to remove the falsely detected projections.

ISSN: 2278 – 7798

Image Processing uses the techniques of smoothing, also called blurring, for excellent smoothing to remove various noises and preserve details. These algorithms have the ability of preserving details, such as gradient weighting filtering, self-adaptive median filtering, robust smoothing and edge preserving filtering. A Gaussian blur (known as Gaussian smoothing) is the result of blurring an image by a Gaussian function which is a widely used in graphics software, typically to reduce image noise and reduce detail.

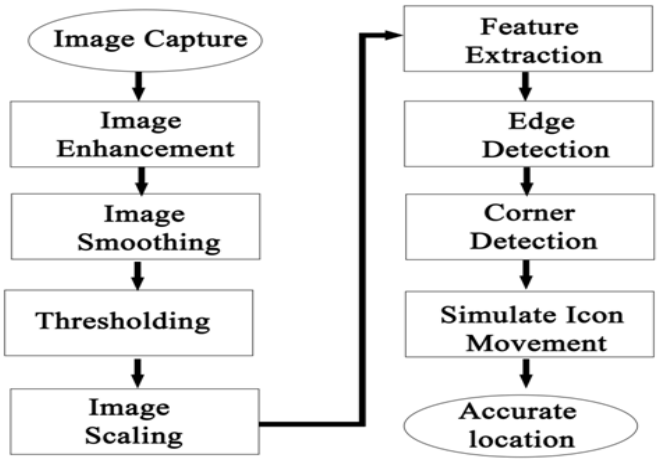


Fig. 2 System Modules

OpenCV offers five different smoothing operations. Among them GaussianBlur() is used with parameters are:

GaussianBlur(InputArray src**, OutputArray** dst**, Size** ksize**, double** sigmaX**, double** sigmaY=0**, int** borderType**=BORDER_DEFAULT)**
where,

- src – input image; with depth of CV_8U, CV_16U, CV_16S, CV_32F or CV_64F.
- dst – output image of the same size and type as src.
- ksize – Gaussian kernel size computed from sigma.
- sigmaX – Gaussian kernel standard deviation in X direction.
- sigmaY – Gaussian kernel standard deviation in Y direction; if sigmaY is zero, it is set to be equal to sigmaX, if both sigmas are zeroes, they are computed from ksize.width and ksize.height , respectively to fully control the result regardless of possible future modifications of all this semantics, it is recommended to specify all of ksize, sigmaX, and sigmaY.
- borderType – pixel extrapolation method.

B. Feature extraction

Detection of facial features is a very important step in number of applications like face recognition, face tracking etc. Feature Extraction [7] which relates to dimensionality reduction starts from initial set of measured data and builds the derived features which are intended to be informative, non-redundant, facilitating the subsequent learning and generalization steps, and also leads to better human interpretations.

Some image processing techniques extract feature points like eyes, nose and mouth and then use as input data for various applications. Various approaches have been proposed to extract these facial points or features from the images. Due to its importance feature selection is used in number of fields like computer vision, signal processing, statistics, neural network, pattern recognition and machine learning.

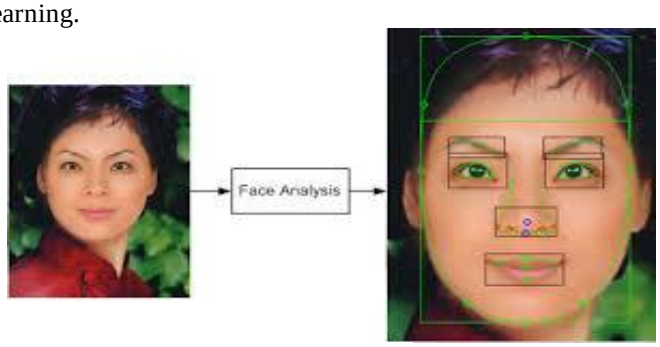


Fig. 3 Face Feature Extraction

There are number of feature extraction techniques like template based technique, face circle fitting and Eigen faces out of which Eigen faces are considered to have higher accuracy as compared to the other techniques.

1. ***Face circle fitting***: The technique [5] initializes a circular model and placed near the estimated center of the face adapted step by step to fit the face. The initialization of the circular model influences the accuracy of face circle fitting and the speed of the adoption process.
2. ***Template based technique***: It is a technique which extracts facial features based on previously designed templates [3] using appropriate energy functions and best match of template in facial image yields the minimum energy. The template matching is done with the image, by altering its parameter values to minimize the energy function, therefore deforming itself to find the best fit.
3. ***Eigen faces and Eigen vectors***: Eigen faces are the Eigen vectors of a set of faces, they do not necessarily correspond to features such as eyes, ears and noses. The projection operation characterizes an individual face by a weighted sum of the Eigen face features, and so to recognize a particular face it necessary only to compare these weights to those of known individuals. Mathematically we calculate the Eigen image from this set is considered as an Eigen vector and any changes between the images are characterized by these face images. These Eigen vectors when they are represented are known as Eigen faces. Each face can be represented as a linear combination of the Eigen faces.

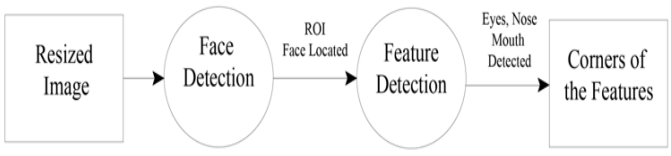


Fig. 4 Model of Feature Extraction from face

Proposed system is implemented using the method below:
Input : Digital RGB image.
Output : Simulate icon movement.

1. Image is captured from the web cameras.
2. Image conversion from RGB to Grayscale using formula (X=0.299R+0.587G+0.114B).
3. Enhancement of the image.
4. Image smoothing is performed using "***GaussianBlur***"

ISSN: 2278 – 7798

5. Thresholding image.
6. Image scaling for threshold image.
7. Extracting features from an image using Haar cascade classifier technique.
8. Edge detection and corner detection.
9. Mathematical computations to locate the points of the features of the face using like two point formula.
10. Icon on the screen is moved to the appropriate

C. Head tracking

A head tracking system is a system which tracks the movement of the head using the features extracted from the image in a particular pattern. The pattern can be detected from the image using two methods that are pattern detection or tracking using previously stored information. Normally the system will try to track the position of the pattern if the system fails in tracking the position or if it does not have enough information then the system will start the process again from pattern detection.

Tracking starts from first the position of the target ; While using color images the problem faced is to vector of covariance matrix of set of face images. Each the color of the object without prior knowledge especially in environment where the light condition changes. The features which are usually used in head tracking are the eyes, nose and lips. Tracking of the eye is different as compared to the nose and lips because these features are steady while the eyes are not (eg. Opening, closing, blinking, wicking etc). The eyes are usually tracked using a BTE (Between the Eyes) constant. The BTE is considered as the reference point to find the location of the eyes.

IV. System Model

A. Assumptions made for HCI system

The simple rectangular features of an image are calculated using an intermediate representation of an image, called the integral image. The integral image is an array containing the sums of the pixels' intensity values located directly to the left of a pixel and directly above the pixel at location (x, y) inclusive. So if A[x,y] is the original image and AI[x,y] is the integral image then the integral image is computed.

The features rotated by forty-five degrees require another intermediate representation called the rotated integral image or rotated sum auxiliary image. The rotated integral image is calculated by finding the sum of the pixels' intensity values that are located at a forty five degree angle to the left and above for the x value and below for the y value. So if A[x,y] is the original image and AR[x,y] is the rotated integral image.

Using two passes both integral image arrays computed and by taking the difference between six to eight array elements forming two or three connected rectangles, a feature of any scale can be computed. Thus calculating a feature is extremely fast and efficient. The detection of various sizes of the same object requires the same amount of effort and time as objects of similar sizes since scaling requires no additional effort. Detecting human facial features, such as the mouth, eyes, and nose require that Haar classifier [6] cascades first be trained. In order to train the classifiers a gentle AdaBoost algorithm and Haar feature algorithms must be implemented.

B. Implementation Environment

The computer vision system implemented using Open Computer Vision Library (OpenCV). The OpenCV library is designed to be used in conjunction with applications that pertain to the field of HCI, robotics, biometrics, image processing, and other areas where visualization is important and includes an implementation of Haar classifier detection and training.

The first step in facial feature detection is detecting the face by analyzing the entire image. The second step is using the isolated face(s) to detect each feature. Since each the portion of the image used to detect a feature is much smaller than that of the whole image, detection of all three facial features takes less time on average than detecting the face itself. Using a 1.2GHz AMD processor to analyze a 320 by 240 image, a frame rate of 3 frames per second was achieved. Since a frame rate of 5 frames per second was achieved in facial detection only by using a much faster processor, regionalization provides a tremendous increase in efficiency in facial feature detection.

C. Result Analysis

The cameras are connected to the system and the video stream is captured from the camera to the system. The snapshot shows of an instance of the video stream captured by the system from the camera. Once the image is captured from the cameras the features are detected and extracted from the image. The features extracted are the eyes nose and the mouth.

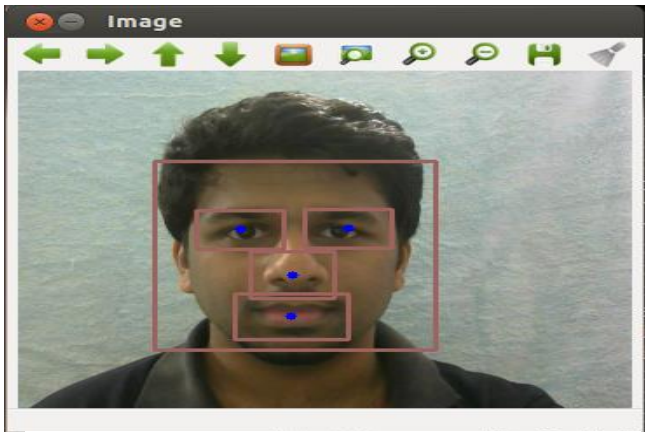


Fig. 5 Image features detected and extracted

Once the features are detected the distance between every feature is measured and the values are saved as in Fig 6.

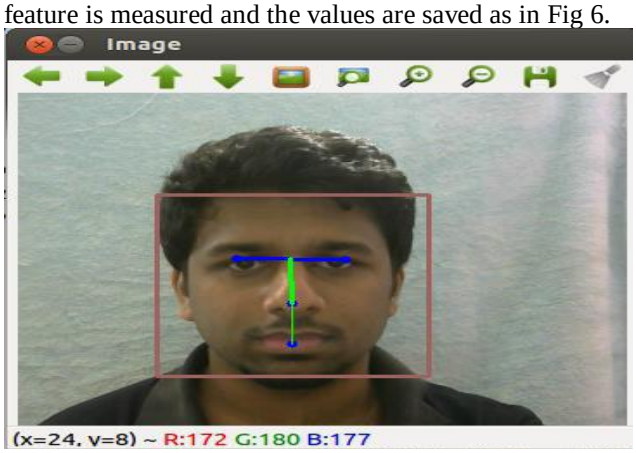


Fig. 6 Distance between Features measured

ISSN: 2278 – 7798

The position of the face from different angles of the cameras positioned at the different sides of the system is taken and the deviation in the distance between the features is noted. It is noticed that when you look down the distance between the nose and the eyes increases and if you look upwards the distance between the eyes and the nose decreases. The snapshots shown below are examples of the same.

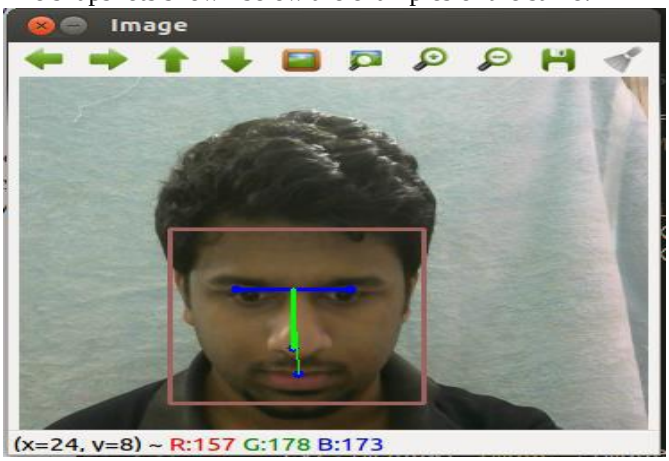


Fig. 7 Distance increases as we look down

The distance between the eyes and the nose decreases as we look up. The camera detects the face from different angles as shown in Fig 8.

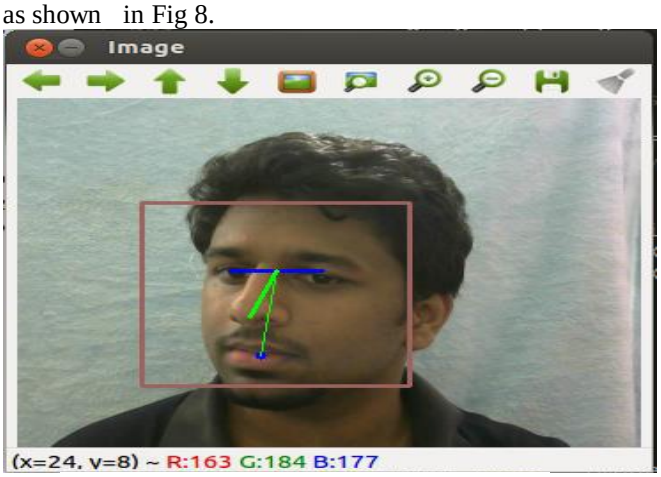


Fig. 8 Angle based detection

As all the cameras capture different images the features are extracted from them and the distance between each of the features in each of the images is calculated at first. Next the average of the different images is calculated using mathematical formulae and the icon is placed on the screen in the position of the average of all the cameras.

The features of the image that is eyes, nose and mouth are detected and that will be presented by point on the output screen. For four cameras the same scenario is followed same as one camera's scenario. The four points are merged together to represent a single point on the output screen.

The features of the image are extracted using various techniques which are in turn used for finding the location of the focus. Once the merged point appears on the screen, it is saved. The grey scale window which will present on the screen after saving will move according to the way the user move the head.

V. Conclusion and Future Work

The proposed system allows its users to interact with it in least efforts on the user's side which makes it more preferable compared to other existing technologies. The system can be envisaged as a system which can be developed and used by physically disabled and paralyzed patients who can move only their heads. For example euthanasia patients who can move only their head as rest of the body parts are will not work. This system uses web cameras which is economically feasible as compared to other systems.

This system can be further enhanced to provide virtual reality displays using head tracking.

REFERENCES

[1] Fakhreddine karray, Milad alemzadeh, Jamil abou saleh and Mo Nours Arab , "Human-Computer Interaction: overview on state of the art" International Journal on Smart Sensing and Intelligent systems, vol. 1, No. 1, March 2008

[2] Ibraheem Fraisleer , Dept of Comp Sc. University of the Western Cape. "Moving the Mouse Pointer Using Eye Gazing", 2012.

[3] Gaurav Mandloi, "A Survey on Feature Extraction Techniques for Colour Images", International Journal of Computer Science and Information Technologies (IJCSIT) , Vol. 5 (3), 2014, 4615-4620.

[4] Nayana P B , Sanjeev Kubakaddi, "Implementation of Hand Gesture Recognition Technique for HCI Using Open CV", International Journal of Recent Development in Engineering and Technology, ISSN 2347 - 6435 (Online) , Volume 2, Issue 5 , May 2014.

[5] Daw-Tung Lin and Chian-ming yang "Real-time eye Detection using Face Circle Fitting and Dark-Pixel Filtering", 2004.

[6] Phillip Ian Wilson, John Fernandez , "Facial feature detection using Haar classifiers", Journal of Computing Sciences in Colleges, Volume 21, Issue 4, April 2006 , Pages 127-133

[7] Sanjeev Dhawan, Himanshu Dogra, "Feature Extraction Techniques for Face Recognition", International Journal of Engineering, Business and Enterprise A pplications (IJEBEA), Aug-Nov 2012, pp. 1-4.

[1]**Nitty Sarah Alex,** Senior Assistant Professor in the Dept. of Information Science and Engg, New Horizon College of Engg, Bangalore , India with work experience of 11 years. My publications include IEEE conference in visual cryptography, three Journal publications in various areas like Sybil attack in WSN, Mail client etc. Along with this I have presented a paper in national conference during my M.tech degree. I have completed a project under VGST grant in Robotics and submitted.

ISSN: 2278 – 7798