M. Buss
A. Peer
T. Schauß
N. Stefanov
U. Unterhinninghofen

Institute of Automatic Control Engineering,
Technische Universität München, Munich,
Germany
{mb, angelika.peer, schauss, nikolay.stefanov,
ulrich.unterhinninghofen}@tum.de

S. Behrendt
J. Leupold

Institute for Real Time Computer Systems,
Technische Universität München, Munich,
Germany
{behrendt, leupold}@rcs.ei.tum.de

M. Durkovic
M. Sarkis

Institute for Data Processing,
Technische Universität München, Munich,
Germany
{durkovic, michel}@tum.de

Development of a Multi-modal Multi-user Telepresence and Teleaction System

Abstract

The presented multi-user telepresence and teleaction system enables two teleoperators, which are independently controlled by two human operators, to perform collaborative actions in a remote environment. The system provides the operators with visual, auditory, and haptic feedback and allows them to physically interact with objects in the remote environment. The interaction between the two human operators is enhanced by augmenting visual and auditory feedback. The paper is focused on the amalgamation of the individual subsystems, which handle the different modalities, into one tightly integrated system, which creates a common workspace for two operators. The overall system architecture and the appropriate design of the auditory, visual, and haptic subsystems are discussed. The audio system makes use of a novel high-fidelity interpolation technique to render three-dimensional sound scenes for both human operators, which enhances their interaction. The design of the video system allows modeling and rendering of the remote environment in real time while regarding changes in the scene by perpetually updating the model. The haptic system is based on admittance-type devices, which are best fitted for applications involving large workspaces and high interaction forces. In addition, the implementation of locomotion techniques for telepresence in large-scale environments are presented. Finally, an application example shows that the system can be successfully employed in a remote maintenance task, which consists of exploring a large-scale environment, moving to the target area, and finally repairing a broken pipe by attaching a sealing clamp. The example demonstrates the necessity of multi-user telepresence and teleaction systems and supports the benefits of consistent multi-modal feedback.

KEY WORDS—multi-modality, multi-user, telepresence, teleaction

Figures 1, 3, 4, 6–10, 13–17 appear in color online: http://ijr.sagepub.com
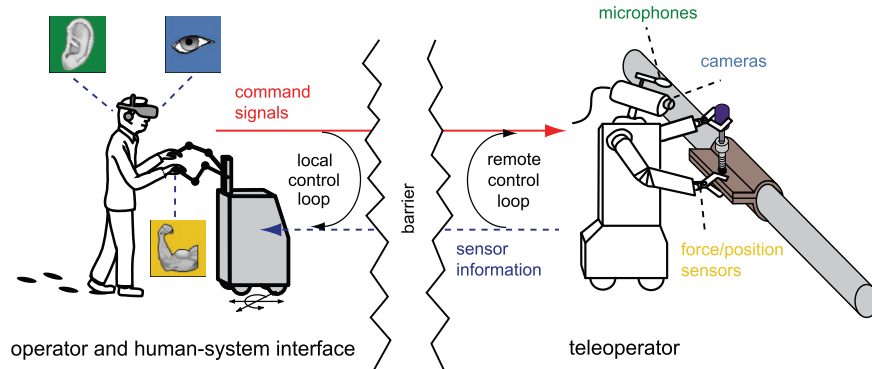
Fig. 1. Principle of a multi-modal telepresence and teleaction system.

1. Introduction

When humans have to perform tasks in dangerous, distant or simply inaccessible environments telepresence and teleaction systems (see Figure 1) play an important role. They combine human skills such as reasoning and decision-making with the advantages of robotic manipulation. Using such a system the human operator is no longer in direct contact with the environment, but interacts with it by means of technical systems.

In past decades several application areas for telepresence and teleaction systems have been presented: space and underwater exploration (Reintsema et al. 2007; Ridao et al. 2007), surgery (Cavusoglu et al. 2003; Ortmaier et al. 2004), plant maintenance, micro-assembly (Sitti et al. 2003; Reinhart et al. 2004), toxic waste cleanup, telemanufacturing, training as well as entertainment.

Most telepresence and teleaction systems focus on single operator, single robot systems only, whereby a single human operator is responsible for carrying out the task. We introduce the *haptic-visual-auditory (HVA) workspace*, a common workspace that allows collaboration of distributed human operators by means of visual, auditory, and haptic modalities; see Figure 2. Since, in such a collaborative HVA workspace, different facilities and capabilities are shared among a group of operators, very complex tasks become feasible.

One of the main demands on such a HVA workspace is the ability of realistically displaying the remote environment. In order to achieve such a high-quality teleoperation, we have developed appropriate design, control, and rendering concepts and incorporated human factors in the development process. In this article we will present the resulting highly integrated teleoperation system that combines the newly developed visual, auditory, and haptic rendering systems.

This article is structured as follows. Section 2 reports the state of the art in multi-modal teleoperation systems. In Section 3, a brief overview of the newly developed teleoperation system and its system architecture is given. Sections 4, 5, and 6 provide a detailed description of the implemented algorithms related to the auditory, visual, and haptic components. In Section 7, the mobile components are described that enable operation in extensive remote environments. These sections are followed by a typical application scenario of the developed telepresence and teleaction system and the results achieved for each single modality are presented. Finally, the article concludes with a summary and outlook on future research.

2. State of the Art

Although nowadays many telepresence and teleaction systems exist, most of them are very limited in their functionalities and performance. The literature provides a considerable number of systems that are limited to a few degrees of freedom (DOF) and/or a relatively small workspace so that full spatial immersion is not achieved (Backes and Tso 1990; Hayati et al. 1990; Yoon et al. 2000; Chong et al. 2002; Kron and Schmidt 2003). Only a few more advanced telemanipulation systems operating in six DOF and using a telemanipulator with human scaled workspace are known, see e.g. Chang et al. (1999). Many systems also suffer from the fact that no multi-modal feedback is provided.

The by far the most advanced telepresence and teleaction system given in the literature, described in Hasunuma et al. (1999) and Tachi et al. (2003), is restricted to a single operator, single robot system; it allows control of a humanoid telerobot by means of a teleexistence cockpit. Although, the cockpit is equipped with multi-modal feedback systems, they are not as advanced as the systems presented in this article.

The newly developed telepresence and teleaction system presented here is characterized by highly advanced visual, auditory, and haptic feedback systems. The auditory modality is enhanced by three-dimensional (3D) sound synthesis, which renders the voice of the fellow operator at the spatial position of the respective teleoperator. The methods employed utilize
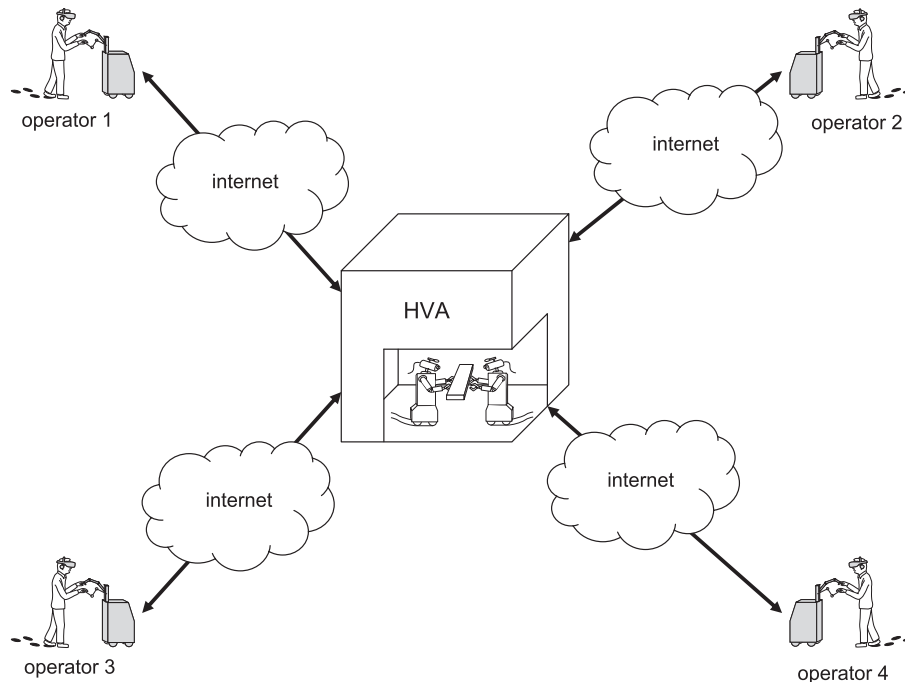
Fig. 2. The haptic-visual-auditory (HVA) workspace.

head-related transfer function (HRTF) databases in combination with an advanced interpolation technique that outperforms existing and well-established techniques, such as e.g. the bilinear method (Begault 1994). The relation between HRTF database resolution, interpolation, and audible errors has recently been explored (Minaar et al. 2005).

The visual modality is enhanced by a hybrid display, which makes it possible to render views from perspectives other than those of the actual cameras without requiring explicit knowledge of the remote environment. The system generates a photo-realistic 3D geometrical model of the remote environment and thus allows one to select an arbitrary view point. Several attempts at using computer generated views in telepresence applications have been described in the literature. However, they are restricted in their hardware prerequisites (Kanade et al. 1997), or triangles are processed individually during the visualization process, which can decrease performance significantly when dealing with detailed sensor models (Ogawa et al. 2005; Burkert 2006).

The system further allows simultaneous locomotion of multiple operators in an extended remote environment and is equipped with highly sophisticated manipulation capabilities.

With respect to state-of-the-art systems, the newly developed teleoperation system is significantly improved in certain points.

In Hasunuma et al. (1999) and Tachi et al. (2003), visual information was provided on several screens. In contrast, we integrate online generated model information with live camera streams into a single visual display. Consequently, the operator is relieved from collecting data from multiple screens, and thus the cognitive load is reduced.

Besides simply replaying sound signals from the remote environment, the system explicitly allows multiple operators to talk to each other and perceive the voice of the peer in a spatially correct manner.

In contrast to exoskeletons which are very fatiguing (Schiele and Visentin 2003), we prefer ceiling or platform mounted haptic interfaces that interact with the human operator only at the end-effector point and do not restrict natural arm movements. Moreover, our system allows simultaneous locomotion and manipulation as the human operator does not need to switch between different operating modes. We use a mobile haptic interface in contrast to a simple motion base which only passively transports the human operator through the environment. This mobile haptic interface allows the human operator to move around and thus gives them a natural feeling of locomotion (see the definition of locomotion interfaces by Hollerbach (2002)). Finally, besides the advancements in the single modalities, the scenario of multiple human operators collaborating in a common HVA has not been addressed before and can be considered unique worldwide.

3. System Architecture

The newly developed and highly integrated telepresence and teleaction system is composed of two teleoperators (see Fig-
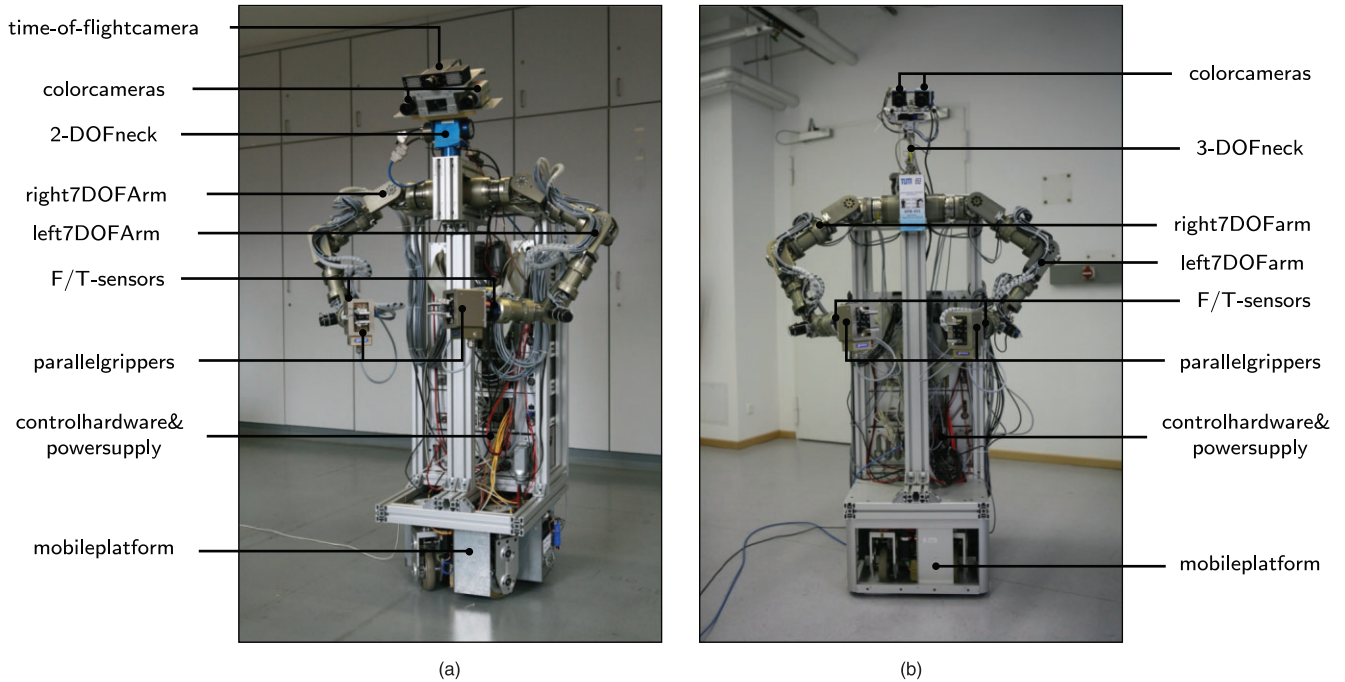
time-of-flightcamera
colorcameras
2-DOFneck
right7DOFArm
left7DOFArm
F/T-sensors
parallelgrippers
controlhardware&
powersupply
mobileplatform

colorcameras
3-DOFneck
right7DOFarm
left7DOFarm
F/T-sensors
parallelgrippers
controlhardware&
powersupply
mobileplatform

(a)   (b)

Fig. 3. (a) Mobile Teleoperator 1 (MTO1), equipped with additional depth imaging camera. (b) Mobile Teleoperator 2 (MTO2).

headtrackingsystem
head-mounteddisplay
leftViSHaRD10
rightViSHaRD10
CyberGlove&
CyberGrasp
F/T-sensors&
magneticcouplings
3-DOFpedal

head-mounteddisplay
F/T-sensors&
magneticcouplings
fngerpotentiometers
rightViSHaRD7
leftViSHaRD7
controlhardware&
powersupply
mobileplatform

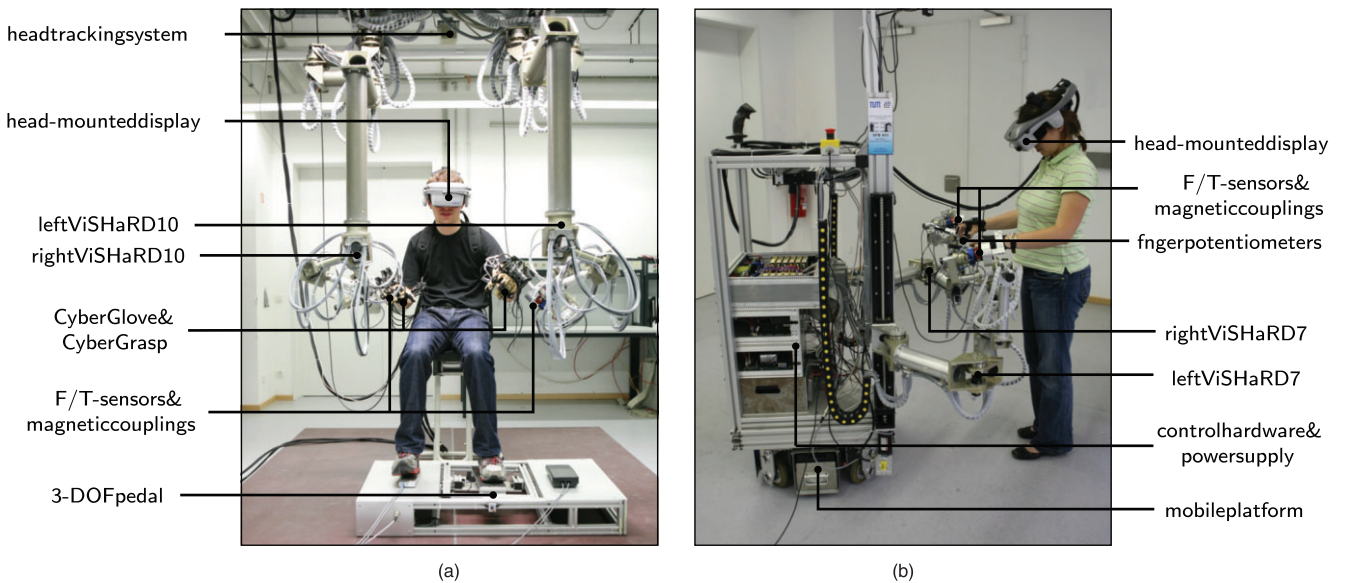(a)   (b)

Fig. 4. (a) Stationary human–system interface (SHSI). (b) Mobile human–system interface (MHSI).

ure 3), which operate in the collaborative HVA workspace, two human–system interfaces (see Figure 4), which can be located far apart from each other, some centralized components for augmented visual and auditory rendering, and a communication network, which transmits data between these com-

ponents. The design of the two teleoperators is almost identical with the exception of the additional 3D vision sensors mounted on one teleoperator. The differences in the design of the two human–system interfaces are more prominent, because the way in which the motion of the teleoperator is commanded

is fundamentally different. One human–system interface is stationary, and thus the velocity of the assigned teleoperator is commanded by means of a pedal; the second human–system interface is mobile such that the operator can freely move around and thereby command the motion of the teleoperator intuitively by his own motion.

In the following subsections, a description of the teleoperators, the two types of human–system interfaces, and the centralized components is given.

3.1. Teleoperators

The *Mobile Teleoperator 1 (MTO1)* is shown in Figure 3(a), and its cousin the *Mobile Teleoperator 2 (MTO2)* in Figure 3(b). Their respective main components are the camera head, which is more sophisticated in the case of MTO1, the two arms with grippers, and the mobile platform.

The heads are each equipped with two cameras and two microphones. The head of MTO1 is additionally equipped with a PMDTec 19K *time-of-flight (TOF)* camera. This camera emits modulated infrared light and determines the time of flight for each pixel by measuring the phase shift, which can be directly transformed into a depth image. Both heads are mounted on pan-tilt platforms with two and three rotational DOF, respectively.

To enable haptic interaction with the remote environment, both teleoperators have two anthropomorphic robot arms designed by Stanczyk and Buss (2004). Each arm possesses seven DOF. A six-DOF force-torque (F/T) sensor at the end-effector measures the interaction forces between robot and environment. The grippers, which are used as end-effectors, are also equipped with force sensors, which can be used to limit or feed back the gripping forces in order to avoid damage to the grasped objects.

The mobile platforms enable the teleoperators to freely move around in the remote environments. Therefore, it is possible to conduct telepresent operations in an arbitrarily large remote environment. The omnidirectional mobile platforms employed enable the teleoperator to approximately mimic the locomotion of the human operator.

In order to increase the autonomy of the teleoperators and to reduce the cabling, all power supplies and control hardware are integrated into the teleoperators.

3.2. Stationary Human–System Interface

Figure 4(a) shows the *stationary human–system interface (SHSI)*. Its components closely correspond to those of the teleoperator.

The images from the stereo cameras are displayed via a *Head-mounted display (HMD)*, and the sound captured by the microphones can be replayed via headphones. The head orientation of the operator is tracked and used to control the orientation of the robot head accordingly.

The robot arms and grippers are commanded by means of haptic interfaces. The two hyper-redundant haptic interfaces VISHARD10 are employed to measure motions and reflect forces. Details on the design, which features high interaction forces and a large, singularity-free workspace, can be found in Ueberle (2006). The commercially available CyberGlove system by Immersion Corp (2007) is used to capture the flection of the fingers and the corresponding CyberGrasp system to exert forces on the fingers. A magnetic coupling mechanism between the haptic interface and the operator ensures a safe operation.

As the operator uses both hands to control the robotic arms, the motion of the mobile platform of the teleoperator is commanded via a pedal. To allow the command of the velocity with two translational and one rotational DOF, an appropriate three-DOF pedal is employed. By tilting and rotating the pedal, the velocity of the teleoperator can be adjusted.

The system is completed by the necessary power supplies, electronics, control hardware, and computers.

3.3. Mobile Human–System Interface

Figure 4(b) shows the *mobile human–system interface (MHSI)*.

Just like the SHSI, its mobile version has an HMD for presenting visual and auditory information. As the operator can move around in a large operating area, an appropriate tracking system, which covers the whole working area, is used to track the position and orientation of the head and body of the operator. The tracking data is used to control the position of the mobile platform and the orientation of the camera head of the teleoperator.

As the VISHARD10 arms used for the stationary interface are too large and heavy for the mobile interface, the smaller seven-DOF version VISHARD7, designed by Peer and Buss (2008a), is mounted on a mobile platform. The functionality, however, is the same and allows the exchange of force and position information with the human operator. The commands to the gripper are sensed through a linear potentiometer which is attached between thumb and index finger. No force feedback of the gripping force is provided. Again, safety is ensured by a magnetic coupling mechanism.

The omnidirectional mobile platform which carries the two haptic interfaces is controlled in such a way that the manipulability of the haptic interfaces is optimized.

3.4. Overall System Architecture

Figure 5 offers an overview of the complex telepresence system. There are two main differences between the two operator–
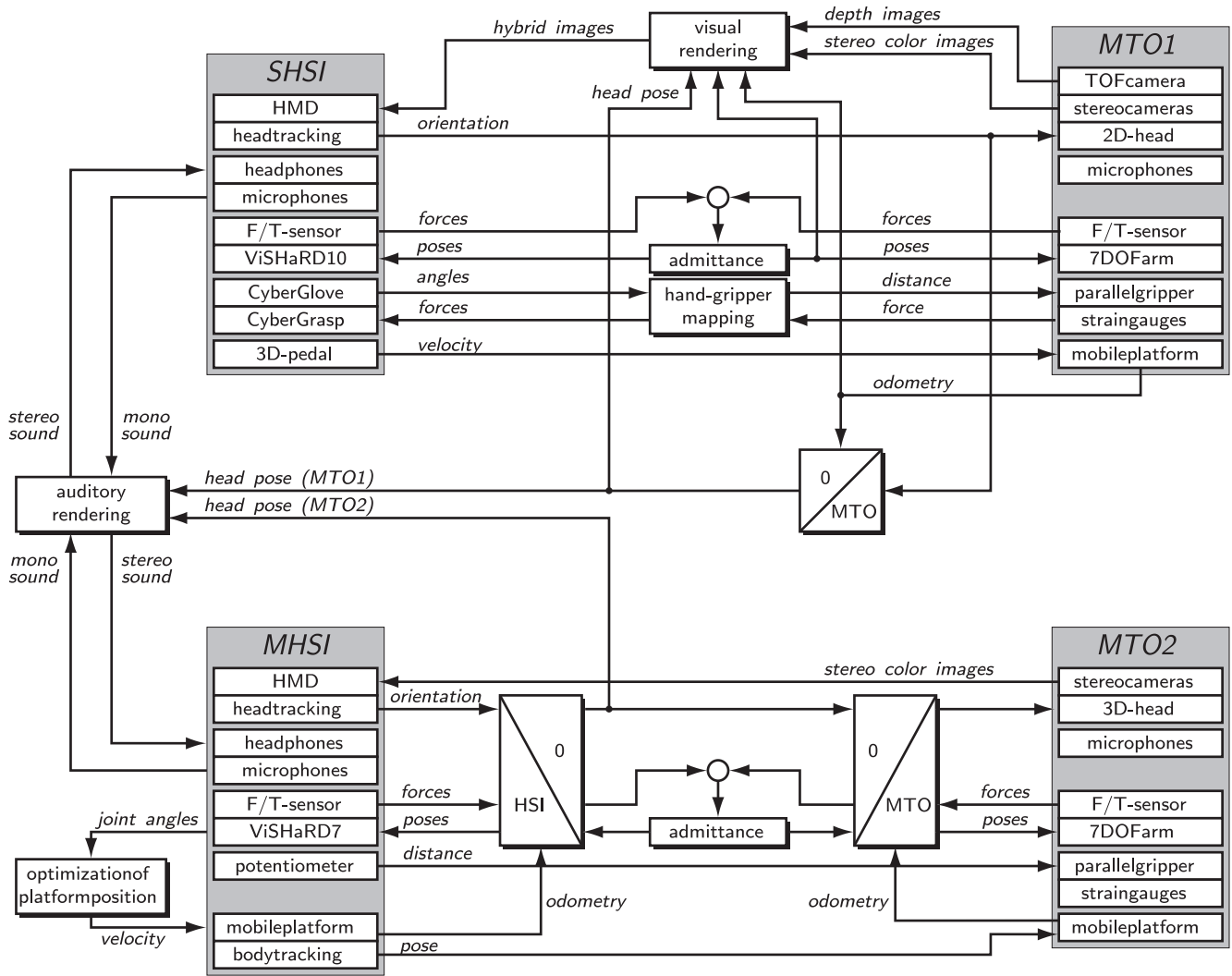
Fig. 5. Overview of the system architecture: the main components and their interconnections.

teleoperator pairs. On the one hand, the pair *SHSI–MTO1* is equipped with additional hybrid visual feedback capabilities and more accurate gripping capabilities. On the other hand, the pair *MHSI–MTO2* is designed for intuitive wide-area telepresence. In principle, the capabilities of both systems could be merged, but this, however, would increase the system complexity even more without any benefits in terms of research outcome.

The subsystems are pairwise coupled, where data is typically exchanged directly between the corresponding components of the human–system interface and the teleoperator: the positions and forces of the arms and grippers are exchanged between the haptic interfaces and the teleoperator; the mobile platforms of the teleoperators receive their commands either from the pedal or the body tracking system; sensed head orientations of the human operators are sent to the teleoperators.

For the MHSI–MTO2 pair, some coordinate transformations have to be applied.

In order to exchange sound signals directly between the two operators, an intelligent auditory rendering system, which produces spatially correct sound impressions, is employed. The auditory renderer is provided with the sound inputs from both operators and the locations and orientations of the respective teleoperators in the HVA. From this data, it renders the sounds for both operators in such a way that they are auditorily immersed in the remote environment.

Augmented visual feedback is generated by a video system, which generates location unbound views of the remote environment. This system is provided with camera images and depth images from the remote environment as well as odometry data, and pose data from MTO1.
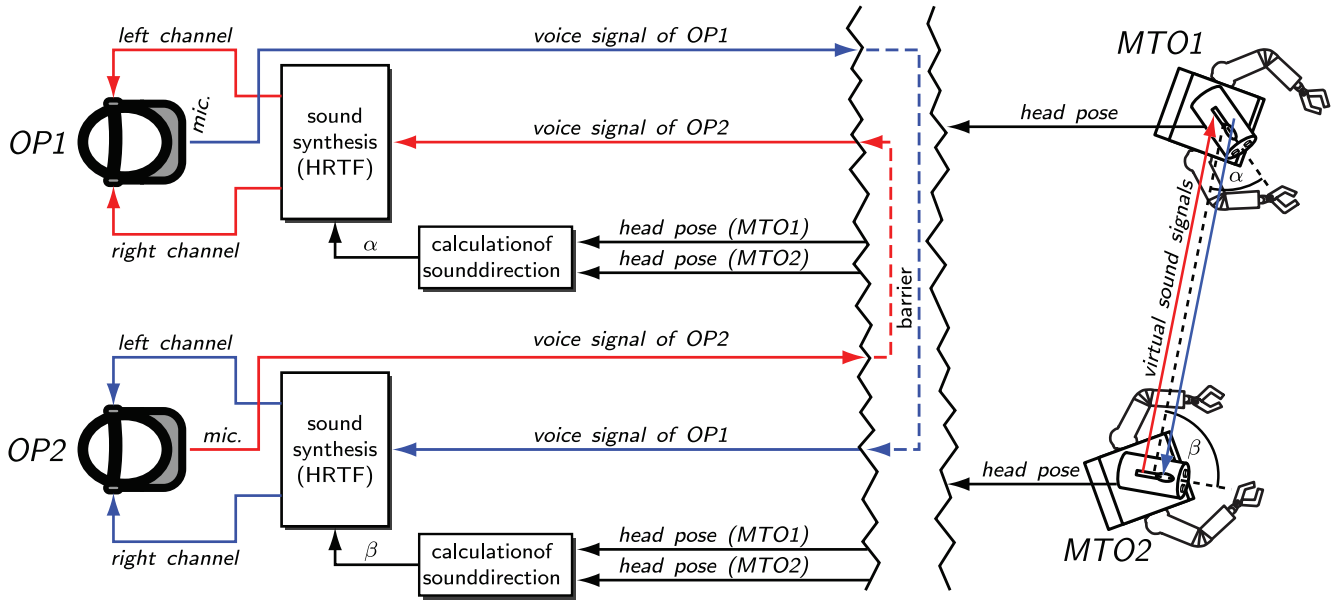
Fig. 6. Auditory telepresence scenario. Real-time rendering of the speech signals of both operators using HRTFs.

4. Audio System

Instead of simply replaying the sounds recorded at the remote site, we have developed techniques for high-fidelity real-time 3D sound synthesis that are able to render accurate virtual sound scenes without noticeable discontinuities for moving sound sources. This section is divided into two parts: the first is concerned with our auditory telepresence scenario, and the second describes the rendering method we have developed.

4.1. Auditory Telepresence Scenario

When several teleoperators are simultaneously working in the same remote environment, 3D sound synthesis can help to deliver auditory cues about the relative positions of the operators. This form of audio communication is very intuitive, since it resembles the feeling of a natural conversation. For example, listeners are able to detect movements of other operators even if those are outside their field of view.

Every operator is equipped with a microphone and headphones and the audio system is primarily used for the communication between operators. In addition to the recorded speech signals the audio system also exchanges information about the positions in the virtual telepresence environment. The virtual auditory spaces depend on the specific position of each teleoperator in the remote environment. The system renders an individual 3D sound scene for every operator using sound and position data. Figure 6 shows a teleoperation scenario with the generation of virtual auditory spaces for two teleoperators.
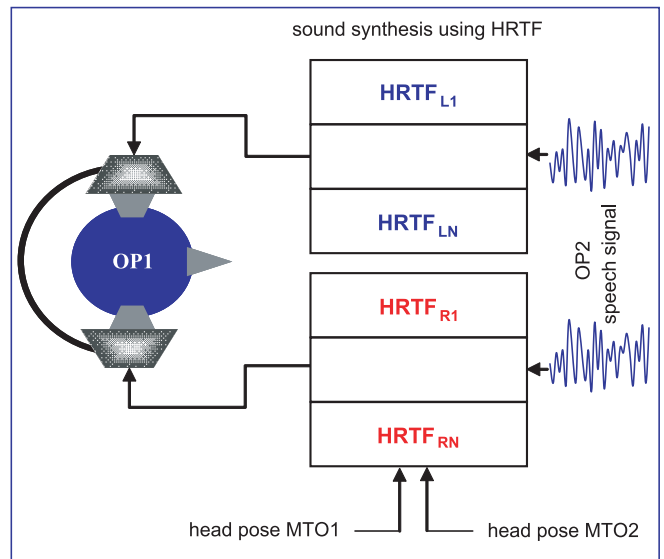

Fig. 7. The 3D sound image reconstruction process involves convolving the sound signal of the other operator with HRTFs.

4.2. Rendering Method

High-quality 3D sound rendering can be performed by utilizing HRTFs. Figure 7 illustrates the adapted process of rendering the incoming sound signal from operator 2 (OP2) for the virtual auditory space of operator 1 (OP1) using HRTFs.

HRTFs are specific to every person and are usually obtained by a complicated measurement process. In order to accom-

plish real-time high fidelity sound synthesis, a densely sampled HRTF database is necessary. When addressing the problem of moving sound sources or when the perceived direction of the sound changes due to the listener's head movements, the switching between different HRTFs has to be addressed. Unless the lowest directional resolution for sampling HRTFs is sufficiently small, a direct switching between them will cause an audible artifact that is heard as a click.

Since the procedure used for measuring a dense database of HRTFs is complex, time-consuming, and requires expensive specialized equipment, only a sparse spatial grid of HRTFs is recorded. It is well known that in real life head turns are often used unconsciously, e.g. to resolve front/back confusions (Blauert 1997). The achievable auditory immersion is affected by the ability to accomplish this switching in a fast and smooth way without creating audible artifacts. Therefore, a good algorithm is necessary to interpolate missing intermediate grid points in our HRTF database as needed. Our group has proposed a binaural impulse–response interpolation algorithm based on the solution for the rational state-space (RSS) interpolation problem (Keyrouz and Diepold 2006) that enables us to estimate missing HRTF database entries. The method is based on the factorization of a block Loewner matrix into a product of generalized observability and controllability matrices. Certain properties to be satisfied by the Loewner matrix are recollected and used to construct a minimal state-space realization of an interpolating matrix transfer function. Motivated by the resulting outstanding interpolation accuracy, and towards reducing the overall synthesis time, we have first reduced the HRTFs length using principal component analysis (PCA) before applying the new rational interpolation method. For more details regarding our interpolation method, the reader can refer to Keyrouz and Diepold (2006).

5. Video System

Most existing telepresence systems provide visual feedback by mounting cameras on the teleoperators and streaming the captured video images to the operator side. In such systems, the mounting positions of the cameras fix the viewing position of the teleoperation.

The presented system goes a step further and provides a stereoscopic display of the remote scene, in which the operator has the ability to choose an arbitrary point of view. Therefore, a 3D geometrical model of the remote environment is constructed and updated over time. The model is transmitted to the operator side along with the video streams where it is finally displayed by rendering the scene graph from the desired view point in a photo-realistic quality.

The 3D scene generation, transmission, and rendering are described in the following sections.

5.1. 3D Scene Generation

The two major components of the stereo-based reconstruction scheme are *stereo matching* and *surface simplification*. Both of these processes are computationally intensive, but they must both be performed to generate the 3D mesh that is necessary to render the scene for the operator. Stereo matching and surface simplification are two different problems which are usually treated independently. Stereo matching is performed using a dense matching algorithm that computes a depth value for each pixel in the image. Many algorithms exist in the literature and good reviews on this topic can be found in Scharstein and Szeliski (2002). Once the dense depth map is estimated, a surface simplification scheme is carried out to generate a simplified mesh of the scene viewed by the teleoperator.

In the scenario presented in this work, stereo matching and mesh simplification must both be applied to obtain a 3D mesh for the virtual display. The stereo matching algorithm computes the depth values at every pixel in the image, and the surface simplification scheme reduces the number of points to generate the simplified 3D mesh. Treating each one of these components independently as is usually done leads to a decrease in the overall performance.

With the algorithm that is presented here, computing the dense disparity map of the scene is not required since only the depth values at the nodes of the reduced mesh are needed to construct the 3D mesh. An image content adaptive meshing scheme is applied instead of surface simplification. The idea is to directly approximate an image with an adaptive 2D mesh of a reduced size and then invoke a sparse stereo matching algorithm that computes the depth values only at the nodes of the corresponding mesh. This will make the matching process avoid all the depth computations for the unused pixels and, therefore, reduce the computational requirements. Consequently, by combining the 2D mesh and the computed depth values from the sparse stereo matcher, we obtain the required 3D mesh that will be used for rendering in Section 5.3. The idea is depicted in Figure 8. As an example, an image adaptive meshing scheme might result in a mesh that has only less than 40% of the number of pixels in the image. Thus, the computational requirement of stereo matching will also be reduced to less than 40% by employing our sparse stereo strategy instead of dense stereo.

To obtain the adaptive mesh, we propose the Tritree scheme introduced in Sarkis and Diepold (2009). It consists of first dividing an image into two main triangles along the diagonal. Then, the approximated intensity values of the image are constructed from the nodes of each of these two triangles. The peak signal-to-noise ratio (PSNR) between the original intensity values of the image and the ones obtained from each triangle is computed. If the PSNR of the reconstructed intensity values of one of the triangles is higher than a predefined threshold, it is left intact. Otherwise, it is divided into two equal smaller triangles along the longest edge. This process is re-
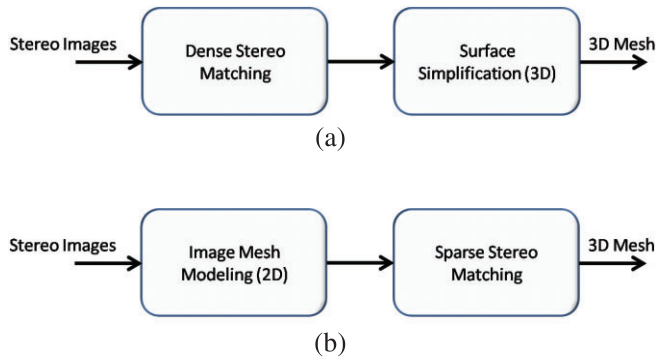
(a)

(b)

Fig. 8. Process of generating a 3D mesh from stereo images: (a) the standard strategy; (b) the technique used in our system.

peated for every new triangle until no further subdivisions are possible. Once the mesh is obtained, we compute the depth values of its nodes via the sparse dynamic programming (DP) strategy, see Sarkis and Diepold (2008b). Note that it is also possible to use other sparse stereo matching strategies as for example the belief propagation proposed in Sarkis and Diepold (2008a). However, the DP scheme is much faster and can attain real-time performance while resulting in good quality disparity maps at the same time.

When the teleoperator navigates in remote scenes where there are not enough textures, the quality of the generated 3D meshes deteriorates since the stereo matching algorithm fails to produce the correct depth values in such sceneries. In these cases, it is possible to integrate the models of the stereo camera system with additional models generated by the photonic mixer device (PMD) camera in order to increase the fidelity and robustness of the system.

5.2. Scene Transmission

Once a new part of the scene geometry is reconstructed, or an older part has changed, this information has to be transmitted to the operator side. Here, the concept of a distributed shared scene graph is used for this purpose for two reasons. First, the display process needs to render views with a high frame rate. A client–server architecture cannot comply with this requirement. In the distributed case each peer has a local copy of the scene graph, and therefore the rendering process is not retarded by network traffic. Second, the reconstructed data should be accessible for all operators. The applied distributed shared scene graph system allows an arbitrary number of peers and transparently exchanges scene graph changes among all peers. In contrast to other known implementations of distributed shared scene graphs (Naef et al. 2003; Schaeffer et al. 2007), the presented system uses MPEG-4 binary format for scenes (BIFS) commands as message protocol. The MPEG-4 BIFS syntax allows high compression rates, as typically a lot

of multi-modal data is exchanged between operators and tele-operators bandwidth resources have to be conserved.

5.3. Rendering the Remote Scene

A photorealistic representation of the remote teleoperator environment must be generated using the received data, namely geometry models and color information from video streams. The major challenges are the high number of triangles sensor-based geometric models consist of and the need to display actual sensor data as fast as possible.

Projective texture mapping provides a much better solution than conventional texture mapping when dealing with a six-figure or seven-figure sum of triangles. This type of texture mapping was used to produce fast shadows in (Segal et al. 1992), and it is well known in image-based rendering applications (Debevec et al. 1998). Using this technique, almost no preprocessing of the camera images is necessary for their use in coloring the scene[1]. By projecting the video streams in a correct way, we generate a colored representation of the currently viewed part of the scene. To present the whole coverage of the scene with color information already gathered by the cameras, further processing is necessary.

We store image information of the camera streams as texture objects in the graphic card memory. Today, 512 megabytes of memory are a common size on high-end graphic boards. That allows us to store enough images to colorize wider areas of the scene. Because of the large number of triangles it is not possible to search for the best image for each triangle in an appropriate time. By the use of programmable shader hardware we modified the fragment processing stage of our graphics board in a way that it accepts a number of images as textures and decides to use one or more of them to colorize the currently processed fragment. The total number of stored images cannot be passed to our modified fragment processing stage. The number is limited in two ways: first of all the number of texture units supported by the graphics card is limited; second, the processing speed is reduced dramatically by passing more images to the fragment shader for decision.

To cope with this limitation we need to restrict the number of images passed to the fragment shader. However, the number of images that are commonly needed to colorize the whole scene exceeds the above-mentioned limits. Therefore, we divide the scene into smaller portions which can be colorized by the use of one or two images. It is important to decide in which way to divide the scene. We use portions of a maximum spatial extent. In particular, we chose cubes of a certain size which are rendered consecutively.

Having many small portions of the scene, *good candidates* must be found to colorize each of them. An ideal candidate for

---

1. In order to minimize distortion artifacts the incoming images must be undistorted. Fortunately it is possible to integrate this operation in our rendering process quite efficiently by the use of programmable shader hardware.

a certain cube is an image fully covering the cube. To render the scene a set of candidates has to be found for each of the cubes[2]. The process of finding the candidates is time critical.

Besides finding candidates we need a second important component. While new images are received from the teleoperator's side and stored in the graphics card memory, we need to delete older images. A simple first in, first out (FIFO) approach is not optimal in this case. Rather than simply delete the oldest images, we should delete images that show parts of the scene that are already covered by newer images. Due to tough timing constraints it is not possible to analyze the content of the images using image processing techniques. We developed two different heuristics, facing these two problems.

The first method restricts movements of the camera to rotational movements around the focal point. In practice this is a viable constraint when mounting a camera on a pan-tilt unit. Even if the camera's focal point does not exactly coincide with the intersection of the two rotational axes the introduced errors are negligible. Candidates are selected by searching for images with little difference in azimuth and elevation of the according cube. We find the two images, which are closest with respect to their directions. From these two images, we delete the older one. For this method we use a tree structure to optimize the search times for selection and deletion. Therefore, the calculation times for both search operations are almost independent from the number of stored images.

The second method allows the camera to be moved in six DOF. Candidates are found by searching for images with a minimum distance from their optical axis to the center of the cube. If an image has to be superseded, the oldest image not used for the current rendering is chosen.

By using the described visualization method, we are able to capture the teleoperator's surroundings and display the captured data immediately. Furthermore, it is possible to integrate perpetually updated sensor data into the rendered representation.

6. Haptic System

Many application scenarios such as disaster operations, rescue, and maintenance tasks require highly dextrous manipulations. In order to avoid operator fatigue and to ease the interaction with the technical system, one of our main efforts was the development of an intuitive telemanipulation system, which does not restrict the human operator when performing manipulations and requires only minor adaptations of the user. In the following sections details of the bimanual and dextrous telemanipulation systems are described.

6.1. Bimanual Telemanipulation System

To achieve a natural interaction with the remote environment, both design and control issues have to be appropriately addressed. In the following paragraphs several requirements of a high fidelity telemanipulation system as developed and described in this article are formulated.

6.1.1. Design Issues

First of all, the workspace of the telemanipulation system plays an important role. If master and slave devices with different workspace sizes are used, motions and feedback forces need to be scaled appropriately. In case the workspace of the slave exceeds the workspace of the master device, indexing and shifting techniques can be applied. However, these techniques are mostly unintuitive, time-consuming, and fatiguing. Thus, to realize an intuitive interaction with the telemanipulation system, haptic interfaces as well as telemanipulators are required, which enable manipulations in six DOF and have a *workspace of the human arm reach* so that fatiguing indexing and shifting techniques can be avoided. The workspace should further be *free of singularities*. Finally, to make full use of human manipulation capabilities, also *bimanual manipulation* tasks should be supported. This can be achieved by providing devices for left and right hands with intersecting workspaces.

Another way to improve the interaction with the remote environment is to build a telemanipulation system that mimics the anthropomorphic manipulation capabilities of a human when performing bimanual six-DOF tele-assembling and telemanipulation tasks. According to Held and Durlach (1992), self-identification with an avatar and thus the feeling of telepresence increases with the similarity of the visual appearance of avatar and human operator. Moreover, the mapping of human operator movements to teleoperator movements plays an important role for the feeling of telepresence. Nash et al. (2000) stated that the better the mapping, the better the feeling of telepresence, because the human operator does not have to think about how to create control actions. Taking this into account, an *anthropomorphic kinematic design* and a human-scaled workspace of the telemanipulator are desirable.

Beside these human factors, the desired application poses restrictions on the design and control concepts as well. While e.g. in applications such as minimal invasive surgery the focus is on the interaction with soft environments such as tissues, in the maintenance field the manipulation of stiff objects such as tubes, handwheels, and metal parts as well as the usage of tools such as screwdrivers and pincers is required. Thus, one of the additional requirements is the *ability to display stiff remote environments*. Boundaries for the maximum displayable stiffness are hereby given by the mechanical rigidity of the single devices as well as by the control architecture used for teleoperation.

---

2. The cubes are rendered consecutively while the fragment shaders need to be parameterized accordingly.

Finally, if the telemanipulation system should be used to perform dextrous manipulation tasks, a *high payload* of the robotic arms is required to mount task specific end-effectors. Such end-effectors include e.g. an exoskeleton or data glove system for the human hand at the master side and a robotic gripper or hand at the slave side. Consequently, *extensibility* of the system plays an important role.

At present, no telemanipulation system which meets all the above presented requirements, namely a human-scaled workspace free of kinematic and algebraic singularities, the possibility of operating in full six DOF using both human arms, an anthropomorphic design of the telemanipulators, the possibility of mounting task-specific end-effectors, and the capability of displaying stiff environments, is known. On this account, the telemanipulation system, presented here, has been developed.

6.1.2. Control Issues

Besides the design, the *implemented control architecture* significantly contributes to the performance of a telemanipulation system. Admittance-type devices as used in the presented telemanipulation system are typically controlled by using a so-called position-based admittance control architecture, see Hannaford and Okamura (2008). This architecture is based on a low-level position controller that compensates for the high inertia and damping of the devices. Depending on the application, such an architecture can be used either to render target dynamics, as e.g. the mass of a tool, or to achieve a compliant behavior when being in contact with the environment. In both cases, the desired behavior is achieved by implementing admittances in the form of mass–spring–damper systems,

$$f = M\ddot{x} + B\dot{x} + Cx, \tag{1}$$

whereby x are positions, f are forces, M stands for the mass, B the damping, and C the stiffness matrix.

In a typical teleoperation scenario, the human operator interacts with a variety of objects which are characterized by different mechanical properties, ranging e.g. from very stiff to soft objects. Depending on the actual task, the human operator also changes their behavior by e.g. adapting the stiffness of their arm. Thus, one of the main challenges in telerobotics is the selection of control architectures and control parameters that are able to robustly stabilize the overall teleoperation system despite changing environment and human operator impedances.

For the presented telemanipulation system a position-based admittance controller with force–force exchange is implemented. In Peer and Buss (2008b) we showed that this control architecture allows stable interaction despite the varying human operator and remote environment impedances. This, however, requires adequate selection of the admittance parameters
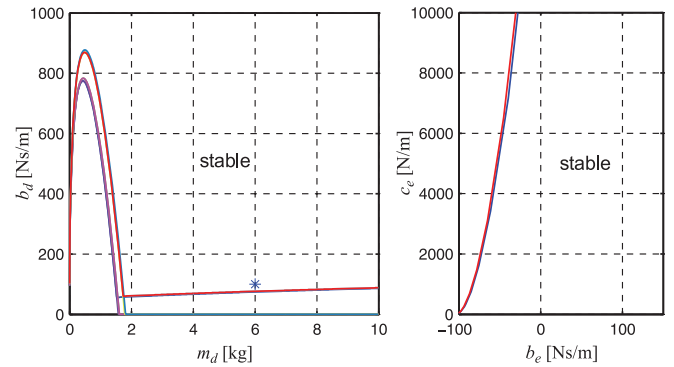

Fig. 9. Stability boundaries of the vertices of the operating domain in the (m_d, b_d)-plane and robustness analysis in the (b_e, c_e)-plane when damped behavior is required, see Peer and Buss (2008b). The parameters m_d and b_d are the mass and damping coefficients, respectively, implemented according to (1) and b_e and c_e stand for the damping and stiffness, respectively, of the environment the human operator interacts with.

as shown in Figure 9. The depicted stability regions are the result of performing stability analysis using linear models for the human–system interface, teleoperator, human operator, and environment, and of applying the parameter space approach (Ackermann 2002).

6.2. Dextrous Telemanipulation

As also dextrous telemanipulation tasks are required, the telemanipulator is additionally equipped with robotic grippers or robot hands, which allow the human operator to manipulate objects of different form and size. The developed dextrous telemanipulation system is illustrated in Figure 10. The data gloves are used to measure human hand and finger motions, which are then mapped to the grippers. When the grippers are in contact with the environment, interaction forces are measured, fed back to the operator, and displayed through the exoskeletons. Since human hand and grippers have different kinematic structures, appropriate mappings for forces and motion between the fingers and the gripper are required.

Two different types of robotic grippers were used in the experiments: a BarrettHand and a two finger parallel gripper (PG70, Schunk). For the BarrettHand the implemented position and force mappings are reported in Peer et al. (2008). For the parallel gripper the distance between index finger and thumb is mapped to the distance of the two fingers of the gripper. The gripping force measured at the teleoperator site is directly fed back to the operator site and applied at both index finger and thumb.

As different human operators have different hand sizes, an appropriate calibration algorithm is needed for the data glove
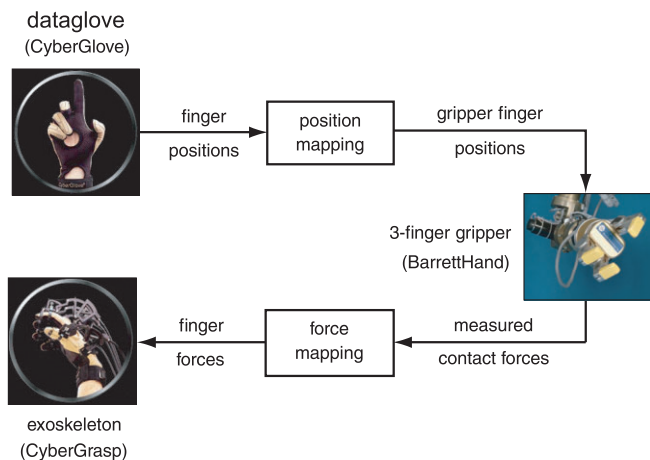
Fig. 10. System architecture of the multi-fingered telemanipulation system.

system. Since the software provided by the manufacturer assumes constant finger link lengths among different operators and does not take into account effects resulting from missing sensors as well as coupling effects, we developed a new calibration method. We introduced a virtual sensor that compensates for the missing ones and selected a hand model that takes into account the coupling effects. The unknown parameters such as gains, finger link lengths, and joint angle offsets were finally determined by an optimization algorithm.

7. Locomotion

Many applications of telepresence systems require the operator to travel considerable distances in the remote environment, e.g. when exploring the remote environment. To meet this demand, the teleoperator must be able to move around in the remote environment, which is achieved by placing it on a mobile platform. In addition, the human–system interface must be endowed with appropriate devices to control the movement of the teleoperator. In this context, it is important to maintain the natural sense of direction of the human operator. If the teleoperator is controlled by a pedal, the sense of direction gets lost as shown in Bakker et al. (1999). A more natural way of controlling the movement of the teleoperator is by replicating the movement of the operator themselves. This requires the human–system interface to be mobile as well such that it can follow the operator.

The two main concepts of haptic interfaces for large-scale environments are worn-type and grounded-type haptic interfaces. In the former case, the operator carries the whole interface, which is designed as an exoskeleton. In this way, unrestricted motion is possible, but the operation is very fatiguing because the operator must support the full weight of the system. In the latter case, the haptic interface is mounted on a mobile platform, which actively follows the motion of the human operator. The advantage is that the weight of the system is carried by the mobile platform; in some cases, however, the motions of the mobile platform may be noticeable to the operator.

7.1. Mobile Human–System Interface

Two haptic interfaces of type ViSHARD7 (see Section 6) are mounted on a mobile platform resulting in one overall MHSI. A given pose of the two end-effectors in world coordinates does not directly result in a given pose of the MHSI, i.e. the overall system is redundant. An optimal pose for the MHSI must be determined. The optimization is performed in such a way that the minimum velocity manipulability of the two end-effectors is maximized. In order to simplify the optimization problem, only the planar DOF are considered. Also, the planar DOF are decoupled in the kinematics of the haptic interfaces, i.e. the planar position only depends on the first two rotational joints.

For the given kinematic structure, ideal manipulability is given if both end-effectors are located on circles with fixed radius r_{opt} around the shoulder joints. In addition, the first rotational joint of each interface should be controlled in such a way that the minimum distance to their joint limits is maximized, i.e. the two joint angles must be equal.

The resulting configuration is symmetric and the corresponding platform position can be obtained by simple geometric calculations: the mobile platform must be aligned parallel to the connecting line from x_L to x_R and its center point must lie on the perpendicular bisector of the connecting line (see Figure 11). The distance n_{opt} can be computed using

$$n_{\mathrm{opt}} = \sqrt{r_{\mathrm{opt}}^2 - \left(\frac{\|d\| - d_P}{2}\right)^2} + n_P \qquad (2)$$

where $d = x_R - x_L$ is the vector connecting the two end-effectors in world coordinates, r_{opt} is the radius of the circle with maximum manipulability, d_P is the distance of the first rotational joints of the haptic interfaces, and n_P is the distance of the platform center from the connecting line between the first rotational joints of the haptic interfaces.

For details on the hardware setup, a derivation of this solution, a more sophisticated approach incorporating the human arm workspace, and experimental results, see Unterhinninghofen et al. (2008).

7.2. Mobile Teleoperator

In contrast to the mobile haptic interface, the position and orientation of the mobile teleoperator is directly commanded by the operator, whose position and orientation is tracked and
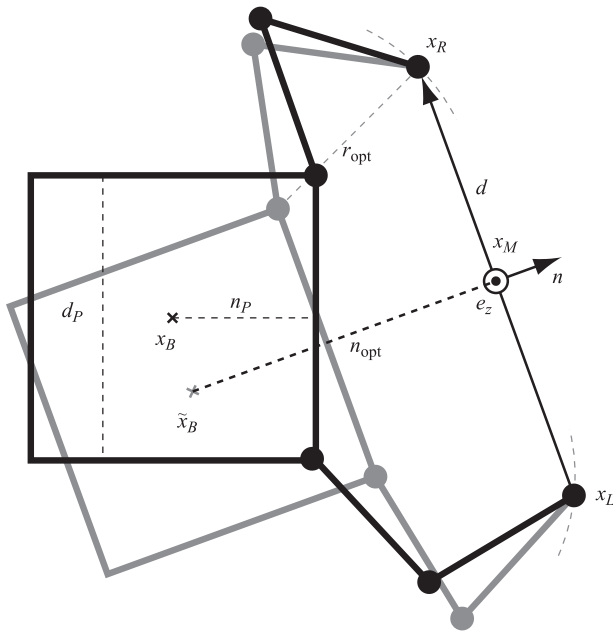
Fig. 11. Geometric solution for optimal positioning of the mobile platform.

Fig. 12. Schematic view of the scenario.

used as command input. As the velocity and acceleration capabilities of the mobile platform are inferior to those of a human, some delays in replicating the operator pose can occur. In order to minimize the effects of these delays, end-effector poses and head orientation are computed in world coordinates (see Figure 5) such that the robot arms and the robot head can compensate for the insufficient dynamics of the mobile platform.

8. Application Example

As a benchmark scenario for the overall system, a cooperative maintenance task is chosen. A broken pipe in a remote environment should be repaired using two mobile teleoperators (MTO1, MTO2). Each teleoperator is controlled by a human operator (OP1, OP2). To repair the pipe the two teleoperators must first move to the maintenance site. The broken pipe is then fixed by fastening a clamp at the breakage. A schematic view of the scenario, where several important positions are marked, is depicted in Figure 12. The starting positions for MTO1 and MTO2 are marked by A and B. MTO1 is controlled by OP1 using the SHSI. MTO2 is controlled by OP2 using the MHSI.

First MTO1 is moved to the maintenance site. As the view through the stereo camera system is limited and the size of the teleoperator is not apparent to the operator, navigating through the narrow doorway would be difficult. Therefore, a 3D model of the passage and doorway is created by scanning the environment (see Section 5). To enhance the quality of the 3D model, the wallpapers on the passage (C) are textured. By choosing
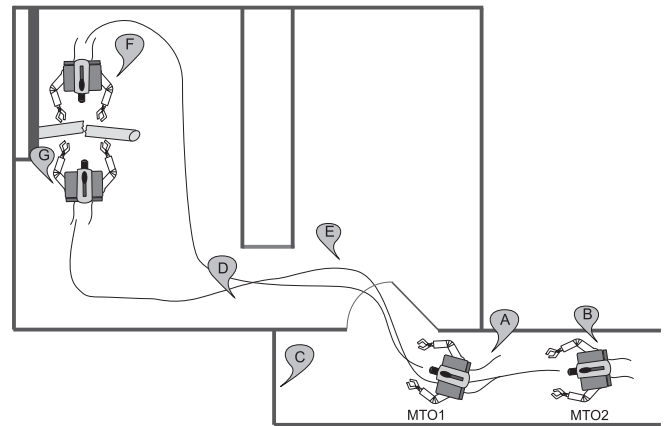
a third person view through the hybrid display, OP1 can then navigate through the door easily.

When the doorway has been passed the view is switched back to the stereo camera images, and MTO1 is moved to the position marked by D. Here, a 3D model of the maintenance site is created and rendered with the method presented in Section 5. Using the hybrid display, the maintenance site can be inspected from an arbitrary angle. Sudden changes of the environment can be determined immediately, as the live video feed is integrated into the rendered representation. Suitable goal positions can be determined. Here, OP1 decides to move to position F while they tell OP2 to move to position G. The two operators communicate using the audio system described in Section 4, i.e. sound is perceived as if originating at the correct location in 3D space.

By simply walking, OP2 controls the movement of MTO2 (E). Due to space restrictions, the operator side motions of OP2 can not be mapped directly to motions of MTO2. Instead, motion compression, a concept introduced in Nitzsche et al. (2004), is used to enable operation in arbitrarily large remote environments. As MTO2 is not equipped with a hybrid display, navigation through the door is difficult and a human at the remote site must guide OP2. As soon as OP2 reaches position G the actual maintenance task can be performed.

The actual maintenance task is solved slightly differently for different trials, i.e. strategies are communicated and then carried out. Here, OP2 grasps the two ends of the broken pipe at the breakage and moves them apart slightly. OP1 then slides the loose clamp over one of the ends. OP2 aligns the two ends of the pipe and OP1 positions the clamp over the breakage. To actually tighten the clamp, the two operators hold the pipe on either end of the clamp with their left grippers. With the right gripper MTO2 holds the clamp in a fixed position. OP1 then uses a hex key in their right gripper to fasten the three screws of the clamp. After the clamp has been tightened the two operators let go of the pipe and clamp, and move away from the maintenance site.

Fig. 13. Photo story of the benchmark scenario.

During the manipulation phase, both OP1 and OP2 naturally use the grippers of the respective teleoperators, as described in section 6.2. Bimanual manipulation with high interaction forces, as described in Section 6.1, is performed. Although closed kinematic chains are formed, and the two teleoperators directly interact over the pipes and (very stiff) clamp, no instabilities are observed.

A photostory illustrating the complete task is shown in Figure 13. Please refer to the multimedia Extension 1 for further insights.

While the actual maintenance task was performed, a high degree of immersion could be observed. For example, when talking to each other, an operator turned their head to look at the other teleoperator. Manipulation seemed natural, i.e. two people would probably have performed the task in a similar way when not using a telepresence and teleaction system, if their hands were constrained to the same capabilities of the teleoperator grippers.

9. Results

In the following, brief results for the different subsystems are presented.

Audio System

Compared with state-of-the-art techniques, our RSS interpolation method allowed fast and very precise reconstruction of HRTFs, and ensured a high-fidelity virtual auditory space over the headphones of both operators.

To verify the performance of the rational interpolation technique, we used 72 Kemar HRTFs measured $5°$ apart for the left ear in the horizontal plane. Additionally, we used 56 Kemar HRTFs measured $6.43°$ apart in the two $±40°$ elevation planes. We then took each plane at a time and removed one of the HRTFs. The now-missing HRTF was reinserted by the above-mentioned interpolation technique. The interpolation result was compared with the corresponding available measurements and the signal-to-distortion ratio (SDR) was computed according to the equation

$$SDR = 10 \log \frac{\sum_{n=0}^{N_h-1} h^2(n)}{\sum_{n=0}^{N_h-1} [h(n) - \hat{h}(n)]^2}, \tag{3}$$

where $h(n)$ denotes a measured HRTF at a certain azimuth for the left ear, $\hat{h}(n)$ denotes an interpolated HRTF at the same azimuth, and N_h is the HRTF length, e.g. $N_h = 512$.

Figure 14 shows the interpolation SDR for the Kemar HRTFs taken on the horizontal plane as well as on the
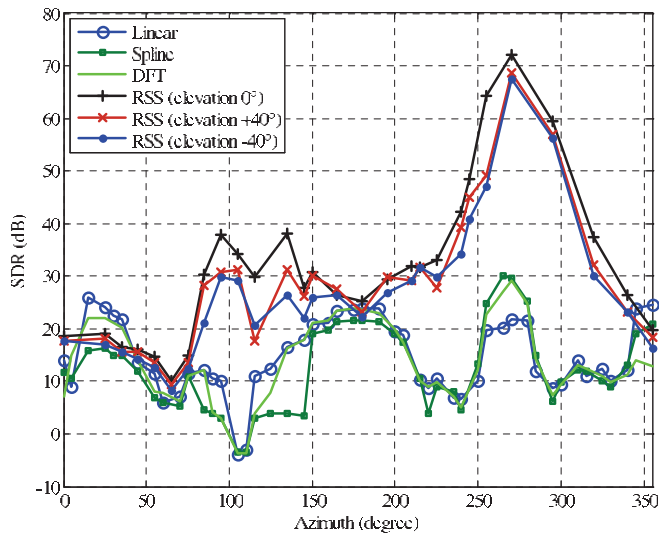
Fig. 14. Interpolation accuracy of state-of-the-art methods compared to the RSS technique for three elevation planes: horizontal (elevation 0°) plane, and ±40° planes.

two ±40° elevation planes. Compared with bilinear, discrete fourier transform (DFT), and spline interpolation methods using time correction, the RSS interpolation shows comparable performance from 0° to 45° azimuths and from 345° to 360°; this is due to the fact that in these azimuthal intervals the sound source is in the contralateral position with respect to the left ear and therefore diffraction effects, caused by the head and pinnae, are added to the transfer functions. Over all the remaining azimuth range, the SDR for the rational method is higher than that for the other three methods.

Video System

The presented approach allows rendering photorealistic views of the teleoperator's surroundings from arbitrary viewing positions. Changes in the scene observed by both color and geometry sensors are displayed as well. Figure 15 shows two examples which were generated during the approach phase of the teleoperation described in Section 8. Note that the areas within the borders show a live camera stream while geometry within coverage of the geometry sensor is updated perpetually.

The operator is able to choose an arbitrary perspective to observe the remote environment. In the presented application, the third-person view assists in passing narrow openings. This is especially helpful at the doorway while approaching the maintenance site, see Figure 16(a). Similar conclusions were also drawn by Salamin et al. (2006), who showed that third-person perspectives can enhance evaluation of distances.

In addition, the rendered representation can help the operator to inspect the teleoperation site in detail before the teleoperator has physically entered the teleoperation site. Critical

passages can be detected earlier and, therefore, they can be avoided.

By using different types of sensors for scene generation, we are able to provide geometry information in places where the teleoperator's surroundings contain much structure by the use of the proposed stereo system and by using the PMD sensor if stereo reconstruction is not feasible. Both methods are compared in Figure 16(b).

Haptic System

For applications similar to the presented example, a haptic system with high force/torque-output capabilities is necessary for unscaled interaction with the remote environment. The performance of the presented system was assessed for the case of free space motion and haptic exploration of soft and stiff materials. Figure 17 shows good position and force tracking results in the direction of the z-axis. The inertia and damping, which are rendered by the master and the slave device, are indicated in Figure 9. Thus, robust stability of the overall telemanipulation system is guaranteed. Interested readers are referred to Peer and Buss (2008b) for a detailed analysis of different control architectures and their stability analysis.

Locomotion

To allow manipulation in an unrestricted workspace two different types of interfaces were developed. In the case of the MHSI two haptic interfaces are mounted on a mobile base allowing control of the mobile teleoperator by simply walking.

In the ideal case, the human operator does not notice the repositioning of the mobile platform at the end-effectors, because the movements of the mobile platform should be exactly compensated by opposite movements of the haptic interfaces. However, due to some structural elasticity, sensor noise, and resulting compensation errors, some unwanted forces at the end-effectors result, which can be felt by the operator. The magnitude of these disturbances was experimentally assessed by making the platform move according to a sinusoidal reference trajectory while the operator was holding the end-effectors. The recorded forces at the end-effector are shown in Figure 18(a) and the estimated positions of platform and end-effectors are shown in Figure 18(b). While some of the high-frequency components of the force errors can be attributed to the trembling of the operator, the remainder is created by the mobile haptic interface. The force errors caused by repositioning the mobile base are in the range of ±3 N. These force errors are relatively high compared to other implementations which benefit from the use of impedance-type haptic interfaces. As shown by Formaglio et al. (2008), the force tracking performance of impedance-type haptic interfaces is maintained when they are integrated in a mobile haptic interface.

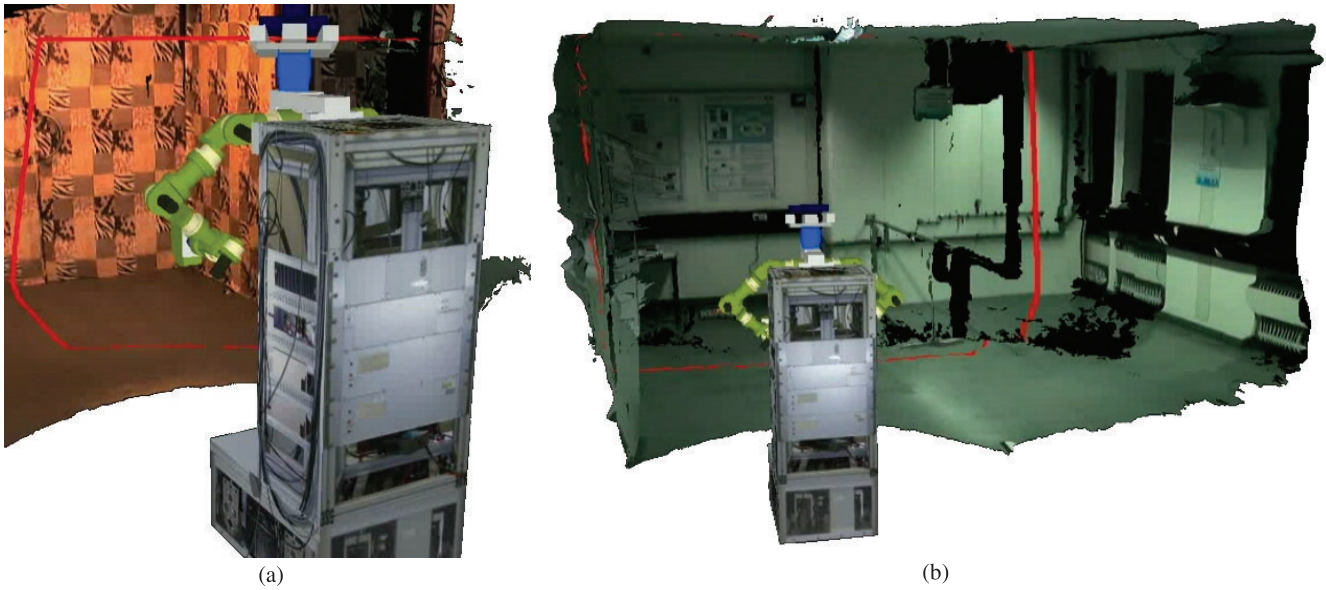(a)   (b)

Fig. 15. Two examples for computer-generated views using the described method. The teleoperator in the foreground is modeled *a priori* and rendered using conventional texture mapping. The model is parametrized using the current arm poses while the position is based on the odometry data.

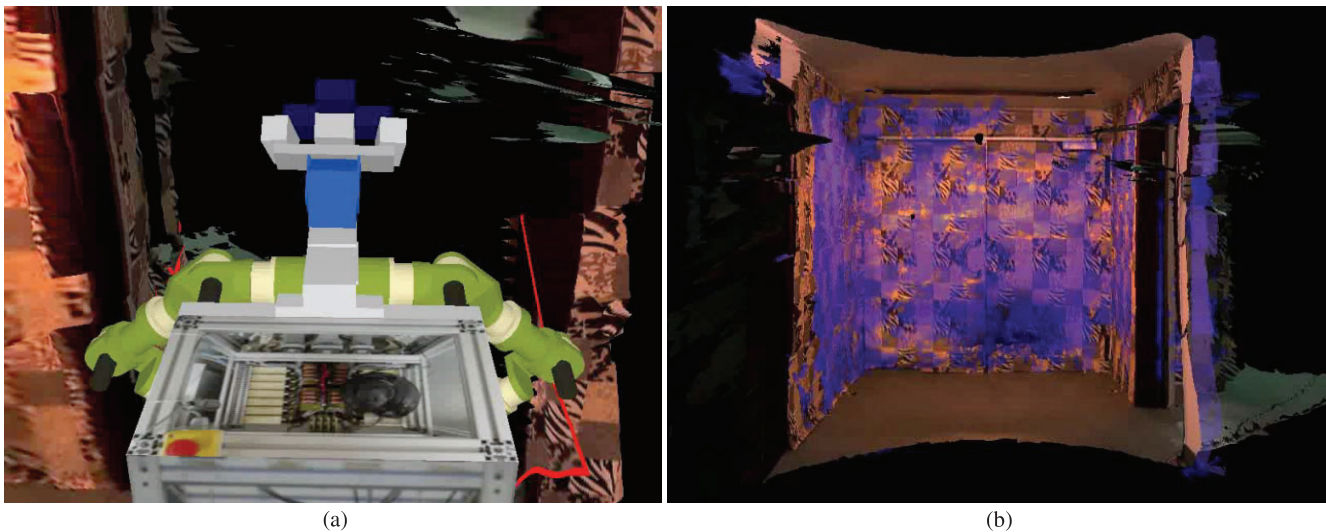(a)   (b)

Fig. 16. Robot before passing the narrow door (a). Scan of the start of the presented application. What can be seen is the output of the proposed stereo system (shaded) overlapped with that of the PMD camera (b).

For admittance-type interfaces, the force errors can be reduced by employing accelerometers in order to measure and to compensate position disturbances induced by the repositioning of the mobile base.

10. Conclusion

A multi-modal multi-user telepresence and teleaction system was presented in this work. The system was specifically developed to enable collaboration of multiple operators in a shared remote environment. The three implemented modalities, haptics, vision, and acoustics, were designed to support this goal
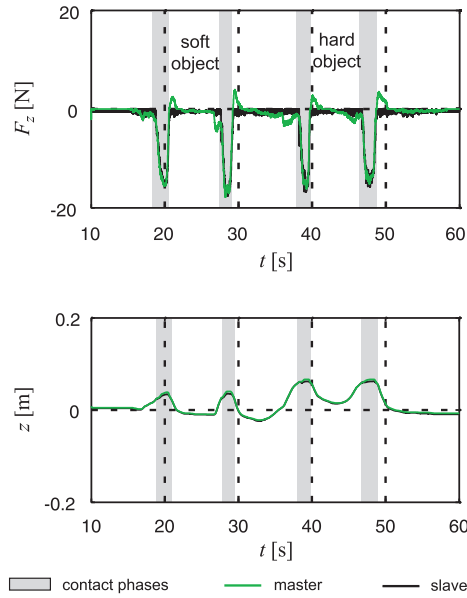
Fig. 17. Force and position tracking performance in the z-axis for the position-based admittance control with force–force exchange ($M_p = 6$ kg, $B_p = 70$ N m/s, $C_p = 0$ N/m, $M_o = 0.2$ kg m^2, $B_o = 1$ N m s/rad, $C_o = 0$ N m/rad). The index p refers to (1) and the index o to the corresponding rotational part.
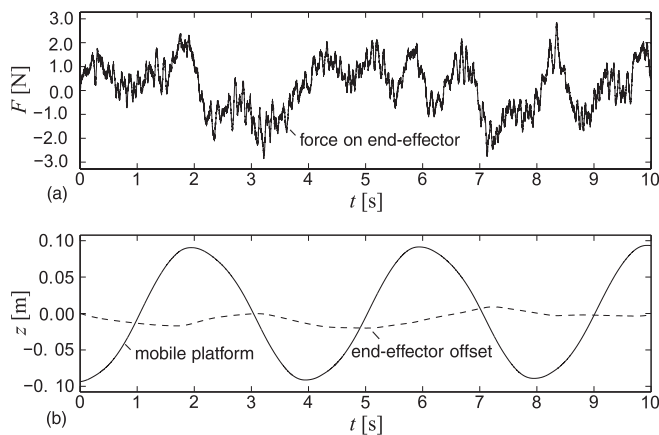

Fig. 18. Force on end-effector during repositioning of the mobile platform: (a) measured force; (b) position of the mobile platform and offset of the end-effector from the starting point.

by providing stable haptic interaction between operators, augmented views of the remote scene, and spatially correct voice signals. Thereby, the overall system creates a common HVA workspace for two operators.

The haptic system uses admittance-type devices to enable stable interaction, high interaction forces, and dextrous manip-

ulation in six DOF. The human-sized workspace of the haptic interfaces and telemanipulators is extended by mounting them on mobile platforms. The video system generates and renders a 3D model of the remote environment in real time, thereby enabling a larger field of view from an arbitrary point of view. The audio system utilizes high-fidelity interpolation techniques to render the speech of the collaborating operator at the correct position in 3D space. These measures improve the interaction between the two operators.

The system is used to perform a benchmark scenario: a remote maintenance task, in which the goal is to move to a target area, locate a broken tube, and repair the breakage. Due to the outstanding features of each subsystem, the task is achieved and a more natural interaction is attained.

Appendix: Index to Multimedia Extensions

The multimedia extension page is found at http://www.ijrr.org.

Table of Multimedia Extensions

Extension	Type	Description
1	Video	The course of the application example, visualizing some properties of the system that are hard to convey in writing.

Acknowledgment

References

Ackermann, J. (2002). *Robust Control, The Parameter Space Approach*, 2nd edn. London, Springer.

Backes, P. G. and Tso, K. S. (1990). UMI: An interactive supervisory and shared control system for telerobotics. In *IEEE International Conference on Robotics and Automation*, IEEE, Cincinatti, Ohio, pp. 1096–1101.

Bakker, N., Werkhoven, P., and Passenier, P. (1999). The effect of proprioceptive and visual feedback on geographical orientation in virtual environments. *Presence: Teleoperators and Virtual Environments*, **8**: 36–53.

Begault, D. R. (1994). *3D Sound for Virtual Reality and Multimedia*. Cambridge, MA, Academic Press, chapter 4, pp. 132–136.

Blauert, J. (1997). An introduction to binaural technology. In *Binaural and Spatial Hearing*, Gilkey, R. and Anderson, T. (eds). Hilldale, NJ, Lawrence Erlbaum, pp. 593–609.

Burkert, T. (2006). Hardware-Beschleunigte Textur-Extraktion für ein Fotorealistisches Prädiktives Display. *Ph.D. Thesis*, Technische Universität München, Lehrstuhl Für Realzeit-Computersysteme.

Cavusoglu, M. C., Williams, W., Tendick, F., and Sastry, S. S. (2003). Robotics for telesurgery: second generation Berkeley/UCSF laparoscopic telesurgical workstation and looking towards the future applications. *Industrial Robot (Special Issue on Medical Robotics)*, **30**(1): 22–29.

Chang, S., Kim, J., Borm, J. H., and Lee, C. (1999). KIST teleoperation system for humanoid robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Kyongju, Korea, pp. 1198–1203.

Chong, N. Y., Kawabata, S., Ohba, K., Kotoku, T., Komoriya, K., Takase, K., and Tanie, K. (2002). Multioperator teleoperation of multirobot systems with time-delay: Part 1 Aids for collision-free control. *Presence: Teleoperators and Virtual Environments*, **11**(3): 277–291.

Debevec, P., Yu, Y., and Borshukov, G. (1998). Efficient view-dependent image-based rendering with projective texture-mapping. In *9th Eurographics Rendering Workshop*, Vienna, Austria, pp. 105–116.

Formaglio, A., Prattichizzo, D., Barbagli, F., and Giannitrapani, A. (2008). Dynamic performance of mobile haptic interfaces. *IEEE Transactions on Robotics and Automation*, **24**(3): 559–575.

Hannaford, B. and Okamura, A. M. (2008). Haptics. In *Springer Handbook of Robotics*, Siciliano, B. and Khatib, O.(eds). Berlin, Springer, chapter 30, pp. 719–739.

Hasunuma, H., Kagaya, H., Takatori, M., Fujimori, J., Mifune, F., Shikoda, S., Kobayasi, M., Itoko, T., and Tachi, S. (1999). Development of teleoperation master system with a kinesthetic sensation of presence. In *International Conference on Artificial Reality and Telexistence*, Waseda, Japan.

Hayati, S., Lee, T., Tso, K., and Backes, P. (1990). A testbed for a unified teleoperated-autonomous dual-arm robotic system. *IEEE Transactions on Robotics and Automation*, **2**: 1090–1095.

Held, R. and Durlach, N. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments*, **1**(1): 109–112.

Hollerbach, J. M. (2002). Locomotion interfaces. In *Handbook of Virtual Environments: Design, Implementation, and Applications*, Stanney, K. (ed). Hilldale, NJ, Lawrence Erlbaum, pp. 239–254.

Immersion Corp (2007). CyberForce Tactile Feedback System. http://www.immersion.com/3d/products/cyber_force.php.

Kanade, T., Rander, P., and Narayanan, P. (1997). Virtualized reality: constructing virtual worlds from real scenes. *Multimedia, IEEE*, **4**(1): 34–47.

Keyrouz, F. and Diepold, K. (2006). Efficient state-space rational interpolation of HRTFs. In *AES 28th International Conference: The Future of Audio Technology–Surround and Beyond*, Piteå, Sweden, pp. 185–189.

Kron, A. and Schmidt, G. (2003). A bimanual haptic telepresence system design issues and experimental results. In *International Workshop on High-Fidelity Telepresence and Teleaction jointly with the Conference HUMANOIDS*, Karlsruhe, Munich, Germany.

Minaar, P., Plogsties, J., and Christensen, F. (2005). Directional resolution of head-related transfer functions required in binaural synthesis. *Journal of the Audio Engineering Society*, **53**(10): 919–929.

Naef, M., Lamboray, E., Staadt, O., and Gross, M. (2003). The Blue-c distributed scene graph. In *Proceedings of the IPT/EGVE Workshop*, ACM, Zurich, Switzerland.

Nash, E., Edwards, G., Thompson, J., and Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of Human–Computer Interaction*, **12**(1): 1–41.

Nitzsche, N., Hanebeck, U., and Schmidt, G. (2004). Motion compression for telepresent walking in large target environments. *Presence: Teleoperators and Virtual Environments*, **13**(1): 44–60.

Ogawa, T., Kiyokawa, K., and Takemura, H. (2005). A hybrid image-based and model-based telepresence system using two-pass video projection onto a 3d scene model. In *ISMAR '05: Proceedings of the 4th IEEE/ACM International Symposium on Mixed and Augmented Reality*. Washington, DC, IEEE Computer Society, pp. 202–203.

Ortmaier, T., Weiss, H., and Hirzinger, G. (2004). Telepresence and teleaction in minimally invasive surgery. In *Proceedings of Robotik 2004*. VDI-Verlag.

Peer, A. and Buss, M. (2008a). A new admittance type haptic interface for bimanual manipulations. *IEEE/ASME Transactions on Mechatronics*, **13**(4): 416–428.

Peer, A. and Buss, M. (2008b). Robust stability analysis of a bilateral teleoperation system using the parameter space approach. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, Nice, France.

Peer, A., Einenkel, S., and Buss, M. (2008). Multi-fingered telemanipulation mapping of a human hand to a three finger gripper. In *International Symposium on Robot and Human Interactive Communication*, pp. 465–470.

Reinhart, G., Clarke, S., Petzold, B., and Schilp, J. (2004). Telepresence as a solution to manual micro-assembly. *CIRP Annals*, **53**(1): 21–24.

Reintsema, D., Landzettel, K., and Hirzinger, G. (2007). DLR's advanced telerobotic concepts and experiments for on-orbit servicing. In *Advances in Telerobotics: Human System Interfaces, Control, and Applications (STAR series)*, Ferre, M., Buss, M., Aracil, R., Melchiorri, C. and Balaguer, C. (eds). Berlin, Springer.

Ridao, P., Carreras, M., Hernandez, E., and Palomeras, N. (2007). Underwater telerobotics for collaborative research. In *Advances in Telerobotics: Human System Interfaces,*

Control, and Applications (STAR series), Ferre, M., Buss, M., Aracil, R., Melchiorri, C. and Balaguer, C. (eds). Berlin, Springer, pp. 347–359.

Salamin, P., F. V., and Thalmann, D. (2006). The benefits of third-person perspective in virtual and augmented reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '06)*, pp. 27–30.

Sarkis, M. and Diepold, K. (2008a). Sparse stereo matching using belief propagation. In *IEEE International Conference on Image Processing*.

Sarkis, M. and Diepold, K. (2008b). Towards real-time stereo using non-uniform image sampling and sparse dynamic programming. In *International Symposium on 3D Data Processing, Visualization and Transmission*, Chapel Hill, North Carolina.

Sarkis, M. and Diepold, K. (2009). Content adaptive mesh representation of images using binary space partitions. *IEEE Transactions on Image Processing*, **18**(5): 1069–1079.

Schaeffer, B., Brinkmann, P., Francis, G., Goudeseune, C., Crowell, J., and Kaczmarski, H. (2007). Myriad: scalable VR via peer-to-peer connectivity, PC clustering, and transient inconsistency. *Computer Animation and Virtual Worlds*, **18**(1): 1–17.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, **47**(1): 7–42.

Schiele, A. and Visentin, G. (2003). The ESA human arm exoskeleton for space robotics. In *International Symposium on Artificial Intelligence, Robotics and Automation in Space*.

Segal, M., Korobkin, C., van Widenfelt, R., Foran, J., and Haeberli, P. (1992). Fast shadows and lighting effects using texture mapping. *SIGGRAPH*, **26**(2): 249–252.

Sitti, M., Aruk, B., Shintani, H., and Hashimoto, H. (2003). Scaled teleoperation system for nano-scale interaction and manipulation. *Advanced Robotics*, **17**(3): 275–291.

Stanczyk, B. and Buss, M. (2004). Development of a telerobotic system for exploration of hazardous environments. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2532–2537.

Tachi, S., Komoriya, K., Sawada, K., Nishiyama, T., Itoko, T., Kobayashi, M., and Inoue, K. (2003). Teleexistence cockpit for humanoid robot control. *Advanced Robotics*, **17**(3): 199–217.

Ueberle, M. (2006). Design, control, and evaluation of a family of kinesthetic haptic interfaces. *Ph.D. Thesis*, Technische Universität München, Lehrstuhl für Steuerungs- und Regelungstechnik.

Unterhinninghofen, U., Schauß, T., and Buss, M. (2008). Control of a mobile haptic interface. In *IEEE International Conference on Robotics and Automation*, pp. 2085–2090.

Yoon, W. K., Tsumaki, Y., and Uchiyama, M. (2000). An experimental teleoperation system for dual-arm space robotics. *Journal of Robotics and Mechatronics*, **12**(4): 378–383.