Towards Sensor Enhanced Virtual Reality Teleoperation in a Dynamic Environment

Muthukkumar S. Kadavasal
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/me_conf

Part of the Computer-Aided Engineering and Design Commons, and the Graphics and Human Computer Interfaces Commons

Recommended Citation
Towards Sensor Enhanced Virtual Reality Teleoperation in a Dynamic Environment

Abstract
A teleoperation interface is introduced featuring an integrated virtual reality based simulation augmented by sensors and image processing capabilities on-board the remotely operated vehicle. The virtual reality system addresses the typical limitations of video-based teleoperation caused by signal lag and limited field of view, allowing the operator to navigate in a continuous fashion. The vehicle incorporates an on-board computer and a stereo vision system to facilitate obstacle detection. It also enables temporary autonomous operation of the vehicle for local navigation around obstacles and automatic re-establishment of the vehicle's teleoperated state. Finally, the system provides real time update to the virtual environment based on anomalies encountered by the vehicle. System architecture and preliminary implementation results are discussed, and future work focused on incorporating dynamic moving objects in the environment is described.

Keywords
Virtual Reality Applications Center, Human Computer Interactions Program, virtual reality, augmented reality, mixed reality, teleoperation

Disciplines
Computer-Aided Engineering and Design | Graphics and Human Computer Interfaces

Comments
TOWARDS SENSOR ENHANCED VIRTUAL REALITY TELEOPERATION IN A DYNAMIC ENVIRONMENT

Muthukkumar S. Kadavasal*  
Human Computer Interaction Program  
Virtual Reality Applications Center  
Iowa State University  
Ames, IA 50010

James H. Oliver  
Department of Mechanical Engineering  
Virtual Reality Applications Center  
Iowa State University  
Ames, IA, 50010

ABSTRACT

A teleoperation interface is introduced featuring an integrated virtual reality based simulation augmented by sensors and image processing capabilities on-board the remotely operated vehicle. The virtual reality system addresses the typical limitations of video-based teleoperation caused by signal lag and limited field of view, allowing the operator to navigate in a continuous fashion. The vehicle incorporates an on-board computer and a stereo vision system to facilitate obstacle detection. It also enables temporary autonomous operation of the vehicle for local navigation around obstacles and automatic re-establishment of the vehicle’s teleoperated state. Finally, the system provides real time update to the virtual environment based on anomalies encountered by the vehicle. System architecture and preliminary implementation results are discussed, and future work focused on incorporating dynamic moving objects in the environment is described.

CR Categories and Subject Descriptors: virtual reality, augmented reality, mixed reality, teleoperation.

1. INTRODUCTION

Teleoperation can be broadly defined as controlling a system from a distance. One of the primary motivations behind teleoperation research is the need to perform tasks in places that are unsuitable for human presence. For example, using unmanned aerial vehicles (UAV) for reconnaissance in hostile regions [1], researching the ocean floor without risking a diver’s life [2], and exploring a damaged nuclear reactor using a teleoperated ground vehicle, are a few examples in which teleoperation can play a vital role. These scenarios require spontaneous and critical decision making which cannot be carried out by autonomous agents, so human participation is important [3].

Teleoperation of a vehicle typically involves direct visual feedback (such as the hobbyists RC airplane), or indirect visual feedback from an onboard video cameras or laser scanners [4]. Figure 1 presents a schematic architecture for a video-based vehicle teleoperation system. Such systems are subject to time lag in the transfer of the video feed as well as the commands. Assuming that a system has a constant lag of $t$ seconds, then every command sent to the vehicle and every frame of video sent to the operator station are received $t$ seconds after the actual event. When the vehicle environment is dynamic then the operator input and vehicle reaction are not

* Author of Correspondence: Ph: (515) 294 3092, email: ksmkumar@iastate.edu

Figure 1: Architecture for Video Based Teleoperation
intuitively linked in time resulting in the operator’s potential loss of situational awareness [5]. Moreover, the video feed obtained from the camera provides a limited or “soda straw” view of the environment due to the camera’s limited field of view (FOV). Lack of peripheral vision due to the limited FOV can be compensated by adding more cameras and sensors [6]. This approach requires the operator to pay attention to several different video feeds simultaneously and create a consistent mental image of the world [7]. This increases the operator’s stress and distracts her/his focus away from the task at hand. Researchers have been working on providing integrated environment data by augmenting multiple sensors [8]. Ricks et al [9] developed “ecological” displays which allow users to navigate in 3D worlds with integrated range and camera information. However, the various data streams from multiple sensors and cameras may be subject to different and variable lag, so synchronizing them before they are presented to the user can be challenging.

These limitations are motivating research to improve teleoperation interfaces. Augmented reality (AR) and virtual reality (VR) technologies are enabling some of the most novel new teleoperation interfaces. The most common AR technique, sometimes referred to as synthetic imagery, involves overlaying and registering text and images (e.g., generated from sensor information) onto a live video feed or computer generated scene. AR interfaces have been applied in assembly and maintenance processes, where the instructions and reference lines can be superimposed over video or graphics representation of models [10]. AR interfaces have also been effective for teleoperation of vehicles for tunnel inspections [11]. The sensor data, registered and overlaid onto video images help operators identify cracks and holes. Although AR helps in providing more information to the operator, it cannot compensate for the loss of situational awareness due to the time lag in the system and the lack of peripheral vision. Fong et al [12], provides a flexible user interface that uses a Personal Digital Assistant (PDA) for teleoperating a ground robot. The PDA navigation tool displays a fusion of collected sensor data overlaid on a map in order to improve operator situational awareness and employs an event driven selective display of images to limit the effects of bandwidth consumption. This interface, however, is subject to the effects of lag caused by transfer and fusion of sensor data, and lack of field of view. Virtual reality based interfaces can play vital role in such situations.

2. VIRTUAL REALITY TELEOPERATION

Milgram et al [13] reported a review on VERO (Virtual Environments for Remote Operations), a virtual reality interface for controlling and manipulating telerobotic systems. The virtual model in the system is updated periodically using sensor data. The review suggests that VR interfaces provide variable perspective viewing and are far more flexible than AR interfaces. Walter et al [14] presents a virtual reality-based teleoperation system. Walter’s approach was developed for ground vehicles and uses a large-scale immersive virtual environment as the primary visual context for the operator which is augmented with sensor-generated meta-data. This provides a broad FOV that fosters situational awareness. The system accommodates lag by essentially enabling the operator to control a simulated vehicle in the future of the actual vehicle: providing it a time series of goal states.

A schematic of Walter’s virtual teleoperation architecture is depicted in Figure 2. The operator’s commands are sent to a vehicle simulation that predicts the dynamic state of the virtual vehicle including its position, velocity, acceleration and heading. The vehicle dynamics simulation produces the simulated state, which is used to position the virtual vehicle and provide a desired location for the teleoperated vehicle. The idea of driving the simulated vehicle and making the teleoperated vehicle follow is based on the wagon tongue path planning algorithm [15]. The teleoperated vehicle uses the simulated states as a series of goal states. A simulation run locally on the vehicle determines the inputs required to get the vehicle to approach the simulated state from its current state.

![Figure 2: Architecture for VR Teleoperation](image)

To calculate these inputs, the current state of the real vehicle (real state) is required. The “observer”, an optical tracking system in Walter’s implementation, provides the real state. To assist the operator in assessing the deviation between virtual and physical manifestations of the vehicle, an “informed state” is computed as the difference in vehicle positions between the simulated state and the real state. The informed state is used to generate a virtual box surrounding the simulated vehicle that grows or shrinks depending on the magnitude of this discrepancy. This wire-frame envelope shown in Figure 3, allows operators to adjust their control to obtain higher fidelity with the remote vehicle, closing the loop between the human and the computer controlling the remote vehicle. Since the operator drives the simulation instead of the real vehicle he/she will not be required to accommodate for the lag that leads to the loss of situational awareness. Walter’s tests compared the camera-based teleoperation to VR teleoperation with imposed one-way signal lag times of 1, 5 and 10 seconds during a task involving navigation through a set of cones. Metrics included total navigation time and gates missed (or cones hit). Results indicate that as lag increased,
camera-based teleoperation became much slower overall, since successful navigation demanded small inputs followed by a wait for results. For the same level of navigation performance, the time to complete the course using VR teleoperation was independent of lag.

![Figure 3: Graphical Representation VR Teleoperation](image)

3. SENSOR ENHANCED VR TELEOPERATION

Virtual reality teleoperation separates the real and simulated states, and thereby ameliorates the interface challenges caused by signal lag time. The system enabled far better navigation performance than video based teleoperation. Of course, state separation assumes the primacy of the virtual world created a priori, and that the operator believes what is perceived through the simulation. Research in terrain simulation and modeling has evolved sufficiently to provide three dimensional graphics model from satellite data, just short of real time [16]. Hence, the virtual terrain can be very accurate. However, the possibility of an operating environment being different from its virtual representation is high in dynamic environments and change might occur in both time and space.

Consider a nuclear facility that maintains a teleoperated vehicle to do daily maintenance. The entire nuclear facility is modeled but the vehicle lacks any mechanism to perceive the environment in real time. If a situation arises in which the facility is damaged, then part of the vehicle’s world model may be inaccurate leading to false assumptions from its remote operator and potentially to an accident. Further research is required to address virtual teleoperation in a dynamic environment. The challenge lies in identifying ways to detect environmental change relative to the virtual model of the environment, use this information to enable the vehicle to adapt to the change, and provide the operator with the dynamically updated environment.

The multibehavior-based mobile robot developed by Luo et al [17] can be teleoperated based on video feed. The robot’s onboard computation enables obstacle detection and path planning in circumstances where the operator cannot intervene. However, the entire teleoperation is carried out for a preplanned path with known destinations. In addition, the path planning is computed based on an a priori environment model, with pre-stored goal states when the robot becomes autonomous. Although, such a system is essentially a path navigator with obstacle avoidance rather than a teleoperated vehicle, it demonstrates the idea of vehicle adaptation to accommodate surprises.

The research presented in this paper builds on Walter’s VR teleoperation approach by integrating on-board vehicle sensors to enable it to adapt to dynamic environments. In addition, the world model is subsequently modified to provide the operator with a dynamically updated virtual environment. The system retains all the components of Walter’s VR teleoperation system, thus maintaining the advantages of accommodating lag and limited FOV. However, the real vehicle in this system is augmented with sensors and significant onboard computational power to support obstacle detection system and decision making. The resulting system is essentially a fusion of VR teleoperation with autonomous obstacle avoidance.

Sensor augmentation is the prerequisite for any vehicle to perceive the surrounding environment in real time. Considerable research has been reported on sensor fusion interfaces, where multiple sensor data from the real vehicle are integrated and presented to the operator. NASA Ames Research Center [18], has conducted an extensive study on developing interfaces using real time sensor data. The images from the surface stereo imager fitted on the vehicle, is processed to provide photo realistic terrain models of the interior. The terrain model developed will facilitate future mission planning and analysis. Research by Jarvis et al [8] and Ricks et al [9] suggests sensors varying from CCD cameras to laser range finders for acquiring information real time. However, it is noteworthy to understand that these proposed systems are modeled for teleoperating vehicles in completely unknown environments, where the teleoperator relies entirely on the lagged data and images. In the proposed approach, an overview of which is presented in Kadavasal et al [19], a sensor augmented vehicle is teleoperated based on an a priori model in a virtual environment. The immediacy of the sensory data coupled with a certain degree of vehicle autonomy will not only help the vehicle adapt to dynamic environments, but retain the edge over other teleoperation systems in overcoming time lag and limited FOV.

There are a wide range of sensors with varying characteristics that are available for depth measurement and it is necessary to understand their advantages and limitations before making a selection. Meier et al [20] and Fong et al [12] present comparative reviews on a range of depth measurement techniques including stereo vision, laser range finders and sonar. The papers suggest that stereo vision provides good angular resolution with low cost and high speed. The disparity map technique using coordinated stereo images is effective for detecting small objects. However, it is unfit for detecting objects that are too close or far away from the cameras. Moreover, lack of textures in the scene and low lighting may result in extremely noisy depth resolution. Sonar, on the other hand can detect objects that are far away and are not affected.
by environmental lighting. However, sonar has poor angular resolution and is prone to error caused by non perpendicular and off axis targets. Further, specular reflections may result in range errors and poor depth resolution. Laser scanners are predominantly used in various teleoperation systems for obstacle avoidance. They have good depth resolution and are not affected by the environmental limitations. But they do have low update rates when compared to other vision systems and cannot detect smaller obstacles. Our current prototype system is developed for a lighted indoor environment with small static and moving obstacles. Stereo vision based sensor systems are suitable for such situations. The system architecture explained in the following paragraphs employs a stereo vision system for obstacle avoidance.

### 3.1. Architecture

Figure 4 shows the proposed system architecture for sensor enhanced VR teleoperation. The simulated and real states are separated and the dynamics engine acts as the vehicle control and generates the control inputs for the simulated vehicle. The real vehicle follows the simulated vehicle using the wagon tongue technique. However, the vehicle is now augmented with two onboard synchronized cameras, a tracking system and onboard computation for image processing. These components act as the vehicle’s senses.

Synchronized stereo vision allows the vehicle to identify any object within a stipulated distance and creates a warning. The warning informs the operator about the new object in the travel path along with the distance to the object and its dimensions and coordinate positions in state space. It also provides an estimated time to collision. The new object is computed as the difference between the real and pre-modeled environment and placed in context in the virtual environment. This update is intended to provide the operator with visual reference for the next time the vehicle is operated in the vicinity of the new object.

With the new object detected and a warning issued, the real vehicle detaches from the wagon tongue algorithm and becomes autonomous. The nearest goal position that is along the actual path but sufficiently clear of the new object is identified using the vehicle adaptation system. The autonomous vehicle reaches the intermediate goal position and reattaches itself to the wagon tongue, i.e., the vehicle again follows the simulated vehicle’s path and is no longer autonomous. The operator will be informed about the new path and the wire frame box around the simulated vehicle will be updated to denote the degree of the vehicle’s deviation its simulation. The individual components of this architecture are described further below.

#### 3.1.1. Obstacle Avoidance System - Stereo Vision

The stereo vision system is comprised of two Unibrain firewire cameras that are connected in series and synchronized. It simulates a low level human eye, which can see and perceive the 3D world [21].

![Figure 4: Architecture for Sensor Enhanced VR Teleoperation](image-url)
The images from the two cameras produce different perspectives of the same scene, which helps in calculating the difference in relative displacement of the objects in the scene. This relative displacement is referred to as disparity. Simple projective geometry shows that the amount of disparity is inversely proportional to the depth of a point in the scene [22]. For example, a cross section of the imaging geometry is illustrated in the Figure 5. The optical centers of the two cameras are aligned and parallel to the horizontal X-axis. The focal lengths \( f \) of both the cameras are assumed to be equal. The distance between the optical centers is \( b \).

\[
Z = \frac{f \cdot b}{D}
\]

Where \( D \), the disparity is calculated as the difference between \( d \) and \( d' \).

A fast stereo matching algorithm is necessary to calculate the disparities between the images in real time. The resultant disparity map should have object surfaces detailed and distinguished as separate regions with minimal depth discontinuities. There are numerous stereo matching algorithms in the literature, but they do not satisfy the requirements imposed by VR teleoperation completely.

Zitnick et al [23] presents a cooperative algorithm to compute disparity using correspondence. This iterative algorithm identifies the match within the predefined 3D space and accounts for occlusion. However, the algorithm in practice takes about 8 seconds per iteration for a 256 x 256 image size. The maximum flow formulation N-Stereo algorithm by Roy et al [24] is another stereo correspondence algorithm that computes precise depth maps albeit with relatively large computational time. Such high time costs may not be suitable for a teleoperation system where sensory data is required to perceive the environment around the vehicle in real time.

The stereo correspondence method adopted in our vision system is based on Birchfield at al’s [25] pixel by pixel stereo matching algorithm. The algorithm estimates the disparity values by matching the pixel intensities of the images. The effective pruning technique (to remove unlikely search nodes) proposed in this approach, coupled with dynamic programming reduces the computational time significantly. The algorithm introduces methods to identify non-textured regions and, achieves a balance between computational time and depth map precision.

In order to provide faster stereo matching, the algorithm assumes that the images from the left and right camera are aligned along the horizontal axis. This can be achieved by image rectification. The intrinsic and extrinsic camera calibration parameters are computed and the images are rectified. The stereo correspondence algorithm computes the disparity map from the rectified images, example results of which are shown in Figure 6. The figure shows the camera image along with the computed disparity map. The process rate for the disparity map is approximately 2 Hz. The disparity results are calculated for environments that contain solid, transparent, curved shape and/or textured objects. The algorithm proves to be effective enough to provide precise object surfaces with distinguished separate regions.

However, there is generally a considerable amount of streaking and noise in the images. The disparity map is then converted into a depth map using projective geometry. An optimum threshold is computed to identify the nearest objects.

![Figure 5: Stereo Image Formation](image)

Consider a point \( P (X, Y, Z) \) in the scene that is imaged by the left and right cameras. Using similar triangles the depth of the point \( P \) can be computed as

\[
Z = \frac{f \cdot b}{D}
\]

Where \( D \), the disparity is calculated as the difference between \( d \) and \( d' \).

![Figure 6: Disparity Maps for sample images](image)
The objects are segmented using a region growing method and its dimensions are calculated. Since the stereo vision system is affected by the environmental lighting and the camera setup, the resulting depth maps may be noisy. To differentiate between noise and the real objects, a Kalman filter [26] tracking system is developed. Currently the stereo vision system is operational and is under testing. However, increasing the reliability of this system, tracking the objects in the vicinity based on the vehicle’s kinematics are some of the research challenges that are currently being addressed.

3.1.2. Vehicle Adaptation

The real vehicle does not have knowledge about the state space and is completely controlled by the operator. Hence, in the autonomous state, the vehicle has to either rely on its on-board sensors for determining appropriate control inputs or the information it received previously from the operator’s control or some combination of both. There has been considerable research in sensor augmentation and vehicle autonomy. However, the prime research goal here is not for developing an autonomous vehicle that can survive in an unknown environment, but to develop a system that can be teleoperated using virtual reality as a tool (to accommodate lag and provide field of view) and at the same time adapt to the partially unknown dynamic environment and increase the operator’s degree of confidence.

This research proposes an optimized path finding method that identifies paths after correlating and synchronizing the previously available terrain knowledge and risks, with the new environment data. The architecture for the proposed vehicle adaptation system is provided in Figure 7a. In the proposed system, the a priori model state space is classified into various zones depending on the level of risks as shown in Figure 7b. It is assumed that the terrain data and risks are continuously updated within the operator’s environment from various information resources (e.g., newly found enemy assets).

The vehicle operation can be classified before hand with respect to the level of caution that is necessary. For example, driving a remote vehicle in a terrain that has suffered from an earthquake will have a different caution level when compared to driving a remote vehicle inside an enclosed space like a building. The caution level indicates the degree to which the vehicle can take chances in precarious situations. This free parameter will be preset by the human operator for the specific operation.

The virtual world provides the real vehicle with a risk map of the neighboring region corresponding to each goal state. The method will account for the real state error. The autonomous vehicle will have the risk map for the current position and relate the new object position to the risk map. The path planning method will then identify the new path for the vehicle based on the actual goal state (simulated state), risk levels of the neighboring zones and the preset caution level. Depending on the preset caution level, the vehicle will either consider a high or moderate or low risk neighboring zone as the alternate path. The autonomous vehicle then reaches the intermediate goal position and reattaches itself to the wagon tongue, i.e., the vehicle follows the simulated vehicle’s path and is no longer autonomous. Although, the vehicle acts autonomously while adapting to the environment, the system follows the strategic hierarchy devised by the operator. In such a system, the operator can predict the autonomous vehicle’s actions easily, thereby reducing the reaction time of both the operator and the vehicle considerably.

![Figure 7a: Vehicle Adaptation](image1)
![Figure 7b: Risk Zone Classification](image2)
When the autonomous vehicle fails to make a decision, the vehicle stops and informs the operator about the obstacle position and coordinates in state space. The operator may then take over the vehicle control and arrive at a decision. This situation results in a context switch from simulation to real in the operator’s environment. Research will be carried out to identify the best possible way to provide the new data to the operator.

3.2. Implementation

Figure 8 shows a schematic representation of the proposed sensor enhanced VR teleoperation system. In the schematic, the simulated vehicle is shown in green and the real vehicle in black. The teleoperator drives the simulated vehicle from the CAVE [27]. The way points are sent to the real vehicle, denoted by red dots and the real vehicle follows the simulation. The dark gray objects are obstacles present in the a priori VR environment model and the brown object is the newfound obstacle unknown to the teleoperator. The real vehicle operation is divided into two time steps. The stereo vision system on board the vehicle detects the new obstacle in time step one. The on board computation assists the vehicle in identifying the corrected path in time step two. The path planning process for this time step is explained in section 3.1.2. This new path is shown in the figure as black dots.

The initial prototype vehicle is built on a toy radio controlled car platform, and controlled by a motor servo control phidgets and a Microsoft sidewinder force feedback wheel. Two Unibrain synchronized Fire-wire cameras are connected to the onboard mini-ixt mother board. Currently the vehicle is tested for camera based live video teleoperation using wireless internet connectivity. The video is transmitted at 30 fps at 620 X 480 resolution with an average lag of 1 second. The stereo vision obstacle detection system is tested for depth reliability and object tracking. Figure 9 presents the depth reliability results for the stereo vision camera. The data is collected in static camera conditions for two different light settings. The results show that the stereo vision system is reliable for identifying small obstacles in indoor conditions. The depth resolution decreases for increasing distance between the camera and the object. The stereo vision reliability is currently under testing for moving camera conditions, with potential problems being motion blur and reduced frame rate. Further, a VR environment of an open building space is modeled and the image processing algorithm is designed using Open Computer Vision (OpenCV [28]). The prototype of this system is operational and currently in testing. The researchers are also planning to extend this model further to accommodate moving objects in state space.

Figure 8: Schematic Representation
4. DISCUSSION AND CONCLUSION

The proposed system for sensor enhanced VR teleoperation retains the human control necessary for decision making while providing considerable autonomy for the vehicle to accommodate surprises encountered in its immediate vicinity. In this way, the operator will not experience the loss of situational awareness due to lag and lack of peripheral vision while navigating a dynamic environment, thereby increasing accuracy and utility. By allowing the vehicle to temporarily detach from the simulated state during the warning period, the operator continues driving in the simulated state with additional knowledge about the real state in the form of the wire frame box. Preliminary results indicate that the operator experiences marginal disruption when the vehicle temporally assumes autonomous operation. Hence, this sensor enhanced VR teleoperation interface is expected to be more adaptable and intuitive when compared to other interfaces.

ACKNOWLEDGEMENTS

This research is supported by United States Air Force Research Laboratory, Grant Number FA9550-05-1-0384. The authors are grateful for the support of colleagues Derrick Parkhurst, Eliot Winer, Alex Stoytchev, Bryan Walter, Jared Knutzon, and Tom Batkiewicz and for the research infrastructure provided by the Virtual Reality Applications Center.

REFERENCES


[16] Collins, R. T., “Site model acquisition and extension from aerial images,” Technical Report, University of Massachusetts Department of Computer Science, Amherst, M.A.


