Music Controller based on Hand Gestures using Webcam

[1]Abin Pal, [1]Hitesh Kumar and [2]K. John Singh
[1]School of Information Technology and Engineering,
VIT University, Vellore, Tamil Nadu, India.
[2]Assistant Professor (Selection Grade), School of Information Technology and Engineering
VIT University, Vellore, Tamil Nadu, India.

***Abstract*--Interaction of human beings with computers can be done without any physical interface and this can be done using gestures. The paper addresses a solution to access the music player in the computer via the webcam. This paper mainly deals with detecting and identifying the hand, the skin and the gesture to perform some specific tasks in the music player.**

Keywords -- interaction of human beings with computers, skin identification.

I. INTRODUCTION

Hand recognition and detection is a important research field by which a set of gestures can control a computer. The new generation laptops with built in webcam, touch mobile phones with front camera has increased the use of applications working on gestures. This paper is mainly focused in controlling the music player without the use of keyboard and mouse. Different algorithms are used to achieve the best performance from the system.

II. CLASSIFICATION OF GESTURES

Gestures can be static (certain fixed position), dynamic as well as static and dynamic. Sign languages follow static as well as dynamic gestures. For example, a static position of our hand can symbolize something, where as movement of our hand in a particular manner or direction can symbolize something different. Gestures can be divided into these following types:-

Head, face Gestures: Nodding of head in a particular direction is a head movement, while eye movements, eye brow movement, expressing happiness, sadness, surprised etc. are all examples of facial gestures. Figure 1(below) shows the various types of facial gestures given to computers as inputs.

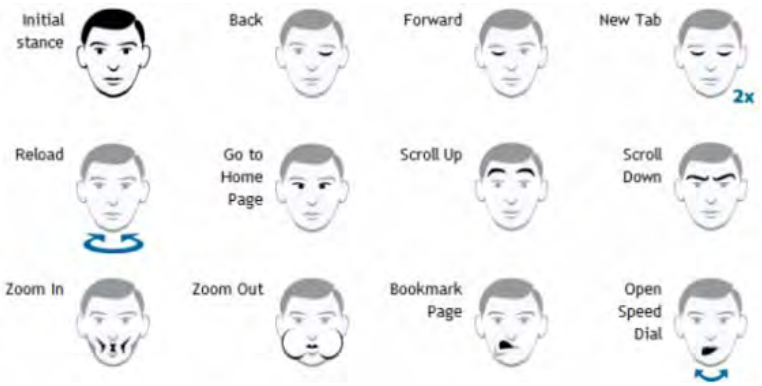

Figure 1 of Facial Expressions

Body Gestures: Movement of your body in expressing joy, fear, frustration, anger is all body gestures. The figure 2 (below) shows various body gestures.

Figure 2 of different body gestures

Hand Gestures: The different positions of the hand to show different signs and symbols are hand gestures. The various hand gestures areshown in Figure 3 (below).

Figure 3 of different hand gestures

III. EXISTING SYSTEMS

In the present situation of human-computer interaction, gesture is a major way to interact with a computer device. A gesture is a type of non-verbal communication which is especially done by the actions performed by the body parts. Gesture can be performed alone or together with verbal communication. It includes the movements of specially hands, eyes, face, arms, fingers, and head with the intention to convey some meaningful information or to interact with the system. In our paper we are mainly focused on the hand gestures. It means different kind of hand movements will lead to different types of input for a system.

Gesture is used to build a richer bridge between humans and machines than the old, primitive command line interface or even the recent graphical user interfaces. Using gestures, systems can be given different inputs and according to those defined inputs to can execute some particular tasks. Normally to capture these gestures a restricted background is required and primarily a camera to capture those images.

Gestures have been given as input to computer over the past 50 years. The use of mouse and pen input is one of the first examples of gesture based input. First a light pen was used and later its successor came as a mouse. The gesture and the movements done with the mouse is reflected in the screen object and hence direct manipulation can be done. Using light pen and mouse provided a lighter and simple way to give inputs to the interface.

Next came the Touch and Pressure inputs. The touch screen which came in the mid 80's was a transformed form of gesture inputs. These touch screen helped to give direct input to the interface saving a lot of time and making the task of giving input much more easy. Nowadays these types of touch screens are not only used in mobile computing but also in large desktops, table computers and many more.

Later came the electronic sensing systems using a suit that used to be mounted on the body. This helped to track the actions, space, orientation and position using some electro-magnetic sensors. These types of systems are used to give human movements to a robot which reflected human movements. This type of system improved the use of gestures to give varieties of inputs to machine.

The electronic sensing gloves were the next stage of this technology with the use of sensor enabled gloves. This type of gloves gave a detailed representation of the palm movement, individual finger movements. The sensor enabled gloves gave accurate hand position, the movements of the fingers and its orientation. The first commercial electronic sensing gloves was the Dataglove described by Zimmerman, Lanier, Blanchard, Bryson and Harvill (1987), and is illustrated in the video by Zacharey, G. (1987). In this type of technology optical fibers were used which were placed at the back of each hand and hence a light bent in the hand leaked the light out of a small crack present in the fiber, and upon the measurement of the loss of light an accurate reading of the hand pose was measured. The Dataglove (Figure 4) was able to measure a hand bent of 10 degrees but the sideways movement of hands was not measurable. But later the development of Cyberglove (Figure 5) by Kramer was a much successful one as it was able to measure the bends of the hand much more accurately.

Figure 4 Dataglove

Although the electronic sensing gloves gave accurate results but were very costly and hence vision techniques came into use. This technique helped to capture not only hand movements but also body movements. There were many drawbacks in this technology such as less accuracy, low resolution cameras and low frame capture rate were unable to track the quick motions of the hands.

Sensor embedded object and tracking devices using sensors also came into use. Infrared rays were emitted from the devices which were tracked by a camera and the actions, gestures performed by the device are translated into different pre defined system behaviors.[4]The use of remote control is a similar type of operation performed in regular use. Some types of receiving devices are kept on the other side which receives the gesture inputs given by the device.

Another type of sensor is the audio sensors which are used to detect the location of a knock or tap on the public display screen. This mainly gestures are given by the finger or hand movements. All these gestures were given by physical contact to a physical device but there are some systems which re designed to take inputs not by any physical contact.

Figure 5 Cyberglove

IV. CHALLENGES FACED IN GESTURE RECOGNITION

To recognize a gesture the most important requirement is the tracking technology used to capture the gestures given as inputs and finally process the inputs and perform some specific task. This technology mainly is based on the positions of the gestures given as inputs.

Most common gestures are hand gestures and the most common technique used to capture different hand gestures are by wearing a glove on the hands which gives an accurate position of the hand, fingers and its orientation. This hand tracking gloves were known as Data-gloves.[10] Although the instrumented gloves have been must more costly but the results were highly accurate. But in vision based gestures, such as eye ball movement could only be tracked using a three-dimensional mapping of the eye pose.[11] A more efficient gesture is recorded using proximity sensors which maps the natural, dynamic gestures and hence gives accurate inputs to the machine.

The main challenge faced in image based gesture is that the videos or images should always be shot in a prominent light environment. The more the background noises the difficulty in recognizing the gesture becomes tough. Further, the camera's quality, resolution and the distance from the object to the camera depends on the recognition of the gesture.

V. ARCHITECTURE OF THE PROPOSED SYSTEM

A basic gesture input device consists of several parts. Firstly the hand gestures given by human beings are taken as input by an input device. Then that input is recognized by the central processing unit of the machine which is compared with the set of gestures present in the memory or with the type of action performed in the gesture, and if the gesture given as input is a gesture recognized by the system then the system gives some specific output. This looks simple but when capturing only the gestures becomes a tough part.[8]The hand gesture is captured through a camera connected to the computer and that gesture is processed and the resultant output is reflected.

The most important thing in hand detection is detecting the skin pixels and the non- skin pixels in the image captured.[2] Normally human skin color lies within a specified range. We are using HSV color space representation for detection the hand. We are removing the V component and we are focusing on the H component (Hue component) and the S component (Saturation component).[6] 10 or many snapshots are taken of the movement of the hand and then is converted into binary mask where the black pixels represent the non-skin pixels and the white pixels represent the skin pixels. Every time the snapshots are taken, the images are converted into HSV color space representation and is checked whether there is a change in the gesture or not.[3] If there is a change in the gesture, the specified task for that gesture is performed.
The hand gestures used in this paper are:-

1. Play / Pause

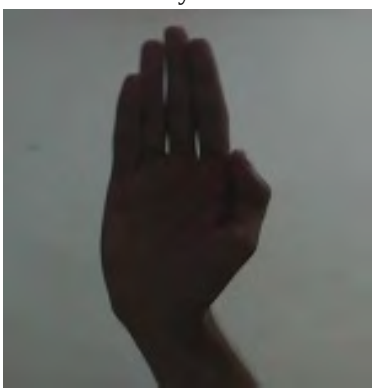

Figure 6 Play/Pause gesture

2. Stop

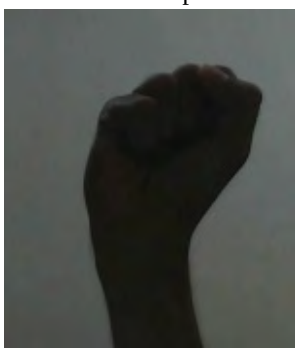

Figure 7 Stop gesture

3. Volume-Up

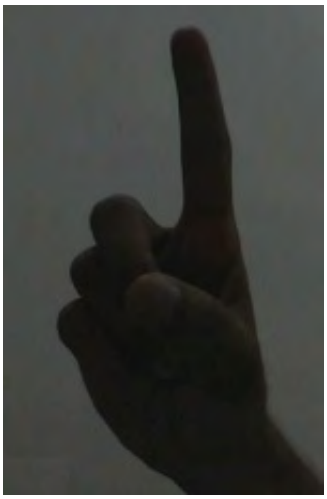

Figure 8 Volume up gesture

4. Volume Down

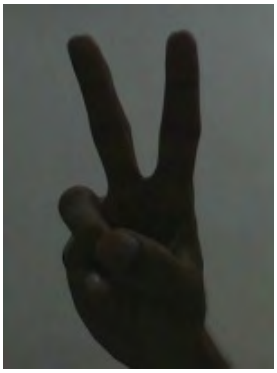

Figure 9 Volume down gesture

5. Full screen

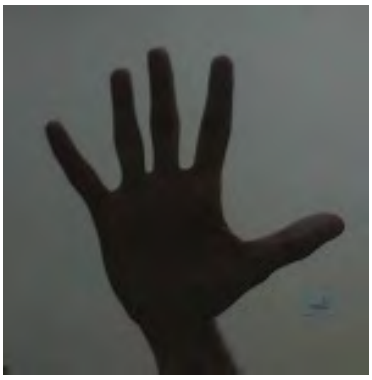

Figure 10 Full screen gesture

The flowchart followed by the technology in this paper is as follows:-

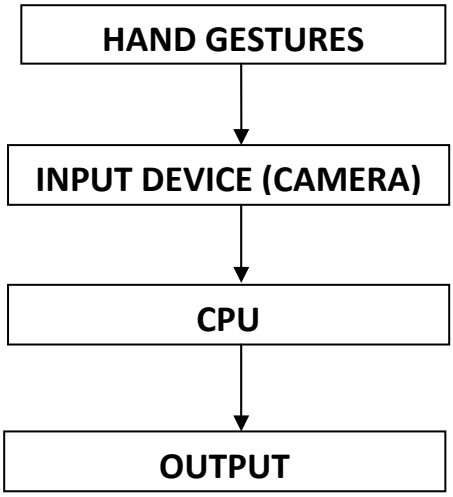

Figure 11 Flowchart

VI. NEED OF GESTURE RECOGNITION TECHNOLOGY

The traditional two-dimensional user interface is not flexible enough for virtual environments. Virtual Environments utilizes different technologies and hence integrate them to create a user interface. It uses different devices to sense the body positions, movements, speech, and sound to communicate to the device. Optical Character Reader (OCR) was the first mechanical or electronic conversion of human written paper which was later used for a reading machine for the blind. In fact nowadays the OCR technology is used for real-time hand writing in the mobile phones, tablets and touch screen devices.

Virtual Environments (VE) are to provide efficient, flexible and strong interactive interface so that the user can be able to give different types of inputs. Human Gestures are very much natural and hence to give much more flexibility to the interface of the system VE's are used. The traditional physical mouse and keyboard are not well enough suited for VE's. This type of environments provide the opportunity to utilize several various technologies and to integrate them into the user interface for much more effective output.[8] Nowadays many devices are present by which the orientation, position of the body, the eye movements, the facial expressions, the gestures and postures are captured create a communication between the system environment and humans. Combinations of communication modalities and different sensor technologies produces a wide range of uni-modal and multi-modal[9] interface techniques[7]. The potential of this types of systems are proving to be very high and impressing to new world of science.

Gesture is used for control and navigation in many Virtual Environments (VEs) and in CAVEs (Cave Automatic Virtual Environments) such as virtual work stations, smart rooms, and many more. In addition, gesture may be perceived by the environment in order to be transmitted elsewhere (e.g., as a compression technique, to be reconstructed at the receiver derived from those measurements, such as velocity and acceleration.

TESTING AND ANALYSIS

We have used the HSV color representation on the captured images, and then have removed V and mainly focused on H component (Hue Component) and S component (Saturation Component). We have captured more than 10 images to recognize the gesture and the test results are as follows:-

Type of Gesture	Images Captured	No. of successful hits	No. of missed hits	Rate of recognition (%)
Play/Pause	10	8	2	80
Stop	10	9	1	90
Volume Up	10	7	3	70
Volume Down	10	8	2	80
Full screen	10	8	2	80

Table1: Recognition of Hand gestures results

The rate of recognition is not 100% because of the background in which the gesture is given. Due to the noisy background or any skin interference in the background other than the hand stops the system from being 100% accurate.[1]

VII. CONCLUSIONS

In the present world many ways are available for providing input to the system. Many of them requires physical touch and some requires physical objects like mouse, keyboard. But very few applications are available that controlled and working with using hand gesture. By using this hand gesture technique user can interact with the system without any physical touch. In this paper we have developed a method to recognize the hand movement, gesture by using webcam. Since there are variety of hand gesture and the transition between states is necessary in each gesture for an effective hand interaction we are capturing more than 10 images of the gesture and by the use of HSV color representation we are detecting the hand in the image and its movement. This paper reveals a system which uses the webcam for capturing the gestures, identifying them and finally accessing the music player.

VIII. REFERENCES

[1] Erdem Yoruk, Ender Konukoglu, Bulent Sankur, and Jerome Darbon, “Shape-Based Hand Recognition”, IEEE Transactions of Image Processing, VOL. 15, NO. 7, July 2006.

[2] R. Lockton and A. Fitzgibbon, “Real-time gesture recognition using deterministic boosting”, In Proc. of British Machine Vision Conference, 2002

[3] Mathias Kolsch and Matthew Turk, “Analysis of Rotational Robustness of Hand Detection with a Viola-Jones Detector”, Department of Computer Science, University of California Santa Barbara, CA

[4] P. Viola and M. Jones, “Robust Real-time Object Detection”, Int. Journal of Compuer Vision, 2002.

[5] Buxton, W., Hill, R., and Rowley, P. 1985. Issues and techniques in touch-sensitive tablet input. In Proceedings of the 12th annual conference on Computer graphics and interactive techniques. ACM Press, 215–224.

[6] Eng-Jon Ong and Richard Bowden, “A Boosted Classifier Tree for Hand Shape Detection”, Center for Vision, Speech and Signal Processing, University of Surrey, Guildford, GU2 7XH.

[7] Eisenstein, J. and Davis, R. 2004. Visual and linguistic information in gesture classification. In Proceedings of the 6th international conference on Multimodal interfaces. ACM Press, 113–120.

[8] Shamaie, A, Hai, W and Sutherland, A. 2001. Hand gesture recognition for HCI”, *ERCIM News (on line edition),http://www.ercim.org/publication/Ercim News*, no.46, (2001).

[9] Bolt, R. A. and Herranz, E. 1992. Two-handed gesture in multi-modal natural dialog. In Proceedings of the 5th annual ACM symposium on User interface software and technology.ACM Press, 7–14.

[10] Digital Image Processing by Rafael C. Gonzalez (Book)

[11] Gesture Recognition: A Survey, IEEE transactions on systems, man and cybernetics by Tinku Acharya and Sushmita Mitra