Headmounted Wireless Video: Computer-Supported Collaboration for Photojournalism and Everyday Use

Steve Mann, University of Toronto

ABSTRACT Traditionally, video has been either part of the environment, such as video surveillance cameras mounted on or inside a building or video conferencing systems based on fixed cameras within a special room, or the domain of large organizations such as broadcast television stations. Recently, however, a new field of research called "personal imaging" has emerged. Personal Imaging systems are based on wireless video technology, and are typically characterized by video from a first-person perspective by way of a head-mounted camera and display together with an image processing computer worn on the body of the user. The possibilities afforded by personal imaging include a personal safety device for crime reduction, a new kind of videoconferencing system for computer-supported collaboration, as well as a new tool for photojournalism. This article briefly describes work in personal imaging as it has evolved over the past 20 years, and then attempts to set forth a future vision for wireless video in a head-mounted context. Most notably, the notion of computer-supported collaborative wireless video is presented.

Personal imaging is a new application domain for wireless video in which the video source is a wearable camera system also comprising a computational element and visual display means typically concealed inside ordinary-looking eyewear [1]. Personal imaging systems are based on a new computational framework called *WearComp* [2], and a new visual user interface paradigm called *mediated reality* (MR) [3].

WearComp, which originated in the 1970s as a tool for the visual arts [2], is a new form of human–computer interaction comprising a computer that is subsumed into the personal space of the user (wearer), controlled by the wearer, and has both operational and interactional constancy (i.e., is always on and always ready and accessible) [4]. Typical embodiments of WearComp comprise a body-worn computer system, a visual display over one or both eyes with text and graphics display capability, and an input device typically consisting of five or more pushbutton switches that may be operated by one hand.

Other input devices typically include a microphone and video camera positioned such that it provides a view of the same subject matter the wearer sees.

DEFINITION OF WEARCOMP

The new computational paradigm, WearComp, on which personal imaging is based is defined, with reference to Fig. 1 in terms of its eight attributes, as follows.

Constant — Always ready. It may have "sleep modes" but is never "dead." Unlike a laptop computer which must be opened up, switched on, and booted up before use, it is always on and always running.

Unrestrictive to the User — Ambulatory, mobile, roving, "you can do other things while using it" (type while jogging, etc.).

Unmonopolizing of User Attention — It does not cut one off from the outside world like a virtual reality game or the like. You can attend to other matters while using the apparatus. It is built with the assumption that computing will be a secondary activity, rather than a primary focus of attention. In fact, ideally it will provide enhanced sensory capabilities. It may, however, mediate (augment, alter, or deliberately diminish) sensory capabilities.

Observable by the User — It can get your attention continuously if you want it to. Almost always observable: within reasonable limitations (e.g., you might not see the screen while you blink or look away momentarily) the output medium is constantly perceptible by the wearer.

Controllable by the User — Responsive. You can grab control of it any time you wish. Even in automated processes you can manually override to break open the control loop and become part of the loop at any time you want (e.g., a big Halt button you want as an application mindlessly opens all 50 documents that were highlighted when you accidently pressed Enter), making a computer more controllable. Infinitely often controllable: the constancy of user interface results from almost always observability and infinitely often controllability in the sense that there is always a potential for manual override which need not always be exercised.

Attentive to the Environment — Environmentally aware, multimodal, multisensory; as a result, this ultimately gives the user increased situational awareness.

Communicative to Others — Can be used as a communications medium when you want it to. Expressive: allows the wearer to be expressive through the medium, whether as a direct communications medium to others, or as a means of assisting the production of expressive media (artistic or otherwise).

Personal — Human and computer are inextricably intertwined.
- *Prosthetic* — You can adapt it so that it acts as a true extension of mind and body; after some time you forget you are wearing it.
- *Assertive* — It can be a barrier to prohibition or requests by others for removal during times when you wish such a barrier. This is in contrast to a laptop computer in a briefcase or bag that could be separated from you by a "please leave all bags and briefcases at the counter" policy at a department store or library.

Private — Others cannot observe or control it unless you let them. Others cannot determine system status unless you want them to; for example, a clerk at a refund counter in a department store where photography is prohibited cannot tell whether or not you are transmitting wireless video to a spouse for a remote advice, in contrast to camcorder technology, where it is obvious you are taking a picture when you hold it up to your eye.

Other Properties that Result from the Eight Basic Properties

Connected — Insofar as technically feasible, wireless communication for a network connection is desirable. That it is *attentive* (can receive inputs from other systems and networks) and *communicative* (can send messages to other systems and networks) implies connectivity.

Anticipatory (Proactive) — That it is *attentive* and *constant* in its processing of incoming information implies that it can not only respond to events as they happen, but also anticipate events so that the apparatus can "learn" from the environment in a manner influenced by prior conditions. Moreover, that it can be constantly *observable* means it can take initiative and "make things happen" (e.g., it can get your attention). That it is *communicative* means it can also take initiative with respect to interacting with others. By others, we mean other people as well as other entities. For example it may take the initiative to send a message to the heater in the building to switch off when it senses you are on the verge of a "sweatdown."[1]

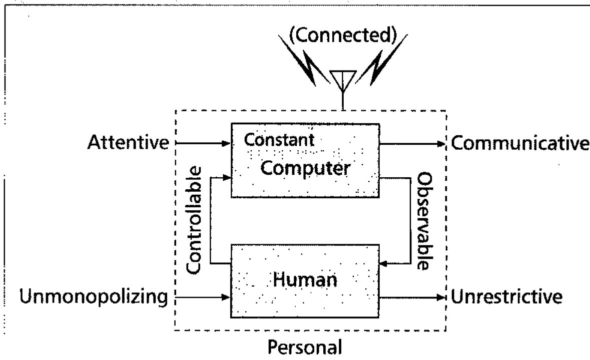
Retroactive — That it is *attentive* and *controllable* through processes which are *constant* means it may continually collect data that might be of use later. Thus, it may "learn" on subsequent recall.

Humanistic — This embodies human intellect as a part of the feedback loop of its processing. This means also that automatic (or autonomous) aspects of the system may be overridden, and therefore it is possible for the human operator to "step inside" the feedback loop of a process when and if desired. This arises from the fact that it is *controllable* and *observable*. In this way WearComp distinguishes itself from entities such as the obedience cuffs [5] worn by prisoners around their ankles, even obedience cuffs which have some form of *observable* property (e.g., in the form of pain-giving electric shocks produced by some obedience cuffs). What the obedience cuff lacks is the *controllable* attribute (in the sense that the prisoner does not have control of the apparatus).

Note that these are options rather than requirements. For example, *private* and *communicative* are in some ways opposite, yet together they facilitate an expanded range of options to move further in one direction or the other on a continuum between these two opposites. The important aspect here is choice for the user to move in an increased affordance space rather than a specific definition of how the user should act while using the apparatus.

Implicit in unrestrictive and unmonopolizing is that use of the apparatus will be a secondary rather than primary task.

[1] *A "sweatdown" is a sweat-induced computer shutdown resulting from excessive sweat adversely affecting system components which have not been sealed from the effects of moisture.*



■ **Figure 1.** *WearComp: a new form of human–computer interaction.*

Mediated Reality

The other important foundation technology for personal imaging is MR, which will now be described.

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 [6, 7], giving rise to *augmented reality* (AR). The general spirit of AR is to *add* computer graphics or the like to the real world.

A more general framework, MR, of which AR is a special case, is proposed. The intent of MR, like typical AR, includes *adding* virtual objects to visual reality, but also includes the ability to *remove*, *augment*, deliberately *diminish*, and significantly *alter* the perception of visual reality. MR attempts to visually "mediate" real objects, using a body-worn wireless video 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 (e.g., wireless video) medium.

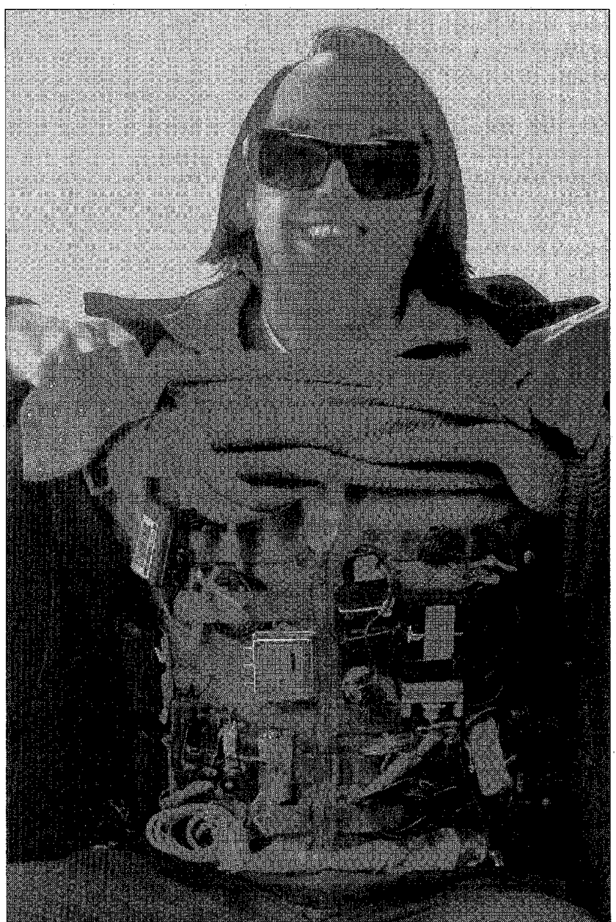
A means of *mediating* (augmenting, deliberately diminishing, or otherwise altering) reality in real time through a wireless video apparatus worn over the eyes has been proposed and reduced to practice [3]. An unobtrusive (covert) version of the apparatus has also been built into ordinary-looking sunglasses connected to an image processing computer concealed under ordinary clothing (Fig. 2). This entire means and apparatus will be referred to as a *reality mediator* (RM). Unlike early 1980s embodiments, characterized by hat-mounted antennas, the RF radiating surfaces are also concealed inside ordinary clothing. The RM allows one to augment, as well as deliberately diminish and otherwise alter, one's perception of reality, as well as to allow others to do so remotely through a wireless video communications link. The system depicted in Fig. 2 therefore leads to a new form of communication where, for example, 3D collaborative space may be shared with a remote viewer.

The ability to deliberately diminish reality is an important capability of the RM. For example, when looking through the special sunglasses at a brightly colored scene, there may not exist a unique color to use for the overlays. However, the sunglasses, instead of creating a perfect illusion of transparency, can be programmed to create an illusion of being *achromat transparent*. Being *achromat transparent* means that each incoming ray of light is absorbed and quantified, and its wavelength ignored. A ray from the same location is sent out the other side of the special sunglasses in the same direction, at the same time, but with a flat (grey) spectrum. This makes the wearer colorblind to real objects, making the real world appear less "busy" when combined with some colorful computer-generated overlays where color was used more effectively to accentuate the virtual objects. This was an effective way to prevent computer-generated objects from being "lost" in the clutter of the real world.

It has been found that color-reduced reality mediation was quite useful, for example, when comfortably seated on a commercial airline or commuter train; reading text on the virtual screen created by the sunglasses (e.g., e-mail), one could "tone down" the surroundings so that they take on a lesser role. It was not desired to be blind to the surroundings, as is someone who is reading a real paper newspaper (newspapers can easily end up covering most of a person's visual field). Other forms of chromatic mediation were also explored.

Wearable Tetherless Computer-Mediated Reality

The compute power required to perform general-purpose manipulation of color video streams was too unwieldy to be worn in a backpack (although body-worn video processing systems and other hardware to facilitate very limited forms of reality mediation have been constructed by the author). Accordingly, a large processing system with good video-processing capability may be accessed remotely by establishing a full-duplex video communications channel between the RM and the host video processor.[2] In particular, the full-duplex communications link comprises a high-quality communications link called the *inbound channel*, which is used to send the stereo video signal from the wearable cameras to the remote computer(s), while a lower-quality communications link called the *outbound channel* is used to carry the processed signal from the computer back to the HMD. This apparatus is depicted in a simple diagram (Fig. 3), in which a stereo video signal is sent to one or more computer systems over a high-quality microwave communications link (the inbound channel). The computer system(s) send back the processed image over a UHF communications link (the outbound channel). Note that in this figure i denotes inbound (e.g., iTx stands for inbound transmitter) and o outbound (e.g., oRx stands for outbound receiver). The term *visual filter* refers to the process(es) that mediate(s) the visual reality and possibly insert(s) "virtual" objects into the reality stream.



■ **Figure 2.** *The author's unobtrusive (covert) RM concealed in ordinary sunglasses with a complete multimedia video production studio and television station which can be concealed underneath ordinary clothing.*

Ideally both channels would be of high quality, but the machine-vision algorithms were found to be much more susceptible to noise than the author's own vision (ability to see through an apparatus that provided only a noisy illusion of transparency); and due to a need to separate the frequency bands, with an inferior frequency band being necessary for one of the channels, it was decided that the higher-quality channel should be the inbound channel. Since the illusion of transparency is degraded by the total noise level of inbound plus outbound artifacts, it was found to be equally poor regardless of the order of these two degradation artifacts. Therefore, it was decided to at least allow the remote video processing hardware to receive the best-quality picture.

Human-Centered Wireless Video

The main reason the MR paradigm was formulated was to make the camera take on a primary and central role in the context of a personal wireless video system. With MR, a camera *must* be present and the communications channels operative in order for the wearer to see anything. Since the camera is the only means by which the wearer of the RM can see (since the apparatus covers the eyes completely), the camera actually becomes the wearer's eye, for all practical purposes. Thus, if, for example, the camera is not adjusted properly, or the video transmission is not of sufficient quality, the wearer will be unable to see properly. In this way, the wearer is part of the feedback loop that keeps the camera, and the rest of the system, adjusted properly. Therefore, whether through conscious effort (manual adjustment of the camera) or subconscious effort (choice of gaze angle, positioning of the head, body, antennas, etc., and general conduct and behavior), the camera will always end up being situated for the best picture. In this way the synergy between human and machine operates as a system which always seeks the best picture. The wearer's entire body posture, mannerisms, and general conduct evolve to become part of a natural control system that constantly adjusts itself to obtain the clearest image on the wearer's own screen, with a side effect that the transmitted video signal as observed by any other entity (be it other humans or other machine vision algorithms) was optimal. After an extended period of time, this form of interaction no longer required conscious thought or effort. This characteristic gives rise to the MR genre of documentary cinematography, as well as more effective computer-supported collaboration.

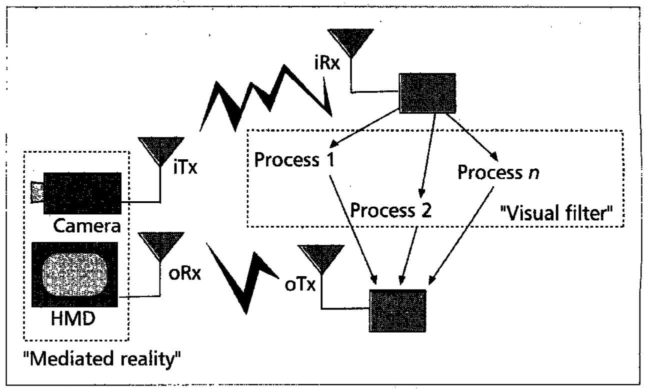
[2] *At times, as many as 20 to 30 state-of-the art workstations working simultaneously have been used to process the video, linked together with large file servers over an FDDI network, so by no means does "processor" imply a need to be limited to a single computer. However, it is hoped that a scaled-down version of a special-purpose computer [8] might soon enable higher-speed photo rendering.*

Infinite Screen Resolution with Camera-Based Head-Tracking

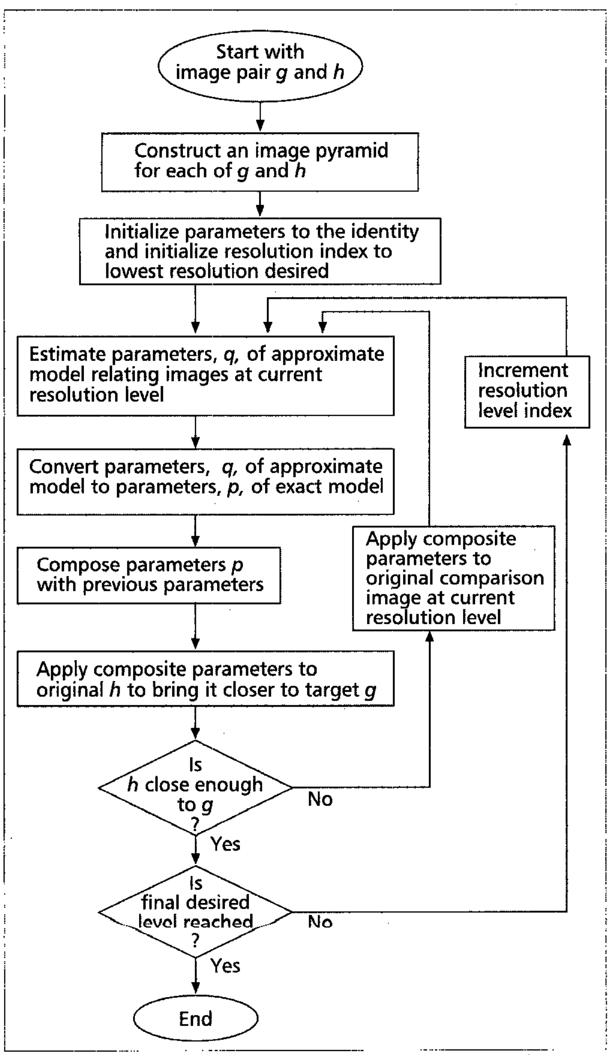
As described in the previous section, the RM must have a camera in order for the wearer to see. Accordingly, other uses of this camera are proposed in this section. These other uses are as a head-tracker, and to make the wearer into a producer as well as a consumer of virtual information.

The proposed methodology, MR, is quite different from other related work where the assumption is often that there is a controlled environment such as the assembly line of a factory, where VR headsets might be used to make employees more productive, and where head-tracking and the like may therefore be done quite easily with a special device fixed in the environment. By tethered cables, workers are, in effect, imprisoned in their work cells, unable to roam freely without taking off their VR headsets.

Quite the opposite is true with personal imaging, where there is no assumption regarding a fixed location. Indeed, the goal of personal imaging is that the apparatus function in nearly any environment, with no special preparation of the environment required.



■ **Figure 3.** *A simple implementation of an RM.*



■ **Figure 4.** *The VideoOrbits head-tracking algorithm.*

The VideoOrbits algorithm, based on the environment [9] and depicted in Fig. 4, requires no special devices installed in the environment. The camera in the personal imaging system simply tracks itself based on its view of objects in the environment. The algorithm is based on algebraic projective geometry, and provides an estimate of the true projective coordinate transformation, which for successive image pairs is composed using the projective group [9]. Successive pairs of images may be estimated using the Lie algebra of the group, while absolute head tracking is done using the exact group by relating the approximate parameters q to the exact parameters p in the innermost loop of the process. The algorithm typically runs at 5–10 frames/s on a general-purpose computer, but its simple structure makes it easy to implement in hardware for full-motion video.

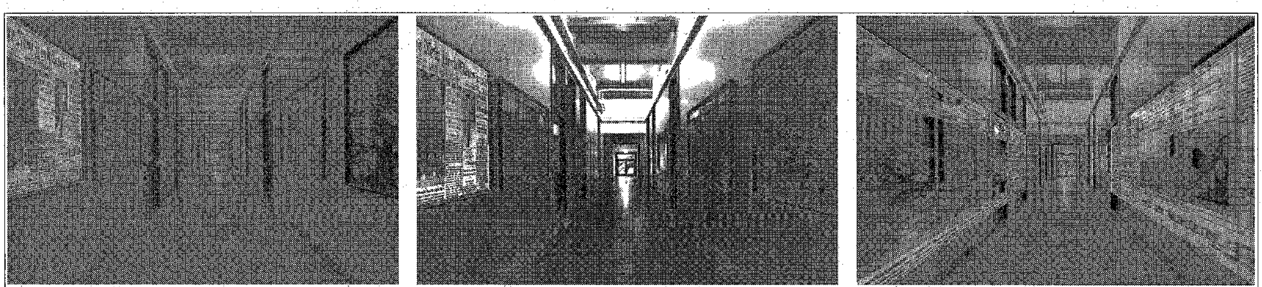
This new method of head-tracking based on the use of a video camera was proposed in [1]; it performs head-tracking visually and works without the need for object recognition. Instead, it builds on the image processing framework proposed by Venetsanopoulos [10] and Stockham [11], and combines these ideas with the Horn and Schunk equations [12] and some ideas in algebraic projective geometry and homometric imaging, using a spatiotonal model, $\tilde{p}$, which works in the neighborhood of the identity

$$\left(\sum_{x,y}(\phi(x,y)\phi^T(x,y))\right)\tilde{\mathbf{p}} = -\sum_{x,y} F_t\phi(x,y) \qquad (1)$$

where $\phi^T = [F_x(xy, x, y, 1), F_y(xy, x, y, 1), F, 1]$, $F(x, t) = f(q(x))$ at time t, $F_x(x, t) = (df/dq)(dq(x)/dx)$ at time t, and $F_t(x, t)$ is the frame difference of adjacent frames. This "approximate model" is used in the innermost loop of the diagram depicted in Fig. 4 and then related to the parameters of an exact projectivity and gain group of transformations so that the true group structure is preserved throughout, and the orbit of this group of transformations is followed by the video stream (hence the name "VideoOrbits"). The method is based on the fact that the unknown nonlinearity of the camera, f, can be obtained from differently exposed images $f(q)$ and $f(kq)$, and so on, and that these can be combined to estimate the actual quantity of light entering the imaging system:

$$\hat{q}(x) = \frac{\sum_i c_i\left(\frac{Ax+b}{cx+1}\right)\frac{1}{k_i}f^{-1}\left(F_i\frac{Ax+b}{cx+1}\right)}{\sum_i c_i\left(\frac{Ax+b}{cx+1}\right)} \qquad (2)$$

where c_i is the derivative of the recovered nonlinear response function of the camera, f, and A, b, and c are the parameters of the true projective coordinate transformation of the light falling on the image sensor. This method allows the actual quantity of light entering the eye due to the real and virtual



a b c

■ **Figure 5.** *The virtual wireless video screen has essentially infinite resolution, but is of course viewed through a finite-resolution eyeglass-based device.*

objects to be placed on an equal footing. This allows the wearer of the apparatus to explore a virtual world of infinite screen resolution, because any flat object in the environment becomes a virtual screen. Each virtual screen is based on projective geometry (Fig. 5).

Figure 5a shows the image of a newspaper appropriately transformed, using the appropriate operator of the projective group of coordinate transformations, and added to the wearer's most recently transmitted image to sustain the illusion of a rigid planar patch in 3D space; this process is called *homographic modeling*. By calculating the quantity of light using a recovered estimate of the nonlinear response function of the camera, a more realistic-looking tone scale results, as shown in Fig. 5b. Fig. 5c shows a true mathematical projection, which provides a more intuitive mode of user interface to the more experienced user. The metaphor used is one of transparency of the head, as if rays of light could pass through the back of the head and enter into the back of the eye. This mode of operation requires long-term adaptation to an often upside-down and reversed image, but after adaptation the sens of the image, reversed or unreversed, encodes a distinction between rays of light coming through the front of the eyes and those going through the back of the head.

It is well known that humans can remember much more information by pretending that each piece of information is in a large building, and imagining ourselves walking around in the rooms of the building. With the proposed framework, we can now make that mental trick a reality, and leave messages for ourselves on various flat surfaces in the building. In this way, an airport lounge or train station may be turned into a virtual office, where we can place our calendar on one wall, our e-mail on another, and just walk around and look at these walls one at a time to manage several windows. VideoOrbits thus provides us with a new window manager.

Moreover, images encountered by another wearer of a similar apparatus may be recognized as belonging to the same orbit of the projective group of coordinate transformations, so the apparatus can be used as a communications device.

Seeing Eye to Eye: Video Consumers as Video Producers

As the wearer looks around at a virtual object which has been inserted into the visual reality stream by way of the RM, a side-effect is that information from the camera is collected, and, in addition to its use as a head-tracker, the camera transmits this video wirelessly to remote sites. Thus, while consuming information (virtual objects) the wearer is also producing information. The video the wearer takes in is assembled into a so-called lookpainting image composite [2] that someone else at another location can navigate in a manner similar to a QuickTime VR environment map. However, the important difference is that no special photographic tripod and apparatus were needed — the environment map was generated just by looking around. Moreover, the exact same mathematical calculations that were used to do the head-tracking are used to assemble the successive frames of video into an environment map.

A remote viewer may navigate the environment map using a desktop computer, or a VR headset with head-tracker. Alternatively, the remote viewer may wear another RM and navigate the environment map using the proposed camera-based head-tracker. In this case, two people at different locations each generate an environment map as a side-effect of navigating the other person's environment map. The same mathematical framework used for generating the environment map (Fig. 6a) is used for attaining subpixel accuracy in tracking (Fig. 6b–d), even in the crowded department store shown in Fig. 6, with other people moving behind the cashier. The proposed MR framework differs from virtual reality or augmented reality applications in the sense that tracking (registration of the real and virtual worlds) remains accurate to this level. Note that only every 10th frame of the image sequence is shown in Fig. 6.

Personal Imaging for Symbiotic Wireless Video

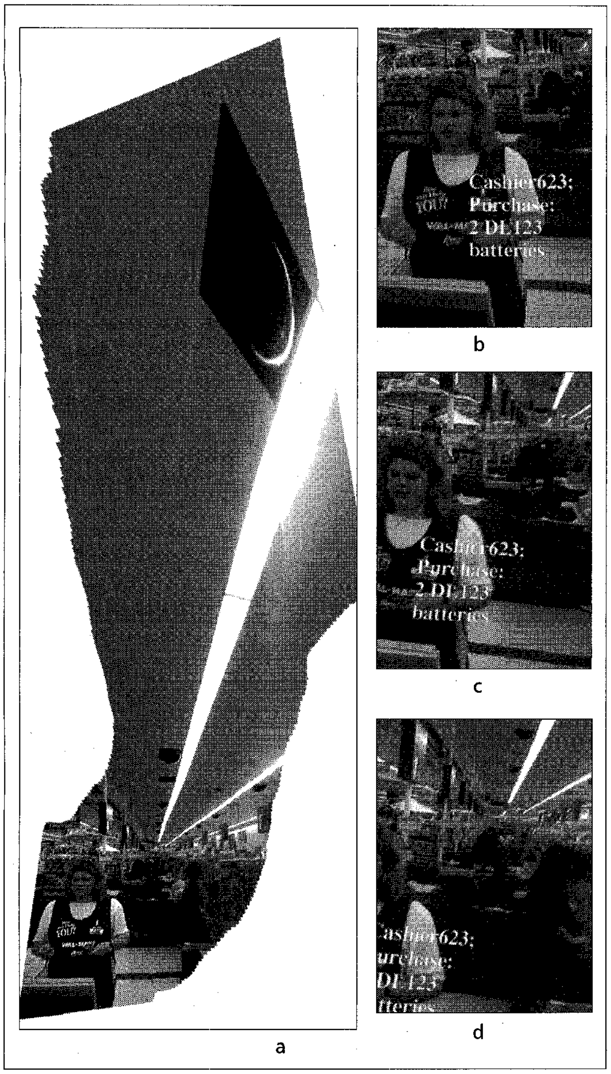
Blind Vision

If both the machine vision and artificial intelligence problems were solved, a blind person with a personal imaging system (body-worn image processing system and camera) and headset could be guided by high-level scene understanding algorithms.

However, rather than attempting to solve the machine vision and artificial intelligence problems, imagine, instead, a blind person wearing a version of a wearable wireless webcam (WearCam) [13] with headset and boom microphone (WearTel) [14], and being guided by a viewer at a remote location who is looking at the received signal from the wearer's camera and transmitter. The remote viewer could send verbal suggestions or other audio cues back into the wearer's headset. In this way, the wearer is guided by human intelligence at a remote location. Now suppose that there was a room full of volunteers, each providing this human intelligence. With such a human intelligence network, it might be possible for a number of blind people to have the opportunity to experience a new form of increased mobility that goes far beyond the self-contained devices built in the 1980s to assist the visually challenged [15].

Virtual Body

Consider the convalescent, infirm, or wheelchair-bound who might wish to live vicariously through the eyes of an able-bodied person. Such is possible with an apparatus such as a wearable wireless webcam [13], in which an able-bodied person wears a camera and transmits navigable environments directly



■ **Figure 6.** *a) Environment maps are generated by simply looking around a room. b, c) Notice how the virtual "name tag" (and grocery list) appears to stay attached to the cashier, even d) when the cashier is no longer within the field of view of the video camera and transmitter.*

to the World Wide Web. Such a system was tested on a very large scale for New Year's Eve 1995, in which a large audience of people unable to get out to celebrate the festivities were able to do so vicariously through the author's eyes. Moreover, any one of these people was able to correspond with the author by e-mail (sometimes suggesting to the author a possible site or event of interest, or commenting on the very image within the author's field of view). Such an apparatus as WearTel, which provides a first-person videoconferencing capability, may give rise to a host of such applications in which an able-bodied volunteer provides "telexperience" — a telematic vicarious experience.

Connected Collective Humanistic Intelligence

Now suppose that the convalescent, infirm, or wheelchair-bound live vicariously through the able bodies of the blind, while at the same time providing the human intelligence needed to interpret the video being sent from these blind people with WearCams. Such is an example of a new kind of symbiotic human intelligence.

Wireless video, together with an audio back-channel, becomes the enabling technology. This use of WearTel differs from traditional video conferencing in the sense that it is wireless, and provides a first-person perspective; that is, it provides a view not of the wearer of the apparatus, but rather of what the wearer is looking at.

Computer-Supported Cooperative Living

A young child with perfect vision does not need remote visual intelligence, but might nevertheless benefit from the wisdom and experience of the elderly. Grandparents, for example, might offer their advice in various social matters, or might be able to tell, over a wireless voice link, an interesting story about an old building they remember from their own childhood, as they see pictures of the building on the child's Web page. They may welcome the opportunity to vicariously spend time with their grandchildren and, in some sense, relive a portion of their childhood. Friends and relatives may be interconnected in a new web of connected collective humanistic intelligence.

It has been the author's experience that such symbiotic relationships must be built in such a way that there is a manner for equilibrium to be reached. This is done by giving those at each side of the communications link the ability to have three settings:

- A privacy setting that prevents the establishment of a link (this is like taking a telephone off the hook to prevent it from ringing)
- A normal setting that allows a link to be established after notification (analogous to a normal telephone call where the telephone rings, and the recipient decides whether or not to lift the handset off its hook)
- An "open" setting that allows a connection to be established at any time (analogous to an open or always-on intercom or a "video wall" that runs continuously)

In this way, if too much advice is given, to the point where the wearer of the apparatus feels overwhelmed or feels intruded upon, the wearer will switch to "privacy," or at least to "normal." In time, both sides learn how much interaction is wanted, and the interaction tends toward a level that is symbiotic. Such an apparatus may also be used for crime reduction.

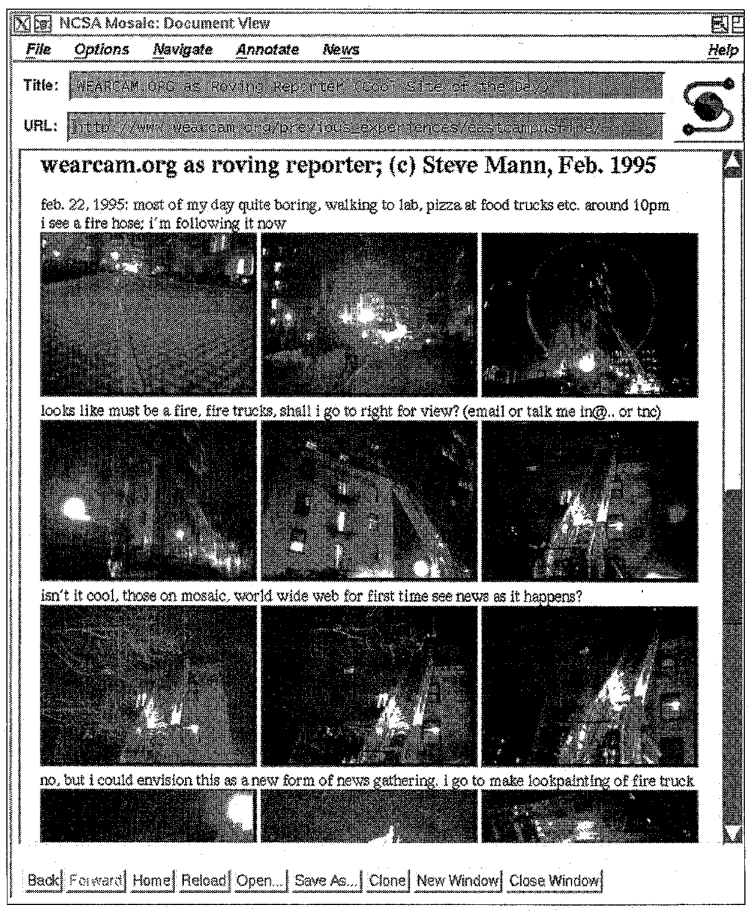
Personal Imaging as a Tool for Photojournalists and Reporters

Throughout the 1990s the author has experimented with personal imaging as a means of creating personal documentary, and sharing this personal documentary video on the Web, in the form of a *wearable wireless webcam* [16].

On several occasions, new and interesting forms of collaboration emerged. On various occasions the author serendipitously encountered newsworthy events while wearing the apparatus. On some such occasions where the events were natural disasters or the like, there was no other coverage of these events (e.g., traditional journalists were unavailable to cover these events on the short notice involved, despite great desire that there be coverage of these events).

An example of how the author functioned as a "roving reporter" is illustrated in Fig. 7. This shows how computer-supported collaborative photojournalism (CSCP) emerged from a wearable wireless webcam. The author encountered an event through ordinary everyday activity. As it turned out later, a newspaper had very desperately wanted to have this event covered, but could not reach any of its photojournalists in time to cover the event. The author, however, was able to offer hundreds of pictures of the event, wirelessly transmitted while the event was still occurring. Furthermore, interaction with a large number of remote viewers enabled a new form of real-time collaboration. In actual practice, multiple images are "stitched together" to make a picture good enough for a full-page newspaper-sized photograph despite the fact that each image has relatively low resolution. The manner in which pictures are combined is with the so-called lookpainting (VideoOrbits) algorithm previously described.

In another example of CSCP the author encountered a flood in the basement of a building and notified the police of the flood. However, without really any conscious thought or effort, the fact that the author walked past the event and looked at it also resulted in its having been recorded and transmitted wirelessly to remote sites. Thus, the author was subsequently able to notify a newspaper of this transmission (wirelessly notifying the newspaper's editorial offices through e-mail), and the very high resolution images of tremendoulsy high dynamic range and tonal fidelity were retrieved by the newspaper's editorial office and published in the newspaper. The quality of the images was higher than typical for the newspaper, suggesting that the wearable wireless webcam can rival the photographic technical quality and resolution of professional photographers armed with the best cameras available on the market.



■ **Figure 7.** *Serendipitously arising computer-supported collaborative photojournalism.*

Conclusions on Personal Imaging and Wireless Video

Wearable tetherless computer-mediated reality, with its origins in a seemingly obscure visual aesthetic and photographic technique of the 1970s and 1980s, has evolved from computer-supported collaborative photography into computer-supported collaborative videography, toward defining a new genre of personal documentary.

In summary, this form of videography, called *lookpainting*, arises in a symbiotic way through the use of the camera as a head-tracking device. Not only does this liberate the wearer from the traditional work cell so that the wearer can use the head-tracker anywhere;, it also means that a side-effect of the very head-tracking operation is to produce a so-called lookpainting. In this way, the act of navigating in the virtual space created by another person at a remote location creates a virtual space reciprocally navigable by the person at the remote location. Thus, two or more people may exchange viewpoints in a new form of interaction called "seeing eye to eye."

Moreover, extremely high-resolution/high-definition pictures can be rendered. Lookpainting also provides a look-around user interface which is even more natural than the point-and-click user interface of modern cameras. Furthermore, lookpainting affords the user total control of the process, makes the process of capturing an image more engaging and fulfilling, and results in environment maps that can be shared remotely with others who have access to the Web or a similar visual communications media. This application has also been successfully demonstrated for photojournalism, resulting in a useful tool for the reporter of the future. Moreover, lookpainting provides a new metaphor for *computer-supported cooperative living*, as well as a new era in which ordinary everyday experience can become newsworthy material — an era in which the job of the reporter may be spread out among ordinary people who will eventually, given enough people wearing enough wireless video units, encounter events far more newsworthy than those captured by a limited number of reporters covering only as much as they can. Lastly, as mediated presence gives rise to mediated telepresence, a new communications medium arises with applications in personal safety and crime reduction as well as interactional capabilities that go far beyond portable IP video telephony.

Applications for the disabled have also been proposed, for example, in which a blind person may form a symbiotic relationship with another person who is infirm or otherwise confined to a wheelchair or hospital bed.

Acknowledgments

The author wishes to thank Woodrow Barfield, Don Norman, Randy Pausch, Simon Haykin, Rosalind Picard, Steve Feiner, Charles Wyckoff, Hiroshi Ishii, Chuck Carter, Kent Nickerson, Kris Popat, Jonathan Rose, and Chris Barnhart for much in the way of useful feedback, constructive criticism, and so on as this work has evolved. Gerald Q. "Chip" Maguire, Jr., Vaughan Pratt, and Bruce Macdonald made valuable

suggestions toward arriving at a clear definition of wearable computing.

Additional thanks to Kopin, Kodak, ViA, Xybernaut, Digital Equipment Corporation, VirtualVision, Hewlett Packard Laboratories, Chris Barnhart, Compaq, Miyota, Chuck Carter, Bob Kinney, Thought Technologies Limited, and the Institute for Interventional Informatics for lending or donating additional equipment that made these experiments possible.

REFERENCES

[1] S. Mann, "Wearable Computing: A First Step toward Personal Imaging," *IEEE Comp.*, vol. 30, no. 2, Feb. 1997; http://n1nlf-1.eecg.toronto.edu/ieeecomputer/index.html.

[2] S. Mann, "An Historical Account of the 'WearComp' and 'WearCam' Projects Developed for "Personal Imaging," *Int'l. Symp. Wearable Comp.*, Cambridge, MA, Oct. 13–14, 1997.

[3] S. Mann, "Mediated Reality," TR 260, M.I.T. Media Lab Perceptual Computing Section, Cambridge, MA, http://wearcam.org/mediated-reality/index.html, 1994.

[4] S. Mann, "Smart Clothing: The Wearable Computer and Wearcam," *Pers. Tech.*, Mar. 1997, vol. 1, issue 1.

[5] J. Hoshen, J. Sennott, and M. Winkler, "Keeping Tabs on Criminals," *IEEE Spectrum*, Feb. 1995, pp. 26–32.

[6] R. A. Earnshaw, M. A. Gigante, and H Jones, *Virtual Reality Systems*, Academic Press, 1993.

[7] I. Sutherland, "A Head-Mounted Three Dimensional Display," *Proc. Fall Joint Comp. Conf.*, 1968, pp. 757–64.

[8] Rose *et al.*, "The Transmogrifier-2 Field Programmable System," http://www.eecg.toronto.edu/\~jayar/research/tm2.html.

[9] S. Mann and R. W. Picard, "Video Orbits of the Projective Group; A Simple Approach to Featureless Estimation of Parameters," TR 338, MIT, Cambridge, MA; http://n1nlf-1.eecg.toronto.edu/tip.ps.gz 1995; also appears in *IEEE Trans. Image Proc.*, vol. 6, no. 9, Sept. 1997.

[10] A. N. Venetsanopoulos, "Nonlinear Mean Filters in Image Processing," *IEEE Trans. ASSP*, vol. 34, no. 3,pp. 573–84, June 1986.

[11] T. G. Stockham, Jr. "Image Processing in the Context of a Visual Model," *Proc. IEEE*, vol. 60, no. 7, July 1972, pp. 828–42.

[12] B. Horn and B. Schunk, "Determining Optical Flow," *Artificial Intelligence*, 1981.

[13] S. Mann, "Wearable, Tetherless Computer-Mediated Reality: WearCam as a Wearable Face-Recognizer, and Other Applications for the Disabled," TR 361, MIT Media Lab Perceptual Computing Section, Cambridge, MA, Feb. 2, 1996; also appears in *AAAI Fall Symp. Developing Assistive Tech. for People with Disabilities*, Nov. 9–11, 1996.

[14] S. Mann, "The Telephone of the Future as Embodiment of Humanistic Intelligence: A New Perspective on Videoconferencing, 1997; http://weartel.com/weartel.html.

[15] S. Mann, "VibraVest/ThinkTank: Existential Technology of Synthetic Synesthesia for the Visually Challenged," *Proc. 8th Int'l. Symp. Elect. Art*, Chicago, IL, Sept. 22–27, 1997.

BIOGRAPHY

STEVE MANN (mann@eecg.toronto.edu), inventor of WearCam (reality mediator) and WearComp, is currently a faculty member at the University of Toronto, Department of Electrical and Computer Engineering. He has been inventing, designing, and building personal imaging systems as a hobby since his high school days in the 1970s and early 1980s. More recently, in 1991 he brought his invention to the Massachusetts Institute of Technology and continued this work there, defining "personal imaging" as a new field of research, in which he received his Ph.D. degree from MIT in 1997. His previous degrees are an M.Eng. in electrical engineering, and undergraduate degrees in physics (B.Sc.) and electrical engineering (B.Eng.) from McMaster University. He was guest editor of a special issue on wearable computing and personal imaging in *Personal Technologies Journal*, one of four organizers of the ACM's First International Workshop on Wearable Computing, and publications chair for the IEEE International Symposium on Wearable Computing (ISWC '97). His present research interests include quantagraphic imaging, lightspace rendering, and wearable tetherless computer-mediated photography. He has currently just begun setting up a new Humanistic Intelligence laboratory to "invent the camera of the future," and is now looking for graduate students and staff for this project. He is also interested in the visual arts and has exhibited his quantagraphic image composites and lightspace renderings (the means and apparatus for which he was recently awarded U.S. Patent 5,706,416) in numerous art galleries from 1985 to the present. See http://www.wearcomp.org/research.html or http://genesis.eecg.toronto.edu/research.html