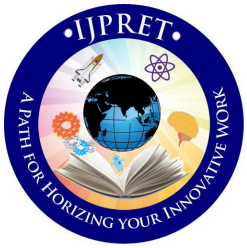
Research Article *ISSN: 2319-507X*
Ashwini Parate, IJPRET, 2013; Volume 1(8): 754-761 *IJPRET*



INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

ROBOT AND HOME APPLIANCES CONTROL USING GESTURE AND SPEECH RECOGNITION

ASHWINI PARATE, PRIYANKA CHANDORE, KRISHNA KANT, GAURAV RAUT

UG Scholar, Department of Information Technology, SITRC, Sandip Foundation, Nasik (M.S.)

Accepted Date:

27/02/2013

Publish Date:

01/04/2013

Keywords

Hand gestures,

Microsoft's SAPI,

XML.

Corresponding Author

Ms. Ashwini Parate

Abstract

Human computer interaction has greatly evolved through a number of technologies like *gesture recognition, speech recognition, voice recognition*, etc. Hand gesture recognition and speech recognition technology are the technologies which are intuitively used today. Hand gestures are easy to use as hand is the exposed part of the human body most of the times. It is more convenient to use hand gestures as the aim can be achieved by sitting at a single place. Hand gestures have been already implemented in many technologies today, like, gaming, sign language recognition, Human-robot interaction, etc. Speech recognition is also a widely used technology. Implementing speech recognition is not at all an issue since all that it requires is a headphone. These two technologies are being integrated. The aim it at managing the motions of a robot and the system cursor by tracking two colors namely red and green in the hand gestures another one is implementing a system which aims at controlling electrical appliances and operating system using speech recognition using SAPI.

I. INTRODUCTION

In section 1, a brief is being told about the two technologies namely, *Gesture Recognition and Speech Recognition*, and a brief introduction of the proposed system. In section 2, the related work done in these fields in early period is mentioned. In section 3, the actual methodology by which it can be implemented is mentioned. Section 4 gives us the analysis of the results that are obtained.

Gesture Recognition is a technology used for Human Computer Interaction. Gesture recognition includes human being making gestures are recognized by the system. Communication with the system through gestures is quite easy as compared to the traditional and obsolete interaction methods such as keyboard and mouse. Hand gestures has a natural ability to represent ideas and actions very easily, thus using these different hand shapes, being identified by gesture recognition system and interpreted to generate corresponding event, has the potential to provide a more natural interface to the computer system [1].

For controlling the home appliances and the operating system we are using speech recognition technology. Speech recognition is actually a technology in which the user gives some command to the system and the system performs some specific action according to the grammar files. A database is maintained in which a mapping of command to action is prepared. In our work, user command is used to activate the home appliances as well as to manage operating system. SAPI (Microsoft's Speech Application Programming Interface) is used for this purpose. SAPI is a middleware that provides an API and a Device Driver Interface (DDI) for speech engines to implement. A person may find it difficult to handle a system by traditional means such as mouse or keyboard. And in such situations Speech Recognition technology can be very much easier.

II. LITERATURE SURVEY

Since past decades there are many techniques by which gesture recognition has been implemented. One of them is the wired technology. In this technology the

user is tied up to a wire. The wire acts as the interface between the human and the user. A major drawback of the system is that there is a physical medium between the user and the system hence the user will not be able to move freely in the room. Instrumented gloves or data gloves are used for wired communication. Some type of electrical sensors is fitted /mounted into/upon the gloves. These sensors provide us the information regarding the position, length of the hand. Data gloves also provide great amount of efficiency but a major drawback is that they are expensive to be used in a wide range of applications. Data gloves have been replaced by optical markers. These optical markers project Infra-Red light and reflect this light on screen to provide the information about the location of hand or tips of fingers wherever the markers are wear on hand, the corresponding portion will display on the screen. These systems also provide the good result but require very complex configuration [2]. There were even some advanced techniques that were introduced which were based on image processing. Like processing the image based on its color, texture, etc. But this technique did not come out to be as efficient because the color or texture of skin may vary from person to person. Results may also vary based on certain illumination conditions.

There is another technique in which the K-means clustering algorithm is used. In this, the input sequence of RGB images gets converted to YCbCr images. Image segmentation is typically performed to locate the hand object in a particular image. The image is segmented into K clusters by means of the K clustering algorithm. The cluster is mainly divided into two clusters where Cluster1 represents the hand object in which all pixel values are set to 1, whereas the second cluster represents the background portion where all the pixel values are set to 0.And then apply filling of holes on binary image. After hand segmentation is done we need to calculate the boundary contours of the image so as to locate the hand region in image. For this purpose the image is scanned from left to right and top to bottom. While scanning left to right the first white pixel that is encountered is considered as the left boundary. While scanning right to left the

first white pixel encountered is considered as the right boundary. And same is followed while scanning from top to bottom and bottom to top.

Another algorithm called Fuzzy-C-means clustering algorithm for classifying both dynamic and static gestures is present. A static hand gesture recognition algorithm for robot control was proposed by Juan P. Wachs et al [4] [5]. Jae-Ho Shin [6] used entropy analysis to extract hand region in complex background for hand gesture recognition system. In [7], a vision-based hand pose recognition technique using skeleton images is proposed, in which a multi-system camera is used to pick the centre of gravity of the hand and points with farthest distances from the centre, providing the locations of the finger tips, which are then used to obtain a skeleton image, and finally for gesture recognition.

VoiceXML is used as one of the recent approaches to use voice recognition as the technique to implement human computer interaction. The origins of VoiceXML began in 1995 as an XML-based dialog design language intended to simplify the speech recognition application development process within an AT&T project called Phone Markup Language (PML) [3].As AT&T reorganized teams at AT&T, Lucent and Motorola continued working on their own PML-like languages [3].In 1998, W3C hosted a conference on voice browsers. By this time, AT&T and Lucent had different variants of their original PML, while Motorola had developed VoiceXML, and IBM was developing its own SpeechML. Many other attendees at the conference were also developing similar languages for dialog design; for example, such as HP's TalkML and PipeBeach's VoiceHTML [3].The VoiceXML Forum was then formed by AT&T, IBM, Lucent, and Motorola to pool their efforts. The mission of the VoiceXML Forum was to define a standard dialog design language that developers could use to build conversational applications [3]. They chose XML as the basis for this effort because it was clear to them that this was the direction technology was going [3].In 2000, the VoiceXML Forum released VoiceXML 1.0 to the public. Shortly thereafter, VoiceXML 1.0 was submitted to the W3C as the basis for the creation of a new international

standard [3]. VoiceXML 2.0 is the result of this work based on input from W3C Member companies, other W3C Working Groups, and the public [3].The only limitation of VoiceXML is it works only with websites. Hence SAPI is the best choice for it.

III. METHODOLOGY

Integration of two technologies namely gesture recognition and speech recognition is proposed, block diagram of system is as shown in the figure 1.

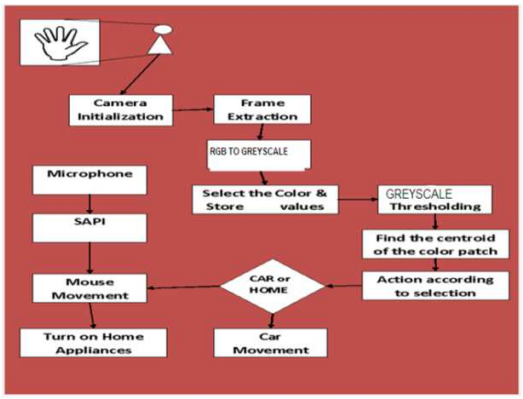


Fig 1- Block Diagram of proposed system

Fig 1 describes that a single camera or a webcam has been used in the system; the webcam can be of any type. But the main problem with the webcams is that they come with different resolutions. This may create some complexions. Hence it has been decided to convert every image into a fixed resolution. Any resolution of the webcam has to be converted into this fixed resolution. We convert the RGB color space to HSV color space. RGB color space is mostly used by many of the computing devices. But HSV is the color space which corresponds more to the human perceptions. The formulas used to convert RGB to HSV depend on which of the RGB components is largest and which is smallest, and are described at Transformation from RGB to HSV. The HSV color space is user-friendlier [1]. After the conversion from RGB to HSV select a single color from the image which is used to track the hand movement. A color threshold is performed on the image.

User gives the input through a microphone. Microphone processes the audio stream given by the user and converts it into the some form of digital data. This digital data is converted into phonemes and then we get a text command. Phonemes are linguistic units. They are the sounds that group together to form our words, although how a phoneme converts into sound depends on many factors including the surrounding

phonemes, speaker's accent and age[8][9].These phonemes are extracted by Microsoft speech SDK[8]. Hidden Markov Model (HMM) is used to convert the phonemes into commands. Phonemes are extracted from the given input. A Markov Model (in a speech recognition context) is basically a chain of phonemes that represent a word [8]. A Markov Model can be shown in Figure 2. The command which is generated from speech recognition is synthesized further for context searching. The command is compared to the context database and then the action is performed accordingly.

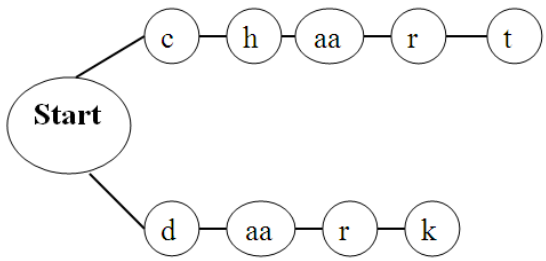


***Fig2*- A Markov Model**

IV. RESULT AND ANALYSIS

Here with two colors for tracking the movement of hand i.e. Red and Green. Red for the up-down motion and green for left-right motion. Our results have been shown in the following figure 2, figure 3 and figure 4.

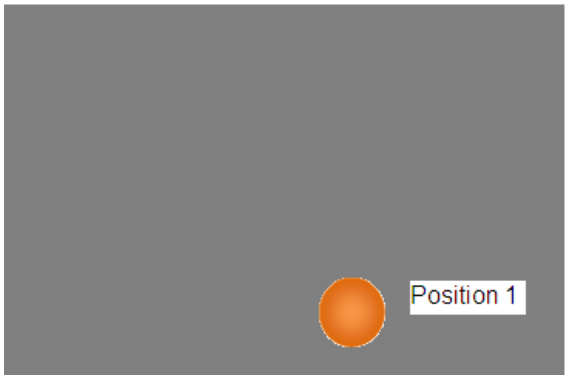


Fig- 2: Position 1(position of hand)

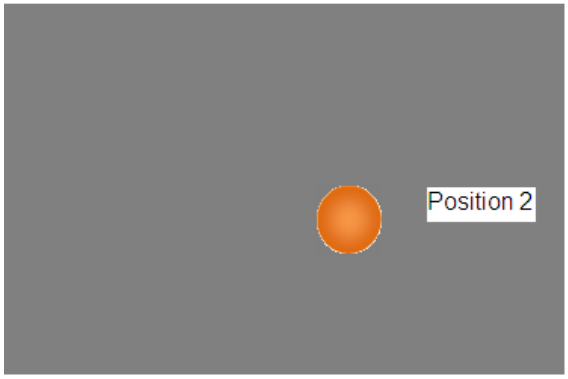


Fig- 3: Position 2(position of hand)

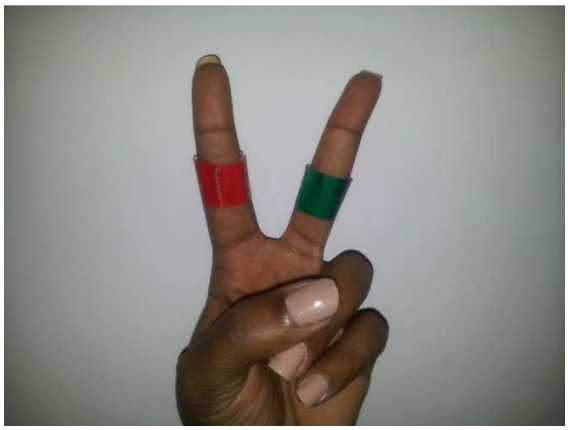


Fig 4- Colors used from tracking

For Representation purpose, we have used an orange spot in our GUI to track the motions of our hand. And as shown in Figure 4, two colors namely red and green are used.

V. CONCLUSION

To overcome many drawbacks of the traditional systems proposed system can be of great use for disabled people by successful implementation of gesture and speech recognition with control the mouse pointer and robot movement by gesture recognition. Operating system and the home appliances are managed using speech recognition.

VI. REFERENCES

1. Prof. Yuvraj V. Parkal "Gesture Based Operating System Control" ,Electronics and Telecommunication department ,College of Engineering, Malegaon (Bk),Maharashtra, India

2. Meenakshi Panwar "Hand Gesture Recognition based on Shape Parameters", Centre for Development of Advanced Computing, Noida, Uttar Pradesh, India

3. Voice Extensible Markup Language (VoiceXML) Version, 16 March 2004.

4. Juan P. Wachs, Helman Stern and Yael Edan, "Cluster Labeling and Parameter Estimation for the automated setup of a Hand gesture Recognition System," IEEE Trans. Systems and Humans, vol. 35, no. 6, pp. 932-944, Nov. 2005.

5. J. Wachs, H. Stern and Y Eden, "Parameter search for an image Processing Fuzzy c-Means hand gesture recognition system," Proc.IEEE Int.Conf.Image Processing, Barcelona, Spain, vol. 3, pp. 341-344, 2003.

6. Didier Coquin, Eric Benoit, Hideyuki Sawada, and Bogdan Ionescu, “Gestures Recognition Based on the Fusion of Hand Positioning and Arm Gestures”, Journal of Robotics and Mechatronics Vol.18 No.6, 2006. pp. 751-759.

7. A. Utsumi, T. Miyasato and F. Kishino, “Multi-Camera Hand Pose Recognition System processing Skeleton Image", IEEE International Workshop on Robot and Human Communication, pp. 219-224, 1995.

8. Anil J Kadam, Pallavi Deshmukh, Amita Kamat, Neelam Joshi, Ritika Doshi,” Speech Oriented Computer System Handling”, International Conference on Intelligent Computational Systems (ICICS'2012) Jan. 7-8, 2012 Dubai

9. Md. Abdul Kader, et al “Speech Enabled Operating System Control” (ICCIT 2008), IEEE.