Motion Tracking Using HCI for Interactive Gaming

T.Siron Anita Susan[1], S.Vikas[2], R.Sathish[3], T.Hari Ganesh[4], S. Renganatha Varun[5]
[1]Assistant Professor, [2,3,4,5] B.Tech Scholar,
Saranathan College of Engineering, Trichy

***Abstract*- Gaming involves heavy concentration, presence of mind, quicker responses from the gamers. Many users use conventional input devices like keyboard, mouse,for interacting with the gaming application.These conventional devices reduce the physical work and increase the chances of obesity, mental tiredness, etc., and are considered unhealthy practices. These devices can be replaced by using HCI technology which provides functionalities to computer that can recognize hand gestures and body motions of the user. Human Computer Interaction (HCI) techniques like hand gesture recognition, voice identification etc., can increase the physical work and can make the user more interactive thus reducing health related risks.**

***Index Terms* - Gesture Recognition, Kinect, Gaze Tracking**

I. INTRODUCTION

Hand Gesture Recognition systems are becoming popular now-a-days because of their simplicity and ease of use. Many hand recognition systems are proposed for better interaction of human with the computer. HCI has wide range of applications like video games, medical areas, training military personnel, analyzing complex chemical compounds in high-tech laboratories, robotics etc.,

One of the earliest gesture recognition included a special glove embedded with various sensors for tracking the motion of the hand and for understanding the gesture produced by the user. These high-tech gloves embeddedwith sensors tend to be more expensive than other systems and cannot be used by all the users.

KINECT[1] is a sensor developed by Microsoft for motion sensing. It is compatible with Windows PC and is designed around a Webcam Style (Plug and Play). Rather than using conventional input devices like keyboard, mouse for interacting with the system, KINECT can be used for gesture Recognition which gives appropriate commands to the system. KINECT uses color sensor for skin detection and an IR based depth detection mechanism. IR Emitter emits IR radiations and it is reflected back to the sensor in KINECT. By the amount of time of reflection of the IR signals, depth of each object from the screen is calculated using some mathematical calculations. This depth sensor and color sensor are combined to recognize only the gesture from user and filters any unwanted noise from the background. After gesture recognition, appropriate commands are issued to the system by the KINECT. These commands are not only useful for controlling the computer system and can also be used to play video games.

The main disadvantage of KINECT is that it incurs additional hardware cost and IR radiations are not suitable in all the environments.

Gaze tracking[2] is a mechanism by the gazing location of the eye is spotted and the mouse pointer is moved to the gazing location. This reduces continuous tracking of the mouse pointer by the user. When the user gazes at a location on the screen, the webcam is used to track the relative position of the eye gazing location. This gives an approximate gazing location to the system. The mouse pointer is moved immediately to the approximate gazing location. From that location, user can use the mouse slightly to get to the exact gazing location. This method of gaze tracking significantly reduces the fatigue effects.

Though it reduces fatigue effects on the user, it increases the error rate of getting to the gazing location. First, the webcam should be able to capture the sudden movement of the eye, which is a difficult task because webcam should have a high frame rate of capturing images. Secondly, the user should synchronize with the system. This is also difficult as various users have different response time. Third, the exact gazing location cannot simply be calculated using a webcam as the movement of eye is not always steady.

Hand Gesture spotting and Recognition[3] is always a complex task to do because of the variety of users and environments. Hand gesture spotting and recognition involves calculating the starting and end points of the hand gesture using image processing or with some complex mathematical calculations. This system proposes a system where the hand motion is detected with some complex calculations and mathematical models.

This system executes gesture segmentation and recognition simultaneously without any delay in the system. It can be used for real-time implementation of hand gesture recognition. It uses Hidden Markov Model (HMM) for Gesture Segmentation and recognition. Hidden Markov Model is capable of modeling spatio-temporal time series of gestures effectively and can handle non-gesture patterns as well.

After gesture segmentation and recognition, the motion of the hand should be tracked to find the relative location of hand to the screen. This can be done using a special algorithm called Mean Shift Algorithm. Mean Shift Algorithm is a non-parametric density estimator that optimizes the smooth similarity function to find the direction of hand gesture path. In order to spot the key gesture effectively, this system proposes a new model called non-gesture model which is created by collecting all the gesture models to the system. This system also uses a stereo camera to capturing the gestures.

This system uses complex calculations for gesture segmentation and tracking which degrades the performance of the gesture recognition system.

Webcam can be used for gesture recognition and tracking of gestures, but the performance of the system gets degraded when various illumination environments are used while gesture recognition. This system proposes a technique for improving the performance of the gesture recognition system by reducing the error rate of the system in various illuminated environments[4].

The system uses an algorithm known as K-Nearest Neighbor algorithm. This algorithm calculates the similarity between two neighboring pixels and classifies the pixels accordingly. The webcam captures the frames of images from the environments and illumination of the environment is done for reducing unwanted noise. Motion and skin is detected using the frames from the webcam. The face is segmented and relative motion of the eye alone is tracked from the skin detected frame. According to the relative location of the eye, the gesture recognition system issues the control signals to the system. These control signals move the mouse cursor to the relative coordinates on the screen. This movement of mouse cursor can be utilized or exploited to control the computer system.

K-Nearest Neighbor algorithm involves many calculations for detecting the skin from the background noises. This reduces the performance of the system but provides more accuracy when it comes to various illuminated environments.

Hand Gestures can be recognized by comparing the gesture video stream[5] with predefined motion templates. This system proposes a simple and robust system of recognizing the gesture and direction of motion using video analytic method. The video of the gesture environment is captured and analyzed with templates to find the particular gesture. In this method, Skin pixel detection and background detection is also done to improve the performance of the gesture recognition software.

Since, motion templates are employed here, complex gestures cannot be identified and new gestures which are not saved previously cannot be detected using this method.

Considering all the disadvantages in the previous systems, the cost of implementation of a gesture recognition system should be reduced to a negligible amount for usage by all the users.

This paper presents an approach for recognizing the gestural actions of the user using a simple webcam and a normal PC. This system is aimed to recognize the user actions and to track motion of the human hand and to produce control signals to the system according to the direction of motion of the hand. This control signals can be used to control the gaming application for interactive gaming by the system.

II. WEBCAM BASED MOTION TRACKING

A standard webcam is cheaper for any user and has the potential to replace the conventional input devices like keyboard andmouse. Webcam can take images in the 640x480 resolution at a moderate speed for effective gesture processing. This webcam can be used in a normal Windows PC with Intel Dual Core processor and 2GB RAM and Java 7.0 or above. Webcam captures number of frames per second.These frames can be processed by using various image processing algorithms to find the relative movement of object in the frame and can also be used to give commands to the computer system. This type of understanding the user motions by the system is known as motion tracking using Webcam.

III. IMPLEMENTATION

The implementation involves the following steps

1. Capturing Frame
2. Skin Detection
3. Segmentation
4. Tracking Motion
5. Control Signal

PROCESS FLOW

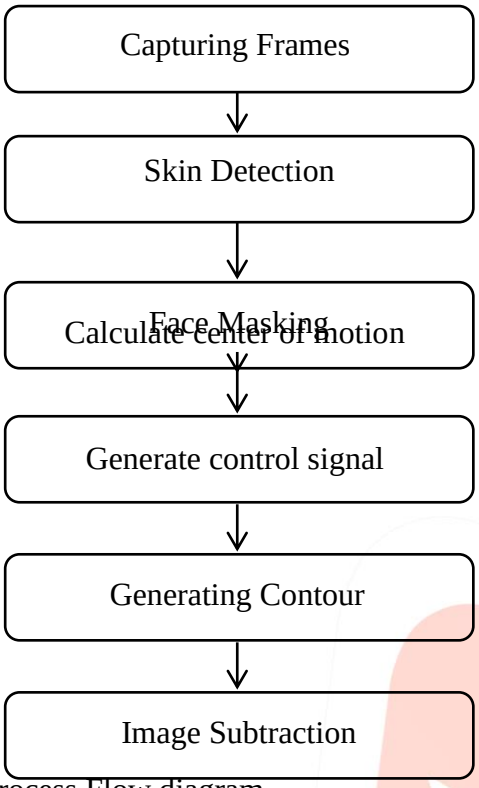

Fig.1 Process Flow diagram

1. CAPTURING FRAMES

Capturing the frames involve capturing the user actions as images using a webcamera. Capturing each frame depends on the frame rate of the web camera. Each frame is to be processed by the Gesture Recognition system. The optimized resolution for a standard webcam is 640x480 pixels. This Resolution gives optimized quality and frame rate for faster response from the system (Real-time Gaming Experience).

Fig.2 Captured Frame

2. SKIN DETECTION

Skin detection involves generating a binary image containing skin colored pixels in a color and non-skin colored pixels as a different color. Skin detection uses many color models for effective hand tracking.

The three color models used here are RGB,YCbCr and HSI.RGB color model is the most familiar color model in present scenario which has the basic colors red, green and blue to represent each pixel in an image. RGB color model is well suited for displaying images in the screen and is not very useful in image processing because it has less difference for skin color and background.

YCbCr[6] color model separates the Luminance or brightness and the Chrominance from the image and has high difference for the background and the skin area. So, it is very much used in image processing.HSI color model can also be used for improving the performance of the skin detection algorithm.

Fig 3. Skin Detected Image

3. SEGMENTATION

Segmentation is the process of finding the face of the user and masking the face for removing unwanted skin colored pixels in the image.

This segments or separates the face of the user and the hand improving the effectiveness of the motion tracking system. This can be done using Haar-cascade face detection algorithm.

Face Detection is the process of identifying the bounding rectangle of the face. This can be done using Haar-Cascade Face Detection algorithm which uses a set of Positive images (images containing face) and negative images (images which does not contain face) to train a set of qualifiers to detect faces in the frame.

Haar-Cascade includes various templates that are matched to the image to find the face area in the frame. Various rectangles are created in various sizes to check the size of the face.

4. TRACKING MOTION

When the skin area is detected and the face is masked in the image.It contains only the hand, this image can be used to find the center of the hand by applying the following steps,

a. REFERENCE IMAGE

Reference image is an image generated for tracing the contour of the hand.This reference image contains grid of lines with 40 pixels spacing between the lines.This is used for identifying the relative position of the hand in the skin detected image.

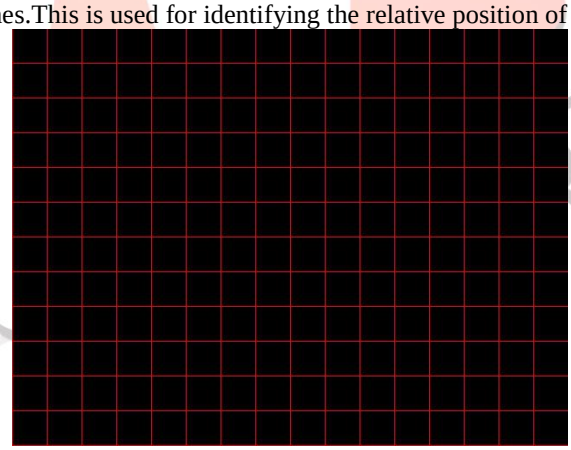

Fig 4. Reference Image

b. IMAGE SUBTRACTION

The reference image is then subtracted from the segmented image to get the line of pixels.

c. CONTOUR EXTRACTION

The image subtraction gives number of lines according to the reference image and the skin detected image. These lines represent the hand area.

Fig 5.Subtracted Image

d. CALCULATE CENTER OF MOTION

Once the hand area or contour is generated, the center of motion of the hand can be calculated by finding the midpoint of the middle line in the image.This can be calculated by counting the number of pixels in each line and counting the no of lines in the image.

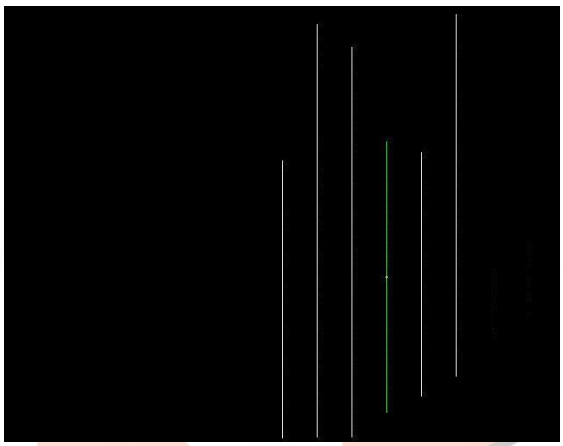

Fig 6. Contour Extraction

5. CONTROL SIGNALS

Since the resolution of the image is 640x480, it can only contain center of motion from 0,0 to 640,480. This image coordinates should be mapped to the screen coordinates by using some conversion factor according to thescreen coordinates.

IV. CONCLUSION

The proposed system provides a better performance in terms of response time and calculation reduction. Since, the proposed system uses various easy and less complex calculations, It will be easier for modification for further development in the future. It also increases the gaming experience in terms of Physical work and Virtual Reality.

V. FUTURE ENHANCEMENT

The project can be extended for motion of two hands for more real-time games like car racing, etc.,

REFERENCES

[1] Bo Yuan and Yanmin Zh,Real-Time Hand Gesture Recognition with Kinect for Playing Racing Video Games.

[2] David Rozado ,Mouse and Keyboard Cursor Warping to Accelerate and Reduce the Effort ofRoutine HCI Input Tasks.

[3] Jyh-Horng Chen, Yuan-Pin Lin, Yi-Ping Chao, Chung-Chih Lin, Webcam Mouse Using Face and Eye Tracking in Various Illumination Environments.

[4] Mahmoud Elmezain, Ayoub Al-Hamadi, Bernd Michaelis., A Novel System for Automatic Hand Gesture Spotting and Recognition in StereoColor Image Sequences.

[5] ShrikantKulkarni, Manoj H and SumamDavid,Robust Hand ,Gesture Recognition System Using Motion Templates.

[6] SumedhaSirsikar, VarshaPowar, AditiJahagirdar, Skin Detection in YCbCr Color Space.

[7] Siddharth S. Rautaray and AnupamAgrawal, Interaction with Virtual Game through Hand Gesture Recognition.