Eurographics Italian Chapter Conference (2007)
Raffaele De Amicis and Giuseppe Conti (Editors)

Mobile Presence - a Computer Graphics Utopia?

Andre Stork[1] and Daniel Danch[1,2]

[1]Fraunhofer IGD Darmstadt, Technical University Darmstadt, Germany
[2]Bauhaus-University Weimar, Germany

Abstract

In this paper we set out a future scenario of mobile presence. We will define our understanding of the term mobile presence. We will derive the requirements imposed by mobile presence with respect to computer graphics and computer graphics research. Discussing the state of the art in various fields, we like to identify and stimulate approaches and areas of research relevant to mobile presence now and in the future.

Categories and Subject Descriptors (according to ACM CCS): H.5.1 [Information Systems]: Artificial, augmented, and virtual realities

1. Introduction

In computer graphics and especially in virtual reality the term presence was introduced to express the sensation of *being there*. Presence is not strictly defined. Presence goes beyond pure immersion (being located within a virtual scene) and interactivity - two key characteristics of virtual reality. Presence is achieved if a user perceives a virtual environment - through his or her different senses - as if it was real. Presence is given if he or she has the impression of *really* being there. Thus, presence is a perceptional effect. Its degree depends on the visual fidelity of the scene, the realistic behavior of obstacles in the scene and the kind of feedback provided to the user. Certainly, presence in VR is not yet fully understood and presence is an active multi-disciplinary research field (`http:www.cs.ucl.ac.uk/research/vr/Projects/Presencia/`).

Moving from a controllable virtual reality setup to a dynamically changing (uncontrollable) mobile mixed reality setup - going from presence to *mobile presence* - has major implications not only on the devices being used and device development but also onto the appropriateness of approaches (algorithms) and performance requirements.

In the remainder of the paper we will describe our vision of *mobile presence*. We will identify the requirements and research areas involved. Reviewing and assessing the state of the art, we like to stimulate various areas of research relevant to *mobile presence*.

1.1. The Plot

Imagine you as a mobile user are entering an unknown physical environment and you want to interact with real props in the scene. You want to create or change virtual props in the scene and you want to interact with virtual and real objects in a similar manner. In addition, the scene can be populated with other local or remote mobile users - subjects.

Of course interaction with objects and subjects should be natural. Objects can interact with each other showing phys-

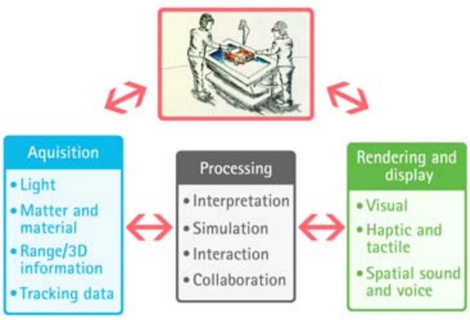

Figure 1: *CG research fields in mobile presence structured into three stages following an input, processing, output principle*

ically correct behavior. The scene should be visualized correctly taking into account the current lighting conditions. Note that real world lighting conditions are changing dynamically. Light matter interaction between real and virtual objects and mutual occlusion should be handled properly.

2. Mobile presence requirements

Taking into account that in *mobile presence* we want to have:

- correct visualization / rendering
- natural interaction with objects and subjects
- physically-based behavior

all in a mobile fashion, we can identify the below stated research areas and we end up with a *tour de force* in computer graphics when trying to achieve *mobile presence*.

We have grouped these research topics into three stages: acquisition, processing and rendering. In each stage both the development of improved, more sophisticated and/or dedicated hardware and software is needed as we will see in the remainder of the paper. First, we will look into the different stages in more detail.

2.1. Stage 1: Acquisition

Driven by the fact that we want to have correct rendering of the mixed reality scene, we need to acquire:

- the lighting situation (dynamically)
- material properties of objects existing in the scene

This will allow for augmented reality rendering where virtual objects are correctly lit in the environment and light matter interaction between real and virtual objects can be calculated. Also, for the correct display of mutual occlusions it is needed to acquire range information and derive 3D reconstruction enabling object and subject recognition in a later stage. Thirdly, we need to track the mobile user itself, the position, pose and gesture of the full body or at least its extremities.

As we will see in the state of the art review, various technologies (software and devices) exist, many times not suited for *mobile presence* where acquisition has to happen *on-the-fly* without bulky devices. In contrast to most of today's approaches, in case of *mobile presence* acquisition needs to be done from the user's perspective. The equipment has to be miniaturized and user worn. Acquisition has to be done dynamically in time. Processing need to be done real time.

2.2. Stage 2: Processing

Stage 2 is kind of diverse since here most of the processing of the inputs, the application logic and the control of the output resides.

To enable meaningful interaction, a *mobile presence* system needs to build up an understanding of the scene and the actions taken by the different players. Although there are decades of research in image recognition and image understanding, the interpretation of the scene data is the research challenge in stage 2, esp. considering the potential limitations in computing resources in a mobile setup. Interpretation of interactions, physically based simulation of objects' behavior and support for collaboration of multiple users need also to be addressed/handled in the processing stage. Interpretation of interactions heavily depends on the application logic.

Physically-based simulation of behavior is possible once an object is recognized and a behavioral model can be associated to the object - still not a trivial task. Execution of the behavior model in realtime might also be an issue based on the complexity of the behavior model. For seamless collaboration of different mobile users, platforms and standards are needed that support ad hoc networking.

2.3. Stage 3: Rendering

Research in multi-modal rendering addressing all senses of human beings (sight, hearing, touch, smell and taste) has a long tradition in VR [FFM*04]. For communication and interaction, the most used senses are sight, hearing and touch. For verbal communication within groups of human beings spatial hints are important carriers of information. Thus, in stage 3 we focus on rendering of: visual, haptic / tactile, spatial speech and sound information.

For visual rendering algorithms are needed to photo-realistically render in real time - maximum visual fidelity in real time - and display technologies (devices) are needed that are able to embed virtual objects into real scenes with correct occlusion effects between real and virtual objects and vice versa. The mobile scenario implies mobile display devices - or a technology that can display anywhere. For natural interaction and feedback, tactile and haptic rendering needs to be provided *on the move*. Again immobile installations of force-feedback devices are inappropriate for *mobile presence*, instead mobile haptic / tactile rendering is required. For speech and sound rendering, spatial information about the sound source has to be acquired in stage 1 and correctly displayed in stage 3 utilizing software and hardware appropriate.

3. State of the art review and assessment

In this section we review and assess the state of the art with respect to the before mentioned requirements of *mobile presence*. We focus on stage 1 *acquisition* and stage 3 *rendering*. The talk will elaborate on stage 2 in more detail.

3.1. Stage 1: Acquisition of appearance, shape and pose information

Most of today's light acquisition techniques rely on light probes and high dynamic range images [DRW*06]. Digital cameras combined with mirrored balls are used to capture omnidirectional images of the real environment which contain the illumination information. To simplify the capturing process significantly, fish-eye cameras and single- or multi-viewpoint catadioptrical systems/cameras can be used. However, the dynamic range of light intensity in real scenes exceeds the capabilities of CCD sensors in standard cameras. Since reproduction of the real dynamic range is essential for scene fidelity, the dynamics is captured by taking multiple images of the scene with varying exposure times - this process is automated by HDRI cameras. A common approach, is to change the exposure sequentially over time. Other approaches use mosaicing with spatially varying filters or utilize multiple image detectors. Nayar et al. trade resolution for multiple exposure by suggesting detectors with a generalized Bayer Grid layout.

Measuring the appearance of real objects is an immanent prerequisite to simulate the interaction between real and virtual objects in a mixed reality scene. A 8D reflectance field or the closely related BSSRDF are the most general representations for the appearance of translucent materials. For less complex materials, lower dimensional versions of the BSSDRF exist. The limitation to homogeneous isotropic BRDFs results in a simple 4D representation. However, many real-world objects consist of heterogeneous materials with spatially varying BRDFs. Restricting to opaque materials and disregarding interreflectivity, one can reduce the number of parameters to six. Current appearance measuring approaches [LG05] can be classified by the way illumination is handled during the measurement process. The vast majority of techniques are based either on point light sources, directional light sources, or arbitrary light sources located within the scene; the latter category demands an expensive inverse rendering approach. In any case, the acquisition is time consuming since a complex and heavily controlled measurement environment and/or a demanding post-processing step is required. Image- and video-based techniques utilize the captured light field data of objects in order to render without the need of geometrical information. The huge amount of data implies the usage of elaborated compression techniques leading to additional processing steps.

In order to properly place virtual objects into a real environment and to resolve occlusions correctly, a 3D model of the scene or at least distance information from the current point of view to the surrounding real objects is needed. Automatic techniques for reconstructing geometry from multiple images or video sequences are still either not robust or generate inadequate triangulation results. Semi-automatic approaches yield better result's but are inappropriate for our dynamic mobile scenario. Alternatively, depth images can be generated using a laser scanner or a depth video camera. Laser scanners tend to generate huge amounts of point set data, whereas depth video cameras are still quite limited in resolution. When trying to combine individual range images, the problem is to automatically identify matching feature points. Projecting structured light and capturing the image of the light structure on a curved surface is another well-known technique for 3D reconstruction. It has been shown, that beamer can be used for displaying both scene content and structured light in a time-multiplexed manner. Again, composing a reconstructed 3D object out of the single relief images is not trivial, esp. in real-time - also the process of projecting different structures takes some fractions of a second. From today's perspective, depth video cameras seem most promising for *mobile presence*.

An indispensable need in mixed reality scenarios is the estimation of pose (the user's body, hands, head). Tracking allows the user to interact with the augmented world in a natural way. Mechanical devices show the best possible precision and minimum latency but significantly restrict the workspace. Most of them also provide force feedback, e.g. the well known Phantom (for haptic rendering see also section on stage 3). In recent years, magnetic tracking was - wherever possible - more and more replaced by optical tracking. Optical tracking is distinguished into marker-based and markerless approaches. Marker-based approaches can deliver very high accuracy but are not suited for *mobile presence*. Markerless approaches have made significant progress in the past years. Today, computationally still quite demanding, they are most appropriate for *mobile presence*. Model-based markerless approaches require a reference model for the tracking to be initialized.

3.2. Stage 3: Rendering

Visual rendering of virtual objects based on a simple local illumination model is not able to meet the requirements imposed by our understanding of *mobile presence*. More sophisticated global illumination techniques have to be applied to seamlessly blend virtual object into real scenes. However, traditional global illumination methods like ray and photon tracing or radiosity [WSB*03] are still quite time consuming and either require considerable preprocessing or high per frame computation costs. Moreover, adapting to dynamically changing environments is not well supported by these methods. With respect to this, recent developments such as Precomputed Radiance Transfer (PRT) is better suited for *mobile presence*. Introduced only in 2002, PRT has stimulated a whole new area in research. Initially designed just for static scenes in dynamically changing lighting environments, PRT has meanwhile been extended towards dynamic scenes and dynamic local light sources. Still - even if supporting dynamic environments - PRT is a compromise in visual fidelity and it is quite data intensive.

Today, for mobile mixed reality display technology provides video and optical see-through head mounted displays (HMDs). Video see-through combines an image of the real scene - updated with each frame - with virtual augmentation. The video capture reduces the 'quality' of the real world - resolution, color, dynamic range. The field of view to the real world is limited by the display. Occlusion can be handled quite easily. With optical see-through displays reality maintains mostly unaffected by the optics, the field of view onto the real scene can be large. To display occlusions correctly - to occlude real objects by virtual ones - light from the real world has to be blocked before it reached the user's eyes. Kiyokawa et al. address this problem by introducing additional LCD panels into the path of light. A special type of HMDs are retinal displays which utilize laser beams to 'write' images directly onto the user's retina. The view to the real world is largely unaffected. A way to handle mutual occlusion with retina displays has not yet been presented. OLEDs in HMDs have considerably reduced the amount of power consumption and enhanced their suitability for mobile use. Laser-based retina displays are not yet available for battery operation. Alternatively, projector-based systems utilizing user worn mini projectors could be used in the future to project images directly onto physical objects rather than on a virtual image plane. Projection based approaches are prone to radiometric distortion since the material of the background affects color presentation and can only be compensated within given limits. In combination with cameras even a geometric correction can be performed. To cope with extended color spaces and dynamic ranges is another future challenge on HMDs.

For interacting with real objects we are used to receive natural tactile and haptic feedback. Force feedback devices and collision detection algorithms to control haptic feedback devices are a classical research topic in VR. Most haptic feedback devices today consist of some mechanical parts that give low degrees of freedom feedback (typically less than 20 degrees) compared to the millions of receptors human beings have to detect various kinds of stimuli. There is only very little work on mobile force feedback devices [ZLKK06], which are also very limited in the possibilities they can display. All immobile force feedback devices are inappropriate for mobile presence. However, there is more and more work on unconventional user interface [BK04] which explores e.g. interaction through bio control. In parallel retina implants have been developed to help the visually impaired. It is our belief that mobile force feedback will only be convincingly possible through nervous cell stimulation.

Spatial sound is an important carrier of presence information. Through the ability to detect time differences in sound waves, we are able to precisely locate sound sources. We can hear where somebody is located before we see him/her. The accordance of visual spatial and aural spatial information contributes to the presence experience. Spatial sound processing, distributing spatial sound information and displaying spatial sound even via mobile ear phones can already be done today even if not in a highest sophisticated manner (wave field synthesis).

4. Summary

In this paper we introduced the vision of *mobile presence*. We discussed the requirements and implications of *mobile presence*. We reviewed the state of the art in relevant fields and reflected onto the before stated needs in *mobile presence*.

Looking into the state of the art and the many open issues - which we touch only partially - that arise in *mobile presence*, the vision still sounds more like a utopia. Taking into account the enourmous advancements achieved by computer graphics community over the last 40 years, one can easily recognize that many trends and developments are clearly hinting at presence and mobility. We are convinced that CG community will continue to develop towards *mobile presence*.

The talk will further elaborate on the topics presented in this short paper.

5. Acknowledgment

The authors want to acknowledge the EC project IMPROVE - Improving Display and Rendering Technology for Virtual Environments -, which partially contributed to related topics of this paper.

References

[BK04] BECKHAUS S., KRUIJFF E.: Unconventional human computer interfaces. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Course Notes* (New York, NY, USA, 2004), ACM Press, p. 18.

[DRW*06] DEBEVEC P., REINHARD E., WARD G., MYSZKOWSKI K., SEETZEN H., ZARGARPOUR H., MCTAGGART G., HESS D.: Introduction some course 5 presentations are available on the introduction citation page. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Courses* (New York, NY, USA, 2006), ACM Press, p. 1.

[FFM*04] FISHER B., FELS S., MACLEAN K., MUNZNER T., RENSINK R.: Seeing, hearing, and touching: putting it all together. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Course Notes* (New York, NY, USA, 2004), ACM Press, p. 8.

[LG05] LENSCH H. P. A., GOESELE M.: Realistic materials in computer graphics bibtex. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Course Notes* (New York, NY, USA, 2005), ACM Press.

[WSB*03] WALD I., SCHMITTLER J., BENTHIN C., SLUSALLEK P., PURCELL T. J.: Realtime ray tracing and its use for interactive global illumination. Montani C., Pueyo X., (Eds.), Eurographics Association. Eurographics 03 STAR.

[ZLKK06] ZHANG R., LOCHMATTER P., KUNZ A., KOVACS G.: Dielectric elastomer spring roll actuators for a portable force feedback device. *haptics 0* (2006), 53.