

REAL TIME HAND GESTURE RECOGNITION SYSTEM FOR DYNAMIC APPLICATIONS

Siddharth S. Rautaray¹, Anupam Agrawal²

Indian Institute of Information Technology Allahabad, India
{sr.rgpv@gmail.com¹, anupam69@gmail.com²}

ABSTRACT

Virtual environments have always been considered as a means for more visceral and efficient human computer interaction by a diversified range of applications. The spectrum of applications includes analysis of complex scientific data, medical training, military simulation, phobia therapy and virtual prototyping. Evolution of ubiquitous computing, current user interaction approaches with keyboard, mouse and pen are not sufficient for the still widening spectrum of Human computer interaction. Gloves and sensor based trackers are unwieldy, constraining and uncomfortable to use. Due to the limitation of these devices the useable command set based diligences is also limited. Direct use of hands as an input device is an innovative method for providing natural Human Computer Interaction which has its inheritance from text-based interfaces through 2D graphical-based interfaces, multimedia-supported interfaces, to full-fledged multi-participant Virtual Environment (VE) systems. Conceiving a future era of human-computer interaction with the implementations of 3D application where the user may be able to move and rotate objects simply by moving and rotating his hand - all without help of any input device.

The research effort centralizes on the efforts of implementing an application that employs computer vision algorithms and gesture recognition techniques which in turn results in developing a low cost interface device for interacting with objects in virtual environment using hand gestures. The prototype architecture of the application comprises of a central computational module that applies the camshift technique for tracking of hands and its gestures. Haar like technique has been utilized as a classifier that is creditworthy for locating hand position and classifying gesture. The patterning of gestures has been done for recognition by mapping the number of defects that is formed in the hand with the assigned gestures. The virtual objects are produced using Open GL library. This hand gesture recognition technique aims to substitute the use of mouse for interaction with the virtual objects. This will be useful to promote controlling applications like virtual games, browsing images etc in virtual environment using hand gestures.

KEYWORDS

Hand gesture, virtual objects, virtual environment, tracking, recognition.

1. INTRODUCTION

The impendent of virtual environments brings in a whole new set of problems for user interfaces. The unveiling of 3D objects and worlds in which the user is engrossed allows such people as

scientists, engineers, doctors and architects to envision composite structures and systems with eminent degrees of quality and naturalism. Shutter glasses furnish a stereo or 3D view of the scene, which is no longer confined to a desktop monitor, but may be a large table, projection screen or room. The limiting component in these systems currently is the fundamental interaction. Virtual environments seek to produce a world where the interaction experiences are real. Current mechanical, acoustic and magnetic input devices track the user and provide control of movement, selection and manipulation of objects in virtual scenes. Several tools are purported and used so far to make such interaction more and more prompt and effortless. Touch screens are the most widely used example: though the ramification of the underlying system is hidden from the user, and makes it possible for a user to point to the choices as he could do in real life. The cost associated to it is the major limitations of the aforesaid technology other limitations may be size, requirement of a physical location, and other intrinsic limitation to 2D. Other more innovative devices proposed for virtual reality include gloves or wearable tools such as mechanical sensors, actuators and micro cameras [1]. They are capable to handle 3D worlds, making it natural and realistic, and also provide in some implementations tactile sensations. Regrettably, their cost is usually very high, and thus the user acceptance confined, hence making them more desirable for professional applications such as a flight simulator or remote surgery equipment. However these interactions are often limited and non rational, while the devices are awkward, unmanageable and prone to distortion from the physical environment. We are interested in formulating an alternative, natural interface that more intimately models the way we interact with the real world. The user should be able to reach out, grab, point and move 3D objects just as we do with real objects.

These challenges open a new direction for human-computer interaction which combined with computer vision techniques and it is possible to build an advanced input devices. The computer vision devices can be implemented and upgrade to the new input devices in the future. It gives the input command to the computer rather than just a function of taking photo or record video. We can do more implementation to transform the computer vision devices to become an input command device to reach the function as keyboard or mouse. One of the ways to give signal to computer vision devices is by using hand gesture. More specifically hand gesture is used as the signal or input modality to the computer. Certain signal can be recognized by computer as an input of what computer should do. These will benefits the entire user without using a direct device and can do what they want as long as the computer vision device can sense it. These make computer user easier than using the keyboard or mouse. The future computer or laptop may eliminate the use of keyboard and mouse by substituting with a vision-based interpretation devices.

Interaction between humans comes from different sensory modes like gesture, speech, facial and body expressions [2]. The main advantage of using hand gestures is to interact with computer as a non-contact human computer input modality. The state of art of human computer interaction presents the facts that for controlling the computer processes gestures of various types of hand movements have been used. The present research effort defines an environment where a number of challenges have been considered for obtaining the hand gesture recognition techniques in the virtual environment. Being an interesting part of the Human computer interaction hand gesture recognition needs to be robust for real life applications, but complex structure of human hand presents a series of challenges for being tracked and interpreted. Other than the gesture complexities like variability and flexibility of structure of hand other challenges include the shape of gestures, real time application issues, presence of background noise and variations in illumination conditions. The specifications also involve accuracy of detection and recognition for

real life applications [3].

The present research effort has a goal of developing an application using vision based hand gestures for manipulation of objects in virtual environment. Our application presents a more effective and user friendly methods of human computer interaction intelligently with the usage of hand gestures. Functions of mouse like controlling of movement of virtual object have been replaced by hand gestures. The complexity involved is with the detection and recognition phases of the simulated virtual application. The challenges encountered are noisy environment which creates a big impingement on the detection and recognition performance of human hand gestures. The application has been designed to be cost effective and uses low cost input tools like webcam for capturing hand as input. Manipulation of virtual objects has been done through modeling of some predefined command based hand gestures.

2. STATE OF ART

In earlier days hand gesture detection was done using mechanical devices to obtain information of the hand gesture [4]. One of the most widely used and accepted examples for hand gestures recognition is data glove. Evolution of computer hardware improved a lot of in present scenario this also effects the performance of computing. Enhancements of gesture recognition has replaced the role of data gloves to non wearable devices due to its naturalness without using any device this is quite user friendly in human computer interaction. One of the major drawbacks of data glove is that it is cumbersome with the limitation of hand movement.

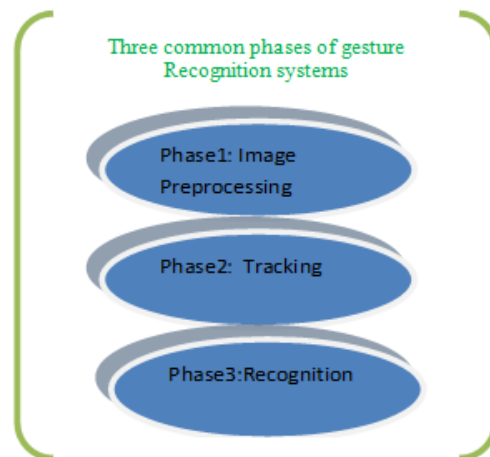


Figure 1. Three common stages of gesture recognition systems

Also vision is one of the major six physical senses that computer must be instantiated perceptibly when communicated to humans [1]. So vision based approaches are preferred more than wearable devices in hand gesture recognition. Generally there are three stages in most of the gesture recognition systems. The three stages may be enumerated as image pre –processing tracking and recognition stage [16] as shown in Figure 1. In tracking, there are several researchers who have done the similar research like Viola-Jones based cascade classifier, commonly used for face

tracking in rapidly image processing [5]. Cascade classifiers are currently considered more robust pattern detection against the noises and lighting conditions as well [6].

For tracking Viola-Jones and several other researchers have developed algorithms used for face tracking in rapid image processing like HAAR cascade classifier. This is presently one of the robust detection techniques under different constraints like noise [6]. Gesture as input of human computer interaction based applications is an emerging field in which many researchers have worked and proposed different practical techniques. Jain [7] implemented a vision based hand gesture pose estimation based application for mobile devices. Pavlovic et al. [8] accomplished in their work that the gestures of users must be explained logically for developing a good human computer interaction based system. Though the present technologies of gesture recognition are not feasible in providing the logical explanations to the human gestures. Some of the major challenges evolved in due course of time are the complexness and robustness of human gestures. Another hand gesture recognition method based on input-output Hidden Markov Models of tracking skin color blobs was proposed by Marcel et al. [9]. Controlling VLC media player using hand gesture recognition is done in real time environment using vision based techniques [10]. The sign language tutoring tool studied by Aran et al. [11] which their research designed to teaching the fundamental of the sign language in interactive way.

Liu and Lovell [12] implemented a technique for real time tracking of hand capturing gestures with the use of a web camera, personal computer and image processing algorithms making it more users friendly. Chen et al. [13] implemented hidden Markov model technique for training the hand gesture for recognizing the hand postures. Nevertheless, this model is more complicated in training the hand gesture equated with Cascade classifiers. Lee et al. [14] developed a Virtual Office Environment System (VOES), in which avatar is used navigate and interact with other participants.

Contemporary works in hand gesture recognition by many researchers show that hand gesture system can also be practically implemented into several type of application systems and various environment. Ahn et al. [15] developed an interactive way of slide show presentation system in the virtual environment. Research in hands, gestures and movement helps in developing models of the human body. This makes it possible to solve the challenges from mathematical viewpoint. However, these techniques proposed are excessively complex and sophisticated for typical application scenarios. Generally, pattern recognition methodologies are capable of solving the problem with humbler hardware and computation necessities. In the present research effort, we will consider these aspects by taking it as a reference to a smart interaction environment of virtual object manipulation and control. Here the user can execute different actions that translate into a command in an intelligent system and further execute the user requirements into practical actions.

3. APPLICATION ARCHITECTURE DESIGN

The application uses a combination of different computer vision techniques for hand gesture recognition. It recognizes static hand gestures. Figure 2 shows the application architecture design for manipulating virtual objects using hand gestures.

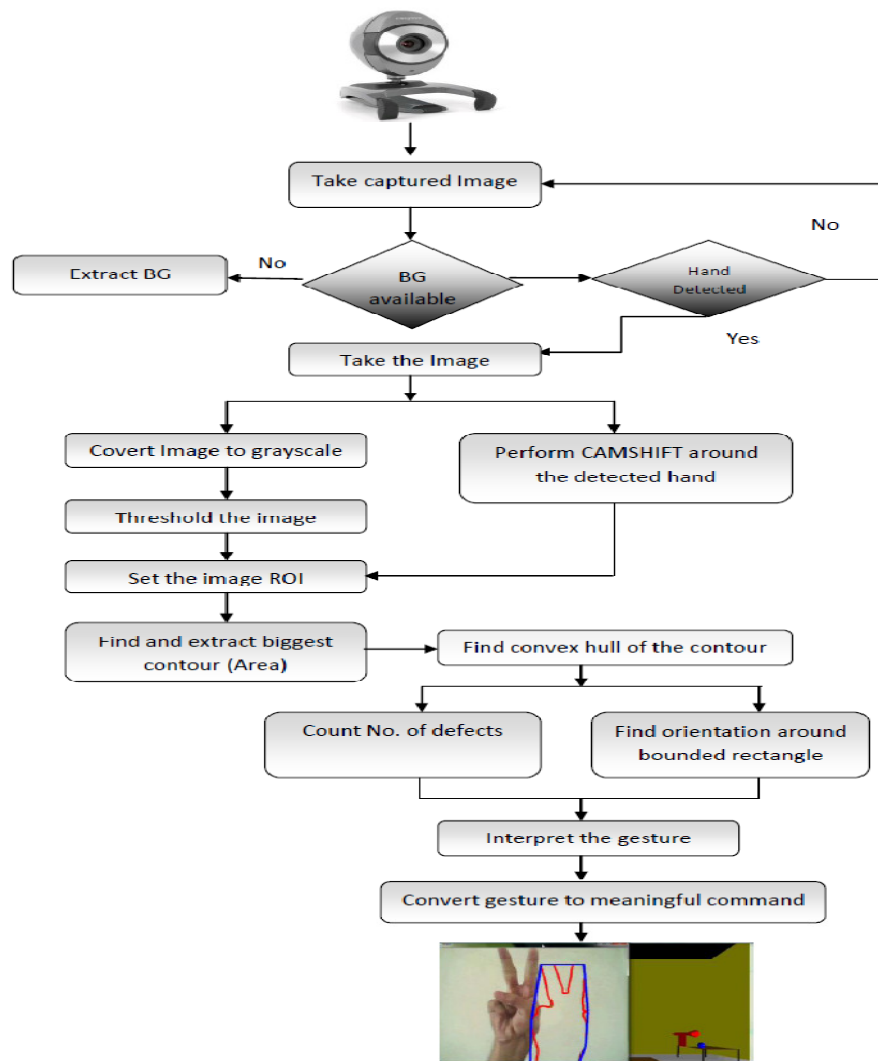


Figure 2. Application architecture design

Images are captured from camera and passed through following phases/techniques .Starting with the acquisition phase that is the first phase. Since a standard input peripheral (keyboard, pointing device) will be unacceptable in this application context. So we have focused on possible alternatives by considering smart interfaces that have been inspired by natural behaviour of the users in real-world actions. The choice of the capturing device is being done in accordance with the idea of spreading the installation in homes, labs, play stations etc, hence maintaining the resulting costs low. For this reason, special care has been taken to ensure good performance even in the presence of low-cost cameras. The camera is supposed to be fixed, and illumination slowly varying. Real-time constraints are being imposed for a careful design of the processing system. To this purpose, the unnecessary information is first removed. In particular, a background suppression procedure has been performed in the HSV colour space, in which the scene can be modelled discarding illumination variations .Thus focusing the attention on areas corresponding to human skin colour. The next section deals with the computer vision techniques/algorithms used for hand tracking and recognition.

4. APPLICATIONS AND ANALYSIS

- a) *Virtual Reality*: Gestures for virtual and augmented reality applications have experienced one of the greatest levels of uptake in computing. Virtual reality interactions use gestures to enable realistic manipulations of virtual objects using one's hands, for 3D display interactions or 2D displays that simulate 3D interactions
- b) *Games*: When we look at gestures for computer games, Freeman tracked a player's hand or body position to control movement and orientation of interactive game objects such as cars. Konrad et al. [10] used gestures to control the movement of avatars in a virtual world, and PlayStation 2 has introduced the Eye Toy, a camera that tracks hand movements for interactive games
- c) *Sign Language*: Sign language is an important case of communicative gestures. Since sign languages are highly structural, they are very suitable as test beds for vision algorithms [12]. At the same time, they can also be a good way to help the disabled to interact with computers. Sign language for the deaf (e.g. American Sign Language) is an example that has received significant attention in the gesture literature [13, 14, 15 and 16].

4.1 Analysis Parameters

In order to find out the performance and viability of the proposed gesture recognition system following testing and analysis parameters could be considered

- a) *Robustness*: In the real-world, visual information could be very rich, noisy, and incomplete, due to changing illumination, clutter and dynamic backgrounds, occlusion, etc. Vision-based systems should be user independent and robust against all these factors.
- b) *Scalability*: The Vision-based interaction system should be easily adapted to different scales of applications. For e.g. the core of Vision-based interaction should be the same for desktop environments, Sign Language Recognition, robot navigation and also for VE.
- c) *Computational Efficiency*: Generally, Vision based interaction often requires real-time systems. The vision and learning techniques/algorithms used in Vision-based interaction should be effective as well as cost efficient.
- d) *User's Tolerance*: The malfunctions or mistakes of Vision-based interaction should be tolerated. When a mistake is made, it should not incur much loss. Users can be asked to repeat some actions, instead of letting the computer make more wrong decisions.

The computer vision techniques used in the application for manipulation of objects in virtual environment have been implemented in C++ with the use of Open CV Library. The virtual objects (front end) have been designed using OpenGL library. The hardware requirements of the application to be implemented include computer with 1.99 GHz processor. The web cam used in the experimental setup captures image sequences at the resolution of 320x240. Practical experiments show that our application is implemented well in environments with little noises (i.e., existence of objects whose color is similar to human skin) and with the balanced lighting condition.

First, the user places his hand in front of the webcam. The webcam then detects the user's hand by creating a rectangle around it as shown in figure 5.

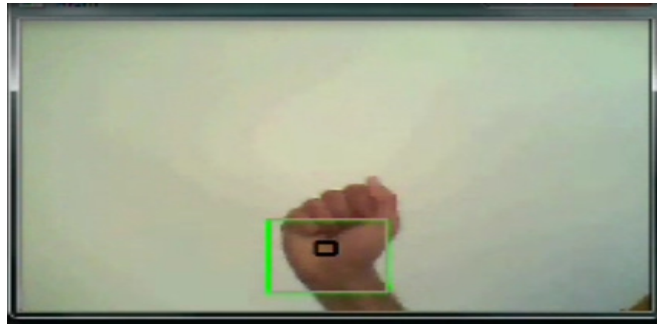
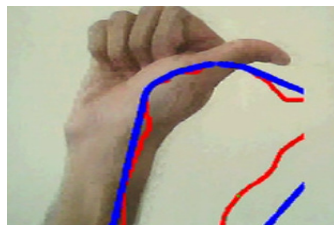
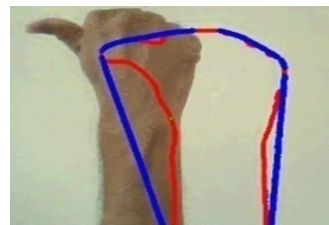


Figure 5. Hand Detected

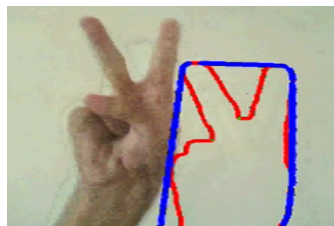
Once the hand has been detected the application further tracks different gestures of the user performed by his hand and generates contour around it.



i. Move Left



ii. Move Right



iii. Move Up



iv. Move Down

Figure 6. Gestures for manipulating objects in virtual environment.

The application uses seven hand gestures defined within the application for manipulation with objects in virtual environment. Figure 6 shows the different gestures along with their assigned commands (functions) to manipulate the objects in virtual environment.

5. RESULTS

Following figures shows the results obtained from different gestures used for manipulating objects in virtual environment.

Depicting the objects in virtual environments where different objects are manipulated by hand gestures. The red stick having a red ball is moving left direction as shown in the following figure 7.

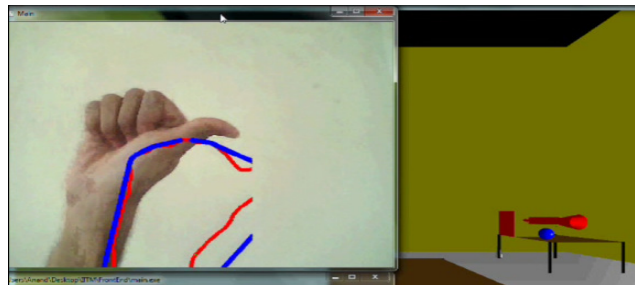


Figure 7. Gesture for moving left

- Depicting the objects in virtual environments where different objects are manipulated by hand gestures. The red stick having a red ball is moving right direction (away from blue ball) as represented in figure 8



Figure 8. Gesture for moving right

- Depicting the gesture of moving up for manipulating the object in virtual environment in up direction as shown in figure 9.

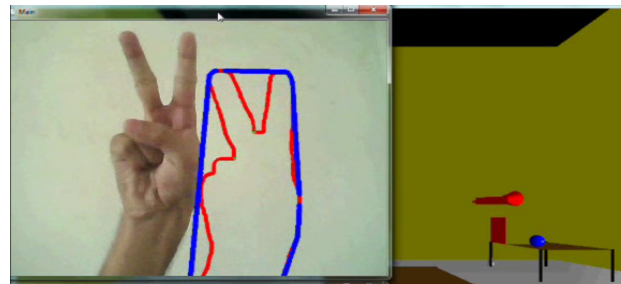


Figure 9. Gesture for moving up

- Depicting the gesture of moving down for manipulating the object in virtual environment in down direction shown in figure 10.

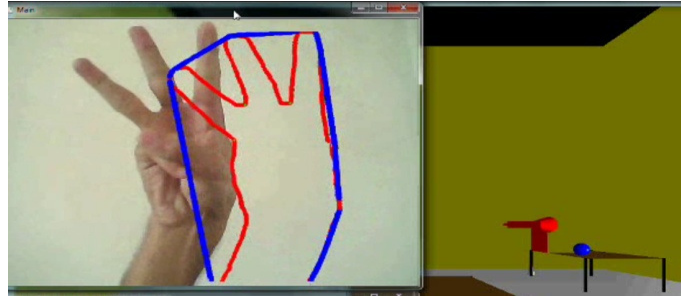


Figure10. Gesture for moving down

6. CONCLUSION

In present environment a number of facilities and various modes for providing input to any application are available. It is though unfortunate that with the ever increasing smart environments and corresponding input technologies still not many applications are available which are controlled using current and smart facility of providing input which is by hand gesture. The most important advantage of the usage of hand gesture based input modes is that using this method the user can interact with the application from a distance without using the keyboard or mouse. The application of manipulating objects through hand gestures in virtual environment is being proposed and implemented in the present paper provides a suitable efficient and user friendly human computer interface. With the help of this application the user can interact with the virtual objects using hand gesture instead of any other physical input devices. As the application provides the flexibility to the users and specifically physically challenged users to define the gesture according to their feasibility and ease of use.

7. FUTURE WORK

The present application though seems to be feasible and more user friendly in comparison to the traditional input modes but is somewhat less robust in recognition phase. An attempt to make the input modes less constraints dependent for the users hand gestures has been preferred. But robustness of the application may be increased by applying some more robust algorithms that may help to reduce noise and blur motion in order to have more accurate translation of gestures into commands.

Another important aspect for the related development could be design of an independent gesture vocabulary framework. This framework though could be independent of the application domain. Also the framework may be useful for controlling different types of games and other applications dependent on the controlled through user defined gestures.

REFERENCES

- [1] Conic, N., Cerseato, P., De & Natale, F. G. B., (2007), “Natural Human- Machine Interface using an Interactive Virtual Blackboard”, In Proceeding of ICIP 2007, pp.181-184.
- [2] Ismail, N. A., & O’Brien, A., (2008), “ Enabling Multimodal Interaction in Web-Based Personal Digital Photo Browsing”, Proceedings of the International Conference on Computer and Communication Engineering , Kuala Lumpur, Malaysia, May 13-15, pp. 907-910.
- [3] Pang, Y. Y., Ismail, N. A., & Gilbert, P. L. S., (2010), “ A Real Time Vision-Based Hand Gesture Interaction”, Fourth Asia International Conference on Mathematical Analytical Modelling and Computer Simulation, pp. 237-242.
- [4] Kortum, P., (2008) “HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces” Morgan Kaufmann Publishers, pp. 75-106.
- [5] Viola & Jones, (2001), “Rapid object detection using boosted cascade of simple features”, In Proceedings of Computer Vision and Pattern Recognition, pp. I-511 - I-518.
- [6] Chen, Q., Coredea, M. D., Petriu, E. M., Varkony, A. R., Koczy, I., & Whalen, T.E., (2009), “ Human Computer Interaction for Smart Applications Using Hand Gesture and Facial Expressions,” International Journal of Advanced Media and Communication, vol. 3c.1/2, pp. 95-109.
- [7] Jain, G. (2009), “Vision-Based Hand Gesture Pose Estimation for Mobile Devices”, University of Toronto.
- [8] Pavlovic, V., Sharma, R., & Huang, T.S. (1997), “Visual interpretation of hand gestures for human-computer interaction: A review.” IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 7(19):pp. 677–695.
- [9] Marcel, S., Bernier, O., Viallet, J. E., & Collobert, D (2000), “Hand Gesture Recognition using Input-Output Hidden Markov Models.” In Proc. of the FG’2000 Conference on Automatic Face and Gesture Recognition.
- [10] Rautaray, S.S., & Agrawal, A. (2010), “A Novel Human Computer Interface Based On Hand Gesture Recognition Using Computer Vision Techniques”, In Proceedings of ACM IITM’10, pp.292-296.
- [11] Aran, O., Ari, I., Benoit, F., Campr, A., Carrillo, A. H., Fanard, Akarun, L., Caplier, a., Rombaut, M., & Sankuru, B., (2006) “ Sign Language Tutoring Tool”, eNTERFACE 2006, The Summer Workshop on Multimodal Interfaces, Croatia.
- [12] Liu, N., & Lovell, B. (2001) “Mmx-accelerated realtime hand tracking system” In Proceedings of IVCNZ.
- [13] F. Chen, C. Fu, & C. Huang, 2003 , “Hand gesture recognition using a real-time tracking method and hidden Markov models” Image and Vision Computing, pp. 745-758.
- [14] Lee, C. S., Ghyme, S. W., Park, C. J., Wohn, K., (1998) “The Control of avatar motion using hand gesture”, In Proceeding of Virtual Reality Software and technology (VRST), pp. 59-65.
- [15] Ahn, S. C., Lee, T. S., Kim, I. J., Kwon, Y. M., & Kim, H. G. (2004),“ Computer Vision-Based Interactive Presentation System,” Proceedings of Asian Conference for Computer Vision.

- [16] Moeslund, T. B., & Norgaard, L. (2002) "A brief overview of hand gestures used in wearable human computer interfaces", Technical report, Aalborg University, Denmark.

Authors

Mr. Siddharth Swarup Rautaray is a PhD student in the Department of Information Technology at Indian Institute of Information Technology, Allahabad, India. His research interests are in human-computer Interactions, user interface design, and computer vision image processing and graphics. He is a member of the IEEE Computer Society, ACM SIGCHI and CSI. Contact him at sr.rgpv@gmail.com



Dr Anupam Agrawal received his Ph.D (Information Technology) degree from IIIT, Allahabad (in association with IIT, Roorkee). He carried out his Post-Doctoral Research work at the Department of Computing and Information System, University of Bedfordshire (UK). He is presently working as an Associate Professor at Indian Institute of Information Technology, Allahabad since 2006. His research interests include Interactive Technologies and Human Computer Interaction, Computer Vision, Pattern Recognition, Image Processing, AI and Soft Computing. He is a senior member of the ACM, the IEEE Computer Society and CSI.

