Experimental Validation of Operator Aids for High Speed Vehicle Teleoperation

Alonzo Kelly, Nicholas Chan, Herman Herman, and Randy Warner

Abstract. Although fully autonomous robots continue to advance in ability, all points on the spectrum of cooperative interfaces between man and machine continue to have their place. We have developed a suite of operator assist technologies for a small (1 cubic meter volume) high speed robot that is intended to improve both speed and fidelity of control. These aids include fast stability control loops that run on the robot and graphical user interface enhancements that help the operator cope with lost peripheral vision, unstable video, and latency. After implementing the driving aids, we conducted an experiment where we evaluated the relative value of each from the perspective of their capacity to improve driving performance. Over a one week period, we tested 10 drivers in each of four driving configurations for three repetitions of a difficult test course. The results demonstrate that operators of all skill levels can benefit from the aids and that stabilized video and predictive displays are among the most valuable of the features we added.

1 Introduction

Teleoperation is a control concept that is as old as robotics — and for fundamental reasons. When the motivation for the use of a robot is to keep a human out of harm's way (e.g. nuclear servicing) or to place a robot where a human could never go (e.g. inside a blood vessel), the robot and the human are separated by assumption. Given that separation in space, the question of how they can effectively work together arises naturally. As research in autonomous robots has advanced, teleoperation has become merely one of many options, but those advances have neither rendered teleoperation irrelevant nor solved many of its fundamental challenges. Nonetheless, a robot which is more autonomous could potentially use its awareness of its surroundings and its state to conform to the needs and limitations of humans. Autonomy therefore has the potential to render the man-machine system more effective. This is hardly a new idea but different applications give rise to different

---

Alonzo Kelly · Nicholas Chan · Herman Herman · Randy Warner
Robotics Institute, Carnegie Mellon University
#10 Fortieth Street, Pittsburgh Pa, 15201
e-mail: {alonzo,ncchan,herman,rwarner}@rec.ri.cmu.edu

J.P. Desai et al. (Eds.): *Experimental Robotics*, STAR 88, pp. 951–962.
DOI: 10.1007/978-3-319-00065-7_63

realizations of the concept. This paper explores the potential of operator aiding in the context of small high speed wheeled mobile robots, with a particular emphasis on experimental validation.

Despite the promise of teleoperation, its history has been characterized by a constant struggle to solve many fundamental and difficult issues. For vehicles, the operator is removed from the rich sensory experience (visual, audio, olfactory, inertial) of sitting in the driver seat. The intrinsic limitations of the associated sensing, communication, and display technology then deprive the operator of all stimuli except a video. Unfortunately, video and associated displays is a poor surrogate for biological vision according to almost any chosen basis of comparison.

These technology limitations are responsible for both the reduced situation awareness of a remotely located driver, and the difficulty of driving competently at high speed. There is no inexpensive and effective means of remotely conveying the sensation of all of the acceleration, impact, and vibration associated with sitting in the driver seat. Likewise, it is well known that latency in the video (or any sensing) makes it very difficult to respond effectively to unpredictable disturbances in a feedback setting, whether there is a human in the loop or not.

1.1 Problem Statement

Nonetheless, displays can be annotated to include knowledge (available to the robot) that may not be discernible in the raw video. Some control loops can be closed on the vehicle where there is far less latency and others can include a predictive component that allows the operator to remove, in a classical feedforward manner, predictable errors before they occur.

Therefore, assistive technologies should be able to improve performance in a measureable way and our goal in this work was not only to implement these technologies on a challenging platform, but to measure their effectiveness in a controlled empirical setting.

1.2 Related Work

This work fuses ideas for teleoperation from the earliest days of robot manipulator research with techniques for video stabilization, ideas from modern gaming interfaces, and elements of electronic stability control (ESC). ESC is now available on most recently manufactured automobiles. We will use predictive displays, gyro-stabilized video, annotations over live video, and speed governing based on yawrate feedback. The effort described here was motivated in part by a desire to produce a second version of our teleoperation system described in [7]. We wanted this second version to be less expensive to produce and to be suitable for robots without lidar perception.

Almost three decades ago, the field of *telerobotics* was a subfield of robotics pursuing a scientific understanding of the man-machine system. Numerous techniques for supervisory control and teleoperation of manipulators were outlined as early as the mid 1980s [12]. Virtual displays that are either predictive or used for preview have often been used to compensate for both delay and limited data bandwidth when remotely operating manipulators. The concept of *teleprogramming* was an early form of model-based teleoperation [6] that used models to mitigate the effects of latency. More recently, more intuitive and task-centric interfaces have been used to operate manipulators over thousands of miles of separation [8].

Latency compensation in space applications has also been accomplished with motion preview and predictive displays [2]. In some cases, stereo graphics viewed in a stereoscopic display have been used to improve operator depth perception [9]. While all of the telerobotics work described so far has been applied to stationary manipulators in a stationary scene, the principles are extendable to moving sensors in a dynamic scene, if the image processing is efficient. For example, Ricks et al. used a predictive method that they dubbed *quickening* to compensate for latency when teleoperating a mobile indoor robot [11].

The use of gyros to stabilize video was originally driven by the need to stabilize camcorders [10]. More recently, numerous techniques have been borrowed from computer vision which use the image data itself in order to estimate the motion of the camera [3]. Of course, inertial and visual cues can also be used at the same time [14].

One early use of augmented video like ours is augmented reality — the introduction of synthetic components into a largely real view. The display used may be head worn, handheld, or projected on a display surface. Numerous applications have been explored in medicine, manufacturing, visualization, entertainment and the military [1]. In our case, the live video is the real part and the graphical and textual annotation is the augmentation. When the augmentations are based on a rich underlying model, they are said to be knowledge-based [5]. Augmented video is also a favorite form of display in computer gaming. In that case, the video is also virtual.

Electronic stability control and roll stability control systems have been the subject of intense development by the automotive industry in the last 20 years. Our approach to ESC is a governor based on yaw rate error whereas active automotive systems are based on using braking to generate restoring moments [13]. Our approach to roll stability control is based on early work from legged mobile robots as realized in the algorithms in [4].

2 Technical Approach

The work described here investigates a number of techniques that promise to improve the performance of the man-machine system. These techniques can be grouped into those related to communications, control, and operator display. After describing the hardware and the overall rationale, the more important algorithms are described below.

2.1 Hardware Design

Our work was conducted on the Forerunner remote-controlled vehicle developed by RE^2 Inc (Figure 1). We chose this vehicle for its size and speed regime (max speed 25 km/hr/). The base platform provides control interfaces and computing to support

Fig. 1 Forerunner Remote Controlled Vehicle. All wheels are driven and steered. The vehicle is easily capable of 25 km./hr. speed so such ballistic motion is achievable.

line-of-sight remote control. To this, we added a Core 2 Duo ULV 1.2 GHz CPU computer, a high dynamic range driving camera, an automotive obstacle avoidance radar, an inertial mfeasurement unit, and 802.11-n wifi. The main computer was used to implement control algorithms and video compression.

Inertial navigation was performed in dedicated FPGA hardware. Another remote computer at the operator control station generated the graphical user interface. Our communications date rate was limited to a mere 0.75 megabits (not bytes) per second. Our approach for tolerating this limit was to use the latest (MPEG-4) video compression and to tune it for this application. Doing so permitted us to optimize video quality within the available resources.

2.2 Control Techniques

Several control techniques were used to assist the operator. A stability control system was used to help reduce risk of loss of yaw stability and of rollover, and a path following controller was used to reject associated disturbances at the vehicle level. The overall rationale for the use of these systems is as follows:

- *Feedback Control.* One of the most basic techniques is to close control loops locally on the vehicle where latency is low and reaction time is short. In this way, the vehicle is able to reject all disturbances that its feedback renders it competent to reject. The operator then has to deal only with the disturbances that remain, and these tend to be lower frequency and somewhat discernible from the operator display. In the case of safety systems like stability governing, the vehicle can be empowered to take control locally to prevent a mishap that would otherwise occur before the operator is even aware of the situation.
- *Model Predictive Control (MPC).* MPC has several uses. The most basic is the use of predictive models in multi-state control algorithms like path following. In this case the predictable effects of terrain following and wheel slip can be modelled to eliminate what would otherwise become error disturbances. Data bandwidth can be reduced as well. On-vehicle processing can be used to perform data-intensive predictions and then transmit only the results to the operator. For example, prediction calculations can include the effects of terrain slope, without having to transmit the terrain data to the operator.
- *Prediction Through Latency*. MPC can also be configured to account for predictable latencies - both uplink (of state) and downlink (of commands). Trajectory predictions performed on the vehicle can be shifted forward in time to reflect

uplink delay and predictions performed on the operator console can account for the command downlink delay. The delay itself can be predicted from past experience and/or from the measured delay of the most recent messages.

2.2.1 Inertial Navigation

We have implemented our own inertial navigation (INS) solution on other programs so we were able to re-use it for this application. The navigation system was used to provide position feedback for control purposes, to provide attitude feedback for video stabilization, and to provide specific force and angular velocity data for stability control. Use of our own INS permits us to optimize for GPS-denied performance and to integrate the solution for navigation and stability control in one package.

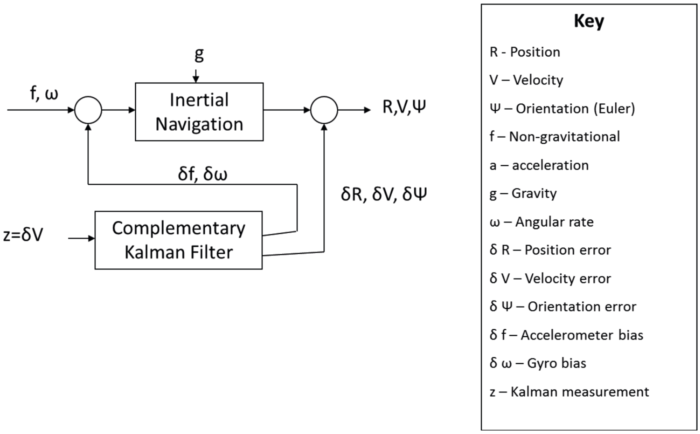

Fig. 2 Inertial Navigation System. A 15 state complementary filter configuration is used.

A relatively high performance Honeywell HG-1930 IMU was used. The design is a classical complementary Kalman filter (Figure 2) with 15 error states (position, orientation, velocity, accelerometer biases, gyro biases) and it is aided by measurements of the 4 wheel rotation rates and steer angles.

2.2.2 Stability Control

Exactly how a vehicle responds to high horizontal acceleration levels depends on at least the wheel support polygon, the center of gravity position, slope, terrain shape, and terrain friction. While the original intention was to implement a rollover prevention system, experimentation revealed that this vehicle is prone to spin out of control before wheel liftoff occurs. Accordingly, we implemented a yaw stability governor as well. It compared the commanded yawrate to the actual (as measured by the gyros in the IMU) and then imposed a computed limit on vehicle speed when the percent yawrate error exceeded a threshold.

2.2.3 Path Following

In addition to actuator level controls, and the operator's own display-based adjustments, a model predictive path following controller was implemented on the vehicle. The operator's driving commands, when converted to predicted response, are interpreted as a path to be followed. A clothoid is a path whose curvature function is of the form $\kappa(s) = a + bs$ for initial curvature a and curvature gradient b. The algorithm (Figure 3) seaches a discretized space of clothoids for the one which minimized the integrated pose error along the path and then sends the associated optimal control to the platform controller.

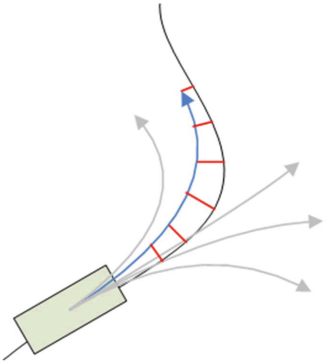

Fig. 3 Path Follower. An MPC algorithm finds the optimal clothoid.

2.3 *Display Techniques*

Several display techniques were also important for assisting the operator. The overall rationale for the use of these systems is as follows:

- *Inputs for Continuous Driving.* Of course, a very effective technique for remote driving with latency is to designate waypoints one at a time and wait until the vehicle achieves them. However, when driving continuously, the operator does not have the luxury of waiting for the display to stabilize (after motion stops) before injecting the next input — it is supposed to be a continuous process. One way to provide an intuitive input mechanism is to have the operator specify an instantaneous goal point in body coordinates. This input is static in a vehicle-fixed display and it can correspond precisely to the control horizon in MPC.
- *Predicted Path Display.* We furthermore chose to interpret the goal point as the desired endpoint of a vehicle trajectory. A predictive model of the vehicle is inverted as described above to produce the control that corresponds most closely to the desired path. The lateral position of the endpoint affects curvature and its distance downrange affects speed. The operator experience is that of literally steering this predicted endpoint. In this way, the mapping from what the operator wants to what the platform is commanded is automatic, state and terrain adaptive, and well calibrated. The result is a man-in-the-loop MPC system where the human continually adjusts the controls, optimizing on the fly, in the context of good predictions. The display discussion below reveals how the predictions are visually placed in the context of the vehicle surroundings on the screen.
- *Video Stabilization.* The context of a small vehicle driving fast over uneven terrain leads to a bumpy ride for the vehicle and a jumpy display for the operator. Accordingly, video stabilization was used to provide the operator with a

synthetically stable camera view. This feature was complementary to path prediction by providing a smoothly varying image of where the vehicle was headed, regardless of terrain following attitude changes.

2.3.1 Video Stabilization

Numerous options exist for the design of this feature. Our approach was based on our own prior efforts elsewhere because the software already existed. Each video frame was precisely tagged with the pose of the camera at the instant that the frame was acquired. The video was then rendered, based on the associated real camera attitude, onto a virtual billboard positioned a few meters in front of vehicle. The billboard was then viewed with a virtual camera at the true camera position — whose attitude was locally level (Figure 4) . The operator perspective is that the video frames may move up and down slightly on the billboard but individual features remain fixed on the display. An added feature rendered the nose of the vehicle so that its attitude could be viewed in the same display.

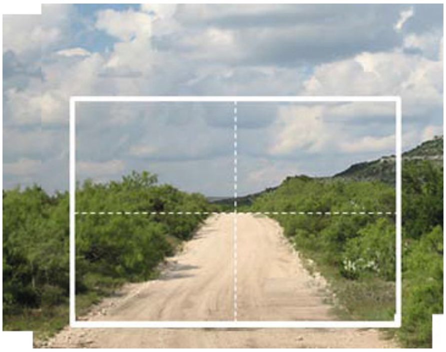

Fig. 4 Video Stabilization. A synthetic locally-level camera view is used so that terrain features remain fixed between video frames.

2.3.2 GUI with Video Overlay

The user interface (Figure 5) included the stabilized video as well as numerous overlays to convey such information as wheel slip (detected as yawrate error), radar-detected obstacles, attitude, speed, and proximity to rollover. Video overlays provide good use of screen real estate and convey extra information while permitting the operator to focus on the rapidly changing video. The predicted vehicle path was also overlaid on the video as shown. This technique allowed the operator to position the goal point precisely with respect to the objects in the scene.

Steering Mode: Ackermann
Power Mode: Velocity
Voltage: 98.85 V

Fig. 5 Graphical User interface. A video game concept is used where annotations are overlaid with some transparency on the stabilized video.

3 Experiments

It was already anecdotally clear that the assistive technology was valuable, so the test was designed to try to quantify that value. The principles of the experimental design included control of as much variables as possible while varying only the aiding modes that were available to the operator. During the testing, the speed was limited to 5 m/s and many operators achieved this speed, at times, due to the acceleration capability of the vehicle. We had damaged (and repaired) the vehicle with higher limits in earlier tests.

Communications latency was typically low, under 100 millseconds for a round trip. Path prediction reduced the effects of latency by giving the operator the means to specify a destination, rather than an immediate velocity command. A destination in front of the vehicle is still valid even if it is received somewhat late, so the controller could still attempt to reach it.

Multiple driving modes were tested in random order to remove bias associated with learning the vehicle response and the test course. Operators were unable to see the course during the test though all could see it briefly before the test. Each operator was given the same briefing on the course and the technology before the test. Effects of cloud cover, precipitation, terrain friction etc. were mitigated by testing operators in all modes in a short period of time. Effects of vegetation were mitigated by driving the course often enough to trample the tall grass before any testing.

The test course (Figure 6) was designed to be short, but quite difficult to drive without the assistive technology. Narrow driving gates were constructed from bright cardboard boxes to enhance their visibility but they were designed to collapse easily on collision to avoid damaging the robot. Their precise positions were marked on the ground to ensure repeatability of course setup, because the gates were often hit by the vehicle. Half the course was grass and the other half was (old) pavement. Operators were told to drive as fast as possible without hitting the sides of any

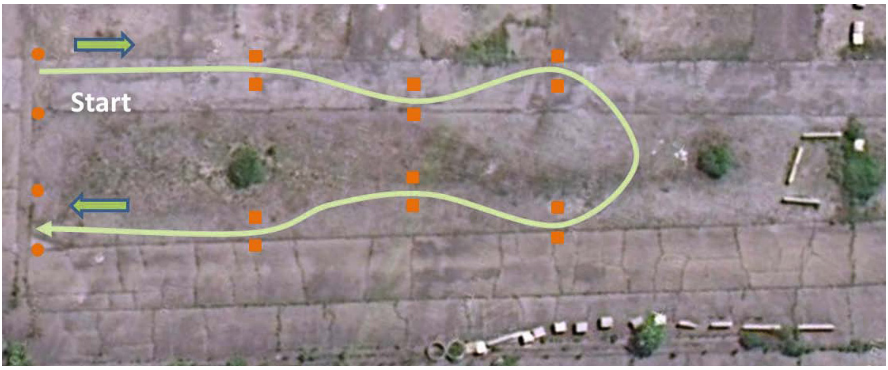

Fig. 6 Test Course - Overhead View. The small squares denote driving gates slightly wider than the vehicle. It took about 1 minute to drive the course when trying to drive fast. The course bounding rectangle is 185 ft along its longest dimension.

gates. None were given an opportunity to learn the feel of the system before the test began, although two drivers already knew the system well. All other drivers were robotics engineers with no knowledge of the system and varying experience in vehicle teleoperation. It was not our intention to evaluate learning curve. Rather, we concentrated on the effect the technology had on each driver as an individual, in the hope that it would help — regardless of skill level. Furthermore, the effects of cloud cover and sun angle, and perhaps other effects, could not be controlled over the course of the entire test. Therefore, comparisons of drivers to each other are not entirely free of such effects.

4 Results

Once initial tests determined that video stabilization and predictive display were the two most useful features, the final tests were designed to investigate these features more fully in order to produce a manageable number of tests. Ten subjects were tested and each drove the course three times in each of four configurations of the driving aids. That is, there were 12 tests performed for each of the 10 people. The configurations are summarized in Table 1. Video compression and stability control were on at all times.

Tests were conducted in mid summer at a test site in Hazelwood in Pittsburgh. We measured four principle observables: the total time to complete the course, the number of times a gate was hit by the vehicle, the number of times that the driver missed the gate entirely, and the curviness (integral of squared curvature with distance) of the path followed. The results averaged over all users are summarized in Table 2.

Table 1 Test Configurations. These four combinations of assistive features were tested.

Attribute	Basic	Stabilized	Predictive	Both
Video Stabilization	no	yes	no	yes
Path Prediction	no	no	yes	yes
Path Follower	no	no	yes	yes

Table 2 Test Results Averaged Over All Users. A clear trend of improved performance is evident with assistive features enabled, both individually and in combination.

Attribute	Basic	Stabilized	Predictive	Both
Time (secs)	49.9	44.6	39.1	36.7
# Hits	1.7	1.4	0.8	0.6
# Missed	0.3	0.2	0.0	0.0
Curviness	0.07	0.06	0.05	0.05

5 Main Experimental Insights

It is important to recognize that this was a difficult course to drive quickly and the vehicle could easily be driven beyond the speed threshold of stable control. Without video stabilization, the gates would jump around significantly in the field of view and it became less clear where the gate actually was in relation to the vehicle. Without prediction and path following, the perception of control fidelity was surprisingly low - meaning the vehicle appeared not to do what it was told to do. Latency was large enough to cause inexperienced operators to overcorrect, enter oscillation, and occasionally lose control entirely. Once the vehicle spun out of control, much time could be lost if it was already close to a gate and it had to be reversed to go through it. There was no rear camera for reverse driving. In any case, once the gate left the field of view due to a violent loss of control, the operator had to turn the vehicle in order to search the periphery of the camera field of view in order to find the gate again.

While the two most significant features added value (both individually and in combination) a fielded system would (based on our results) probably have all of them turned on, so we will concentrate on interpreting this case. With all features turned on a) 8 out of 10 users showed $> 20\%$ improvement in time, b) 7 of 10 users showed a 25% improvement in the smoothness of the path driven and c) 8 of 10 users hit fewer obstacles. Whereas 4 users missed gates entirely with all features off, no users missed gates with all features on.

In considering the assistive features independently, the following results are noteworthy. Paths were smoother with path prediction only enabled and times were faster by 5% on average with video stabilization only enabled. Also, the two experienced users showed definite improvements with the use of the assistive technology, though the improvements were less pronounced in relative terms. It is difficult to determine to what degree this reflects reduced effectiveness of the technology with more experienced users or the fact that their unaided scores were already pretty good, and therefore harder to improve upon. The two users that did not hit fewer obstacles already hit very few so the relative improvements are less meaningful.

Users were also asked to complete an informal survey to provide their impressions of the usefulness of each control model. Users found that the path prediction feature made it easier to judge the motion of the vehicle. Times were measured to be faster with video stabilization turned on and users found that the feature made it easier to see the gates. In short, the two primary operator aids were found to be both individually useful and complementary.

6 Conclusion

In this work, we have produced empirical validation of the conjecture that semi-autonomous teleoperation of (even high speed) mobile robots can produce benefits both in terms of productivity and of safety. While that is not so surprising, we have conducted experiments to try to quantify the value of such improvements and we have also assessed their value relative to each other. Our application context is that of a small, high speed, mobile robot, operating on nonflat terrain. Within that context, we have some evidence that all of the features we added were valuable. We left video compression on at all times because it is an established technique that were were not particularly interested in studying, though interesting studies have been done elsewhere. We left stability control on at all times because we felt it was too dangerous to the vehicle to do otherwise based on our preparations for the experiment. In a sense, both of these features were considered necessities for our context.

The remaining assistive features can be summarized as a control aid (path following and prediction) and a visualization aid (video stabilization). For all of our operators, regardless of skill level, these features were both individually valuable when used alone and complementary when used together. It is noteworthy that the ”all features on” configuration can be viewed as a model predictive control system with a human in the loop. Not only was the path predcted well but it was presented in the context of live video of the objects in the scene. This made it possible for the operator to literally line up the robot path with the gap between obstacles, well in advance, and then refine the path based on a continuously updated, calibrated prediction of the ”fit” of the robot to the gate. In this way, the problem becomes reduced to gently adjusting the path endpoint in a stable video rather than guessing the inputs required to make the obstacle gap appear in the center of the screen, at just the right time, as the robot drives through it.

After the tests, all operators expressed a preference to use the system with all features on at all times. While the level of improvement was not not extraordinary, we also did not try to maximize it. There are many realistic situations where the enhanced safety, higher speed, and more robust and precise control will all add up to an improved capacity to get a job done.

Acknowledgments. This CMU effort was sponsored by the U.S. Government under Other Transaction number W15QKN-08-9-000 1 between the Robotics Technology Consortium, Inc, and the Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government.

References

1. Azuma, R.: A survey of augmented reality. Presence 6(4), 355–385 (1997)
2. Bejczy, A., Kim, W., Venema, S.: The phantom robot: predictive displays for teleoperation with time delay. In: Proceedings of the 1990 IEEE International Conference on Robotics and Automation, 1990., vol. 1, pp. 546–551 (1990), doi:10.1109/ROBOT.1990.126037
3. Censi, A., Fusiello, A., Roberto, V.: Image stabilization by features tracking. In: Proceedings of International Conference on Image Analysis and Processing, pp. 665–667 (1999), doi:10.1109/ICIAP.1999.797671

4. Diaz-Calderon, A., Kelly, A.: On-line stability margin and attitude estimation for dynamic articulating mobile robots. Int. J. Rob. Res. 24, 845–866 (2005), `http://dl.acm.org/citation.cfm?id=1086867.1086870`, doi:10.1177/0278364905057865
5. Feiner, S., Macintyre, B., Seligmann, D.: Knowledge-based augmented reality. Commun. ACM 36, 53–62 (1993), `http://doi.acm.org/10.1145/159544.159587`
6. Funda, J.: Teleprogramming: Towards delay-invariant remote manipulation (1991), `http://search.proquest.com/docview/304008149?accountid=9902`, 304008149; 1991; 9125646; 746704491; 66569; 141 p.; Funda, Janez; 30906811; English; Copyright UMI - Dissertations Publishing 1991; M1: Ph.D.; M3: 9125646
7. Kelly, A., Chan, N., Herman, H., Huber, D., Meyers, R., Rander, P., Warner, R., Ziglar, J., Capstick, E.: Real-time photorealistic virtualized reality interface for remote mobile robot control. I. J. Robotic Res. 30(3), 384–404 (2011)
8. Lloyd, J.E., Beis, J.S., Pai, D.K., Lowe, D.G.: Model-based telerobotics with vision (1997)
9. Milgram, P., Zhai, S., Drascic, D., Grodski, J.: Applications of augmented reality for human-robot communication. In: Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 1993, vol. 3, pp. 1467–1472 (1993), doi:10.1109/IROS.1993.583833
10. Oshima, M., Hayashi, T., Fujioka, S., Inaji, T., Mitani, H., Kajino, J., Ikeda, K., Komoda, K.: Vhs camcorder with electronic image stabilizer. IEEE Transactions on Consumer Electronics 35(4), 749–758 (1989), doi:10.1109/30.106892
11. Ricks, B., Nielsen, C., Goodrich, M.: Ecological displays for robot interaction: a new perspective. In: Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 3, pp. 2855–2860 (2004), doi:10.1109/IROS.2004.1389842
12. Sheridan, T.: Human supervisory control of robot systems. In: Proceedings of the 1986 IEEE International Conference on Robotics and Automation, vol. 3, pp. 808–812 (1986), doi:10.1109/ROBOT.1986.1087506
13. Van Zanten, A., Erhardt, R., Pfaff, G.: Vdc, the vehicle dynamics control system of bosch. Tech. Rep. 950759, SAE (1995)
14. Zhu Xu, G., Yang, Y., Jin, J.: Camera stabilization based on 2.5d motion estimation and inertial motion filtering. In: Proceedings of 1998 IEEE International Conference on Intelligent Vehicles, vol. 2, pp. 329–334 (1998)