Teleoperation with Haptics

Aishanou Osha Rait  
ME, IIT Delhi  
aorait@gmail.com

Sudipto Mukherjee  
ME, IIT Delhi  
sudipto@mech.iitd.ac.in

Abstract— In a teleoperation system, an operator controls the movements of a robot that is located some distance away. Some teleoperated robots are limited to simple tasks, such as aiming a camera and sending back visual images. In a more sophisticated form of teleoperation known as telepresence, the human operator has a sense of being located in the robot’s environment. Haptics now makes it possible to include touch cues in addition to audio and visual cues in telepresence models. In this article we discuss haptic aspects to be considered while designing a wearable exoskeleton for pick and place operations.

I. INTRODUCTION

Teleoperation is the process of operating a machine at a distance. In situations like in space applications or disasters we encounter situations where the environment is hazardous to human health. It is expedient to deploy robots in such locations. Their motion is controlled by human operators positioned at a safe distance from hazard centers with radiation and chemicals. In space programs, the rovers on other planets are controlled remotely from earth. Recently in Japan, after the damage caused by the tsunami and the earthquake, robots were deployed for shutting down the nuclear plants and carrying out rescue missions.

However, even for seemingly innocuous tasks like inserting a peg in a hole, it is difficult to provide instructions in the form of a predefined algorithm. Since, we are to go a long way before we can make task planning systems matching the human brain, telepresence provides us with a convenient approach to deal with certain class of applications. The human operator can maneuver a device whose motion will be emulated by a remote machine. The controlling device may be a mouse, a joystick or a steering wheel. To make the interface more intuitive, it can be a wearable robot or exoskeleton as shown in Figure 1.

II. HARDWARE STRUCTURE

Our work focuses on an exoskeleton which has been developed (Fig 1) to be worn on a human arm, and provide the necessary information for effectively controlling an industrial robot through hand motion. The application targeted is a peg in a hole insertion, which is to be carried out by a KUKA KR5 ARC manipulator (Fig 3). For effective functioning, the end-effector’s motion of the master device, in this case the exoskeleton, is reproduced in the slave device, in this case the KUKA robot, with power amplification as desired. This type of control is called position-position teleoperation.

The KUKA KR5 is a six degree of freedom (DOF) manipulator, whereas the human arm, and correspondingly the anthropometric exoskeleton (shown schematically in Figure 2), has seven DOFs. There are three degrees of freedom at the shoulder joint, two at the elbow joint and two at the wrist. Sensors are mounted on seven different places on the exoskeleton in order to sense the variations of the joint angles.

Figure 1: Upper body exoskeleton in IIT Delhi

Figure 2: An exoskeleton robot to be worn on human arm (Master)

Figure 3: KUKA KR5 industrial robot (Slave)
Tekscan’s FlexiPot Potentiometers (Fig 4) are used for shoulder joint and pronation–supination (twisting) of the elbow joint. Rotary Potentiometers (Fig 5) are used for the measurement of angle of rotation of flexion–extension (forward-backwards) and abduction–adduction (inside-outside) motions.

Figure 4: Flexipot Potentiometers

Figure 5: Rotary Potentiometers

III. KINEMATIC MAPPING

Figure 6 shows the exoskeleton and the KUKA in one instance. Very clearly, other than the ends being in the same location, there is no structural similarity. A one-on-one mapping from joint measurements in the master device to that of the KUKA is not possible as the degrees of freedom vary. Secondly, the master device has adjustable link lengths to allow customization based on the operators arm size, while the KUKA has invariant kinematics. In order to map the position and orientation of the tip of human hand to the position of the end-effector of the manipulator, we first compute the position and orientation of the distal end of the master from the sensor readings. This is called computing the forward kinematics in robotics parlance and is done by successive matrix computations. Once the stance of the end-effector is known, the joint positions and orientations that the KUKA must attain to reproduce the same stance are calculated. This process is called inverse kinematics and is sometimes tricky as the solution is not unique and has to be carefully managed for stability and physical feasibility.

IV. NATURE OF HAPTIC FEEDBACK

Inserting a peg in a hole is a high precision task, requiring careful maneuvering by the operator. The tolerance routinely goes down to 5 micron, which is smaller than the accuracy or repeatability of industrial robots. Visual servoing may not always be feasible as the hole/peg is often occluded on the approach. Also, till now we have been dealing only with transfer of information from the master to the slave and not vice versa, other than the visual information through the camera. It is required that the manipulator also relays us information about the forces in its own environment.

Experiments have been done elsewhere [2] to measure the forces arising during the insertion of a peg in a hole. The results obtained, indicate that for high tolerance insertion, an unskilled operator takes longer, produces larger lateral forces and requires larger insertion force. Moreover, impulsive forces are measured even after the insertion is complete. Skilled operator on the other hand produces negligible lateral forces, and the insertion force does not exceed self weight of the object being inserted. The sequence followed is that of rough alignment without exerting a significant force in any direction, followed by fine alignment involving a search motion to align the two centre axis. The oscillations along the direction of motion are minimum for a skilled operators suggesting decoupling of lateral motions. The insertion direction force drops to zero as axes are aligned, remains zero during the insertion, and increases to support self weight of the peg on completion. In this process, the maximum inclination is about a degree and consequently haptic guidance rather than visual feedback should be the preferred option. Skilled operators utilize haptic clues which enable them to approach the task in a highly consistent and efficient way, minimizing the unnecessary forces, as compared to unskilled operators who blindly search for a solution.

V. HAPTIC FEEDBACK TECHNOLOGY

Haptic feedback had numerous applications. For example, in gaming industry, a more realistic experience of the virtual environment is achieved by introduction of force feedback devices along with existing visual and auditory feedbacks. The Novint Falcon’s, 3-D force feedback controller can be
controlled to render a pistol report and a shotgun blast differently. Some so called 4D theatres, have force feedback embedded in the seat, to provide the audience with a more enriching experience, including collisions and free-fall. In medical applications students are trained on delicate surgical techniques on the computer, feeling what it's like to suture blood vessels in an anastomosis or inject BOTOX into the muscle tissue of a virtual face. In mobile phones haptic feedback is being embedded in the touch screen to allow the user to get an experience of pressing a physical button while typing. The presence of haptic feedback in everyday applications is growing by the day and hopefully soon we might be able to feel the rocks and soil texture of some distant planet’s surface, remotely from the earth [3, 4, 5].

While inserting a peg in a hole, the hand experiences sensations in the form of kinesthetic and tactile feedback. Using a force-toque sensor, we measured the forces and torques arising in such an insertion. The Fourier transform of the data indicates that the response is primarily of low frequency, with time constant of 4 to 7 seconds. In teleoperation, these forces and torques will be physically felt by the manipulator and for the human operator to react effectively, it is required that he feels similar forces, scaled to an appropriate value. The force to be fed-back is mainly of restraining type, that is opposite in direction to the motion, similar to the action of brakes.

Electric motors can provide the required torque but their weight and space requirements limits their feasibility of being mounted on a wearable robot. Further, being active devices, they can be harmful to the operator. Hysteresis and eddy current brakes, cannot maintain the desired torque when the device is in static configuration [6, 7]. Piezoelectric and electrostatic devices can produce only small torques. In our application, force to be fed-back is mainly of restraining type and thus, electrically actuated magnetic particle brakes, similar in working principle to the clutch shown in Figure 8 are suitable haptic force generators. They have high torque to weight ratio and low time constants (of the order of msec). Secondly, they are typically actuated by 24 V, 0.5 amp power sources. The added advantage is that being passive devices they will not harm the user in case the system becomes unstable or the control mechanism fails. Based on the measurements above, magnetic particle brakes with peak torque capacity of 1.0 and 0.5 N-m capacity have been selected to provide the feedback.

Figure 8: Electromagnetic Particle Clutch (http://magnetic-clutch.com)

VI. CONCLUSION AND FUTURE SCOPE

Haptic feedback provides the operator with a better understanding of the remote environment, especially in situations where visual feedback does not have sufficient resolution. For enhanced telepresence, the exoskeleton shown in Figure 1 is currently being redesigned to include three axis haptic feedback, and more compact electronics. The reaction forces sensed by the robot during the insertion procedure, after
visual positioning of the end-effector will be used to actuate the magnetic particle brake. In the context of the peg-in-hole insertion, this feedback assumes significance when only small magnitude adjustments are needed in the insertion process. This feedback to the exoskeleton device should augment the user experience. In the long term, we also hope to incorporate fingers with haptic feedback at the end-effectors for further enhancement of telepresence.

ACKNOWLEDGEMENT

We thank Mr. Dharamender Jaitly, Mr. Kush Prasad, Mr. Gaurav Agarwal and Mr. Rajeevlochana C.G for their contributions.

REFERENCES


AUTHOR INFORMATION

Aishnau Osha Rait completed her B.Tech in Electrical and Electronics Engineering from BIT, Mesra, Ranchi in 2011. Her B.Tech project dealt with the implementation and comparison of different PWM techniques using FPGA, for controlling the speed of electric motors. She is currently working as a JRF in Mechanical Engineering Department of IIT Delhi, under the project "Tele-operation of an Industrial Robot". Her research interests include image processing, autonomous robotics and embedded systems.

Sudipto Mukherjee did B.Tech in Mechanical Engineering from IIT Kanpur in 1985 and M.S and Ph.D from Ohio State University, USA in 1988 and 1992 respectively. His Ph.D thesis was on dexterous grasp and manipulators. He has been involved in designing & developing robotic arm and grasping devices, mobile robots, micro flap-wing devices, under-carriage vehicle inspection systems, etc. He is a recipient of INAE Young Engineer Award in 1998 and AICTE Career Award for Young Teachers in 1995. He is a fellow of Institution of Engineers. He is a Council member of the International Research Council for The Biomechanics of Impact and a Member of the Sectoral Monitoring Committee for the Innovative Engineering Technologies of the Engineering Sciences Cluster of CSIR. Currently, he is Mehra Chair Professor of Design and Manufacturing at IIT Delhi. He has more than 120 research publications in reputed journals and conference proceedings.

Expertise:
The importance of virtual reality in robotics is on the rise in today's world. This is primarily because virtual reality offers a wide range of application in planning, designing, macro and micro assembly designing, rehabilitation therapy, and in remote monitoring of areas by robots. For robotics, there is a growing need for developing virtual reality simulators wherein the robots and their neurocontrol systems can be prototyped and trained. These robots can then be tested within the virtual 3D models of their intended mission environments. Virtual reality in robotics is also an inherent part of immersive environments and teleoperations.

**Sponsored Research Projects**

2. "Acquisition, Representation, Processing and Display of Digital Heritage", funded by the Department of Science and Technology (DST) (2010 - 13)
3. "Exploratory Visualization in 3D Virtual Environment", funded by the Department of Science and Technology (DST)(2010 - 13).
5. "Interrogative Synthetic Environments", funded by International Division, Department of Science and Technology (DST) (2000 - 03).
6. "Compression and access of video sequences and visualization of remote environments", funded by Department of Science and Technology (DST), India and Ministry of Science and Technology, Israel, (1998 - 2000)
7. "Virtual Intelligence", funded by All India Council of Technical Education (AICTE), (1996 - 99)

**Recent Student Research Projects**

4. Amit Aagarwal, "In the area of Augmented Reality", Ongoing M.Tech thesis, Department of Computer Science and Engineering, IIT Delhi, under the guidance of Dr. Subodh Kumar.

**Selected Publications**


**Faculty**

1. Prem Kalra, Professor, C.S.E
2. Subhashis Banerjee, Professor, C.S.E
3. K. K. Biswas, Professor, C.S.E
4. Subodh Kumar, Assoc. Professor, C.S.E

**Facilities**

1. HD Cameras
2. Kinects
3. Range Scanners
4. 3D TVs
5. Head Mounted Device
6. Laser scanner
7. Bassler Pilot Cameras
Virtual Reality in Robotics

Ayesha Choudhary  Prem Kumar Kalra  Subodh Kumar  Subhashis Banerjee
CSE, IIT Delhi  CSE, IIT Delhi  CSE, IIT Delhi  CSE, IIT Delhi
ayesha@cse.iitd.ac.in  pkalra@cse.iitd.ac.in  subodh@cse.iitd.ac.in  suban@cse.iitd.ac.in

Abstract—Virtual reality is gaining an important role in robotics especially with the increase in demand for teleoperation and immersive environments. Teleoperation has important applications since it allows robots to carry out tasks in areas which are hazardous for human operators. For teleoperation to be successful, creating accurate virtual 3D models of the remote environment is necessary. Haptics also plays an important role in creating a realistic feeling of immersion. In our research in the field of immersive environments for teleoperation, we are currently focusing on all three aspects, virtual reality, immersion and haptics. Our research in virtual reality is currently focusing on developing tools for creating the virtual 3D models both offline and online. For the haptics and control of the robot or exoskeleton in the remote environment, we are working towards developing tools for simulating the actions that the robot has to be trained to perform in the remote environment.

I. INTRODUCTION

Virtual reality plays an important role in robotics, specially in the case of immersive environments for teleoperation [1]. Teleoperation or remote-controlled robots is an important manner in which intelligence and maneuverability of human beings can be combined with the precision and durability of the robots. Robots governed by remote human operators are excellent candidates for working in hazardous or uncertain environment, such as nuclear power plants. Creation of virtual immersive environment through visual and haptic feed back is important for creating the feeling of immersion for the user. Teleoperation, is of great importance especially in cases where robots are used for operation in areas where it is dangerous for human beings to enter, and in medical operations that are carried out remotely. Successful implementation of teleoperation system demands creation of environments that make the user feel as if they are present at the remote site. However, though the applications and advantages of teleoperation are many, it is not an easy task.

Teleoperation is complex primarily due to the complexity of the system and and information management that is required. In case of immersive environment for teleoperation, a 3D model of the remote environment and the feeling of immersion are also very important. Creation of the virtual world of the remote area as well as modeling of the objects in the remote area are important for proper feeling of immersion and for teleoperation. Both have to go hand-in-hand and has to be facilitated by a feedback loop also.

We need to focus on re-creating the feeling of being in the remote environment by using immersive technology, be able to manipulate the robot in the remote environment and correct the action by having a feedback loop. Immersive virtual environment should allow the user to "feel" or "perceive" that he/she is in the remote environment. Therefore, the virtual environment created should be realistic enough for such perception. Moreover, the realism in the virtual environment has to be of good quality because the user should be able to act in the virtual environment such that the robot in the remote environment should be able to correctly perform the the user’s action. Therefore, virtual reality in robotics plays an important role and is an active area of research. Immersive environments for teleoperation is also an active area of research and requires research in various areas since all the senses are required to create a sense of real immersion in the virtual environment.

Panoramic 3D displays, surround sound acoustics, and haptic and force feedback devices create a sense of immersion that is close to the actual remote environment. Once the user “feels” immersed in the environment, the user should be able to interact within this in a natural and intuitive manner.

Haptics also plays an important role in immersive teleoperation applications [2] and goes a long way in contributing to the sense of presence in a variety of applications. An important factor in “presence” is that of resonance or the synergy between human cognition and the sensory information from the environment. When there is a sensory flow that supports the action, it is claimed that the system is in resonance. Human haptics plays an important role in presence since it has the unique ability to perceive sensory inputs from the environment along with the ability to act upon these inputs. Therefore, for the true feeling of immersion there has to be a resonance in the system. However, currently there exists a large visual-haptic gap between the level of technological development in this area and currently researchers are focusing on means and ways to reduce this gap also.

In our research in the field of immersive environments for teleoperation, we are currently focusing on all three aspects, virtual reality, immersion and haptics. Our research in virtual reality is currently focusing on developing tools for creating the virtual 3D models. We have developed techniques for creating virtual 3D models wherein we have some a priori knowledge of scene information and a single view of the scene. Moreover, we have also developed methods for creating 3D models of the scene using multiple images of the remote area. In this manner, we can create models using the view form the multiple cameras in the remote location online, whenever required.

For the haptics and control of the robot or exoskeleton in the remote environment, we are working towards developing tools for simulating the actions that the robot has to be trained with.
These simulations shall be used for both training the robots as well as for receiving a feedback to correct the action carried out by the robot in the remote environment. In the sections below, we first discuss our work on creating 3D models of scene using various computer vision techniques. Then, we discuss the other aspects of our research related to the area of immersive environments, teleoperation and robotics.

II. 3D VIRTUAL MODELS

Creating virtual models of the remote site is among the first steps towards developing immersive environments for teleoperation. We have developed methods that allow creating 3D models of the structured scene from single view of the real scene along with some \textit{a priori} knowledge of the scene. This is used for creating off-line 3D models of the remote environment and can be augmented as changes in the scene occur, using our multi-view based 3D reconstruction and virtual modeling tool.

In our work on single image based reconstruction of structured scene, we have developed a GUI based interactive tool that allows the user to specify the constraints on the structure of the scene such as the reference planes, coordinate axes and scale of the scene. It also implements a grammar based method for collective processing of the input scene constraints. We use texture mapping on the rendered scene and it also allows reconstruction of surfaces of rotation and symmetric replication of the rendered faces, whenever required. We have also developed a hierarchical tool for integrating the 3D information for several different views, that allows merging of separately reconstructed models from different views to obtain a detailed and unified model of the structured scene. This is performed by registering images corresponding to two or more different views and obtaining the corresponding rotation and translation matrices for importing one view to another. Figures 1(a) and 1(b) shows the input image and 3D virtual model of a building created using our interactive tool, respectively. Figures 2(a), 2(b) show two 3D models reconstructed from single views of each of these and Figure 2(c) shows the merged model from these two. Therefore, although a single view is not able to capture all the details of a large area, we have shown that it is possible to create 3D models of large areas by merging models created from single views of parts of the large area.

Another aspect of reconstruction of 3D scenes is when no \textit{a priori} knowledge of the scene is available. This case arises often in real-world applications, specially when the remote environment is not favorable for human beings to interact with. We assume that the remote environment is covered by multiple cameras and that continuous video data is available from these cameras. In this case, we develop techniques for dense 3D reconstruction of the scene. The first step in this case is developing techniques for self-calibration of the cameras. We extract the SIFT [3] features from the multiple input images and refine these using computer vision techniques for self-calibration of the cameras. This helps recover the parameters for each of the cameras in the multi-camera network. Then, dense reconstruction of the scene is carried out by first projecting the 3D points onto a reference view and creating a mesh. Then, we use a technique known as space carving [4] that assigns a color to each element (voxel) of the 3D space such that color consistency is maintained. Figures 3 and 4 shows the multiple input images and the sparse point cloud formed, respectively. Figure 5 gives the dense reconstruction of the scene from the multiple input images using the technique outlined above.

Kinect sensors have been recently developed by Microsoft and have the ability to compute the depth of a scene in real-time. Instead of using only multiple cameras, if multiple Kinects are used, it gives us far more information to create 3D virtual models of the remote scene. We experimented with multiple Kinects to extract dense point clouds by merging point clouds extracted from each of the Kinect sensors. In this case, we first calibrate each of the Kinect sensors and then extract the individual point clouds from each of the Kinect

![The single input image of a building.](image1)

![The 3D reconstructed model of the building.](image2)

![The input images for dense 3D reconstruction.](image3)
sensors by using its depth and color images. Then, we align
the point clouds from two or more Kinects by first finding
the 3D points of the 2D correspondences obtained from the
color images of each view and then finding the relative pose
between pairs of views. Figures 6(a) and 6(b) show the color
images before and after alignment from two Kinect data sets
that have 60% overlap in their views. Figures 7(a) and 7(b)
show the difference images before and after alignment of the
same data.

We also experimented with HD cameras and Kinect to see
how information fusion of the images from HD camera along
with the depth information from Kinect improves the quality of
3D reconstruction of the virtual model. Although the Kinect
sensor is capable of generating a 3D point cloud using an
RGB camera and an IR camera but the resolution and the
amount of details in the point cloud is low as it picks up
textures from an image of resolution $640 \times 480$. To enhance
the resolution and therefore, the amount of detail, we use the
textures from two HD cameras and are able to get the details
which are lost by using only the Kinect color images. We first perform cross calibration of the color and depth cameras of Kinect and calibrate and register the two HD cameras also, and then using this information we combine the views from the Kinect sensor and the two HD cameras. Figures 8(a) and 8(b) show the Kinect input images from the RGB and IR camera, respectively while Figure 9 shows the result of cross calibration of the color and depth images from the Kinect sensor. Figure 10(a) shows the result of registration between the Kinect and the first HD camera while Figure 10(b) shows the result of registration between Kinect and the second HD camera. Figure 11(a) and 11(b) shows the epipolar rectification of the two HD cameras respectively. Figures 12(a) and 12(b) shows the result of registration before and after combining data from Kinect and the two HD cameras, respectively.

Reconstruction of the 3D environment is important for creating an immersive environment and giving the user the perception of being in the remote environment. We have seen that there are various methods for reconstruction and that information fusion from multiple cameras as well as sensor of various modalities help in recreating a detailed virtual model of the remote environment. However, for teleoperation, we also need to reconstruct the detailed structures of the objects in the remote environment, to enable the user to command the robot or exoskeleton to manipulate these objects. Once the robot has manipulated the object, it should also be possible to get the correct 3D view of the object in its new orientation. We explore the use of an image-based range scanner using structured light for getting the detailed 3D information from small objects to be able to reconstruct their 3D models. It requires calibrating the scanner and the work environment to ensure that the exposure settings are good and that the object is focused properly. The object should be scanned from multiple sides for complete coverage of the object. This results in a point cloud, or a mesh or a solid surface. Then these meshes or point clouds need to be aligned to get the complete 3D model of the object. We carried out experiments with two small objects. Figure 13 shows multiple scans of a doll from various orientations.

III. IMMERSION AND HAPTICS

As part of the immersive environments for teleoperation, virtual and augmented reality play an important role. As discussed before, haptics also plays an important role since it is necessary to have the haptic and force feedback senses for the user to control and react to the environment in which he/she is immersed. For immersion, we have experimented with 3D displays wherein the user needs to wear special goggles to
view in 3D and are looking into head-mounted displays for a deeper feeling of being immersed in the environment. For haptics and force feedback, we first developed a joystick based simulation tool wherein the user controls a peg that has to be put into a hole using the joystick. Currently, we are in the process of including force-feedback into this system, so that the user gets a more precise feel of the movement and collision avoidance can be improved. The objective of this simulation is to create a simulation for manipulation of objects in the remote environment via a robotic arm/gripper, which can convey the sense of force, contact and touch of gripper in the immersed virtual environment.

In our simulation, we have incorporated collision avoidance
to enable the checking of possible collision between various objects during teleoperation. In the particular experiment, we ensure that the peg moves into a hole by checking if it is colliding with a solid surface or not and avoid collisions using this module. For the haptic interface for teleoperation, currently we use a joystick and phantom omni-haptic device. An operation of the device will lead to teleoperation in the remote environment. Moreover, force feedback is being incorporated so that if objects collided in the remote environment, the force is felt by the operator. This conveys the sense of contact and touch.

For a better perception of immersion, our system supports stereo vision to provide for more realistic 3D viewing. As the user wears an HMD device, and manipulates the haptic device, he/she will feel as if they are present in the remote environment and will be able to carry out the desired operations more robustly. The simulation is also used in a training and feedback loop so that if the exoskeleton makes a mistake, it can be corrected by the user in the virtual environment using the haptic device and real-time reconstruction of the action performed by the exoskeleton/robot. Figures 14 and 15 give snapshots of the simulation of a user putting a peg into a hole using an haptic device.

IV. Conclusion

Virtual reality for robotics is an active area for research, specially since it is important for creating immersive environments for teleoperation. Teleoperation allows the user to remotely control a robot or an exoskeleton, to perform action in the remote environment based on the action that the user initiates. This is specially useful in areas where it is difficult and dangerous for human operator to enter such as for disaster management, in nuclear power plants, etc. To be able to create immersive environments for teleoperations, virtual models of the remote environment has to be created and we have looked into creating virtual 3D models using both presence of a priori scene knowledge and single view as well as using multiple views where there is no a priori scene knowledge. We have also looked at information fusion from depth camera, HD cameras, image based range scanners and normal color images for enhanced and detailed reconstruction of the virtual 3D environments and objects in these environments. For teleoperation and a complete feeling of immersion, we are looking into using haptic and force feedback devices which allow the user the feeling of manipulating real devices in the remote area and also allows the user to control the robot in the remote environment with better precision. We are also exploring the development and interaction in the immersive environments using 3D displays and head mounted devices,
to give the user a complete immersive experience.

V. ACKNOWLEDGEMENT

We would like thank the Board of Research in Nuclear Sciences (BRNS), India for supporting our research work and providing a common platform to nurture cutting edge research in the area of virtual reality and robotics. We also thank Brojeshwar Bhowmick, Abhinav Shukla, Suvam Patra, Neeraj Kulkarni, Krishna Chaitanya and Amit Agarwal, Ph.D and M.Tech. students in the Department of Computer Science and Engineering, IIT Delhi for their valuable contribution towards the research in the areas of virtual reality, 3D reconstruction and use of haptic devices and 3D displays for creating immersive environments.

REFERENCES


AUTHOR INFORMATION

Ayesha Choudhary did her B.A (Honors) in Mathematics from Delhi University, New Delhi in 1998, M. Sc. (Mathematics) and M. Tech. (Computer Applications) both from Indian Institute of Technology, Delhi in 2000 and 2002 respectively. She did her Ph.D. (Computer Science) from the Indian Institute of Technology Delhi in 2011 in computer vision and machine learning. The title of her thesis is “Automated Analysis of Surveillance Videos”. She has been working as a research scientist in the Department of Computer Science and Engineering, IIT Delhi since 2008. Her research interests include computer vision, machine learning, activity and event analysis, linear algebra and robotics. For more information please visit http://www.cse.iitd.ac.in/~ayesha

Prem Kumar Kalra is a professor in Department of Computer Science and Engineering at the Indian Institute of Technology, Delhi. Earlier, he was at MIRALab, University of Geneva (Switzerland). He obtained his PhD in computer science from Swiss Federal Institute of Technology, Lausanne in 1993. His research interests include computer vision based modeling and rendering, 3D visualization and animation, and image/video super-resolution. He has published over 60 papers in international journals and conferences. He is in the Editorial Board of the journal The Visual Computer (Springer) and has been program committee members of many international and national conferences such as ICCV, ACCV, CGI, and ICVGIP. For more information about his research activities, please visit http://www.cse.iitd.ac.in/~pkalra

Subodh Kumar is an associate professor in Department of Computer Science and Engineering at the Indian Institute of Technology, Delhi. He did his Ph.D from the University of North Carolina, Chapel Hill. His research interests are in the areas of 3D interactive graphics and large scale walkthroughs. He has developed a number of Computer Graphics software. He has been on the programme committee of a number of International Conferences. He has extensive experience in designing virtual reality systems. For more information about his research activities, please visit http://www.cse.iitd.ac.in/~subodh

Subhashis Banerjee did his B. E. (Electrical Engineering) from Jadavpur University, Calcutta in 1982, and M. E. (Electrical Engineering) and Ph. D. from the Indian Institute of Science, Bangalore, in 1984 and 1989 respectively. Since 1989 he has been on the faculty of the Department of Computer Science and Engineering at I.I.T. Delhi where he is currently a Professor. He is also the Head of the Computer Services Centre. His research interests include Computer Vision and Real-time Embedded Systems. For more information about his research activities, please visit http://www.cse.iitd.ac.in/~suban
Vision-based Robotic Control

Vision is an important source of information for controlling a robotic manipulator. It is assumes greater importance for situations where it may not be possible for a human being to physically control an industrial task, such as perform surveillance, to identify objects of interest, pick them up, and place them in a controlled manner. The group has worked in this area for quite some time. The work in this area can be grouped into the following areas:

1. **3-D object recognition with an active camera**
   In this area, the group first considered recognising objects which cannot be uniquely recognised from a single view alone. One has to plan a suitable set of views around the object, to recognise it uniquely. To this end, the group examined two possible cases: first, when the complete object lies within the field of view of the camera, and second, when only a part of the object may be visible. In the first case, the work involved an efficient aspect graph (an encoding of different views of an object)-based planning strategy to uniquely identify the given object. In the second problem, the robot may view only a small part of a large 3-D object. The group proposed a part-based strategy to identify the given object uniquely. In both cases, the probabilistic algorithms handle uncertainty better than what a deterministic algorithm would.

2. **Analysis of unusual activities using a pan-tilt-zoom camera network**
   An automated system is advantageous over a manual one, wherein an operator may have to go through hours and hours of canned video, in order to detect unusual activities. This is off-line. In an on-line case, this needs attentive human operators, which may sometimes not be possible. The group's work in this area includes building distributed algorithms for calibration and composite event recognition of events observed by a network of PTZ (pan-tilt-zoom) cameras, possibly mounted on robotic vehicles. The work also proposes new strategies for automatically learning usual events and automatically detecting unusual events using unsupervised and semi-supervised learning as well as using video epitomes and probabilistic latent semantic analysis (pLSA).

3. **Optimal camera placement for surveillance of large spaces**
   A robotic vehicle with a camera (among many such entities) needs to align itself in line with many others of the same type, in order to have proper coverage (for surveillance) of large spaces. The task is further complicated when the objects in the scene move around. The group has worked on theoretical aspects of camera placement for surveillance of large spaces.

4. **Vision-guided control of a robotic manipulator**
   This project aims to develop vision based robust object detection and pose identification system which can handle occlusion. The project will start with vision-based systems alone – to use visual information to identify and estimate the pose of realistic industrial objects. This will be subsequently extended to using a laser range sensor, in addition to a visual sensor (camera). The project will examine the relative accuracy of both cases, in view of the different capabilities of the two sensors. A related task is to develop intelligent view planning by the robotic manipulator, and visual servoing mechanism for vision-guided control of the manipulator. This has a few sub-tasks. One will be with a camera fixed on the manipulator. The final one will have a hand-eye coordination for picking up a pellet and pushing it into a cylindrical hole.

---

**Sponsored Research Projects**

2. “Vision-Guided Control of a Robotic Manipulator”, funded by BRNS (2010-15)

**Recent Student Research Projects**


**Facilities**

1. KUKA KR5 Arc Industrial Robot
2. Bassler Pilot Cameras
3. Laser Range Scanner

**Selected Publications**


**Faculty**

1. Santanu Chaudhury, Professor, E.E
2. Subhashis Banerjee, Professor, C.S.E
3. Sumantra Dutta Roy, Assoc. Professor, E.E
Active 3-D Object Recognition through Next View Planning

Sumantra Dutta Roy
EE, IIT Delhi
sumantra@ee.iitd.ac.in

Santanu Chaudhury
EE, IIT Delhi
schaudhury@gmail.com

Subhashis Banerjee
CSE, IIT Delhi
suban@cse.iitd.ac.in

Abstract—This article presents an overview of work on active 3D object recognition at the Indian Institute of Technology, Delhi. We have concentrated on the use of simple features and suitably planned multiple views to recognise a 3-D object with an uncalibrated camera. We use isolated planned views, without incurring the overhead of tracking the object of interest, across views. Our work has focussed primarily on two areas: aspect graph-based modelling and recognition using noisy sensors; and recognising large 3-D objects using Inner Camera Invariants. We have proposed new hierarchical knowledge representation schemes in both cases. A common thread in both is a novel robust probabilistic reasoning-based reactive object recognition strategy which is scalable to memory and processing constraints, if any. We present results of numerous experiments in support of our proposed strategies.

I. INTRODUCTION

3-D object recognition is a difficult task primarily because of the loss of information in the basic 3-D to 2-D imaging process. Most model-based 3-D object recognition systems consider features from a single image, using properties invariant to an object, and preferably, invariant to the viewpoint. We often need to recognise 3-D objects which because of their inherent asymmetry (in any set of features: geometric, photometric, or colour-based, for example), cannot be completely characterised by an invariant computed from a single view. In order to use multiple views for an object recognition task, one needs to maintain the relationship between different views of an object. In single-view recognition, systems often use complex feature sets, which are not easy to extract from images. In many cases, it may be possible to achieve unambiguous recognition using simple features, and suitably planned multiple views [1], [2].

A single view of a 3-D object often does not contain sufficient features to recognise it unambiguously. Objects which have two or more views in common with respect to a feature set, may be distinguished through a sequence of views. As a simple example [1], [2], let us consider the set of features to be the number of horizontal and vertical lines, and a model base of polyhedral objects. Fig. 1(a) shows a given view of an object. All objects in Fig. 1(b) have at least one view which correspond to two horizontal and two vertical lines.

A further complication arises if the given 3-D object does not fit inside the camera's field of view. Fig. 2 shows an example of such a case. The view in Fig. 3(a) could have come from any of the objects in Fig. 3(b), (c) and (d). Even if the identity of the object were known, one may often like to know what part of the object the camera is looking at – the pose of the camera with respect to the object. Single-view recognition systems often use complex feature sets, which are associated with high feature extraction costs, which in itself, may be noisy. A simple feature set is more applicable for a larger set of objects. In many cases, it may be possible to achieve recognition using a simpler feature set and suitably planned multiple observations.

An Active Sensor is one whose parameters can be varied in a purposive manner. For a camera, this implies a purposive control over the external parameters (the parameters \( \mathbf{R} \) and \( \mathbf{t} \) describing the 3-D Euclidean transformation between the camera coordinate system and a world coordinate system) and the internal parameters (given by the internal camera parameter matrix \( \mathbf{A} \), composed of parameters: the focal lengths in the \( x- \) and \( y- \) image directions, the position of the principal point, and the skew factor). Our work in active 3-D object recognition has primarily considered two areas namely,

1) Aspect Graph-based Modelling and Recognition using Noisy Sensors
2) Recognition of Large 3-D objects through Next View Planning using Inner Camera Invariants

The following sections give an overview of our work in this area.