EyeTap Technology for Wireless Electronic News Gathering*

Steve Mann
mann@eecg.toronto.edu

James Fung
fungja@eecg.toronto.edu

Eric Moncrieff
ericm@eecg.toronto.edu

ECE Department, University of Toronto,
Toronto, Ontario, Canada

The results of actual field tests of a new class of Personal Imaging systems, the Electronic News Gathering WearComp (ENGwear), are presented. These incorporate EyeTap Technology to cause the eye itself to function as if it were a camera, and to allow the eye to be wirelessly tapped out of and tapped into, capturing exactly what the user is viewing. This brings the point of view of the camera to the same point as the user's eye, and also allows the user to allow others to wirelessly modify his/her perception of visual reality. Images thus captured are simultaneously sent and received live to and from the internet using IP-over-wireless technology. This novel application of wireless technology empowers reporters: ENGwear systems allow for a great degree of mobility, and are constantly operational and recording to a circular buffer allowing the reporter to never 'miss a shot'.

I. Introduction

The "Wearable Wireless Webcam" [1] was used for electronic news gathering in 1994, and in the production of the first online video documentary, running almost continuously from 1994 to 1996 [2]. During that time, specific newsworthy events, such as the East Campus Fire (http://wearcam.org/eastcampusfire.htm) and the Edgerton House Flood were reported in real time. All of these events, (including material covered in other media such as newspaper and television) were gathered serendipitously, having simply been in the right place at the right time, by coincidence. Making use of amateur radio with wireless communications, images were transmitted from day–to–day travels in the context of ordinary living.

The transmitted video captured a first–person perspective, in which the apparatus of the invention effectively caused the eye itself to function as if it were a camera. This unique perspective allowed the wearer to capture images of a more natural sort — every person in his or her own element, without the disturbing influence a camera might have on the subject matter.

Several new applications of the wireless EyeTap technology have recently been explored in further detail, with positive results. One of these new developments is the ENGwear project, which explored Wireless EyeTap technology as it relates to Electronic News Gathering. Carried out by Mann and his students during orientation week at the University of Toronto, the project allowed the EyeTap cyborgs to transmit images from the events and celebrations live to the group's website. This provides a hint of the potential of this technology to empower the common person to report on his own life, and to show the world the daily events in each of our lives which are newsworthy.

The ENGwear project relied on several well-established pieces of technology, and introduced several new ones to the field of ENG. This paper reviews some of the basic technologies, EyeTap and WearComp, and introduces the new application of Electronic News Gathering to the field of Personal Imaging, showing how it emerges directly from fundamental characterisitcs of EyeTap technology. The ability to document everyday life is available to the wearer with almost no further mental adjustment, and has the potential to change the very way we interact as individuals in society, as well as within emerging media.

II. Background

ENGwear systems are comprised of three major components: 1) EyeTap Technology, 2) WearComp, and 3) Wireless Communications.

II.A. EyeTap Technology

EyeTap technology allows the user's eye to function as a camera. Rays of light entering the eye are absorbed and quantified. This quantified form of visual

*Thanks to Kodak, DEC, ClearNET, and NSERC for assisting with this project.

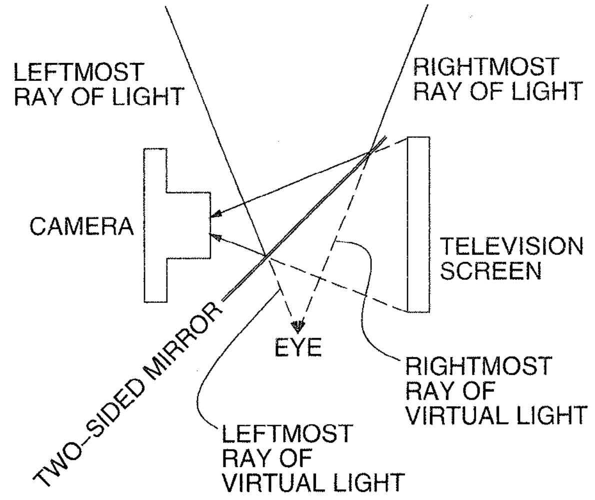

Figure 1: **A diverter embodiment of the reality–mediator:** A diverter (in this case, a double–sided mirror) diverts incoming rays of light to a camera, while at the same time providing the eye with a view of a television screen, connected to the wearable computer system. The television screen appears backwards to the eye, but since the computer captures a backwards stream of images (the camera's view of the world is also through a mirror), display of that video stream will create an illusion of transparency that is not backwards. Typically, such an apparatus may operate with photoquantigraphic image processing capability [3].

information can then be processed by a WearComp, with information added (augmented reality), removed (diminished reality) or replaced (mediated reality, of which augmented and diminished realities are special cases). The filtered version of the light (mediated light) is then redisplayed to the wearer. If the information is allowed to pass unaltered, EyeTap essentially functions to replace a portion of the wearer's visual field of view with exactly the same subject matter. In this case, the WearComp captures exactly what the user sees, and the apparatus merely functions as a viewfinder. Of course, various "visual filters", which are simply computer programs, can be inserted into the visual reality stream to mediate the wearer's visual perception.

The ENGwear systems used in this project were based on the 45 degree diverter implementation of the EyeTap principle, as illustrated in Fig. 1. This embodiment was chosen for its ease of implementation, and is the form used in the teaching of the Personal Cybernetics course (ECE1766, http://wearcam.org/ece1766.html) at the University of Toronto.

While a photographer or cameraman must look into the viewfinder of his or her camera, ENGwear reporters need only look at the subject matter in question, assured that the image is captured just as they are viewing it, since one of their eyes functions as both the viewfinder and camera.

When the mediated area is smaller than the total field of view of the reporter's eye, some sort of visual cue is used to indicate the area captured by the EyeTap. In these systems, though images could be captured in full colour, the virtual light generated was monochrome green. Thus, the mediated field of the wearer's vision appeared as a combination of what was normally seen from their untapped eye, and a monochrome green version from their tapped eye. In these cases, the EyeTaps functioned as viewfinders, with the reporter setting her gaze such that she overlayed the mediated version of reality atop the subject matter she wished to capture. It was found that reporters naturally tended to place the mediation zone over the subject matter of interest, so that the camera functioned as a true extension of the mind and body, often being operable without conscious thought or effort.

Since rays of light in the mediated field of view are collinear with rays of light entering the EyeTap device, ENGwear systems are suitable for long term use. Conventional viewfinders on television cameras often have fields of view which do not match the user's eye's field of view. Thus, while a cameraman may experience disorientation when walking and filming for long periods of time looking through a viewfinder, an ENGwear reporter can use their system constantly without experiencing the disorienting effects associated with non-collinear viewfinders.

II.B. WearComp

As mentioned above, incoming light is passed in a quantified form through a WearComp. A WearComp has three defining characteristics:

1. *Within the personal space of the wearer*
   A WearComp is recognized by others as being within the personal space of the wearer, just as clothes, shoes, bellybags, or a purse would be regarded. This allows for ENGwear to be a more covert, ubiquitous form of news gathering equipment since, while one may be asked to put down a bag or a camera, it is less often that one is requested to remove one's shirt or WearComp.

2. *Interactional Constancy*
   WearComps are constantly inside the feedback

loop of the wearer's consciousness. The computer is inextricably set between the user and the world; it is a filter which is always operational. There is no need to look at it; information vital to everyday life is always being presented whether or not traditional computing tasks are in progress. This leads to a literally transparent user interface.

3. *Operational Constancy*
WearComps are constantly in operation, and constantly presenting information to the user. All information within and entering the WearComp/human feedback loop is captured, processed, preserved and presented. This mode of constant operation contrasts with traditional media such as a broadcast cameras, which must be held up to the eye when the cameraman wishes to film.

   A benefit of operational constancy is the decreased likelihood of ever "missing the shot." Images captured by the WearComp are placed in a circular buffer. Such a buffer is constantly updated, but continues to hold all the images up to a few minutes or hours into the past. When a significant event occurs, one can be assured it is captured, so long as the reporter sees it himself. Thus, even though the wearer may not be transmitting at the time a critical event occured, the images can still be retrieved from the buffer and transmitted. In this sense, no time or effort needs to be expended to ready a camera and place subject matter in the viewfinder. Whereas the conventional cameraman must make a conscious effort to record images, ENGwear reporters, by witnessing events, record them.

II.C. Wireless Communications Link

When a WearComp and EyeTap are coupled with wireless communications, the ENGwear system becomes a highly mobile news gathering device. Most traditional news media relies on the reporter/cameraman team. Additionally, the team carries with it a large array of broadcasting equipment such as cameras, microphones, tape, cables, and, for live broadcasts, a satellite feed mounted on a news van. This equipment overhead affords them the capability to capture high quality images and sound, but in an obtrusive and cumbersome manner.

The ENGwear project, by comparison, required no trucks, no satellites, and very little broadcasting infrastructure. An ENGwear system empowers an individual to act as reporter, cameraman, and transmitter, while giving her more mobility than a traditional media team. Furthermore, ENGwear reporters are afforded the opportunity to observe the scene from the perspective of an eye witness. The interference caused by the presence of a news team is greatly reduced, if not removed, with the use of EyeTap technology.

In contrast with broadcast television equipment, which is designed to produce high quality full motion video, ENGwear systems capture events as a sequence of still images, or as video from which individual frames can be easily extracted. The rate of image transmission and the image resolution can be chosen based on the nature of the event being reported on. For instance, in some cases, it may be desirable to transmit several high resolution images over the span of a few minutes, while in others, it may be desirable to capture an event in a more live capacity, transmitting lower resolution pictures at 30 frames per second. The ENGwear system allows the user to make a tradeoff between frame rate and image resolution. This tradeoff can also be made after the data is transmitted (e.g. by the viewer of the event combining multiple images into a Photoquantigraphic Image Composite described below). The desired resolution and frame rate is also determined, to some extent, by the capacity of the wireless connection to transmit images.

When the ENGwear project was conducted, it was found that these factors resulted in a reporting style consisting of data that could be rendered as video, or as a sequence of extremely high resolution still images, potraying an event. Even when images captured encompassed a limited field of view, and were individually of low resolution, multiple images could later be combined into a single, large composite called a Photoquantigraphic Image Composite (PIC). The result was a single, high resolution image made up of a sequence of individual low resolution images seamlessly integrated with one another using the VideoOrbits algorithm developed by Mann [4]. This provided a means to capture larger, more expressive still images with video cameras.

Most of the live footage, however, consisted of sequences of still images which captured events as they occured. This sort of video journal lay somewhere between traditional print media, which gives few, but very descriptive images of an event, and live video.

III. The Edgertonian image sequence in ENGwear systems

Such image sequences are Edgertonian in nature, reflecting a fundamental principle of EyeTap operation. Traditional image sequence compression, such as MPEG, is based on processing frames of video as a continuum. The integrity of motion is often regarded as being more important than, or at least as important as the integrity of each individual frame of the image sequence. However, it can be argued that temporal integrity is not always of the utmost importance, and can, in fact, often be sacrificed, with good reason.

III.A. Edgertonian versus Nyquist thinking

Consider the very typical situation when the frame rate of a picture acquisition process vastly exceeds the frame rate at which it is possible to send pictures of satisfactory quality over a given bandwidth–limited communications channel. This situation arises, for example, with Web based cameras, including the "Wearable Wireless Webcam."[5]

Suppose, for example, that the camera provides 30 pictures per second, but the channel only allows us to send one picture per second (ignore for the moment the fact that we can trade spatial resolution, temporal resolution, and compression quality to adjust the frame rate). In order to downsample our 30 pictures per second down to one picture per second, the "Nyquist school of thought" would suggest that we should temporally low pass filter the image sequence, in order to remove any temporal frequencies that would exceed the Nyquist frequency.

To apply this standard "lowpass filter then downsample" approach, we might average together each 30 successive pictures, to obtain each output picture. Thus fast moving objects would be blurred to prevent temporal aliasing.

We might be tempted to think that this blurring is desirable, given temporal aliasing that would otherwise result. However, cinematographers and others who produce motion pictures often disregard concepts like temporal aliasing. Most notably, Harold E. Edgerton, the inventor of the electronic flash, who is known for his movies of high speed events in which objects are "frozen" in time, has produced movies and other artifacts that defy any avoidance of temporal aliasing. Edgerton's movies provide us with a temporal sampling that is more like a Dirac comb (downsampling of reality) than a lowpass filtered and then downsampled version of reality.

Returning to the example of downsamping from 30 frames per second to one frame per second, an Edgertonian thinker would likely advocate simply taking every 30th frame from the original sequence, and throwing all the others away.

The Edgertonian downsampling philosophy gives rise to image sequences in which propeller blades or wagon wheel spokes appear to spin backwards or stand still. The Nyquist philosophy gives rise to image sequences in which the propeller blades or wagon wheel spokes would visually disappear.

We believe that it is preferable that the propeller blades and wagon wheel spokes appear to spin backwards, or stand still, rather than visually disappear. In the context of ENGwear, it is preferable to have a series of crisp well–defined "snapshots" of reality, rather than the blur of images that one would get by following the antialiasing approach of traditional signal processing.

Personal experience with EyeTap video tends to lead to a more intimate understanding of how the world looks through Web–based video, rather than traditional photographic media. On this system, it was possible to choose from among various combinations of Edgertonian and Nyquist sampling strategies. It was found that experiencing the world through "Edgertonian eyes" was generally preferable to the Nyquist approach.

III.B. Frames versus rows, columns, and pixels

There is a trend now toward processing sequences of images as spatiotemporal volumes, e.g. as a function $f(x, y, t)$. Within this conceptual framework, motion pictures are treated as static $3d$ volumes of data. So–called *spatiotemporal filters* $h(x, y, t)$ are applied to these spatiotemporal volumes $f(x, y, t)$.

However, this unified treatment of the three dimensions (discretized to row, column, and frame number) ignores the fact that the time dimension has a much different intuitive meaning than the other two. Apart from the progressive (forward–only) direction of time, there is the more important fact (even for stored image sequences) that a snapshot in time (e.g. a still picture selected from the sequence) often has immediate meaning to the human observer. A single row of pixels across a picture, or a single column of pixels down a picture do not generally have similar significance to the human observer. Likewise a single pixel means little to the human observer in the absence of surrounding pixels.

Notwithstanding their utility, slices of the form

$f(x,t)$ or of the form $f(y,t)$ are often confusing at best, compared to the still picture $f(x,y)$ which remains as an extraction from a picture sequence that is far more meaningful to a typical human observer.

Thus we believe that downsampling across rows or down columns of an image should be preceeded by lowpass filtering, whereas temporal downsampling should not.

There is therefore a special significance to the notion of a "snapshot in time" and the processing, storage, transmission, etc., of a motion picture as a sequence of such "snapshots". The relationship between individual sharply defined frames of an "Edgertonian" sequence of pictures gave rise to the development of a new framework for transmitting images across the wireless links used in the ENGwear project.

III.C. Picture Transfer Protocol (PTP)

When applying data compression to a stream of individual pictures that will be viewed in realtime (for example in videoconferencing, such as the first–person–perspective videoconferencing of the wearable Eye Tap device), it is helpful to take into account the manner in which the data will be sent.

Most notably, pictures are typically sent over a packet–based communications channel. For example, Wearable Wireless Webcam used the AX25 Amateur Radio protocol. Accordingly, packets typically either arrive intact or are corrupt. Packets that are corrupt, traditionally, would be re–sent.

An interesting approach is to provide data compression on a per–image basis, and to vary the degree of compression so that the size of each picture in the image sequence is exactly equal to the length of one packet.

Together with the prior assumption (that images are acquired at a rate that exceeds the channel capacity), it will generally be true that by the time we know that a packet (which is a complete picture) is corrupt at the receiver, there will have already been acquired a newer picture than the one that was sent. Suppose, for example, the round trip time (RTT) is 100 ms (which is equal to the time it takes to generate three pictures), there would be little sense in re–sending a picture that was taken three pictures ago.

The commonly arising situation in which pictures are captured at a rate that exceeds the RTT suggests that there will always be newer picture information at the transmit site than what would be re–sent in the event of a lost packet.

This approach forms the basis for the proposed "Picture Transfer Protocol" (PTP). In particular, PTP is based on the idea of treating each "snapshot" in time as a single entity, in isolation, and compressing it into a single packet, so that it will have either arrived in its entirity or not arrived at all (and therefore be discarded). It should be clear that the philisophical underpinnings of PTP are closely related to those of Edgertonian downsampling.

III.D. Best case imaging and fear of functionality

A direct result of Edgertonian sampling is that a single picture from a picture sequence has a high degree of relevance and meaning even when it is taken in isolation.

Similarly, a direct result of PTP is that a single packet from a packet sequence has a high degree of relevance and meaning even when it is taken in isolation (for example when the packets before and after it have been corrupted).

It is therefore apparent that if a system were highly unreliable, to the extent that pictures could only be transmitted occasionally, and unpredictably, then the Edgertonian sampling combined with PTP would provide a system that would degrade gracefully.

Indeed, if we were to randomly select just a few frames from one of Harold Edgerton's motion pictures, we would likely have a good summary of the motion picture, since any given frame would provide us with a sharp picture in which subject matter of interest could be clearly discerned. Likewise, if we were to randomly select a few packets from a stream of thousands of packets of PTP, we would have data that would provide a much more meaningful interpretation to the human observer than if all we had was randomly selected packets from an MPEG sequence.

III.E. ENGwear as Personal Safety Device

Accordingly, an object of the Personal Imaging project is to provide a system that transmits pictures in harsh or hostile environments. One such application of such a system is the Personal Safety Device described in [6].

The Personal Safety Device differs from other wireless data transmission systems in the sense that it was designed for "Best Case" operation. Ordinarily, wireless transmissions are designed for worst case scenarios, such as might guarantee a certain minimum level of performance throughout a large metropolitan area. The Personal Safety Device, however, is designed to only make it hard for an adversary to guarantee total

(a) (b)

Figure 2: Lookpaintings from the ENGwear project (a) During the St. George Street Carnival, an ENGwear reporter obtained a lookpainting of a recently finished street painting. The curving gaze of the ENGwear reporter caused the irregularly shaped border of this image. (b) A sequence of images captures the St. George Street Carnival, with the ENGwear reporter starting his gaze at the top of a nearby building and panning down to street level.

nonperformance.

It is not a goal of the Personal Safety Device to guarantee connectivity in the presence of hostile jamming of the radio spectrum, but, rather, the goal is to make it difficult for the adversary to guarantee the absence of connectivity. Therefore, an otherwise potential perpetrator of a crime would never be able to be certain that the wearer's device was non–operational and would therefore need to be on his or her best behaviour at all times. This is the Fear of Functionality (FoF) model.

Since there is, and would be, the possibility of just one packet, which contains just one picture, providing incriminating evidence of wrongdoing, an individual can simply wear a personal safety device to obtain protection from criminals, assailants, and attackers.

The personal safety device need not work constantly, but, rather, must merely present criminals with the possibility that it could work sometimes. This scenario forms the basis for best-case design as an alternative to the usual worst-case design paradigm.

The Personal Imaging system therefore transmits video, but the design of the system is such that it will, at the very least, occasionally transmit a meaningful still image.

ENGwear Personal Imaging systems can thus allow transmission from environments where there may be active resistance to news gathering efforts. With fully covert systems, it becomes possible for ENGwear reporters to operate with little chance of detection from areas where other reporters would spotted and evicted. Once transmitted, destruction of the images becomes difficult as images are duplicated and retransmitted by servers across the internet. While video tapes or photographic negatives could be confiscated or destroyed, images transmitted wirelessly can propogate quickly amongst internet servers, making their containment or destruction difficult.

Typically the wearer's jacket functions as a large low frequency antenna, providing transmission capability in a frequency band that is very hard to stop. For example, the 10 meter band is a good choice because of its unpredictable performance (owing to various "skip" phenomena, etc.).

IV. The ENGwear Project

Two different architectures were experimented with in this first trial of ENGwear systems. One consisted of an Intel based 486 archicture, capturing images with a commercial, black and white CCD. The second system was based on a 275 MHz StrongArm architecture which incorporated NTSC input from a broadcast quality, full colour television camera. Both systems captured images with the 45 degree diverter implementation of the EyeTap system.

The experiment lasted four days during the Engineering Orientation Week at the University of Toronto. It was intended to disguise the non-covert eyetap devices in Engineering faculty hard hats, which have a long standing tradition of decoration by upper year Engineering students. Despite the fact that the EyeTap diverters used were large and visible, since they were mounted on hard hats, they blended into the festivities naturally. In this way, they were used in a manner similar to the smaller versions of the EyeTap technology [4], to obtain similarly candid interaction.

Just as the work of a photojournalist can be regarded as an artistic form of expression, VideoOrbits allows ENGwear reporters to create high resolution expressive images of their subject matter. This occurs when an ENGwear reporter captures a sequence of images in rapid succession, while sweeping her gaze across her total field of view. Essentially, this sequence captures information of greater spatial extent, with the temporal differences between images remaining small. When many such images are combined into a Photoquantigraphic Image Composite, what results is a high resolution lookpainting. The shape of the border of the lookpainting will reflect the gaze of the wearer. Thus, if the wearer captures a sequence moving their head in a curving motion, so too will the border of the lookpainting have the same curving motion. Some compelling lookpaintings were captured during the span of the ENGwear project, depicted in Figure 2.

As local media became interested in the project, the ENGwear students found themselves providing media coverage of the traditional media, who had come to provide coverage of the ENGwear reporters. A sort of "meta-media" situation resulted in which the news gatherers became news themselves, for both ENGwear and traditional reporters alike.

V. Conclusions

Electronic News Gathering has emerged as a novel new application in the field of Personal Imaging. Wireless communications and WearComps now make a new genre of news gathering possible – a new sort of photojournal lying somewhere between live television coverage and newsprint, captured from the very eyes that witness events as they happen. New capabilites are afforded to ENGwear reporters stemming from the Edgertonian nature of WearComp video input capture, best case scenario models, and the operational constancy of WearComps. New capabilities include: the potential to broadcast, using PTP, from areas unreachable by traditional media teams; the capability to record in an unobtrusive manner; the ability to instantly record images and retrieve them from the recent past if necessary; and a greater degree of mobility. ENGwear systems remove the requirement for expensive video equipment and large mobile transmitters, allowing any WearComp with meagre computational power and even a low bandwidth transmitter to provide live coverage of events – a privledge which can now be afforded not only to large broadcasting corporations, but to common individuals. With many ENGwear systems functional and worn day–to–day by private citizens, the face of news coverage could be changed drastically as the view of the common citizen comes increasingly into focus.

VI. Further information

For further information on the licensing of EyeTap technology, email mann@eecg.toronto.edu, and cc admin@eyetap.org

References

[1] S. Mann. 'mediated reality'. TR 260, M.I.T. M.L. vismod, Cambridge, Massachusetts, http://wearcam.org/mr.htm, 1994.

[2] Steve Mann. Personal imaging and lookpainting as tools for personal documentary and investigative photojournalism. *ACM Mobile Networking*, 4(1):23–36, 1999. Special Issue on Wearable Computing.

[3] S. Mann. Compositing multiple pictures of the same scene. In *Proceedings of the 46th Annual IS&T Conference*, pages 50–52, Cambridge, Massachusetts, May 9-14 1993. The Society of Imaging Science and Technology.

[4] Steve Mann. Humanistic intelligence/humanistic computing: 'wearcomp' as a new framework for intelligent signal processing. *Proceedings of the IEEE*, 86(11):2123–2151+cover, Nov 1998. http://wearcam.org/procieee.htm.

[5] S. Mann. Wearable Wireless Webcam, 1994. http://wearcam.org.

[6] Steve Mann. Smart clothing: The wearable computer and wearcam. *Personal Technologies*, pages 21–27, March 1997. Volume 1, Issue 1.