Flying a Manta with Gesture and Controller: An Exploration of Certain Interfaces in Human-Robot Interaction

Brian Ballsun-Stanton and Jon Schull

Information Technology Department, B. Thomas Golisano College of Computing and Information Sciences, Rochester Institute of Technology, 102 Lomb Memorial Drive, Rochester, New York, United States

{bpb9521, jis}@it.rit.edu

Abstract. We document and discuss a number of interface explorations conducted while building a cable-array robotic sculpture in the form of a fiberglass bio-mimetic flying manta ray. Our two primary interface devices were a gesture sensing Essential Reality P5 glove and a Microsoft Xbox 360 game controller. We also offer a conceptual design space comprising three axes: mapping, frame of reference, and interface location. This design space lets us discuss and compare all of our explorations and envision other potentially interesting interfaces.

Keywords: Data Glove, Essential Reality P5 Glove, Game Controllers, Gesture, Gesture Control, Human-Robot Interaction, Information Visualization, Interface Mappings, Interfaces, Microsoft Xbox 360, Robotics, Teleobservation

1 Introduction

Over the past year, students and faculty at the Rochester Institute of Technology collaborated on a cable-array robotic[1] kinetic sculpture in the form of a manta ray (Fig. 1) that maneuvered through the B. Thomas Golisano College of Computing and Information Sciences' three-story atrium. By lengthening and shortening computer-controlled suspension lines, we maneuvered the manta, simulating swimming through the air on its own accord.

The manta installation was intended to express the idea that modern life is increasingly embedded in an unseen ocean of information [7], controlled by barely perceptible computer-controlled constraints. We sought to capture this vision via an incongruously graceful manta ray flying through the three-story college atrium under the control of nearly transparent computer-controlled fishing lines. Projected on to the wings of the manta would be web pages and other digital information sniffed [2] from

---

[1] For more information on cable array robots, see [6], [15]

local unsecured wireless internet traffic, thus revealing the omnipresent but invisible communications that suffuse modern life and expressing the idea that we are all swimming through an invisible sea of cyberspace.

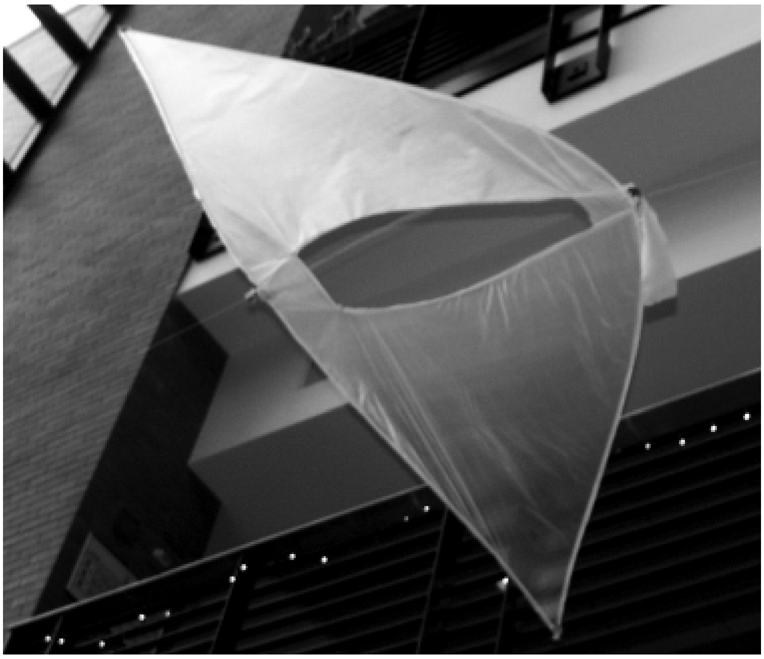


Fig. 1. A view of the kite-like Manta ray suspended from computer-controlled fishing lines in the three-story atrium of Rochester Institute of Technology's B. Thomas Golisano College of Computing and Information Sciences. The manta is comprised of fiberglass rods with a translucent skin upon which we can project digital images. Because the rods are joined in a state of dynamic tension, they flex and bend when the lines are moved, producing a lifelike swimming motion.

This paper will discuss a number of human-robotic interface issues [3], [14] that arose during the design, construction, and programming of the system. In our designs, we explored gesture sensing [5], [9] and joystick driven interfaces. These explorations allowed us to formulate a potentially useful design space for the investigation of HRI issues comprised of three conceptual axes: Motor Mapping, Frame of Reference, and Location. This paper will use the design space to discuss and compare our interface schemes and to illuminate some potentially interesting issues and new interfaces for future consideration.

1.1 Three Human Robotic Interface (HRI) Conceptual Axes

The Motor Mapping axis of our conceptual framework (Fig. 2.) has two poles: natural (human-intuitive) mapping and mechanical (machine-informed) mapping. Consider a hand puppet. When wearing one, flexing your thumb and pinky fingers move its arms, while the other three fingers control the mouth and head: this is a natural mapping [12]. A puppeteer with multiple marionettes who uses difficult, arcane, and precise gestures, constrained by the kinematics of dangling strings, to control the marionettes, exemplifies the other pole of the Motor Mapping axis. This axis thus captures the range of cognitive effort required to translate the operator's intentions into machine-executable instructions.

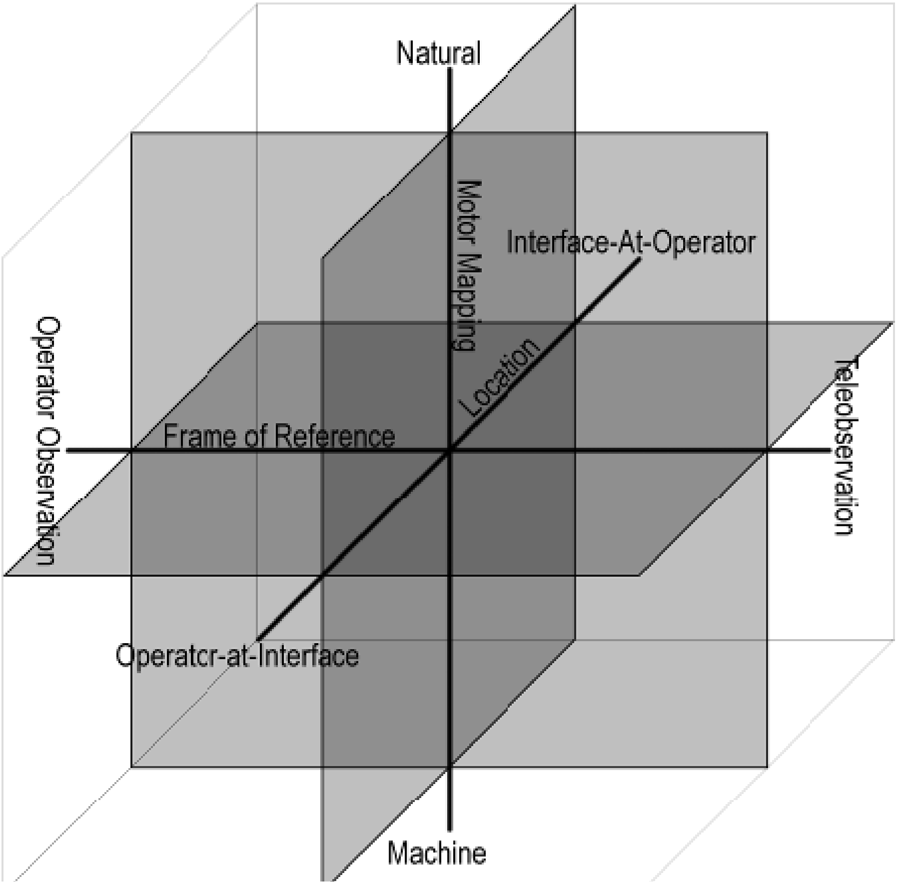


Fig. 2. This is a three-dimensional representation of our three-axis HRI design space. Each axis varies between the two specified poles and is orthogonal to the others. The first axis, Motor Mapping, varies between two poles: natural and mechanical. The second axis, Frame of Reference, varies between operator observation and teleobservation. The third axis, Location, varies between operator-at-interface and interface-at-operator.

The two poles of the Frame of Reference axis (Fig. 2.) are operator observation and teleobservation. An example of an operator-observed frame of reference is a radio-controlled plane whose pilot simply controls and observes the plane from the ground. Conversely, a teleobserved frame of reference is a pilot who adopts the plane's point of view with the aid from a wireless camera mounted in the plane's cockpit.[1] There is a pronounced difference in frames of reference when the airplane is flying toward the pilot. The teleobserving pilot perceives a right turn of the airplane as a turn to the pilot's right. Conversely, the operator-observing pilot perceives a right turn of the airplane as a turn to the pilot's left.

The third orthogonal axis of our HRI design space, Location, (Fig. 2.) spans from operator-at-interface, where an operator must go to the interface to operate the device, to interface-at-operator where the operator's physical location is relatively unconstrained. A traditional airplane is operator-at-interface: the pilot must be in the cockpit. In contrast, a motorized web-camera that can be pointed and controlled by anyone from anywhere on the Internet is fully interface-at-operator.

These three orthogonal axes thus map out the three-dimensional space of Fig 1 within which a variety of interface control schemes may be situated. We will suggest below that there are some regions within this space are considerably more user-friendly than others, and that there are sweet spots where the operator's actions take on the character of nuanced gesture rather than gross manipulation, and in which the device becomes an extension of the actor rather than merely the object of his actions.[2]

2 The Manta and our Explorations

As we sought to reveal the invisible sea of cyberspace [17] through the manta, some other themes emerged regarding the role of the interaction designers [11] and the nature of the designers' creative processes. The mission of the interaction designers in this project was to ensure that we experienced the emerging world of pervasive, ubiquitous [1], and ambient computing as a natural habitat where we could express our intentions and achieve our goals with minimal effort and maximal efficacy. Furthermore, in this project, the interaction designers' creative process was exploratory movement through a multidisciplinary multidimensional design space in search of locations real and virtual where man and machine could comfortably coexist. However, while this search was the ideal to which we aspired, the search itself was not without its failures: each teaching an important lesson.

---

[1] This is not to be confused with first-person vs third-person perception as discussed in the the computer gaming literature [4], [13]. From our point of view, both first and third person perception in games is teleobservation because the computer-mediated action is happening remotely and is transmitted via the monitor.

[2] The philosopher Martin Heidegger [10] has created the concepts of present-at-hand and ready-to-hand. These concepts explore the distinction between a tool becoming an extension of our mind versus the direct object of our actions.

2.1 Interface Explorations with a P5 Glove and Counterweighted Cable Array

Our first cable array robot's infrastructure was a counterweight based four line system in which four computer-controlled motors and pulleys reeled fishing lines in and out to alter the robot's position under control of a P5 data glove. Within this arrangement, we will discuss two interfacing schemes used to control the sculpture and the projections. First, we will explore a system intended to control the digital imagery projected on to the manta's wings. Then we will discuss our first cable-array control scheme.

Interface Scheme One: from Hand Position to Page Position

By extracting data from the ambient Wi-Fi network, we were able to create a large dynamic virtual mosaic in the form of a 9 x 9 matrix of web pages. By continuously projecting a small part of that mosaic on to the wings of the flying manta, we sought to create the illusion that the manta was flying through a continuous sea of web pages, revealing them briefly as it passed. We intended our first interfacing scheme to allow us to control the drift of the sea of information independently of the movement of the manta itself. To this end, the P5 glove's position was sampled every five milliseconds and used to pan the oversize web page as the operator moved their hand, e.g., so that the page would scroll right when the operator's hand moved right.

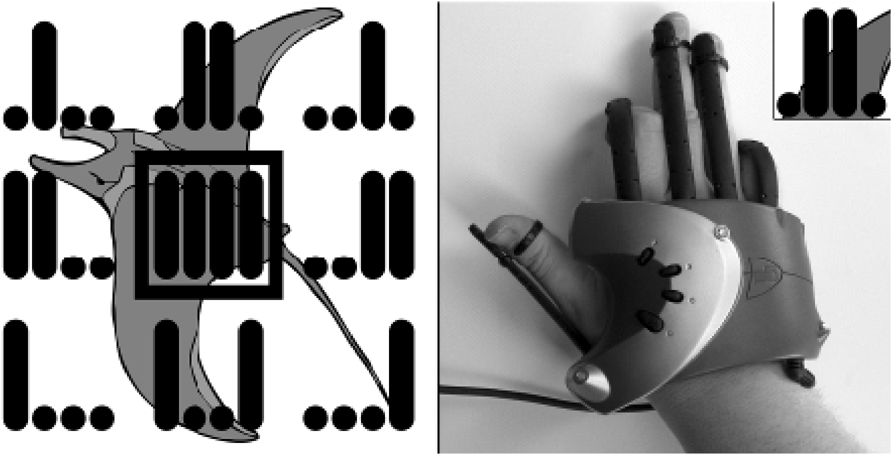
A natural mapping of hand-position to page-position made great sense in theory (c.f. Norman), but in practice, it proved unworkable. The P5's sensor platform did not reliably detect the LEDs embedded in the P5 glove, and this caused rapid and disruptive shifts in the visual display. The P5 handled the failure in detection poorly: it guessed. These guesses used the visible LEDs without any prior knowledge, resulting in frequent cross-screen jumps of the viewport without any user input. Less drastic, but we observed more frequent shifts of over one hundred units. These shifts also caused corresponding jumps in the viewport. We tried to compensate for this jitter error by using error correction algorithms, including Richard Hachem's[8] implementation of a Kalman filter[16] for P5 Glove location data, but these filters introduced unacceptable time delays without any sufficient decrease in jitter error.

A second source of error was far more insidious. Again, due to the poor sensing of the LEDs on the glove, movement away from the center of the sensor bar caused unpredictable skewing effects, which in turn caused intuitive movements to produce non-intuitive results. For example, the sensor platform sometimes interpreted a movement up as a movement away from the sensor platform, and the correlation between hand motions and page motions varied with distance from the sensor tower. Under time pressure, we abandoned the search for a combined solution.

In terms of the design space illustrated in Fig. 2, this interface scheme fits in the forward upper right quadrant. The Mapping was natural (albeit error prone), the Frame of Reference was teleobserved, and the Location (painfully) operator at interface: the system failed if we moved more than a few feet from the sensor tower. This interface had a very natural mapping: the display echoed movement of the glove. Our location-sensing interface was fully teleobserved because all the action was happening in an electronic space. In lieu of perceiving the electronic space directly, we could only perceive the local copy rendered on our monitor. However, despite this pure teleobservation, our location was fully at operator-at-interface since the page required a specific P5 Glove and computer running a copy of our custom software.

Interface Scheme Two: from Finger Bends to Motor Vectors

In the second interface scheme, we chose to forgo natural mapping in favor of a relatively easy-to-implement machine-oriented linkage based on bend sensors built into the fingers of the P5 glove. As the operator curled and extended their fingers the interface software sent commands to the motors to reel out, stop, or reel in. By combining these finger manipulations into a gesture, the operator could fly the manta, after a fashion. As seen in Figs. 3a and 3b, very specific gestures were required to approximate movement in one of the eight directions by commanding each motor to change into one of three states: reel out, reel in, or stop. This movement, combined with the rapid vertical movements required by the system, created the illusion of flight, complete with flapping wings.



Figs. 3a and 3b. Operators used the stylized instruction sheet (*Left*) as a guide to the manta's flight commands while using the finger-sensing interface. The eight patterns around the edge indicate in which lateral direction to move relative to the operator's position. The dots indicate a bent finger while the lines indicate an extended finger. For example, the P5 Glove (Right) is gesturing for the robot to move laterally away from the operator. Inset is the gesture the operator is trying to replicate.

The problems encountered with this scheme stemmed from the non-intuitive mapping as well as from electro-mechanical issues: because of idiosyncrasies of each motor and winch, the motor speed induced by each full forward and full backward position was different for each motor. With the fingers in a fixed configuration, the manta would typically drift away from the desired direction of travel. Since this scheme did not include variable speed control, corrective actions by the operator produced saw tooth motions in which the manta lurched dramatically but drunkenly through a three-story atrium above the heads of onlookers.

Ironically, the visual effect of this performance was considerably more graceful than this description might lead one to expect. Partly by design, partly through trial and error, the manta itself was composed of fiberglass rods, which flexed and bent gracefully in response to jerky changes in motor speed, producing a surprisingly good simulation of a manta flapping its way through an invisible medium. Equally clear, was our pattern of design evolution. Our jerky but effective pattern of design evolution became clear here too: somewhat like the manta, we were lurching from one corner of our design space to another, in an ongoing attempt to find a combination of technologies an intuitive control scheme for the operator and a corresponding natural pattern of movement in the manta.

This interface occupies a location in the front bottom left quadrant of the design space shown in Fig. 2. Our finger-sensing interface was machine mapped. Despite the possibility of a natural mapping from our first scheme, we chose machine mapping for a quicker and less error prone implementation. In general, during development, one the fastest routes to mechanical functionality is by conforming to machine-imposed constraints. Furthermore, our physical package was prone to frequent failure due to tangled fishing lines and had no remote control functionality. This tendency towards failure required a constant presence at the package, rendering moot the lack of remote control. The debugging requirements make it fully operator observed and operator-at-interface.

2.2 Interface Explorations with an Xbox 360 Controller and Winched Cable Array

While audiences' appreciated our early performances, our first physical arrangement of motors and lines had no remote control functionality, suffered from frequently tangled lines, and required constant attention. For this reason, and because the P5 sensors required that we stay within a few feet of the sensor platform, it was very operator-at-interface. We therefore decided to redesign most aspects of the system.

Fig. 4. A picture of the Microsoft Xbox 360 controller driving the manta diagonally upwards to the operator's left.

We replaced the four counterweighted motors with three more powerful motors designed to winch the lines onto internal spools. As further protection, we enclosed these motors in an enclosure intended to reduce the probability and viciousness of entanglements. We replaced the P5 Glove with a Microsoft Xbox 360 controller (Fig. 4.) chosen for its easy computer connectivity. We also developed a client/server architecture for our software that increased flexibility and modularity of design.

With this setup, we experimented with two more interfacing schemes. In one scheme, we adopted a machine-mapped interface metaphor similar to the one we had previously used: three axes of joystick movement (up-down on one joystick, and up-down and left-right on the other) mapped to each of our three motors. In the last scheme, the mapping was considerably more natural for our operators as forward and backward movements of the left joystick controlled the manta's altitude whereas lateral movements of the right joystick made the manta move along the horizontal plane. In both instances, we also took advantage of the buttons on the controller to control other aspects of the software.

Interface Scheme Three: from the P5 Glove to an Xbox 360 Controller

By replacing finger-bends with joystick movements and implementing variable motor speed controls, our interface scheme afforded much more nuanced control of the manta's speeds. On the other hand, the arrangement still required that the operator continuously translate desired changes in the manta's location to desired changes in line length, and translate changes in direction to changes winch speed. Due to the shift from four lines to three, this control scheme was virtually unusable. While the interface was more ergonomic, the mappings now required the operator to input fractional speeds: something almost impossible to do on a real-time basis.

Much more successful was our adoption of the client/server architecture and use of the more portable Xbox 360 controller. Now, the operator could move around, connected wirelessly to the server controlling the motors. This allowed the operator to view the manta and the scene from various perspectives. While true teleobservation escaped us (as the control scheme could not orient itself to the manta's frame of reference), we were no longer tied to the motors as in previous schemes.

This interface scheme can be located firmly at the center of the bottom four quadrants. While this was fully machine mapped, as discussed above, the increased freedom of the client/server architecture allowed us to divorce ourselves from being completely operator observed and operator-at-interface. However, the lack of a control scheme which would adapt to the manta's frame of reference and the infrastructure tired to the Xbox 360 controller (namely that the operator had to lug a Wi-Fi enabled laptop around) prevented us from moving firmly to the other poles. At heart, while these changes reduced the amount of control over the manta, they increased the operator's sense of control. This sense of control, a feeling that we were steering the manta instead of just manipulating it, was due to the operator's ability to modulate speeds, increased freedom from the motor platform, and the ability to stop the manta reliably by simply releasing the joysticks.

Interface Scheme Four: from Motor Control to Vector Control

Our last interface scheme allowed us to move from machine-mapped thumb movements to naturally mapped manipulations: thumb movements that translated intuitively into manta movements, and corresponded well with the operator's sensible perception of manta movements. By mapping the three axes of joystick movement in a certain game controller orientation (Fig 4) to the three spatial axes of the rectangular atrium and by mapping degree of joystick deflection to speed of the manta, it became possible to use thumb movements that made sense to move the manta naturally. Of course, doing so required that the computer perform the trigonometric computations to convert these gestures into changing rates of winch rotation for the three motors. That however, is precisely what gesture recognition systems and human-oriented interfaces must do: allow human operators to express intentions with intuitive movements, and to get intuitive feedback from those movements.

Indeed, while this last scheme was far from perfect, we found that we could overcome a variety of sins by the combination of intuitive movement and intuitively perceived feedback. For example, while we are still coping with mechanical idiosyncrasies as well as with variations in friction from one line to another and from one moment to the next, compensating for these variations is now relatively easy and unconscious on the part of the operator. Thus, much as an automobile driver pushes down or pulls back on an automobile accelerator to compensate for fluctuating headwinds, the manta operator simply pushes the appropriate joystick more or less forcefully in the intended direction to compensate for mechanical and other fluctuations in the system he is controlling.

This interface achieved a major distinction from the prior interface by programming the computer to translate naturally expressed human intentions to machine-executable motor commands so that the computer allowed mostly intuitive actions to express the human's intent: "I want it to go over there". Therefore, while the prior scheme was in the middle of the four quadrants (machine mapped, neutral frame of reference and neutral location); this scheme is in the middle of the top four quadrants, due to the increased human mapping. This scheme is neutral on the frame of reference axis because it provided provision for remote control. With only a few software modifications, this scheme could achieve a full teleobservation mode. With the change to a wireless Xbox 360 controller, this scheme would also firmly be interface-at-operator.

3 Discussion and Future Research

Several general observations seem pertinent. In discussions of this nature, it is not always obvious what is intuitive and what is not, and there may well be additional orthogonal dimensions to the three we have identified. With this in mind, we offer our framework as a useful way of exploring and discussing alternative schemes, not as a comprehensive taxonomy.

Furthermore, we suspect that some quadrants in the design space are more promising when used in certain types of projects. In performative robotics, our own experience leads us to speculate that as one approaches the natural and interface-at-operator poles, the operator will feel more like a performer than an operator, and the robotic system will feel more like an extension of one's body. We can report, for example, that during one deployment a Digital Performance Conference dance party[1] [13], the operator had occasion to move the manta down floor level, and make it interact with human dancers. Although the operator was twenty feet way from the dance floor, the experience of interacting with the other dancers was more salient than the experience of interacting with the machinery or the manta.

We should also note that we are still far from realizing our ideal visual or performative vision. Thumb twitching while tethered to a computer is likely to be, and to feel, less expressive of the performer's intentions than arm waving and body swaying while roaming freely. We intend to explore this question using a wireless Nintendo Wii Remote, or a gyration air mouse, both of which promise a natural, interface-at-operator arrangement that should be as reliable (within receiver range) as the Microsoft Xbox 360 controller was.

Finally, we would like to defend explicitly the proposition that a manta ray dancing at a disco (and other socio-technical scenarios not usually encountered in engineering labs) provides a remarkably good opportunity for exploration of gestural human-computer interaction and for discovery and exploration of novel engineering challenges. For hundreds of millions of years, primates and mammals have been manipulating each other, from birth, via gesture-based social communication. They have been manipulating relatively unintelligent beasts of burden through gesture for tens of millions of years. This is a time-honored working model with much to recommend it.

The emergence of non-biological robots and ambient, pervasive computing infrastructures distinguish this common era. These emergent technologies are increasingly going to be a significant part of the human environment. Engineers, interaction designers, and their clients will have more success (and fun) if they try doggedly to move away from machine-constrained and human-constraining control schemes for normal interaction methodologies and attempt instead to emulate, simulate, study, and situate their work in what people do when they are doing what comes naturally. If we can get a robotic manta to dance, and to be welcome at the party, we will have created something that is wondrous, strange, but which promotes trouble-free and transparent interaction.

Acknowledgements

We would like to thank Dianne Bills, Steven Jacobs, and the Information Technology Department's Innovation and Information class. Furthermore, we would like to thank the Golisano College of Computing and Information Sciences as well as the First International Conference on Digital Live Art for allowing us to fly our manta in their spaces. Finally, we would like to thank the Department of Planning and Natural Resources, Division of Fish and Wildlife, of the Virgin Islands for their line drawing of a Manta Ray, used here with permission.

---

[1] For a number of videos showing the manta at various locations, go to http://vega.it.rit.edu/~bpb9521/

References

1. Abowd, G. D., & Mynatt, E. D.: Charting Past, Present, and Future Research in Ubiquitous Computing. ACM Trans. Comput. -Hum. Interact., 7 (2000) 29-58
2. Ansari, S., Rajeev, S. G., Chandrashekar, H. S.: Packet Sniffing: A Brief Introduction. Potentials, IEEE, 21 (2002) 17-19; 17- 19
3. Benko, H., Ishak, E. W., Feiner, S.: Cross-Dimensional Gestural Interaction Techniques for Hybrid Imrnersive Environments. vr, 00 (2005) 209
4. Bruce Thomas, Ben Close, John Donoghue et al.: First Person Indoor/Outdoor Augmented Reality Application: ARQuake. Personal and Ubiquitous Computing, V6 (2002) 75-86
5. Chalhoub, N. G.: Control of a Leadscrew Driven Flexible Robot Arm. (1986)
6. German, J. J., Jablokow, K. W., Cannon, D. J.: The Cable Array Robot: Theory and Experiment. 3 (2001) 2804-2810 vol.3; 2804- 2810 vol.3
7. Gershon, N., Eick, S. G., Card, S.: Information Visualization. IEEE Comput. Graphics Appl., 17 29-31
8. Hachem, R.: Kalman Filtering for P5.
9. Hasanuzzaman, M., Zhang, T., Ampornaramveth, V. et al.: Gesture-Based Human-Robot Interaction using a Knowledge-Based Software Platform. Industrial Robot, 33 (2006) 37-37-49
10. Heidegger, M.: Being and time. Harper & Row, (1962)
11. Jan O. Borchers ER -: A Pattern Approach to Interaction Design. AI & Society, V15 (2001) 359-376
12. Norman, D.: The design of everyday things. Doubleday, (1990)
13. Ryan, M.: Beyond Myth and Metaphor - the Case of Narrative in Digital Media. Game Studies: The international Journal of Computer Game Research, 1 (2001)
14. Stefan Waldherr, Roseli Romero, Sebastian Thrun: A Gesture Based Interface for Human-Robot Interaction. Autonomous Robots, V9 (2000) 151-173
15. Usher, K., Winstanley, G., Carnie, R.: Air Vehicle Simulator: An Application for a Cable Array Robot. (2005) 2241-2246; 2241- 2246
16. Welch, G., & Bishop, G.: The Kalman Filter.
17. Wertheim, M.: The pearly gates of cyberspace: A history of space from dante to the internet. W. W. Norton \& Co., Inc, New York, NY, USA (1999)