Web Browser Control Through Hand Gestures using Web RTC API

Indrajit Bastapure
Department of Computer Engineering
AISSMS Institute Of Information Technology
Pune411001, Maharashtra ,India

Tanmai Aurangabadkar
Department of Computer Engineering
AISSMS Institute Of Information Technology
Pune411001, Maharashtra ,India

Revati Tavare
Department of Computer Engineering
AISSMS Institute Of Information Technology
Pune411001, Maharashtra ,India

Abstract- Gestures are a major form of human communication. A primary goal of gesture recognition is to create a system which can identify specific human gestures and use them to convey information for device control and by implementing real time gesture recognition a user can control a computer by doing a specific gesture in front of a video camera linked to the computer. Till now the concept of hand gesture was implemented using the different languages like C, C++ and Python which were quite complicated. This paper mainly emphasizes on the research based concept of web control using hand gestures through different hand recognition methods. There are many new technologies available for normal users of internet like HTML-5, JavaScript, and JQuery which are used in this research based concept which is implemented with the help of Canvas Tag to capture the live feed of Webcam. Many applications and techniques are discussed here with the explanation of the system recognition framework and its main phases.

Keywords- Hand Gesture, Gesture-Control, Hand Gesture Recognition, Human-Computer Interaction (HCI),DegreesOfFreedom(DOFs),Metacarpophalangeal(MCP,DistalInterphalangeal(DIP),Proximal Interphalangeal (PIP),Web RTC(Web Real Time Communication), API (Application Programming Interfac),RGBA(Red,Green,Blue,Alpha).

INTRODUCTION

Gesture recognition is a part of computer science and language technology which interprets human

gestures using different mathematical algorithms. Gestures can be originated mainly from bodily motion or state. It is a physical movement of hands, arms or body which conveys meaningful information regarding gestures.

Many approaches have been made using cameras and computer vision algorithms to interpret sign language. Gesture recognition has been studied in widely topics, and has a wide range of applications such as recognizing of human computer interaction (HCI), robot control, machine vision. There are many applications which are based on hand gesture control.

 PC or laptops and hardware devices like mouse or keyboard are used for accessing Internet.In order to access web efficiently and in easier way then gesture based concept can be used. Although the hand is a complex biological organ, it can be thought as a mechanical machine and mechanical principles can be applied to it. In this context, three elements are involved: muscles serve as the motor for providing driving force, tendons, bones, and joints transmit the motors driving
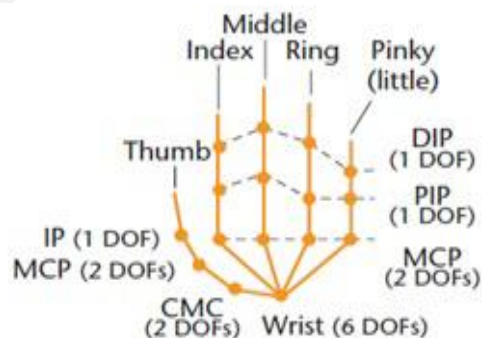

 Fig.1 Simplified representation of the hand structure[1][2]. force, skin and pulp tissues apply the force[1][1].In the given figure below, the lines represent the hands bones and thedots are its joints. Hand motions at joints are marked with degrees of freedom (DOFs), for example, at the middle finger,

metacarpophalangeal (MCP) joint, two different movements are possible.

Other joints depicted include the interphalangeal (IP), carpometacarpal (CMC),distal interphalangeal (DIP), and proximal interphalangeal (PIP).

Nowadays, the majority of the human-computer interaction (HCI) is based on mechanical devices such as keyboards, mouse, joysticks or game pads. In recent years there has been a growing interest in a class of methods based on computational vision due to its ability to recognise human gestures in a

natural way. These methods use as input the images acquired from a camera or from a stereo pair of cameras. The main goal of these algorithms is to measure the hand configuration in each time instant.In Human computer interaction (HCI) mostly hand is used than any other part of body for interaction. Human hand is having many joints and can be transformed into many gestures with the help of muscles. There are various implementations
using sophisticated algorithms which include modeling with underlying anatomical structure, data-driven algorithms in hand animation. Human hand structure is studied in these gesture based concept.

In the whole process, first the video stream is
retrieved which contains a particular gesture. Then the gesture in the video stream is identified, the meaning related to it is then interpreted and thus the final action is performed .Gesture detection from the video streams is an important task.
There are many methods available for detection like vision based approaches and the data glow approaches .In the data glow approach, user has to wear devices that connects it to the computer which reduces the natural level of interaction with computer. In the vision based approach, different properties such as color and texture are used in analyzing the gestures.

TABLE I.   HAND GESTURE RECOGNITION SYSTEM

Category	Type
Application	sign language, games, Robot control.
Motion	static, dynamic.
Image acquired data	Cameras, video, coloured gloves.
Data dimensions	2D, 3D.
Number of hands used	one hand.
Input features	3D hand model ,Appearance based low level features.
Gesture modality	communicative, manipulative.

RELATED WORKS

Gesture based system can be of three states gloves
based, vision based and low level features based. Vision based approach is most promising but also depends on costly sensors and wires [2][1]. Vision based approach need robots with camera, gesture recognition algorithm translates human gesture into command .In appearance based approach models hand using the intensity of 2D images defines gesture as sequence of views. This approach is easier than 3D model approach. Low level features based approach requires mapping between input video and finding the centroid of the hand region, an elliptical bounding region of the hand, edges, regions, silhouettes, moments and histograms.
Gesture recognition system has phases which are segmentation, feature detection, extraction etc. Segmentation is an important phase that affects accuracy of gesture detection. After obtaining input gesture first step is segmentation to extract hand region from image [2][4][1] [2].

Fig. 2. Hand Segmentation (Transformed image from RGB to HSV)

Fig. 3. Normalized hand(left) and Gray scale image(right)

Temporal gesture[6] can be regarded as a continuous change sequence of gesture trajectory. A standard gestures database is being built. Before this, hand is detected by acquiring the images through the web-cam. And then the gesture extraction of features on standard gesture videos is being performed, thus by setting up the rules. In real-time recognition process, when a gesture input is given, the hand position is detected by a hand detector. The consecutive gesture segmentation process and feature extraction process is being followed. Finally, the nearest neighbour method[6] is used to decide in which class of gestures the given input gesture belongs to. This process of segmentation is very important in the sense of hand detection.

Fig. 4. Application Programming Interface

API is nothing but the elements used for building software applications and these elements are the set of routines ,protocols and different tools. It shows the interaction between the different software components and can be used in building of the graphical user interface. API is somewhat like building blocks which help in the construction of the programs. Technology package is built by the Google and is openly available via webrtc.org . It is an open community which defines and standardizes low cost, high quality audio and video communication [4][1].Also no licences or other fees are required. It's easy for people to use these services [4][2]:
-There are no downloads, no installations
-Only one has to just surf through the right address.
-No need to write the implementation code, just call it.

services like real-time streaming audio, video or data communication in the real-time streaming audio, video or data. Also exchange session control messages and media information are available.

Algorithm

getUserMedia

Algorithm getUserMedia
{
    Step 1-Get streaming video with getUserMedia.
    Step 2-Set the attribute video as true.
    Step 3-Pass the video stream to the function.
    Step 4-Select the query with video attribute.
    Step 5-Show error message if it fails.
    Step 6-Success.
    }

This specification defines the 2D Context for the HTML canvas[5] element. The 2D Context provides objects, methods, and properties to draw and manipulate graphics on a canvas drawing surface [4][1] The Canvas 2D Context specification
is supported in:
-Safari 2.0+
- Chrome 3.0+
- Firefox 3.0+
- Internet Explorer 9.0+
- Opera 10.0+
- iOS (Mobile Safari) 1.0+

- Android 1.0+

Canvas

Algorithm Canvas
{
 Step 1-Create a html page with canvas element.
 Step 2-If browser is not supporting canvas element then display error.
 Step 3-Create canvas id.
 Step 4-Set height and width attributes.
 Step 5-Select the 2D or 3D context.
 Step 6-Draw images with it.
}

Proposed Model

According to the system architecture, the user interacts with the system with the help of hardware devices such as web camera or microphone, this live video feed of user interaction with webcam is captured with the help of Web RTC API[4].
This structure will be the main interface for user and browser to use the Gesture API. This API will incorporate a custom algorithm that is going to decide the nature of hand motion captured through web cam using [4]Web RTC API. The motion captured will be considered as a live

video feed which will be piped to the Canvas tag[5]. An algorithm will analyse every frame as to detect the hand motion and translate it into a custom trigger which will be incorporating a function of user's choice.
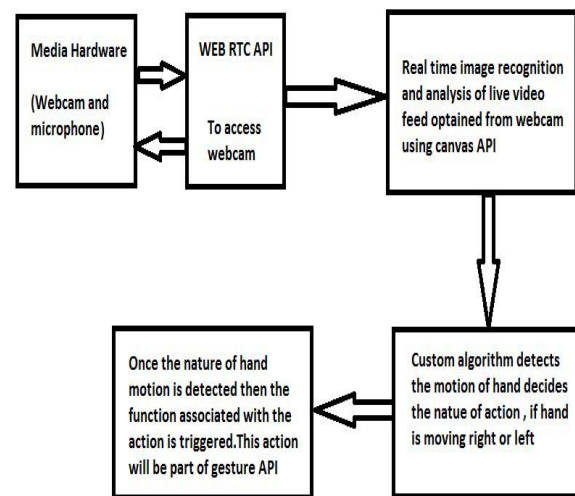

Fig. 5. System Architecture

To implement the algorithm, Canvas tag[5] will divide the individual frame captured from video feed in nine quadrants. Now the algorithm will convert the RGB frame into black and white frame.
The white part will represent the skin of hand and rest of the objects will be considered as the black region. Now every frame will be analysed for white region and the difference obtained in the nine quadrants will decide the up, down, left and right actions.

Blend mode difference method

There are many algorithmic methods used for converting the RGB image into the binarized imaged. But here, Blend mode difference method is used. In this, the difference between the next and the previous images is calculated. The difference obtained can be of 3 types: positive, negative and zero. The positive and the negative difference is considered as a change in pixel whereas zero is considered as no change in pixel. When there is change in pixel value (difference is positive or negative) then the value of the pixel is changed to white colour (FFF). And whenever no change in the pixel is seen, the value of that pixel is changed to black(000).

Difference = Next image - Previous image
The pixel is associated with mainly 4 different
values which are: red, green, blue and alpha.(value of alpha is also always (255H)).

Fig. 6. Pixel structure

Pseudocode

Step 1: Retrieve previous and next RGB images I1 and I2.
Step 2: Select a pixel from both the images i.e P1 and P2.
Step 3: Find the RGB values of the pixel which are
R0,G0,B0,A0
R1,G1,B1.A1.
Step 4: Calculate the difference of pixels between them.
D= (D1,D2,D3,D4)
D1= R1-R0.
D2= G1-G0.
D3= B1-B0.
D4= A1-A0.
Step 5: If D is changed then P1 or P2 = FFF(white).
Step 6: If D is not changed then P1 or P2 = 000(black).
Step 7: Thus, the binaries image is obtained.
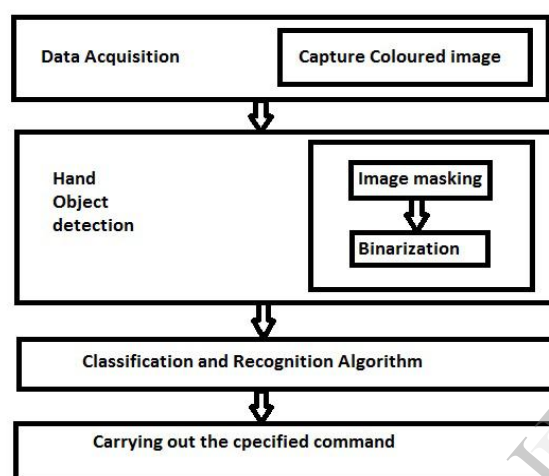

Fig. 8. Image Analysis [4][3]

Fig. 7. The flow of gesture recognition system

A. Image Analysis

The technique of image analysis combines techniques that compute statistics and measurements based on the gray-level intensities of the image pixels. One can use the image analysis functions to determine whether the image quality is good enough for your inspection task. One can also analyze an image to understand its content and to decide which type of inspection tools are to be used by the application. Image analysis functions also provide measurements you can use to perform basic inspection tasks such as presence or absence verification.
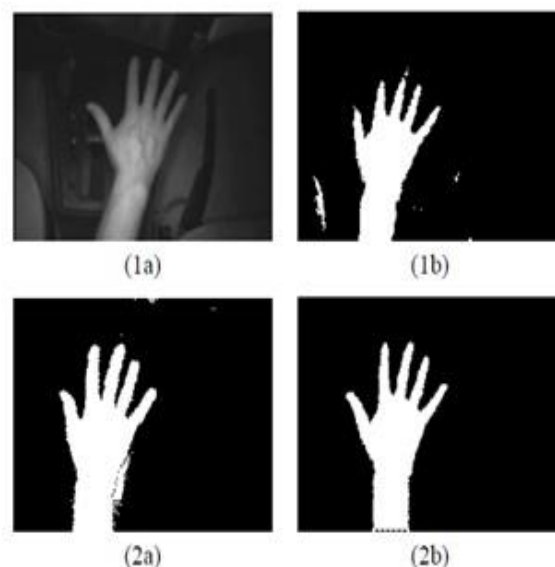
Future Scope

Since use of hand gestures is increasing in the day-to-day technical solutions. Refinements such as by adding more gestures, where my computer operations like Cut, Copy, Paste and Undo can be implemented. Thus, by integrating this system with voice recognition system and robots being embedded in it interactive computing gaming may be possible. Enhancement of our system to control power-point application can be done. Further, dynamic image processing and event handling procedures can also be implemented through hand gesture control in case of CAD simulation. Moreover, with the security concerns gesture based passwords can be developed. Hence, in the later stages technologies which completely remove mouse dependency can be developed with the help of this hand gesture based control system. The system implemented for future applications such as Sign Language Recognition, Robotics, Human Manipulation and Instruction, Virtual Reality, Gestureto- speech, Games[7], Television Control.
 -Highly Interactive Computer gaming[7]
 -Removing mouse dependency completely from the system
-CAD Simulation
- Interactive presentation[8]
- Gesture based passwords
- Car driving [5]

CONCLUSION

Gesture recognition system has been prevalent since the past decade which allows control on computer games, robots, medical visualization devices and also crisis management systems. This provides an expressive, natural and intuitive way of interaction through the use of hand gesture interfaces. Such systems should have:

-Fast response time.

-High recognition accuracy.

- Ease of interactive learning.

- CAD/CAM Simulation.

- Cross platform software.

- A high degree of user satisfaction.

This system makes use of webcam only, there is no

additional hardware is required. This paper review highlights on the potential of hand gestures as a natural model for human-machine interaction and indicates the need for additional research and evaluation procedures to gain widespread acceptance.

REFERENCES

1. R. Bowen Loftin, Jim X,"REAL-TIME NATURAL HAND GESTURES" IEEE computing in science and engineering,2005

2. Rafiqul Zaman Khan , Noor Adnan Ibraheem "Survey on Gesture Recognition for Hand Image Postures" Faculty of Science, Department of Computer Science, Aligarh Muslim University, Aligarh 202002,India URL: http://dx.doi.org/10.5539/cis.v5n3p110

3. Martin Zobl, Michael Geiger, Bjorn Schuller, Manfred Lang, Gerhard Rigoll "A REALTIME SYSTEM FOR HAND GESTURE CONTROLLED OPERATION OF IN-CAR DEVICES" Institute for Human-Machine Communication, Munich University of Technology, D-80290 Munchen
URL: gerhard.rigoll@ei.tum.de

4. Sam Dutton "Web RTC" Google Chrome Developer Relations URL: http://samdutton.com

5. Rik Cabanier, Eliot Graff, Jay Munro, Tom Wiltzius HTML Canvas 2D Context, Level 2 W3C Working Draft Google, Inc,29October2013URL:http://www.w3.org/TR/2013/WD2dcontext 220131029/

6. Jing Lin, Yingchun Ding "A temporal hand gesture recognition system based on hog and motion trajectory" Department of Physics, Beijing University of Chemical Technology, Beijing 100029, China ,23 May 2013 URL: www.elsevier.de/ijleo

7. Marco Roccetti , Gustavo Marfia, Angelo Semeraro "Playing into the wild: A gesture-based interface for gaming in public spaces" University of Bologna, Mura Anteo Zamboni 7, 40127 Bologna, Italy, 28 December 2011 URL: www.elsevier.com/ locate/jvci

8. Hanjie Wanga, Jingjing Fu , Yan Lu b, Xilin Chen a, Shipeng Li"Depth sensor assisted real-time gesture recognition for interactivepresentation" Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China, Microsoft Research Asia, Beijing 100080, China, 21 October 2013 URL: www.elsevier.com/ locate/jvci