Personal Panoramic Perception

Terry Boult   tboult@eecs.lehigh.edu
Vision and Software Technology Lab, EECS Department Lehigh University

Abstract

For a myriad of military and educational situations, video imagery provides an important view into a remote location. These situations range from remote vehicle operation, to mission rehearsal, to troop training, to route planning, to perimeter security. These situations require a large field of view and most would benefit from the ability to view in different directions.

Recent research has led to the development of new technologies that may radically alter the way we view these situations. By combining a compact omni-directional imaging system and a body-worn display, we can provide a new window into the remote environment: personal panoramic perception (P^3). The main components of a P^3 system are the omni-directional camera, a body-worn display and, when appropriate, a computer for processing the video.

This paper discusses levels of immersion and their associated display/interface "needs". It also looks at the capture system issues, including resolution issues, and the associated computational demands. Throughout the discussion we report on details of and experiences from using our existing P^3 systems.

1   Introduction

The ability to generate panoramic video has been around for years, e.g. see [1, 2], but it has seen limited usage. What has changed recently, and is driving a growing interest, is the combination of simultaneous decreased size and increased quality in collection systems, coupled with low-cost means of presenting/processing this data to provide perspective images. This paper looks at the component technologies and the systems issues involved in supporting a *Personal panoramic perception* (P^3) system, where a user has a personal system for viewing different areas within a panoramic video stream. Unlike remote pan-tilt camera based systems, P^3 supports multiple users simultaneously looking in different directions, which makes it ideal for team oriented exercises.

The main components of a P^3 system are the omni-directional camera (with video recording or transmission), a body-worn display and, when appropriate, a computer for processing the video. Let us begin with an overview of these components of the system.

The *paracamera* based collection systems, pioneered by S. Nayar, is a compact camera system that images a hemisphere or more while maintaining a single perspective viewpoint, [3]. The images can be processed to produce a proper perspective image in any direction capturing the entire viewing hemisphere in a single image, see
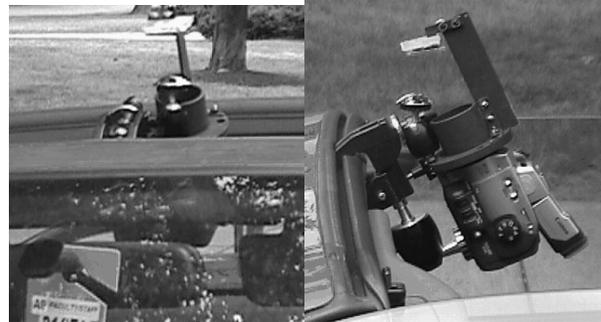

Figure 1. An example car-mount paracameras.

figure 3. The paracameras can vary in size from small transmitting systems (about 9cm tall by 6cm in diameter), to compact recording systems, to self contained underwater recording, to intensified night vision systems, see figures 1–2 for some examples. Supporting geometrically correct, live omni-directional video in a small package is a key constraint for most of the aforementioned applications.
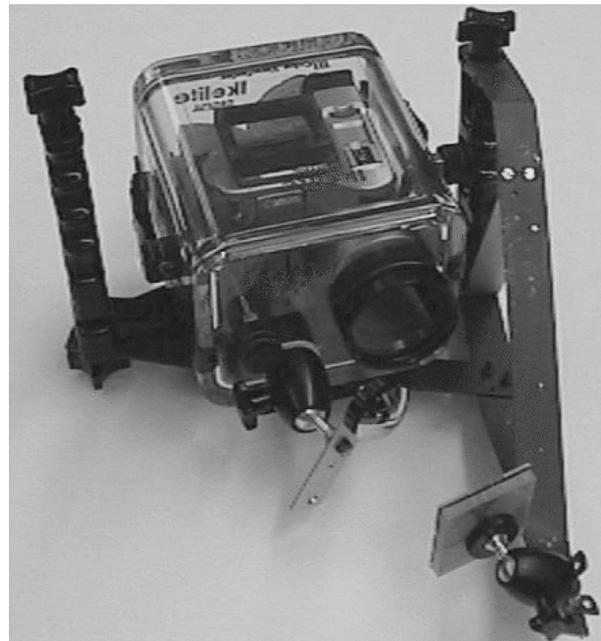

Figure 2. Second generation underwater paracamera. System dimensions are 25cm x 20cm x 18cm (plus 16cm for arm).

For the body-worn display we have been experimenting with different ways of displaying the information including direct paraimages, panoramic views, and user-directed perspective windows. The display device can

range from an immersive HMD with head-tracking, to a small monocular HMDs, to hand-held displays or even commercial TVs.

For current systems use COTS frame grabbers/processors. On a 233Mhz x86 processor our *Remote Reality* software allows the HMD to view 30 frame-per-second (fps) video of the remote site in whatever "direction" the user looks or directs. The system is capable of updating its viewing directions with only a 30 to 60 millisecond (15-30fps) delay.

This paper begins by examining different levels of immersion and their associated display /interface "needs", then looks at the capture system issues (including resolution issues) and ends with a looks at computational demands.

2 Levels of Immersion and User Interface

While there are many potential applications, we use the desired level of "immersion" to separate our discussion into three main groups:

- highly immersive: giving the user the impression they are at the remote application.

- informative: giving the user access to remote "information" in any or all directions, while still maintaining the user's local situational awareness.

- augmentive: enhancing either of the above interface with overlayed contextual information. This reduces immersion and adds complexity to the system, but it can increase situational awareness.

We briefly discuss each of these approaches.

2.1 High Immersion: Remote Reality

Our first interface is a immersive, like in many virtual reality system, but because it provide video access to a remote location we refer to it as *Remote Reality*. This interface uses a a bi-ocular HMD with a head tracker, see figure 4. The head tracker provides orientation information which is used to determine the viewing direction for the unwarping map. As the HMD turns (or if the users request a software "zoom") the virtual viewpoint is stationary; only the direction of the virtual "imaging array" is moved. We briefly look at the significant issues for this type of interface.

While any panoramic image generation process might be used for this type of immersive display, our work has concentrated on paracamera systems. In principle any other collection system *that maintains a single perspective viewpoint*, e.g. [4], could be used but most of them are larger, more difficult to calibrate or build.[1] If the viewpoint is not constant (or at least constrained to a be

[1] In [5] the complete class of possible lens & (single) mirror based systems that produce omni-directional image was investigated to see which satisfy the single-viewpoint assumption.

in a very small volume), the result is a lurching or bending in the images as the HMD changes orientation. Such artifacts significantly reduce the immersion.

With the single viewpoint imaging and an HMD with head-tracking, we can produce a system that provides a very smooth and very natural visual change. However maintaining the illusion of immersion also depends on acceptable system response time. Making the system fast enough took a few, but straight forward tricks: fixed point math for most computations and table lookup for the expensive operations. Because we can bound the size of all inputs and addresses we can bound calculation operations, including table-lookup-based division, can limit errors to less than 1/16 pixels using only 32 bit integer operations. With this, a 233Mhz x86 processor can update the view maps at 15-30fps (depending on other system load).

Figure 4. An immersive interface: Remote Reality head-tracked HMD. User is holding an early car-mounted para-camera.

To maintain the immersion, the displayed horizontal field of view (HFOV) needs to be reasonably matched to the display's visual extent and the user should see nothing but the projected remote video. Since most HMDs only have a 30-50 degree HFOV, the result is a little like looking through goggles. If a significantly larger physical HFOV is mapped into the small display, the user will perceive an unnatural warping or wobbling as they change their head position. While our prototype setup approximately matches the visual and physical sensations it does limit the situational awareness since there is no peripheral vision. With better HMDs, the potential exists to have a much larger FOV and include peripheral vision.

We also note that, the users need to turn their head, not just their eyes, to see in a new direction. While this initially distracts from their immersion, the user very quickly

Figure 3. An Omnidirectional (ParaImage) taken from a car. Note the "struts" are from an early version of the car mount, newer versions have only 1 (smaller) strut.

becomes acclimated to this constraint.

The high immersion of Remote Reality precludes the user from seeing their local environment, thus this is appropriate only for applications where the user is active in their observation but passive with respect to their own environment. If used in a tele-operation scenario, the user can control a remote vehicle's motion. For other users, it is as if they are passengers at the remote location. Some obvious applications for immersive remote reality are tele-operation, education, training and mission rehearsal. Except for the tele-operation, the point is to acquaint the user with a remote environment, acquiring knowledge and experience, and hence these applications lend themselves to recorded remote reality. A few less obvious applications include recording/replaying for: cataloging the state/contents of complex facilities such as ships or pipe complexes and security surveys of a route or building.

2.2 Informative

For other situations, e.g. police or military operations in urban terrain, is not acceptable for the user to be completely immersed. Instead the user must be aware of, and often moving within, their local environment while they simultaneously expand their situational awareness of the remote location. Thus we have been investigating different types of informative, but minimally invasive, interfaces. These interfaces use one of two display devices. The first is a small unobtrusive monocular HMD, see figure 5. The second is a hand-held device such as the portable TV in figure 6. (Of course, higher price/quality models of both of these types of display devices exist).

In the immersive interface the head-tracker provided a very natural means for the user to choose a direction to view. Even if the display was unobtrusive, as in figure 5,

the need to use one's head to choose a viewing direction is impractical while walking or taking part in almost any "local" event. One of the most difficult aspects of the informative displays is how, or if, to choose a viewing direction.

Figure 5. An informative monocular display with (a track-ball pointer).

A direct analogue of the head-tracked display is to provide the user with some type of a pointing device, e.g. the belt-worn track-ball in figure 5. With the pointing device the user can choose any direction of interest. The advantages of this is that they can maximize the displayed resolution (many small LCD can only display 320x240 true pixels), and, when needed, can choose new viewpoints. The disadvantage is that choosing a view requires a free hand and some practice to get used to the interface. It can be effective for team operations where someone is tasked with a particular view direction. Since this interface requires both an interaction device and reasonable CPU power, a machine supporting this can also support the following two interfaces, and one could tradeoff between the three.

The remaining informative displays are what we call information overview, they provide information on the entire scene at one time. The most obvious informative overview display is to generate a panoramic view. Unfortunately the aspect ratio of a panorama is far from that of most display technologies and direct display would result in very poor visible resolution. There is also the question of the type of panorama to show (spherical, cylindrical, or some custom version). To help with the resolution issues we display the scene in a split view, with a

panorama for the forward (with respect to vehicle) and one for the rear-view (with left-right reverse as in a rear-view-mirror). These are then stacked to provide full coverage in a 4x3 aspect ration display. We have experimented with various types of panorama and are currently using one where the azimuth angle grows linearly. We have found this provides a good tradeoff between resolution in regions of most importance and perceived image distortion. Note that this interface requires little training and no user interaction, but places the highest demands on the computing and I/O subsystem (we warp the full 640x480 image) and display resolution.

The "simplest" interface, is simply to broadcast the paraimage to a display device. This approach has three primary advantages:

1. There is no user need to "point" as the display shows all directions at once.

2. There is no added "computational" requirements.

3. The direction within the image is the actual direction from the camera to the object of interest.

The primary disadvantage is that the interpretation of the image is not as intuitive. As can be seen in figures 3 and 6, the lower part of the image is relatively easy to understand (front of vehicle), but objects behind the vehicle are upside down. With a little training, however, it becomes quite understandable (and is now the preferred interface by my students and I for operations in complex environments). If upside-down viewing is a problem, hand-held displaces can be rotated if needed, or inexpensive video flippers could be used.

2.3 Augmentive displays

The final type of interface, or more appropriately interface option, is being developed for applications where the user needs to augment their reality, rather than supplant it. The goal here is to add information, based on additional sensors and collateral data, to the video stream the user is seeing. The applications here include remote vehicle operation and urban police actions. Both ground and helicopter-based systems are being developed/tested.

For vehicle operation (as opposed to remote observation) it is generally not sufficient to immerse oneself in the video at the remote location. While the head-tracking interface is natural for view pointing, the user needs additional information such as speed and status, at a minimum they should be able to see their "dashboard". In addition it might be helpful if they could see vehicle position and direction with respect to a map. This type of augmentation is what one would expect in vehicle operation and like existing systems we are developing system to use remote GPS (or DGPS) and inertial navigation. Initially we anticipate the vehicle pilot will be at a safe location and will use the bi-ocular HMD with head tracking for setting

Figure 6. A hand-held display (low cost TV) showing a raw paraimage

view direction, leaving their hands free to operate the vehicle.

An added type of augmentation, currently only effective when the vehicle has stopped, is for us to provide a tracking system to warn the user of motion within the scene, see [6] for details on the algorithm. This is currently being added to the informative "overview" types of displays. (On a directed view interface we would have to provide a means for the user to locate the target or to understand the new viewing direction if automatically provided). We note that this can add significantly to the computational demands of the system, but can still be accomplished at 15-30fs with COTS hardware (high power drain) or 5-10 fps on more power efficient hardware.

2.4 So what interface to use?
In urban maneuvers, a driver can pilot the vehicles from a relatively safe location, but other team members needs to be following it for the clearing/security activities. The vehicles can transmit (encrypted if needed) omni-directional video while team members use augmenting remote reality to look for potential threats around the vehicle's location. Unlike what could be done with a pan-tilt system, *the team members can simultaneously look in different directions* a soldier can watch his own back. Additionally, no team member needs to transmit to the vehicle to control the pan/tilt viewing direction; the forward team can all be radio silent.

Informal observations show that for simple environments, pilots using the immersive HMD spend most of their time facing directly ahead, but as the environment becomes more complex and the desired path includes many turns, the pilots increasingly use their freedom of viewing direction. Other than the speed of response, using remote reality for a solo pilot is not significantly different than having a remote pan/tilt unit. The difference becomes apparent when the pilot or other team members needs to navigate while also locating significant non-navigational features within the environment.

Preparations are underway for formal evaluations of this hypothesis also a subjective comparison of the different interfaces for a collection of Military Operation in Urban Terrain (MOUT) type tasks. These will include both driving, target localization/identification (by driver) and target localization/identification by teams. The experiments will use a tele-operated vehicle, our RROVer (Remove Reality Omni-Vehicle), see figure 7

3   Systems issues
The first prototype immersive system strove to minimize cost while maintaining acceptable quality. Thus the system uses COTS parts. Our current data collection system was approximately $4K (+$1K for underwater) and the computing/HMD play-back system was about $3K. The system uses a 233Mhz K6 CPU (running Linux) & $300 video capture card. The system computes biocular 320x240 30 fps NTSC video. This resolution is reasonably matched to the HMD used, which is currently Virtual I-O glasses. The VIO built-in head tracker provides yaw, pitch and roll, with updates to the viewing direction at 15-30fps. With a better head tracker (e.g. Intersense IS300) and 300Mhz CPU we can insure consistent 30fps update of both viewpoint and video data. Better HMD's are also commercially available, at costs ranging from $2K to $10K, for low to medium volume usage and $20K very rugged high-volume usage. We are now porting to use a 640x480 resolution HMD and better head trackers and expect to demo this improved system at CISST.

We note that the above described hardware is not "wearable", but suitable for a desktop/remote driver. Unfortunately none of commercially available wearable computers have the video I/O bandwidth and resolution necessary for the 640x480 30fps video processing. We have assembled a wearable versions using a PC104+ based CPU with a BT848 video capture card. This operates at 30fps, but draws significant power (25-30W). A second (lower power, lower speed and lower cost) uses a Netwinder^{TM} and operates at 8fps. The limiting factor in these systems is I/O requirements of full resolution video, not the actual computations needed for the different user interfaces. A wearable version is needed only for the immersive display, for dual driving panoramas, the computer can be on the vehicle and transmit the processed video, or a separate machine can receive the raw video and retransmit the processed views.
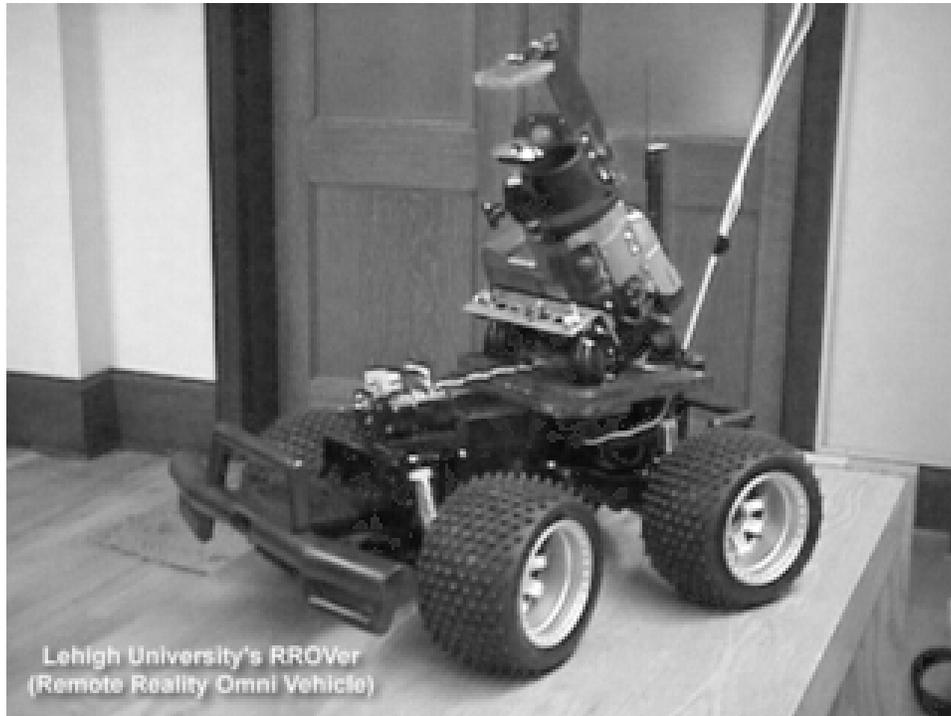
Figure 7. The Remove Reality Omni-Vehicle, a testbed for our study of personal panoramic perception

4   Para-Cameras and Resolution

While remote reality systems could be built with a multi-tude of cameras at the remote location, central in its design was the omni-directional camera designed by Shree Nayar [5]. This camera directly captures a full hemisphere (or more) while maintaining a single perspective viewpoint allowing it to be used for full motion video. Furthermore, placing two of these paracamera systems back-to-back allows a true viewing sphere, i.e. 360 x 360 viewing. Unlike fish-eye lenses, each image in the paracamera system can be processed to generate *geometrically correct* perspective images in any direction within the viewing hemisphere.

The paracamera's omni-directional imager combines an telecentric/orthographic lens and a parabolic mirror with the axis of the parabolic mirror parallel to the optic axis of the lens systems. The orthographic lens results in the entering rays being parallel. Rays parallel to the axis reflect off a parabolic surface at an angle such that they virtually intersect at the focus of the parabolic surface. Thus the focus of the paracamera provides a single "virtual" viewpoint. The single virtual viewpoint is critical for the Remote Reality system as it allows for consistent interpretation of the world with a very smooth transition as the user changes the viewing direction. While there are other systems with large or even hemispheric fields of view, as show in [7], *fish-eye lens and hemispherical mirrors do not satisfy the single viewpoint constraint.*

Because omni-directional imaging compresses a viewing hemisphere into a small image, maintaining resolution and captured image quality is quite important, and takes careful design. While the process scales to any size imager, the current systems use NTSC (640x480) or PAL (756x568) cameras. Note that the "spatial resolution" of the paraimage is not uniform. While it may seem counter intuitive, the spatial resolution of the paraimages is *greatest* along the horizon, just where objects are most distant. While the process scales to any size imager, the current systems use 640x480 NTSC (or 756x568 PAL) cameras. If we image the whole hemisphere, the spatial resolution along the horizon is $\frac{240\text{pixels}*2*\pi}{360\text{degrees}} = 4.2\frac{\text{pixels}}{\text{degrees}}$ (5.1 PAL) which is 14.3 arc-minutes per pixel (11.8 PAL). If we zoom in on the mirror, cutting off a small part of it, to increase the captured mirror diameter to 640 pixels (756 PAL), we can achieve 10.7 arc-minutes per pixel, i.e. 5.5 pixel per degree (6.6 PAL).

As a point of comparison, let us consider a traditional "wide-angle" perspective camera, such as those used in building "multi-camera" panoramic systems. If we allow for a small overlap in fields of view, to support blending at the seam, it would take 3 cameras with about a $150°$ horizontal field-of-view (FOV) to form a panorama. Note that each of these would have $\frac{640\text{pixels}}{150\text{degrees}} = 4.2\frac{\text{pixels}}{\text{degrees}}$, i.e. about the same as the Paracamera. Clearly, the traditional cameras would need more hardware and computation.

The paracamera's unique design yields what may be a new pareto optimal design choice in the resolution/FOV trade-off. We have the horizontal resolution of a $150°$ camera but cover the full $360°$ of the horizon. The "lost pixels" occur in the region above the horizon where the para-camera's resolution goes down, while traditional cam-

eras have increasing overlap.

As an informal point on the "quality", we note that some graphics/VR-oriented people hear about the output resolution, 320x240 16bit color, used in the immersive display, and want to dismiss it as inadequate. However, the initial system has been demonstrated to a large number of people ($>$ 1000), e.g. see [8], [9] and [10], with very positive feedback from most of them. Even the "skeptics" who have tried it admitted they were surprised at the quality. While the resolution is far from that of high end graphics systems, the naturalness of objects, fluidity of motion and the complex/subtle textures (even at low-resolution) of the video seem to make up for the pixel loss.

We note that Cyclovision now sells a 1Kx1K still camera version and we have built a 1Kx1K system that operates (but cannot record) at 5fps system. Higher resolution/speed systems are being developed, though they will be considerable more expensive than those based on consumer cameras.

5 Camera issues

While a number of paracamera models are commercially available from www.cyclovision.com, for most of our remote reality system have developed our own smaller custom designs directly incorporating camcorders rather than cameras, e.g. see figures 1. (Note small 9cm tall systems are now commercially available from cyclovision.) The development of the underwater cameras and vehicle cameras involved solving both optical and mechanical design problems. We are currently working on an omnidirectional system for helicopters and one to be carried underwater by a dolphin.

Figure 1 shows some custom car mounts for omni-cameras. The early vehicle mounts, see (left) used the Cyclovision paracameras and a separate tape-recorder inside the vehicle. They can be attached to the car windshield or roof via suction-cups and straps and, while large and obtrusive, were quite functional. The second generation uses our custom design with optical folding and integrated camcorder. This puts the user and camera back behind the mirror and inside the vehicle. To use it one only needs to "pop-up" the mirror above a sun roof. The inset shows a side view. In both cases, damping vehicular vibrations are an issue.

From our experience there are 3 main issues in omni-directional camera design for these types of applications:

1. Resolution limits imposed by optical components (lenses and mirrors).

2. Resolution limits imposed by camera electronics including pixel counts, light sensitivity and readout electronics. The single most significant camera issue, because of the unwarping, is interlace vs progressive scan. The second is camera pixel counts and general CCD/color "resolution" issues.

3. Mechanical mounting; Even small vibrations introduce blurring.

6 Conclusion

This paper has discussed some of the major issues in developing a personal panoramic perception system. Individual applications will need to tailor the concept to their situations, but the paper should provide a good starting point for the user interface issues, the imaging issue and some systems issues.

When combined with the small size and ease of use of the paracamera-based capture devices, personal panoramic perception and remote reality offers significant advantages in a number of domains where simulated worlds or simple video are currently being used.

[1] D. Rees, "Panoramic television viewing system." United States Patent No. 3,505,465, April 1970.

[2] J. Charles, R. Reeves, and C. Schur, "How to build and use an all-sky camera," *Astronomy Magazine*, April 1987.

[3] S. Nayar, "Omnidirectional video camera," in *Proceedings of the 1997 DARPA Image Understanding Workshop*, May 1997.

[4] V. Nalwa, "A true omnidirectional viewer," tech. rep., Bell Laboratories, Holmdel, NJ 07733, USA, February 1996.

[5] S. Nayar, "Catadioptric omnidirectional camera," in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, pp. 482–488, June 1997.

[6] T. Boult, A. Erkin, P. Lewis, R. Micheals, C. Power, C. Qian, and W. Yin, "Frame-rate multi-body tracking for surveillance," in *Proc. of the DARPA IUW*, 1998.

[7] S. K. Nayar and S. Baker, "Complete Class of Catadioptric Cameras," *Proc. of DARPA Image Understanding Workshop*, May 1997.

[8] T. Boult, C. Qian, W. Yin, A. Erkin, P. Lewis, C. Power, and R. Micheals, "Applications of omnidirectional imaging: Multi-body tracking and remote reality," in *Proc. of the IEEE Workshop on Computer Vision Applications*, Oct. 1998.

[9] T. Boult, "Remote reality demonstration," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1998. Technical Demonstration.

[10] T. Boult, "Remote reality," in *Proc. of ACM SIGGRAPH 1998*, 1998. Technical Sketch.