An Investigation into Virtual Representations of Real Places

Jens Arnspang[6], David Benyon[4], Manfred W. Fahle[5], Erik Granum[1], Claus B. Madsen[1], Tomas Pajdla[3], Shmuel Peleg[2], Michael Smyth[4], Phil Turner[4], Susan Turner[4] and Daphna Weinshall[2]

1 Aalborg Universitet, Aalborg – Denmark, Computer Vision and Media Technology Laboratory

2 The Hebrew University of Jerusalem, Israel, School of Computer Science and Engineering

3 Czech Technical University in Prague Centre for Machine Perception (CMP), Department of Cybernetics

4 Napier University, Edinburgh, School of Computing

5 Universität Bremen, Institute for Brain Research

6 Aalborg Universitet, Esbjerg – Denmark, Department of Computer Science and Engineering

Keywords: Image Based Rendering, Photorealism and Embodied Interaction.

Summary

The paper describes the main themes of a recently funded research project under the European Community's Future and Emerging Technologies 'Presence' initiative. The aim of the research is to develop new tools for empirical and theoretical studies of presence based on the concept of the observer's embodiment in the computationally created virtual environment. The approach will be based on true-to-life visual and auditory information presented in real-time. These will utilise a range of projection technologies from Head Mounted Displays to a six sided CAVE. The project will undertake a series of empirical investigations into the experience of 'being there' in relation to real places and objects. Analysis of presence will be at the level of psychological, physiological and neurological. At this early stage of the research four themes are presented: acquisition and real time rendering of real places; augmentation of such representations; psycho-physiological aspects of presence and establishing a sense of place and embodiment.

1 Introduction

BENOGO is a recently funded project under the European Community's Future and Emerging Technologies 'Presence' initiative. This paper introduces the project and its main research themes. The BENOGO consortium comprises 6 academic institutions from Europe and Israel with expertise in virtual reality, photo-realistic panoramic image acquisition and rendering, the psycho-physics of visual sensing and spatial perception, and the human aspects of new technologies. This project brings together a mixture of novel technologies that will enable real-time visualisation for an observer of recorded real places. The aim of the research is to develop new tools for empirical and theoretical studies of presence based on the concept of the observer's embodiment in the computationally created virtual environment. Furthermore, as real places (possibly known to the observer) with man-made and/or organic objects (like trees, foliage etc.) are otherwise hard to represent in a virtual environment, the objective is to bring about new insight into presence through comparison with the sense of presence experienced in the real world.

The BENOGO experience will be based on true-to-life visual and auditory information presented in real-time. The technology will be designed to support the observer's active exploration of the visual and auditory space, through the addition of a physical dimension to the experience. This physical dimension is paramount to achieving 'embodiment' (i.e. the observer's sensation of being bodily grounded in an environment). The new technology of *Image Based Rendering* does not require a reconstructed geometrical model of the scene (Buehler et al., 2001, Shum et al., 2000). It bypasses an important technological problem and presents a breakthrough, but large amount of image data needs to be stored, and recalled for real-time visualisation. Through visual and auditory augmentation a sense of life can be added including objects for interaction. Projection technologies will range from Head Mounted Display (HMD) to large screens including a 6-sided CAVE.

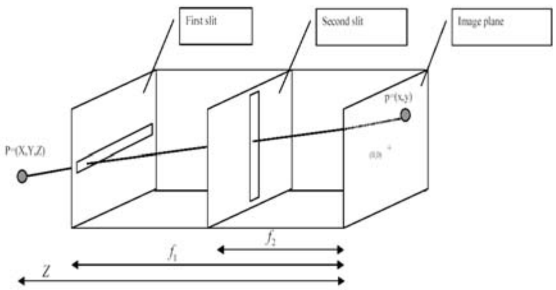
The empirical research will exploit the possibilities to investigate the experience of ‘being there’ in relation to real places and objects The theoretical framework will be based on the concept of embodiment in conjunction with presence and sense of place. This will be investigated in terms of fidelity of experience and presentation as compared to equivalent real-word places, and physiological and neurological aspects like consistency of sensory-motor co-ordination. The framework will be developed in close interaction with, and as a guide for, technical development by focusing on the particular strengths that the technology offers as well as on its weak points. Feedback from empirical studies will form an essential part of the project. The research iterates through 11 demonstrators to achieve these goals. In order to structure the research, four main themes have been identified.

2 Theme 1: Acquisition and real time rendering of real places

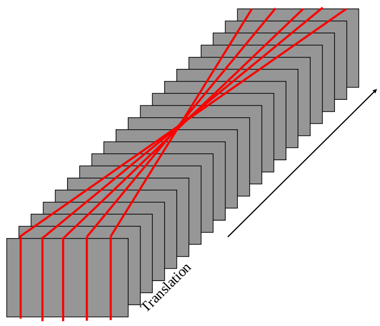
Currently there are two dominant approaches to the generation of new (synthetic) images: In the traditional approach images are generated from complete 3D models of objects and scenes. This approach uses a compact data representation and has the flexibility to generate any new view that may be required. Its main drawback is the reliance on the availability of models: many objects (such as hair and crop fields) and many physical phenomena (such as shadows) are very hard to model; as a result, synthetic images generated from models often appear artificial (like paintings) and lack a sense of realism.

Another approach to image synthesis is collectively called Image-Based-Rendering (IBR), (Buehler et al., 2001, Shum et al., 2000). In this approach a scene is photographed from essentially all possible points of view, and new images are generated through complicated computations, re-sampling and interpolation of this image collection. The main advantage of this approach is the realistic nature of the resulting images. The main drawbacks follow from the necessity to have in storage every possible point of view of every possible point in space. As a result the storage load of this approach is huge, and the image acquisition is very tedious. These drawbacks make image-base-rendering impractical for most applications.

A new approach to image-based-rendering is proposed that will be both practical and feasible. Essentially, the approach will trade-off fidelity to the perspective projection against the simplicity and speed of new view generation. More specifically, all image rendering methods today use the perspective projection, or try to approximate it. A new image-based rendering approach has been developed which is based on a new projection model – the **Two-Slit Camera**. The model is obtained by relaxing the strict requirement of perspective projection, namely, that all projection rays intersect at a single point (the focal-point). Instead, projection rays are required by definition to intersect two curves in space. The Two-Slit Camera model allows the development of image-based rendering techniques that can use much simpler image data, and have reduced storage and computation needs. For example, from an image sequence taken by a side-moving camera, new images can be generated in the plane defined by the line of motion and the camera's optical axis; in particular, we can generate a forward moving sequence by rotating the slicing plane of the **Space-Time Volume**. The Space-Time Volume is the image data bank needed for this approach to Image Based Rendering (Figure 1).



A Two-Slit camera. The horizontal and vertical slits define a unique ray for each scene point.



The Space-Time Volume: Stacked Images

Figure 1: A Conceptual Model of a Two-Slit Camera

Innovations regarding the project's Image Based Rendering (IBR) technique occur in two related areas: 1) the acquisition process for obtaining the images for the space-time volume, and 2) the real-time rendering, (i.e., the process of using the space-time volume for generating new views). It is the size and the construction of the space-time volume, coupled with the design of the rendering process, that define the volume within which images can be synthesized, (the volume within which the observer can visually explore the rendered scenario). The primary driving force behind the innovations will be to *extend* the volume within which it is practical to generate arbitrary new views.

Naive techniques for the acquisition of space-time volumes were used in computer graphics to build complete light-fields (Buehler et al., 2001, Shum et al., 2000). A typical approach uses a robotic arm to position a conventional camera with relatively narrow angle of view. Many images have to be acquired in order to obtain space-time volumes capturing wide view angle and the amount of data is unnecessarily high since many redundant images have to be acquired in order to register narrow fields of view. It is proposed to optimise the acquisition process by using wide-angle optics combined with carefully designed camera paths to reduce the number of acquisition images needed for a particular space-time volume. Not only will fewer images be sufficient but also less computation will be required as the number of images to register will decrease dramatically. New techniques of camera tracking and calibration using highly non-linear image projection will allow extending the existing technology beyond the realm of conventional cameras and will make the acquisition of large space-time volumes feasible.

Tightly coupled with the acquisition is the process of rendering (synthesizing views). The project's IBR technique is innovative in its very nature in that it uses alternative projection models to circumvent practical problems with conventional IBR, but the project will result in novel contributions. New techniques will be developed for re-sampling given space-time volumes in order to synthesize new views efficiently, and with optimal visual result; optimal in the sense that it best supports conveying a feeling of presence. To this end the project will investigate and characterize new and possibly specialized projection models, which in turn will enable more efficient space-time volume acquisition *and* rendering.

3 Theme 2: Augmentation

3.1 Video

The imagery synthesized by the project's IBR techniques will be visually augmented using an adaptation of a conventional computer graphics paradigm. That is, these augmentations will be rendered from 3D models. Being able to *seamlessly* blend computer graphics renderings of virtual objects with the IBR imagery is absolutely paramount in creating a credible visual augmentation. There are two primary reasons for visual augmentation. First, the IBR technology does not allow visualization of objects inside the spatial volume for which images can be synthesized. Secondly, augmentations are used to add dynamics (moving objects) to the otherwise still images.

Regarding visual augmentation the project will produce innovative results on two fronts: 1) methods for utilizing standard computer graphics hardware for rendering synthetic images based on non-perspective projection models, and 2) methods for extracting information concerning lighting conditions from images of real scenes and for applying this information in the computer graphics rendering of virtual objects.

Concerning the first topic the project will need to develop novel methods for rendering virtual objects based on the same projection model (the two-slit projection model) used by the IBR. Without such methods the augmented virtual objects may appear *geometrically* conspicuous merely because they have been projected differently. The project will develop novel techniques for taking advantage of state-of-the-art computer graphics hardware to produce two-slit projections of virtual objects. Currently, two quite different families of techniques are planned, and each will be evaluated in terms of advantages and disadvantages.

Concerning the second topic augmented virtual objects will be *lightingwise* conspicuous if they are not rendered using lighting which is consistent with that of the real scenes, because the IBR will reproduce real scene lighting accurately. Especially the virtual objects must cast credible shadows on the real scene. To this end the project will develop techniques for extracting essential geometrical and spectral lighting properties from the image collection used for IBR. Current state-of-the-art in this field requires a complete *a priori* 3D model of the scene, (Loscos et al., 2000). This project will develop techniques requiring little or no *a priori* scene knowledge, but take advantage of the large number of images available of the scene. This scientific contribution will be of general interest to the computer vision, computer graphics, and augmented reality communities, since the methods are independent of the project's special IBR techniques. This research will also involve establishing experimental results concerning how accurate virtual object lighting has to be in order not to be visually conspicuous to the human observers. For example, which is more conspicuous: geometrically or spectrally incorrect shadows?

3.2 Audio

Audio augmentations will also be made. Three major cases will be dealt with: 1) Ambient soundscapes, not necessarily with any 3D information of the scene, however spatially distributed in the display situation, in order to convey an aural impression of scene (e.g. office sounds, city sounds, ocean sounds, etc.). 2) Directional information, according to either observers’ ego motions (turning of my head etc.), or according to an introduced known object (the motor sound of my car, moving through the scene. 3) More fine tuned material information, influence of room parameters on the quality of the sound. For example, when clapping hands in the room or free field, the sound and the visual scene should be reasonable in mutual context, and not give divergent impressions.

For either of the three themes above, basic issues are: A) How much need in fact be known of the actual 3D geometry, in order to produce spatial sound, that actually augment the overall sensation of being there? B) To what extent can a fine tuned spatial sound repair a somewhat ‘fuzzy’ visual geometry (the visualizations are known not to be perspective, with varying degrees of degradation in the visual geometry)? C) Will moving sound objects produce an augmented presence, in general, and especially in the case of distorted image geometry?

All these basic questions on the sound augmentation, obviously involves technical analysis and synthesis subtasks, and they involve basis psychophysical experimental questions. Both such classes of questions will be addressed. In summary the innovations are in following areas:

- the production of spatial static sound, relating to images, without knowing the entire geometry of the scene in question;
- the combination of spatial sound information with spatial scenes visualised in a distorted perspective;
- the production of moving sound with only little knowledge of the scene geometry at hand, yet with a noticeable contribution to spatial understanding of the scene.

4 Theme 3: Psycho-physiological aspects of presence

The combination of photorealistic scenes with computer generated stimulus elements is a radically new way to perform psychophysical experiments that have been rarely used in the past. The planned experimentation with this type of stimuli will produce insights into the way in which synthetic objects or 'morphed' natural objects (such as actors) should be inserted into natural scenes. This will enable the quantification of parameters that impact on the impression of 'naturality' and should pave the road towards a much better understanding of which features are preferentially used by the visual system when analysing natural scenes and those that are less crucial. It is contended that one of these factors will be the minimal number of different frames required to produce naturalizing motion impressions, or more precisely, how often a moving object has to be 'up-dated' during motion at different speeds during the apparent motion produced by all (raster) displays. These results may - after much further research - lead to design rules for realistic synthesis of natural scenes, making best use of the hardware available based on the sensory physiology of humans and on their often surprising shortcomings in specific perceptual dimensions given their astonishingly good performance in others.

5 Theme 4: Establishing a sense of place and embodiment

At the outset the project will adopt a conceptual framework based on the idea of *embodied interaction* (or embodiment) that is proving to be an important new research area for designing interaction (e.g. Dourish, 2001). The approach will be to focus on patterns of embodied interaction, to develop incrementally the key features of three aspects of embodiment: sound, physical objects and sensory modalities (motion, light, temporal aspects); active movement and awareness (background, foreground, external); purpose and motivation, task complexity and social presence.

5.1 Phenomenology

Embodiment in itself is a development of the phenomenological school of philosophy. The philosopher Edmund Husserl developed the concept of phenomenology to describe the experience or awareness of things in a manner which did not reduce them to scientific data. For Husserl, an individual’s experience was the experience of *something*. By focusing attention on the act of this ‘experiencing of’ rather than on the thing being experienced or the person who was having the experience, he produced a new kind of knowledge. This knowledge could account for things beyond the reach of science. Husserl's student, Martin Heidegger, developed the ideas of phenomenology starting with the common-sense observation that the world exists - a condition he dubbed 'Being'. Beings are those entities which exist in the world and are able to reason about being. The focus of the use of a phenomenological approach should be *average-everydayness* (perhaps anticipating Garfinkel's observations concerning the status of the individual in ethnomethodology).

Continuing in the phenomenologist tradition, Merleau-Ponty's account of 'being-in-the-world' emphasises the importance of the body. He places the body at the centre of our relation to the world and argues that it is only through having bodies that we can truly experience space. In the context of perception Merleau-Ponty (1962) formulated a sense of sight as an embodied vision that is an incarnate part of the flesh of the world: 'our body is both an object among objects and that which sees and touches them'. In art and architecture many argue that this central role of the body is disappearing. The emphasis on the visual sense in Western culture has resulted in 'designs which housed the intellect and the eye, but that have left the body and the senses, as well as our memories and dreams, homeless' (Pallasmaa, 1996). The sense of 'aura', the authority of presence, that Walter Benjamin (1997) regards as a necessary quality of an authentic piece of art, has been lost.

5.2 Space, place, purpose and engagement

Within the physical environment what are the features that gives a space its sense of place? Has the built environment the potential to imbue a space with enough identity that it can become a place in our minds? The interaction between people, who populate a space, and the objects, which comprise it, can result in a variety of interpretations of place and the sense of presence. For example, the design of city centre parks may provide convenient lunchtime seating for office workers but as night falls may also provide unforeseen challenges for the city's skateboarders. While the 'found space' (Borden, 2001) remains the same for each group it is contended that the sense of place is quite different. In terms of presence the role of, and interaction with and through, the objects that populate both physical and virtual spaces and places will constitute a key research question which will be addressed within the project. An extension of this real-world focus will explore the issues of embodiment and sense of place in other media such as text and film. For example, the analysis of the roles of sight and sound in cinematic narrative and the consideration of immersion and interactivity in literature and virtual worlds offer useful starting points.

5.3 Empirical work

The empirical work will draw upon the theory and philosophy of embodiment. As for specific techniques, potential exists at the boundaries between phenomenology, personal constructs and perception and is expected to provide significant leverage. For example, personal constructs and their exploration through the technique of repertory grids will form one aspect of this work. While the initial uses of repertory grids were in the exploration of such things as attitudes and neuroses, they have also been used in knowledge elicitation and to some extent in information system and Human Computer Interaction (HCI) design. Other methods of elicitation such as language and content analysis, self-report questionnaires, observation and longitudinal studies will be utilised as appropriate, drawing upon previous work in this area and comparing the sense of presence and place as experienced in the physical world against that experienced in the virtual world. The output of the early stages of this part of the work will be scenarios to steer technical development; later work will focus on the iterative design and development of the BENOGO demonstrators.

5.4 Patterns of Embodied Interaction

At the time of writing the project is starting empirical research on a sense of presence . For example, we are currently analysing a pilot study concerning self-reports of experience in real-world places of the type to be captured in BENOGO. Very early results suggest that: (a) movement around a place and haptic interaction with aspects of it are very commonly remarked upon; (b) while sight is the dominant sensory modality, senses such as hearing and smell play a significant role in accounts attempting to evoke the ‘atmosphere’ of a place, rather than simply conveying a factual account of its features. There are also considerable individual differences in the salient features reported, which it will be important to clarify for experimental design. This and other similar work feeds into the first in a series of scenarios that will drive the design and evaluation of the technical demonstrators. The arenas available for the experimental work – Head Mounted Displays, fully immersive six-sided CAVE and wide screen panorama – coupled with the photo-realistic imagery, provide a unique opportunity to explore the dimensions of embodied interaction. Empirical work will also compare these experiences with interaction in the real world enabling direct comparisons of alternative representations and interactions in the same setting. This opportunity to undertake such empirical work leads directly to a clearer understanding of the role of the perception-action loop in presence and to the design of optimal user experiences. The scenarios to be chosen will also enable us to explore interactive narratives and the contextual setting for embodied interaction.

Currently two scenarios have been proposed by the consortium. They are as follows:

- **Example scenario, Interactive: "Picnic":** One or more people are standing in a forest and can look around. As the project progresses the area within the space within which the user can move to visually explore the scenarion (Region of Exploration – REX) will expand to support explorative ego motion, - first on a disc and later in a volume. In this demonstrator type the user has complete, interactive control over position in the REX, and over choice of gaze direction within the Field of View – FOV (ie the horizontal and vertical field of view available to the user from a certain position in the REX).

- **Example scenario, Guided (Tour): "Safari":** One or more users are sitting in a vehicle (golf cart, or open land rover), and are being taken for a trip along some path (REX is 1D curve in space). It will be the system which controls the movement along the path (although the user can start and stop). Later in the project the users will be allowed to shift in the seats, and even stand up, effectively making REX a "tunnel".

Scenarios will be augmented with context appropriate virtual objects (and real physical objects in the immersive CAVE), as well as soundscapes. As the project progresses the fidelity of these augmentations will increase, but the demonstrator track is driven by developments in the IBR technology. In terms of the two example scenarios given above, the picnic scenario could be augmented with a physical companion, and a virtual picnic basket, and the safari scenario could be augmented with real seats and a tour guide, and virtual golf cart corner posts and top cover. Analysis of these scenarios will inform the first phase of the planned empirical investigations.

Coupled with the analysis of the scenarios a series of investigations of presence and how it is experienced in the real world will be undertaken by the project such as the pilot study described earlier in this section. It is envisaged that such studies will contribute to the project in a number of ways. Firstly, through the provision of benchmarks of presence against which to measure the articulation of presence within the created synthetic environments. Secondly, by the identification of those features which most contribute to a sense of presence in a given place. Finally, studies of how individuals or groups interact with the environment (built or otherwise) could well inform how interfaces to future information artefacts might be structured. For example, notions of information structure and navigation, the use of formal and informal signs and the demarcation of private and public space.

The project will explore both how presence is sensed and also what is sensed when a feeling of presence is experienced. One of the central themes of phenomenology is that the body is placed at the centre of our relationship with the world, in the words of Merleau-Ponty (1962) 'our body is both an object among objects and that which touches and sees them'. This implies that our experience of presence in the world is a bodily experience; embodied interaction. An associated issue is whether and to what extent presence is associated with the artefacts that populate a place. For example, can the sense of presence associated with returning to an old school building be solely as a result of the place or is it also as a result of the individuals who populated the space during the period of our education? Perhaps it is the case that objects can convey a sense of presence that is intimately bound to both the spaces where they reside and, by implication, with other objects which share that space. Indeed it could be argued that it is the artefacts that populate a space that gives it its sense of place and ultimately presence.

The output from this work will be a number of patterns that capture characteristics of presence. We expect these to range form the very detailed and physiological characteristics such as the refresh rate of graphics, or the resolution of images to much more general concerns of the whole user experience of virtual environments. These patterns will provide examples and design guidance for combining characteristics of the domain, the people, the technologies and the activities to give a sense of unmediated interaction within the information space.

6 Conclusions

The BENOGO project is unusual both in respect of the technology and in respect of the empirical work that is proposed. We will look at psychophysical views and at experiential views of the notion of presence through the concept of embodied interaction. Having photo-realistic imagery, a number of different presentation arenas and access to people's experiences of the real places that are rendered will provide special opportunities for investigating presence. The outcome will not be a whole theory of presence, but will, instead, be generic patterns of embodied interaction that capture design guidance and the importance of different features.

References

Benjamin, W. (1997) Charles Baudelaire Verso: London

Borden, I. (2001) Skateboarding, space and the city: Architecture and the Body, Berg, Oxford and New York.

Buehler, C. Bosse, M. Gortler, S. Cohen, M. and McMillan. L. (2001) Unstructured Lumigraph Rendering. Proc. SIGGRAPH '01, 425-432.

Dourish, P. (2001) Where the action is. The foundations of embodied interaction MIT Press

Loscos, C., Drettakis, G. and Roberts, L. (2000) Interactive Virtual Relighting of Real Scenes. IEEE Transactions on Visualization and Computer Graphics, Vol. 6, No. 4, p. 289 – 305, October-November

Merleau-Ponty, M. (1962) Phenomenology of Perception, trans from French by Colin Smith, Routledge and Kegan Paul Ltd.

Pallasmaa, J (1996) The Eyes of the Skin, Architecture and the Senses, Academy Editions.

Shum, H. and Kang, S. (2000) A Review of Image-Based Rendering Techniques. SPIE Int. Conf. On Visual Communication and Image Processing, 2-13.