Multi Touch Gesture Technology for Interactive HCI

Aditya Mhatre[#1]

#Dept. of Technology, University of Mumbai
M.G. Road, Fort, Mumbai, Maharashtra, India-400 032
adjmhatre@gmail.com[1#]

Abstract— **We propose the combinational use of hand gestures & simulating software which provides an attractive alternative to cumbersome interface devices for human computer interaction (HCI). Many hand gesture recognition methods using visual analysis have been proposed: syntactical analysis, neural networks, the hidden Markov model (HMM). In the pre-processing stage, our approach consists of three different procedures for hand localization, hand tracking and gesture spotting. The hand location procedure detects hand candidate regions on the basis of skin-color and motion. The hand tracking algorithm finds the centroids of the moving hand regions, connects them, and produces a hand trajectory. The gesture spotting algorithm divides the trajectory into real and meaningless segments. To construct a feature database, this approach uses a combined and weighted location, angle and velocity feature codes.**

Keywords -Image processing, Gesture recognition, Colour Thresholding, Matlab, and CIVB.

I. INTRODUCTION

Hand gesture recognition using visual devices has a number of potential applications in HCI (human computer interaction), VR (virtual reality), and machine control in the industrial field. Most conventional approaches to hand gesture recognition have employed such external devices as data gloves and color makers. For a more natural interface, however, hand gesture must be distinguishable from visual images without the aid of any external device.

A. Sixth Sense Technology

Sixth Sense in scientific (or non-scientific) terms is defined as Extra Sensory Perception or in short ESP. It involves the reception of information not gained through any of the five senses. Nor is it taken from any experiences from the past or known. Sixth Sense aims to more seamlessly integrate online information and tech into everyday life. By making available information needed for decision-making beyond what we have access to with our five senses, it effectively gives users a sixth sense.

B. Gesture Recognition

Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Current focuses in the field include emotion recognition from the face and hand gesture recognition. Many approaches have been made using cameras and computer vision algorithms to interpret sign language.

Gestures can exist in isolation or involve external objects. Free of any object, we wave, beckon, fend off, and to a greater or lesser degree (depending on training) make use of more formal sign languages. With respect to objects, we have a broad range of gestures that are almost universal, including pointing at objects, touching or moving objects, changing object shape, activating objects such as controls, or handing objects to others. Gesture recognition can be seen as a way for computers to begin to understand human body language, thus building a richer bridge between machines and humans than primitive text user interfaces or even GUIs (graphical user interfaces), which still limit the majority of input to keyboard and mouse. Gesture recognition enables humans to interface with the machine (HMI) and interact naturally without any mechanical devices.

Gestures can be used to communicate with a computer so we will be mostly concerned with empty handed semiotic gestures. These can further be categorized according to their functionality.

C. Computer Vision Based Algorithm

Computer vision is the science and technology of machines that see. As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner.

Computer vision, on the other hand, studies and describes the processes implemented in software and hardware behind artificial vision systems. The software tracks the user's gestures

using computer-vision based algorithms. Computer vision is, in some ways, the inverse of computer graphics. While computer graphics produces image data from 3Dmodels, computer vision often produces 3D models from image data. There is also a trend towards a combination of the two disciplines, e.g., as explored in augmented reality.

The fields most closely related to computer vision are image processing, image analysis and machine vision. Image processing and image analysis tend to focus on 2D images, how to transform one image to another. His characterization implies that image processing/analysis neither require assumptions nor produce interpretations about the image content.

D. *The Recognition Algorithms*

The computer vision system for tracking and recognizing the hand postures that control the menus is based on a combination of multi-scale color feature detection; view based hierarchical hand models and particle filtering. The hand postures or states are represented in terms of hierarchies of multi-scale color image features at different scales, with qualitative inter-relations in terms of scale, position and orientation. In each image, detection of multi-scale colour features is performed. The hand postures are then simultaneously detected and tracked using particle filtering, with an extension of layered sampling referred to as hierarchical layered sampling.

E. *Hand Segmentation*

Since, efficient hand tracking and segmentation is the key of success towards any gesture recognition, due to challenges of vision based methods, such as varying lighting condition, complex background and skin color detection; variation in human skin color complexion required the robust development of algorithm for natural interface.

II. LITERATURE SURVEY

With the massive influx of computers in society, human computer interaction, or HCI, has become an increasingly important part of our daily lives. It is widely believed that as the computing, communication, and display technologies progress even further, the existing HCI techniques may become a bottleneck in the effective utilization of the available information flow. For example, the most popular mode of HCI is based on simple mechanical devices— keyboards and mice. These devices have grown to be familiar but inherently limit the speed and naturalness with which we can interact with the computer. This limitation has become even more apparent with the emergence of novel display technology such as virtual reality [1], [2], [3].

Thus in recent years there has been a tremendous push in research toward novel devices and techniques that will address this HCI bottleneck. One long-term attempt in HCI has been to migrate the "natural" means that humans employ to communicate with each other into HCI. With this motivation automatic speech recognition has been a topic of research for decades. Tremendous progress has been made in speech recognition, and several commercially successful speech interfaces have been deployed. However, it has only been in recent years that there has been an increased interest in trying to introduce other human-to-human communication modalities into HCI. This includes a class of techniques based on the movement of the human arm and hand, or hand gestures. Human hand gestures are a means of non-verbal interaction among people. They range from simple actions of using our hand to point at and move objects around to the more complex ones that express our feelings and allow us to communicate with others. To exploit the use of gestures in HCI it is necessary to provide the means by which they can be interpreted by computers. The HCI interpretation of gestures requires that dynamic and/or static configurations of the human hand, arm, and even other parts of the human body, be measurable by the machine.. This has spawned active research toward more "natural" HCI techniques.

III. PROBLEM DEFINITION

This project will design and build a man-machine interface using a video camera to interpret the hand gestures (plus others for additional keyboard and mouse control).
The keyboard and mouse are currently the main interfaces between man and computer.
Humans communicate mainly by vision and sound, therefore, a man-machine interface would be more intuitive if it made greater use of vision and audio recognition. Another advantage is that the user not only can communicate from a distance, but need have no physical contact with the computer. However, unlike audio commands, a visual system would be preferable in noisy environments or in situations where sound would cause a disturbance.
The visual system chosen was the recognition of hand gestures. The amount of computation required to process hand gestures is much greater than that of the mechanical devices, however standard desktop computers are now quick enough to make this project —hand gesture recognition using computer vision — a viable proposition.
In this project will be using next generation features i.e. hand gestures for controlling computer functions like navigation, zoom in/out, changing of slides etc. without actual touch to the screen or device.
These features can be implemented using image processing.

IV. CONSTRUCTION AND WORKING

A. Software And Hardware Requirements

1) Webcam

Figure 1: TWO DIMENSIONAL WAVELET TRANSFORM

A webcam captures and recognizes an object in view and tracks the user's hand gestures using computer-vision based techniques.

It sends the data to the computer. The camera, in a sense, acts as a digital eye, seeing what the user sees. It also tracks the movements of the thumbs and index fingers of both of the user's hands. The camera recognizes objects around you instantly.

Single camera system:

This system would provide considerably less information about the hand. Some features (such as the finger against a background of skin in the example above)would be very hard to distinguish since no depth information would be recoverable.

Essentially only "silhouette" (Detection of all skin within the hand without any feature detection information could be accurately extracted. The silhouette data would be relatively noise free (given a background sufficiently distinguishable from the hand) and would require considerably less processor time to compute than either multiple camera system. It is possible to detect a large subset of gestures using silhouette information alone and the single camera system is less noisy, expensive and processor hungry. Although the system exhibits more ambiguity Stereographic system and multiple two dimensional view system this disadvantage is more than outweighed by the advantages mentioned above. Therefore, it was decided to use the single camera system.
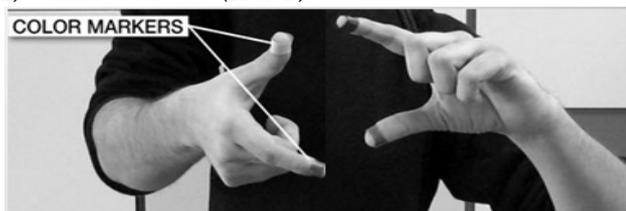
2) Color stickers (R G B)

Figure 2: COLOR MARKERS

It is at the tip of the user's fingers. Marking the user's fingers with red, green, and blue tape helps the webcam recognize gestures. The movements and arrangements of these Markers are interpreted into gestures that act as interaction instructions for the projected application interfaces.

3) Matlab Software

MATLAB (matrix laboratory)   is   a numerical computing environment and fourth-generation programming language. Developed by Math Works, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran.

Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the Mu-PAD symbolic engine, allowing access to symbolic computing capabilities. An additional package, Simu-link, adds graphical multi-domain simulation and Model-Base design for dynamic and embedded systems.

4) Image processing tools

These tools can be used in the area of image processing which involves using algorithms to detect and isolate various desired portions or shapes (features) of a digitized image or video stream. It is particularly important in the area of optical character recognition.
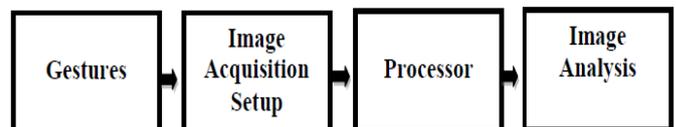
B. Construction

Figure 3: BLOCKS DIAGRAM

Image acquisition setup: It consists of a web camera with suitable interface for connecting it to PC.

Processor: It consists of personal computer or a dedicated image processing unit.

Image analysis: Certain tools are used to analyses the content in the image captured and derive conclusions e.g. Matlab 7.0

C. Working

The Sixth Sense prototype comprises computers, stickers and a camera. The camera recognizes and tracks users' hand gestures and physical objects using computer-vision based techniques. The software program processes the video stream data captured by the camera and tracks the locations of the colored markers at the tips of the user's fingers. We use colored caps on

fingers so that it becomes simpler for the software to differentiate between the fingers, demanding various applications. The movements and arrangements of these fiducials are interpreted into gestures that act as interaction instructions for the projected application interfaces.

The software program analyses the video data caught by the camera and also tracks down the locations of the coloured markers by utilizing single computer vision techniques. One can have any number of hand gestures and movements as long as they are all reasonably identified and differentiated for the system to interpret it, preferably through unique and varied fiducials.

The software recognizes Multi-touch gestures, like the ones you see in Microsoft Surface or the IPhone where you touch the screen and make the map move by pinching and dragging. The technology is mainly based on hand gesture recognition, image capturing, processing, and manipulation, etc.

This project lets the user zoom in, zoom out or drag and navigate using intuitive hand movements.
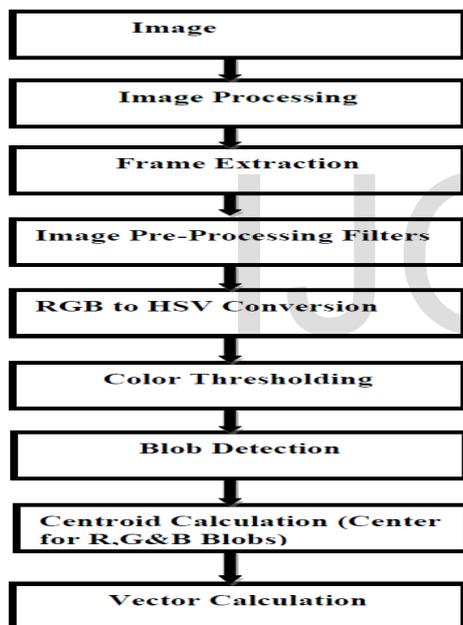
V. FLOWCHART

Figure 4: FLOW CHART

A. Frame Extraction

• In each picture we will see three basic colors viz. red, green, and blue. The process of frame extraction is used to separate these frames according to their colour.

• Extracting images from a video depends upon the frames we are considering per second and then using that frame to output an image. So, here we need to control the frame rate,

image format and in case you want a specific resolution of the image, you can do that by setting the frame size.

B. Image Pre-Processing Filters

• Contains numerous image filters for image optimization
• Miscellaneous filters for edge enhancement, noise suppression, character modification etc.
• Includes several functions for image processing
• Contrast increase by static or dynamic binarisation, look-up tables or image plane separation
• Resolution reduction via binning
• Image rotation
• Conversion of colour images to gray value images.

C. RGB to HSV Conversion

• The RGB color space, used directly by most computer devices, expresses colors as an additive combination of three additive primary colors of light: red, green, and blue.
• A commonly used color space that corresponds more naturally to human perception is the HSV color space, whose three components are hue, saturation, and value.
• The raw data provided by the video card was in the RGB (red, green, blue) format. However, since the detection system relies on changes in color (or hue), it could be an advantage to use HSV to permit the separation of the hue from luminosity (light level).
• To test this, the maximum and minimum HSV pixel color values of a small test area of skin were manually calculated. These HSL ranges were then used to detect skin pixels in a subsequent frame (detection was indicated by a change of pixel color to white). The test was carried out three times using either hue, saturation or
• Luminosity colour ranges to detect the skin pixels.
• RGB to HSV conversion is used in computer vision and image analysis for feature detection or image segmentation.

D. Color Thresholding

• Color Thresholding is used for defining the ranges of colors used.
• The Color Threshold module is used to remove parts of the image that fall within a specified color range. This module can be used to detect objects of consistent color values.

Figure 5: COLOR THRESHOLDING

E. *Blob Detection*

•In the area of computer vision, blob detection refers to visual modules that are aimed at detecting points and/or regions in the image that differ in properties like brightness or color compared to the surrounding.

•It reduces the noise of picture frame.

F. *Centroid Calculation*

• Centroid means the geometric center of the object's shape.

•The position of the colors is recognized using the centroid calculation in which the centroid of each colour stickers is calculated separately.

G. *Vector Calculation*

• Vector calculation basically means the coordinates of last few locations that give us information about patterns.

• On calculation of vectors proper actions or events can be executed or performed.

VI. ALGORITHMS

A. *Navigation*

• The setting for navigation is such that only one colour is detected while the others are not.

•To move the cursor the index finger movement is tracked and the cursor moves accordingly.

•For single click if index finger is in upper position twice and downward position once in succession it is detected as a single click.

•For double click if index finger is in upper position thrice and downward position twice in succession it is detected as a double click.

VII.   FEATURES AND APPLICATIONS

A. *Features*
1) *Navigation*
Navigation is a field of study that focuses on the process of monitoring and controlling the movement of the cursor from one place to another on the screen.

B. *Application Areas Of Hand Gestures System*
Hand gestures recognition system has been applied for different applications on different domains, as mentioned in including; sign language translation, virtual environments, smart surveillance, robot control, medical systems etc. overview of some hand gesture application areas are listed below

1)   *Sign Language Recognition:*

Since the sign language is used for interpreting and explanations of a certain subject during the conversation, it has received special attention. A lot of systems have been proposed to recognize gestures using different types of sign languages [6]. For example [6] recognized American Sign Language ASL using boundary histogram, MLP neural network and dynamic programming matching.

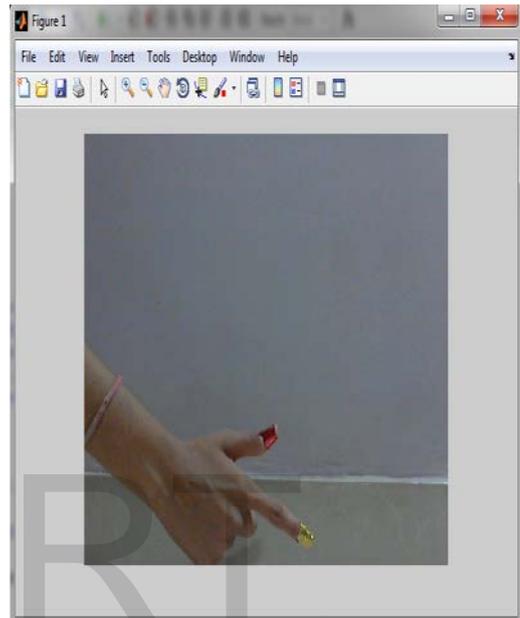## VIII. EXPERIMENTAL OBERVATIONS & RESULTS
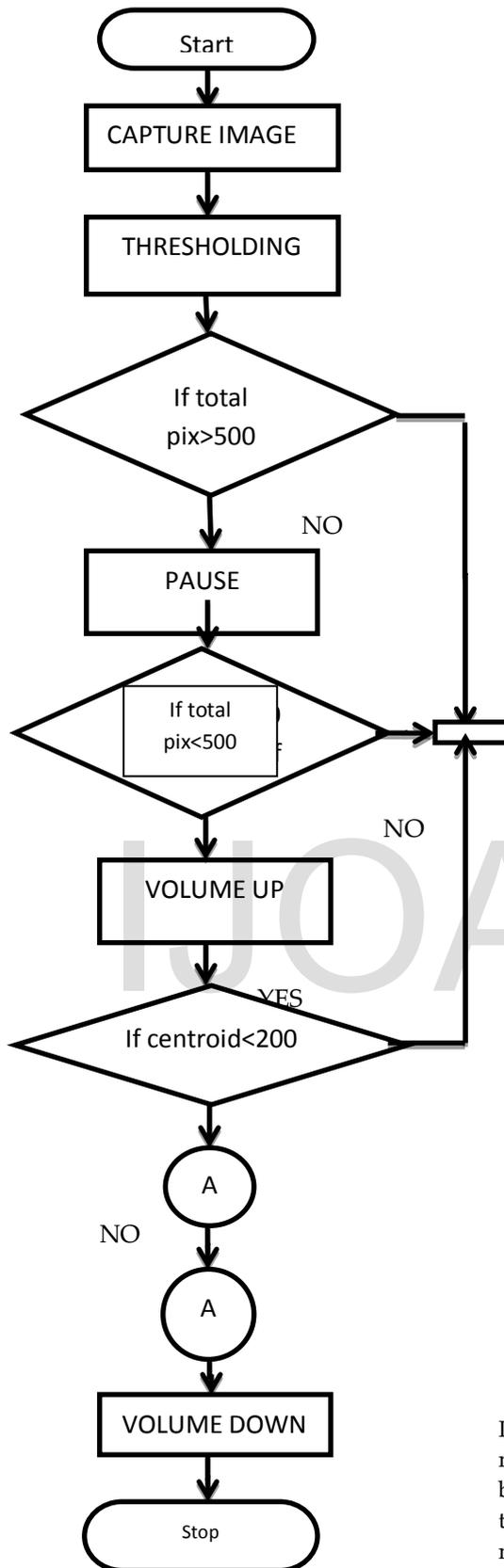

Figure 6: IDENTIFICATION OF COLOR MARKERS
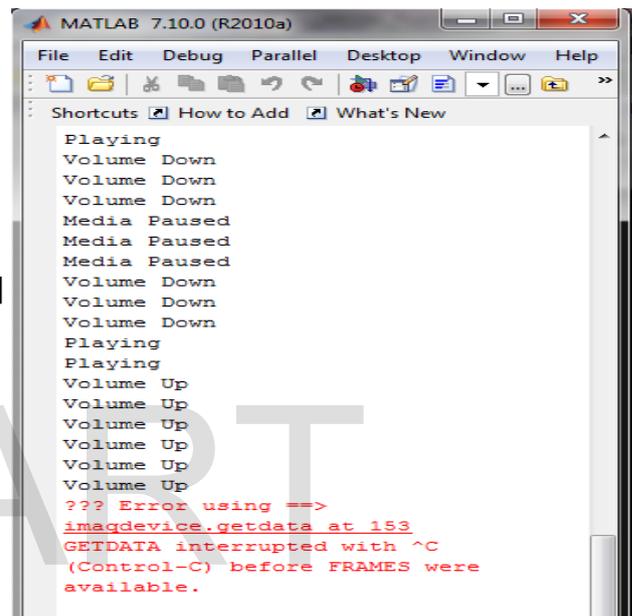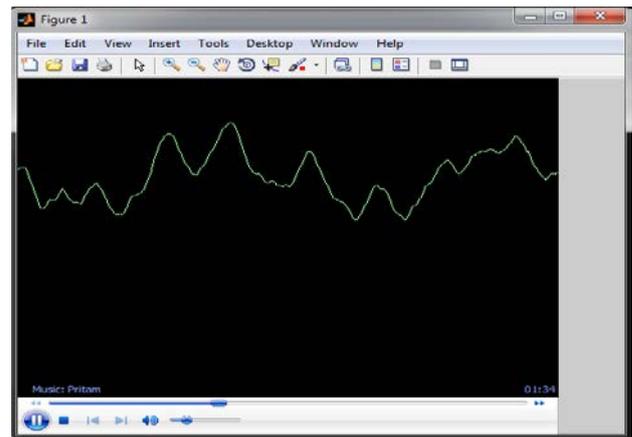
Figure 7: FLOW CHART OF THE PROGRAM

Figure 8: IMPLEMENTATION OF MEDIA PLAYER USING COLOR MARKERS

In this application, at first the webcam identifies the color markers & the user selects the pixel density of the markers. For this application red & yellow markers will be used. Now the application program is debugged using Matlab as the simulating software. On completion of the debugging process the user can control the Media player through his hand gestures which interface with the GUI.

In this flowchart at first the image of the required colour markers is taken, it is followed by the colour thres-holding& blob detection process. At first the location of the centroid of the markers is found out then if the total pixels selected on the markers is greater than 500 pixel then song being played in the media player gets paused otherwise it plays. Now if the centroid of the marker is less than 200 then volume of the song

can be increased otherwise if it is more than 200 pixel volume will decreased.

IX. FUTURE SCOPE

The future of this technology is very bright and in this year, Microsoft will launch its breakfast table top which will be considered a huge achievement. Progress will be made in recognizing gestures through the abovementioned hand gesture method and hopefully all its limitations would be overcome. Hence, in the coming 4-6 years, this technology will attain great heights.

X.CONCLUSION

So we conclude that through the combinational use of color markers & simulating software, we can control any PC s' GUI. Thus we are eliminating the use of traditional I/O devices like keyboards etc. This technique is more effective and friendly for human computer interaction. It can be used for advertising. It is a new emerging technology.

XI.ACKNOWLEDGEMENT

XI.REFERENCES

[1] J.A. Adam, "Virtual Reality," IEEE Spectrum, vol. 30, no. 10, pp. 22-29, 1993

[2] A.G. Hauptmann and P. McAvinney, "Gesture With Speech for Graphics Manipulation," Int'l J. Man-Machine Studies, vol. 38, pp. 231-249, Feb. 1993.

[3] H. Rheingold, Virtual Reality. Summit Books, 1991

[4] Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.

[5] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems,Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi:

10.1109/TSMCC.2007.893280.

[6] Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi, Akira Iwata, (2002). "A Rotation Invariant Approach On Static-Gesture Recognition Using Boundary Histograms And Neural International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012

[7] Niket Amoda, Ramesh K Kulkarni, "Efficient Image Retrieval using Region Based Image Retrieval", International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868, Foundation of Computer Science FCS, New York, USA, International Conference & workshop on Advanced Computing 2013 (ICWAC 2013)

[8] Niket Amoda, Ramesh K Kulkarni, "Image Segmentation and Detection using Watershed Transform and Region Based Image Retrieval", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)-Volume 2, Issue 2, March - April 2013

[9] Niket Amoda, Ramesh K Kulkarni, "Efficient Image Segmentation Using Watershed Transform", INTERNATIONAL JOURNAL OF COMPUTER SCIENCE & TECHNOLOGY (IJCST)-VOL IV ISSUE II, VER. 2, APR. TO JUNE, 2013

[10] Min B., Yoon, H., Soh, J., Yangc, Y., &Ejima, T. (1997). "Hand Gesture Recognition Using Hidden Markov Models". IEEE International Conference on computational cybernetics and simulation. Vol. 5, Doi: 10.1109/ICSMC.1997.637364

[11] M.D. Yacoub, Wireless Technology: Protocols, Standards, and Techniques, CRC Press, London (2002).

[12] K. McMenemy and S. Ferguson, A Hitchhiker's Guide to Virtual Reality, A K Peters, Wellesley (2007).

[13] Global Positioning System, "Home page", http://www.gps.gov/, visited on 10/10/2007.

[14] S.G. Burnay, T.L. Williams and C.H. Jones, Applications of Thermal Imaging, A. Hilger, Bristol (1988).

[15] J.A. Adam, "Virtual Reality," IEEE Spectrum, vol. 30, no. 10, pp. 22-29, 1993

[16] A.G. Hauptmann and P. McAvinney, "Gesture With Speech for Graphics Manipulation," Int'l J. Man-Machine Studies, vol. 38, pp. 231-249, Feb. 1993.

[17] H. Rheingold, Virtual Reality. Summit Books, 1991

[18] Xingyan Li. (2003). "Gesture Recognition Based on Fuzzy C-Means Clustering Algorithm", Department of Computer Science. The University of Tennessee Knoxville.

[19] S. Mitra, and T. Acharya. (2007). "Gesture Recognition: A Survey" IEEE Transactions on systems,Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311- 324, doi:10.1109/TSMCC.2007.893280.

[20] Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi, Akira Iwata, (2002). "A RotationInvariant Approach On Static-Gesture Recognition Using Boundary Histograms And Neural International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.4, July 2012173Networks," IEEE Proceedings of the 9th International Conference on Neural InformationProcessing, Singapura.

[21] Joseph J. La Viola Jr. (1999). "A Survey of Hand Posture and Gesture Recognition Techniques andTechnology", Master Thesis, Science and Technology Center for Computer Graphics and ScientificVisualization, USA.

[22] Mahmoud E., Ayoub A., J¨org A., and Bernd M., (2008). "Hidden Markov Model-Based IsolatedAnd Meaningful Hand Gesture Recognition", World Academy of Science, Engineering andtechnology 41.

AUTHOR PROFILES

Aditya Mhatre received his diploma in electronics & telecommunication from Vidyalankar Polytechnic. Currently he is pursuing Bachelor's degree in Electronics & Telecommunication from University of Mumbai. He has also published many technical papers in international journals & conferences. His research areas are wireless communication & image processing. He is also pursuing an associate degree in electrical engineering & computer science from Massachusetts institute of Technology, USA.