M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 260
Submitted to PRESENCE Special Issue on Augmented Reality

‘Mediated reality’

Steve Mann, N1NLF
steve@media.mit.edu
MIT Media Lab, Building E15-389,
20 Ames Street, Cambridge, MA02139
Tel. (617) 253-9610; Fax. (617) 253-8874

Abstract

The general spirit and intent of Augmented Reality (AR) is to *add* virtual objects to the real world. A typical AR apparatus consists of a video display with partially transparent visor, upon which computer-generated information is *overlayed*. The general spirit of what is proposed, like typical AR, includes *adding* virtual objects, but also includes the desire to *take away*, *alter*, or more generally to visually ‘mediate’ real objects, using a body-worn apparatus where both the *real* and *virtual* objects are placed on an equal footing, in the sense that both are presented together via a synthetic medium. Successful implementations have been realized by *viewing* the real world using a head-mounted display (HMD) fitted with video camera(s), body-worn processing, and/or bidirectional wireless communications. This portability enabled various forms of the apparatus to be tested extensively in everyday circumstances, such as while riding the bus, or shopping. The proposed approach shows promise in applications where it is desired to have the ability to reconfigure reality. For example, color may be deliberately diminished or completely removed from the real world at certain times when it is desired to highlight parts of a virtual world with graphic objects having unique colors. The fact that vision may be *completely* reconfigured also suggests utility to the visually handicapped.

1 Introduction

Ivan Sutherland, a pioneer in the field of computer graphics, described a head-mounted display with half-silvered mirrors so that the wearer could see a virtual world superimposed on reality [Earnshaw et al., 1993] [Sutherland, 1968], giving rise to “Augmented Reality (AR)”.

Others have adopted Sutherland’s concept of a Head-Mounted Display (HMD) but generally without the see-through capability. An artificial environment in which the user cannot see through the display is generally referred as a Virtual Reality (VR) environment. One of the reasons that Sutherland’s approach was not more ubiquitously adopted is that he did not merge the virtual object (a simple cube) with the real world in a meaningful way. Feiner’s group was responsible for demonstrating the viability of AR as a field of research, using sonar (Logitech 3D trackers) to track the real world so that the real and virtual

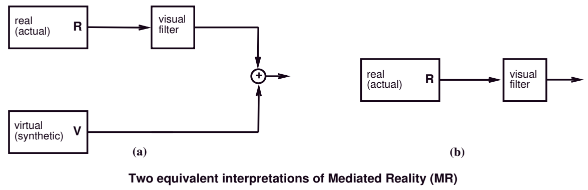


Figure 1: **Two equivalent interpretations of mediated reality (MR):** (a) In addition to the ability to add computer-generated (synthetic) material to the wearer’s visual world, there is potential to alter reality, if desired, through the application of a ‘visual filter’. The coordinate transformation embodied in the ‘visual filter’ may either be inserted into the virtual channel as well, or the graphics may be rendered in the coordinate system of the filtered reality channel, so that the real and virtual channels are in register. (b) The ‘visual filter’ need not be a *linear system*. In particular, the ‘visual filter’ may itself embody the ability to create computer-generated objects and therefore subsume the “virtual” channel.

worlds could be registered [Feiner et al., 1993b] [Feiner et al., 1993a]. Other research groups [Fuchs et al.,] also contributed to this development. Some research in AR arises from work in telepresence [Drascic, 1993].

AR, although lesser known than VR, is currently used in some specific applications. Helicopter pilots often use a see-through visor that superimposes virtual objects over one eye, and the F18 fighter jet, for example, has a beam-splitter just inside the windshield that serves as a heads-up display (HUD), projecting a virtual image that provides the pilot with important information.

The general spirit of AR is to *add* computer graphics or the like to the real world. A typical AR apparatus does this with beam splitter(s) so that the user sees directly through the apparatus while simultaneously viewing a computer screen.

The goal of this paper is to consider a wireless (untethered) apparatus worn over the eyes that, in real time, computationally *reconfigures* reality in addition to adding to it. This ‘mediation’ of reality may be thought of as a *filtering* operation applied to reality and then a combining operation to insert *overlays* (Fig 1(a)). Equivalently, the addition of computer-generated material may be regarded as arising from this filtering operation itself (Fig 1(b)).

A means of *mediating* (augmenting, enhancing, deliberately diminishing, or otherwise altering) reality, in real time, through an apparatus worn over the eyes, will first be described using an idealized implementation based on a hypothetical ‘lightspace glass’, and later in a more practical implementation, using video camera(s), a head-mounted video display, and a combination of body-worn and untethered remote processing hardware. In either case (idealized or practical), the entire apparatus will be referred to as a 'Reality Mediator' (RM).

1.1 'lightspace glass'

In what follows, a simplified model [Mann, 1994a] – *rays* of light – will be used. This simplified model neglects both wave-like properties (such as diffraction) and particle-like properties (such as quantum effects) of light. (The model also assumes the existence of the abstract notion of *instantaneous* frequency). A special window that can both measure and produce rays of light is hypothesized as a conceptual framework upon which the concept of 'mediated reality (MR)' is developed.

Consider a hypothetical window that would absorb and quantify every ray of light incident upon it. The information obtained from such a window — the location on the glass where each ray of light struck, its direction of arrival, its wavelength, and the exact instant in time it struck — would be sufficient to reconstruct all the light passing into the window. Suppose this hypothetical window could also produce any ray of light desired — that it could send out any number of light rays from specified locations on the glass, in specified directions, and of specified wavelengths. This hypothetical 'lightspace glass', would be essentially an ideal holographic video camera and ideal holographic video display in one.

Holographic video cameras and *holovideo* displays have actually been built [Hilaire et al., 1990] [Lucente et al., 1992], and the two have been connected, though the apparatus occupied some large optical benches and a room full of equipment. It would not even fit in a small room, let alone be small enough to wear in a pair of eye glasses. Nevertheless, perhaps in years to come, these technologies could become feasible.

The use of a hypothetical 'lightspace glass' to measure the way a scene responds to light has been previously proposed [Mann, 1994a], together with a more practical realization over a very limited domain ('lightspace' of a static monochromatic scene), though this apparatus has not been realized for moving scenes, and certainly not in a small enough package to be body-worn. However, the 'lightspace glass' has been a useful abstraction for the purposes of understanding the concepts underlying MR. Note that if it were possible to create a sufficiently realistic implementation of 'lightspace glass', it could hypothetically be used to surround an object, as well as all of the support circuitry for the glass itself, and render that object invisible, by virtue of its ability to absorb any ray of light before it got to the object and then re-produce that ray correctly at the other side of the object.

1.2 'lightspace glasses'

Suppose that we were to make a visor from this glass. Clearly, it could be used as a VR display, because the *holographic camera* functionality of it could absorb and quantify all the incoming rays of light and then simply ignore this information, while the *holovideo* portion of the glass could create a virtual environment for the user. (See Fig 2 (VR).)

The glass would also have the capability of functioning like an ordinary window in the sense that the *holographic camera* functionality of it could absorb and quantify all the rays of light incident upon it, and then the *holovideo* portion of it could send exactly those same rays of light out the other side. For the moment, assume that the ideal nature of the glass permits it to perfectly sustain the 'illusion of transparency'.

This 'illusion of transparency' would have many uses of its own. Obviously it could be used to make the visor function like a pair of sunglasses darkening rays of light coming out the other side. Because of the computer control, the darkening could even vary, in accordance with some gradient that would be darker up where the sun was, resulting in 'smart sunglasses'. The 'smart sunglasses' would use machine vision to track the sun and adjust the position of the darkening mask.

Many conventional sunglasses have a fixed gradient, typically being darker at the top than at the bottom, which creates an annoying artificial percept of motion when the wearer tilts his or her head back or forward, even though nothing in the scene is moving. "Smart sunglasses" would eliminate this problem by fixing the darkening mask with respect to the scene rather than the wearer.

Now in addition to creating the illusion of allowing light to pass right through, the visor could also create new rays of light, having nothing to do with the rays of light coming into it. The combined 'illusion of transparency' and the new light would provide the wearer with a typical AR experience (Fig 2 (AR)).

In an AR environment, graphics sometimes fail to stand out from the real objects. For example, when looking through the glasses at a brightly colored scene, there may not exist a unique color to use for the overlays. Suppose, however, that the glass, instead of creating an illusion of transparency, creates an illusion of being *achromat transparent*. Being *achromat transparent* means that each incoming ray of light is absorbed and quantified, and its wavelength is ignored. A ray from the same location, is sent out in the same direction, at the same time, but with a flat (grey) spectrum. This would make the user colorblind to real objects, making the real world appear less "busy" when combined with some colorful computer-generated overlays where color could be used, more effectively, to accentuate the virtual objects. This would prevent computer-generated objects from being "lost" in the clutter of the real world.

Using practical (non-holographic) implementations of MR (to be described in Sec 2), I have found color-reduced reality mediation to be quite useful. For example, when I am comfortably seated on a commercial airline or commuter train and wish to read text on my screen (e.g. read email), I like to "tone down" my surroundings so they take on a lesser role. I do not wish to be blind to my surroundings, as is someone who is reading a newspaper (newspapers can easily end up covering most of a person's visual field).

This form of reality mediation allows me to focus primarily on the virtual world which might, for example, be comprised of email, a computer source file, and other miscellaneous work, running in emacs19, with colorful text,

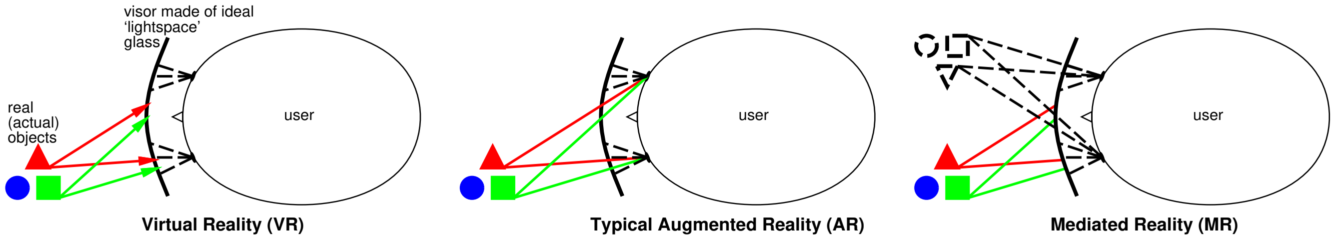


Figure 2: Consider a hypothetical glass that absorbs and quantifies every ray of light that hits it, and is also capable of generating any desired rays of light. Such a glass, made into a visor, could produce a virtual reality (VR) experience by ignoring all rays of light from the real world, and generating rays of light that simulate a virtual world. Rays of light from real (actual) objects indicated by solid shaded lines; rays of light from the display device itself indicated by dashed lines. The device could also produce a typical augmented reality (AR) experience by creating the 'illusion of transparency' and also generating rays of light to make computer-generated "overlays". Furthermore, it could 'mediate' the visual experience, allowing the perception of reality itself to be altered. In this figure, a non-useful (except in the domain of psychophysical experiments) but illustrative example is shown: objects are *left-right reversed* before being presented to the viewer.

where the text colors are chosen so that no black, white, or grey text (text colors that would get 'lost' in the new reality) is used. My experience is like reading a newspaper printed in brightly colored text on a transparent material, behind which the world moves about in black and white. I am completely aware of the world behind my "newspaper" but it does not distract from my ability to read the "paper".

Alternatively, the real world could be left in color, but the color mediated slightly so that unique and distinct colors could be reserved for virtual objects and graphics overlays. In addition to this 'chromatic mediation', other forms of 'mediated reality' are often useful.

1.2.1 Registration between real and virtual worlds

Alignment of the real and virtual worlds is very important, as indicated in the following quote [Azuma, 1994]:

> Unfortunately, registration is a difficult problem, for a number of reasons. First, the human visual system is very good at detecting even small misregistrations, because of the resolution of the fovea and the sensitivity of the human visual system to differences. Errors of just a few pixels are noticeable. Second, errors that can be tolerated in Virtual Environments are not acceptable in Augmented Reality. Incorrect viewing parameters, misalignments in the Head-Mounted Display, errors in the head-tracking system, and other problems that often occur in HMD-based systems may not cause detectable problems in Virtual Environments, but they are big problems in Augmented Reality. Finally, there's system delay: the time interval between measuring the head location to superimposing the corresponding graphic images on the real world. The total system delay makes the virtual objects appear to "lag behind" their real counterparts as the user moves around. The result is that in most Augmented Reality systems, the virtual objects appear to "swim around" the real objects...
>
> *Until the registration problem is solved, Augmented Reality may never be accepted in serious applications.*

(emphasis added)

The problem with many implementations of AR is that even once registration is attained, if the glasses slip down your nose, ever so slightly, the real and virtual worlds will not generally remain in perfect alignment.

Using the 'illusory transparency' approach, the illusion of transparency is perfectly coupled with the virtual world once the signals (video or *holovideo*) corresponding to the real and virtual worlds are put into register and combined into one signal. Not all applications lend themselves to easy registration at the signal level, but those that do (such as the finger-tracking mouse to be discussed in Sec 6.1) call for the 'illusory transparency' approach. In this case, I find that when the glasses slip down my nose a little (or a lot for that matter), both the real and virtual worlds slip down together in a unified way, and remain in perfect register. Since they are both the same medium (e.g. video or *holovideo*), once registration is attained between the real and virtual video signals themselves, the registration problem remains solved regardless of how the glasses might slide around on the wearer, or how the wearer's eyes are focused or positioned with respect to the glasses.

Another important point is that even with perfect registration, when using a see-through visor (with beam splitter or the like), real objects may lie in a variety of different depth planes, while virtual objects are generally flat (in each eye that is), to the extent that their distance is at a particular focus (apart from the variations in binocular disparity of the virtual world). This is not too much of a problem when all of the objects are far away as is often the case in aircraft (e.g. in the fighter jets using HUDs), but in many other applications (such as in a typical building interior) the differing depth planes destroy the illusion of unity between real and virtual worlds.

With the 'illusory transparency' approach, however, the real and virtual worlds exist in the same medium and therefore are not only registered in location but also in depth, since the depth limitations of the display device affect both the virtual and real environments in exactly the same way.

1.2.2 Transformation of the perceptual world

Even without any graphics overlays, mediated realities are still interesting and useful. For example, the colorblinding glasses in themselves might be useful to an artist trying to study relationships between light and shade. While it is certain that the average person might not want any part of this experience, especially given the cumbersome nature of the current realization of the RM, without question, there are at least a small number of users who would be willing to wear an expensive and cumbersome apparatus in order to see the world in a different light. Consider, for example, the artist who travels halfway around the world to see the morning light in Italy. As the cost and size of the RM decreases, no doubt there would be a growing demand for glasses that alter (enhance or diminish) tonal range, allowing artists to manipulate contrast, color, and the like.

MR glasses could (in principle) be used to synthesize the effect of ordinary glasses, but with a computer-controlled prescription that would modify itself automatically, while conducting automatically scheduled eye tests on the user.

The RM might also, for example, reverse the direction of all outgoing light rays to allow the wearer to live in an "upside-down" world (Fig 2 (MR)), perhaps being useful for experiments in psychology. Although the vast majority of RM users of the future will no doubt have no desire to live in an upside-down, left-right-reversed, or sideways rotated world, these visual worlds serve as illustrative examples of extreme reality mediation.

In his 1896 paper [Stratton, 1896], George Stratton reported on experiments in which he wore eyeglasses that inverted his visual field of view. Stratton argued that since the image upon the retina was inverted, it seemed reasonable to examine the effect of presenting the retina with an "upright image".

His "upside-down" glasses consisted of two lenses of equal focal length, spaced two focal lengths, so that rays of light entering from the top would emerge from the bottom, and vice-versa. Stratton, upon first wearing the glasses, reported seeing the world upside-down, but, after an adaptation period of several days, was able to function completely normally with the glasses on.

Dolezal [Dolezal, 1982] (page 19) describes "various types of optical transformations", such as the *inversion* explored by Stratton, as well as *displacement*, *reversal*, *tilt*, *magnification*, and *scrambling*. Kohler [Kohler, 1964] also discusses "transformation of the perceptual world".

Each of these "optical transformations" could be realized by selecting a particular *linear time-invariant system* as the visual filter in Fig 1. (A good description of *linear time-invariant systems* may be found in a communications or electrical engineering textbook such as [Haykin, 1983].)

The optical transformation to greyscale, described earlier, could also be realized by a 'visual filter' (Fig 1 (a)) that is a linear time-invariant system, in particular, a *linear integral operator* [Arfken, 1985] (page 669) that, for each ray of light, collapses all wavelengths into a single quantity giving rise to a ray of light, having a flat spectrum, emerging from the other side.

Of course, the 'visual filter' of Fig 1 (b) may not, in general, be realized through a *linear system*, but there exists an equivalent *nonlinear* filter arising from incorporating the generation of virtual objects into the filtering operation.

One final and somewhat amusing note on 'lightspace glasses' is in order. When I am wearing a practical implementation of the RM, (with head-mounted display) in my day-to-day social interactions in public, people often complain of a loss of eye contact with me. However, if I were wearing the hypothetical 'lightspace glasses' I would also be wearing my own personal holographic video display, because these glasses would be able to produce any collection of light rays. This would allow me to present any desired 3D image to those who look into the glasses. In particular, the loss of eye-contact that people complain about when they try to talk to me could be eliminated by using the glasses to generate a *holovideo* of my eyes. Of course if I were sleeping through a boring meeting, I could still present others with a *holovideo* of wide open eyes dancing in attentive saccades rather than the actual view of closed eyes.

2 Non-holographic realizations of MR

2.1 'Video transparency'

The idealized *holographic video* camera is difficult to make in practice, since one would require a microscopic lattice of photocells or the like with unrealizably fast response. Even a discrete realization of a holographic video camera made from a dense array of miniature video cameras is costly and bulky. The *holovideo* display is even more costly and bulky.

However, since the visor may be relatively well fixed with respect to the wearer, there is not really a great need for full parallax holographic video. The fixed nature of the visor conveniently prevents the wearer from having *look-around* with respect to the visor itself (e.g. look-around is accomplished when the user and the visor move together to explore the space). Thus two views, one for each eye, suffice to create a reasonable 'illusion of transparency'.

Others [Fuchs et al.,] [Drascic, 1993] [Nagao, 1995] have also explored video-based 'illusory transparency', augmenting it with virtual overlays. Nagao, in the context of his hand-held TV set with single camera [Nagao, 1995] calls the it "video see-through".

It is worth noting that whenever 'illusory transparency' is used, as in the work of [Fuchs et al.,] [Drascic, 1993] [Nagao, 1995] reality will be 'mediated', whether or not that mediation was intended. At the very least this mediation takes on the form of limited dynamic range and color gamut, as well as some kind of distortion, which may be modeled as a 2D coordinate transformation. Since this mediation is inevitable, it is worthwhile to attempt to exploit it, or at least plan for it, in the design of the apparatus. A 'visual filter' may even be used to attempt to mitigate the distortion.

A practical color stereo 'reality mediator (RM)' may be made from video cameras and display. One example, made from a display having 480 lines of resolution, is depicted in Fig 3.

It is desired to have the maximum possible 'visual bandwith', even if the RM is going to be used to conduct experiments on *diminished reality*. For example, the apparatus of Fig 3 may be used to experience colorblindness and reduced resolution by applying the appropriate 'visual filter'

Figure 3: 'Reality mediator' as of late 1994, showing a color stereo head-mounted display (VR4) with two cameras mounted to it. The inter-camera distance and field of view match approximately my interocular distance and field of view with the apparatus removed. The components around my waist comprise radio communications equipment (video transmitter and receiver). The antennas are located at the back of the head-mount to balance the weight of the cameras, so that the unit is not front-heavy.

to select the desired degree of degradation in a controlled manner that can also be automated by computer. (For example, color can be gradually reduced over the course of a day, under program control, to the extent that I can become color blinded but not even realize it.)

I mounted the cameras the correct interocular distance apart, and used cameras that had the same field of view as the display devices. With the cameras connected directly to the displays, the illusion of transparency will be realized to some degree, at least to the extent that each ray of light entering the apparatus (e.g. absorbed and quantified by the cameras) will appear to emerge at roughly the same angle (by virtue of the display).

Although I had no depth-from-focus capability there was, enough depth perception remaining on account of the stereo disparity for me to function somewhat normally with the apparatus. Depth-from-focus is what is sacrificed in working with a non-holographic RM.

A first step in using a reality mediator is to wear it for a while to become accustomed to its characteristics. Unlike in typical beam-splitter implementations of augmented reality, transparency, if desired, is synthesized, and therefore only as good as the components used to make the RM.

I wore the apparatus in identity map configuration (cameras connected directly to the displays) for several days. I could easily walk around the building, up and down stairs, through doorways, to and from the lab, etc. I did, however, experience difficulties in scenes of high dynamic range, and also in reading fine print (such as a restaurant menu or a department store receipt printed in faint ink when the ribbon was near the end of its useful life).

The unusual appearance of the apparatus was itself a hinderance in my daily activities (for example when I wore it to a formal dinner), but after some time people appeared to become accustomed to seeing me this way.

The attempt to create an illusion of transparency was itself a useful experiment because it established some working knowledge of what can be performed when vision is *diminished* or *degraded* to RS170 resolution and field of view is somewhat limited by the apparatus.

Knowing what can be performed when reality is mediated (e.g. diminished) through the limitations of a particular HMD (e.g. the VR4) would be useful to researchers who are designing VR environments for that HMD, because it establishes a sort of *upper bound* on "how good" a VR environment could ever hope to be when presented through that particular HMD. A reality mediator may also be useful to those who are really only interested in designing a traditional beam-splitter-based AR system because RM could be used as a development tool, and could also be used to explore new conceptual frameworks.

2.2 Mediated presence

McGreevy [McGreevy, 1992] explored the use of a "head-mounted camera/display/recorder system" that deprived the user of both color and stereo vision. Even though it had 2 cameras: "The lack of stereo vision provided in the head-mounted display prompted both subjects to use alternative cues to make spatial judgements [McGreevy, 1992]" (His subjects would move their heads back and forth to and perceive depth from the induced motion parralax.)

He referred to the experience as a "mediated presence".

His head-mounted camera/display system was a very simple form of reality mediator, where the mediation was fixed (e.g. not computer controllable).

McGreevy's work was important because it showed that despite the fact that his two subjects had essentially immediate response (essentially no delay) and also had the luxury of perfect (e.g. un-mediated) touch, hearing, smell, etc., they still had much difficulty adapting to the mediated visual reality in which they were placed. This showed that no matter how good a VR or telepresence simulation could be, there would still be significant limitations imposed by the interface between that simulation and the user – the HMD.

The system of Fig 3 overcame some of the problems associated with McGreevy's system – having 34 times more 'visual bandwidth' than McGreevy's system, it was just at the point where it was possible to conduct much of my daily life through this illusion of transparency. I could degrade my RM down to the level of McGreevy's system, and beyond, in a computer-controlled fashion to find out how much 'visual bandwidth' I needed to conduct my daily affairs. (The 'visual bandwidth' is calculated as the number of pixels times two for stereo, times another three for color, although one might argue that there is redundancy because left and right, as well as color channels are quite similar to one another.) I found, for various tasks, a certain point at which it was possible to function in the RM. In particular,

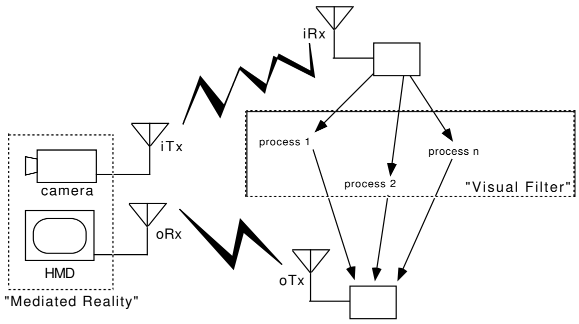


Figure 4: **Simple implementation of a 'reality mediator (RM)'.** The camera sends video to one or more more computer systems over a high-quality microwave communications link, which I refer to as the 'inbound channel'. The computer system(s) send back the processed image over a UHF communications link which I refer to as the 'outbound channel'. Note the designations "i" for inbound (e.g. iTx denotes inbound transmitter), and "o" for outbound. 'visual filter' refers to the process(es) that mediate(s) the visual reality and possibly insert "virtual" objects into the reality stream.

I found that anything below about 1/8 of my system's full bandwidth (about four times that of McGreevy's system) made most tasks very difficult, or impossible.

Once it is possible to live within the shortcomings of the RM's ability to be 'transparent', new and interesting experiments can be performed.

2.3 Video mediation

Once the apparatus is worn long enough to be comfortable with the 'illusory transparency', mediation of the reality can begin.

The compute-power required to perform general-purpose manipulation of color video streams is too unwieldy to be worn in a backpack (although I've constructed body-worn computers and other hardware to facilitate very limited forms of reality mediation). In particular, a system with good video-processing capability, such as Cheops [V. M. Bove and Watlington, 1995] or one or more SGI Reality Engines, may be used remotely by establishing a full-duplex video communications channel between the RM and the host computer(s). In particular, a high-quality communications link (which I call the 'inbound-channel') is used to send the video from my cameras to the remote computer(s), while a lower quality communications link (the 'outbound channel') is used to carry the processed signal from the computer back to my HMD. This apparatus is depicted in a simple diagram (Fig 4). Ideally both channels would be of high-quality, but the machine-vision algorithms were found to be much more susceptible to noise than was my own vision (e.g. I could still find my way around in a "noisy" reality, and still interact with "snowy" virtual objects).

To a very limited extent, looking through a camcorder provides a 'mediated reality' experience, because we see the real world (usually in black and white, or in color but with a very limited color fidelity) together with virtual text objects, such as shutter speed and other information about the camera. If, for example, the camcorder has a black and white viewfinder, the 'visual filter' (the colorblindness one experiences while looking through the viewfinder with the other eye closed) is unintentional in the sense that the manufacturer would rather have provided a full-color viewfinder. This is a very trivial example of a mediated reality environment where the filtering operation is *unintentional* but nevertheless present.

Although the colorblinding effect of looking through a camcorder may be undesirable most of the time, there are times when it is desirable. The 'diminished reality' it affords may be a desired artifact of the 'reality mediator' (for example in the case where the user chooses to remove color from the scene either to "tone-down" reality or to accentuate the perceptual differences between light and shade). This simple example points out the fact that a 'mediated reality' system need not function as just a 'reality enhancer', but rather, it may enhance, alter, or deliberately *degrade* reality.

Stuart Anstis [Anstis, 1992], using a camcorder that had a "negation" switch on the viewfinder, experimented with living in a "negated" world. He walked around holding the camcorder up to one eye, looking through it, and observed that he was unable to learn to recognize faces in a negated world. His negation experiment bore a similarity to Stratton's inversion experiment mentioned in Sec 1.2.2, but the important difference within the context of this paper is that Anstis experienced his mediated visual world through a video signal. In some sense both the regular eyeglasses that people commonly wear, as well as the special glasses researchers have used in prism adaptation experiments [Kohler, 1964] [Dolezal, 1982] are reality mediators, but it appears that Anstis was the first to explore, in detail, an electronically mediated world.

2.4 The Reconfigured Eyes

Using my 'reality mediator', I repeated the classic experiments like those of Stratton and Anstis (e.g. living in an upside-down or negated world), as well as some new experiments, such as learning to live in a world rotated 90 degrees. However, in this sideways world, I found that I could not adapt to having each of the images rotated by 90 degrees separately, but had to rotate the cameras together (Fig 5).

The video-based RM (e.g. Fig 3) permits me to experience any coordinate transformation that can be expressed as a mapping from a 2D domain to a 2D range, in real time (30frames/sec = 60fields/sec) in full color, because a full-size remote computer (e.g. SGI Reality Engine) is used to perform the coordinate transformations. This apparatus allows me to experiment with various computationally-generated coordinate transformations both indoors and outdoors, in a variety of different practical situations. Examples of some useful coordinate transformations appear in Fig 6.

Researchers at Johns Hopkins University have been experimenting with the use of cameras and head-mounted displays for helping the visually handicapped. Their approach has been to use the optics of the cameras for magnification, together with the contrast adjustments of the video display to increase apparent scene contrast [jhu, 1995]. They also talk about using *image remapping* in the future:

(a) (b)

Figure 6: **Living in coordinate-transformed worlds:** Color video images are transmitted, coordinate-transformed, and then received back at 30 frames per second – the full frame-rate of the VR4 display device. (a) This 'visual filter' would allow a person with very poor vision to read (due to the central portion of the visual field being hyper-foveated for a very high degree of magnification in this area), yet still have good peripheral vision (due to a wide visual field of view arising from demagnified periphery). (b) This 'visual filter' would allow a person with a *scotoma* (a blind or dark spot in the visual field) to see more clearly, once having learned the mapping. The visual filter also provides edge enhancement in addition the coordinate transformation. Note the distortion in the cobblestones on the ground and the outdoor stone sculptures.

Figure 5: **Living in a "Rot 90" world:** It was found to be necessary to rotate both the cameras rather than just rotate each one. Thus, it would not seem possible to fully adapt to, say, a prism that rotated the image of each eye, but the use of cameras allows the up-down placement of the "eyes". The parallax, now in the up-down direction, affords a similar sense depth as we normally experience with eyes spaced from left to right together with left-right parallax.

> One of the most exciting developments in the field of low vision is the Low Vision Enhancement System (LVES). This is an electronic vision enhancement system that provides contrast enhancement... *Future enhancements* to the device include text manipulation, autofocus and *image remapping*.

(quote from their WWW page [jhu, 1995], emphasis added). This research effort suggests the utility of the real-time visual mappings (Fig 6) successfully implemented using the apparatus of Fig 3.

The idea of living in a coordinate transformed world has been explored extensively by other authors [Kohler, 1964] [Dolezal, 1982], using optical methods (such as prisms and the like). Much could be written about my experiences in various electronically coordinate transformed worlds, but a detailed account of all of the various experiences is beyond the scope of this paper. Of note, however, I observed that visual filters differing slightly from the identity (e.g. rotation by a few degrees) had a more lasting impression on me when I removed my apparatus (e.g. left me incapacitated for a greater time period upon removal of the apparatus), than visual filters that were far from the identity (e.g. rotation by 180 degrees – upside-down). Furthermore, the visual filters close to the identity tended to leave me with an opposite aftereffect (e.g. I'd consistently reach too high after taking off the RM where the images had been translated down slightly, or reach too far 'clockwise' after removing the RM that had been rotating images a few degrees counterclockwise). Visual filters far from the identity (such as reversal or upside-down mappings) did not leave me with an opposite aftereffect: I would **not** see the world as being upside down upon removing upside-down glasses. I think of this phenomenon as being analogous to learning a second language (either a natural language or computer language). When the second language is similar to the one we already know, we make more mistakes switching back and forth than when the two are distinct. When two (or more) adaptation spaces were distinct, for example, in the case of the identity map and the rotation operation ('rot 90'), I could sustain a dual adaptation space and switch back and forth between the 'portrait' orientation of the identity operator and and 'landscape' orientation of the 'rot 90' operator without one causing lasting aftereffects in the other.

Regardless of how much care is taken in creating the illusion of transparency, there will be a variety of flaws, not the least of which is limited resolution, lack of dynamic range, limited color (mapping from the full spectrum of visible light to three responses of limited color gamut), and improper alignment and placement of the cameras. In Fig 3, for example, the cameras are mounted *above* the eyes. Even if they are mounted in front of the eyes, they will extend, putting me in the visual world of some hypothetical organism that has eyes that stick out of its head some 3 or 4 inches (except in the case of a blind person in the future when sufficient technological advances permit using cameras for artificial eyes). Thus some adaptation is almost always needed.

After wearing my apparatus for an extended period of time, I eventually adapted, despite its flaws, whether these be unintended (e.g. limited dynamic range, limited color gamut, etc.), or intended (e.g. deliberately presenting myself with an upside-down image). In some sense I subsume the visual reconfiguration induced by the apparatus into my brain, so that, in a sense, the apparatus and I act as a single unit. Manfred Clynes uses the example of a person riding a bicycle to describe this sort of synergism [Clynes, 1960] where, after sufficient adaptation time, conscious effort is no longer needed in order to use the machine. He refers to this state as *cyborgian*. In some sense, after adapting to the RM, one becomes a *cyborg*.

2.4.1 Giant's Eyes

I found that having the cameras above the display (as in Fig 3) induced some parallax error for nearby objects, so I tried mounting the cameras at the sides of my head (Fig 7(a)). This gave me an interocular distance of approximately 212mm, resulting in an enhanced sense of depth. Objects appeared smaller and closer than they really were – the world looked like a size-reduced scale-model of reality. While walking home that day (wearing the apparatus), I felt that I had to duck down to avoid hitting what appeared to be a low tree branch. However, my recollection from previous walks home had been that there were no low branches on the tree, and, removing my RM, I noticed that the tree branch that appeared to be within arm's reach was several feet in the air. After some time I got used to this enhanced depth perception, and then tried mounting the cameras on a 1 meter baseline. Crossing the street, I had the illusion of small toy cars moving back and forth very close to my nose, and I had the feeling that I could just push them out of my way, but my better judgement served to make me wait until there was a clearing in the traffic before crossing the road to get to the river. Looking out across the river, I had the illusion that the skyscrapers on the other side were within my arm's reach in both distance and height.

2.4.2 'Slowglasses'

Suppose that we had a hypothetical glass of very high refractive index. (Science fiction writer Bob Shaw refers to such glass as *slowglass* [Shaw, 1966]. In Shaw's story, a murder is committed and a piece of slowglass is found at the scene of the crime – the glass being turned around as curious onlookers wait for the light present during the crime to emerge from the other side.) Every ray of light that enters one side of the glass comes out the other side unchanged, but simply delayed. A visor made from *slowglass* would present the viewer with a full-parallax delayed view of a particular scene, playing back with the realism of the idealized *holovideo* display discussed in Sec 1.2.

A practical (non-holographic) implementation of this illusion of *delayed transparency* was created using the reality mediator (Fig 3) with a video delay.

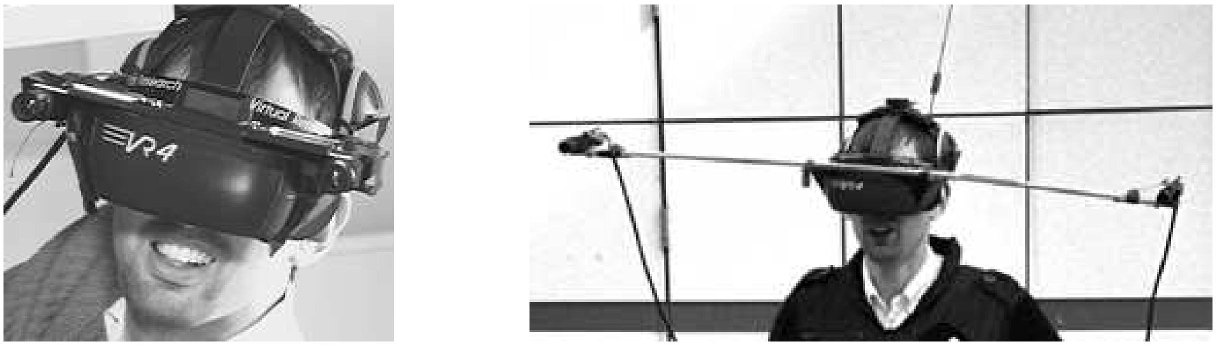
As is found in any poor simulation of virtual reality, wearing 'slowglasses' induces a similar dizziness and nausea to reality. After experimenting with various delays one will develop an appreciation of the importance of moving the information through the RM in a timely fashion to avoid this unpleasant delay.

2.4.3 'Edgertonian' Eyes

Instead of a fixed delay of the video signal, I experimented by applying a repeating freeze-frame effect to it (with the cameras' own shutters set to 1/10000 second). With this video *sample and hold*, I found that nearly periodic patterns would appear to freeze at certain speeds. For example, while looking out the window of a car, periodic railings that were a complete blur without my RM would snap into sharp focus with the RM. Slight differences in each strut of the railing would create interesting patterns that would dance about revealing slight irregularities in the structure. (Regarding the nearly periodic structure as a true periodic signal plus noise, the noise is what gave rise to the interesting patterns). Looking out at another car, traveling at approximately the same speed as me, I could read the writing on the tires, and easily count the number of bolts on the wheel rims. Looking at airplanes, I could see the number of blades on the spinning propellers, and, depending on the sampling rate of my RM, the blades would appear to rotate slowly backwards or forwards, in much the same way as objects do under the stroboscopic lights of Harold Edgerton [Edgerton, 1979]. By manually adjusting the processing parameters of my RM, I could see many things that escape normal vision.

2.4.4 Virtual 'smart strobe'

By applying machine vision (some rudimentary intelligence) to the incoming video, the RM should be able to decide what sampling rate to apply. For example, it should recognize a nearly periodic or cyclostationary signal and adjust the sampling rate to lock onto the signal much like a phase-locked loop. A sufficiently advanced RM with eye tracking and other sensors might make inferences about what you'd like to see, and, for example, when looking at a group of airplanes in flight would freeze the propeller on the one you were concentrating on.



(a) (b)

Figure 7: **Giant's eyes: extended baseline**. (a) With a 212mm baseline, I could function in most everyday tasks, but would see crosseyed at close conversational distances. (b) With a 1m baseline, I could not function in most situations, but had a greatly enhanced sense of depth for distant objects. Wires from the cameras go down into my waist bag containing the rest of the apparatus. Inbound transmit antenna is just visible behind my head.

2.4.5 Wyckoff's world

One of the problems with the RM is the limited dynamic range of CCDs. One possible solution is to operate at a higher frame rate than needed, while underexposing, say, odd frames and overexposing even frames. The shadow detail may then be derived from the overexposed stream, the highlight detail from the underexposed stream, and the midtones from a combination of the two streams. The resulting extended-response video may be displayed on a conventional HMD by using Stockham's *homomorphic filter* [T. G. Stockham, Jr., 1972] as the 'visual filter'. The principle of extending dynamic range by combining differently exposed pictures is known as the Wyckoff principle [Mann and Picard, 1994a], in honor of Charles Wyckoff. Using a Wyckoff composite, I could be outside on bright sunny days and see shadow detail when I looked into open doorways to dark interiours, as well as see detail in bright objects like the sun.

The Wyckoff principle is also useful in the context of night vision because of the high contrasts encountered at night. In the Wyckoff world, one can read rating numbers printed on a bright mercury vapor arc lamp and also see into the darkness off in the distance behind the lamp, neither brightness extreme of which is visible to the naked eye.

2.5 Conclusion of Sec 2

With high-quality cameras and display devices, the illusion of transparency was found to be sufficiently good that I was able to function comfortably in many of my day-to-day tasks. Further improvements in the technology suggest the possibility for creating an even more comfortable illusion of transparency. Once this illusion is assimilated and accepted, for all practical purposes, the cameras have replaced the functionality of the eyes, except now signals may be both extracted from them, and inserted into where they were connected. Furthermore, the signal path may now be conveniently interrupted so that a 'visual filter' may be installed, to some degree, behaving as though it were positioned between the eye and the brain.

3 Partially mediated reality

Artificial Reality is a term defined by Myron Krueger to describe video-based, computer-mediated interactive media [Earnshaw et al., 1993]. His apparatus consisted of a video display (screen) with a camera above it that projected a 2D outline of the user together with sprite-like objects. Myron Krueger's environment is a partially mediated reality, in the sense that within the screen the reality is mediated, but the user is also free to look around the room and see unmediated real objects. For example, the part of the user's visual field that shows his or her "reflection" (a left-right reversed video image of the camera is superimposed with computer graphic objects) is a mediated-reality zone, while the periphery (e.g. the user's own feet which can be seen by looking straight down) is outside this mediation zone.

The Artificial Life Interactive Video Environment (ALIVE) [Maes et al.,] is similar to Myron Krueger's environment. In the ALIVE, a user sees him/her self in a "magic mirror" created by displaying a left-right reversed video image from a camera above the screen. Virtual objects, appear, for example, a virtual dog will come over and greet the user. ALIVE is also a partially mediated reality.

3.1 Monocular mediation

A camera and a display device completely covering only one eye (Fig 8(a)) can be used to create a partially mediated reality. In the apparatus of Fig 8(a) my right eye sees a green image (processed NTSC on a VGA display) which becomes fused with the unobstructed (full-color) view through my left eye.

Often the mediated and unmediated zones are in poor register and I cannot fuse them. The poor register may even be deliberate, e.g. I often like to have my right eye in a rotated ('rot 90') world even though this means that I cannot see in stereo in a meaningful way. However, I can still switch my concentration back and forth. I am able to selectively decide to concentrate on one or the other of these two worlds.

An RM made from a camera and a Virtual Vision television set permits a mediation of even lesser scope to take place. Not only does it play into just one eye, but

(a) (b)

Figure 8: **Partially mediated reality:** (a) Half MR: My right eye is *completely* immersed in a mediated reality environment arising from a camera on my right, while my left eye is free to see unmediated real-world objects. (b) Substantially less than half MR: My left eye is *partially* immersed in a mediated reality environment arising from a camera also on my left.

the field of view of the display only covers part of that eye (Fig 8(b)). The visor is transparent so that both eyes can see the real world (although my left eye is partially blocked). With these glasses, I might see an object with both eyes, through the transparent visor, and then look over to the 'mediation zone' where my left eye sees, "through" the illusion of transparency in the display. Again, I can switch my attention back and forth between the mediated reality and ordinary vision. I see a double-vision effect (e.g. when I look at someone's face through the glasses of Fig 8(b), I often see two replicas of their face, the one that is mediated, and the one that is not). This doubling effect, due to imperfect registration between the mediated and unmediated zones, may or may not be a problem depending on how the RM is used. For example, if I present the mediated world as grey, it remains distinct from the unmediated world, and I am able to mentally switch back and forth between seeing directly, and living in the mediated world, even though the two overlap almost exactly. I often even have the camera present the images in 'rot 90' and then switch my concentration back and forth, having a dual adaptation space. Thus, depending on the application or intent, there may be desire to register or to deliberately misregister the possibly overlapping direct and mediated zones.

4 Seeing 'eye-to-eye'

With two reality mediators of the kind depicted in Fig 8(b), I would set the output frequency of one to the input frequency of the other, and vice versa, so that someone else would see through my eyes and me through the other person's eyes. The Virtual Vision glasses allowed me to concentrate mainly on what was in my own visual field of view (because of the transparent visor), but at the same time have a general awareness of the other persons's visual field. This 'seeing eye-to-eye' as I called it, allowed for an interesting form of collaboration. Seeing eye-to-eye through the apparatus of Fig 3 requires a *picture in picture* process (unless one wishes to endure the nauseating experience of looking *only* through the other person's eyes), usually having the wearer's own view occupy most of the space, while using the apparatus of Fig 8(b) does not require any processing at all.

Usually when we communicate (e.g. by voice or video) we expect the message to be received and concentrated on, while when 'seeing eye-to-eye' there is not the expectation that the message will *always* be seen by the other person. Serendipity was the idea, where each of us would sometimes pay attention and sometimes not.

4.1 'Safety net'

Now suppose that instead of just two people, we have a community (network) of individuals wearing RMs. In some sense these people pay attention mostly to their immediate surroundings, but may, at times, get an image from someone who thinks there might be danger. This fear of danger might be triggered by a 'maybe I'm in distress' button pressed by the wearer, or automatically (e.g. by a heart rate monitor and activity meter such as a pedometer, where the heart rate divided by the physical activity gives a 'non–exertion–arousal' index). A community of individuals networked in this way would look out for each others' safety in the form of a 'neighbourhood watch'. This 'safety net' could be used for a 'virtual safewalk': a participant, about to walk home or enter an underground parking garage late at night, sees 'eye-to-eye' with one or more people (perhaps in a different time zone, say somewhere in the world where it is morning, so the virtual escort has fresh alert eyes).

4.2 On the "safety versus privacy" argument

The ease with which wearable wireless video cameras allow one to roam about and share viewpoints with others raises many privacy issues [Mann, 1994b], and it is important to look at these issues within the broader context of video privacy in general.

When I first joined the Media Lab, I expressed concern regarding the possible development of surveillance technologies, such as ubiquitous use of video cameras, face recognition and the like. My advisor, trying to relieve my concerns regarding a possible Big-Brother future, presented me with the argument of her advisor (Sandy Pentland) who was the director of the research on face recognition:

> Cameras make the world a smaller place, kind of like a small town. You give up privacy in exchange for safety. In a small town, if you were suffering from a heart attack and collapsed on the floor of your kitchen, chances are better that someone would come to your rescue. Perhaps a neighbour would come over to borrow some sugar, and, since your door would be unlocked, would just come right in and see you had collapsed and come to your aid.

Although this analogy makes perfect logical sense, there was something that bothered me about it: On the *safety versus privacy* axis, the small town of the past and the Orwellian future I feared are very similar. However, if we look along a different dimension, characterized by symmetry, the small town and the Orwellian future are exact opposites. In a small town, the sheriff knows what everyone's up to, but everyone also knows what the sheriff is up to.

Phil Patton [Patton, 1995] discusses the surveillance dilemma, making reference to the ubiquitous "ceiling domes of wine-dark opacity", making mention that "many department stores use hidden cameras behind one-way mirrors in fitting rooms", and in general, that there is much more video surveillance than we might at first think. Sheraton's use of hidden cameras in employee changerooms [Hancock et al., 1995] takes surveillance to new heights. The number of companies selling devices with hidden cameras inside (for example, smoke-detector cameras, fire sprinkler cameras, exit-sign cameras, etc) is growing rapidly.

With so much video surveillance in place, and growing at a tremendous rate, one wonders if privacy is a lost cause. If we are going to be under video surveillance, we may as well keep our own "memory" of the events around us, analogous to a contract in which both parties keep a signed copy. Falsification of video surveillance recordings is a point addressed in the movie *Rising Sun*, and in William Mitchell's book, *The Reconfigured Eye* [Mitchell, 1992]. However, if there is a chance that individuals might have their own account of what happened, organizations using surveillance would not even consider falsifying surveillance data. Even though it is easy to falsify images [Mitchell, 1992], when accounts of what happened differ, further investigation would be called for. Careful analysis (e.g. kinematic constraints on moving objects in the scene, the way shadows reflect in shiny surfaces, etc) of two or more differing accounts of what happened would likely uncover falsification that would otherwise remain unnoticed. The same technology that is used to demonstrate a person has removed an item from a department store without paying may be used by a person to demonstrate that he or she did, in fact, pay. One can only imagine what would have happened if the only video recording of the Rodney King beating were one that had been made by by police, using a police surveillance camera. Of course, most officials are honest, and would have no reason to be any more paranoid of the proposed virtual small-town than of the Orwellian world we might otherwise be heading towards.

4.3 Conclusion of Sec 4

'Safety net' offers an alternative to video surveillance. It suggests a future in which people, through prosthesis, might have both improved visual memory and improved ability to share it. But it also suggests a hope that the visual memory be distributed among us, and be less likely to be abused than if it exist in a centralized form, as is more common with a network of surveillance cameras, such as is commonly used on the streets in the UK. The proliferation of hidden cameras everywhere has the possibility to threaten our privacy, but suppose the only cameras were the prosthetic elements of other individuals. Then at least one would still have privacy when one was alone.

5 People looking at...

ALIVE (mentioned in Sec 3) is part of a larger research effort of Pentland's group at the MIT Media Lab, entitled *Looking at People*. Fixed cameras are pointed at people. It is interesting to consider what happens when this paradigm is reversed, and instead, people *wear* cameras.

This of course raises some more complicated issues, because the camera is no longer fixed in space. In order to make sense of a wearable camera, one might first consider *image stabilization*. Image motion may be resolved into two components, that due to parallax (e.g. moving from one place to another), and that due to rotation of the camera about its center of projection.

I find that I can quickly twist my neck and induce quite a large image motion, primarily due to rotation of the camera about its center of projection[1]. To induce the same amount of image motion by pure translation of the camera would be difficult or impossible, for most scenes, as I would need to expend much more energy to obtain the same amount of image motion by parallax as I do by rotating my neck. Thus the simple physics of the RM suggests that images can be dramatically stabilized using only a homographic (projective) coordinate transformation.

A typical image sequence from the RM (e.g. from the camera, which, through adaptation, is functioning for all practical purposes as my eye) appears in Fig 9. Note that the images are rotated 90 degrees because I had spent the previous week or so living in a "rot 90" world, adapting to see in this world (as Stratton did with the upside-down world). As one can see by the quick but careful composition of each successive frame of video (e.g. the camera encompasses the important parts of the building), I had fully adapted to the "rot90" world, and, in a sense, become one with this particular instance of the apparatus.

All of the images may be brought into a single coordinate system [Mann and Picard, 1994c] (Fig 10). Once the images are stabilized, they can be assembled together to make an *image mosaic* (Fig 11).

The *image mosaicing* principle (e.g. figures like Fig 11) has been previously explored [Hirose et al., 1994] [Mann and Picard, 1994b] [Szeliski and Coughlan, 1994].

Of particular interest is the work using *image mosaicing* in a virtual reality environment [Hirose et al., 1994] [HIROSE, 1994]. The system depicted in Fig 3 allows me to explore the video mosaics of Hirose as I walk around outdoors, in my day to day activities.

Because the wireless apparatus allows me to use the remote compute-power of a large number (sometimes as many as 20 or 30) of state-of-the-art workstations, I can interact with the world around me by simply turning my head, "painting with looks". Head rotation is tracked using the featureless *video orbits* method[Mann and Picard, 1994b]. This interactive video environment allows me to explore the world in new and interesting ways, and to see how well different views of, say, a building, will fit together into an image mosaic. I can, for example, live in a no-AGC world, and then switch on the AGC and obtain almost immediate feedback on how the AGC affects the final "painting".

I use my apparatus as an artist's sketch pad of sorts, useful for taking down visual "notes", and helping me overcome my memory disability. Thad Starner [Starner, 1995] uses his wearable computer with a program he calls a *rem-*

[1] a small amount is also due to parallax since the camera is mounted out away from the exact point of rotation of my neck

(a) (b) (c) (d) (e)

Figure 9: Video from my reality mediator (through which I had been seeing the world for an extended period of time). In this case, I had adapted, for several days, to the 'rot90' (sideways) world. Video frames shown here are 'unrotated' back, hence the tall (1 : 0.75) aspect ratio.

(a) (b) (c) (d) (e)

Figure 10: Images after a coordinate transformation to bring them into register with frame (c). The coordinate-transformed images are alike except for the region over which they are defined.

Figure 11: **Living in a motion-stabilized world:** Frames of Fig 10 "cemented" together on single image "canvas" may be transmitted back to the RM.

brance agent[Starner, 1993]. This program continually runs in the background and helps him remember text that he has previously typed. It is not hard to imagine a visual remembrance agent. Being someone who had trouble remembering faces, supplementing part of the brain with computer memory accessible on the Internet, there is no reason why one should ever need to forget a face. In addition to video input on my wearable apparatus, I have a variety of other devices, such as biosensors. With my biosensors, I hope it will be possible to have a visual *rembrance agent* that has an awareness of my *affective state* [Picard, 1995] and operates without conscious thought or effort [Clynes, 1960].

6 Wearable Interactive Video Environment (WIVE)

6.1 Drawing in the air

Video environments like Myron Krueger's and the ALIVE are useful because they recognizes the user's gestures. Similarly, the RM can be used to allow my body-worn computer[2] to recognize my own jestures. For example, I might draw in free space (Fig 12), using my finger as a mouse to outline actual objects in the scene. In order to track my finger, I attach a small IR LED so that it will be brighter than anything else in view and then threshold the images to obtain a cluster of pixels that corresponds to my finger (the pointing device), or I attach some colored tape whose color is unique from the background. Because I am drawing right on top of the video stream, registration is, for all practical purposes, exact to within the pixel resolution of the devices. In this work, the apparatus used differs from that of Fig 3. In particular, there is only one channel instead of two, because the goal is to annotate images with no regard to depth.

[2]The compute power is remote but in a virtual sense is worn on my body via the full-duplex video communications link.

6.2 Equipment repair

In a current collaborative project with Thad Starner, a wearable system is being developed (Fig 13) to allow a technician to repair a piece of equipment and see a computer graphics repair manual with drawings superimposed on the real world. Because of the ease with which exact registration is possible using active light sources or the like, the registration problem may be solved at the video signal level, resulting in an environment where the real and virtual worlds fuse together as one.

The approach presented here might also be useful within the context of work done by other researchers, such as Knowledge-based Augmented Reality for Maintenance Assistance (KARMA) [Feiner et al., 1993b] [Feiner et al., 1993a], seeing architectural anatomy of buildings [Feiner et al., 1995], and combining ultrasound with virtual reality in obstetrics [Fuchs et al.,].

(a) (b) (c)

Figure 12: **Drawing in the air:** (a) Monocular reality mediator used in fingertracking. The video signal from the camera is fed to the short antenna (*inbound transmit* channel, operating at microwave frequencies), while the *outbound receive* signal (UHF frequencies) arrives via the longer antenna and is fed to my right-eyed display (modified Virtual Vision system using a high-resolution CRT instead of the original LCD). (b) View of apparatus (note copper mesh cap on my head acting as ground plane for the antennae) and object being outlined (luxo lamp). (c) What I see through the glasses: red outline of the object that I made by moving my finger around in the space between the camera (taking the role of my eye) and the object. The rest of the scene is displayed grey so distinct colors may be reserved for virtual objects.

Figure 13: Equipment repair in a 'mediated-reality' environment. In this case, a laser printer is being serviced; note the colored tape installed in the printer and on my fingertip.

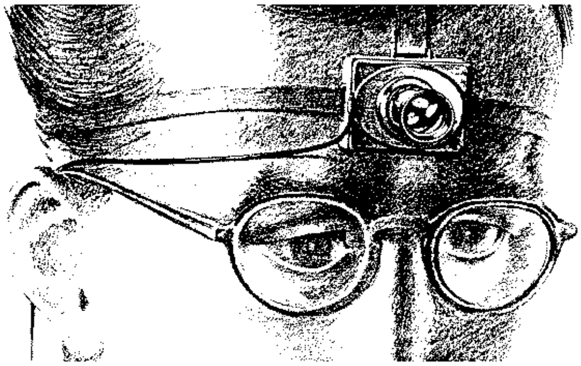


Figure 14: Vannevar Bush's camera: original caption reads "A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBJECT"

7 A new cinematographic reality

In 1945, Vannevar Bush described a wearable camera (Fig 14) that would record whatever the wearer was looking at onto microfilm [Bush, 1945].

In many ways Bush's proposed camera is similar to so-called point-of-view (POV) cameras, that are used extensively in sports (e.g. mounted in the helmet of a football player), as well as the headcams of David Letterman (e.g. "monkeycam"), or the hidden body-mounted cameras used in TV shows like *60 minutes*, as well as by individuals who are trying to protect themselves from harassment or from false harassment charges.

Recording the output of my RM gives rise to an interesting method of cinematography, similar in some ways to the cameras mentioned above, but also quite distinct in other ways. The obvious difference is that if my camera is set wrong, I will not see properly, and will trip and fall.

The traditional body-mounted cameras do not provide exactly the same field of view that the wearer experiences. Bush attempted to address this problem by proposing that a small square in the eyeglass would be used to sight the camera. However this small square is really not in the spirit of MR, and so does not close the loop through the wearer. In addition to imprecise sighting as the glasses slide around, if the exposure were wrong, the wearer would not be immediately aware. Hence there is a need to depend on an automatic gain control which is seldom as good as having human in the loop. Using the RM, however, puts the wearer in the loop because it acts as *both* a recording and a seeing device — imperfect adjustment of the camera unfavorably mediates the wearer's vision in a way that causes him or her to adjust it toward a more optimal setting.

Once I have worn my RM for some time, and become fully accustomed to experiencing the world through it, there is a certain synergy between me and the machine that is not experienced with a head-mounted camera alone. More subtle differences between a recording made from the output of an RM and that made from a conventional body worn camera include the way that when I am talking to two people the closing of the loop forces me to turn my head back and forth as I talk to one person, and then to the other. This need arises from my limited peripheral vision.

After wearing the apparatus for some time, I learn how to compensate for defficiencies such as limited peripheral vision and limited dynamic range. The resulting video [Mann, 1994b] is much more like having extracted a signal from my eye than the signal arising from the traditional point-of-view methods, because I live with the RM over an extended period of time and, after time, it begins to behave as though it truly were an extension of my own body. What others see is exactly what I see, no more or no less. In some sense I've become what Manfred Clynes terms a *cyborg*.

8 Conclusions

A simple implementation of a 'reality mediator' (RM) using camera(s) and head mounted display has been presented. The implications of an extremely powerful body-worn computer, as might be available in years to come, were explored using a full duplex video communications link sending the video from the camera(s) to one or more remote computers via a microwave communications link, and viewing the result on the computer's screen remotely via a UHF link.

Many different aspects and implications of mediated reality (MR) were explored. A reality mediator can be used as an artist's tool, allowing him or her to 'see the world in a different light' – experiencing a mediated world where awareness of light and shade is increased, or an Edgertonian world where spatio-temporal periodicity and cyclostationarity can be explored, or the world of image mosaics where the movement of the head acts as a paintbrush sweeping out seamless renditions of Hockney's photographic collages of a single subject. The RM, perhaps acting as a front-end to programs like Adobe Photoshop, lets the artist manipulate images by outlining objects with the tip of the finger. The artist, in a sense, lives inside the video space he/she creates. The RM also allows us to share our mediated experiences with others, to see 'eye-to-eye' and to form a network of individuals looking out for one another's safety. With the RM, the visually handicapped who still have a small amount of sight remaining could see better; those with memory disabilities might rely on external 'visual memory'. MR allows us to reconfigure our visual reality in new, useful, and interesting ways.

While a full practical implementation of MR is some years away, current implementations could be useful in specific domains. In particular, in applications where registration is extremely critical yet can be solved at the signal level, or where it is desirable to be able to alter as well as augment reality, MR has been shown to have great promise.

9 Acknowledgements

I'd like to thank Thad Starner for use of the tracking software and helping me get it and the generalized 'visual filter' software running with my reality mediator. Also thanks to Chuck Oman for pointing out references [Dolezal, 1982] and [Kohler, 1964]. Thanks to Ted Adelson, Roz Picard, Neil Gershenfeld, Sandy Pentland, and Jennifer Healey for many useful discussions, to Kris Popat for making the realization that 'lightspace' glass might, in principle endow one with invisibility, and to Matt Reynolds (KB2ACE) for help in an improvement to the outbound ATV system. Thanks also to VirtualVision, Virtual Research, Ed Gritz, BelTronics, and Compaq for lending or donating equipment that made my experiments possible.

References

[jhu, 1995] (1995). Lions vision research and rehabilitation center. http://www.wilmer.jhu.edu/low_vis/low_vis.htm.

[Anstis, 1992] Anstis, S. (1992). Visual adaptation to a negative, brightness-reversed world: some preliminary observations. In Carpenter, G. and Grossberg, S., editors, *Neural Networks for Vision and Image Processing*, pages 1–15. MIT Press.

[Arfken, 1985] Arfken, G. (1985). *Mathematical Methods for Physicists*. Academic Press, Orlando, Florida, third edition.

[Azuma, 1994] Azuma, R. (1994). Registration Errors in Augmented Reality: NSF/ARPA Science and Technology Center for Computer Graphics and Scientific Visualization . http://www.cs.unc.edu/~azuma/azuma_AR.html.

[Bush, 1945] Bush, V. (1945). As we may think. *Atlantic Monthly*. http://www2.theatlantic.com/atlantic/atlweb/flashbks/computer/bushf.htm.

[Clynes, 1960] Clynes, M. (September 1960). Cyborgs and space. *Astronautics, pages 26,27,and 74-75*, pages 26,27,and 74–75.

[Dolezal, 1982] Dolezal, H. (1982). *Living in a world transformed*. Academic press series in cognition and perception. Academic press, Chicago, Illinois.

[Drascic, 1993] Drascic, D. (1993). David drascic's papers and presentations. http://vered.rose.utoronto.ca/people/david_dir/Bibliography.html.

[Earnshaw et al., 1993] Earnshaw, R. A., Gigante, M. A., and Jones, H. (1993). *Virtual reality systems.* Academic press.

[Edgerton, 1979] Edgerton, H. E. (1979). *Electronic flash, strobe.* MIT Press, Cambridge, Massachusetts.

[Feiner et al., 1993a] Feiner, MacIntyre, and Seligmann (1993a). Karma (knowledge-based augmented reality for maintenance assistance). http://www.cs.columbia.edu/graphics/projects/karma/karma.html.

[Feiner et al., 1993b] Feiner, MacIntyre, and Seligmann (Jul 1993b). Knowledge-based augmented reality. Communications of the ACM, 36(7).

[Feiner et al., 1995] Feiner, Webster, Krueger, MacIntyre, and Keller (1995). Architectural anatomy. Presence, 4(3), 318-325.

[Fuchs et al.,] Fuchs, H., Bajura, M., and Ohbuchi, R. Teaming ultrasound data with virtual reality in obstetrics. http://www.ncsa.uiuc.edu/Pubs/MetaCenter/SciHi93/1c.Highlights-BiologyC.html.

[Hancock et al., 1995] Hancock, L., Kalb, C., and Underhill, W. (July 17, 1995). You don't have to smile. *Newsweek.*

[Haykin, 1983] Haykin, S. (1983). *Communication Systems.* Wiley, second edition.

[Hilaire et al., 1990] Hilaire, P. S., Benton, S., and Lucente, M. (1990). Electronic display system for computational holography. In *SPIE Proceedings #1212 "Practical holography IV"*, pages 174–182.

[HIROSE, 1994] HIROSE, M. (1994). Welcome to our hirose laboratory !! http://ghidorah.t.u-tokyo.ac.jp/index.html.

[Hirose et al., 1994] Hirose, M., Takahashi, K., Koshizuka, T., and Watanabe, Y. (1994). A study on image editing technology for synthetic sensation. ICAT '94 Proceedings, pp.63-70.

[Kohler, 1964] Kohler, I. (1964). *The formation and transformation of the perceptual world*, volume 3 of *Pshchological issues.* International university press, 227 West 13 Street. monograph 12.

[Lucente et al., 1992] Lucente, M., Hilaire, P. S., and Benton, S. (1992). A new approach to holographic video. *SPIE Proceedings #1732 "Holography '92".*

[Maes et al.,] Maes, Darrell, Blumberg, and Pentland. The alive system: Full-body interaction with animated autonomous agents. TR 257, M.I.T. Media Lab Perceptual Computing Section, Cambridge, Ma.

[Mann, 1994a] Mann, S. (1994a). Lightspace: Recording varying shadows and highlights that reproduce with same varying illumination during reconstruction. Technical Report 348, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts.

[Mann, 1994b] Mann, S. (1994b). Personal WWW page: 'put yourself in my shoes and see the world from my perspective'. http://wearcam.org.

[Mann and Picard, 1994a] Mann, S. and Picard, R. (1994a). Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. Technical Report 323, M.I.T. Media Lab Perceptual Computing Section, Boston, Massachusetts. Also appears, IS&T's 46th annual conference, May 1995.

[Mann and Picard, 1994b] Mann, S. and Picard, R. W. (1994b). 'virtual bellows': Assembling video into high quality still images. TR 259, M.I.T. Media Lab Perceptual Computing Section, Cambridge, Ma.

[Mann and Picard, 1994c] Mann, S. and Picard, R. W. (1994c). Virtual bellows: constructing high-quality images from video. In *Proceedings of the IEEE first international conference on image processing*, Austin, Texas.

[McGreevy, 1992] McGreevy, M. W. (1992). The presence of field geologists in mars-like terrain. *PRESENCE*, 1(4):375–403. MIT Press.

[Mitchell, 1992] Mitchell, W. J. (1992). *The Reconfigured Eye.* The MIT Press.

[Nagao, 1995] Nagao, K. (1995). Ubiquitous talker: Spoken language interaction with real world objects. http://www.csl.sony.co.jp/person/nagao.html.

[Patton, 1995] Patton, P. (1995). Caught. *WIRED.*

[Picard, 1995] Picard, R. W. (1995). Affective computing. TR 321, MIT Media Lab.

[Shaw, 1966] Shaw, B. (1966). *Light of Other Days.* Analog.

[Starner, 1993] Starner, T. (1993). The remembrance agent. Class project for intelligent software agents class of Patti Maes.

[Starner, 1995] Starner, T. (1995). The cyborgs are coming or the real personal computers. http://www-white.media.mit.edu/vismod/publications/tech_reports/abstracts/TR-318-ABSTRACT.html.

[Stratton, 1896] Stratton, G. M. (1896). Some preliminary experiements on vision. *Psychological Review.*

[Sutherland, 1968] Sutherland, I. (1968). A head-mounted three dimensional display. In *Proc. Fall Joint Computer Conference*, pages 757–764.

[Szeliski and Coughlan, 1994] Szeliski, R. and Coughlan, J. (1994). Hierarchical spline-based image registration. *CVPR.*

[T. G. Stockham, Jr., 1972] T. G. Stockham, Jr. (1972). Image processing in the context of a visual model. *Proc. IEEE*, 60(7):828–842.

[V. M. Bove and Watlington, 1995] V. M. Bove, J. and Watlington, J. A. (1995). Cheops: A reconfigurable data-flow system for video processing. *IEEE Transactions on Circuits and Systems for Video Processing*, 5:140–149.