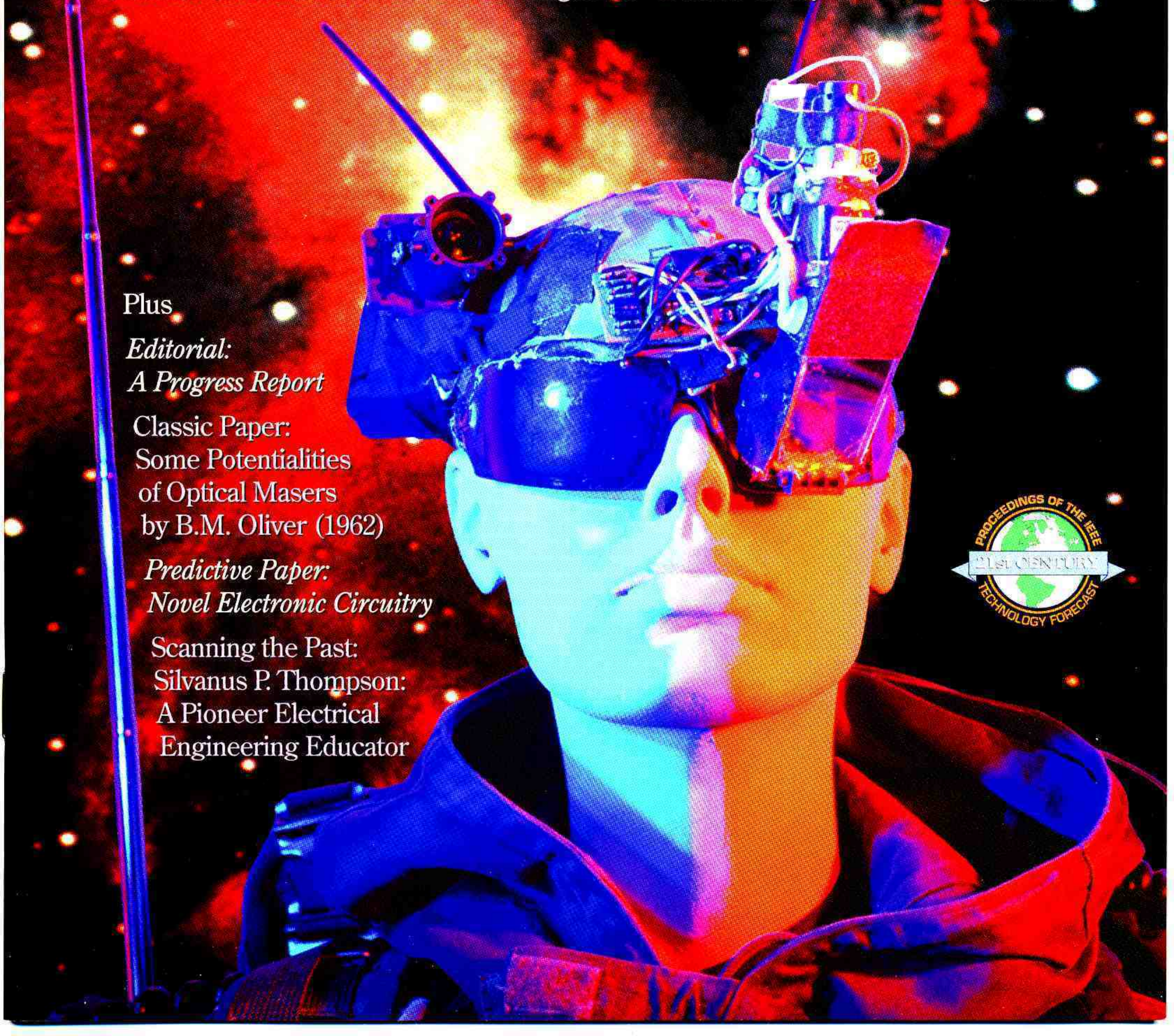
PROCEEDINGS OF THE IEEE

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC. NOVEMBER 1998

Special Issue on:

INTELLIGENT SIGNAL PROCESSING

*Papers on: Scanning the Technology * Humanistic Computing * Adaptive Stochastic Resonance * Noise Processing * Virtual Examples * Deterministic Annealing * Map Modeling * Neural Net Framework * Document Recognition * Pattern Classification & Recognition*



Plus

Editorial:
A Progress Report

Classic Paper:
Some Potentialities
of Optical Masers
by B.M. Oliver (1962)

Predictive Paper:
Novel Electronic Circuitry

Scanning the Past:
Silvanus P. Thompson:
A Pioneer Electrical
Engineering Educator

PROCEEDINGS OF THE IEEE
21st CENTURY
TECHNOLOGY FORECAST

Humanistic Computing: "WearComp" as a New Framework and Application for Intelligent Signal Processing

STEVE MANN, MEMBER, IEEE

Invited Paper

Humanistic computing is proposed as a new signal processing framework in which the processing apparatus is inextricably intertwined with the natural capabilities of our human body and mind. Rather than trying to emulate human intelligence, humanistic computing recognizes that the human brain is perhaps the best neural network of its kind, and that there are many new signal processing applications (within the domain of personal technologies) that can make use of this excellent but often overlooked processor. The emphasis of this paper is on personal imaging applications of humanistic computing, to take a first step toward an intelligent wearable camera system that can allow us to effortlessly capture our day-to-day experiences, help us remember and see better, provide us with personal safety through crime reduction, and facilitate new forms of communication through collective connected humanistic computing. The author's wearable signal processing hardware, which began as a cumbersome backpack-based photographic apparatus of the 1970's and evolved into a clothing-based apparatus in the early 1980's, currently provides the computational power of a UNIX workstation concealed within ordinary-looking eyeglasses and clothing. Thus it may be worn continuously during all facets of ordinary day-to-day living, so that, through long-term adaptation, it begins to function as a true extension of the mind and body.

Keywords**—***Consumer electronics, cybernetic sciences, human factors, humanistic property protection, image processing, machine vision, mobile communication, photoquantigraphic imaging, signals.*

I. INTRODUCTION

What is now proposed is a new form of "intelligence" whose goal is not only to work in extremely close synergy with the human user, rather than as a separate entity, but more importantly to arise, in part, because of the very existence of the human user. This close synergy is achieved through a user-interface to signal processing hardware that is both in close physical proximity to the user and is constant.

The constancy of user-interface (interactional constancy) is what separates this signal processing architecture from other related devices such as pocket calculators and personal digital assistants (PDA's).

Not only is the apparatus operationally constant in the sense that although it may have power saving (sleep) modes, it is never completely shut down (or "dead," as is a calculator worn in a shirt pocket but turned off most of the time). More important is the fact that it is also interactionally constant. By interactionally constant what is meant is that the inputs and outputs of the device are always potentially active. Interactionally constant implies operationally constant, but operationally constant does not necessarily imply interactionally constant. Thus, for example, a pocket calculator, worn in a shirt pocket and left on all the time is still not interactionally constant because it cannot be used in this state (e.g., one still has to pull it out of the pocket to see the display or enter numbers). A wrist watch is a borderline case; although it operates constantly in order to continue to keep proper time and is conveniently worn on the body, one must make a conscious effort to orient it within one's field of vision in order to interact with it.

A. Why Humanistic Computing?

It is not, at first, obvious why one might want devices such as pocket calculators to be operationally constant. However, we will later see why it is desirable for personal electronics devices such as cameras and signal processing hardware to be on constantly, for example, to facilitate new forms of intelligence that assist the user in new ways.

Devices embodying humanistic computing are not merely intelligent signal processors that a user might wear or carry in close proximity to the body, but instead they are devices that turn the user into part of an intelligent control

---

Manuscript received November 1, 1997; revised April 17, 1998.

The author is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ont. M5S 3G4 Canada (e-mail: mann@eecg.toronto.edu).

Publisher Item Identifier S 0018-9219(98)07857-8.

system where the user becomes an integral part of the feedback loop.

B. Humanistic Computing Does Not Necessarily Mean "User-Friendly"

Devices embodying humanistic computing often require that the user learn a new skill set and therefore do not always allow for ease in adaptation. Just as it takes a young child many years to become proficient at using his or her hands, some of the devices that implement humanistic computing have taken years of use before they began to truly behave as if they were natural extensions of the mind and body. Thus, in terms of human-computer interaction [1], the goal is not just to construct a device that can model (and learn from) the user, but, more importantly, to construct a device in which the user also must learn from the device. Therefore, in order to facilitate the latter, devices embodying humanistic computing should provide a constant user-interface—one that is not so sophisticated and intelligent that it confuses the user. Although the device may implement very sophisticated signal processing algorithms, the cause and effect relationship of this processing to its input (typically from the environment or the user's actions) should be clearly and continuously visible to the user, even when the user is not directly and intentionally interacting with the apparatus. Accordingly, the most successful examples of humanistic computing afford the user a very tight feedback loop of system observability (ability to perceive how the signal processing hardware is responding to the environment and the user), even when the controllability of the device is not engaged (e.g., at times when the user is not issuing direct commands to the apparatus). A simple example is the viewfinder of a wearable camera system, which provides framing, a photographic point of view, and facilitates the provision to the user of a general awareness of the visual effects of the camera's own image processing algorithms, even when pictures are not being taken. Thus a camera embodying humanistic computing puts the human operator in the feedback loop of the imaging process at all times, even when the operator only wishes to take pictures occasionally. A more sophisticated example of humanistic computing is a biofeedback-controlled wearable camera system, in which the biofeedback process happens continuously, whether or not a picture is actually being taken. In this sense, the user becomes one with the machine over a long period of time, even if the machine is only directly used (e.g., to actually take a picture) occasionally.

Humanistic computing attempts to both build upon, as well as recontextualize, concepts in intelligent signal processing [2], [3], and related concepts such as neural networks [2], [4], [5], fuzzy logic [6], [7], and artificial intelligence (AI) [8]. Humanistic computing also suggests a new goal for signal processing hardware, that is, in a truly personal way, to directly assist, rather than replace or emulate human intelligence. What is needed to facilitate this vision is a simple and truly personal computational signal processing framework that empowers the human intellect. It should be noted that this framework which the author developed in the 1970's and early 1980's is in many ways similar to Engelbart's vision that arose in the 1940's while he was a radar engineer, but there are also some important differences. Engelbart, while seeing images on a radar screen, envisioned that the cathode ray screen could also display letters of the alphabet, as well as computer-generated pictures and graphical content, thus he envisioned computing as an interactive experience for manipulating words and pictures. Engelbart envisioned the mainframe computer as a tool for augmented intelligence and augmented communication with which a number of people in a large amphitheater could interact with one another using a large mainframe computer [9], [10].

While Engelbart himself did not realize the significance of the personal computer, his ideas are certainly embodied in modern personal computing. What is now described is a means of realizing a similar vision, but with the computing resituated in a different context, namely, the truly personal space of the user. The idea here is to move the tools of augmented intelligence and augmented communication directly onto the body, giving rise to not only a new genre of truly personal computing, but to some new capabilities and affordances arising from direct physical contact between the computational apparatus and the human body. Moreover, a new family of applications arises, such as "personal imaging," in which the body-worn apparatus facilitates an augmenting of the human sensory capabilities, namely, vision. Thus the augmenting of human memory translates directly to a visual associative memory in which the apparatus might, for example, play previously recorded video back into the wearer's eyeglass mounted display, in the manner of a so-called visual thesaurus [11].

II. "WEARCOMP" AS MEANS OF REALIZING HUMANISTIC COMPUTING

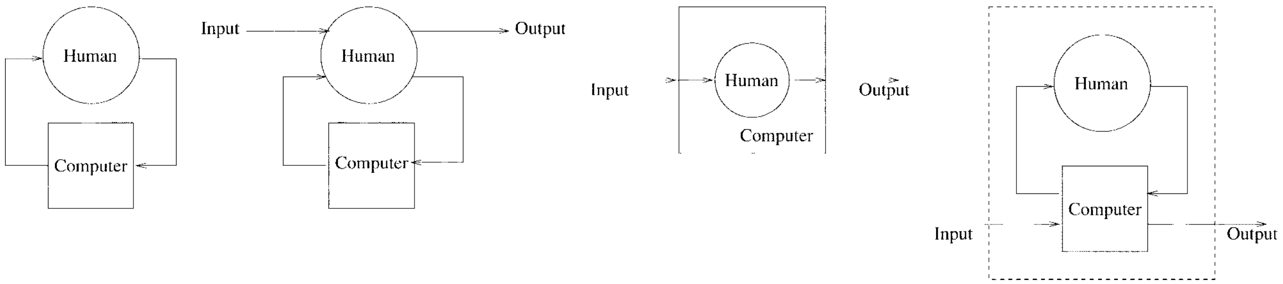
WearComp [12] is now proposed as an apparatus upon which a practical realization of humanistic computing can be built, as well as a research tool for new studies in intelligent signal processing.

A. Basic Principles of WearComp

WearComp will now be defined in terms of its three basic modes of operation.

1) Operational Modes of WearComp: The three operational modes in this new interaction between human and computer, as illustrated in Fig. 1, are as follows.

1) Constancy: The computer runs continuously and is "always ready" to interact with the user. Unlike a hand-held device, laptop computer, or PDA, it does not need to be opened up and turned on prior to use. The signal flow from human to computer and from computer to human depicted in Fig. 1(a) runs continuously to provide a constant user-interface.
2) Augmentation: Traditional computing paradigms are based on the notion that computing is the primary task. WearComp, however, is based on the notion that



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

Fig. 1. The three basic operational modes of WearComp. (a) Signal flow paths for a computer system that runs continuously, is constantly attentive to the user's input, and is constantly providing information to the user. Over time, constancy leads to a symbiosis in which the user and computer become part of each other's feedback loops. (b) Signal flow path for augmented intelligence and augmented reality. Interaction with the computer is secondary to another primary activity, such as walking, attending a meeting, or perhaps doing something that requires full hand-to-eye coordination, like running down stairs or playing volleyball. Because the other primary activity is often one that requires the human to be attentive to the environment as well as unencumbered, the computer must be able to operate in the background to augment the primary experience, for example, by providing a map of a building interior, or by providing other information, through the use of computer graphics overlays superimposed on top of the real world. (c) WearComp can be used like clothing to encapsulate the user and function as a protective shell, whether to protect us from cold, protect us from physical attack (as traditionally facilitated by armor), or to provide privacy (by concealing personal information and personal attributes from others). In terms of signal flow, this encapsulation facilitates the possible mediation of incoming information to permit solitude, and the possible mediation of outgoing information to permit privacy. It is not so much the absolute blocking of these information channels that is important; it is the fact that the wearer can control to what extent and when these channels are blocked, modified, attenuated, or amplified by invarious degrees, that makes WearComp much more empowering to the user than other similar forms of portable computing. (d) An equivalent depiction of encapsulation (mediation) redrawn to give it a similar form to that of (a) and (b), where the encapsulation is understood to comprise a separate protective shell.

computing is NOT the primary task. The assumption of WearComp is that the user will be doing something else at the same time as doing the computing. Thus the computer should serve to augment the intellect, or augment the senses. The signal flow between human and computer, in the augmentational mode of operation, is depicted in Fig. 1(b).

3) Mediation: Unlike hand held devices, laptop computers, and PDA's, WearComp can encapsulate the user [Fig. 1(c)]. It does not necessarily need to completely enclose us, but the basic concept of mediation allows for whatever degree of encapsulation might be desired since it affords us the possibility of a greater degree of encapsulation than traditional portable computers. Moreover, there are two aspects to this encapsulation, one or both of which may be implemented in varying degrees, as desired.

   a) Solitude: The ability of WearComp to mediate our perception can allow it to function as an information filter, and allow us to block out material we might not wish to experience, whether it be offensive advertising or simply a desire to replace existing media with different media. In less extreme manifestations, it may simply allow us to alter aspects of our perception of reality in a moderate way rather than completely blocking out certain material. Moreover, in addition to providing means for blocking or attenuation of undesired input, there is a facility to amplify or enhance desired inputs. This control over the input space is one of the important contributors to the most fundamental issue in this new framework, namely that of user empowerment.

   b) Privacy: Mediation allows us to block or modify information leaving our encapsulated space. In the same way that ordinary clothing prevents others from seeing our naked bodies, WearComp may, for example, serve as an intermediary for interacting with untrusted systems, such as third party implementations of digital anonymous cash, or other electronic transactions with untrusted parties. In the same way that martial artists, especially stick fighters, wear a long black robe that reaches to the ground in order to hide the placement of their feet from their opponent, WearComp can also be used to clothe our otherwise transparent movements in cyberspace. Although other technologies like desktop computers can, to a limited degree, help us protect our privacy with programs like Pretty Good Privacy (PGP), the primary weakness of these systems is the space between them and their user. It is generally far easier for an attacker to compromise the link between the human and the computer (perhaps through a so-called Trojan horse or other planted virus) when they are separate entities. Thus a personal information system owned, operated, and controlled by the wearer can be used to create a new level of personal privacy because it can be made

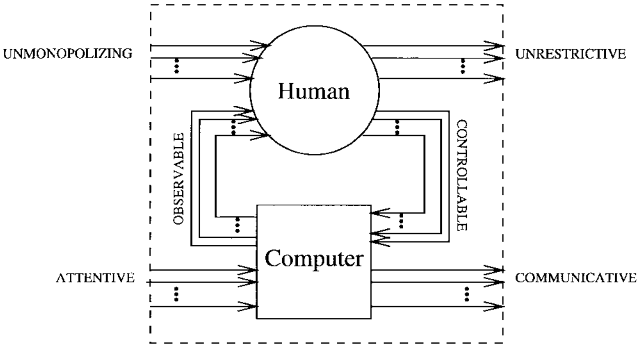


Fig. 2. The six signal flow paths for the new mode of human–computer interaction provided by WearComp. These six signal flow paths each define one of the six attributes of WearComp.

much more personal, e.g., so that it is always worn (except perhaps during showering) and is therefore less likely to fall prey to attacks upon the hardware itself. Moreover, the close synergy between the human and computers makes it harder to attack directly, e.g., as one might look over a person's shoulder while they are typing, or hide a video camera in the ceiling above their keyboard.[1]

Because of its ability to encapsulate us, e.g., in embodiments of WearComp that are actually articles of clothing in direct contact with the flesh, it may also be able to make measurements of various physiological quantities. Thus the signal flow depicted in Fig. 1(a) is also enhanced by the encapsulation, as depicted in Fig. 1(c). To make this signal flow more explicit, Fig. 1(c) has been redrawn in Fig. 1(d), where the computer and human are depicted as two separate entities within an optional protective shell, which may be opened or partially opened if a mixture of augmented and mediated interaction is desired.

Note that these three basic modes of operation are not mutually exclusive in the sense that the first is embodied in both of the other two. Also, these other two are not meant necessarily to be implemented in isolation. Actual embodiments of WearComp typically incorporate aspects of both augmented and mediated modes of operation. Thus, WearComp is a framework for enabling and combining various aspects of each of these three basic modes of operation. Collectively, the space of possible signal flows giving rise to this entire space of possibilities is depicted in Fig. 2. The signal paths typically comprise vector quantities. Thus, multiple parallel signal paths are depicted in Fig. 2 to remind the reader of this vector nature of the signals.

B. The Six Basic Signal Flow Paths of WearComp

There are six informational flow paths associated with this new human–machine symbiosis. These signal flow paths each define one of the basic underlying principles of WearComp, and are each described in the following from the human's point of view. Implicit in these six properties is that the computer system is also operationally constant and personal (inextricably intertwined with the user).

1) *Unmonopolizing of the User's Attention*: WearComp does not necessarily cut one off from the outside world like a virtual reality game does. One 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, facilitate mediation (augmenting, altering, or deliberately diminishing) of these sensory capabilities.
2) *Unrestrictive to the User*: While being ambulatory, mobile or roving, one can do other things while using WearComp, e.g., one can type while jogging, running down stairs, etc.
3) *Observable by the User*: WearComp can get the user's attention continuously if the user wants it to. The output medium is constantly perceptible by the wearer. It is sufficient that it be almost always observable, within reasonable limitations, such as the fact that a camera viewfinder or computer screen is not visible during the blinking of the eyes.
4) *Controllable by the User*: WearComp is responsive. The user can take control of it at any time the user wishes. Even in automated processes the user should be able to manually override the automation to break open the control loop and become part of the loop at any time the user wants to. Examples of this controllability might include a "Halt" button the user can invoke as an application mindlessly opens all 50 documents that were highlighted when the user accidently pressed "Enter."
5) *Attentive to the Environment*: WearComp is environmentally aware, multimodal, and multisensory. (As a result, this ultimately gives the user increased situational awareness.)
6) *Communicative to Others*: WearComp can be used as a communications medium when the user wishes. WearComp allows the wearer to be expressive through the medium, whether as a direct communications medium to others, or as a means of assisting the user in the production of expressive or communicative media.

III. PHILOSOPHICAL ISSUES

There are many open questions in this new area of research. Some of these include the following.

1) Is WearComp good?
2) Do we need it?

---

[1] For the purposes of this paper, privacy is not so much the absolute blocking or concealment of personal information, but it is the ability to control or modulate this outbound information channel. Thus, for example, one may wish certain people, such as members of one's immediate family, to have greater access to personal information than the general public. Such a family-area-network may be implemented with an appropriate access control list and a cryptographic communications protocol.

3) Could it be harmful to the user?
4) Could it be harmful to society?
5) Are humans prepared for such a close synergy with machines, or will nature "strike back?"
6) Will the apparatus modify the behavior of the wearer in an undesirable way?
7) Will it become a cause of irritation to others?

As with many new inventions, such as clothing, the bicycle, hot air balloons, etc., there has been an initial rejection, followed by scientific study and experimentation, followed by either acceptance or modification of the invention to address specific problems.

For example, the adverse effects of the bright screen constantly projected onto the eye were addressed by building the apparatus into very dark sunglasses so that a much lower brightness could be used.

More important is perhaps the danger of becoming dependent on the technology in the same way that we have become dependent on shoes, clothing, the automobile, etc. For example, the fact that we cannot survive naked in the wilderness, or that we have become sedentary because of the automobile, must have its equivalent problems within the context of the proposed invention.

Many of these issues are open philosophical questions that will only be answered by further research. However, the specific framework and some of the various ideas surrounding it will hopefully form the basis for a further investigation of some of these questions.

A. Fundamental Issues of WearComp

The most fundamental paradigm shift that WearComp has to offer is that of personal empowerment. In order to fully appreciate the magnitude of this paradigm shift, some historical examples of tools of empowerment will now be described to place WearComp in this historical context.

1) Historical Context: In early civilization, individuals were all roughly equal, militarily. Wealth was generally determined by how many heads of cattle or how many "mounts" (horses) a person owned. In hand-to-hand combat or fighting with swords each individual was roughly an equal. Since it was impossible to stay on a horse while fighting, horses provided little in the way of military power. Thus, even those too poor to afford to keep a horse were not at a tremendous disadvantage to others from a fighting standpoint.

It was the invention of the stirrup, however, that radically changed this balance. With the stirrup, it became possible to stay on a horse while fighting. Horses and heavy armor could only be afforded by the wealthy, and even a large group of unruly peasants was no match for a much smaller group of mounted cavalry. However, toward the Middle Ages, more and more ordinary individuals mastered the art of fighting on horseback, and eventually the playing field was leveled.

Then, with the invention of gunpowder, the ordinary civilian was powerless against soldiers or bandits armed with guns. It was not until guns became cheaper that everyone could own one—such as in the "Old West." The Colt 45, for example, was known as the "equalizer" because it made everyone roughly equal. Even if one person was much more skilled in its use, there would still be some risk involved in robbing other civilians or looting someone's home.

2) The Shift from Guns to Cameras and Computers: In today's world, the handgun has a lesser role to play. Wars are fought with information, and we live in a world in which the appearance of thugs and bandits is not ubiquitous. While there is some crime, we spend most of our lives living in relative peace. However, surveillance and mass media have become the new instruments of social control. Department stores are protected with security cameras rather than by owners who keep a shotgun under the counter or who hire armed guards to provide a visible deterrent. While some department stores in rough neighborhoods may have armed guards, there has been a paradigm shift where we see fewer guns and more surveillance cameras.

3) The Shift from Draconian Punishment to Micromanagement: There has also been a paradigm shift, throughout the ages, characterized by a move toward less severe punishments, inflicted with greater certainty. In the Middle Ages, the lack of sophisticated surveillance and communications networks meant that criminals often escaped detection or capture, but when they were captured, punishments were extremely severe. Gruesome corporeal punishments where criminals might be crucified or whipped, branded, drawn and quartered, and burned at the stake were quite common in these times.

The evolution from punishment as a spectacle in which people were tortured to death in the village square toward incarceration, in which people were locked in a cell and forced to attend church sermons, prison lectures, etc., marked the first step in a paradigm shift toward less severe punishments [13]. Combined with improved forensic technologies like fingerprinting, this reduction in the severity of punishment came together with a greater chance of getting caught.

More recently, with the advent of so-called "boot camp," where delinquent youths are sent off for mandatory military-style training, the trend continues by addressing social problems earlier before they become large problems. This requires greater surveillance and monitoring, but at the same time it is characterized by less severe actions taken against those who are deemed to deserve these actions. Thus there is, again, still greater chance of being affected by smaller punishments.

If we extrapolate this trend, what we arrive at is a system of social control characterized by total surveillance and micropunishments. At some point, the forces applied to the subjects of the social control are too weak to even justify the use of the word "punishment," and perhaps it might be better referred to as "micromanagement."

This "micromanagement" of society may be effected by subjecting the population to mass media, advertising, and calming music played in department stores, elevators, and subway stations.

Surveillance is also spreading into areas that were generally private in earlier times. The surveillance cameras that were placed in banks have moved to department stores. They first appeared above cash registers to deal with major crimes, such as holdups. But then they moved into the aisles and spread throughout the store to deal with petty theft. Again, more surveillance for dealing with lesser crimes.

In the United Kingdom, cameras installed for controlling crime in rough areas of town spread to low-crime areas as well in order to deal with problems like youths stealing apples from street markets or patrons of pubs urinating on the street. The cameras have even spread into restaurants and pubs—not just above the cash register, but throughout the pub, so that while going out for pints one may no longer have privacy.

Recently, electronic plumbing technology, originally developed for use in prisons, for example, to prevent all inmates from flushing the toilets simultaneously, has started to be used in public buildings. The arguments in favor of it go beyond human hygiene and water conservation, as proponents of the technology argue that it also reduces vandalism. Their definition of vandalism has been broadened to include deliberately flooding a plumbing fixture and deliberately leaving faucets running. Thus, again, what we see is greater certainty of catching or preventing people from committing lesser transgressions of the social order.

One particularly subtle form of social control using this technology are the new hands-free electronic showers developed for use in prisons where inmates would otherwise break off knobs, levers, and pushbuttons. These showers are just beginning to appear in government buildings, stadiums, health clubs, and schools. The machine watches the user from behind a tiled wall and through a small dark glass window. When the user steps toward the shower, the water comes on, but only for a certain time, and then it shuts off. Obviously the user can step away from the viewing window and then return to receive more water, thus defeating the timeout feature of the system, but this need to step away and move back into view is enough of an irritant to effect a slight behavioral modification of the user. Thus what we see is that surveillance has swept across all facets of society, but it is being used to deal with smaller and smaller problems. From dealing with mass murderers and bank robbers to people who threaten the environment by taking long showers, the long arm of surveillance has reached into even the most private of places, where we might have once been alone. The peace and solitude of the shower, where our greatest inspirations might come to us, has been intruded upon not with a major punishment, but with a very minor form of social control, one that is too small, in fact, to even be called a punishment.

These surveillance and social control systems are linked together, often to central computer systems. Everything from surveillance cameras in the bank to electronic plumbing networks is being equipped with fiber optic communications networks. Together with the vast array of medical records, credit card purchases, buying preferences, etc., we are affected in more ways, but with lesser influence. We are no longer held at bay by mounted cavalry. More often than being influenced by weapons, we are influenced in very slight, almost imperceptible ways, e.g., through a deluge of junk mail, marketing, advertising, or a shower that shuts off after it sees that we have been standing under it for too long.

While there are some (the most notable being Bentham [13]) who have put forth an argument that a carefully managed society results in maximization of happiness, there are others who argue that the homogenization of society is unhealthy and reduces humans to cogs in a larger piece of machinery, or at the very least, results in a certain loss of human dignity. Moreover, just as nature provides biodiversity, many believe that society should also be diverse, and people should try to resist ubiquitous centralized surveillance and control, particularly to the extent where it homogenizes society excessively. Some argue that micromanagement and utilitarianism, in which a person's value may often be measured in terms of usefulness to society, is what led to eugenics, and eventually to the fascism of Nazi Germany. Many people also agree that, even without any sort of social control mechanism, surveillance in and of itself still violates their privacy and is fundamentally wrong.

As with other technologies, like the stirrup and gunpowder, the electronic surveillance playing field is also being leveled. The advent of the low-cost personal computer has allowed individuals to communicate freely and easily among themselves. No longer are the major media conglomerates the sole voice heard in our homes. The World Wide Web has ushered in a new era of underground news and alternative content. Thus, centralized computing facilities, the very technology that many perceived as a threat to human individuality and freedom, have given way to low-cost personal computers that many people can afford. This is not to say that home computers will be as big or powerful as the larger computers used by large corporations or governments, but simply that if a large number of people have a moderate degree of computational resources, there is a sense of balance in which people are roughly equal in the same sense that two people, face to face, one with a 0.22 calibre handgun and the other with a Colt .45 are roughly equal. A large bullet hole or a small one both provide a tangible and real risk of death or injury.

It is perhaps modern cryptography that makes this balance even more pronounced, for it is so many orders of magnitude easier to encrypt a message than it is to decrypt it. Accordingly, many governments have defined cryptography as a munition and have attempted, with only limited success, to restrict its use.

4) Fundamental Issues of WearComp: The most fundamental issue in WearComp is no doubt that of personal empowerment through its ability to equip the individual with a personalized, customizable information space which is owned, operated, and controlled by the wearer. While home computers have gone a long way to empowering the individual, they only do so when the user is at home. As the home is perhaps the last bastion of space not yet touched by the long arm of surveillance (i.e., space that one can call one's own), the home computer, while it does provide an increase in personal empowerment, is not nearly so profound in its effect as the WearComp which brings this personal space (space one can call one's own) out into the world.

Although WearComp, in the most common form we know it today (miniature video screen over one or both eyes, body worn processor, and input devices such as a collection of pushbutton switches or joystick held in one hand and a microphone), was invented by the author in the 1970's for personal imaging applications, it has more recently been adopted by the military in the context of large government-funded projects. However, as with the stirrup, gunpowder, and other similar inventions, it is already making its way out into the mainstream consumer electronics arena.

An important observation to make with regards to the continued innovation, early adopters (military, government, large multinational corporations), and finally ubiquity, is the time scale. While it took a relatively longer time for the masses to adopt the use of horses for fighting and hence level the playing field, later on the use of gunpowder became ubiquitous in a much shorter time period.

Then, sometime after guns had been adopted by the masses, the spread of computer technology, which in some situations even replaced guns, was much faster still. As the technology diffuses into society more quickly, the military is losing its advantage over ordinary civilians. We are entering a pivotal era in which consumer electronics are surpassing the technological sophistication of some military electronics. Personal audio systems like the Sony Walkman are just one example of how the ubiquity and sophistication of technology feed upon each other to the extent that the technology begins to rival, and in some ways exceed, the technical sophistication of the limited-production military counterparts such as two-way radios used in the battlefield.

Consumer technology has already brought about a certain degree of personal empowerment, from the portable cassette player that lets us replace the music piped into department stores with whatever we would rather hear to small hand held cameras that capture police brutality and human rights violations. However, WearComp is just beginning to bring about a much greater paradigm shift, which may well be equivalent in its impact to the invention of the stirrup or that of gunpowder. Moreover, this leveling of the playing field may, for the first time in history, happen almost instantaneously should the major consumer electronics manufacturers beat the military in raising this invention to a level of perfection similar to that of the stirrup or modern handguns. If this were to happen, this decrease of the time scale over which technology diffuses through society will have decreased to zero, resulting in a new kind of paradigm shift that society has not yet experienced. Evidence of this pivotal shift is already visible in, for example, the joint effort of Xybernaut Corp. (a major manufacturer of wearable computers) and Sony Corp. (a manufacturer of personal electronics) to create a new personal electronics computational device.

B. Aspects of WearComp and Personal Empowerment

There are several aspects and affordances of WearComp. These are as follows.

1) *Photographic/Videographic Memory*: Perfect recall of previously collected information, especially visual information (i.e., visual memory [14]).
2) *Shared Memory*: In a collective sense, two or more individuals may share in their collective consciousness, so that one may have a recall of information that one need not have experienced personally.
3) *Connected Collective Humanistic Computing*: In a collective sense, two or more individuals may collaborate while one or more of them is doing another primary task.
4) *Personal Safety*: In contrast to a centralized surveillance network built into the architecture of the city, a personal safety system is built into the architecture (clothing) of the individual. This framework has the potential to lead to a distributed "intelligence" system of sorts, as opposed to the centralized "intelligence" gathering efforts of traditional video surveillance networks.
5) *Tetherless Operation*: WearComp affords and requires mobility and the freedom from the need to be connected by wire to an electrical outlet or communications line.
6) *Synergy*: Rather than attempting to emulate human intelligence in the computer, as is a common goal of research in AI, the goal of WearComp is to produce a synergistic combination of human and machine in which the human performs tasks that it is better at while the computer performs tasks that it is better at. Over an extended period of time, WearComp begins to function as a true extension of the mind and body and no longer feels as if it is a separate entity. In fact, the user will often adapt to the apparatus to such a degree that when taking it off, its absence will feel uncomfortable in the same way that we adapt to shoes and clothing to such a degree that without them most of us would feel extremely uncomfortable (whether in a public setting, or in an environment in which we have come to be accustomed to the protection that shoes and clothing provide). This intimate and constant bonding is such that the combined capability resulting in a synergistic whole far exceeds the sum of its components.
7) *Quality of Life*: WearComp is capable of enhancing day-to-day experiences, not just in the workplace, but in all facets of daily life. It has the capability to enhance the overall quality of life for many people.

IV. PRACTICAL EMBODIMENTS OF WEARCOMP

The WearComp apparatus consists of a battery-powered wearable Internet-connected [15] computer system with miniature eyeglass-mounted screen and appropriate optics to form the virtual image equivalent to an ordinary desktop multimedia computer. However, because the apparatus is tetherless it travels with the user, presenting a computer screen that either appears superimposed on top of the real world or represents the real world as a video image [16].

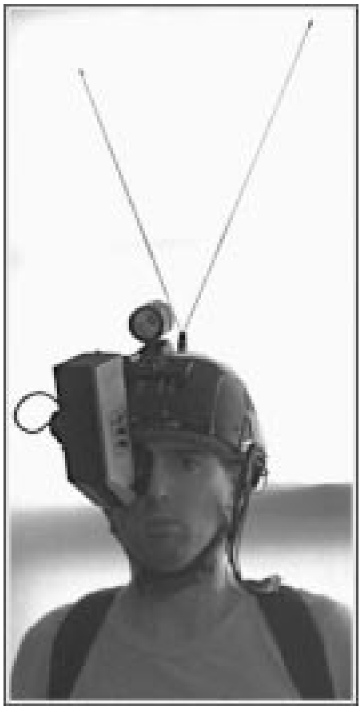
Due to advances in low-power microelectronics [17] we are entering a pivotal era in which it will become possible for us to be inextricably intertwined with computational technology that will become part of our everyday lives in a much more immediate and intimate way than in the past.

Physical proximity and constancy were simultaneously realized by the WearComp Project[2] of the 1970's and early 1980's (Fig. 3), which was a first attempt at building an intelligent "photographer's assistant" around the body. It was comprised of a computer system attached to the body, a display means constantly visible to one or both eyes, and a means of signal input, including a series of pushbutton switches and a pointing device (Fig. 4) that the wearer could hold in one hand to function like a keyboard and mouse but would still be able to operate the device while walking around. In this way, the apparatus resituated the functionality of a desktop multimedia computer with mouse, keyboard, and video screen as a physical extension of the user's body. While the size and weight reductions of WearComp over the last 20 years have been quite dramatic, the basic qualitative elements and functionality have remained essentially the same, apart from the obvious increase in computational power.

However, what makes WearComp particularly useful in new and interesting ways, and what makes it particularly suitable as a basis for humanistic computing, is the collection of other input devices, not all of which are found on a desktop multimedia computer.

In typical embodiments of WearComp these measurement (input) devices include the following:

1) ultraminiature cameras concealed inside eyeglasses and oriented to have the same field of view as the wearer, thus providing the computer with the wearer's "first-person" perspective;
2) one or more additional cameras that afford alternate points of view (e.g., a rear-looking camera with a view of what is directly behind the wearer);
3) sets of microphones, typically comprising one set to capture the sounds of someone talking to the wearer (typically a linear array across the top of the wearer's eyeglasses), and a second set to capture the wearer's own speech;
4) biosensors, comprising not just heart rate but full ECG waveform, as well as respiration, skin conductivity, sweat level, and other quantities [20], each available as a continuous (sufficiently sampled) time-varying voltage; typically these are connected to the wearable central processing unit through an eight-channel analog to digital converter;
5) footstep sensors typically comprising an array of transducers inside each shoe;



(a)

(b)

Fig. 3. Early embodiments of the author's original "photographer's assistant" application of personal imaging. (a) Author wearing WearComp2, an early 1980's backpack-based signal processing and personal imaging system with right eye display. Two antennas operating at different frequencies facilitated wireless communications over a full-duplex radio link. (b) WearComp4, a late 1980's clothing-based signal processing and personal imaging system with left eye display and beam splitter. Separate antennas facilitated simultaneous voice, video, and data communication.

---

[2] For a detailed historical account of the WearComp Project, and other related projects, see [18] and [19].

(a)

(b)

Fig. 4. Some early input devices ("keyboards" and "mice") designed and built by the author for WearComp: (a) 1970's: input device comprising pushbutton switches mounted to a wooden hand-grip and (b) 1980's: input device comprising microswitches mounted to the handle of an electronic flash. These devices also incorporated a detachable joystick (controlling two potentiometers), designed as a pointing device for use in conjunction with the WearComp project.

6) wearable radar systems in the form of antenna arrays sewn into clothing; these typically operate in the 24.36 GHz range.

The last three, in particular, are not found on standard desktop computers, and even the first three, which often are found on standard desktop computers, appear in a different context here than they do on a desktop computer. For example, in WearComp, the camera does not show an image of the user, as it does typically on a desktop computer, but, rather, it provides information about the user's environment. Furthermore, the general philosophy, as will be described in Sections V and VI, will be to regard all of the input devices as measurement devices. Even something as simple as a camera will be regarded as a measuring instrument within this signal processing framework.

Certain applications use only a subset of these devices, but including all of them in the design facilitates rapid prototyping and experimentation with new applications. Most embodiments of WearComp are modular, so that devices can be removed when they are not being used.

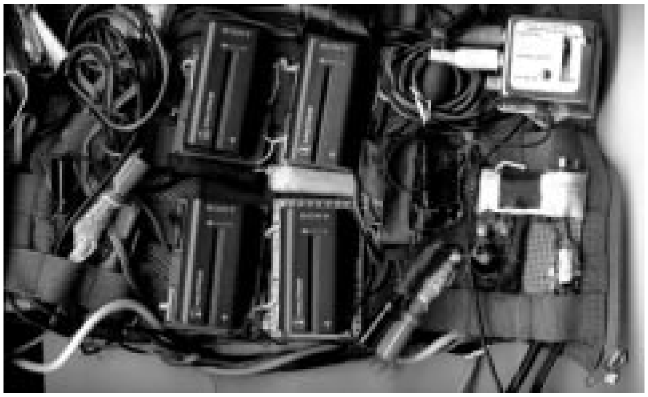
A side-effect of this WearComp apparatus is that it replaces much of the personal electronics that we carry in our day-to-day living. It enables us to interact with others through its wireless data communications link, and therefore replaces the pager and cellular telephone. It allows us to perform basic computations, thus replacing the pocket calculator, laptop computer, and PDA. It can record data from its many inputs, and therefore it replaces and subsumes the portable dictating machine, camcorder, and the photographic camera. And it can reproduce ("play back") audiovisual data, so that it subsumes the portable audio cassette player. It keeps time, as any computer does, and this may be displayed when desired, rendering a wristwatch obsolete. (A calendar program which produces audible, vibrotactile, or other output also renders the alarm clock obsolete.)

However, it goes beyond replacing all of these items, because not only is it currently far smaller and far less obtrusive than the sum of what it replaces, but these functions are interwoven seamlessly, so that they work together in a mutually assistive fashion. Furthermore, entirely new functionalities and new forms of interaction arise, such as enhanced sensory capabilities, as will be discussed in Sections V and VI.

A. Building Signal-Processing Devices Directly into Fabric

The wearable signal processing apparatus of the 1970's and early 1980's was cumbersome at best, so an effort was directed toward not only reducing its size and weight, but, more importantly, reducing its undesirable and somewhat obtrusive appearance. Moreover, an effort was also directed at making an apparatus of a given size and weight more comfortable to wear and bearable to the user [12] through bringing components in closer proximity to the body, thereby reducing torques and moments of inertia. Starting in 1982, Eleveld and Mann [19] began an effort to build circuitry directly into clothing. The author coined the term "smart clothing" to refer to variations of WearComp that are built directly into clothing, and are characterized by (or at least an attempt at) making components distributed rather than lumped whenever possible or practical.

(a)



(b)

Fig. 5. The "underwearable" signal processing hardware (a) as worn by author (stripped to the undershirt, which is normally covered by a sweater or jacket); (b) close-up of underwearable signal processor, showing webbing for routing of cabling.

It was found [19] that the same apparatus could be made much more comfortable by bringing the components closer to the body which had the effect of reducing both the torque felt bearing the load, as well as the moment of inertia felt in moving around. This effort resulted in a version of WearComp called the "Underwearable Computer" [19] shown in Fig. 5.

Typical embodiments of the underwearable resemble an athletic undershirt (tank top) made of durable mesh fabric, upon which a lattice of webbing is sewn. This facilitates quick reconfiguration in the layout of components, and rerouting of cabling. Note that wire ties are not needed to fix cabling as it is simply run through the webbing, which holds it in place. All power and signal connections are standardized, so that devices may be installed or removed without the use of any tools (such as soldering iron) by simply removing the garment and spreading it out on a flat surface.

Some more recent related work by others [21] also involves building circuits into clothing, in which a garment is constructed as a monitoring device to determine the location of a bullet entry. The underwearable differs from this monitoring apparatus in the sense that the underwearable is totally reconfigurable in the field, and also in the sense that it embodies humanistic computing (the apparatus reported in [21] performs a monitoring function but does not facilitate human interaction).

In summary, there were three reasons for the signal processing hardware being "underwearable."

1) By both distributing the components throughout the garment and by keeping the components in close physical proximity to the body, it was found that the same total weight and bulk could be worn much more comfortably.
2) Wearing the apparatus underneath ordinary clothing gave rise to a version of the WearComp apparatus which had a normal appearance. Although many of the prototypes were undetectable (covert) to visual inspection, early prototypes could be detected upon physical contact with the body by others. However, it was found that by virtue of social norms the touching of the body by others (and therefore discovery of the apparatus) was seldom a problem. Thus making certain that the apparatus did not have an unsightly or unusual appearance was found to be sufficient to integrating into society in a normal way. Unobtrusiveness is essential so that the apparatus does not interfere with normal social situations, for one cannot truly benefit from the long-term adaptation process of humanistic computing unless it is worn nearly constantly for a period of many years. Two examples of underwearables, as they normally appear when worn under clothing, are depicted in Fig. 6, where the normal appearance is quite evident.
3) The close proximity of certain components to the body provided additional benefits, such as the ability to easily integrate measuring apparatus for quantities such as respiration, heart rate and full ECG waveform, galvanic skin resistance, etc., of the wearer. The fact that the apparatus is worn underneath clothing facilitated direct contact with the body, providing a much richer measurement space and facilitating new forms of intelligent signal processing.

1) Remaining Issues of Underwearable Signal Processing Hardware: This work on "underwearable signal processing hardware" has taken an important first step toward solving issues of comfort, detectability, and detectability on physical contact, which remains as a greater technological challenge. Issues such as shielding of electromagnetic

(a) (b)

Fig. 6. Covert embodiments of WearComp suitable for use inordinary day-to-day situations. Both incorporate fully functional UNIX-based computers concealed in the small of the back, with the rest of the peripherals, analog to digital converters, etc., also concealed under ordinary clothing. Both incorporate camera-based imaging systems concealed within the eyeglasses, the importance of which will become evident in Section V in the context of personal imaging. While these prototype units are detectable by physical contact with the body, detection of the unusual apparatus was not found to be a problem, since normal social conventions are such that touching of the body is normally only the domain of those the wearer knows well. As with any prosthetic device, first impressions are important to normal integration into society, and discovery by those who already know the wearer well (e.g., to the extent that close physical contact may occur) typically happens after an initial acceptance is already established. Other prototypes have been integrated into the clothing in a manner that feels natural to the wearer and to others who might come into physical contact with the wearer. (a) Lightweight black and white version completed in 1995. (b) Full-color version completed in 1996, which included special-purpose digital signal processing hardware based on an array of TMS 320 series processors connected to a UNIX-based host processor, concealed in the small of the back.

radiation have already been largely solved through the use of conductive undergarments that protect the body from radiation (especially from the transmitting antennas that establish the connection to the Internet). In addition to protecting the wearer who would otherwise be in very close proximity to the transmitting antennas, such shielding also improves system operation, for example, by keeping RF from the transmitter as well as ambient environmental RF out of the biosensors and measurement instrumentation that is in direct contact with the body. Many of these practical issues will be dealt with in future design in moving from early prototypes of the "underwearable signal processing system" into production systems.

B. Multidimensional Signal Input for Humanistic Computing

The close physical proximity of WearComp to the body, as described earlier, facilitates a new form of signal processing.[3] Because the apparatus is in direct contact with the body, it may be equipped with various sensory devices. For example, a tension transducer (pictured leftmost, running the height of the picture from top to bottom, in Fig. 7) is typically threaded through and around the underwearable, at stomach height, so that it measures respiration. Electrodes are also installed in such a manner that they are in contact with the wearer's heart. Various other sensors, such as an array of transducers in each shoe [23] and a wearable radar system (described in Section VI) are also included as sensory inputs to the processor. The ProComp 8 channel analog-to-digital converter, along with some of the input devices that are sold with it, is pictured in Fig. 7 together with the CPU from WearComp6.

1) Safety First!: The importance of personal safety in the context of this new paradigm in computing must be emphasized. Since electrical connections are often in close proximity to the body, and are, in fact, often made directly to the body (such as when fitting the undergarment with electrodes that connect to the bare flesh of the user, in the chest area, for ECG waveform monitoring), the inputs to which these connections are made must be fully isolated. In the present prototypes, this isolation is typically accomplished by using a length of fiber-optic cable between the measurement hardware and the host computer. This is particularly essential in view of the fact that the host

---

[3] The first wearable computers equipped with multichannel biosensors were built by the author during the 1980's and were inspired by a collaboration with Ghista of McMaster University. More recently, in 1995, the author put together an improved apparatus based on a Compaq Contura Aero 486/33 with a ProComp 8 channel analog-to-digital converter, worn in a Mountainsmith waist bag, and sensors from Thought Technologies Limited. The author subsequently assisted Healey in duplicating this system for use in trying to understand human emotions [22].

Fig. 7. Author's personal imaging system equipped with sensors for measuring biological signals. The sunglasses in the upper right are equipped with built-in video cameras and display system. These look like ordinary sunglasses when worn (wires are concealed inside the eyeglass holder). At the left side of the picture is an eight-channel analog-to-digital converter together with a collection of biological sensors, both manufactured by Thought Technologies Limited, Canada. At the lower right is an input device called the "twiddler," manufactured by HandyKey, and to the left of that is a Sony Lithium Ion camcorder battery with custom-made battery holder. In the lower central area of the image is the computer, equipped with special-purpose video processing/video capture hardware (visible as the top stack on this stack of PC104 boards). This computer, although somewhat bulky, may be concealed in the small of the back underneath an ordinary sweater. To the left of the computer is a serial-to-fiber-optic converter that provides communications to the eight-channel analog-to-digital converter over a fiber-optic link. Its purpose is primarily one of safety, to isolate high voltages used in the computer and peripherals (e.g., the 500 volts or so present in the sunglasses) from the biological sensors which are in close proximity, typically with very good connection, to the body of the wearer.

computer may be connected to other devices, such as head-mounted displays containing voltages from 500 V to as high as 9 kV (e.g., in the case of head-mounted displays containing cathode ray tubes).

Moreover, the presence of high voltages in the eyeglasses themselves, as well as in other parts of the system (as when it is interfaced to an electronic flash system which typically uses voltages ranging from 480 V to 30 kV), requires extra attention to insulation since it is harder to free oneself from the apparatus when it is worn than when it is merely carried. For example, a hand-held electronic flash can be dropped to the ground easier should it become unstable, while a wearable system embodies the danger of entrapment in a failing system.

Furthermore, since batteries may deliver high currents, there would be the risk of fire, and the prospect of being trapped in burning clothing, were it not for precautions taken in limiting current flow. Thus, in addition to improved high voltage insulation there is also the need to install current-limiting fuses and the like throughout the garment.

As an additional precaution, all garments are made from flameproof material. In this regard, especially with the development of early prototypes over the last 20 years, it was felt that a healthy sense of paranoia was preferable to carelessness that might give rise to a dangerous situation.

2) More Than Just a Health Status Monitor: It is important to realize that this apparatus is not merely a biological signal logging device, as is often used in the medical community, rather, it enables new forms of real-time signal processing for humanistic computing. A simple example might include a biofeedback-driven video camera. Picard also suggests its possible use to estimate human emotion [24].

The emphasis of this paper will be on visual image processing with the WearComp apparatus. The author's dream of the 1970's, i.e., that of an intelligent wearable image processing apparatus, is just beginning to come to fruition.

V. THE PERSONAL IMAGING APPLICATION OF HUMANISTIC COMPUTING

A. Some Simple Illustrative Examples

1) Always Ready: From "Point and Click" to "Look and Think": Current commercial personal electronics devices we often carry are just useful enough for us to tolerate but not good enough to significantly simplify our lives. For example, when we are on vacation, our camcorder and photographic camera require enough attention that we often either miss the pictures we want, or we become so involved in the process of video or photography that we fail to really experience the immediate present environment [25].

One ultimate goal of the proposed apparatus and methodology is to "learn" what is visually important to the wearer and function as a fully automatic camera that takes pictures without the need for conscious thought or effort from the wearer. In this way it might summarize a day's activities and then automatically generate a gallery exhibition by transmitting desired images to the World Wide Web, or to specific friends and relatives who might be interested in the highlights of one's travel. The proposed apparatus, a miniature eyeglass-based imaging system, does not encumber the wearer with equipment to carry or with the need to remember to use it, yet because it is recording all the time into a circular buffer [19] merely overwriting that which is unimportant, it is always ready. Although some have noted that the current embodiment of the invention (still in the prototype stage) is cumbersome enough that one might not wear it constantly, it is easy to imagine how, with mass production and miniaturization, smaller and lighter units could be built, perhaps with the computational hardware built directly into ordinary glasses. Making the apparatus small enough to comfortably wear at all times will lead to a truly constant user-interface.

In the context of the always ready framework, when the signal processing hardware detects something that might be of interest, recording can begin in a retroactive sense (e.g., a command may be issued to start recording from 30 seconds ago), and the decision can later be confirmed with human input. Of course this apparatus raises some important privacy questions discussed previously and also addressed elsewhere in the literature [26], [27].

The system might use the inputs from the biosensors on the body as a multidimensional feature vector with which to classify content as important or unimportant. For example, it might automatically record a baby's first steps as the parent's eyeglasses and clothing-based intelligent signal processor make an inference based on the thrill of the experience. It is often moments like these that we fail to capture on film: by the time we find the camera and load it with film, the moment has passed us by.

2) Personal Safety Device for Reducing Crime: A simple example of where it would be desirable for the device to operate by itself, without conscious thought or effort, is in an extreme situation such as if the wearer were attacked by a robber wielding a shotgun and demanding cash.

In this kind of situation it is desirable that the apparatus functions autonomously, without conscious effort from the wearer, even though the wearer might be aware of the signal processing activities of the measuring (sensory) apparatus he or she is wearing.

As a simplified example of how the processing is typically done, we know that the wearer's heart rate, averaged over a sufficient time window, would likely increase dramatically[4] with no corresponding increase in footstep rate (in fact, footsteps would probably slow at the request of the gunman). The computer would then make an inference from the data, and predict a high visual saliency. (If we simply take heart rate divided by footstep rate, we can get a first-order approximation of the visual saliency index.) A high visual saliency triggers recording from the wearer's camera at maximal frame rate and also sends these images together with appropriate messages to friends and relatives who look at the images to determine whether it is a false alarm or real danger.[5]

Such a system is, in effect, using the wearer's brain as part of its processing pipeline, because it is the wearer who sees the shotgun, and not the WearComp apparatus (e.g., a much harder problem would have been to build an intelligent machine vision system to process the video from the camera and determine that a crime was being committed). Thus, humanistic computing (intelligent signal processing arising, in part, because of the very existence of the human user) has solved a problem that would not be possible using machine-only intelligence.

Furthermore, this example introduces the concept of "collective connected humanistic computing" because the signal processing systems also rely on those friends and relatives to look at the imagery that is wirelessly sent from the eyeglass-mounted video camera and makes a decision as to whether it is a false alarm or real attack. Thus, the concept of humanistic computing has become blurred across geographical boundaries and between more than one human and more than one computer.

3) The Retro-Autofocus Example—Human in the Signal Processing Loop: The above two examples dealt with systems which use the human brain, with its unique processing capability, as one of their components in a manner in which the overall system operates without conscious thought or effort. The effect is to provide a feedback loop of which subconscious or involuntary processes become an integral part.

An important aspect of humanistic computing is that the conscious will of the user may be inserted into or removed from the feedback loop of the entire process at any time. A very simple example, taken from everyday experience, rather than another new invention, is now presented.

One of the simplest examples of humanistic computing is that which happens with some of the early autofocus single lens reflex (SLR) cameras in which autofocus was a retrofit feature. The autofocus motor would typically turn the lens barrel, but the operator could also grab onto the lens barrel while the autofocus mechanism was making it turn. Typically the operator could "fight" with the motor and easily overpower it, since the motor was of sufficiently low torque. This kind of interaction is particularly useful, for example, when shooting through a glass window at a distant object where there are two or three local minima of the autofocus error function (e.g., focus on particles of dust on the glass itself, focus on a reflection in the glass, and focus on the distant object). Thus, when the operator wishes to focus on the distant object and the camera system is caught in one of the other local minima (for example, focused on the glass), the user merely grasps the lens barrel, swings it around to the approximate desired location (as though focusing crudely by hand on the desired object of interest), and lets go so that the camera will then take over and bring the desired object into sharp focus. This very simple example illustrates a sort of humanistic intelligent signal processing in which the intelligent autofocus electronics of the camera work in close synergy with the intellectual capabilities of the camera operator. It is this aspect of humanistic computing that allows the human to step into and out of the loop at any time and makes it a powerful paradigm for intelligent signal processing.

B. Mathematical Framework for Personal Imaging

The theoretical framework for humanistic computing is based on processing a series of inputs from various wearable sensory apparatus in a manner that regards each one of these as belonging to a measurement space; each of the inputs (except for the computer's keyboard-like input device comprised of binary pushbutton switches) is regarded as a measurement instrument to be linearized in some meaningful continuous underlying physical quantity.

Since the emphasis of this paper is on personal imaging, the treatment here will focus on the wearable camera (dis-

---

[4] Perhaps it may stop, or "skip a beat" at first, but over time, on average, in the time following the event, experience tells us that our hearts beat faster when frightened.

[5] It has been suggested that the robber might become aware that his or her victim is wearing a personal safety device and try to eliminate it or perhaps even target it for theft. In anticipation of these possible problems, personal safety devices operate by continuous transmission of images so that the assailant cannot erase or destroy the images depicting the crime. Moreover, the device itself, owing to its customized nature, would be unattractive and of little value to others, much as are undergarments, a mouthguard, or prescription eyeglasses. Furthermore, devices may be protected by a password embedded into a programmable logic device that functions as a finite state machine, making them inoperable by anyone but the owner. To protect against passwords being extracted through torture, a personal distress password may be provided to the assailant by the wearer. The personal distress password unlocks the system but puts it into a special tracking and distress notification mode.

cussed here in Section V) and the wearable radar (discussed later in Section VI). The other measurement instruments are important, but their role is primarily to facilitate harnessing and amplification of the human intellect for purposes of processing data from the imaging apparatus.

The theoretical framework for processing video is based on regarding the camera as an array of light-measuring instruments capable of measuring how the scene or objects in view of the camera respond to light.[6] This framework has two important special cases, the first of which is based on a quantifiable self-calibration procedure and the second of which is based on algebraic projective geometry as a means of combining information from images related to one another by a projective coordinate transformation.

These two special cases of the theory are now presented in Sections V-B1 and V-B2 respectively, and both are brought together in Section V-B3. The theory is applicable to standard photographic or video cameras, as well as to the wearable camera and personal imaging system.

1) Quantigraphic Imaging and the Wyckoff Principle: It should be noted that the goal of quantigraphic imaging, to regard the camera as an array of light-measuring instruments, is quite different from the goals of other related research [29] in which there is an attempt to separate the reflectance of objects from their illumination. Indeed, while Stockham's effort was focused on separately processing the effects due to reflectance and scene illumination [29], quantigraphic imaging takes a camera-centric viewpoint, and it does not attempt to model the cause of the light entering the camera but merely determines, to within a single unknown scalar constant, the quantity of light entering the camera from each direction in space. Quantigraphic imaging measures neither radiometric irradiance nor photometric illuminance (since the camera will not necessarily have the same spectral response as the human eye, or, in particular, that of the photopic spectral luminous efficiency function as determined by the Commission Internationale de l'Éclairage (CIE) and standardized in 1924). Instead, quantigraphic imaging measures the quantity of light integrated over the particular spectral response of the camera system in units that are quantifiable (e.g., linearized) in much the same way that a photographic light meter measures in quantifiable (linear or logarithmic) units. However, just as the photographic light meter imparts to the measurement its own spectral response (e.g., a light meter using a selenium cell will impart the spectral response of selenium cells to the measurement) quantigraphic imaging accepts that there will be a particular spectral response of the camera which will define the quantigraphic unit of measure.

A field of research closely related to Stockham's approach is that of colorimetry, in the context of the so-called color constancy problem [30], [31] in which there is an attempt made to determine the true color of an object irrespective of the color of the illuminant. Thus, solving the color constancy problem, for example, might amount to being able to recognize the color of an object and ignore the orange color cast of indoor tungsten lamps, or ignore the relatively bluish cast arising from viewing the same object under the illumination of the sky outdoors. Quantigraphic imaging, on the other hand, makes no such attempt, and it provides a true measure of the light arriving at the camera without any attempt to determine whether color effects are owing to the natural color of an object or the color of the illumination.

Quantigraphic imaging, however, may be an important first step to solving the color constancy problem—once we know how much light is arriving from each direction in space in each of the spectral bands of the sensor, and we have a measure of these quantities that is linearized, we can then apply to these quantigraphic images any of the traditional mathematical image processing frameworks such as those of Stockham [29], Venetsanopoulos [32], or those from color theory. Quantigraphic imaging may also be a first step to other uses of the image data, whether they be for machine vision or simply for the production of a visually pleasing picture.

The special case of quantigraphic imaging presented here in Section V-B1 pertains to a fixed camera (e.g., as one would encounter in mounting the camera on a tripod). Clearly this is not directly applicable to the wearable camera system, except perhaps in the case of images acquired in very rapid succession. However, this theory, when combined with the Video Orbits theory of Section V.B-2, is found to be useful in the context of the personal imaging system, as will be described in Section V-B3.

Fully automatic methods of seamlessly combining differently exposed pictures to extend dynamic range have been proposed [33], [34] and are summarized here.

Most everyday scenes have a far greater dynamic range than can be recorded on a photographic film or electronic imaging apparatus (whether it be a digital still camera, consumer video camera, or eyeglass-based personal imaging apparatus as described in this paper). However, a set of pictures that are identical except for their exposure, collectively show us much more dynamic range than any single picture from that set and also allow the camera's response function to be estimated within a single constant scalar unknown.

A set of functions

$$E_n(\mathbf{x}) = f(k_n q(\mathbf{x})) \tag{1}$$

where k_n are scalar constants, is known as a Wyckoff set [35], [19], and describes a set of images E_n when $\mathbf{x} = (x, y)$ is the spatial coordinate of a piece of film or the continuous spatial coordinates of the focal plane of an electronic imaging array, q is the quantity of light falling on the sensor array, and f is the unknown nonlinearity of the camera's response function (assumed to be invariant to $\mathbf{x} \in \mathbb{R}^2$).

Because of the effects of noise (quantization noise, sensor noise, etc.), in practical imaging situations, the dark (often underexposed) pictures show us highlight details of the scene that might have been overcome by noise (e.g., washed out) had the picture been properly exposed. Similarly, the

---

[6] This "lightspace" theory was first written about in detail in 1992 [28].

light pictures show us some shadow detail that might not have appeared above the noise threshold had the picture been properly exposed.

A means of simultaneously estimating f and k_n, given a Wyckoff set E_n, has been proposed [33], [35], [19]. A brief outline of this method follows. For simplicity of illustration (without loss of generality), suppose that the Wyckoff set contains two pictures, $E_1 = f(q)$ and $E_2 = f(kq)$, differing only in exposure (e.g., where the second image received k times as much light as the first). Photographic film is traditionally characterized by the so-called "D logE" (density versus log exposure) characteristic curve [36], [37]. Similarly, in the case of electronic imaging, we may also use logarithmic exposure units $Q = \log(q)$ so that one image will be $K = \log(k)$ units darker than the other

$$\log\left(f^{-1}(E_1)\right) = Q = \log\left(f^{-1}(E_2)\right) - K. \tag{2}$$

The existence of an inverse for f follows from the semimonotonicity assumption [35], [19]. (We expect any reasonable camera to provide a semimonotonic relation between quantity of light received, q, and the pixel value reported.) Since the logarithm function is also monotonic, the problem is reduced to that of estimating the semimonotonic function $F() = \log(f^{-1}())$ and the scalar constant K given two pictures E_1 and E_2

$$F(E_2) = F(E_1) + K. \tag{3}$$

Thus

$$E_2 = F^{-1}(F(E_1) + K) \tag{4}$$

provides a recipe for "registering" (appropriately lightening or darkening) the second image to match the first. This registration procedure differs from the image registration procedure commonly used in image resolution enhancement (to be described in Section V-B2) because it operates on the range (tonal range) of the image $E(\mathbf{x})$ as opposed to its domain (spatial coordinates) $\mathbf{x} = (x, y)$. (In Section V-B3, registration in both domain and range will be addressed.)

Once f is determined, each picture becomes a different estimate of the same true quantity of light falling on each pixel of the image sensor

$$\hat{q}_n = \frac{1}{k_n} f^{-1}(E_n). \tag{5}$$

Thus one may regard each of these measurements (pixels) as a light meter (sensor element) that has some nonlinearity followed by a quantization to a measurement having typically 8-bit precision.

It should be emphasized that most image processing algorithms incorrectly assume that the camera response function is linear (e.g., almost all current image processing, such as blurring, sharpening, unsharp masking, etc., operates linearly on the image) while in fact it is seldom linear. Even Stockham's homomorphic filtering [29], which advocates taking the log, applying linear filtering, and then taking the antilog, fails to capture the correct nonlinearity [19], [35] as it ignores the true nonlinearity of the sensor array. It has recently been shown [35], [19] that in the absence of any knowledge of the camera's nonlinearity an approximate one-parameter parametric model of the camera's nonlinear response is far better than assuming it is linear or logarithmic. Of course, finding the true response function of the camera allows one to do even better, as one may then apply linear signal processing methodology to the original light falling on the image sensor.

2) Video Orbits: A useful assumption in the domain of "personal imaging" is that of zero parallax, whether this be for obtaining a first-order estimate of the yaw, pitch, and roll of the wearer's head [20], or making an important first step in the more difficult problem of estimating depth and structure from a scene.[7] Thus, in this section, the assumption is that most of the image motion arises from that of generating an environment map, zero-parallax is assumed.

The problem of assembling multiple pictures of the same scene into a single image commonly arises in mapmaking (with the use of aerial photography) and photogrammetry [44], where zero-parallax is also generally assumed. Many of these methods require human interaction (e.g., selection of features), and it is desired to have a fully automated system that can assemble images from the eyeglass-based camera. Fully automatic featureless methods of combining multiple pictures have been previously proposed [45], [46], but with an emphasis on subpixel image shifts; the underlying assumptions and models (affine and pure translation, respectively) were not capable of accurately describing more macroscopic image motion. A characteristic of video captured from a head-mounted camera is that it tends to have a great deal more macroscopic image motion and a great deal more perspective "cross-chirping" between adjacent frames of video, while the assumptions of static scene content and minimal parallax are still somewhat valid. This assumption arises for the following reasons.

1) Unlike the heavy hand-held cameras of the past, the personal imaging apparatus is very lightweight.
2) Unlike the hand-held camera which extends outward from the body, the personal imaging apparatus is mounted close to the face. This results in a much lower moment of inertia so that the head can be rotated quickly. Although the center of projection of the wearable camera is not located at the center of rotation of the neck, it is much closer than with a hand-held camera.

It was found that the typical video generated from the personal imaging apparatus was characterized by rapid sweeps or pans (rapid turning of the head), which tended to happen over much shorter time intervals and therefore dominated over second-order effects such as parallax and scene motion [19]. The proposed method also provides an indication of its own failure, and this can be used as a feature rather than a "bug" (e.g., so that the WearComp system is aware of scene motion, scene changes, etc., by

---

[7] First modeling the motion as a projective coordinate transformation, and then estimating the residual epipolar structure or the like [38]–[43].

virtue of its ability to note when the algorithm fails). Thus the projective group of coordinate transformations captures the essence of video from the WearComp apparatus.[8]

Accordingly, two featureless methods of estimating the parameters of a projective group of coordinate transformations were first proposed in [33] and in more detail in [35], one direct and one based on optimization (minimization of an objective function). Both of these methods are multiscale (e.g., use a coarse to fine pyramid scheme), and both repeat the parameter estimation at each level (to compute the residual errors). Although one might be tempted to call both iterative, it is preferable to refer to the direct method as repetitive to emphasize that it does not require a nonlinear optimization procedure such as Levenberg–Marquardt [48], [49] or the like. Instead, it uses repetition with the correct law of composition on the projective group, going from one pyramid level to the next by application of the group's law of composition. A method similar to the author's optimization method was later proposed in [35] and [50]. The author's direct method has also been subsequently described in more detail [51].

The author's direct featureless method for estimating the eight scalar parameters[9] of an exact projective (homographic) coordinate transformation is now described. In the context of personal imaging, this result is used with multiple images to seamlessly combine images of the same scene or object, resulting in a single image (or new image sequence) of greater resolution or spatial extent.

Many papers have been published on the problems of motion estimation and frame alignment. (For review and comparison, see [52].) In this section, the emphasis is on the importance of using the "exact" eight-parameter projective coordinate transformation [51], particularly in the context of the head-worn miniature camera.

The most common assumption (especially in motion estimation for coding, and optical flow for computer vision) is that the coordinate transformation between frames is translation. Tekalp *et al.* [46] have applied this assumption to high-resolution image reconstruction. Although translation is less simple to implement than other coordinate transformations, it is poor at handling large changes due to camera zoom, rotation, pan, and tilt. Zheng and Chellappa [53] considered the image registration problem using a subset of the affine model—translation, rotation, and scale. Other researchers [45], [54] have assumed affine motion (six parameters) between frames.

The only model that properly captures the "keystoning" and "chirping" effects of projective geometry is the projective coordinate transformation. However, because the parameters of the projective coordinate transformation had traditionally been thought to be mathematically and computationally too difficult to solve, most researchers have used the simpler affine model or other approximations to the projective model. The eight-parameter pseudoperspective model [39] does, in fact, capture both the converging lines and the chirping of a projective coordinate transformation, but not the true essence of projective geometry.

Of course, the desired "exact" eight parameters come from the projective group of coordinate transformations, but they have been perceived as being notoriously difficult to estimate. The parameters for this model have been solved by Tsai and Huang [55], but their solution assumed that features had been identified in the two frames, along with their correspondences. The main contribution of the result summarized in this section is a simple featureless means of automatically solving for these eight parameters.

A group is a set upon which there is defined an associative law of composition (closure, associativity) which contains at least one element (identity) who's composition with another element leaves it unchanged, and for which every element of the set has an inverse. A group of operators together with a set of operands form a so-called group operation.[10] In the context of a Lie group of spatial coordinate transformation operators acting on a set of visual images as operands, such a group is also known as a Lie group of transformations [57].

Note that Hoffman's use of the term "transformation" is not synonymous with "homomorphism" [56] as is often the case in group theory, such as when a transformation T acts on a law of composition between elements $g, h \in G$, such that $T(gh) = T(g)T(h)$. Instead, what is meant by "transformation," in the context of this paper, is a change in coordinates of a picture (image). Thus, in the context of this paper, transformations act on images, not on elements of the group (which happens to be a group of transformation operators).

As in [57], coordinate transformations are operators selected from a group and the set of images are the operands. This group of operators and set of images form the group action in the sense defined in [56]. When the coordinate transformations form a group, then two such coordinate transformations, $\mathbf{p_1}$ and $\mathbf{p_2}$, acting in succession on an image (e.g., $\mathbf{p_1}$ by doing a coordinate transformation, followed by a further coordinate transformation corresponding to $\mathbf{p_2}$, acting on that result) can be replaced by a single coordinate transformation. That single coordinate transformation is given by the law of composition in the group. The orbit of a particular element of the set, under the group operation [56], is the new set formed by applying to it all possible operators from the group.

Thus, the orbit is a collection of pictures formed from one picture through applying all possible projective coordinate transformations to that picture. This set is referred to as the "video orbit" of the picture in question [51]. Equivalently, we may imagine a static scene in which the wearer of the personal imaging system is standing at a single fixed location. He or she generates a family of images in the same

---

[8] The additional one-time download of a lens distortion map into WearComp's coordinate transformation hardware eliminates its lens distortion, which would otherwise be very large owing to the covert (and therefore small) size of the lens, and in engineering compromises necessary to its design. The Campbell method [47] is used to estimate the lens distortion for this one-time coordinate transformation map.

[9] Published in detail in [51].

[10] Also known as a group action or G-set [56].

orbit of the projective group of transformations by looking around (rotation of the head).[11]

The coordinate transformations of interest, in the context of this paper

$$\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{\mathbf{A}[x,y]^T + \mathbf{b}}{\mathbf{c}^T[x,y]^T + d} = \frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^T\mathbf{x} + d} \tag{6}$$

define an operator $\mathbf{P_1}$ that acts on the images as follows:

$$P_{\mathbf{A_1},\mathbf{b_1},\mathbf{c_1},d_1} \circ E(\mathbf{x}) = E(\mathbf{x}') = E\left(\frac{\mathbf{A_1}\mathbf{x} + \mathbf{b_1}}{\mathbf{c_1^T}\mathbf{x} + d_1}\right) \tag{7}$$

where the operator $\mathbf{P_1}$ is parameterized by $\mathbf{A}_1 \in \mathbb{R}^{2\times 2}$, $\mathbf{b}_1 \in \mathbb{R}^{2\times 1}$, $\mathbf{c}_1 \in \mathbb{R}^{2\times 1}$, and $d_1 \in \mathbb{R}$.

These operators may be applied to an image in succession, this succession being defined, for example, with such operators, $\mathbf{P_1}$ and $\mathbf{P_2}$, as

$$\mathbf{P_2} \circ \mathbf{P_1} \circ E(\mathbf{x}) = \mathbf{P}_{\mathbf{A_2},\mathbf{b_2},\mathbf{c_2},\mathbf{d_2}} \circ \mathbf{P}_{\mathbf{A_1},\mathbf{b_1},\mathbf{c_1},\mathbf{d_1}} \circ E(\mathbf{x}) \tag{8}$$

$$\mathbf{P_2} \circ \mathbf{P_1} \circ E(\mathbf{x}) = E\left(\frac{A_2\left(\frac{\mathbf{A_1}\mathbf{x}+\mathbf{b_1}}{\mathbf{c_1^T}\mathbf{x}+d_1}\right) + b_2}{c_2^T\left(\frac{\mathbf{A_1}\mathbf{x}+\mathbf{b_1}}{\mathbf{c_1^T}\mathbf{x}+d_1}\right) + d_2}\right) \tag{9}$$

$$= E\left(\begin{bmatrix} \mathbf{A_2}, \mathbf{b_2} \\ \mathbf{c_2^T}, d_2 \end{bmatrix} \begin{bmatrix} \mathbf{A_1}, \mathbf{b_1} \\ \mathbf{c_1^T}, d_1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}\right) \tag{10}$$

from which we can see that the operators can be represented as 2×2 matrixes, and that the law of composition defined on these operators can be represented by matrix multiplication. Associativity follows from matrix multiplication. The matrix

$$\begin{bmatrix} \mathbf{I}_2, & \mathbf{0} \\ \mathbf{0}, & 1 \end{bmatrix} \tag{11}$$

represents the identity operation where I_2 is the identity matrix [1, 0; 0, 1].

The "video orbit" of a given two-dimensional (2-D) frame is defined to be the set of all images that can be produced by applying operators from the 2-D projective group of coordinate transformations (6) to the given image. Hence, the problem may be restated: given a set of images that lie in the same orbit of the group, find for each image pair that operator in the group which takes one image to the other image.

If two frames of the video image sequence, say, f_1 and f_2, are in the same orbit, then there is a group operation $\mathbf{p}$ such that the mean-squared error (MSE) between f_1 and $f_2' = \mathbf{p} \circ f_2$ is zero. In practice, however, the element of the group that takes one image "nearest" the other is found (e.g., there will be a certain amount of error due to violations in the assumptions due to noise such as parallax, interpolation error, edge effects, changes in lighting, depth of focus, etc.).

The brightness constancy constraint equation [59], which gives the flow velocity components, is

$$\mathbf{u_f^T}\mathbf{E_x} + E_t \approx 0. \tag{12}$$

As it is well known [59], the optical flow field in 2-D is underconstrained. The model of pure translation at every point has two parameters, but there is only one equation (12) to solve, thus it is common practice to compute the optical flow over some neighborhood, which must be at least two pixels but is generally taken over a small block, 3× 3, 5×5, or sometimes larger (e.g., the entire image, as in the video orbits algorithm described here).

However, rather than estimating the two-parameter translational flow, the task here is to estimate the eight parameter projective flow (6) by minimizing

$$\begin{aligned} \varepsilon_{\text{flow}} &= \sum \left(\mathbf{u_m^T}\mathbf{E_x} + E_t\right)^2 \\ &= \sum \left(\left(\frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^T\mathbf{x} + d} - \mathbf{x}\right)^T \mathbf{E_x} + E_t\right)^2. \end{aligned} \tag{13}$$

Although a sophisticated nonlinear optimization procedure, such as Levenberg–Marquardt, may be applied to solve (13), it has been found that solving a slightly different but much easier problem allows us to estimate the parameters more directly and accurately for a given amount of computation [51]

$$\begin{aligned} \varepsilon_w = \sum \Big(&\left(\mathbf{A}\mathbf{x} + \mathbf{b} - (\mathbf{c}^T\mathbf{x} + d)\mathbf{x}\right)^T \\ &\cdot \mathbf{E_x} + \left(\mathbf{c}^T\mathbf{x} + d\right)E_t \Big)^2. \end{aligned} \tag{14}$$

(This amounts to weighting the sum differently.)

Differentiating (13) with respect to the free parameters $\mathbf{A}, \mathbf{b}$, and $\mathbf{c}$, and setting the result to zero gives the following linear solution:

$$\begin{aligned} &\left(\sum \phi\phi^T\right)[a_{11}, a_{12}, b_1, a_{21}, a_{22}, b_2, c_1, c_2]^T \\ &\qquad = \sum (\mathbf{x}^T\mathbf{E_x} - E_t)\phi \end{aligned} \tag{15}$$

where $\phi^T = [E_x(x,y,1), E_y(x,y,1), xE_t - x^2E_x - xyE_y, yE_t - xyE_x - y^2E_y]$.

In practice, this process has been further improved by making an initial estimate using methods such as described in [60], [61], [62], as well as [63].

3) Dynamic Range and "Dynamic Domain": The contribution of this section is a simple method of "scanning" out a scene, from a fixed point in space, by panning, tilting, or rotating a camera, whose gain (automatic exposure, electronic level control, automatic iris, automatic gain control (AGC), or the like)[12] is also allowed to change of its own accord (i.e., arbitrarily).

Nyquist showed how a signal can be reconstructed from a sampling of finite resolution in the domain (e.g., space or time), but he assumed infinite dynamic range (e.g., infinite precision or word length per sample). On the other hand, if we have infinite spatial resolution but limited dynamic range (even if we have only 1 bit of image depth), Curtis and Oppenheim [64] showed that we can also obtain perfect reconstruction using an appropriate modulation function.

---

[11] The resulting collection of images may be characterized by fewer parameters through application of the Hartley constraint [58], [35] to an estimate of a projective coordinate transformations.

[12] For simplicity, all these methods of automatic exposure control are referred to as AGC in this paper, whether or not they are actually implemented using an AGC circuit or otherwise.

In the case of the personal imaging system, we typically begin with images that have very low spatial resolution and very poor dynamic range (video cameras tend to have poor dynamic range, and this poor performance is especially true of the small charge coupled devices (CCD's) that the author uses in constructing unobtrusive lightweight systems). Thus, since we lack both spatial and tonal resolution, we are not at liberty to trade some of one for more of the other. Thus the problem of "spatiotonal" (simultaneous spatial and tonal) resolution enhancement is of particular interest in personal imaging.

In Section V-B1, a new method of allowing a camera to self-calibrate was proposed. This methodology allowed the tonal range to be significantly improved. In Section V-B2, a new method of resolution enhancement was described. This method allowed the spatial range to be significantly enhanced. In this section, a method of enhancing both the tonal range and the spatial domain resolution of images is proposed. It is particularly applicable to processing video from miniature covert eyeglass-mounted cameras because it allows very noisy low-quality video signals to provide not only high-quality images of great spatiotonal definition, but also a rich and accurate photometric measurement space which may be of significant use to intelligent signal processing algorithms. That it provides not only high-quality images but also linearized measurements of the quantity of light arriving at the eyeglasses from each possible direction of gaze follows from a generalization of the photometric measurement process outlined in Section V-B1.

Most notably, this generalization of the method no longer assumes that the camera need be mounted on a tripod, but only that the images fall in the same orbit of a larger group, called the "projectivity + gain" group of transformations.

Thus, the apparatus can be easily used without conscious thought or effort, which gives rise to new intelligent signal processing capabilities. The method works as follows. As the wearer of the apparatus looks around, the portion of the field of view that controls the gain (usually the central region of the camera's field of view) will be pointed toward different objects in the scene. Suppose, for example, that the wearer is looking at someone so that their face is centered in the frame of the camera, f_1. Now suppose that the wearer tips his or her head upward so that the camera is pointed at a light bulb up on the ceiling, but that the person's face is still visible at the bottom of the frame, f_2. Because the light bulb has moved into the center of the frame, the camera's AGC causes the entire image to darken significantly. Thus these two images, which both contain the face of the person to whom the wearer is talking, will be very differently exposed. When registered in the spatial sense (e.g., through the appropriate projective coordinate transformation), they will be identical over the region of overlap, except for exposure, if we assume that the wearer swings his or her head around quickly enough to make any movement in the person he is talking to negligible. While this assumption is not always true, there are certain times that it is true (e.g., when the wearer swings his or her head quickly from left to right and objects in the scene are moving relatively slowly). Because the algorithm can tell when the assumptions are true (by virtue of the error), during the times it is true it uses the multiple estimates of $\hat{q}_n$, the quantity of light received, to construct a high definition environment map.

An example of an image sequence captured with a covert eyeglass-based version of the author's WearComp7 invention and transmitted wirelessly to the Internet appears in Fig. 8.

Clearly, in this application AGC, which has previously been regarded as a serious impediment to machine vision and intelligent image processing, becomes an advantage. By providing a collection of images with differently exposed but overlapping scene content, additional information about the scene, as well as the camera (information that can be used to determine the camera's response function, f) is obtained. The ability to have, and even benefit from AGC is especially important for WearCam contributing to the hands-free nature of the apparatus so that one need not make any adjustments when, for example, entering a dimly lit room from a brightly lit exterior.

The group of spatiotonal image transformations of interest is defined in terms of projective coordinate transformations, taken together with the one-parameter group of gain changes (image darkening/lightening) operations

$$\begin{aligned} p_{\mathbf{A},\mathbf{b},\mathbf{c},d,k} \circ f(q(x)) &= g_k\left(f\left(q\left(\frac{\mathbf{A}\mathbf{x}+\mathbf{b}}{\mathbf{c}^{\mathbf{T}}\mathbf{x}+d}\right)\right)\right) \\ &= f\left(kq\left(\frac{\mathbf{A}\mathbf{x}+\mathbf{b}}{\mathbf{c}^{\mathbf{T}}\mathbf{x}+d}\right)\right) \end{aligned} \tag{16}$$

where g_k characterizes the gain operation. These coordinate transformations admit a group representation

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} & \mathbf{0} \\ \mathbf{c} & d & 0 \\ \mathbf{0} & 0 & k \end{bmatrix} \tag{17}$$

giving the law of composition defined by matrix multiplication.

Two successive frames of a video sequence are related through a group-action that is near the identity of the group, thus the Lie algebra of the group provides the structure locally. As in previous work [34], an approximate model which matches the "exact" model in the neighborhood of the identity is used, together with the law of composition in the group (i.e., all of the manipulation and composition of multiple coordinate transformations is based on the algebra of the "exact" model, even though an approximate model is used in the innermost loop of the computation).

To construct a single floating-point image of increased spatial extent and increased dynamic range, first the images are spatiotonally registered (brought not just into register in the traditional "domain motion" sense, but also brought into the same tonal scale through quantigraphic gain adjustment). This form of spatiotonal transformation is illustrated in Fig. 9, where all the images are transformed into the coordinates of the first image of the sequence, and in Fig. 10, where all the images are transformed into the coordinates of the last frame in the image sequence. It

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

Fig. 8. The "fire-exit" sequence captured using a covert eyeglass-based personal imaging system with AGC. Here, every third frame of the ten-frame image sequence is shown. As the camera pans across to take in more of the open doorway, the image brightens up showing more of the interior, while, at the same time, clipping highlight detail. (a) Frame 0 shows the writing on the white paper taped to the door very clearly, but the interior is completely black. (b) In Frame 3 the paper is obliterated—it is so "washed out" that we can no longer read what is written on it. Although the interior is getting brighter at this point, it is still not discernible in Frame 3—we can see neither what is written on the paper, nor what is at the end of the dark corridor. However, as the author turns his head to the right, pointing the camera into the dark corridor, more and more detail of the interior becomes visible as we proceed through the sequence, revealing the inner depths of the long dark corridor, and showing that the fire exit is blocked by the clutter inside.

should be noted that the final quantigraphic composite can be made in the coordinates of any of the images. The choice of reference frame is arbitrary since the result is a floating point image array (not quantized). Furthermore, the final composite need not even be expressed in the spatiotonal coordinates of any of the incoming images. For example, quantigraphic coordinates (linear in the original light falling on the image array) may be used to provide an array of measurements that linearly represent the quantity of light to within a single unknown scalar constant for the entire array.

Once spatiotonally registered, each pixel of the output image is constructed from a weighted sum of the images whose coordinate-transformed bounding boxes fall within that pixel. The weights in the weighted sum are the so-called "certainty functions," which are found by evaluating the derivative of the corresponding estimated effective characteristic function at the pixel value in question [65].

Although the response function $f(q)$, is fixed for a given camera, the "effective response function" $f(k_i(q))$ depends on the exposure k_i associated with frame i in the image sequence.

The composite image may be explored interactively on a computer system (Fig. 11). This makes the personal imaging apparatus into a remote camera in which viewers on the World Wide Web experience something similar to a QuickTime virtual reality (VR) environment map [66], except with some new additional controls allowing them to move around in the environment map both spatially and tonally.

It should be noted that the environment map was generated by a covert wearable apparatus simply by looking

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

Fig. 9. Images of Fig. 8 expressed in the spatiotonal coordinates of the first image in the sequence: (a) frame 0, (b) frame 3, (c) frame 6, and (d) frame 9. Note both the keystoning and chirping effect of the images toward the end of the sequence, indicating the spatial coordinate transformation, as well as the darkening, indicating the tone scale adjustment, both of which make the images match (a). Prior to quantization for printing in this figure, the images that were most severely darkened [e.g., (b) and (c)] to match (d) contained a tremendous deal of shadow detail owing to the fact that the quantigraphic step sizes are much smaller when compressed into the range of (a).

around, and that no special tripod or the like was needed, nor was there significant conscious thought or effort required. In contrast to this proposed method of building environment maps, consider what must be done to build an environment map using QuickTime VR.

> …despite more than twenty years photographic experience, Charbonneau needed to learn new approaches for this type of photography. First, a special tripod rig is required, as the camera must be completely level for all shots. A 35-mm camera with a lens wider than 28 mm is best, and the camera should be set vertically instead of horizontally on the tripod. Exposure is another key element. Blending together later will be difficult unless identical exposure is used for all views [66].

The constraint of the QuickTime VR method and many other methods reported in the literature [50], [41], [43], i.e., that all pictures be taken with identical exposure, is undesirable for the following reasons.

1) It requires a more expensive camera as well as a nonstandard way of shooting (most low cost cameras have automatic exposure that cannot be disabled, and even on cameras where the AGC can be disabled, AGC is still used so the methods will seldom work with pre-existing video that was not shot in this special manner).
2) Imposing that all pictures be taken with the same exposure means that those images shot in bright areas of the scene will be grossly overexposed, while those shot in dark areas will be grossly underexposed. Normally the AGC would solve this problem and adjust the exposure as the camera pans around the scene, but since it must be shut off, shooting all the pictures at the same exposure will mean that

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

Fig. 10. Images of Fig. 8 expressed in the spatiotonal coordinates of the last image in the sequence: (a) frame 0, (b) frame 3, (c) frame 6, and (d) frame 9. Before requantization for printing in this figure, (d) had the highest level of highlight detail owing to is very small quantigraphic quantization step size in the bright areas of the image.

> most scenes will not record well. Thus, special studio lighting is often required to carefully ensure that everything in the scene is equally illuminated.

In contrast to the prior art, the proposed method allows natural scenes of extremely high dynamic range to be captured from a covert eyeglass-mounted camera by simply looking around. The natural AGC of the camera ensures that: 1) the camera will adjust itself to correctly expose various areas of the scene, so that no matter how bright or dark (within a very large range) objects in the scene are, they will be properly represented without saturation or cutoff and 2) the natural ebb and flow of the gain, as it tends to fluctuate, will ensure that there is a great deal of overlapping scene content that is differently exposed, and thus the same quantities of light from each direction in space will be measured with a large variety of different quantization steps. In this way, it will not be necessary to deliberately shoot at different apertures in order to obtain the Wyckoff effect.

Once the final image composite, which reports, up to a single unknown scalar, the quantity of light arriving from each direction in space, it may also be reduced back to an ordinary (i.e., nonquantigraphic) picture by evaluating it with the function f. Furthermore, if desired, prior to evaluating it with f, a lateral inhibition similar to that of the human visual system may be applied to reduce its dynamic range so that it may be presented on a medium of limited display resolution, such as a printed page (Fig. 12). It should be noted that this quantigraphic filtering process (that of producing 12) would reduce to a variant of homomorphic filtering in the case of a single image $E(\mathbf{x})$, in the sense that E would be treated to a global nonlinearity f^{-1} (to obtain q), then linearly processed (e.g., with unsharp masking or the like), and then the nonlinearity f^{-1} would be undone by applying f

$$E_c = f\big(L\big(f^{-1}(E)\big)\big) \tag{18}$$

where E_c is the output (or composite) image and L is the linear filtering operation. Images sharpened in this way tend to have a much richer, more pleasing and natural appearance [19] than those that are sharpened according to

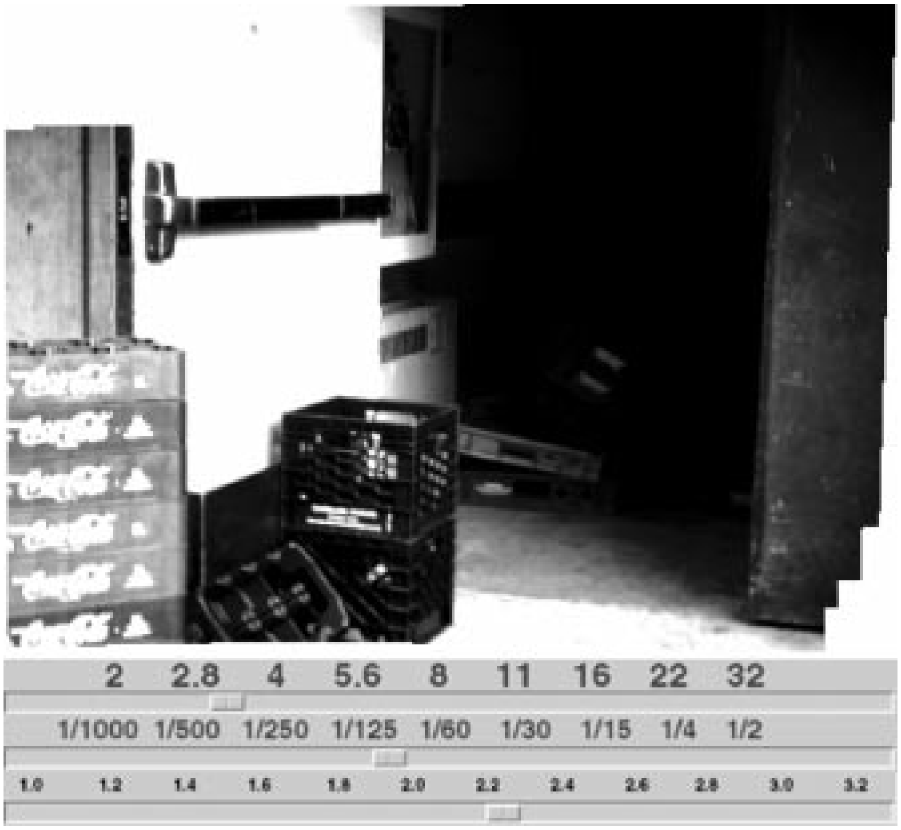


Fig. 11. Virtual camera: floating point projectivity + gain image composite constructed from the fire-exit sequence. The dynamic range of the image is far greater than that of a computer screen or printed page. The quantigraphic information may be interactively viewed on the computer screen, however, not only as an environment map (with pan, tilt, and zoom), but also with control of "exposure" and contrast.

either a linear filter or the variant of homomorphic filtering suggested by Stockham [29].

Perhaps the greatest value of quantigraphic imaging, apart from its ability to capture high-quality pictures that are visually appealing, is its ability to measure the quantity of light arriving from each direction in space. In this way, quantigraphic imaging turns the camera into an array of accurate light meters. Furthermore, the process of making these measurements is activity driven in the sense that areas of interest in the scene will attract the attention of the human operator, so that he or she will spend more time looking at those parts of the scene. In this way, those parts of the scene of greatest interest will be observed with the greatest variety of quantization steps (e.g., with the richest collection of differently quantized measurements) and will therefore, without conscious thought or effort on the part of the wearer, be automatically emphasized in the composite representation. This natural foveation process arises, not because the AI problem has been solved and built into the camera so that it knows what is important, but simply because the camera is using the operator's brain as its guide to visual saliency. Because the camera does not take any conscious thought or effort to operate, it resides on the human host without presenting the host with any burden, yet it benefits greatly from this form of humanistic computing.

4) Bifoveated WearCam: The natural foveation, arising from the symbiotic relationship between human and machine (humanistic computing) described in Section V-B3 may be further accentuated by building a camera system that is itself foveated.

Accordingly, the author designed and built a number of WearComp embodiments containing more than one electronic imaging array. One common variant, with a wide-angle camera in landscape orientation combined with a telephoto camera in portrait orientation, was found to be particularly useful for humanistic computing: the wide camera provided the overall contextual information from the wearer's perspective, while the other (telephoto) provided close-up details, such as faces.

This "bifoveated" scheme was found to work well within the context of the spatiotonal model described in the previous Section (V-B3).

One realization of the apparatus comprised two cameras concealed in a pair of ordinary eyeglasses and is depicted in Fig. 13. It should be noted that there are precedents for display-only systems, such as Kaiser ElectroOptical's head-mounted display product, but that the use of multiple resolution levels within the current invention is new.

Signal processing with respect to bifoveated cameras is a special consideration. In particular, since the geometry of one camera is fixed (in epoxy or the like) with respect to the other, there exists a fixed coordinate transformation that maps any image captured on the wide camera to one that was captured on the foveal camera at the same time. Thus,

Fig. 12. Fixed-point image made by tone-scale adjustments that are only locally monotonic, followed by quantization to 256 greylevels. Note that we can see clearly both the small piece of white paper on the door (and even read what it says—“COFFEE HOUSE CLOSED”), as well as the details of the dark interior. Note that we could not have captured such a nicely exposed image using an on-camera “fill-flash” to reduce scene contrast because the fill-flash would mostly light up the areas near the camera (which happen to be the areas that are already too bright), while hardly affecting objects at the end of the dark corridor which are already too dark. Thus, one would need to set up additional photographic lighting equipment to obtain a picture of this quality. This image demonstrates the advantage of a small lightweight personal imaging system built unobtrusively into a pair of eyeglasses, in that an image of very high quality was captured by simply looking around without entering the corridor. This might be particularly useful if trying to report a violation of fire-safety laws, while at the same time not appearing to be trying to capture an image. The success of the covert, high-definition image capture device suggests possible applications in investigative journalism, or simply to allow ordinary citizens to report violations of fire safety without alerting the perpetrators.

when there is a large jump between images captured on the foveal camera—a jump too large to be considered in the neighborhood of the identity—signal processing algorithms may look to the wide camera for a contextual reference (owing to the greater overlap between images captured on the wide camera), apply the estimation algorithm to the two wide images, and then relate these to the two foveal images. Furthermore, additional signal inputs may be taken from miniature wearable radar systems, inertial guidance or an electronic compass built into the eyeglasses or clothing. These extra signals typically provide ground-truth as well as cross-validation of the estimates reported by the proposed algorithm. The procedure (described in more detail in [20]) is illustrated in Fig. 14.

5) Lightspace Modeling for Humanistic Computing: The result of quantigraphic imaging is that, with the appropriate signal processing, WearComp can measure the quantity of light arriving from each angle in space. Furthermore, because it has display capability (usually the camera sensor array and display element are both mounted in the same eyeglass frame), it may also direct rays of light into the eye. Suppose that the display element has a response function h. The entire apparatus (camera, display, and signal processing circuits) may be used to create an “illusion of transparency” through display of the quantity $h^{-1}(f^{-1}(E_c)) = h^{-1}(q)$, where E_c is the image from the camera. In this way, the wearer sees “through” (i.e., by virtue of) the camera,[13] and would be blind to outside objects in the region over which the apparatus operates if not for the camera.

Now suppose that a filter L is inserted into the “reality stream” by virtue of the appropriate signal processing on the incoming images E_c prior to display on h

$$E_m = h^{-1}\big(L\big(f^{-1}(E_c)\big)\big). \qquad (19)$$

In this context, L is called the “visual filter” [16] and may be more than just a linear spatial filtering operation. As a trivial but illustrative example, consider L such that it operates spatially to flip the image left–right. This would make the apparatus behave like the left–right reversing glasses that Kohler [67] and Dolezal [68] made from prisms for their psychophysical experiments. [See Fig. 15; mediated reality (MR).] In general, through the appropriate selection of L, the perception of visual reality may be augmented, deliberately diminished (e.g., to emphasize certain objects by diminishing the perception of all but those objects) or otherwise altered.

---

[13] In some embodiments of WearComp, only a portion of the visual field is mediated in this way. Such an experience is referred to as “partially mediated reality” [16].

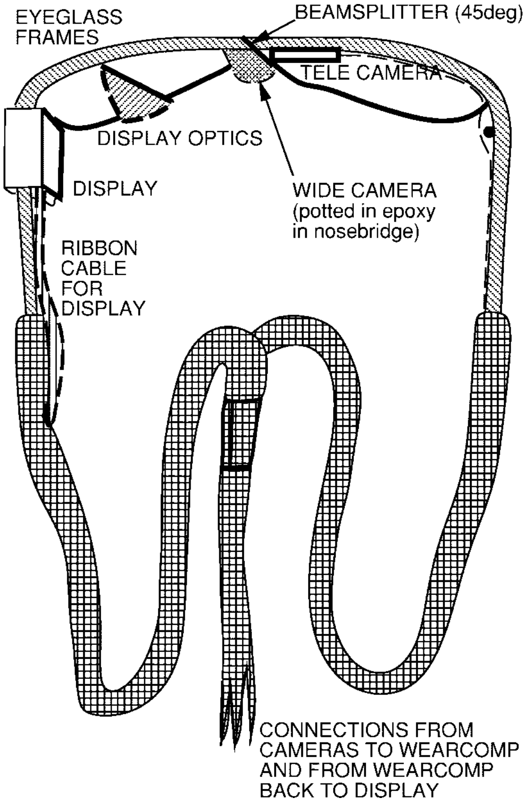


Fig. 13. Author's multicamera personal imaging system with two miniature cameras and display built into ordinary eyeglasses. This bifoveated scheme was found to be useful in a host of applications ranging from crime-reduction (personal safety/personal documentary) to situational awareness and shared visual memory.

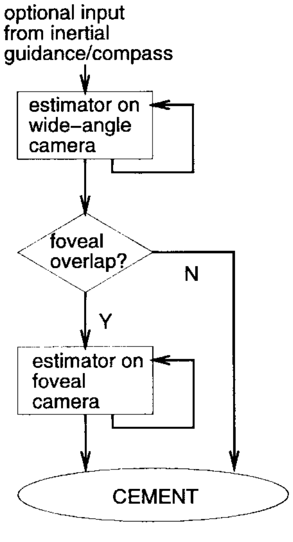


Fig. 14. Signal processing approach for bifoveated WearCam. Note also that the spatial coordinates are propagated according to the projective group's law of composition while the gain parameters between the wide-camera and foveal-camera are not directly coupled.

One feature of this wearable tetherless computer-mediated reality system is that the wearer can choose to allow others to alter his or her visual perception of reality over an Internet-connected wireless communications channel. (An example of such a shared environment map appears in Fig. 16.) This map not only allows others to experience our point of view vicariously (e.g., here a spouse can see that the wearer is at the bank, and send a reminder to check on the status of a loan or pay a forgotten bill), but it can also allow the wearer to allow the distant spouse to mediate the perception of reality. Such mediation may range from simple annotation of objects in the "reality stream" to completely altering the perception of reality.

Other examples of computer-mediated reality include lightspace modeling, so that the response of everyday objects to light can be characterized, and thus the objects can be recognized as belonging to the same orbit of the group of transformations described in this paper. This approach facilitated such efforts as a way-finding apparatus that prevented the wearer from getting lost, as well as an implementation of Feiner's Post-It-note metaphor using a wearable tetherless device, so that messages could be left on everyday objects.

VI. BEYOND VIDEO: SYNTHETIC SYNESTHESIA AND PERSONAL IMAGING

The manner in which WearComp, with its rich multidimensional measurement and signal processing space, facilitates enhanced environmental awareness is perhaps best illustrated by way of the author's effort of the 1980's at building a system to assist the visually challenged. This device, which used radar rather than video as the input modality, is now described.

A. Synthetic Synesthesia: Adding New Sensory Capabilities to the Body

The addition of a new sensory capability for assisting the visually challenged is now described. It has also been found to be of great use to the sighted as well. For example, the author found that increased situational awareness using the system resulted in greater safety in many ordinary day-to-day activities such as riding a bicycle on a busy street.

Mediated reality may include, in addition to video, an audio reality mediator, or more generally, a "perceptual reality mediator." This generalized mediated perception system may include deliberately induced synesthesia.[14] Perhaps the most interesting example of synthetic synesthesia was the addition of a new human sensory capability based on miniature wearable radar systems combined with intelligent signal processing. In particular, the author developed a number of vibrotactile wearable radar systems in the 1980's, of which there were three primary variations.

1) *CorporealEnvelope*: Baseband output from the radar system was envelope-detected to provide a vibrotactile sensation which was proportional to the overall energy of the return.[15] This provided the sensation

---

[14] Synesthesia [72] is manifest as the crossing of sensory modalities, as, for example, the ability (or as some might call a disability) to taste shapes, see sound, etc.

[15] Strictly speaking the actual quantity measured in early systems was that of a single homodyne channel, which only approximated energy. Later, in some systems, energy was measured properly with separate I and Q channels.

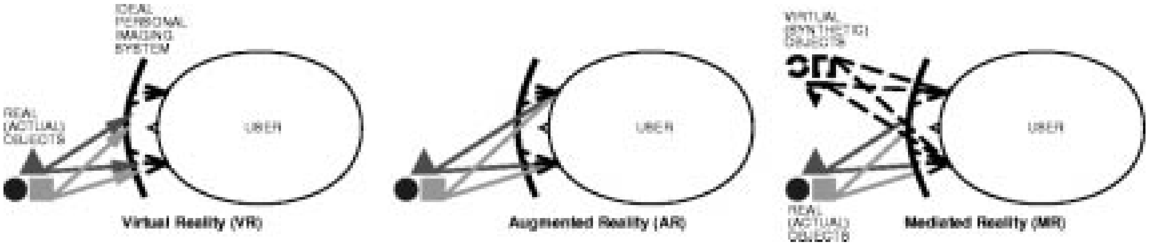


Fig. 15. Lightspace modeling. The WearComp apparatus, with the appropriate quantigraphic signal processing, may be thought of as a hypothetical glass that absorbs and quantifies every ray of light that hits it, and it is also capable of generating any desired bundle of rays of light coming out the other side. Such a glass, made into a visor, could produce a 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 are indicated by solid shaded lines; rays of light from the display device itself are indicated by dashed lines. The device could also produce a typical augmented reality (AR) [69], [70] experience by creating the "illusion of transparency" and also by 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 simple but illustrative example is shown: objects are left–right reversed before being presented to the viewer.

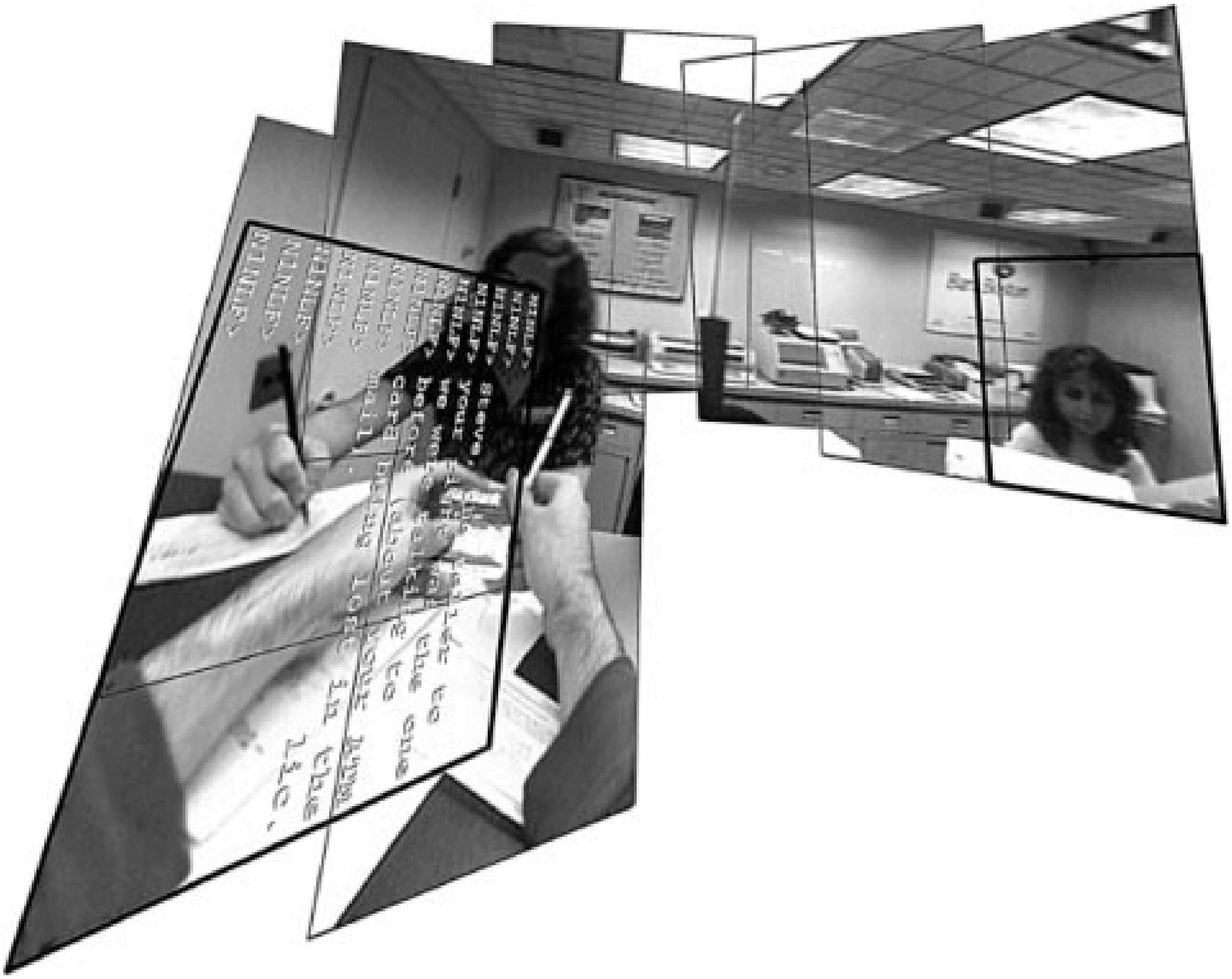


Fig. 16. Shared environment maps are one obvious application of WearComp. Multiple images transmitted from the author's "Wearable Wireless Webcam" [71] may be seamlessly combined together onto a World Wide Web page so that others can see a first-person-perspective point of view, as if looking out through the eyes of the person wearing the apparatus. However, because the communication is bidirectional, others can also communicate with the wearer by altering the wearer's visual perception of reality. This might, for example, allow one to recognize people one has never met before. Thus, personal imaging allows the individual to go beyond vicarious experience toward a more symbiotic relation to a networked collective humanistic computing environment within a mediated visual reality [16]. © Steve Mann, 1995. (Picture rendered at higher-than-normal screen resolution for use as cover for the journal *Presence*.)

of an extended "envelope" around the body in which one could feel objects at a distance. In later (late-1980's) embodiments of CorporealEnvelope, envelope detection was done after splitting the signal into three or four separate frequency bands, each driving a separate vibrotactile device so that each would convey a portion of the Doppler spectrum (e.g., each corresponding to a range of velocities of approach). In another late-1980's embodiment variously colored lamps were attached to the wearer's eyeglasses to provide a visual synesthesia of the radar sense. In one particular embodiment, red, green, and blue lamps were used such that objects moving toward the wearer illuminated the blue lamp, while objects moving away illuminated the red lamp. Objects not moving relative to the wearer, but located near the wearer appeared green. This work was inspired by using the metaphor of the natural Doppler shift colors.

2) *VibroTach (Vibrotactile Tachometer)*: The speed of objects moving toward or away from the wearer was conveyed, but not the magnitude of the Doppler return (e.g., it was not possible to distinguish between objects of small radar cross section and those of large radar cross section). This was done by having a Doppler return drive a motor, so that the faster an object moved toward or away from the wearer the faster the motor would spin. The spinning motor could be felt as a vibration having frequency proportional to that of the dominant Doppler return.
3) *Electric Feel Sensing*: The entire Doppler signal (not just a single dominant speed or amplitude) was conveyed to the body. Thus, if there were two objects approaching at different speeds, one could discern them separately from a single vibrotactile sensor. Various embodiments of Electric Feel Sensing included direct electrical stimulation of the body, as well as the use of a single broadband vibrotactile device.

One of the problems with this work was the processing, which in the early-1980's embodiments of WearComp was very limited. However, today's wearable computers, far more capable of computing the chirplet transform[16] in real time, suggest a renewed hope for the success of this effort to assist the visually impaired.

1) Safety First!: Again, with direct connection to the body, there must of course be extra attention devoted to user safety. For example, direct electrical stimulation of the body has certain risks associated with it, such as nerve damage from excessively strong signals, as well as nerve damage from weaker signals that are not properly conditioned (such as may happen with excessive DC offset). Similarly, vibrotactile devices may also afflict the user with long-term damage, as one might experience with any sensory device in the extreme (just as loud music can cause deafness, and excessively bright head mounted displays can cause blindness, excessive vibrotactile stimulation can cause loss of feeling). While the radar signals themselves tend to be less of a concern, owing to their very low power levels (often below the level of background radiation), there should still be the obvious precautions taken with radar as with any other radio signals (such as the much more powerful transmitters used to establish an Internet connection). The potential problem of radio frequency noise pollution has been addressed through the use of very low-power transmissions from the radar. Because of the homodyne nature of the receiver, a very weak signal may be sent out, since it is known exactly what signal will be expected back. Even in pulsed modes of operation, by virtue of pulse compression, signal levels can often remain below the ambient signal levels. Radar systems operate according to an inverse fourth power law (since what is transmitted falls off as the square of the distance, and what is received back falls off as the square of the distance also). Distances to objects of interest with a personal radar system are typically on the order of only a few meters, in contrast to traditional radar, where distances of interest are many kilometers. Thus, because power output needed is proportional to the fourth exponent of the distance from the objects of interest, power output is very low and thus has not been a hazard.

One must also consider the safety issues of both the effects of synthetic synesthesia, as well as the development of a reliance upon it. Synthetic synesthesia involves a remapping of the human perceptual system, the long-term effects of which should still be studied carefully. Acquired dependence is a concern which might not at first occur to one practicing this art. In many ways, the author discovered that after many years the device began to function as a true extension of the mind and body, as if it were an additional sense. Much like a person who is blind at birth but has his or her sight restored later in life due to medical advancements, there is a strange sense of confusion when a new sense is introduced to the body without the support infrastructure within the brain. After many years of use, one begins to learn the new sense, and internalize the meanings of the new sensory modalities. Together with this remapping of the human perceptual system, and the obvious dangers it might pose (and the question as to whether learning is damage, since learning permanently alters the brain), is the deeper philosophical question as to whether or not acquiring a dependence on a new sensory modality is a good idea.

As a specific example, in the form of a personal anecdote, the author had one time found himself on a long bike trip (180 mile trip along major busy highways) relying greatly on this new sense, when at some point along the trip there was a major thunderstorm which required removal and shutting down of the extra sensory capabilities. In many ways the author felt at great risk, owing to the heavy traffic and the acquired need to have enhanced situational awareness. Removal of a new sensory capability, after an acquired dependency, was much like removal of an existing sense (e.g., like suddenly having to ride a bicycle while blindfolded or wearing ear plugs).

---

[16] The chirplet transform [73] characterizes the acceleration signature of Doppler returns so that objects can be prioritized, e.g., those accelerating faster toward the wearer can be given higher priority, predicting eminent collision, etc.

Thus an important safety issue as we enhance our intellectual and sensory capabilities will be risks involved should this apparatus ever quit functioning after we have become dependent upon it. Acquired dependency is nothing new, of course. For example, we have acquired a dependency on shoes and clothing, and would doubtless have much greater difficulty surviving naked in the wilderness than might those indigenous to the wilderness, especially those who had not invented shoes or clothing.

Thus, as we build prostheses of the mind and body, we must consider carefully their implications, especially as they pertain to personal safety. This new computational framework will therefore give a whole new meaning to the importance of reliability.

2) A True Extension of the Mind and Body: Such simple early prototypes as those discussed already suggest a future in which intelligent signal processing, through the embodiment of humanistic computing, may allow the wearer to experience increased situational awareness. It will then be misleading to think of the wearer and the computer with its associated input/output apparatus as separate entities. Instead it will be preferable to regard the computer as a second brain, and its sensory modalities as additional senses through which synthetic synesthesia are inextricably intertwined with the wearer's own biological sensory apparatus.

VII. CONCLUSION

A new form of intelligent signal processing, called "humanistic computing" was proposed. It is characterized by processing hardware that is inextricably intertwined with a human being to function as a true extension of the user's mind and body. This hardware is constant (always on, therefore its output is always observable), controllable (i.e., is not merely a monitoring device attached to the user, but rather, it takes its cues from the user), and corporeal in nature (i.e., tetherless and with the point of control in close proximity to the user so as to be perceived as part of the user's body). Furthermore, the apparatus forms a symbiotic relationship with its host (the human), in which the high-level intelligence arises on account of the existence of the host (human), and the lower-level computational workload comes from the signal processing hardware itself.

The emphasis of this paper was on personal imaging, to which the application of humanistic computing gave rise to a new form of intelligent camera system. This camera system was found to be of great use in both photography and documentary video making. Its success arose from the fact that it: 1) was simpler to use than even the simplest of the so-called "intelligent point and click" cameras of the consumer market (even though many of these embody sophisticated neural network architectures) and 2) afforded the user much greater control than even the most versatile and fully featured of professional cameras.

This application of humanistic computing took an important first step in moving from the "point and click" metaphor toward the "look and think" metaphor—toward making the camera function as a true visual memory prosthetic which operates without conscious thought or effort, while at the same time affording the visual artist a much richer and complete space of possibilities. Moreover, this work sets forth the basic principles of a photo-videographic memory system.

A focus of humanistic computing was to put the human intellect into the loop but still maintain facility for fail-safe mechanisms operating in the background. Thus the personal safety device was proposed.

What differentiates humanistic computing from environmental intelligence (ubiquitous computing [74], reactive rooms [75], and the like) is that there is no guarantee environmental intelligence will be present when needed, or that it will be in control of the user. Instead, humanistic computing provides a facility for intelligent signal processing that travels with the user. Furthermore, because of the close physical proximity to the user, the apparatus is privy to a much richer multidimensional information space than that obtainable by environmental intelligence.

Furthermore, unlike an intelligent surveillance camera that people attempt to endow with an ability to recognize suspicious behavior, WearComp takes its task from the user's current activity, e.g., if the user is moving, the apparatus is continually rendering new composite pictures, while if the user is not moving it is no longer taking in new orbits. This activity-based response is based on the premise that the viewpoint changes cause a change in orbit, etc.

The following represents characteristics of systems embodying humanistic computing.

1) *Activity driven and attention driven*: Salience is based on the computer's taking information in accordance with human activity. Video orbits are activity driven (i.e., they start when wearer stops at a fixed orbit). In other words the visual salience comes from the human; the computer is doing the processing but taking cue from the wearer's activity. For example, if the wearer is talking to a bank clerk, but takes brief glances at the periphery, the resulting image will reveal the wearer's clerk in high resolution, while the other clerks to the left and right will be quantified at much lesser certainty. Further processing on the image measurements thus reflect this saliency so that the system adapts to the manner in which it is used.
2) *Environmentally aware*: Situated awareness arises in the context of both the wearer's environment and his/her own biological quantities which, through the wearer's own mind, body, and perceptual system, also depend on the environment.
3) *Inextricably intertwined with the human, i.e., situated*: If the user is aroused the system will take more pictures. In this way, the computation makes a departure from that of traditional artificial intelligence. The processor automatically uses the wearer's sense of salience to help it so that the machine and human are always working in parallel.

ACKNOWLEDGMENT

The author wishes to thank S. Haykin, R. Picard, S. Feiner, C. Wyckoff, W. Barfield, S. Zaky, J. Rose, H. Ishii, T. Starner, J. Levine, F. Sparacino, K. Russell, R. Mann, R. Mann, B. Mann, and S. Roberts for much in the way of useful feedback, constructive criticism, etc., as this work has evolved, and Z. Parpia for making some important suggestions for the presentation of this material. Thanks is due also to individuals the author has hired to work on this project, including N. Friedman, C. Cgraczyk, M. Reynolds, etc., who each contributed to this effort.

Dr. Carter volunteered freely of his time to help in the design of the interface to WearComp2 (the author's 6502-based wearable computer system of the early 1980's), and K. Nickerson similarly helped with some of the miniature personal radar units and photographic devices involved with this project throughout the mid-1980's.

Much of the early work on biosensors and wearable computing was done with, or at least inspired by work the author did with Dr. N. Ghista, and later refined with suggestions from Dr. H. DeBruin, both of McMaster University. Dr. M. Wong of McMaster University supervised a course project in which the author chose to design an RF link between two 8085-based WearComp systems which had formed part of the author's "photographer's assistant" project.

Much of the inspiration toward making truly wearable (also comfortable and even fashionable) signal processing systems was through collaboration with J. Eleveld during the early 1980's.

B. Kinney of U.S. Army Natick Research Labs assisted in the design of a tank top, based on a military vest, which the author used for a recent (1996) embodiment of the WearComp apparatus worn underneath ordinary clothing.

Additional thanks to Kodak, Xybernaut Corp., Digital Equipment Corp., ViA, VirtualVision, HP Labs, Compaq, Kopin, Colorlink, Ed Gritz, Miyota, C. Carter, and Thought Technologies Limited for lending or donating additional equipment that made these experiments possible.

REFERENCES

[1] W. A. S. Buxton and R. M. Baecker, *Readings in Human-Computer Interaction a Multidisciplinary Approach*. ch. 1, 2, 1987.

[2] S. Haykin, *Neural Networks: A Comprehensive Foundation*. New York: McMillan, 1994.

[3] ——, "Radar vision," presented at *2nd Int. Specialist Seminar Parallel Digital Processors*, Portugal, Apr. 15–19, 1992.

[4] D. E. Rumelhart and J. L. McMelland, Eds., *Parallel Distributed Processing* Cambridge, MA: MIT, 1986.

[5] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, pp. 1481–1487, Sept. 1990.

[6] B. Kosko, *Fuzzy Thinking: The New Science of Fuzzy Logic* New York: Hyperion, 1993.

[7] B. Kosko and S. Isaka, *Fuzzy Logic*, vol. 269. New York: Scientific American, 1993, pp. 76–81.

[8] P. Laplante and M. Minsky, Eds., "Steps toward artificial intelligence," in *Great Papers on Computer Science*. St. Paul, MN: West 1996.

[9] D. C. Engelbart, "Augmenting human intellect. A conceptual framework," Stanford Research Institute, Menlo Park, CA, Res. Rep. AFOSR-3223, 1962. [Online]. Available WWW: http://www.histech.rwth-aachen.de/www/quellen/ engelbart./ahi62index.html.

[10] ——, "A conceptual framework for the augmentation of man's intellect," in *Vistas in Information Handling*, P. D. Howerton and D. C. Weeks, Eds. Washington, DC: Spartan, 1963.

[11] S. Mann and R. W. Picard, "Method and apparatus for relating and combining multiple images of the same scene or object(s)," U.S. Patent 5706416, Jan. 6, 1998.

[12] S. Mann, "Wearable computing: A first step toward personal imaging," *IEEE Trans. Comput.*, vol. 30, pp. 25–32, Feb. 1997. [Online]. Available WWW: http://hi.eecg.toronto.edu/ieeecomputer/index.html.

[13] M. Foucault, *Discipline and Punish* (translated from *Surveiller et Punir*). New York: Pantheon, 1977.

[14] S. M. Kosslyn, *Image and Brain The Resolution of the Imagery Debate* Cambridge, MA: MIT, 1994.

[15] S. Mann, "'WearStation': With today's technology, it is now possible to build a fully equipped ham radio station, complete with internet connection, into your clothing," *CQ-VHF*, pp. 1–46, Jan. 1997.

[16] ——, "'Mediated reality'," Media Lab Perceptual Computing Section, MIT, Cambridge, MA, Tech. Rep. 260, 1994. [Online]. Available WWW: http://wearcam.org/ mediated-reality/index.html.

[17] J. D. Meindl, "Low power microelectronics: Retrospect and prospect," *Proc. IEEE*, vol. 83, pp. 619–635, Apr. 1995.

[18] S. Mann, "An historical account of the 'WearComp' and 'WearCam' projects developed for 'personal imaging'," presented at *IEEE Int. Symp. Wearable Computing*, Cambridge, MA, Oct. 13–14, 1997.

[19] ——, "Personal imaging," Ph.D. dissertation, MIT, Cambridge, MA, 1997.

[20] ——, "Further developments on 'headcam': Joint estimation of camera rotation + gain group of transformations for wearable bi-foveated cameras," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, Munich, Germany, Apr. 1997, vol. 4, pp. 2909–2912.

[21] E. J. Lind and R. Eisler, "A sensate liner for personnel monitoring application," presented at *IEEE Int. Symp. Wearable Computing*, Oct. 13–14, 1997.

[22] R. W. Picard and J. Healey, "Affective wearables," in *Proc. 1st Int. Symp. Wearable Computers*, Cambridge, MA, Oct. 1997, pp. 90–97.

[23] S. Mann, "Smart clothing: The wearable computer and wearcam," *Personal Technol.*, vol. 1, no. 1, pp. 21–27, Mar. 1997.

[24] R. W. Picard, *Affective Computing*. Cambridge, MA: MIT, 1997.

[25] D. Norman, *Turn Signals Are the Facial Expressions of Automobiles*. Reading, MA: Addison Wesley, 1992.

[26] S. Mann, "'Smart clothing': Wearable multimedia and 'personal imaging' to restore the balance between people and their intelligent environments," in *Proc. ACM Multimedia'96*, Boston, MA, Nov. 18–22 1996, pp. 163–164.

[27] ——, "Humanistic intelligence," in *Proc. Ars Electronica*, Sept. 8–13, 1997, pp. 217–231, [Online]. Available WWW: http://wearcam.org/ars/; http://www.aec.at/fleshfactor.

[28] ——, "Lightspace," submitted for publication.

[29] T. G. Stockham, Jr., "Image processing in the context of a visual model," *Proc. IEEE*, vol. 60, pp. 828–842, July 1972.

[30] E. H. Land, "The retinex theory of color vision," *Sci. Amer.*, pp. 108–129, 1977.

[31] D. Forsyth, "A novel algorithm for color constancy," *Int. J. Comput. Vision*, vol. 5, no. 1, pp. 5–36, Aug. 1990.

[32] A. N. Venetsanopoulos, "Nonlinear mean filters in image processing," *IEEE Trans. Audio, Speech, Signal Processing*, vol. 34, pp. 573–584, June 1986.

[33] S. Mann, "Compositing multiple pictures of the same scene," in *Proc. 46th Ann. IS&T Conference*, Cambridge, MA, May 9–14, 1993, pp. 50–52.

[34] ——, "'Pencigraphy' with AGC: Joint parameter estimation in both domain and range of functions in same orbit of the projective-Wyckoff group," MIT Media Lab., Cambridge, MA, Tech. Rep. 384, Dec. 1994. [Online]. Available WWW: http://hi.eecg.toronto.edu/icip96/index.html. Also appears in *Proc. IEEE Int. Conf. Image Processing (ICIP-96)*, Lausanne, Switzerland, pp. 193–196.

[35] S. Mann and R. W. Picard, "Virtual bellows: Constructing high-quality images from video," in *Proc. IEEE 1st Int. Conf. Image Processing*, Austin, Texas, Nov. 13–16, 1994, pp. 363–367.

[36] C. W. Wyckoff, “An experimental extended response film,” *S.P.I.E. Newslett.*, pp. 16–20, June–July 1962.

[37] ———, “An experimental extended response film,” Edgerton, Germeshausen & Grier, Inc., Boston, MA, Tech. Rep. no. B-321, Mar. 1961.

[38] O. D. Faugeras and F. Lustman, “Motion and structure from motion in a piecewise planar environment,” *Int. J. Pattern Recognition Artificial Intell.*, vol. 2, no. 3, pp. 485–508, 1988.

[39] G. Adiv, “Determining 3D Motion and structure from optical flow generated by several moving objects,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 384–401, July 1985.

[40] A. Shashua and N. Navab, “Relative Affine: Theory and application to 3D reconstruction from perspective views,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, June 1994, pp. 483–489.

[41] H. S. Sawhney, “Simplifying motion and structure analysis using planar parallax and image warping,” in *Proc. 12th ICPR*, vol. 1, Oct. 1994, pp. 403–408.

[42] N. Navab and S. Mann, “Recovery of relative affine structure using the motion flow field of a rigid planar patch.,” presented at Mustererkennung 1994, Tagungsband, Germany.

[43] R. Kumar, P. Anandan, and K. Hanna, “Shape recovery from multiple views: A parallax based approach,” presented at ARPA Image Understanding Workshop, Nov. 10, 1994.

[44] W. A. Radlinski, *Manual of Photogrammetry* (Schaum’s Outline series), 3rd ed., M. M. Thompson, R. C. Eller, and J. L. Speert, Eds. Falls Church, VA: McGraw-Hill, 1966.

[45] M. Irani and S. Peleg, “Improving resolution by image registration,” *CVGIP*, vol. 53, pp. 231–239, May 1991.

[46] A. M. Tekalp, M. K. Ozkan, and M. I. Sezan, “High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration,” in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, San Francisco, CA, Mar. 23–26, 1992, pp. 111–169.

[47] L. Campbell and A. Bobick, “Correcting for radial lens distortion: A simple implementation,” MIT Media Lab., Perceptual Computing Sect., Cambridge, MA, Tech. Rep. 322, Apr. 1995.

[48] K. Levenberg, “A method for the solution of certain nonlinear problems in least squares,” *Quart. J. of Appl. Math.*, vol. 2, pp. 164–168, 1944.

[49] D. W. Marquardt, “An algorithm for the estimation of nonlinear parameters,” *SIAM J.*, vol. 11, pp. 431–441, 1963.

[50] R. Szeliski and J. Coughlan, “Hierarchical spline-based image registration,” *CVPR*, pp. 194–201, June 1994.

[51] S. Mann and R. W. Picard, “Video orbits of the projective group; a simple approach to featureless estimation of parameters,” MIT, Cambridge, MA, Tech. Rep. 338. [Online]. Available WWW: http://n1nlf-1.eecg.toronto.edu/tip.html, 1995. Also appears in *IEEE Trans. Image Processing*, vol. 6, pp. 1281–1295, Sept. 1997.

[52] S. S. Beauchemin, J. L. Barron, and D. J. Fleet, “Systems and experiment performance of optical flow techniques,” *Int. J. Comput. Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[53] Q. Zheng and R. Chellappa, “A computational vision approach to image registration,” *IEEE Trans. Image Processing*, vol. 2, pp. 311–325, July 1993.

[54] L. Teodosio and W. Bender, “Salient video stills: Content and context preserved,” in *Proc. ACM Multimedia Conf.*, New York, Aug. 1993, pp. 39–46.

[55] R. Y. Tsai and T. S. Huang, “Estimating three-dimensional motion parameters of a rigid planar patch,” *Trans. Acoustics, Speech, Signal Processing*, vo. 29, no. 6, pp. 1147–1152, Dec. 1981.

[56] M. Artin, *Algebra*. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[57] W. C. Hoffman, “The lie algebra of visual perception,” *J. Math. Psychology*, vol. 4, pp. 348–349, 1967.

[58] R. I. Hartley, “Self-calibration of stationary cameras,” G.E., Schenectady, NY, Tech. Rep. 12301.

[59] B. Horn and B. Schunk, “Determining optical flow,” *Artificial Intell.*, 1981.

[60] B. Girod and D. Kuo, “Direct estimation of displacement histograms,” presented at *OSA Meeting Image Understanding Machine Vision*, June 1989.

[61] Y. Sheng, C. Lejeune, and H. H. Arsenault, “Frequency-domain Fourier-Mellin descriptors for invariant pattern recognition,” *Opt. Eng.*, vol. 27, no. 5, pp. 354–357, May 1988.

[62] S. Mann, “Wavelets and chirplets: Time-frequency perspectives, with applications,” in *Advances in Machine Vision, Strategies and Applications* (World Scientific series in computer science), 32nd ed., Petriu Archibald, Ed. Singapore: World Scientific, 1992.

[63] R. Wilson, A. D. Calway, and E. R. S. Pearson, “A generalized wavelet transform for Fourier analysis: The multiresolution Fourier transform and its application to image and audio signal analysis,” *IEEE Trans. Inform. Theory*, vol. 38, pp. 674–690, Mar. 1992.

[64] S. R. Curtis and A. V. Oppenheim, “Signal reconstruction from Fourier transform sign information,” MIT Res. Lab. Electron., Cambridge, MA, Tech. Rep. 500, May 1984.

[65] S. Mann and R. W. Picard, “Being ‘undigital’ with digital cameras: Extending dynamic range by combining differently exposed pictures,” MIT Media Lab. Perceptual Comput. Sec.,Boston, MA, Tech. Rep. 323, 1994. Also appears in *Proc. IS&T’s 48th Ann. Conf.*, Washington, DC, May 1995, pp. 422–428. [Online]. Available WWW: http://wearcam.org/ist98.htm

[66] J. Campbell, “QuickTime VR,” Oct. 18, 1996. [Online]. Available WWW: http://www.webreference. com/content/qtvr/.

[67] I. Kohler, *The Formation and Transformation of the Perceptual World*, vol. 3 of *Psychological Issues* [monograph 12]. New York: International Univ. Press, 1964.

[68] H. Dolezal, *Living in a World Transformed* (Academic Press series in cognition and perception). Chicago: Academic, 1982.

[69] S. Feiner, B. MacIntyre, and D. Seligmann, “Knowledge-based augmented reality,” *Commun. ACM*, vol. 36, no. 7, pp. 52–62, July 1993.

[70] S. Feiner, A. Webster, T. Krueger, B. MacIntyre, and E. Keller, “Architectural anatomy,” *Presence*, vol. 4, no. 3, pp. 318–325, 1995.

[71] S. Mann. (1994). Wearable wireless webcam [Online]. Available WWW: http://wearcam.org.

[72] R. E. Cytowic, *Synesthesia: A Union of the Senses*. New York: Springer-Verlag, 1989.

[73] S. Mann and S. Haykin, “The chirplet transform: Physical considerations,” *IEEE Trans. Signal Processing*, vol. 43, no. 11, pp. 2745–2761, Nov. 1995.

[74] M. Weiser. (1988–1994). Ubiquitous computing [Online]. Available WWW: http://sandbox.parc. xerox.com/ubicomp.

[75] J. Cooperstock. Reactive room [Online]. Available WWW: http://www.dgp.toronto.edu/˜rroom/ research/papers/.

Steve Mann (Member, IEEE) received the B.Sc. degree in physics in 1986 and the B.Eng. and M.Eng. degrees in electrical engineering in 1989 and 1991, respectively, all from McMaster University, Hamilton, Ont., Canada. He received the Ph.D. degree from Massachusetts Institute of Technology (MIT), Cambridge, in 1997.

He is currently a Faculty Member at the Department of Electrical and Computer Engineering, University of Toronto, Ont., Canada. He is the inventor of WearComp (wearable computer and personal imaging system), and he is pictured here wearing an early eyeglass-based prototype of his invention. He has been experimenting with wearable computing and personal imaging as a hobby since his high school days in the 1970’s and early 1980’s. In 1991 he introduced his vision to MIT and established a new direction of research there. His present research interests include quantigraphic imaging, formulation of the response of objects and scenes to arbitrary lighting, creating a self-linearizing camera calibration procedure, and wearable, tetherless computer-mediated reality—inventing the camera of the future. He has also exhibited his quantigraphic image composites and lightspace renderings in numerous art galleries from 1985 to the present.

Dr. Mann was Guest Editor of a Special Issue on Wearable Computing and Personal Imaging in *Personal Technologies Journal* as well as one of four organizers of the ACM’s first International Workshop on Wearable Computing. He first proposed and was Publications Chair for the IEEE International Symposium on Wearable Computing (ISWC’97), and he gave the keynote address at the International Conference on Wearable Computing.