

# A Virtual Window On Media Space

*William W. Gaver*<sup>\*†</sup>

*Gerda Smets*<sup>†</sup>

*Kees Overbeeke*<sup>†</sup>

\*Royal College of Art  
Kensington Gore, London SW7 2EU, U.K.  
gaver@rca-crd.demon.co.uk

†Technische Universiteit Delft  
Jafalaan 9, 2628 BX Delft, The Netherlands  
c.j.overbeeke (or) g.j.f.smets@io.tudelft.nl

## ABSTRACT

The Virtual Window system uses head movements in a local office to control camera movement in a remote office. The result is like a window allowing exploration of remote scenes rather than a flat screen showing moving pictures. Our analysis of the system, experience implementing a prototype, and observations of people using it, combine to suggest that it may help overcome the limitations of typical media space configurations. In particular, it seems useful in offering an expanded field of view, reducing visual discontinuities, allowing mutual negotiation of orientation, providing depth information, and supporting camera awareness. The prototype we built is too large, noisy, slow and inaccurate for extended use, but it is valuable in opening a space of possibilities for the design of systems that allow richer access to remote colleagues.

**KEYWORDS:** CSCW, groupwork, media spaces, video

## INTRODUCTION

Media spaces are computer-controlled networks of audio and video equipment designed to support collaboration [2, 4, 6, 14, 17, 22]. They are distinguished from more common videophone, video-conferencing, and video broadcasting systems in that they are continuously available environments rather than periodically accessed services. Because maintaining high-bandwidth connections is costly, current video services are typically used for planned and focused meetings. Media spaces, in contrast, assume a future in which broadband networks are commonplace, the data rates needed for high fidelity video and audio are a trivial fraction of the total available, and thus the systems can be left "on" all the time. In practice, this is usually simulated in-house, using dedicated networks of analog audio and video cables leading from central computer-controlled switches to equipment in offices and common areas.

The constancy of connections that characterizes media spaces has several implications for how they are used. Because they are not associated with special events (e.g. meetings), they become part of the everyday work environment. It is common always to have a video connection somewhere, often to a common area, if only because such views are more pleasant than blank monitors. This implies, finally, that the proportion of time spent using the media space for meetings is relatively low, and instead they often are used to support a more informal, peripheral awareness of people and events. Again, this distinguishes them from more commonly encountered video systems.

## Trouble in Media Space

While there is anecdotal evidence that media spaces can support professional activities [2, 6, 13, 14, 17], and particularly long-term collaborative relationships [1, 3], quantitative data supporting the value of these technologies have been more difficult to find. Typically studies show that adding video to an audio channel makes no significant effect on conversation dynamics or on the performance of tasks that do not rely heavily on social cues [7, 19, 20]. Even in a more naturalistic media space setting, Fish et al. [4] found that people usually used the Bellcore system as a prelude to physically co-present meetings, and concluded that it did not clearly add to the functionality provided by telephone or email systems.

These results seem relevant primarily for the role of video in supporting relatively focused interactions. But one of the motivating intuitions behind early media space research was that they might help create and sustain a more informal sense of shared awareness [e.g., 2, 17, 22]. The fact that much recent research meant to assess media spaces seems focused on more formal uses may be because of the difficulty of finding quantitative data that addresses their informal possibilities. In any case, it has made it difficult to assess these original intuitions.

## *Access to the Task Domain*

Observational and analytic studies, on the other hand, have suggested limitations on the ability for media spaces to support informal shared awareness. For instance, Heath and Luff [9] described how co-present collaborators shape their activities and utterances for their partners, and contrasted this with the difficulties they observed in organizing these sorts of visually mediated activities in and

through media spaces [10]. They concluded that while a great deal of everyday collaboration is mediated by access to colleagues in the context of their tasks, this sort of access is often not provided by current media spaces.

A similar point was made by Nardi et al. [15] in their study of video used during neurosurgery within operating rooms and remote offices. Video is important in such settings because it allows visual access to events that are otherwise inaccessible (e.g. cameras are pointed into the head during brain surgery), and thus provides the awareness necessary for coordination. The emphasis of this sort of application is on visual access to tasks, not faces; thus Nardi et al. [15] recommend "turning away from talking heads" and instead focusing on "video-as-data."

This work is valuable in emphasizing the ability for video to support awareness of task-related artifacts. It is less convincing as a case against face-to-face video, however, since giving access to tasks is not incompatible with giving access to people. From our perspective, the point is not that cameras should be focused away from people towards workbenches (or skulls, as the case may be), but that the narrow focus of video itself must be broadened.

#### *Extending Affordances*

Gaver [5] analysed the affordances of media spaces to understand how the technologies shape perception and interaction. This analysis emphasized several limitations on the visual information media spaces convey:

- Video provides a restricted field of view on remote sites.
- Video has limited resolution.
- Video conveys a limited amount of information about the three-dimensional structure of remote scenes.
- There are discontinuities (or "seams," [13]) on the edges of scenes and between views from different cameras.
- There are also discontinuities between local and remote scenes and their geometries.
- The medium is anisotropic: the discontinuities between local and remote geometries are not reciprocal (and thus not predictable).
- Movement with respect to remote spaces is usually difficult or impossible.

Each of these attributes has implications for collaboration in media space. But the inability to move with respect to remote spaces may be most consequential of all. As Gibson [8] emphasized, movement is fundamental for perception. We move towards and away from things, look around them, and move them so we can inspect them closely. Movement also has implications for the other constraints produced by video. If we can look around, we increase our effective field of view. Moving can provide visual information that is lost because of low resolution [21]. It provides information about three-dimensional layout in the form of movement parallax [8, 16]. Finally,

movement might allow people to compensate for the discontinuities and anisotropies of current media spaces.

#### **Allowing Movement in Remote Spaces**

One approach to approximating movement within remote sites was explored using the MTV (for Multiple Target Video) system, which employed several switched video cameras in each of two offices [7]. Observations of six pairs of partners collaborating on two tasks indicated that the increased access was indeed beneficial. Participants used all the views, and were often creative, finding unexpected ways to gain access to their colleagues and their working environments. In fact, they accessed face-to-face views for much less time than views that included places and objects relevant for the tasks. This supports suggestions that access to task domains may be more useful than access to colleague's faces [10, 15]. However, participants did seem to rely on quick views of their colleagues as a way to assess attention and availability; looking times may be misleading as a basis for judging the importance of these views.

Though multiple cameras provided valuable visual access for collaboration, a number of problems with this strategy became clear. Despite the proliferation of cameras (and associated clutter), there were still significant gaps in the visual coverage provided. In addition, participants seemed to have problems establishing a frame of reference with one another, and in directing orientation to different parts of their environments. One result was that the video images themselves became the shared objects, rather than the physical spaces they portrayed, and participants would point at these images rather than the offices themselves. In general, the greater access provided by multiple cameras seemed outweighed by the addition of new levels of discontinuity and anisotropy.

Despite these problems, increasing visual access to remote environments seems a clearly desirable goal. In this paper, we describe another approach, involving the creation of a Virtual Window that allows true visual movement over time, rather than a series of views from static cameras. By providing an intuitive way to move remote cameras, we believe we can overcome many of the limitations of video for supporting peripheral awareness without introducing the problems that come with multiple cameras.

#### **THE DELFT VIRTUAL WINDOW**

The basic idea of the Virtual Window is that moving in front of a local video monitor causes a remote camera to move analogously, thus providing new information on the display (see Figure 1). To see something out of view to the right, for instance, the viewer need only "look around the corner" by moving to the left; to see something on a desk, he or she need only "look over the edge," and so forth. The result is that the monitor appears as a window rather than a flat screen, through which remote scenes may be explored visually in a natural and intuitive way.

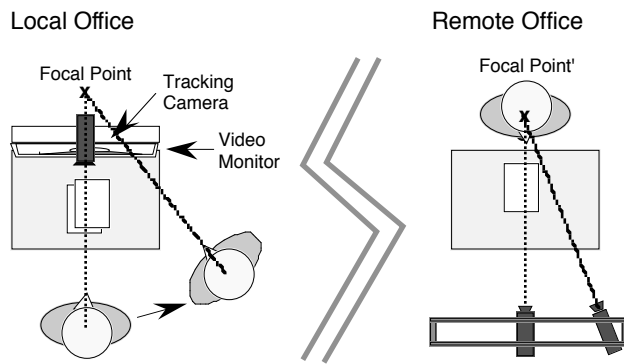


Figure 1. The Virtual Window: Local head locations are detected by a tracking camera and used to control a moving camera in the remote office. The effect is that the image on the local monitor changes as if it were a window.

### Movement Parallax and Depth Television

The Delft Virtual Window was invented by Overbeeke and Stratmann [16] originally as a means for creating *depth television*, allowing information for three-dimensional depth to be conveyed on a two-dimensional screen. The system creates the self-generated optic flow patterns that underlie *movement parallax*. As the head is moved around a focal point (shown in Figure 1), objects appear to move differently from one another depending on their distances (this is easy to see by moving one's head around an object while focusing on it: objects in the background seem to move parallel with the head, while those in the foreground move against it). Movement parallax is well suited for depth television because it does not require different images to be presented to both eyes. Indeed, similar methods have been used for computer graphics [12], but the Delft Virtual Window is the first system that provides movement parallax around a focal point for realtime video [16].

The Virtual Window has been tested experimentally by comparing people's accuracy at judging depth in remote scenes when they were viewed from static cameras, from moving cameras that they did not control, and from the Virtual Window system [16]. A clear advantage was found for the Virtual Window system over static views, and a significant decrease in variability of depth judgements when compared with those made from passively viewed moving scenes. The experimental evidence thus supports the intuitive impression that the Delft Virtual Window can do a good job of conveying depth information.

### Affordances for Increased Access

It is difficult to implement Virtual Window systems with the speed and accuracy necessary to give very good impressions of depth. But the technique gives rise to a number of other, serendipitous affordances that make even less-ambitious versions potentially beneficial for media spaces.

**Field of View** Because the camera moves around a focal point, it provides access to a much larger area of the remote scene than stationary cameras do (see Figure 2). The distance of the focal point from the camera determines the effective field of view. If it is set at infinity, for example, the camera moves only laterally and relatively little is added to the field of view. At the opposite extreme, if the focal point is set at the front of the camera itself, there is no lateral movement and the camera movement is equivalent to that provided by a pan-tilt unit. The field of view is greatly expanded, but parallax information for depth is lost.

**Resolution** As Gaver [5] pointed out, for static cameras there is an inherent tradeoff between field of view and resolution. This conflict does not exist for moving cameras: Not only is the effective field of view increased by allowing movement, but Smets [21] has shown that information for fine details can be obtained over time from a moving camera: effective resolution is increased as well.

**Continuity of the Remote Scene** Although the greater field of view offered by the Virtual Window must be accessed over time, new views are linked continuously. Instead of jumping from one view to another, one moves smoothly among views, making it easy to understand how they relate to one another. This contrasts with the MTV system, in which jumps among views introduced gaps and discontinuities that seemed to impede orientation [7, 11].

**Continuity with the Local Scene** If visual movement within the remote scene appears continuous with movement-induced shifts of perspective on the local one, the sense of continuity in and through media space should be increased. The glass screen of the video monitor will continue to act as a barrier between local and remote spaces, of course, but no longer spaces with different physics (i.e., in which head movements produce different visual consequences).

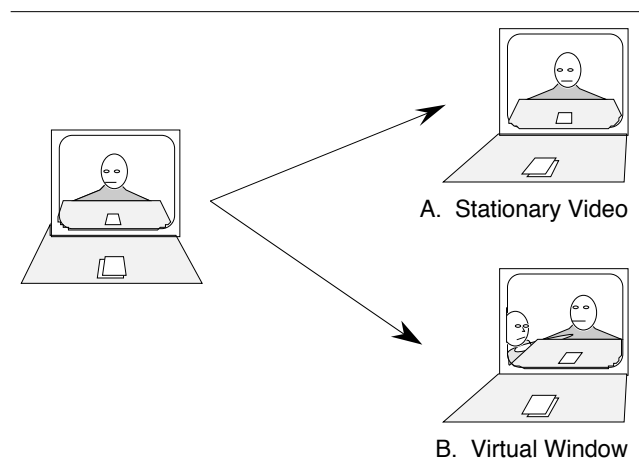


Figure 2. If cameras are stationary (A), local movements do not change the field of view, but do introduce discontinuities between local and remote spaces. In the Virtual Window system (B), local moves can provide a greater field of view continuously with local visual changes.

*Control and Coordination:* Finally, local control over remote cameras has several implications for perception and interaction in media space. Not only does it imply a larger field of view, but one available for active exploration rather than one depending on passive presentation. This may help support coordination with remote colleagues. As a simple example, it is common to hold something up to show a remote colleague, only to misjudge and hold it partially off-camera. Correcting the error usually requires explicit negotiation ("a little to the left...no, *my* left!"). The Virtual Window system allows the remote viewer to compensate for his or her partner's mistake simply by moving, without requiring any explicit discussion about the mechanics of the situation.

The combination of these affordances – the ability to expand the field of view, to raise the effective resolution, to increase the continuity within and between spaces, to support control and coordination, and to provide depth information – make the Virtual Window concept appealing for media space research. In the following sections, we describe our approach to implementing such a system and our experiences with the prototype we built.

#### EXPERIENCES WITH A VIRTUAL WINDOW

We collaborated to design, build, and assess an instantiation of the Virtual Window system. Most of the design, implementation, and initial programming were done at the Delft University of Technology. Two of the three devices were then installed, the software ported and developed, and the results tested at Rank Xerox Cambridge EuroPARC.

There are three separate aspects involved in instantiating a Virtual Window system:

- *Head-tracking* The location of the viewer's head with respect to the monitor must be determined.
- *Camera-moving* The camera must be moved in the remote site.
- *Mapping* The head location must be mapped to a desired camera location.

A number of approaches may be taken to these issues [16]. The prototype we built depended on a combination of idealistic goals (e.g., hands-free operation) tempered – sometimes betrayed – by pragmatic realities (e.g., cost of implementation). In the end, the process of designing, building, and trying it ourselves taught us, at least as much as watching it in use, both about the fundamental issues at stake and about the realities of implementation. Here we describe our tactics in some detail, and discuss some of the implications for our experiences with the system.

#### Head-tracking

We decided at the outset of the project that head-tracking should be accomplished without requiring users to wear any special devices or clothing. This seemed crucial if the Virtual Windows were to be used as casually as the rest of EuroPARC's media space. However, this precluded the

use of commercially available devices such as Polhemus sensors or infrared trackers. Instead, our version of the Virtual Window uses image processing on a video signal to determine head location.

For our implementation, a "tracking camera" is mounted on the local video monitor (Figure 1) and the incoming video stream is processed to extract the viewer's head location. The basic image processing strategy is shown in Figure 3. First, a single frame is digitized from the head-tracking camera when nobody is in view; this is used as the *reference image*. While the system is running, the reference image is subtracted from each incoming video frame, leaving a *difference image* that is processed to find an area of large differences assumed to be the viewer's head.

Finding such an area is at the heart of the image processing algorithm. First the differences along the rows of the image are summed, giving a difference profile for the height of the image. A threshold is set between the overall average of the differences and the greatest difference, and the top of the head is taken to be the first row of the image from the top that crosses the threshold (the head is assumed to be upright in the image). Then a horizontal difference profile is taken from a row on or just below the supposed top of the head, and a new threshold is set. The first cells to exceed this threshold from the right and left are assumed to be the sides of the head, and the center of the head to be halfway between the two.

A number of small variations can be used to improve this basic algorithm. For instance, it is useful to set a threshold for the minimum distance required before moving the camera. This helps to avoid spurious camera jitter caused by small fluctuations between successive frames.

This algorithm is simplistic in a number of ways. For instance, it does not recognise a head per se, but only areas where the incoming image is very different from the reference image. This means that the algorithm will track any

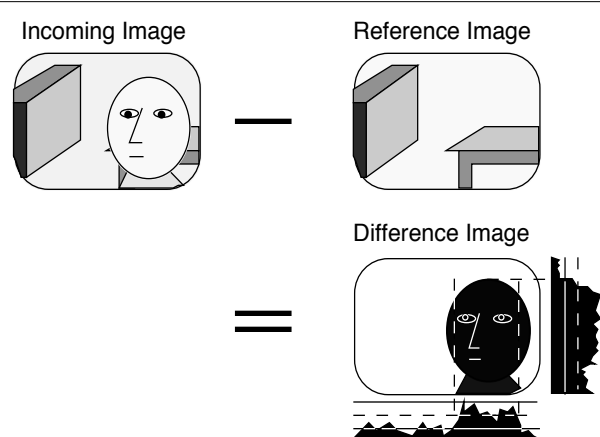


Figure 3. Head-tracking is accomplished by looking for values over threshold in a difference image produced by subtracting a reference image from each incoming frame.

source of change, such as a moving hand. It also means that the algorithm is very sensitive to changes in the ambient light, since these tend to introduce spurious differences between the incoming and reference images. Finally, it implies that more than one source of difference – such as two people in the tracking camera's field of view – may cause it to return inaccurate values (it tends to track whoever is higher in the tracking image, and returns an average horizontal value if they are at the same level). This is a manifestation of the more fundamental problem of scaling the Virtual Window to provide the correct visual information to more than one viewer.

Nonetheless, the algorithm works surprisingly well for all its simplicity. When conditions are good, the algorithm produces generally accurate values allowing a viewer's head to be tracked even against a cluttered background. Clearly there are more sophisticated approaches that might be used for this task, but there are severe constraints on the amount of processing that can be done while maintaining reasonable system latency. Even using this simple algorithm, we only achieved rates of about 3 - 7 frames per second on a Sparcstation 2; more accurate algorithms might not be worth still slower rates.

#### Camera-Moving

To move a camera around a focal point, recreating the optics of looking through a window, it is necessary both to rotate it and to move it laterally. This means that commercially available pan-tilt units are inadequate, unless the focal point is set to the front of the camera and no lateral movement is required.

We constructed our camera-moving apparatus from two A3 size flat-bed plotters that originally used software-controlled stepper motors to move pens over paper. We modified them extensively, cutting away most of the flat bed to reveal the basic frame, moving the control boards,

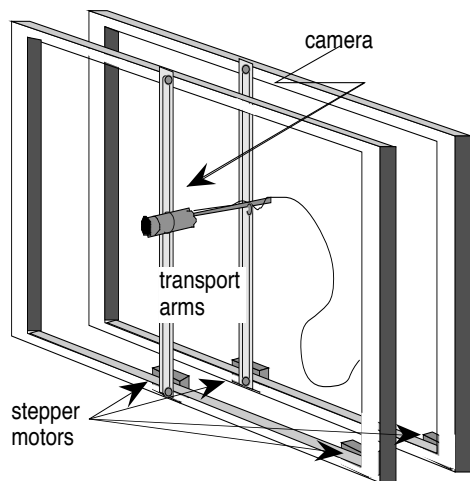


Figure 4. The camera transport mechanism uses two transport arms to move the front and back of a thumb camera separately.

and mounting them together so they would stand vertically (see Figure 4). The two pen transports are used to move the front and back of a Panasonic thumb-sized camera separately; each is powered by two stepper motors controlled over an RS232 link by the host computer. Though we had originally planned to use the built-in hardware and software to control the motors, this produced only instantaneous acceleration and deceleration, which led to unacceptably shaky camera movement. We hired an electronics contractor to develop new control hardware and software, which greatly enhanced the system by allowing smooth acceleration and deceleration of each motor separately.

The camera-moving devices are successful in being able to move a camera relatively quickly and smoothly over an area of about .35 X .2 meters. However, when two of them were moved from the large workshop in Delft where they had been designed and initially tested to the smaller, quieter office environment in Cambridge, it quickly became apparent that they are far too large and noisy to be acceptable for office use. Each of the devices takes up a volume of about .7 X .5 X .2 meters, and has a footprint of roughly .8 X .5 meters, larger than most of the video monitors being used. In addition, the motors cause audible vibrations in the frame. When we changed the system to allow each of the four motors to accelerate and decelerate independently, as described above, the noise problem was greatly exacerbated because each motor introduced its own independently changing frequency component. The resulting noise, though sounding impressively like a science fiction sound effect, is obviously too intrusive to be used in an office environment. In sum, the camera-moving devices have been adequate for our initial research, but a different design would be necessary for longer-term use.

#### Mapping Head Location to Camera Movement

A final issue for implementing a Virtual Window is the mapping between head and camera location. We discuss two aspects of this here: the determination of the focus point and errors caused by the expression of location as a point in the tracking camera's picture plane.

*Determining the Focal Point* One difficulty in implementing the virtual window system is in determining the focal point about which the remote camera is to move. Ideally, the viewer's actual focus could be determined by measuring gaze direction, convergence and accommodation. In practice, this seems difficult at best, not clearly necessary depending on the aims of the system, and almost certainly unfeasible if the system is to be used casually.

For our prototype, then, the focal point was set by the user using a simple graphical interface. We assumed that the focal point is always on a line extending from the center of the camera moving device. By taking the origin of our movement coordinates at that point, we can express the focal point simply as the ratio of front and back camera movements (see Figure 5). If the ratio is 1, the focal length is infinite, the front of the camera moves as much as

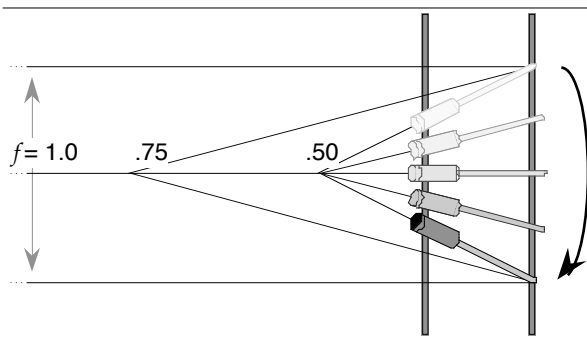


Figure 5. The focal point,  $f$ , can be expressed as the ratio of front to back movement. When  $f$  is 1, the focal point is at infinity and the camera only moves laterally. When  $f$  is 0, the focal point is at the front of the camera and the effect is like a pan-tilt device. Here the camera is shown as it moves around an  $f$  of .5 from top to bottom.

the back, and the effect is one of lateral movement with no rotation. If the ratio is 0, the focal point is the front of the camera, and the camera only rotates without moving laterally, just like a pan-tilt unit. Intermediate ratios give intermediate focal lengths.

**Angular Locations and Visual Information** For our prototype, we simply mapped the pair of coordinates returned by the head-tracking software to a new location for the back of the camera so that the maximum values of each would map to one another. This seems satisfactory in practice, but in reality it leads to systematic differences from the optical changes that movement in front of a window would make. In Figure 6, for instance, the two heads are both on the edge of the tracking camera's field of view, and so would return the same head locations and receive the same view from the remote camera. But if the monitor were really a window, the views would be different, as indicated by the lines of sight shown in the figure. This disparity arises because the edges of the tracking camera's image plane do not map to the edges of the monitor. The practical consequences of this disparity are unclear – again, our simple

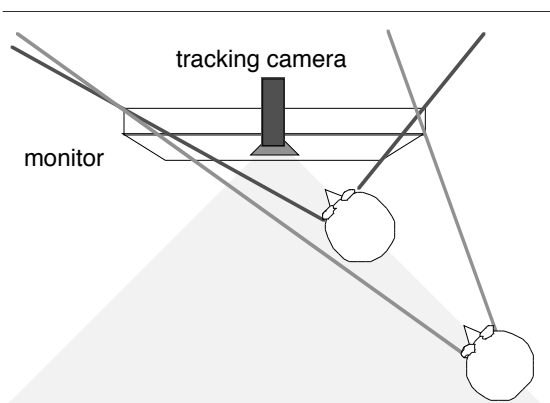


Figure 6. Equal locations in the tracking camera's picture plane should sometimes map to different camera positions.

mapping seems satisfactory – but the issue bears consideration.

As we suggested earlier, implementing a Virtual Window that can move a remote camera with the speed and accuracy necessary for veridical depth perception is difficult; some of the issues we have just discussed should make clear why this is so. We relaxed a number of the requirements for our prototype, since we were less interested in producing convincing depth information than we were in exploring the other affordances offered by the Virtual Window. Nonetheless, in many cases the changing scene provided by our implementation does evoke a good impression of depth (albeit at the wrong scale: often the remote office seems like a relatively small box). More importantly, the prototype has allowed us to explore some of the possibilities of using the Virtual Window to provide greater access to remote sites.

### Observing the Virtual Window in Use

To observe the system in use, we had six pairs of participants use it in pursuing two simple collaborative tasks. Subjects sat in separate offices, each controlling camera movement in his or her partner's office using the Virtual Window. The first task was called the *Room-Draw Task*, and required each participant simply to draw a floor-plan of his or her colleagues' office. The second task was the *Overhead Projector Design Task*, which asked the partners to redesign an overhead projector so that the lens-carrying arm would not block the audience's view. These tasks were modelled after similar ones used previously to assess collaboration in media spaces [7, 11]. They are designed to be simple, easily understood and motivated, and to focus on participants' access to their remote colleagues' environment.

Our observations tended to confirm the advantages, and emphasize the deficiencies, that we had noticed in developing the system. In the following, we briefly describe the problems that participants had with the system, then the advantages it provided.

### When It Was Bad...

The first two pairs of participants used the system on a beautiful spring day, with white clouds racing over a bright blue sky. Unfortunately, this provided a compelling demonstration of the head-tracking algorithm's susceptibility to variations in ambient light. The reference images we used could not be representative of the wide ranges of room illumination, and so the cameras often moved erratically as the head-tracking algorithm located the areas of greatest momentary difference, even though these were often due to the shifting light.

The results were extremely puzzling and frustrating to the participants in the study, who had not used the Video Window before, and who for the most part were relatively naive about media spaces in general. The movements of the view were only partially related to their own movements, and it seemed that because they were new to the

system they had little comprehension of what or whether anything was going wrong. In any case, there was little they could do to correct problems except to take a new reference image, which required ducking under the table so that they would not be in view of the tracking camera. On occasions when the view would show an area of the remote office that was useful, participants would often freeze in an attempt to keep the camera from moving. Ironically, in these circumstances a stationary camera would have given the participants better access to the remote site than a moving one – a point to which we return.

*But When It Was Good...*

Fortunately, the remaining participants were tested on cloudy days more typical of England, which meant that the systems were relatively accurate and stable. In these conditions, several advantages of the Virtual Window became clear. For example, there were several instances in which a participant would move slightly to achieve a better view on something his or her partner was displaying; thus, as we had expected, the system appeared to allow subjects mutually to negotiate orientation. In addition, there were occasions in which the system seemed to help participants maintain awareness of their partner's field of view, by increasing their awareness of the camera and its orientation (though this may in part have been due to the salience of the camera-moving device).

Most importantly, though, the Virtual Window did succeed in allowing participants to explore their partner's office visually, and the mapping between local movements and remote views appeared natural to the users. It seems difficult to convey the force of this result because of its simplicity. For instance, when one participant wanted to look down and to the side, he simply stood up and moved to the side. This sort of observation seems easy to overlook in the midst of the many difficulties people had with the current system. But the fact that this is possible at all, and that it seemed so natural, is a major success of the Virtual Window system.

## CONCLUSIONS

Providing the ability to move with respect to remote spaces seems a clearly desirable goal. But our experiences with the Virtual Window, as well as with the earlier MTV system [7] suggest that the vague notion of "remote movement" should be decomposed. From this perspective, experiencing a monitor as a window requires:

- user access to new views of the remote site
- linked continuously in space and time
- produced by local head movement
- with enough speed and accuracy for movement parallax.

This decomposition is useful in comparing strategies for providing greater access to remote scenes. For instance, the original MTV system [7] provided new views of remote sites, but they were not linked continuously in space or time. A later version, which replaced switching with multiple monitors [11], allowed continuous access

over time, but there were still discontinuities (gaps) in spatial coverage. Pan-tilt-zoom units provide both sorts of continuity, but are typically controlled by joysticks and similar devices. Finally, the Virtual Window we built enables head-tracked camera movement, but not true movement parallax.

Though the prototype we built is too slow and inaccurate to provide good movement parallax, and too large and noisy for everyday use, many of the problems we encountered seem less like inherent failings of the concept and more like challenges for iterative design. We may have been too ambitious in our design, rejecting reliable off-the-shelf equipment and using less-reliable custom solutions in an attempt to avoid compromising our ideals about how the system should work. Nonetheless, the prototype does illustrate some of the potential advantages of the Virtual Window approach. In addition, it opens a space of possibilities for the design of systems that allow much richer access to remote sites.

For instance, the inaccuracy of the head-tracking algorithm was clearly due to its reliance on an accurate reference image. There are several possibilities for increasing the robustness of this algorithm. If the overall differences between the incoming and reference pictures are consistently large, for example, it might be assumed that the reference image is out of date and the user could be notified. Alternatively, the reference frame could be replaced with the results of low-pass filtering the current stream of images; this would have the effect of blurring out any movement (e.g., of the head) and helping to compensate for shifts in light. Finally, other head-tracking techniques might fruitfully be explored, such as passive range-finding devices, including those which require users to wear special devices.

Similarly, we might expect that further iterations of the camera-moving system would greatly help with its size and noise. One possibility is to shift priorities from providing movement parallax towards providing a greater field of view. This would imply that lateral movement is unnecessary and allow the use of a commercially available pan-tilt-zoom unit. An additional advantage of using an off-the-shelf unit would be the opportunity to incorporate zoom as well, so that leaning towards the monitor might cause the camera to enlarge the image around the focal point. In fact, we are currently exploring such a system with Koichi'ro Tanikosi, Hiroshi Ishii, and Bill Buxton at the University of Toronto.

A more radical design option is to avoid moving a camera at all, and instead to produce a shifting view on remote scenes by moving a window over, and then undistorting, the view from a fish-eye lens. Apple Computer has developed a similar strategy for creating Quicktime "virtual reality" [18], but not for use with realtime video. The processing demands of such a strategy are quite high, but it has a number of advantages. Not only would it eliminate the difficult problems of mechanically moving the mass of a camera very quickly with no discernible vibrations, but it

would also do away with the problem of scaling the system to deal with multiple, distributed remote viewers. It is not clear that the strategy could be extended to produce lateral as well as rotary camera movement, but it seems well worth further investigation.

Finally, it is also desirable to design for the enduring differences between Virtual Windows and real ones. For example, a clear finding of our user study was the need to distinguish and allow separate control over movement in local and remote spaces. Once participants had achieved good views of remote spaces, they often seemed reluctant to move for fear of losing them. This problem is partially an effect of the current system's limitations. When working in front of a real window, moving away to achieve some local goal is easily reversed simply by moving back again. Using the current implementation of the Virtual Window, in contrast, moving back is no guarantee of recovering the original view. Though future versions should alleviate this problem, it may actually be desirable to maintain the dissociation. A foot pedal could be added to the system, for instance, allowing people to stop the Virtual Window so that local movement would not disturb a good view of the remote site.

In sum, the prototype Virtual Window is useful in opening up a wide space for the design of new video systems. Perhaps none will succeed in fully creating the experience of looking through a window into an office thousands of miles away, but many are likely to be useful in overcoming the limitations of existing systems. In the end, perhaps the most important contribution the Virtual Window makes is as a concrete reminder that media spaces need not be constrained to single, unmoving cameras left sitting on top of video monitors.

#### ACKNOWLEDGEMENTS

We thank Rank Xerox Cambridge EuroPARC and the Faculty of Industrial Design Engineering at Delft TU for supporting this collaboration, and particularly Bob Anderson and Allan Maclean. Peter Jan Stappers was an invaluable guide to Virtual Window design, particularly the head-tracking algorithm. We thank Ronald Thunessen for work on the camera-moving apparatus and Jeroen Ommering for the "Cameraman" motor-control software. Finally, we are extremely grateful to Abi Sellen for helping with the study reported here, and to her and Christian Heath, Paul Luff, Anne Schlotmann, Paul Dourish, Sara Bly and Wendy Mackay.

#### REFERENCES

- 1 Adler, A., and Henderson, H. (1994). A room of our own: Experiences from a direct office share. *Proceedings of CHI'94*. ACM: New York, 138 - 144.
- 2 Bly, S., Harrison, S., and Irwin, S. (1993). Media spaces: Bringing people together in a video, audio, and computing environment. *Communications of the ACM*, 36 (1), 28 - 47.
- 3 Dourish, P., Adler, A., Bellotti, V. and Henderson, A. (1994). Your place or mine? Learning from long-term use of video communication. Working Paper, Rank Xerox Research Centre, Cambridge Laboratory.
- 4 Fish, R., Kraut, R., Root, R., and Rice, R. Evaluating video as a technology for informal communication. *Proceedings of CHI'92*. ACM, New York, 37 - 48.
- 5 Gaver, W. The affordances of media spaces for collaboration. *Proceedings of CSCW'92*.
- 6 Gaver, W., Moran, T., MacLean, A., Löfstrand, L., Dourish, P., Carter, K., and Buxton, W. Realizing a video environment: EuroPARC's RAVE system. *Proceedings of CHI'92*. ACM, New York, 27 - 35.
- 7 Gaver, W., Sellen, A., Heath, C. and Luff, P. (1993). One is not enough: Multiple views on a media space. *Proceedings of INTERCHI'93*. ACM: New York, 335 - 341.
- 8 Gibson, J. J. (1979). The ecological approach to visual perception. Houghton Mifflin, New York.
- 9 Heath, C., and Luff, P. (1992a). Collaboration and control: Crisis management and multimedia technology in London underground line control rooms. *CSCW Journal*, 1 (1-2), 69 - 94.
- 10 Heath, C., and Luff, P. (1992b). Media space and communicative asymmetries: Preliminary observations of video mediated interaction. *Human-Computer Interaction*, 7, 315 - 346.
- 11 Heath, C., Luff, P., and Sellen, A. (1994). Rethinking media space: The need for flexible access in video-mediated communication. CHI submission.
- 12 Hodges, L., and McAllister, D. (1987). True three-dimensional CRT-based displays. *Information Display*.
- 13 Ishii, H., Kobayashi, M., and Arita, K. (1994). Iterative design of seamless collaboration media. *Communications of the ACM*, 37 (8), 83 - 97.
- 14 Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T., and Wellman, B. Experiences in the use of a media space. *Proceedings of CHI'91*. ACM, New York, 203 - 208.
- 15 Nardi, B., Schwarz, H., Kuchinsky, A. Leichner, R., Whittaker, S. and Scabassi, R. (1993). Turning away from talking heads: The use of video-as-data in neurosurgery. *Proceedings of INTERCHI'93*. ACM: New York, 327 - 334.
- 16 Overbeeke, C., and Stratmann, M. (1988). Space through movement. Unpublished doctoral thesis, TU Delft, The Netherlands.
- 17 Root, R. (1988). Design of a multimedia vehicle for social browsing. In *Proceedings of the CSCW'88*. ACM, New York. 25-38.
- 18 Rose, H. (1994). QuickTime VR: Much more than "virtual reality for the rest of us." *Converge*, August.
- 19 Sellen, A.. Speech patterns in video-mediated conversations. *Proceedings of CHI'92*. ACM, New York, 49 - 59.

- 20 Short, J., Williams, E., and Christie, B. *The social psychology of telecommunications*. London: Wiley & Sons, 1976.
- 21 Smets, G. Designing for telepresence: The interdependence of movement and visual perception implemented. *Proceedings of the IFAC Man-Machine Symposium*, 1992.
- 22 Stults, R. (1986). Media space. Xerox PARC technical report.