GENG, Weidong
STRAUSS, Wolfgang
FLEISCHMANN, Monika
ELISTRATOV, Vladimir
KULESSA, Thomas
KOLESNIK, Marina

Perceptual User Interface in Virtual Shopping Environment

Published on netzspannung.org:
http://netzspannung.org/about/mars/projects
19 October 2004

First published: International Journal of Image and Graphics No. 2, Vol.3 (2003), p. 365 -378.

Fraunhofer Institut Medienkommunikation

The Exploratory Media Lab
MARS Media Arts & Research Studies

International Journal of Image and Graphics

PERCEPTUAL USER INTERFACE IN VIRTUAL SHOPPING ENVIRONMENT

WEIDONG GENG*, WOLFGANG STRAUSS
MONIKA FLEISCHMANN, VLADIMIR ELISTRATOV
THOMAS KULESSA, MARINA KOLESNIK

MARS-Exploratory Media Lab
Fraunhofer Institute for Media Communication (IMK)
D-53754, Sankt Augustin, Germany
E-mail: {geng,strauss,Fleischmann,elistratov, kulessa,kolesnik}@imk.fhg.de

Received (February 22nd, 2002)
Revised (August 15th, 2002)

In this paper we presents our effort towards the goal of perceptual user interface for major interaction tasks, such as navigation/travel, selection/picking and personal data access, in virtual shopping. A set of 3-D navigation devices, vision-based pointing and personal access system are mainly discussed. The motivation and design principles behind these interfaces are also described. A prototype integration solution, which brings these devices together in virtual shopping environment, is given. These interfaces and interaction devices have been implemented and tested for evaluation

Keywords: Perceptual User Interface; Human-computer Interface; Virtual Reality; E-Commerce

1. Introduction

Human-computer interaction has not changed fundamentally for nearly twenty years. Graphical user interface gives users direct control and predictability. The WIMP (Windows, icons, menus, pointer) paradigm has served to provide a stable and global interface to computing. These properties provide the user a clear model of what commands and action are possible and what their affects will be. They allow users to have a sense of accomplishment and responsibility about their interaction with computer application.

As computers are becoming ubiquitous, the critical bottleneck in their use is often the user interface rather than the computing power of the system. It is clear this paradigm will not scale to match the myriad of factors and uses of computers in the future. Interaction with Computing devices is becoming more and more pervasive in our daily life. Conventional GUI techniques appear to be ill-suited for some of the kinds of interactive platform now starting to emerge, with computing devices having tiny and

* Currently in Multimedia Innovation Center, The Hong Kong Polytechnic University, Hong Kong

large displays, virtual environment, mixed reality and public media space. It is assumed that physical interaction between humans and computation will be less like the current desktop keyboard /mouse /display paradigm and more like the way humans interact with the physical world. That is, what we need are interaction techniques well matched with how people will use computers.

Therefore, perceptual user interfaces (PUI) are recently proposed and investigated. [1, 2] Its essence is grounded in how people interact with each other and with the real world. These PUIs are characterized by interaction techniques that combine an understanding of natural human capabilities (particularly communication, motor, cognitive, and perceptual skills) with computer I/O devices and machine perception and reasoning.[2]

PUIs are desired to bring our human capabilities to bear on creating more natural and intuitive interfaces. They seek to make the user interface more intuitive and compelling by taking advantages of the ways in which people naturally interact with each other and with the world—both verbally and non-verbally. Devices and sensors should be transparent and passive if possible, and machine should perceive relevant human communication channels as well as generate output that is naturally understood. This is expected to require integration at multi-levels of technologies such as speech and sound recognition and generation, computer vision, graphical animation and visualization, language understanding, touch based sensing and feedback (haptics), learning, and modeling and dialogue management. A summary of interaction modalities between human and computer is shown in Figure1.

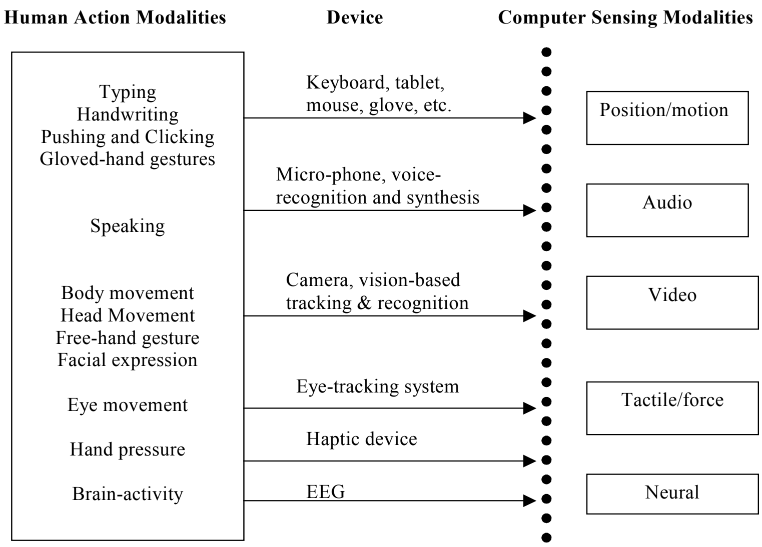

Fig. 1 Modalities Mapping between human action and computer sensing (Adapted from Sharma et al. [3])

In MARS lab, we developed a set of innovative navigation and pointing devices, which support intuitive user interaction for virtual shopping environment, e.g. shopping windows, large showcases or info-communication portals in public space. They will provide an infrastructure for different levels of navigation and interaction in multi-user e-commerce applications situated in public space.

In the following section, we will present several navigation devices developed by MARS lab. Vision-based pointing system and personal access system will be discussed in Section 3, and a prototype integration interface for 3D e-commerce environment is described in Section 4. Finally, some testing results and a brief discussion of future work is given.

2. Intuitive 3-D Navigation

We explored and developed several physical non-standard interfaces, i.e. Treadmill, Virtual Balance and Cyberwheel, for public spaces with large-scale displays. The underlying ideas behind them are to employ the natural human movements, such as walking posture, body motion and hand actions, to drive the navigation in 3D spatial environment.

Basic design parameters for navigation interface are human movements and natural modalities, so that a user does not need to learn unknown equipment and unknown principles of interaction. Intuition is given by using well-known principles familiar to the normal end-users.

2. 1. *Walking on treadmill*

Fig. 2 Virtual walk through Paris source/MARS/coutesy of Canal+

How large is virtual space? Using a treadmill makes the user really walk and keep on running. [4] Its navigation speed relies on the status of rolling cylinders driven by the walking movement of user, and navigation directions are controlled by buttons configured on bar-sticks in front of the user. We found out that users get tired after tens of minutes. It needs additional controls for turning direction, so that it was only partial real walking experience. Some researchers proposed complicated and expensive omni-directional treadmills. [5, 6] In general, a navigation system based on treadmill fits best as a spatial navigator in fitness rooms or in emergency-rescue training, where users are forced to actually walk or run. In fact the treadmill had been in use as test-training device for Berlin fireman and virtual facility management.

2. 2. *Flying with Virtual Balance*

Virtual Balance is a platform reacting on the body movement of a user while standing on it. [7] It is designed for navigation in large virtual environment. This platform consists of weight sensor discs (shown in Figure 3). The sensors receives changes of weight distribution on the disc and transmit them to an analysis unit, which in turn controls the position and orientation of user's viewpoint in the virtual environment. Minimal weight shift on the platform enables navigation. Stepping forward is moving down and walk, leaning backward is going up and flying.

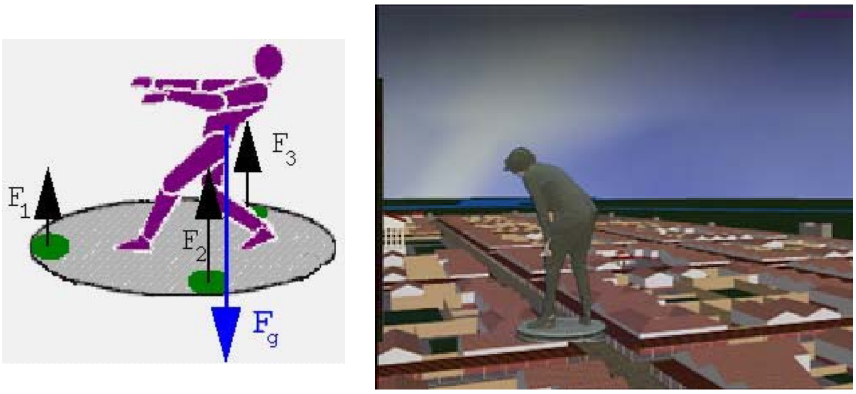

Fig. 3 Physical Principles and navigation of Virtual Balance

One could imagine Virtual Balance as a surfing board or magic carpet from eastern fairy tales. Our test showed that it takes some time to get used to the device. Young people generally mastered it faster than old ones. The major limitation of Virtual Balance is its lacking of full motion control. It is ought to use additional input channel to control the speed of virtual motion.

2. 3. *Traveling via Cyberwheel*

A series of experiments and testing lead us to the idea of inventing a 5 degrees of freedom navigation device, called Cyberwheel, whose major objective in functionality is to replace the navigation panel in VRML browser for web-based virtual environment. Its design and implementation are shown in Figure 4.

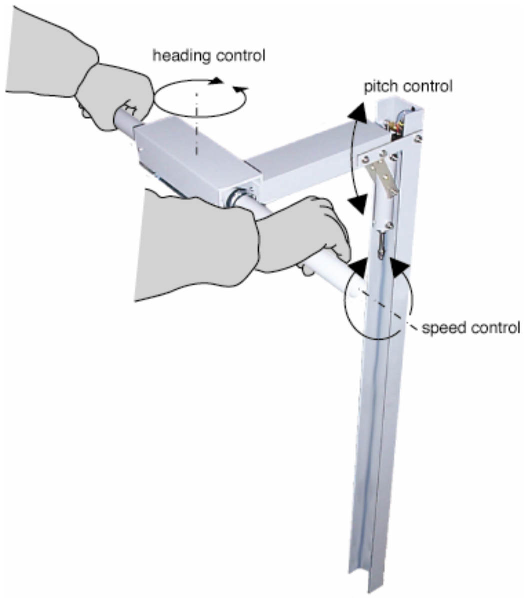

Fig. 4 Illustration of Cyberwheel

Heading control is performed like on a bike or motorcycle. It is convenient and known to everyone in principle. Rotation of the handles controls the speed of virtual motion. If the handles are released they are pushed back to the original position with zero speed. In addition, such a kind of control is simple and requires no efforts. If required, it is possible to control the pitch angle raising and lowering the upper part of the device. If no more pressure are exerted to change the pitch angle, the pitch control component will return to its original position. Compression and torsion spring are used for force-feedback effect. Rotations of 3 controls change a resistance in rotary potentiometers. The resistances are measured by micro-controller and sent to a computer via RS-232 cable.

3. Human-centered Non-obtrusive Interaction

3.1. *Vision-based pointing*

We investigated pointing-based interaction by capturing and tracking the arm-movements via computer vision techniques. The role of vision system is to keep an eye on user's actions, detect a moment when the gesture of pointing has occurred, interpret its direction and follow the pointing hand in real time. The system operates in real time, and runs on a stand-alone computer.

As shown in Figure 5 virtual scene is back-projected on a wall of approximately 3 by 2 meters, and the user moves freely in a real-world space of approximately 4 by 3 meters. A monochrome infrared-sensitive CCD camera at the ceiling of the space captures the top image of the user with the frequency rate of 25 frames per second. The assumption about the overall lighting is that it has to stay constant during the operation, since color calibration is done only once during the system initialization.

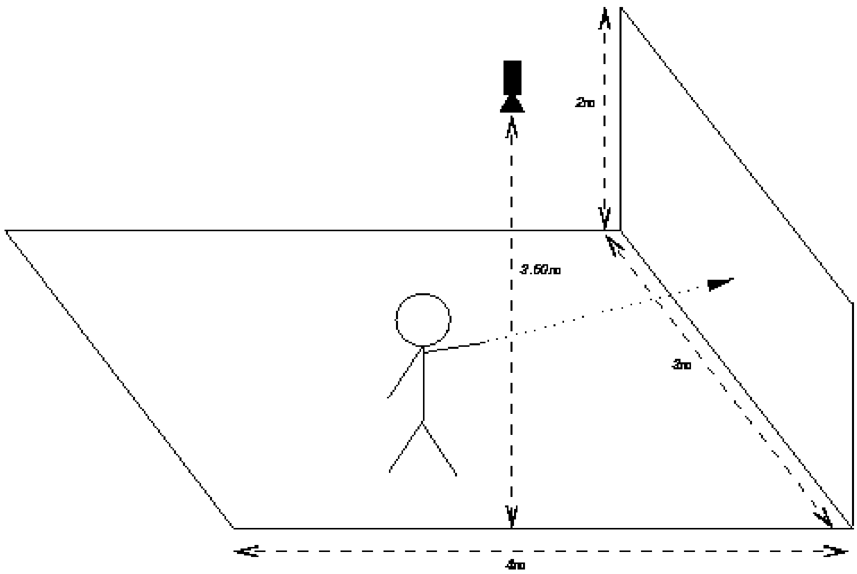

Fig. 5 Vision-based pointing

The pointing gesture occurs when the user spreads one of his arms towards the screen. The pointing direction is restricted to one hemisphere in front of the user, and it is parameterized relative to the projective screen. It is modeled by a 3-D line that connects a point between the user's eyes with the tip of the pointing hand/finger. 3-D line models the pointing gesture. The recognition of pointing position is then equivalent to finding the horizontal and the vertical coordinates of the point where this line intersects with the screen.

3.1.1. *Pointing movement tracking*

Our pointing gesture consists of hand/arm movements which are tracked by a modified Hierarchical Feature Vector Matching (HFVM) algorithm. [8] The tracking operations are performed for the two objects: the tip of the pointing hand/finger and the center of the head. Statistical data characterizing the hand object are computed in the neighborhood of the hand tip, only for those pixels that belong to the extracted hand segment. These data are collected into the hand feature vector for the reference image and compared to the hand feature vector of the search frame. The search area in the search frame is estimated using a prediction assumption about the likely hand location as compared to its current position in the reference frame. The feature vector is then computed for each pixel in the search area of the search frame. Here we exploit a closed-world assumption i.e. we do not expect new objects other than the pointing hand to appear in the search area. The HFVM matching process is applied to compare the reference hand feature vector against each feature vector of the search frame. Similar computations are carried out for the head object, which is defined within a neighborhood of the center pixel of the head segment.

3.1. 2. *Calculation of pointing position*

The 3-D interpretation of the pointing gesture is based on the internal camera parameters and the room geometry. Evidently, it is more economical to employ off-line calculation to compute those values that do not change over time, accordingly reduces the number of real time operations.

Camera configuration is defined by 6 parameters: 3 for rotation and 3 for translation. Camera translation is defined relative to the known room geometry. First we measure the altitude of the camera above the floor. Then we calculate the distance between the screen and some points marked on the floor. The direction perpendicular to the screen is marked on the floor as a line segment. This later is acquired by the camera to identify in the image plane the direction perpendicular to the screen.

To determine the screen coordinates of the point selected by the user, a two-step algorithm is applied. First, we calculate the pointing line in 2-D. In the second step we compute both the perspective projection of the pointing line onto the room floor, and accordingly gets its intersection with the screen. The implementation of stereo-vision system for tracking of pointed (x, y) coordinates is on-going.

3.2. *Wireless Access of Personal Information*

The design principle behind personal wireless access system (shown in Figure 6) is to avoid obtrusive typing of various personal data like name, address, personal preferences, and enable user's instant transmission of personal data. Ideally, once the user enters the application region, the relevant personal profile will be automatically sent to the application from mobile phone or hand-held computer. Of course, its precondition is that the user has authorized to do so.

Fig. 6 Sending business card via hand-held device

In our personal access system, Infra-red Data Association (IrDA) standard is the basis to carry out the desired wireless communication of personal data. We use the infrared adapter supplied by ACTiSYS company. It offers 2.4 meters reliable infra-red link distance under 115.2 Kbps baud rate.

4. Integrated PUI in E-commerce Environment

4.1. *Interaction tasks in 3D shopping environment*

Interaction tasks in 3-D environment typically consist of Navigation/travel, Selection/Picking, Manipulation and System Control. Travel is the motor component of navigation, and just refers to the physical movement from place to place. Selection/picking is simply the specification of an object or a set of objects for some purpose. Manipulation refers to the specification of object properties (most often position and orientation, but also other attributes). Selection and manipulation are often used together, but selection may be stand-alone task. For example, the user may select an object in order to apply a command such as ‘delete’ to that object. System control is the task of changing the system states or the mode of interaction. It is usually done with some types of command to the system (either explicitly or implicitly). It is often the case that a system control technique is composed of the other three tasks.

Regarding the 3D interactions in public space for e-commerce, basic interaction requirements are:

- 3D navigation in the market or shop.
- Picking up objects (products) from the goods shelf in 3D scene/market/shop
- Presentation of Product (It is usually predefined, and relies on the content of scene, we can manipulate the product in real-time if we can capture more parameters).
- Tracking of avatar or multi-avatar ‘movements, postures and gestures
- Communication between multiple avatars (users), it could be text or text-to-speech.
- Shopping decision-making and their response actions such as presenting the personal credit card number etc.

4. 2. *Task-oriented integration*

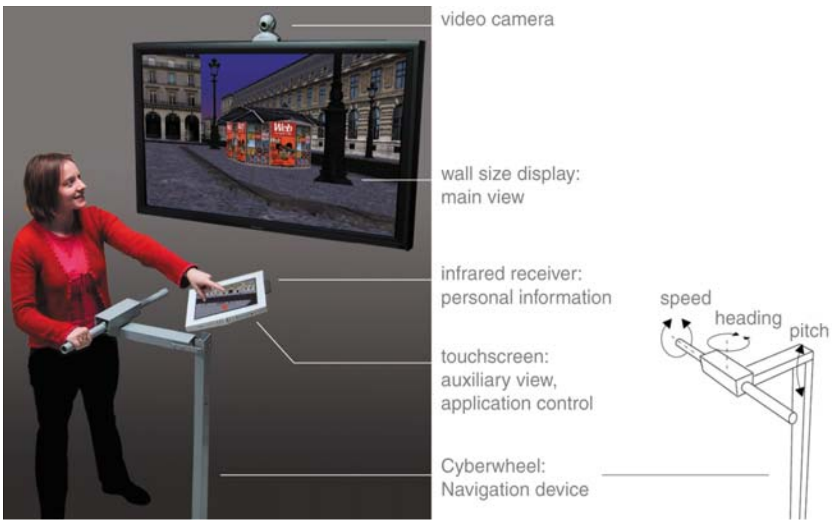

Figure 7 Cyberwheel-based Integrated Interaction for e-shopping

We manage to bring the relevant input/output devices working together, and give a human-computer interface solution for interaction tasks that are commonly involved in e-commerce context. It is assumed that the choice of the input/output device will determine the concrete interaction technique. The prototype installation, shown in Figure 7, consists of the following components/devices:

1) Wall size display, it provides a main view of 3D virtual scene.
2) Cyberwheel, it is a 5 degrees of freedom input device as aforementioned, and here is used for easy and precise navigation in 3D environments.

3) Touch screen, it shows current position of the user via displaying a top-view of the 3D scene. In addition it can also be used as application control to perform some operations or choose objects.
4) Infrared receiver, it is for acquiring user's personal information (name, e-mail, preferences) from hand-held infrared-enabled personal devices (mobile phone, Palm Pilot, etc)
5) Video camera, it is reserved for vision-based pointing interaction or for video conference.

Its potential e-shopping scenario can be described as follows. When the application is launched, the user may at first be asked to beam via infrared link some personal information, such as name and customer ID, from his mobile device, such as Palm Pilot. Cyberwheel performs the navigation in 3D virtual space projected on the wall-size screen. The user can be aware of his current position from the planar map or the top-view displayed on the touch screen. He may jump immediately to a desired point through a single touch. The selection of virtual objects or performing manipulations (e.g. opening a door or presentation of product) are accomplished in the following way. An icon corresponding to the target object, action or operation will be displayed on the touch screen. The user touches the icon to select the object or trigger the presentation/action/operation. The touch screen serves as somewhat remote control for the virtual environment, and the icon is displayed only when the corresponding action is really applicable. The vision-based pointing technique will permit the user to avoid the use of special pointing device that is necessary to carry in a hand. If the user has made the decision of buying something, the paying-related information such as credit-card ID could be input via hand-held mobile devices.

5. Results and Discussion

Perceptual user interfaces will enable multiple styles of interaction via adding human-like perceptual capabilities to the computer, and enhance the computers as tools or appliances, directly enhancing GUI-based applications. Perhaps more importantly, these new technologies will enable broad use of computer assistants, or agents, that will interact in more human-like ways. Towards the goal of setting-up an integrated interface (Figure 8) for e-shopping, we tested current interaction devices in a series of interactive environments including the following prototype application scenarios:

- *Exploration of interactive virtual city*. The system can be installed in a travel agency or in an exhibition with a model of a well-known tourist center. It can be also installed in a museum for exploration of a reconstructed model of an ancient city.
- *Interactive 3D product or company presentation.* The system can be installed in an exhibition or company reception area. The user walks through a virtual office or shop enquiring information about company products.

These evaluations show encouraging results. The proposed installation works well in these testing scenarios, and basically achieves the desired goals. In the programming level, the lessons we learned from the integration are that: to ease the development of more applications with natural interfaces, we must be able to handle other forms of input (audio, video, and sensor input) as easily as keyboard and mouse input

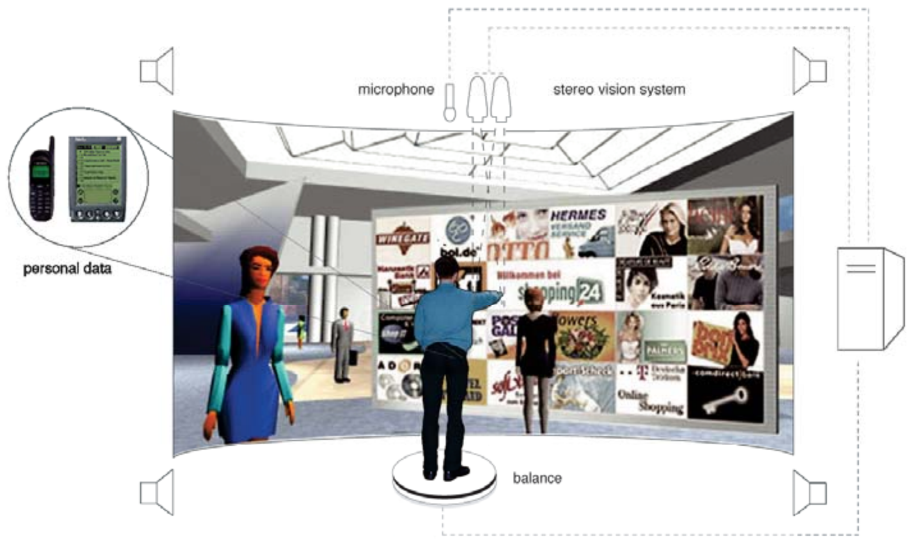

Figure 8 Perceptual User Interface in Virtual Shopping source/MARS/courtesy of blaxxuninteractive

In the future there will be an increasing diversity of user interfaces on an increasing diversity of computerized devices. These devices include hand-held personal digital assistants, cell phones, pagers, computerized pens, computerized notepads, and various kinds of desk and wall-size computers, as well as devices in everyday objects. Interfaces on these computing devices cannot typically use the standard desktop paradigm, and people will not necessarily expect these devices to act like 'regular' computers. PUIs facilitate a richer variety of communications capabilities between humans and computation, and bring our human capabilities to achieve more natural and intuitive interfaces. It is the goal of these perceptual interfaces to support common forms of human expression and leverage more of our implicit actions in the world. They will play more and more important roles in human-computer interaction.

Acknowledgements

The work in this paper was supported by European Commission Project, SoNG (The Portals of Next Generation, IST-1999-10192). Its computing platform is SoNG MPEG-4 player.

References

1. Alex Pentland, "Perceptual Intelligence", Communication of ACM, March 2000, Vol 43, No. 3, 35-44

2. Matthew Turk, George Robertson, "Perceptual User Interface", Communication of ACM, 2000 March, Vol 43 No 3, 33-34

3. Sharma, R., Pavlovic, V. I. & Huang, T. S. , Toward multi-modal human-computer interface. Proceedings of IEEE, 86(5), 853-86, 1998

4. Fleischmann, M. Imagination Systems – Interface Design for Science Art & Education, Proceedings of IMAGINA'96, Monte Carlo

5. Iwata,H. Walking About Virtual Environments on Infinite Floor, Proc of IEEE Virtual Reality Annual International Symposium, 1999

6. Rudolph P. Darken, William R. Cockayne, David Carmein: The Omni-Directional Treadmill: A Locomotion Device for Virtual Worlds. ACM Symposium on User Interface Software and Technology 1997: 213-221

7. Fleischmann, M.; Strauss, W.; Elistratov, V. Interfacing Cultural Heritage,Proceedings of Eurographics-99, Milano, Italy, September 1999

8. Kolesnik, M., G. Paar, A. Bauer and M. Ulm. Algorithmic Solution for Autonomous Vision-Based Off-Road Navigation. In Proc. of SPIE: Enhanced and Synthetic Vision, April 1998, Orlando, Florida.