

Semi-Autonomous Vehicle Teleoperation Across Low-Bandwidth or High-Delay Transmission Links

Jennifer Kay
School of Computer Science
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh PA 15213
phone: (412) 268-6498
fax: (412) 268-5571
email: jennie@cs.cmu.edu

Charles Thorpe
Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh PA 15213
phone: (412) 268-3612
fax: (412) 268-5571
email: cet@cs.cmu.edu

Abstract

Guiding a remote vehicle when real time image transmission is not possible is an important problem in the field of teleoperation. In such a situation, it is impractical to put an operator behind a steering wheel and expect accurate steering. In semi-autonomous teleoperation, an operator designates the path that the vehicle should follow in an image of the scene transmitted from the vehicle, and the vehicle autonomously follows this path. Previous techniques for semi-autonomous teleoperation require stereo image data, or inaccurately track paths on non-planar terrain. STRIPE is a method for accurate semi-autonomous teleoperation using monocular image data. By combining techniques in computer control, artificial intelligence, and intelligent user interfaces, we are designing a unique system for remote operation of a vehicle across low-bandwidth and high-delay transmission links.

1. Introduction

Driving a vehicle, either directly or remotely, is an inherently visual task. When heavy fog limits visibility, we reduce our car's speed to a slow crawl, even along very familiar roads. In teleoperation systems, an operator's view is limited to images provided by one or more cameras mounted on the remote vehicle.

Traditional methods of vehicle teleoperation require that a real time stream of images is transmitted from the vehicle camera to the operator control station, and the operator steers the vehicle accordingly. For this type of teleoperation, the transmission link between the vehicle and operator workstation must be very high bandwidth (because of the high volume of images required) and very low latency (because delayed images cause operators to steer incorrectly).

In many situations, such a high-bandwidth, low-latency communication link is unavailable or even technically impossible to provide. For example, current non-line-of-sight radio transmission technology does not provide the bandwidth required, and inter-planetary distances can delay communication by several minutes or more. A variety of image compression techniques have been developed to provide a higher frequency of images across low-bandwidth data links, but there is a limit to how much an image can be compressed before important data is lost.

As the frequency and immediacy of images transmitted to the operator decreases, the operator's ability to directly control the vehicle diminishes. At some point the operator's accuracy becomes so low that it is not practical to try to control a vehicle in such a manner.

When remote driving is required, but direct control is not possible, a "semi-autonomous" approach must be used. In semi-autonomous control, the operator provides the high-level decisions about where the vehicle should go, and the vehicle takes over the task of actually controlling the vehicle's steering. Because there is limited bandwidth and potential delays in the transmission system, the operator should be able to direct the vehicle based on a single image of the scene, and the vehicle should accurately follow the prescribed path for some distance while a new image is being transmitted. Other semi-autonomous teleoperation systems, which will be discussed in section 2., require stereo image data, or stray from the intended path on non-planar terrain. The STRIPE (Supervised TeleRobotics using Incremental Polyhedral Earth Reprojection) system is being developed to use only monocular image data and accurately follow a designated path on varying terrain.

In STRIPE, a single image from a camera mounted on the vehicle is transmitted to the operator workstation. The operator uses a mouse to pick a series of "waypoints" in the image that define a path that the vehicle should follow. These 2D waypoints are then transmitted back to the vehicle, where they are used to compute the appropriate steering commands while the next image is being transmitted. STRIPE does not require any advance knowledge about the geometry of the terrain to be traversed.

STRIPE is a unique combination of computer and human control. The computer

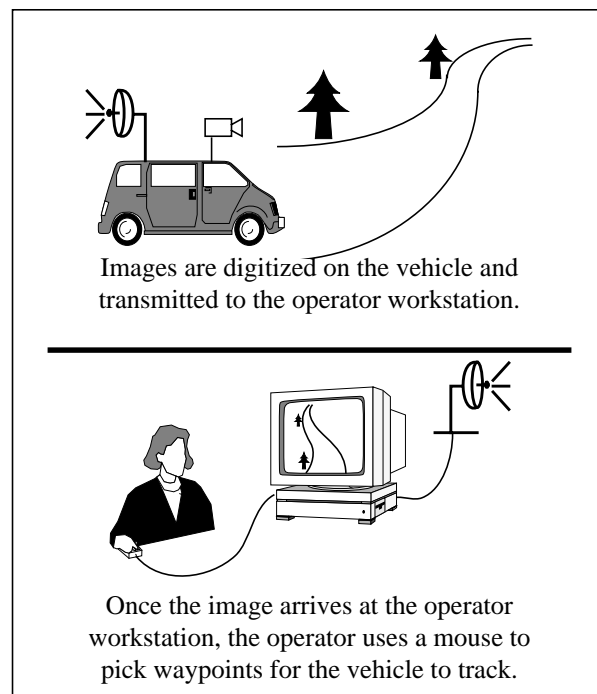


Figure 1. The Basic STRIPE Scenario

must determine the 3D world path designated by the 2D image path and then accurately control the vehicle over rugged terrain. The human issues involve accurate path selection, and the prevention of disorientation, a common problem across all types of teleoperation systems. STRIPE has been implemented on the Carnegie Mellon Navlab II, an army Hmmwv that has been reconfigured to be completely computer controllable. It has been successfully demonstrated on a variety of terrain types and we are beginning a series of experiments to improve the operator's ability to accurately designate waypoints in the images and to reduce operator disorientation.

2. Other Low-Bandwidth, High-Delay Systems

Depiero [2] demonstrated that human operators can directly and effectively control the steering of a remote vehicle if they are provided with approximately five images per second. The vehicle simulator community has found that delays of as little as a few hundred milliseconds in image data can be detrimental to an operators performance and physical well being [3]. Clearly, a transmission delay of one hour between image capture and display on the operator workstation is much too large to enable any sort of realistic steering of a teleoperated vehicle. Even a transmission delay of 1 minute per image is too long. It seems as though

transmission rates of anything over several tenth of a second per image make direct driving impossible, and some kind of automatic vehicle control becomes necessary. This is not simply a theoretical problem. Consider a system for remotely operating a vehicle on Mars, for example, where transmissions to Earth can be delayed by about 15 minutes due to the distance between the planets. Or terrestrial systems that use current military standard tactical communication links, which only transmit 16 to 24 Kbits of data per second [2], a rate several orders of magnitude lower than that needed for real-time image transmission.

Computer Aided Remote Driving (CARD) [6] is a system for semi-autonomous control of a remote vehicle. A stereo pair of images is transmitted to the operator from the vehicle. The operator uses a 3D mouse to pick waypoints on a stereo display. Since these waypoints are chosen in a 3D image, no 2D to 3D conversion is necessary, and no assumptions must be made about the shape of the terrain. These waypoints are transmitted back to the vehicle, where they are used to compute the appropriate steering directions. This method is very effective in situations where transmitting two images has approximately the same cost as transmitting a single image, as in high-bandwidth space applications. However, when the delay is not only due to distance, but also to low-bandwidth transmission, sending two images might take twice as long as sending a single image, and this may make the delays unacceptable.

The Feedback Limited Control System (FELICS) enables an operator to remotely drive a vehicle with a single image transmitted as infrequently as every 3.5 seconds [8]. A triangular “puck”, a cursor that represents the vehicle’s location, is superimposed on the image, initially just in front of the vehicle’s position. The operator uses a knob (to control heading) and a stick (to indicate speed) to “drive” the puck through the image, thereby establishing the path the vehicle should follow. At predefined intervals a marker is placed in the image along the puck’s path to mark a “waypoint” (i.e. a point towards which the vehicle should steer).

In order to steer the vehicle, FELICS must convert the 2D image waypoints into 3D real-world points. This conversion is accomplished as follows: At the time of image capture, the location of the vehicle’s “groundplane”, the plane on which the vehicle’s wheels rest, is noted. Figure 2 shows a side view of the vehicle, the path to be traversed, and a vastly oversized pinhole camera representing the camera on the vehicle. Points on the image plane (i.e. 2D waypoints) can be reprojected through the focus out into the world, and their intersection with the groundplane (represented by the x’s) is considered to be their location in the 3D world. The resulting 3D waypoints are transmitted back to the vehicle, where they are used to control the steering.

While FELICS works well enough at 3.5 seconds per image to keep the vehicle on

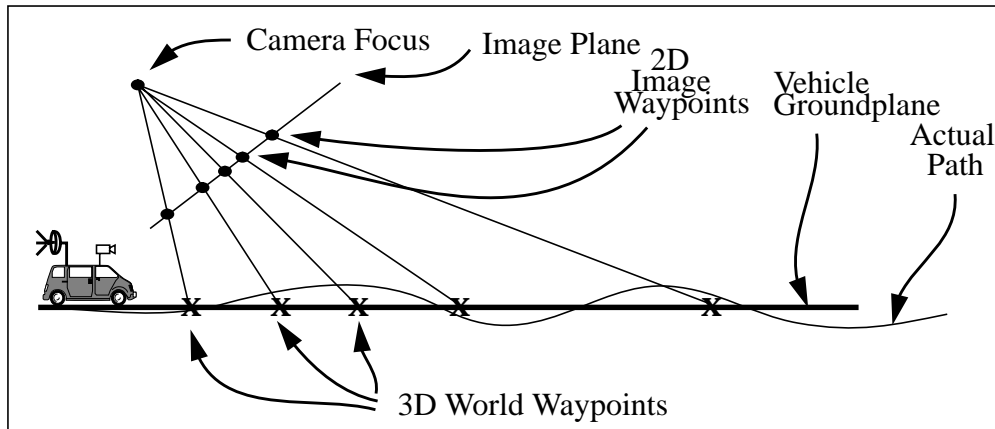


Figure 2. FELICS Flat-Earth Reprojection

a wide road, its flat-earth projection method can cause errors on non-planar terrain. These errors can become increasingly problematic as image frequency decreases. For example, consider the scene depicted in Figure 3. The initial groundplane of the vehicle is horizontal, but the road ahead has a slight slope of 3 degrees. Suppose a sequence of points has been chosen by the operator, and a point 50 meters ahead of the vehicle on the road has been designated as the point where the vehicle should stop. With FELICS's simple flat-earth approach, all points are projected onto the current groundplane, and the vehicle will stop 25 meters ahead of its initial position, i.e. where the image point projects onto the current groundplane.

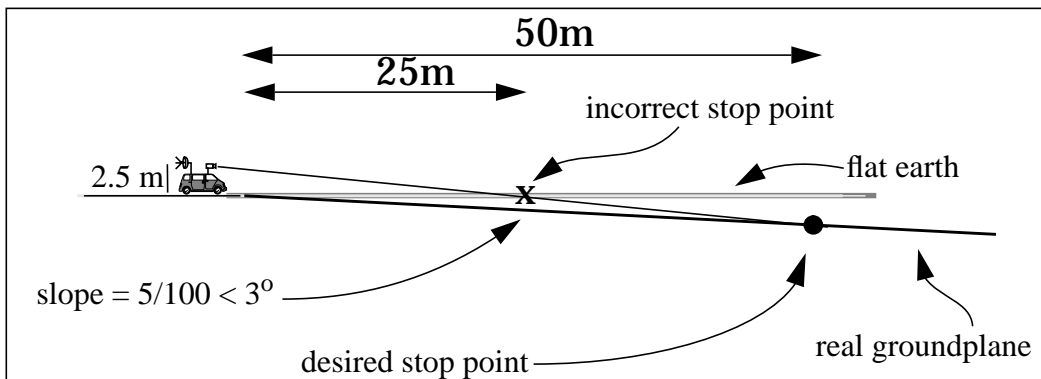


Figure 3. Flat Earth Reprojection

3. The Current STRIPE System

STRIPE [5] is a semi-autonomous teleoperation system that uses only a single image per cycle, but avoids the inaccuracies of FELICS's flat-earth reprojection.

The basic STRIPE system consists of 3 modules:

- The “Image Capture” module, that runs on the vehicle.
- The “Operator Workstation” module, that runs at the base station.
- The “Main STRIPE” module, that runs on the vehicle.

A single image from a camera mounted on the vehicle is digitized by the Image Capture module. The image is then transmitted from the vehicle to the Operator Workstation module, where it is displayed on a monitor. The operator then uses a mouse to pick a series of “waypoints” in the image that designate the path the vehicle should follow (see Figure 1). These 2D points are transmitted back to the Main STRIPE module on the vehicle.

In order to compute the vehicle’s steering, the Main STRIPE module must convert these 2D image waypoints into 3D real-world points. Clearly a simple flat-earth reprojection, such as that used by FELICS, is not sufficiently accurate.

When the 2D waypoints are transmitted to the vehicle, they are initially projected onto the vehicle’s current groundplane. The resulting 3D waypoints are used to initiate steering of the vehicle, and the vehicle starts to move. Several times a second the vehicle re-estimates the location of its current groundplane by measuring vehicle position and orientation. The original image waypoints are then projected onto the new groundplane to produce new 3D waypoints and the steering direction is adjusted appropriately. This reproject-and-drive procedure is repeated until the last waypoint is reached, or new waypoints are received.

For example, consider the side view of a road shown in Figure 4. The vastly oversized pinhole camera represents the camera on top of the vehicle, with the 2D waypoints on the image plane. The intersection of the projections of each of the 2D waypoints with the actual terrain is the location of the “actual” 3D waypoint the operator intended the vehicle to track.

Initially, the STRIPE Main module reprojects the 2D waypoints onto the vehicle groundplane, as shown in Figure 5. After the vehicle has begun to move, it updates its prediction of the waypoints by reprojecting the original waypoints onto the new groundplane as shown in Figure 6. Figure 7 shows the vehicle after it has passed several of the waypoints.

STRIPE has no advance knowledge of the 3D locations of all of the waypoints. However, as the vehicle approaches a particular waypoint, the vehicle’s groundplane becomes an increasingly accurate approximation for the plane that the waypoint lies on. By the time the vehicle needs to steer based on that particular waypoint, it has a precise knowledge of where that point lies in the 3D world. Note that all of this movement was based on a single image (and so our pinhole camera

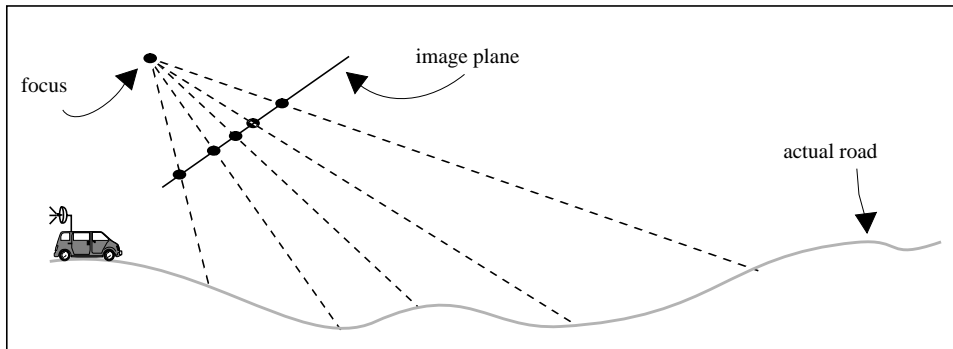


Figure 4. Side view of a road and the corresponding image, with designated waypoints. The giant pinhole camera represents the camera on top of the vehicle. Points on the image plane of the giant camera represent the 2D waypoints chosen by the operator. The dotted lines show where those image points project onto the actual terrain. The intersection of each of the dotted lines with the actual road is the “correct” location of the 3D waypoint.

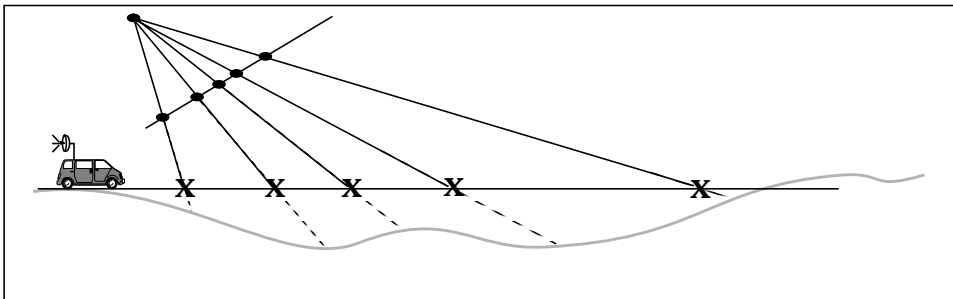


Figure 5. The 2D waypoints are initially projected onto the vehicle's groundplane. The x's represent the location of the projected point.

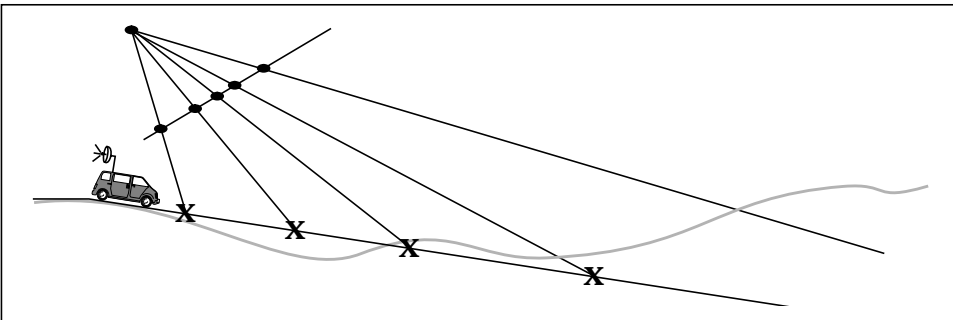


Figure 6. The same waypoints are reprojected onto the vehicle's new groundplane.

does not move throughout this segment).

It is important to emphasize that the flat-earth assumption that STRIPE makes is very different from standard flat-earth systems. Because the vehicle module is directly in charge of steering the vehicle, it does not need to plan all of the steering

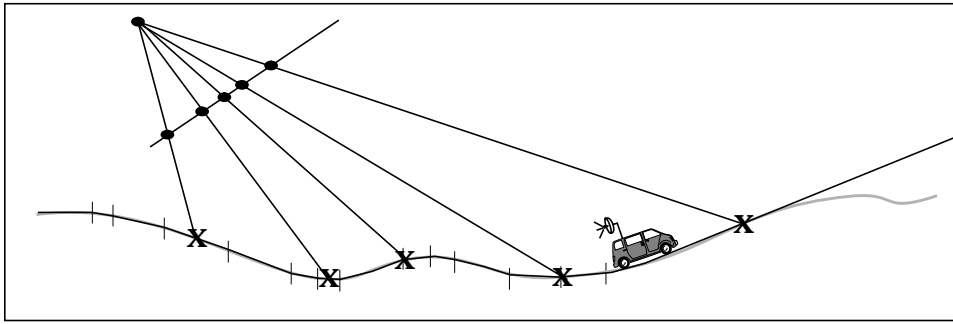


Figure 7. The vehicle continues its reproject-and-drive procedure. The path behind the vehicle has been divided into small planar regions.

commands it will use in advance. Instead, it can compute the steering direction as it needs it. In the STRIPE path tracking algorithm, a “look-ahead distance” is a predefined constant value. The vehicle steers by finding the point on the desired path that is approximately one look-ahead distance away from the vehicle, and heading towards that point.

How does STRIPE’s local flat-earth assumption compare with the global one? Again, consider the scenario depicted in Figure 3. In the STRIPE scenario, we would stop at the appropriate point. Suppose we had a look-ahead distance of 5m. Initially, STRIPE would predict that a point that is really 5.56m ahead on the road was only 5m ahead, and begin to drive. However, as soon as the vehicle had rolled onto the new groundplane, subsequent predictions would be correct, and the vehicle would stop accurately at the stop point. This example is particularly illustrative of the problems that even a very mild change in groundplane orientation can cause. By restricting one’s flat earth assumption to the look-ahead distance, the accuracy can be significantly improved. The “Main STRIPE” module on the vehicle runs at about 80 Hz. on a Sun Sparc 10.

4. Investigating the Interface

4.1. Operator Difficulties

Experience with the more traditional high-bandwidth and low-delay teleoperation shows that operators have a tendency to get disoriented and lost fairly easily, even with landmarks and a map of the area [1][7]. In our case, low-bandwidth and high-delay teleoperation, the problem still exists, and the low frequency of images may actually aggravate it. Because there is a significant delay between images, operators cannot use the image flow to get a feel for the camera orientation. Without knowing the orientation of the camera relative to the vehicle, there is no

way to determine where any given point in the image is relative to the vehicle.

For example, consider an on-road driving scenario. With the camera pointing straight ahead, the operator decides to make a left turn at an upcoming junction. When the vehicle is still some distance from the junction, the images sent back from the camera pointing straight ahead provide a view of the current road as well as the junction, and the operator easily picks points to move the vehicle towards the junction. When the vehicle gets closer to the junction, the camera must be panned to the left to see a better view of the road behind the junction, and an image is digitized. That image appears on the operator work station along with a number on the control panel indicating the camera pan and tilt. New operators seem to have difficulty using the numerical pan/tilt values to help interpret the image that they see. A typical operator comment is, “If I pick this point in the image, how sharply will the vehicle turn?” The problem seems to be one of understanding which point in the real world corresponds to a chosen point in the image. This does not seem to be as much of a problem in higher-bandwidth teleoperation systems, probably because the more frequent display of new image data provides the operator with a better sense of the vehicle’s motion relative to the camera direction.

Fixing the orientation of the camera on the vehicle does not solve the problem of choosing an appropriate path. For example, consider the images in Figure 8 and Figure 9. Both are images taken from a vehicle driving along a road with a fixed camera. In Figure 8, the path to choose to continue driving along the road is quite clear. However, Figure 9 was taken at an intersection where a single image provides few clues as to where to steer.



Figure 8. An “easy” road.



Figure 9. A “difficult” intersection.

4.2. Improving the Interface

There is almost no empirical data available on the performance of low-bandwidth high-delay systems. In order to improve the performance of novice users, as well as create a standard of comparison for other systems, we have developed a series of empirical tests to help us determine how different variables affect an operator's skill at the teleoperation task.

The variables we will be studying include bandwidth, latency, field of view, resolution vs. image frequency, and color images vs. monochrome. Some of these have been studied in the context of high-bandwidth and low-delay teleoperation systems [4][7], but there is no way to extrapolate those results to the low-bandwidth and high-delay scenario.

4.2.1. Bandwidth and Latency

At the core of our system are the difficulties encountered by a low-bandwidth and high-delay transmission link. Our maximum bandwidth is limited by that of the cellular modems that we are using for operator to vehicle communication (9600 bps), and we will also perform tests at reduced bandwidths. Latency (over and above delays due to restricted bandwidth) will be varied from 0s to 120s.

4.2.2. Vehicle Test Course

It is important to conduct these tests over a variety of test courses, both to determine the overall robustness of the system, as well as the influence of the course on other variables. We have developed a set of three test courses that are designed to test an operators skill at obstacle avoidance, map following, and following "obvious" paths. The type of test course has been shown to have an influence in some tests of more traditional teleoperation systems. For example, Glumm et. al. [4] found that operators preferred using a lens with a wider field of view on their obstacle avoidance test course.

4.2.3. Field of View

It seems likely that the influence of the field of view of the camera on a STRIPE operator will be very different from that of an operator of a traditional vehicle teleoperation system. In traditional systems, as in simulators, operators report feelings of discomfort and disorientation [3][4]. In a low-bandwidth and high-delay system, the fact that the images are discrete appears to eliminate these problems.

4.2.4. Operator Interface

The current STRIPE operator interface consists of three windows. One window shows the image transmitted from the vehicle, the second is a message window that provides the operator with instructions, and the third is a control panel that allows the operator to adjust camera angle, and start and stop the system.



Figure 10. The current STRIPE operator interface.

To attempt to reduce the problems of disorientation, and to improve an operator's ability to understand the orientation of the camera with respect to the vehicle, we have developed two additional interfaces.

The idea behind the dashboard interface (Figure 11) is that individuals might better understand the orientation of the camera with respect to the vehicle if they see the image as if through the window of the vehicle. Camera angle could be adjusted by using the mouse to move the outline of the image from left to right or by entering a numerical value on the control panel.

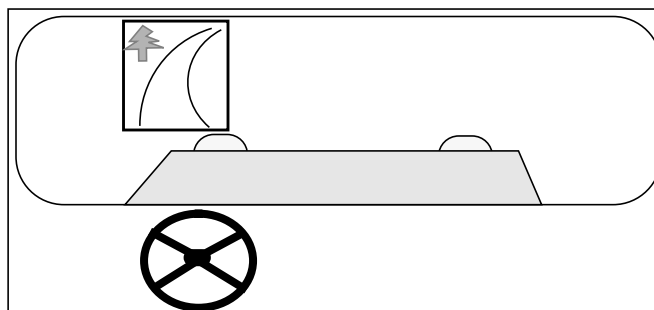


Figure 11. Dashboard Interface.

The compass interface (Figure 12) presents an overhead view of the vehicle with the camera on top, and lines indicating the field of view of the camera. The orientation of the camera can be altered by using the mouse to grab one of the field of view lines and move them (the lines move together) or by entering a number on the control panel. The orientation of the vehicle would remain fixed, regardless of the vehicle's actual orientation.

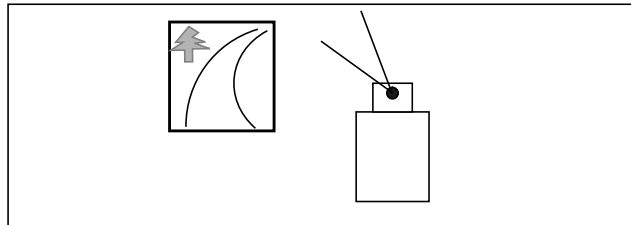


Figure 12. Compass Interface.

4.2.5. Vehicle Control System

In addition to studying the standard STRIPE control system, we will study an operator's ability to use a steering wheel to directly control the vehicle across low-bandwidth and high-latency links. We will also test how quickly and accurately a normal driver in the vehicle can perform the given tasks.

4.2.6. Resolution vs. Image Frequency

When considering low-bandwidth data transmission, there is a trade-off between image resolution and frequency. Lower resolution images can be transmitted with higher frequency, but do not provide as much information for the operator. Higher resolution images provide more detail, but at the cost of reduced image frequency. We will test three different image resolutions: 640x480 pixels, 320x240 pixels, and 160x120 pixels, and their corresponding relative frequencies.

4.2.7. Color vs. Monochrome Image Data

Anecdotal evidence suggests that color can be useful for humans teleoperating vehicles on Earth. However, color data is typically two to three times the size of monochrome data, and would thus reduce image frequency.

4.2.8. Method

Clearly attempting to implement a detailed study of all of the independent variables described above would be a mammoth task. Initially we will perform small-scale tests of the various independent variables to determine which seem to

have the most dramatic effect on an operator's ability to accurately control a remote vehicle. Following these tests, the two variables that seem to have the most dramatic effect on an individual's ability to accurately operate a remote vehicle will be combined with the various vehicle test courses, three alternate STRIPE interfaces, and bandwidth vs. latency, and a more detailed series of empirical tests will be performed.

5. Conclusion

STRIPE is a unique combination of computer control, artificial intelligence, and intelligent user interaction. An important factor in this approach is its technological simplicity and low cost. The technology required for high-bandwidth data transmission is often very complex and expensive. STRIPE uses only simple components: monocular video display, mouse input, inertial sensor, and minimal computing power. This research not only provides a working system for teleoperation across low bandwidth and high latency links, it will also provide important data about a human operators ability to accurately direct a vehicle under such conditions.

6. Acknowledgments

This work is funded under TACOM contract DAAE07-90-C-R059, Topographic Engineering Center contract DACA76-89-C0014, and by a NASA graduate student research fellowship.

7. Bibliography

- [1] Aviles, W.A., Hughes, T.W., Everett, H.R., Umeda, A.Y., Martin, S.W., Koyamatsu, A.H., Solorzano, M.R., Laird, R.T., and McArthur, S.P., "Issues in Mobile Robotics: The Unmanned Ground Vehicle Program Teleoperated Vehicle (TOV)", *Proceedings of Mobile Robots V*, SPIE, Boston, MA, November 1990.
- [2] DePiero, Frederick W., Noell, Timothy E., and Gee, Timothy F., "A Video Transmission System for Low-Bandwidth Remote Driving", *Proceedings of the ANS Fifth Topical Meeting on Robotics and Remote Systems*, Knoxville, TN, April 1993.
- [3] Frank, Lawrence H., Casali, John G., and Wierwille, Walter W., "Effects of Visual Display and Motion System Delays on Operator Performance and Uneasiness in a Driving Simulator", *Human Factors*, vol 30, no. 2, April 1988.

- [4] Glumm, Monica M., Kilduff, Patricia W., and Masley, Amy S., “A Study of the Effects of Lens Focal Length on Remote Driver Performance”, ARL-TR-25, Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD, November 1992.
- [5] Kay, Jennifer, and Thorpe, Charles, “STRIPE: Supervised TeleRobotics Using Incremental Polygonal-Earth Geometry”, *Proceedings of the 3rd International Conference on Intelligent Autonomous Systems*, Pittsburgh, PA, February 1993.
- [6] Lescoe, Paul, Lavery, David, and Bedard, Roger, “Navigation of Military and Space Unmanned Ground Vehicles in Unstructured Terrains”, *Proceedings of the Third Conference on Military Robotic Applications*, September 1991.
- [7] McGovern, Douglas E., “Experiences and Results in Teleoperation of Land Vehicles”, Sandia Report SAND90-0299, Sandia National Laboratories, Albuquerque, NM, April 1990.
- [8] Rahim, Wadi, “Feedback Limited Control System on a Skid-Steer Vehicle”, *Proceedings of the ANS Fifth Topical Meeting on Robotics and Remote Systems*, Knoxville, TN, April 1993.