Motion Compression Applied to Guidance of a Mobile Teleoperator*

Patrick Rößler, Frederik Beutler, and Uwe D. Hanebeck
Intelligent Sensor-Actuator-Systems Laboratory
Institute of Computer Science and Engineering
Universität Karlsruhe (TH)
Karlsruhe, Germany
{*patrick.roessler* | *beutler* | *uwe.hanebeck*}*@ieee.org*

Norbert Nitzsche
Institute of Automatic Control Engineering
Technische Universität München
München, Germany
norbert.nitzsche@ei.tum.de

***Abstract*— Telepresence aims at giving a human user the impression of being present in a remote environment. However, the user is actually situated in a user environment and his motion is tracked. A mobile teleoperator in the remote environment replicates this motion. The user can thus control the mobile teleoperator's locomotion by walking. A stereo-camera system mounted on the mobile teleoperator constantly records live camera images and transfers them to the user environment, where they are presented to the user on a head-mounted display.**

This paper presents a long distance experiment, in which a mobile teleoperator was controlled over a standard internet connection by natural locomotion. Without further processing of the user's motion data, however, only exploration of a remote environment of the same size or smaller than the user environment is possible. As this is not desirable, we use Motion Compression, an optimal nonlinear transformation of the user's path. This algorithm allows controlling free motion in an arbitrarily large target environment from a limited user environment.

***Index Terms*— Telepresence, Motion Compression, Tracking, Mobile Teleoperator**

I. Introduction

Telepresence gives a human user the impression of actual presence in a remote environment. This is achieved by tracking the user's motion and transferring it to a robot, the so called *teleoperator*, which replicates this motion. In return, the teleoperator gathers sensory data, which is presented to the user. In order to achieve a high grade of immersion, it is important that the user only perceives the target environment. Live camera images recorded by the teleoperator are presented to the user on a head-mounted display. As a result the user has the impression of presence in the remote environment, i. e., he identifies with the teleoperator.

In order to allow exploration of large target environments, *mobile teleoperators* are used. These are typically wheel-based platforms with a pan-tilt camera head attached to it, but other systems, like humanoid robots, are also possible.

*This work was supported in part by the German Research Foundation (DFG) within the Collaborative Research Centers SFB 588 on "Humanoid robots – learning and cooperating multimodal robots" and SFB 453 on "High-fidelity telepresence and teleaction".

In most current systems, the user position is fixed and only rotational head motion is tracked and transferred to the teleoperator. The locomotion of the mobile teleoperator is usually controlled by means of joysticks [1], foot pedals [2], steering wheels [3] or similar input devices, while the user is sitting on a chair. The lack of proprioceptive feedback, i. e., the sense of self-motion, gives the user the feeling of flying or driving through the target environment.

There are several systems for virtual reality applications, that aim at giving the user a more realistic impression of walking through a virtual target environment. In [4] a walking-in-place metaphor is used to navigate through the target environment. Other work proposes complex user interfaces like an omnidirectional treadmill [5], moving floor tiles [6], or even more sophisticated devices [7]. In [8] a system is presented, that allows the user to explore a real target environments by walking freely. This is achieved by tracking the head and hip of the human user and controlling the mobile teleoperator accordingly. The user has not only visual but also proprioceptive feedback of his motion, which yields in better performance in localization and navigation tasks [9]–[11].

Without further processing of the motion data or the use of complex hardware, a system like the one described above allows only exploration of a target environment smaller or equal to the size of the user environment. This is especially undesirable, as the target environment may be of arbitrary size and shape, but the user environment is typically limited, for example by the range of the tracking system or simply by the available space. Scaling of the user's motion is not an appropriate solution, as it leads to loss of immersion. This problem can be solved by using Motion Compression [12], an algorithm which allows exploration of large target environments by moving freely in a limited user environment. Motion Compression provides a nonlinear mapping between the user's path and the teleoperator's path. A high grade of immersion is guaranteed by preserving path length, as well as turning angles. Although this algorithm proved highly useful and intuitive for navigation in virtual environments, it had yet to prove its applicability to guidance of mobile teleoperators. This paper presents a telepresence system,

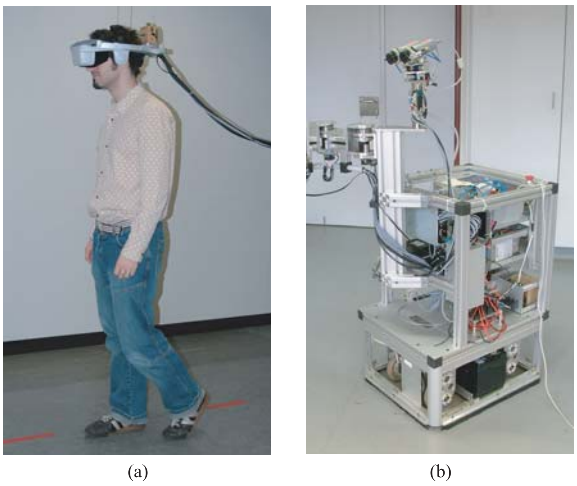

Fig. 1. User in local user environment (a) controlling a mobile teleoperator the remote target environment (b).

that allows a human user (Fig. 1(a)) to control a mobile teleoperator (Fig. 1(b)) by using Motion Compression over an internet connection.

The remainder of this paper is structured as follows. Section II reviews Motion Compression. An overview of the system setup is given in section III. Section IV describes the mobile teleoperator. The setup of the telepresence system is explained in section V and results are discussed in section VI. Finally, conclusions are drawn in section VII.

II. Motion Compression

In order to allow exploration of an arbitrarily large target environment, while moving in a limited user environment Motion Compression provides a nonlinear transformation between the desired path in the target environment, the *target path*, and the user's path in the user environment, the *user path*. The algorithm consists of three functional modules.

Path prediction gives a prediction of the desired target path based on the user's head motion and on knowledge of the target environment. If no knowledge of the target environment is available, path prediction is based completely on the user's view direction.

Path transformation transforms the target path into the user path in such a way, that it fits into the user environment. In order to guarantee a high grade of immersion the user path has the same length and features the same turning angles as target path. This means that a user, who walks on the user path, will control the teleoperator the same distance along the target path. If the user turns to one side, the teleoperator will perform the same turning motion. The two paths differ, however, in path curvature. The nonlinear transformation found by the path transformation module is optimal regarding

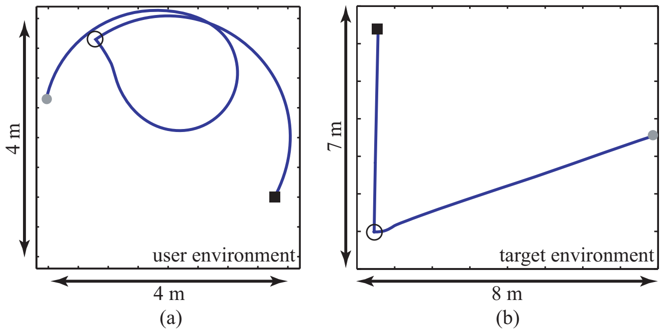

Fig. 2. The corresponding paths in both environments. User path in the user environment (a) and the target path in the target environment (b).

the difference of path curvature. Fig. 2 shows an example of the corresponding paths in both environments.

Finally, *user guidance* guides the user on the user path, while he has the impression of actually walking along the target path. User guidance benefits from the fact, that a human user walking in a goal oriented way constantly checks for his orientation toward the goal and compensates for deviations. By introducing small deviations in the teleoperators posture, the user can be guided on the user path.

In order to clarify this fact, we will give an example. Fig. 3 (a) shows the initial situation, with user environment and target environment superimposed in such a way, that user and teleoperator coincide. The target path is given as a straight line and the user path is curved to the left. When the user now moves straight forward in the direction, where he perceived the goal, the teleoperator follows a path curved to the right (Fig. 3 (b)). The user now perceives his goal to the left, compensates for this deviation by turning to that direction, and thus follows the user path as shown in Fig. 3 (c). In [13] a more in detail view on user guidance is given and major enhancements are proposed. These enhancements, however, are not yet included in the current implementation of the algorithm.

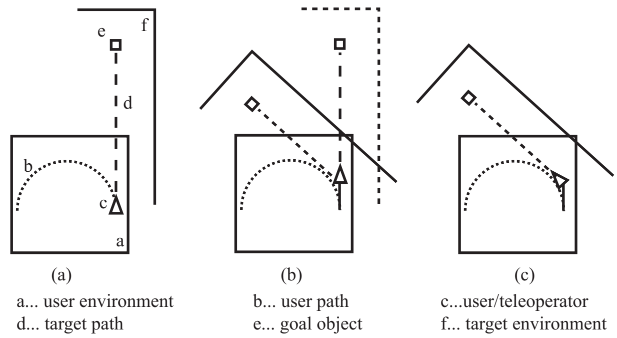

Fig. 3. User guidance in Motion Compression in three stages (a), (b), (c).

All three modules are executed every time an update of the user's posture is available. Motion Compression provides a nonlinear location variant transformation, that can be described by a linear homogeneous transformation at any given time. This transformation changes, however, as the user moves. More details can be found in [12].

III. System Overview

This section gives an overview of the data flow between user and mobile teleoperator. The tracking system tracks the user's head posture. This posture information is now fed into the Motion Compression implementation. Every time an update of the user's posture is available, the three steps of the algorithm are executed as described in section II. The linear transformation calculated by Motion Compression is used to transform the user's posture into the teleoperator's desired head posture.

The desired head posture is then sent to the mobile teleoperator through an internet connection. Camera head and mobile platform are controlled accordingly. The stereo-camera system mounted on the camera head constantly captures live images, which are compressed and sent to the user. These images are presented to the user on a high-quality head-mounted display. Fig. 4 shows the whole data flow.

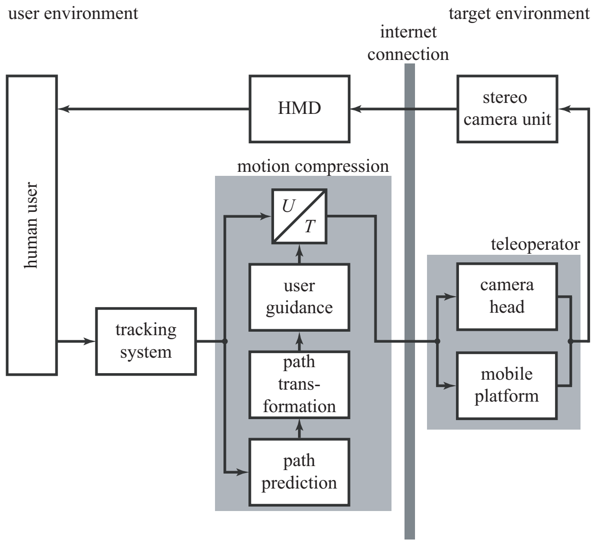

Fig. 4. Data flow in the proposed telepresence system.

IV. Teleoperator

A. Hardware

The mobile teleoperator consists of a mobile platform equipped with a camera head. The camera head is a pan-tilt-roll unit carrying a stereo-camera system.

The mobile platform is an omnidirectional nonholonomic robot, called OmniBase, that is based on four independently driven and steered wheels [14]. The mobile platform has three degrees of freedom (DoF) in cartesian space, that is position on the ground and orientation. Its kinematics, allow it to reach every position in the target environment with an arbitrary orientation. In this work, the mobile platform localizes itself only based on its odometry, which is calculated from the steering angle and distance covered by each individual wheel, as well as gyroscope data for better orientation estimation.

B. Control of Teleoperator

According to the hardware setup, the mobile teleoperator has six DoF in configuration space, while it has only five DoF in cartesian space. There is a redundancy in the azimuth, that is covered by the platform's rotational DoF as well as the camera head's panning angle. This redundancy has to be resolved. The camera head's panning angle is fast but limited, as there are wires running from the cameras to the platform. The platforms rotational DoF, on the opposite, is slow but unlimited. Based on these facts, the redundancy is resolved by frequency and amplitude, i. e., fast and short rotations are handled by the camera head, the remaining rotations are handled by the platform.

Fig. 5 shows, that the platform directly receives the desired position values and the low-pass filtered desired azimuth as command values. Its wheel configurations are then controlled according to [8] in order to match these desired values.

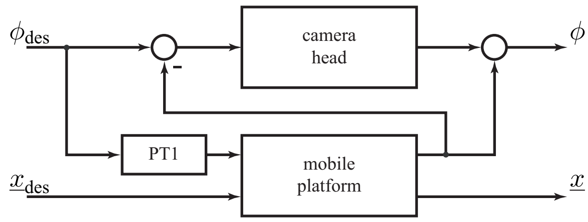

Fig. 5. Control structure of the mobile teleoperator.

The camera head, on the other hand, receives the difference between the original desired orientation and the platforms orientation as command values. As a result of this control structure, the camera head quickly follows the human user's head motion. The platform slowly follows this motion.

V. Experimental Setup

In order to prove the system described above, we conducted a telepresence experiment between the Universität Karlsruhe (TH) in Karlsruhe, Germany and the Technische Universität München in Munich, Germany. While the mobile teleoperator was located in Munich, the user was in Karlsruhe. That means, the experiment covered a distance of approximately 300 km (Fig. 6). User environment and target environment were connected through UDP sockets on a standard shared internet connection. No dedicated communications lines were used. The weakest link in the connection was a 10 Mbit/s LAN-connection. The desired head posture was transferred to the mobile teleoperator at a constant rate of 50 Hz.

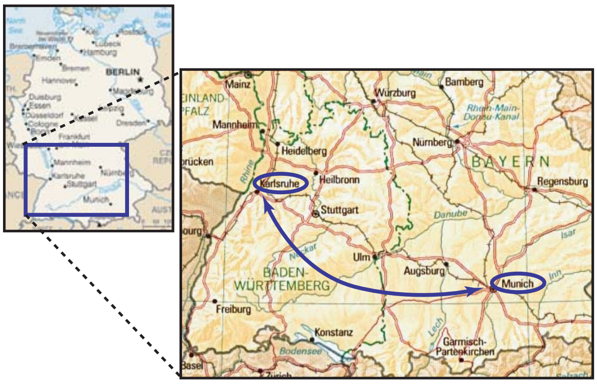

Fig. 6. Map of southern Germany, showing the cities of Karlsruhe and Munich. [15]

A. Tracking System

For the estimation of the user's posture, i. e., translation and orientation, an acoustic tracking system is used. This system consists of four loudspeakers, which are placed in the corners of the user environment. These loudspeakers simultaneously emit different acoustic signals. In order to separate the signals each acoustic signal is spread with an orthogonal Gold Code c^i of the block length $N = 32$. The coefficients of the orthogonal Gold Code are multiplied with an orthogonal carrier frequency. Thus, the generated signal for loudspeaker i is

$$s^i(t) = \sum_{k=0}^{N-1} c_k^i \cdot \cos\left(2\pi \frac{k}{T} t + 2\pi f_s^i t\right) \cdot \mathrm{rect}\left(\frac{t}{T} - \frac{1}{2}\right) , \quad (1)$$

where $\mathrm{rect}(t)$ is the rectangular pulse. The starting frequency is given as f_s^i and T is the signal length. Hence, the end frequency f_e^i is chosen according to

$$f_e^i = f_s^i + \frac{N-1}{T} . \quad (2)$$

The signals are received by four microphones attached to the head-mounted display. The received signals are delayed and attenuated depending on the distance between the loudspeakers and the microphones. Each received signal is filtered with a bandpass filter, according to the given starting and end frequencies. In order to estimate the time delay between sending and receiving the signal, the cross correlation between the filtered signal and the transmitted signal is calculated. The estimated time delay is then converted to the range based on the velocity of sound. Based on the arrangement of four loudspeakers and four microphones 16 estimated ranges are available. A gradient descent algorithm uses these estimates to estimate the posture of the user's head. The initial values for the gradient descent are computed with an algorithm presented in [16]. This algorithm is based on the decoupling of orientation and translation. Both signal generation and signal processing are implemented on a digital signal processor card from Analog Device (Blackfin 533).

The tracking system uses a gyroscope cube to smooth the orientation data. The gyroscope cube consists of three orthogonally arranged gyroscopes from Analog Device. The sensor data is recorded with an MSP430-149 microcontroller from Texas Instruments. The measured angular velocities are sent to a PC, where the posture estimates from the acoustic tracking system is fused with the data from the gyroscope cube by means of a Kalman Filter. Fig. 7 shows the setup for tracking the head-mounted display.

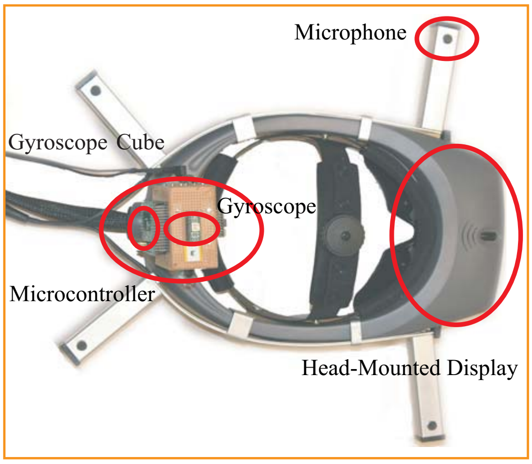

Fig. 7. Top view on the hardware setup consisting of four microphones and a gyroscope cube mounted on a head-mounted display.

B. Processing of Camera Images

In order to give the user a realistic impression of the target environment it is crucial to display the visual information from the target environment, at a high update rate and with low delay. The stereo-camera system synchronously records images from both cameras and sends them to the user. The image size is set to 320×240 Pixels to reduce network load. A further reduction of bandwidth usage is obtained by applying a JPEG-compression on these images.

The JPEG-images are split up into data packages, which are transferred to the user through a UDP socket connection. On the user site the packages are received, decompressed and supersampled to the HMD's resolution of 1280×1024 Pixels using the video card's hardware acceleration.

Decompression and display of the stereo-images is performed incrementally. This means, the image data from the first received data packages is already displayed, while the remaining packages are still arriving. This reduces delay compared to conventional MPEG decoders, which only process complete images.

VI. Experimental Results

A. Tracking System

The acoustic tracking system provides 15 estimates for the posture per second. However, the gyroscope cube provides 50

estimates for the orientation per seconds, which results in an update rate for the posture sufficient to get a good immersion. The performance of the tracking system was evaluated in a test run with a predefined motion trajectory. The user had to move from the start position on a circular path. He was asked to walk in view direction. Fig. 8 illustrates the motion trajectory for the translation. The vectors are directed in view direction. The system provides a good relative accuracy, as can be seen in Fig. 8. This is very important for this application. Absolute accuracy was shown to be sufficient by comparing the tracking results with a few reference positions. However, absolute accuracy is of less importance for the application.

The algorithm detects outliers in the range estimates based on the arrangement of the microphones. In this case the algorithm provides no results for the posture. However, after some acoustic measurements the algorithm provides reasonable ranges. Thus, the algorithm estimates the users posture and the user is tracked again.

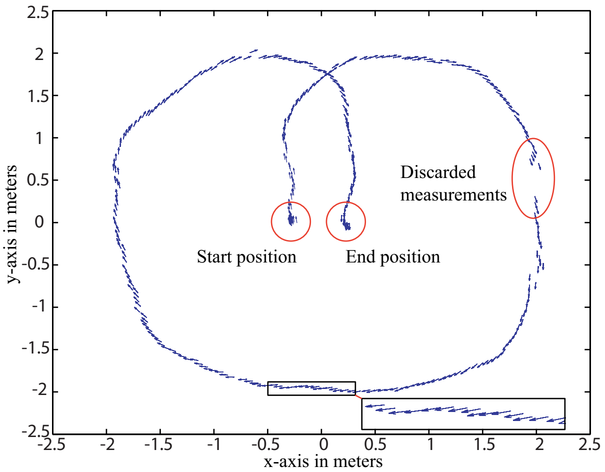

Fig. 8. The estimated translation sequences in a test run with a predefined motion trajectory.

B. Teleoperation

In order to obtain a measure of performance, the user was asked to perform a navigation task. The task was to navigate the mobile teleoperator from a lab room, through a door, and along two hallways. Fig. 9 shows the resulting target path as controlled by the user. Note, that this path is not predefined but a result of deliberate actions performed by the user. A video of the experiment can be found at [17].

The task can be segmented into four subtasks, approaching the door (i), passing the door (ii), moving along the first hallway (iii), and moving along the second hallway (iv). Fig. 10 (a) shows the user's path in the user environment. In Fig. 10 (b) a detail of the path with subtask (ii) is shown. As a result of Motion Compression, the user walks on a curved path, while the teleoperator moves almost straight along. In order to achieve a minimal curvature difference, the curvature switches from right curvature to left curvature, when changing from subtask (ii) to subtask (iii). Previous experiments with Motion Compression show, that the algorithm is extremely robust against errors in path prediction.

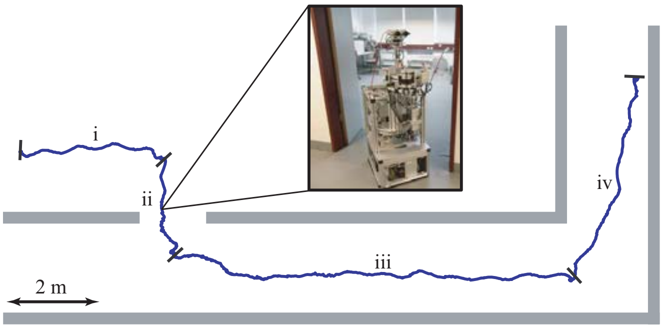

Fig. 9. Target path controlled by the user. Approaching the door (i), passing the door (ii), moving along the first hallway (iii), and moving along the second hallway (iv).

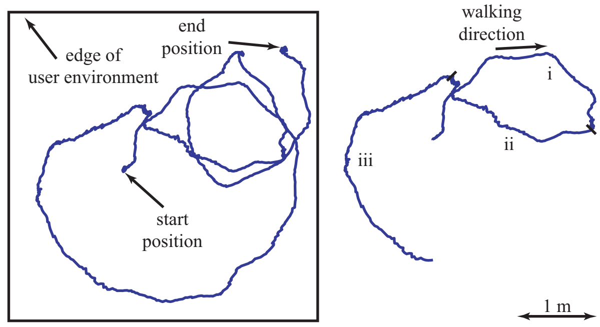

Fig. 10. The complete user path in the user environment (a). Detail of the user path (b). Approaching the door (i), passing the door (ii), and moving along the first hallway (iii).

In this experiment an earlier version of the tracking system, without inertial sensors, was used. As this system does not provide elevation and roll angles, the corresponding DoF of the mobile teleoperator were fixed to zero.

Table I shows the completion times for the four subtasks. The row *time direct* gives the completion times of a human user actually completing the task directly. The ratio of completion times is between 3.0 and 5.3 for all subtasks, which can be considered as a good result. As shown in Fig. 11

TABLE I

DISTANCES COVERED AND COMPLETION TIMES.

subtask	(i)	(ii)	(iii)	(iv)
distance covered	3.0 m	3.2 m	8.4 m	5.1 m
time teleoperator	14.0 s	15.8 s	25.0 s	14.7 s
time direct	3.4 s	3.0 s	8.2 s	4.4 s
ratio teleoperator/direct	4.1	5.3	3.0	3.3

the teleoperator quickly reaches a speed of approximately 0.5 m/s in the hallways, which is a relaxed walking speed. However, it can be clearly seen, that at the subtask changes speed drops to almost zero. This is a result from the user taking a short period of time for regaining orientation after performing a $90°$ turn. As a result from the narrow field of view, which does not allow the user to see the doorframe while walking through it, the speed is especially low when passing the door. This is also an explanation for the high ratio of completion times in subtask (b).

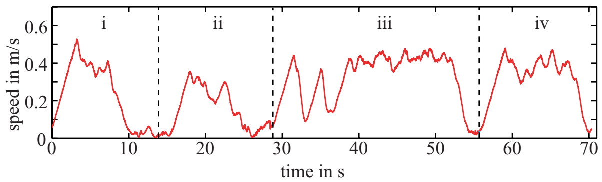

Fig. 11. Low-pass filtered teleoperator speed reconstructed from positions measured by odometry.

There was, however, a substantial delay, which was estimated by the user to be between 500 ms and 1 s. This delay is most probably the reason for the short loss of orientation after performing wide turns.

VII. Conclusions

This paper presents a telepresence system, that allows a human user to control a mobile teleoperator in a natural and intuitive way over an internet connection. The teleoperator's locomotion is controlled by the user's locomotion and, thus, gives the user the sensation of presence in the target environment. In order to enable the user to explore arbitrarily large target environments, while moving in a limited user environment Motion Compression, an optimal nonlinear transformation between user path and target path, is used.

A newly developed acoustic tracking system, equipped with additional inertial sensors, is used to estimate the user's head posture. This posture is transformed into the teleoperator's desired head posture and the teleoperator's camera head is controlled accordingly. Live stereo-camera images are constantly transmitted to the user, giving him an impression of the target environment.

Experiments with a long-distance telepresence application show good results. When walking straight along hallways, user and teleoperator almost reach normal walking speed. However, navigating through narrow passages still proves to be a difficult task. This might be simplified by giving the user more information about his environment, for example by augmenting additional sensor information into the display. The main problem of the system is, that after wide angle turns, users frequently lose orientation for a short period of time. This is caused by a substantial delay experienced by the user.

Future work will focus on eliminating this delay as good as possible. By using appropriate models for human motion and adequate filtering algorithms, the mobile teleoperator can be controlled predictively.

In order to allow real interaction with the target environment and to provide even better immersion, this system will be enhanced with an audio channel and haptic feedback. It will then cover the four human senses: vision, hearing, haptics and proprioception. The resulting system will be a highly intuitive interface for mobile robots.

References

[1] R. Stemmer, R. Brockers, S. Drüe, and J. Thiem, "Comprehensive Data Acquisition for a Telepresence Application," in *Proceedings of Systems, Man, and Cybernetics*, The Hague, The Netherlands, 2004, pp. 5344–5349.

[2] S. M. Goza, R. O. Ambrose, M. A. Diftler, and I. M. Spain, "Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform," in *Proceedings of ACM International Conference on Human Factors in Computing Systems (CHI 2004)*, Vienna, Austria, 2004, pp. 623–629.

[3] C. Bunz, M. Deflorian, C. Hofer, F. Laquai, M. Rungger, F. Freyberger, and M. Buss, "Development of an Affordable Mobile Robot for Teleexploration," in *Proceedings of IEEE Mechatronics & Robotics (MechRob'04)*, Aachen, Germany, 2004, pp. 865–870.

[4] M. Slater, M. Usoh, and A. Steed, "Steps and Ladders in Virtual Reality," in *ACM Proceedings of VRST '94 - Virtual Reality Software and Technology*, 1994, pp. 45–54.

[5] H. Iwata, "The Torus Treadmill: Realizing Locomotion in VEs," *IEEE Computer Graphics and Applications*, vol. 19, no. 6, pp. 30–35, 1999.

[6] H. Iwata, H. Yano, H. Fukushima, and H. Noma, "CirculaFloor," *IEEE Computer Graphics and Applications*, vol. 25, no. 1, pp. 64–67, 2005.

[7] H. Iwata, H. Yano, and F. Nakaizumi, "Gait Master: A Versatile Locomotion Interface for Uneven Virtual Terrain," in *Proceedings of the IEEE Virtual Reality Conference*, 2001, pp. 131 – 137.

[8] N. Nitzsche and G. Schmidt, "A Mobile Haptic Interface Mastering a Mobile Teleoperator," in *Proceedings of IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, Sendai, Japan, 2004.

[9] R. P. Darken, T. Allard, and L. B. Achille, "Spatial Orientation and Wayfinding in Large-Scale Virtual Spaces: An Introduction," *Presence*, vol. 7, no. 2, pp. 101–107, 1998.

[10] B. Peterson, M. Wells, T. A. Furness III, and E. Hunt, "The Effects of the Interface on Navigation in Virtual Environments," in *Proceedings of Human Factors and Ergonomics Society 1998 Annual Meeting*, vol. 5, 1998, pp. 1496–1505.

[11] M. J. Tarr and W. H. Warren, "Virtual Reality in Behavioral Neuroscience and Beyond," *Nature Neuroscience Supplement*, vol. 5, pp. 1089–1092, 2002.

[12] N. Nitzsche, U. D. Hanebeck, and G. Schmidt, "Motion Compression for Telepresent Walking in Large Target Environments," *Presence*, vol. 13, no. 1, pp. 44–60, 2004.

[13] P. Rößler, U. D. Hanebeck, and N. Nitzsche, "Feedback Controlled Motion Compression for Extended Range Telepresence," in *Proceedings of IEEE Mechatronics & Robotics (MechRob'04), Special Session on Telepresence and Teleaction*, Aachen, Germany, 2004, pp. 1447–1452.

[14] U. D. Hanebeck and N. Šaldic, "A Modular Wheel System for Mobile Robot Applications," in *Proceedings of the IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS)*, Kjongju, Korea, 1999, pp. 17–23.

[15] "Maps courtesy of the University of Texas Libraries, The University of Texas at Austin." [Online]. Available: http://www.lib.utexas.edu/maps/germany.html

[16] F. Beutler and U. D. Hanebeck, "Closed-Form Range-Based Posture Estimation Based on Decoupling Translation and Orientation," in *Proceedings of IEEE Intl. Conference on Acoustics, Speech, and Signal Processing (ICASSP05)*, Pennsylvania, PA, USA, 2005, pp. 989–992.

[17] "Telepresence Experiment between Karlsruhe and Munich." [Online]. Available: http://isas.uka.de/site_de/research/TelepresenceKaMuc.avi