The Gestural Joystick and the Efficacy of the Path Tortuosity Metric for Human/Robot Interaction

Richard Voyles
University of Denver
2390 S. York Street
Denver, CO USA
1 (303) 871-2481
rvoyles@du.edu

Jaewook Bae
Samsung Corp.
1st 75 W. Plumeria Drive
San Jose, CA USA
baex0021@umn.edu

Roy Godzdanker
University of Denver
2390 S. York Street
Denver, CO USA
roygod@gmail.com

ABSTRACT
This paper describes the design and evaluation of the "gestural joystick," a wearable 2-D pointing controller for mobile robots in hazardous environments that uses hand gestures. Hazardous environments, such as that of a collapsed building search, require operators to wear a significant amount of protective clothing. This protective clothing, which may include hard hats, suits, gloves, goggles, etc., reduces comfort, mobility, dexterity, load capacity, and ability to interact with conventional computer input devices. The gestural joystick, which is embedded in protective clothing, mitigates some of these impacts, but at the cost of lesser familiarity for the user and, therefore, potentially lesser performance. Effective performance metrics are required to evaluate this interface mechanism. Path tortuosity has been proposed as a performance metric for the evaluation of teleoperation of a robot, but has not been proven to be distinct from time-to-complete metrics. By injecting controlled uncertainty between the user and robot, we show, for the first time, that path tortuosity is a useful and distinct metric for the evaluation of robot teleoperation.

Categories and Subject Descriptors
B.4.2 **[Input/Output Devices]**

General Terms
Human Factors

Keywords
Human/robot interaction, path tortuosity, gestures.

1. INTRODUCTION
Urban search and rescue (USAR) has received much interest, recently, from robotics researchers around the world. Robots have a great deal to offer in that they are more expendable than humans, can be made smaller than humans to enter confined spaces, and can tolerate harsher conditions. Robots are routinely deployed during emergency response exercises and are increasingly being deployed during actual emergencies. For instance, after the September 11th attack on the World Trade Center, robot scientists contributed to USAR teams to search for trapped victims [1]. USAR robots are currently controlled exclusively via teleoperation but traditional input devices have proven inappropriate due to the protective clothing required [2]. Traditional human/computer interaction (HCI) devices are hampered by protective clothing including safety glasses, hard hats, respirators, and, most importantly, gloves. Gloves have been found to reduce operator effectiveness in many tasks [3]. Since emergency responders are required to wear heavy gloves to insulate themselves from the hazardous environment, we have been investigating ways to embed wearable human/robot interfaces into the bulky clothing itself -- exploiting it as an asset, rather then a liability.

The prototype wearable glove-based input interface was

Figure 1: Using the Gestural Joystick to operate a commercial Inuktun robot.

PerMIS'08, August 19–21, 2008, Gaithersburg, MD, USA.

presented in [13] (and shown in Figure 1) for our USAR robot, TerminatorBot [4], which is a miniature robot for core-bored inspection tasks. This interface paradigm grew from the burdens of providing an intuitive interface and controller for the TerminatorBot with conventional interaction devices (Figure 2). Conventional input devices, such as touch screen, mouse, keyboard and joystick, are cumbersome to operators wearing safety gears, even though they are small enough to carry conveniently. Conventional glove-based interfaces, such as the Power Glove, CyberGlove, MIT LED glove and others [5], have shown to be highly accurate for the recognition of various hand gestures [6], [7], [8], [9]. However, these glove interfaces require complex and fragile sensor structures and wires and are not suitable for use in rugged environments. We are developing a new paradigm of gestural joysticking for USAR tasks that reduces encumbrances, yet remains intuitive to non-technical operators. Evaluating novel paradigms is challenging, however, and requires great care.

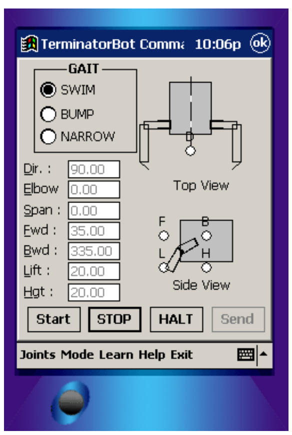

Figure 2: The initial PDA-based interface for the TerminatorBot is too confusing for non-technical users.

1.1 HRI Performance Metrics

In the evaluation of user efficacy, metrics are necessary to measure how well a rescuer can control the robot under specified tasks [10], [11]. The development of metrics for USAR robots is still in its infancy, though the standard “time-to-complete” metric is commonplace. Steinfeld [11] tried to generalize common metrics for task-oriented HRI and built a standardized framework in HRI. However, the framework is difficult to apply to such a specific domain as that represented by USAR and development is still in progress.

The time required to complete a specified, closed-ended task is a common metric for the efficacy of user interfaces. This metric is simple to implement, it results in a single, quantitative value that is easily compared, and it applies to many types of tasks and interfaces. If one interface requires the user to spend more time in completing a standardized, relevant task than another interface, it is reasonable to conclude that the former interface is inferior to the latter.

More recently, researchers have begun to examine other metrics that relate more specifically to the robot teleoperation task. It has been suggested that the degree of smoothness of resulting motion of a human-directed robot control task is a good indicator of the quality of the human/robot interface [10]. Voshell et al termed this metric “path tortuosity” [10] and quantified it through fractal dimension [12]. They used the 2-D projection of a mobile robot’s motion on the ground to show that path tortuosity correlated well with user subjective opinions and time-to-complete. However, to our knowledge, nobody has clearly demonstrated that path tortuosity is actually distinct from time-to-complete. Until now, the question has been left unanswered as to whether or not path tortuosity adds information or merely confirms time-to-complete measurements (which are easier to gather). In fact, path tortuosity often provides similar results to the easier to measure time-to-complete metric.

In this paper, we employ two different metrics to evaluate the human/robot interface for a specific task: time-to-complete and path tortuosity. In a test of non-technical users, we introduced varying degrees of uncertainty between the operator console and the robot, unknown to the operator. In effect, we were varying the quality of control they had over the robot in a controlled way, independent of the interface device. Our results provide evidence that path tortuosity is distinct from time-to-complete and provides a complimentary measure of efficacy in human/robot interaction.

2. SYSTEM CONFIGURATION

The gestural input system (see Figure 3), which we call “WRIST” for Wearable Responder Interface for Search Tasks, forgoes the complex gestural capabilities of other glove interfaces, but also eliminates the encumbrances of these other interfaces such as wires, batteries, and active components. By simplifying the interface to only the pointing task, a robust, wearable, wire-free solution results that is suitable for hazardous environments with none of the difficulties of a complex alphabet to memorize. Based on permanent magnets and sensitive Giant Magnetoresistive (GMR) detectors, analog forward/backward and right/left gestures are encoded into commands that can be relayed to a robot, personal computer, or other device requiring pointing information.

Calibration of the sensitive but highly nonlinear GMR detectors is described in [15]. User evaluation tests are presented below for a simulated search and rescue task to compare controllability to a conventional joystick. These

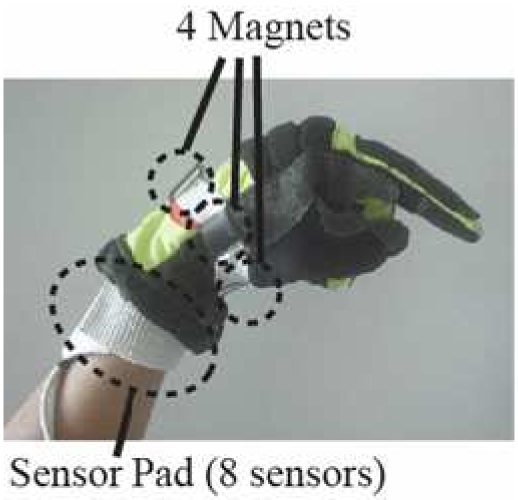

Figure 3: The WRIST system consists of an array of magnets on the glove, an array of sensors on the sleeve, and a data processor.

tests employ not only the standard metric of time-to-complete, but add a metric for path tortuosity based on the fractal dimension of the robot's traced path. Possibly for the first time, we demonstrate the independence of the path tortuosity metric from the time-to-complete metric for pointing controllability.

WRIST provides unencumbered wearability that can be easily equipped on many styles of traditional gloves with minimal modifications. Moreover, the hardware structure of the system should permit intuitive and robust operation.

The embedded rare-earth permanent magnets (Figure 4) have relatively small size and consume no power. The sensor pad is integrated with eight GMR sensors and is worn on the wrist or embedded in the fringe of the sleeve. The magnets and sensors interact with no interconnecting wires allowing the glove to be easily removed.

The signal strength of each magnetic sensor changes with wrist motion. The redundant sensor signals are continually monitored by the CPU where they are translated into the four directional components as described later. The sensor pad contains 8 GMR sensors (AA004, NVE Co., Eden Prairie, Minnesota, USA) and 8 op-amps on a flexible printed circuit board (PCB). The sensor pad can be completely wrapped around the wrist. The PCB is about the thickness of a piece of paper making it easy to embed into a sleeve.

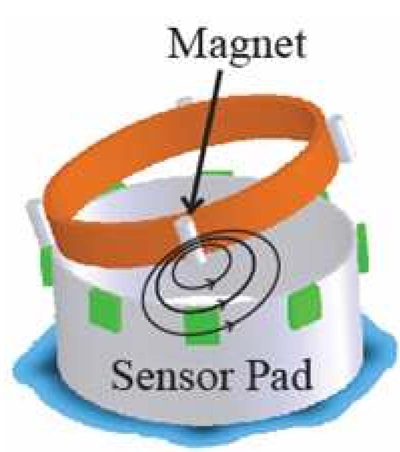

Figure 4: Configuration of the magnets and GMR sensors.

The CPU module is built around the Atmel Corporation's 8-bit AtMega128 RISC microprocessor. Its Analog/Digital converters can process eight analog signals from the eight magnetic sensors. Each Analog/Digital converts a signal into a 10-bit digital output. The glove of WRIST includes nothing but four rare-earth magnets. Therefore, it is washable and virtually unbreakable.

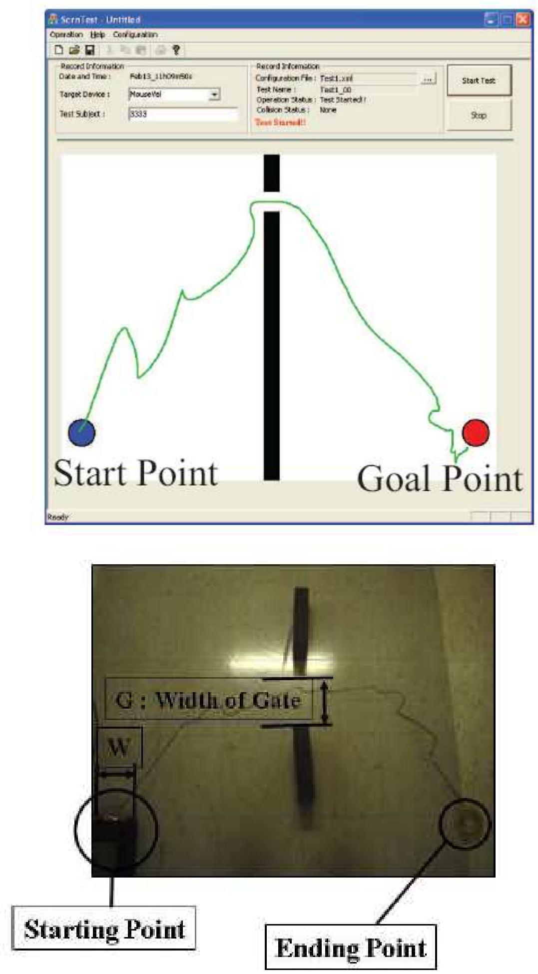

Figure 5: Usability tests based on the NIST gate model. Top is a virtual robot driven through a single gate to a goal position. Bottom is the trace of a physical Inuktun robot driven through a single gate to a goal position.

3. USABILITY EVALUATION

A variety of usability tests were developed to test operator performance with WRIST in comparison to a conventional joystick and a mouse in two different modes (position and velocity mode). Fitz' Law tests were performed as well as physical mobility tests and virtual and physical robot control tests based on the NIST "gate" model. The focus of this paper will be on the virtual and physical robot control tests.

In the NIST gate model, a single robot is teleoperated to navigate through a narrow gate, as shown in Figure 5. While the tests by NIST have used multiple gates, we use only one, for simplicity.

While the path tortuosity metric has been proposed before [10], it has not been proven to provide information that is independent of the time-to-complete metric. Although it is intuitive to use such a metric, many tests show close correspondence between the two. For example, in Figure 6, we show the results of five separate trials for the WRIST system and a conventional joystick. (Time to complete is on the top while fractal dimension is on the bottom.) Both metrics trend downward as gap width increases, meaning that the quality of control increases (task time gets shorter and the path gets smoother) for each interface device as the course gets easier. This is the expected result and shows that path tortuosity is consistent with intuition and time-to-complete, but it does not demonstrate path tortuosity is actually providing new information.

3.1 Verification of the Uniqueness of Path Tortuosity

In an effort to validate the independence of the metric, we deliberately introduced random error into the control software of certain test subjects without the test subject's knowledge. This error manifested itself in the form of a random angular offset in the motion direction of the cursor commanded by the user. In effect, this error controlled the degree of inaccuracy with which the user could control the cursor for the devices tested.

The distribution of the angular error was Gaussian and we made no modification to the commanded magnitude. For our operator pool, we used over one hundred non-engineering undergraduate students in a laboratory class that introduces them to technology. The task was explained to them and they were given one practice trial with each of five different interface devices. The interface devices included the WRIST system, a joystick, a trackball, a mouse in position mode and a mouse in velocity mode. Only a small subset of the data is presented here.

In comparing these "perturbed" runs to normal runs across several users, we found that time-to-complete was not reliably correlated with the actual inaccuracy of the trial, which is assumed to be dictated by the introduced error. The path tortuosity, as computed through Fractal

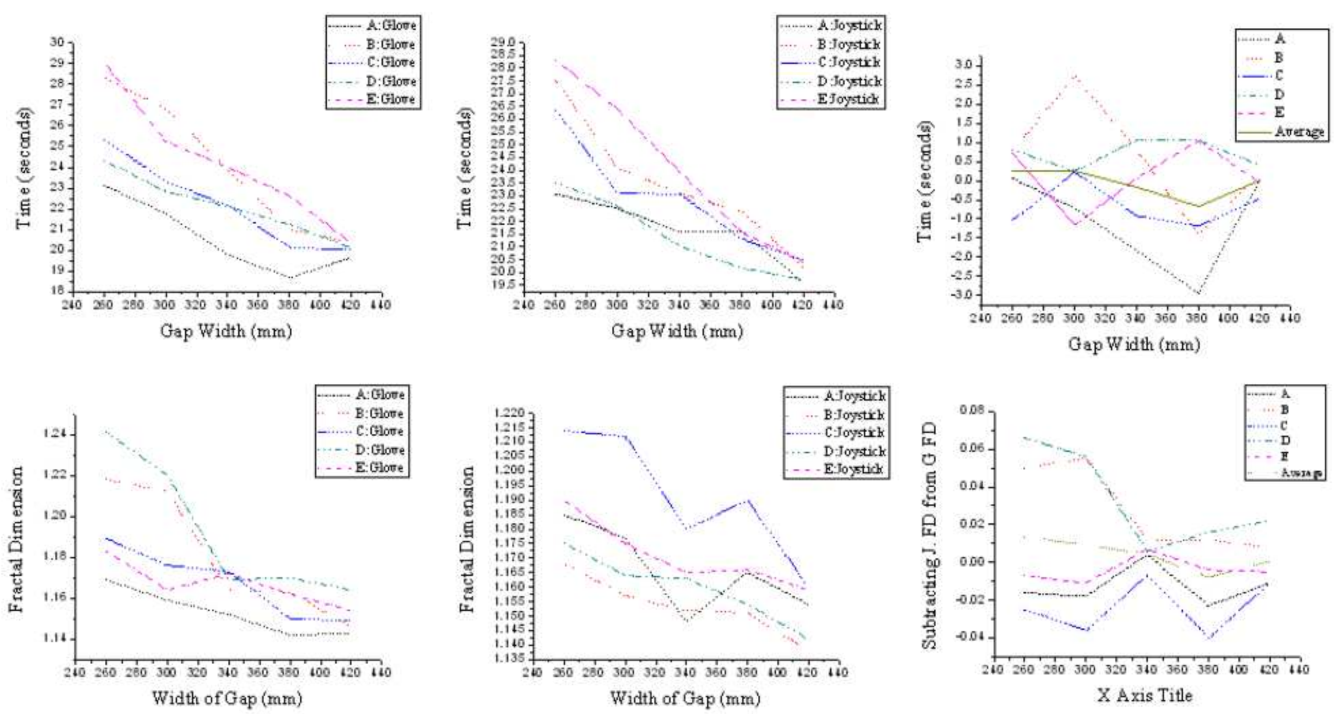

Figure 6: Top set of plots show the time-to-complete metric for five trials each of the WRIST and a conventional joystick as gate width increases. The bottom set of plots show path tortuosity for the same devices and gate tasks. The two metrics trend the same.

dimension, was a more reliable predictor of the introduced error.

The data are difficult to interpret as a pronounced learning effect is evident in the results (Figure 7). The subjects are not particularly tech savvy so many of the interface devices were new to them and they were only given one practice trial. As they moved through the five trials with each device, negative slopes are apparent as time-to-complete and path tortuosity both decrease with practice. Furthermore, the gate width was varied for many of the test subjects, which varies the difficulty of the task. This introduces a negative or positive slope depending on whether the width is increasing or decreasing.

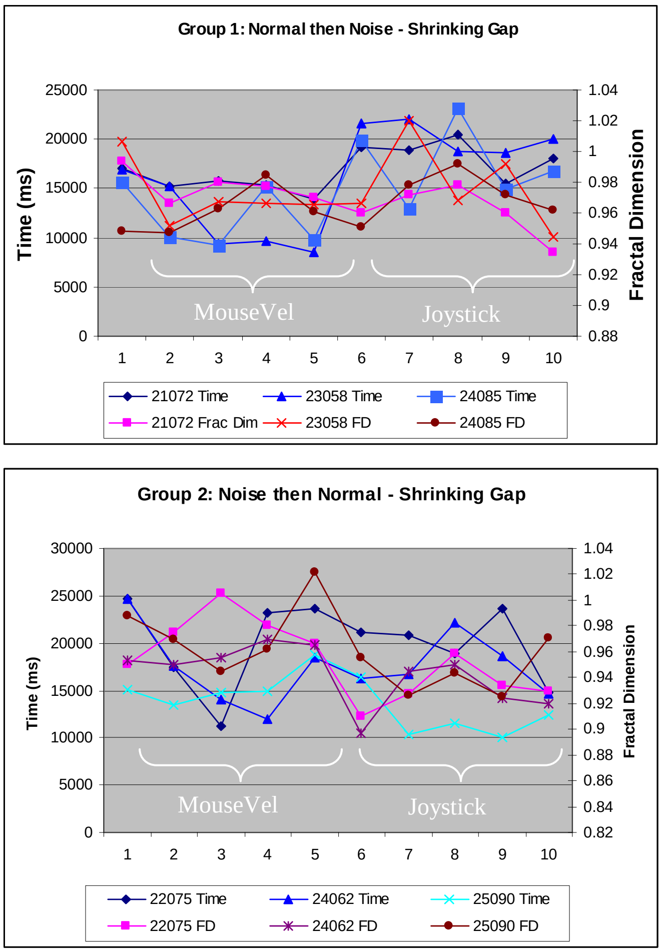

Figure 7: Group 1, top, and Group 2, bottom, indicating the trend over two devices and noise conditions for both time-to-complete metric and path tortuosity. Test subjects are denoted by number.

To counteract these effects, we look at complementary test suites that reverse the effects. For example, we took two groups of test subjects and had them use devices in the same order; in this case, mouse in velocity mode followed by the joystick. (Five trials with each device are completed.) Over the five trials, the gate width is gradually reduced, making the task more difficult and offsetting the learning effect. The complementary tests groups had noise injected at different times, however. By examining the slope and intercept of the trends, we can spot significant artifacts in the metrics.

Group 1 had noise introduced into the mouse tasks but not the joystick tasks while group 2 had noise introduced into the joystick tasks but not the mouse tasks. With the offsetting effects of learning and increasing difficulty, the slopes are near zero.

With the slope near zero, we can focus on the intercept to see which is higher or lower in relation to the device and noise. The intercepts are tabulated below and we see that path tortuosity is a more reliable predictor of the introduced noise than time-to-complete. (Erroneous cases are highlighted in Table 1.)

Table 1: Regression statistics of groups of user data.

Group	Device/Noise	Time Intercept	Fractal Dimension Intercept
Group 1	Mouse/No	17310	0.988
Group 1	Joystick/Yes	20150	0.979
Group 1	Mouse/No	18680	0.992
Group 1	Joystick/Yes	22220	1.000
Group 1	Mouse/No	13960	0.942
Group 1	Joystick/Yes	18790	0.966
Group 2	Mouse/Yes	18950	0.964
Group 2	Joystick/No	22830	0.919
Group 2	Mouse/Yes	22810	0.945
Group 2	Joystick/No	18070	0.919
Group 2	Mouse/Yes	12790	0.959
Group 2	Joystick/No	14620	0.938

To further demonstrate the independence of path tortuosity, we gathered another set consisting of two groups of data: a "Noisy" group of 6 test subjects with random error introduced and a "Normal" group of 22 test subjects. Since it is meaningless to directly compare the performance of individual test subjects, we exhaustively compared the average performance (both time-to-complete and path tortuosity) of every subgroup of four test subjects from each group. This resulted in 15 Noisy subgroups ($^{6}C_{4}$ – 6-choose-4) and 7315 Normal subgroups ($^{22}C_{4}$).

Comparing every Noisy subgroup to every other Normal subgroup involved over 100,000 comparisons. Of these comparisons, the fractal dimension of the noisy subgroup was greater than the fractal dimension of the normal subgroup 99.5% of the time while the time-to-completion of the noisy subgroup was greater than the time-to-completion of the normal subgroup 49.4% of the time (essentially chance).

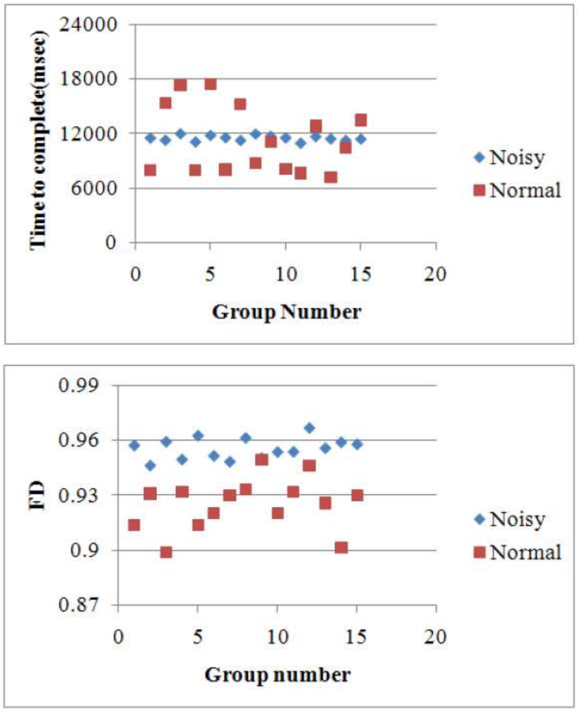

Figure 8: Random sampling of group comparisons with time-to-complete, top, and path tortuosity, bottom, for robot teleoperation tasks with and without noise introduced.

4. CONCLUSION

A large number of tests have been conducted on both technical and non-technical users to determine the usefulness of the WRIST system in human/robot interaction. Over 120 users were tested with various input devices, including the WRIST system, a mouse, a joystick, and a trackball on both simulated robots and physical robots. We gathered both quantitative results and qualitative results. Our quantitative user test results indicate users were not as proficient, in either time-to-complete or path tortuosity, in using the WRIST as compared to the joystick or mouse. This is not surprising as nearly all users had some familiarity with these devices. In fact, learning rates were extrapolated from the data and the WRIST, indeed, shows a much higher improvement rate over the trials. In qualitative evaluations from non-technical users who had little prior exposure to use of a joystick, users preferred the WRIST system for its more intuitive nature. Coupling this result with the portability of the device and other unmeasured benefits of its wire-free nature leads us to believe it will be a successful interface when fully developed.

The discrimination between the time-to-complete and path tortuosity metrics is highly meaningful. It seems intuitive that the two metrics would be distinct, but we believe this is the first result that demonstrates they are, in fact, distinct and separably valuable.

5. ACKNOWLEDGMENTS

Partial funding for this work was provided by the NSF Safety, Security and Rescue Research Center and by NSF through grant 0719306. Thanks to Robin Murphy and the NSF R4 program for the loan of the Inuktun robot.

6. REFERENCES

[1] R. Murphy, “Human-robot interaction in rescue robotics,” IEEE Systems, Man and Cybernetics, vol. 34, no. 2, May 2004.

[2] K. Newman, “The open interface: Beyond keyboards and mice,” e.nz Magazine, May 2002.

[3] G. I. of Technology, “Real-time cooperative behavior for tactical mobile robot teams,” Georgia Tech College of Computing and Georgia tech research Institute: Final Report, vol. A003, 2001.

[4] R. M. Voyles, A. C. Larson, M. Lapoint, and J. Bae, “Core-bored search and rescue applications for an agile limbed robot,” in Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, vol. 1, 2004, pp. 58–66.

[5] D. Sturman and D. Zeltzer, “A survey of glove-based input,” in IEEE Computer graphics and Applications, 1994.

[6] G. Kessler, L. Hodges, and N. Walker, “Evaluation of the cyberglove as a whole-hand input device,” in ACM Transactions on Computer-Human Interaction, vol. 2, no. 4, 1995, pp. 263–283.

[7] B. H. Thomas and W. Piekarski, “Glove based user interaction techniques for augmented reality in an outdoor environment,” Virtual Reality, vol. 6, pp. 167–180, 2002.

[8] B. Thomas and W. Piekarski, “Glove based user interaction techniques for augmented reality in an outdoor environment,” Virtual Reality, vol. 6, no. 3, pp. 167–180, 2002.

[9] R. Voyles and P. Khosla, “A multi-agent system for programming robots by human demonstration,” Integrated Computer-Aided Engineering, vol. 8, no. 1, pp. 59–67, 2001.

[10] M. Voshell, D. Woods, and F. Phillips, “Overcoming the keyhole in human-robot coordination: simulation and evaluation,” in Proceedings of the Human Factors and Ergonomics Society, pp. 26–30, 2005.

[11] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich, “Common metrics for human-robot interaction,” in Human-Robot Interaction Conference, March 2006.

[12] Mandelbrot, B., “How Long is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension,” Science, v. 156, n. 3775, pp. 636-638, 1967.

[13] J. Bae and R. Voyles, “Wearable joystick for gloves-on human/computer interaction,” in Proceedings of the SPIE Defense and Security Symposium, Orlando, FL, April 2006.

[14] M. Goldstein, D. Chincholle, and M. Backstrom, “Assessing two new wearable input paradigms: The finger-joint-gesture palm-keypad glove and the invisible phone clock.” Personal and Ubiquitous Computing, vol. 4, no. 2/3, 2000.

[15] J. Bae, A. Larson, R.M. Voyles, R. Godzdanker, J. Pearce, "Development and User Testing of the Gestural Joystick for Gloves-On Hazardous Environments," in Proc. of the IEEE Intl. Symp. on Robot & Human Interactive Communication, Jeju, Korea, Aug. 2007, pp. 1096-1101.