

Steve Mann

Mann@eecg.toronto.edu

James Fung

fungja@eecg.toronto.edu

University of Toronto

10 King's College Road

Toronto, Canada

EyeTap Devices for Augmented, Deliberately Diminished, or Otherwise Altered Visual Perception of Rigid Planar Patches of Real-World Scenes

Abstract

Diminished reality is as important as augmented reality, and both are possible with a device called the *Reality Mediator*. Over the past two decades, we have designed, built, worn, and tested many different embodiments of this device in the context of wearable computing. Incorporated into the Reality Mediator is an "EyeTap" system, which is a device that quantifies and resynthesizes light that would otherwise pass through one or both lenses of the eye(s) of a wearer. The functional principles of EyeTap devices are discussed, in detail. The EyeTap diverts into a spatial measurement system at least a portion of light that would otherwise pass through the center of projection of at least one lens of an eye of a wearer. The Reality Mediator has at least one mode of operation in which it reconstructs these rays of light, under the control of a wearable computer system. The computer system then uses new results in algebraic projective geometry and comparametric equations to perform head tracking, as well as to track motion of rigid planar patches present in the scene. We describe how our tracking algorithm allows an EyeTap to alter the light from a particular portion of the scene to give rise to a computer-controlled, selectively mediated reality. An important difference between mediated reality and augmented reality includes the ability to not just augment but also deliberately diminish or otherwise alter the visual perception of reality. For example, diminished reality allows additional information to be inserted without causing the user to experience information overload. Our tracking algorithm also takes into account the effects of automatic gain control, by performing motion estimation in both spatial as well as tonal motion coordinates.

I Introduction

Ivan Sutherland, a pioneer in the field of computer graphics, described a head-mounted display with half-silvered mirrors so that the wearer could see a virtual world superimposed on reality (Earnshaw, Gigante, & Jones, 1993; Sutherland, 1968), giving rise to augmented reality (AR).

Others have adopted Sutherland's concept of a head-mounted display (HMD) but generally without the see-through capability. An artificial environment in which the user cannot see through the display is generally referred to as a virtual reality (VR) environment. One of the reasons that Sutherland's ap-

proach was not more ubiquitously adopted is that he did not merge the virtual object (a simple cube) with the real world in a meaningful way. Feiner's group was responsible for demonstrating the viability of AR as a field of research, using sonar (Logitech 3-D trackers) to track the real world so that the real and virtual worlds could be registered (Feiner, MacIntyre, & Seligmann, 1993a, 1993b). Other research groups (Fuchs, Bajura, & Ohbuchi; Caudell & Mizell, 1992) also contributed to this development. Some research in AR also arises from work in telepresence (Drascic & Milgram, 1996).

However, the concept of the Reality Mediator, which arises from the field of humanistic intelligence (HI) (Mann, 1997a, 2001a, 2001b) differs from augmented reality, which has its origins in the field of virtual reality. HI is defined as intelligence that arises from the human being in the feedback loop of a computational process in which the human and computer are inextricably intertwined. Wearable computing has emerged as the perfect tool for embodying HI. When a wearable computer functions as a successful embodiment of HI, the computer uses the human's mind and body as one of its peripherals, just as the human uses the computer as a peripheral. This reciprocal relationship, in which each uses the other in its feedback loop, is at the heart of HI. Within an HI framework, the wearable computer is worn constantly to assist the user in a variety of day-to-day situations.

An important observation arising from this constant use is that, unlike handheld devices, laptop computers, and PDAs, the wearable computer can encapsulate us (Mann, 1998). It can function as an information filter and allow us to block out material we might not wish to experience (such as offensive advertising) or simply replace existing media with different media. Thus, the wearable computer acts to mediate one's experience with the world. The mediating role of EyeTap and wearable computers can be better understood by examining the signal flow paths between the human, computer, and external world as illustrated in figure 1. There exist well known email and Web browser filters that replace or remove unwanted advertising, mediating one's use of the media. Diminished reality extends this mediation to the visual domain.

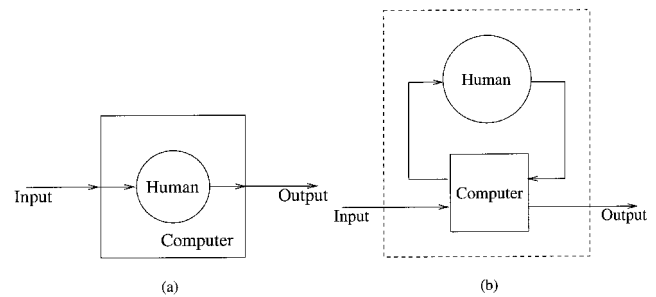


Figure 1. (a) The wearable computer can be used like clothing to encapsulate the user and function as a protective shell, whether to protect us from cold or physical attack (as traditionally facilitated by armor), or to provide privacy (by concealing personal information and personal attributes from 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, in various degrees, that makes wearable computing much more empowering to the user than other similar forms of portable computing. (b) An equivalent depiction of encapsulation (mediation) redrawn where the encapsulation is understood to comprise a separate protective shell.

2 EyeTap Devices

Just as half-silvered mirrors are used to create augmented reality, EyeTap devices are used to mediate one's perception of reality. EyeTap devices have three main components:

- a measurement system typically consisting of a camera system, or sensor array with appropriate optics;
- a diverter system, for diverting eyeward bound light into the measurement system and therefore causing the eye of the user of the device to behave, in effect, as if it were a camera; and
- an aremac for reconstructing at least some of the diverted rays of eyeward bound light. (Thus, the aremac does the opposite of what the camera does and is, in many ways, a camera in reverse. The etymology of the word *aremac* itself arises from spelling the word *camera* backwards (Mann, 1997c).)

A number of such EyeTap devices, together with wearable computers, were designed, built, and worn by the authors for many years in a wide variety of settings and situations, both inside the lab as well as in ordinary day-to-day life (such as while shopping, riding a bicycle, going through airport customs, attending weddings, and so on). This broad base of practical real-life experience helped us better understand the fundamental issues of mediated reality.

Although the apparatus is for altering our vision, in most of the practical embodiments that we built we provided at least one mode of operation that can preserve our vision unaltered. This one mode, which we call the “identity mode,” serves as a baseline that forms a point of departure for when certain changes are desired. To achieve the identity mode requirement, the EyeTap must satisfy three criterion:

- *focus*: The subject matter viewed through the eyetap must be displayed at the appropriate depth of focus.
- *orthospaciality*: The rays of light created by the aremac must be collinear with the rays of light entering the EyeTap, such that the scene viewed through the EyeTap appears the same as if viewed in the absence of the EyeTap.
- *orthotonality*: In addition to preserving the spatial light relationship of light entering the eye, we desire that the EyeTap device also preserves the tonal relationships of light entering the eye.

2.1 Focus and Orthospaciality in EyeTap Systems

The aremac has two embodiments: one in which a focuser (such as an electronically focusable lens) tracks the focus of the camera to reconstruct rays of diverted light in the same depth plane as imaged by the camera, and another in which the aremac has extended or infinite depth of focus so that the eye itself can focus on different objects in a scene viewed through the apparatus.

Although we have designed, built, and tested many of each of these two kinds of systems, this paper describes only the systems that use focus tracking.

In the focus-tracking embodiments, the aremac has focus linked to the measurement system (for example, “camera”) focus, so that objects seen depicted on the aremac of the device appear to be at the same distance from the user of the device as the real objects so depicted. In manual focus systems, the user of the device is given a focus control that simultaneously adjusts both the aremac focus and the “camera” focus. In automatic focus embodiments, the camera focus also controls the aremac focus. Such a linked focus gives rise to a more natural viewfinder experience. It reduces eyestrain as well. Reduced eyestrain is important because these devices are intended to be worn continually.

The operation of the depth tracking aremac is shown in figure 2.

Because the eye’s own lens (L_3) experiences what it would have experienced in the absence of the apparatus, the apparatus, in effect, taps in to and out of the eye, causing the eye to become both the camera and the viewfinder (display). Therefore, the device is called an EyeTap device.

Often, lens L_1 is a varifocal lens or otherwise has a variable field of view (such as a “zoom” functionality). In this case, it is desired that the aremac also have a variable field of view. In particular, field-of-view control mechanisms (whether mechanical, electronic, or hybrid) are linked in such a way that the aremac image magnification is reduced as the camera magnification is increased. Through this appropriate linkage, any increase in magnification by the camera is negated exactly by decreasing the apparent size of the viewfinder image.

The operation of the aremac focus and zoom tracking is shown in figure 3.

Stereo effects are well known in virtual reality systems (Ellis, Bucher, & Menges, 1995) wherein two information channels are often found to create a better sense of realism. Likewise, in stereo embodiments of the devices that we built, there were two cameras or measurement systems and two aremacs that each regenerated the respective outputs of the camera or measurement systems.

The apparatus is usually concealed in dark sunglasses that wholly or partially obstruct vision except for what the apparatus allows to pass through.

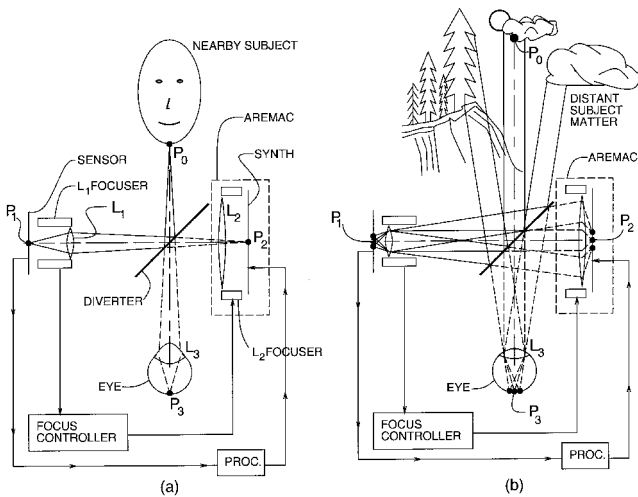


Figure 2. Focus tracking aremac: (a) with a Nearby subject, a point P_0 that would otherwise be imaged at P_3 in the eye of a user of the device is instead imaged to point P_1 on the image sensor, because the diverter diverts eyeward bound light to lens L_1 . When subject matter is nearby, the L_1 focuser moves objective lens L_1 out away from the sensor automatically, as an automatic focus camera would. A signal from the L_1 focuser directs the L_2 focuser, by way of the focus controller, to move lens L_2 outward away from the light synthesizer. In many of the embodiments of the system that we built, the functionality of the focus controller was implemented within a wearable computer that also processed the images. We designed and built a focus controller card as a printed circuit board for use with the Industry Standards Association (ISA) bus standard that was popular at the time of our original design. We also designed and built a PC104 version of the focus controller board. Our focus controller printed circuit layout (available in PCB format) is released under GNU GPL, and is downloadable from http://wearcam.org/eyetap_focus_controller, along with FIFO implementation of serial select servo controller. (Later, we built some other embodiments that use a serial port of the wearable computer to drive a separate focus controller module.) The focus controller drives up to four servos to adjust the position of lenses L_2 of a stereo rig, as well as the two lenses L_1 . In other embodiments, we used automatic-focus cameras and derived the signal controlling the servo position for lens L_2 by extracting the similar servo positioning signal from the focus adjustment of the autofocus camera. At the same time, an image from the sensor is directed through an image processor (PROC) into the light synthesizer (SYNTH). Point P_2 of the display element is responsive to point P_1 of the sensor. Likewise, other points on the light synthesizer are each responsive to corresponding points on the sensor, so that the synthesizer produces a complete image for viewing through lens L_2 by the eye, after reflection off the back side of the diverter. The position of L_2 is such that the eye's own lens L_3 will focus to the same distance as it would have focused in the absence of the entire device. (b) With distant subject matter, rays of parallel light are diverted toward the sensor where lens L_1 automatically

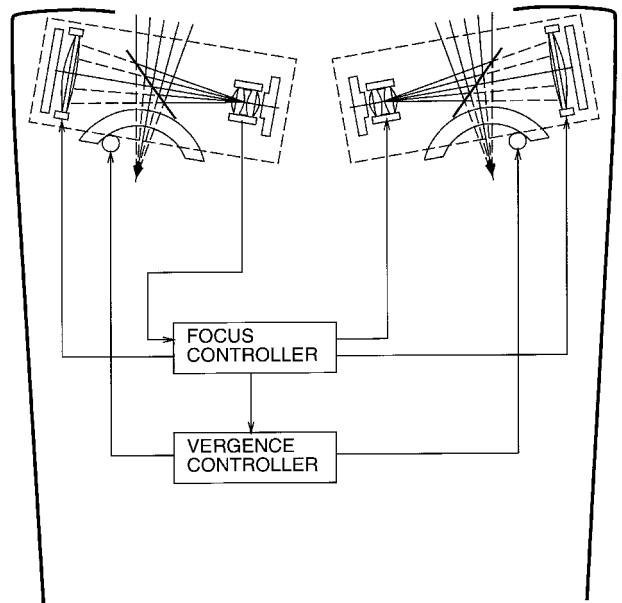


Figure 3. Focus of right camera and both aremacs (as well as vergence) controlled by the autofocus camera on the left side. In a two-eyed system, it is preferable that both cameras and both aremacs focus to the same distance. Therefore, one of the cameras is a focus master and the other camera is a focus slave. Alternatively, a focus combiner is used to average the focus distance of both cameras and then make the two cameras focus at equal distance. The two aremacs, as well as the vergence of both systems, also track this same depth plane as defined by camera autofocus.

2.2 Importance of the Orthospacial Criterion

Registration, which is important to augmented-reality systems (You, Neumann, & Azuma, 1999; Azuma, 2001; Behringer, 1998), is also important in mediated reality.

Of the three registration criteria (focus, orthospaciality, orthotonicity), an important one is the orthospacial criteria for mitigation of any resulting mismatch be-

retracts to focus these rays at point P_1 . When lens L_1 retracts, so does lens L_2 , and the light synthesizer ends up generating parallel rays of light that bounce off the backside of the diverter. These parallel rays of light enter the eye and cause its own lens L_3 to relax to infinity, as it would have in the absence of the entire device.

tween viewfinder image and the real world that would otherwise create an unnatural mapping. Indeed, anyone who has walked around holding a small camcorder up to his or her eye for several hours a day will obtain an understanding of the ill psychophysical effects that result.

The diverter system in EyeTap allows the center of projection of the camera to optically coincide with the center of projection of the eye. This placement of the camera makes EyeTap different from other head-mounted camera systems that place the camera only “near to” the eye’s center of projection. We will now discuss how camera placement of the EyeTap allows it to work without parallax in a variety of situations, without the limitations experienced by head-mounted camera systems.

It is easy to imagine a camera connected to a television screen and carefully arranged in such a way that, when viewed from a particular viewpoint, the television screen displays exactly what is blocked by the screen, so that an illusory transparency results. This illusory transparency would hold only so long as the television is viewed from this particular viewpoint. Moreover, it is easy to imagine a portable miniature device that accomplishes this situation, especially given the proliferation of consumer camcorder systems (such as portable cameras with built-in displays). We could attempt to achieve this condition with a handheld camcorder, perhaps miniaturized to fit into a helmet-mounted apparatus, but it is impossible to align the images exactly with what would appear in the absence of the apparatus. We can better understand this problem by referring to figure 4.

Figure 5 shows, in detail, how, in figure 4, we imagine that the objective lens of the camera, placed directly in front of the eye, is much larger than it really is, so that it captures all eyeward bound rays of light, for which we can imagine that it processes these rays in a collinear fashion. However, this reasoning is pure fiction and breaks down as soon as we consider a scene that has some depth of field.

Thus, the setup of figures 4 and 5 works for only a particular viewpoint and for subject matter in a particular depth plane. Although the same kind of system could obviously be miniaturized and concealed in ordinary-appearing sunglasses, in which case the limitation

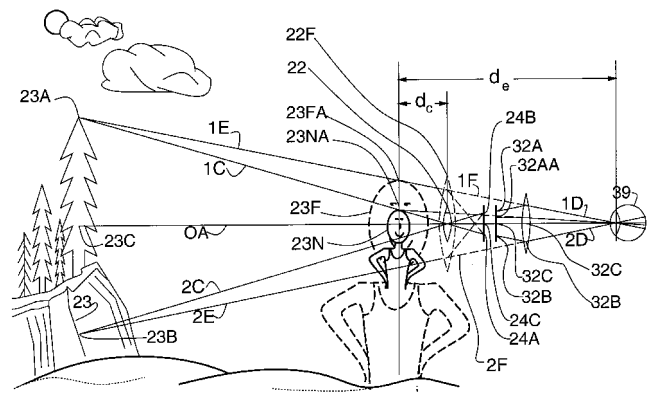


Figure 4. The small lens (22) shown in solid lines collects a cone of light bounded by rays 1C and 2C. Consider, for example, eyeward-bound ray of light 1E, which may be imagined to be collected by a large fictional lens 22F (when in fact ray 1C is captured by the actual lens 22), and focused to point 24A. The sensor element collecting light at point 24A is displayed as point 32A on the camcorder viewfinder, which is then viewed by a magnifying lens and emerges as ray 1D into the eye (39). It should be noted that the top of the nearby subject matter (23N) also images to point 24A and is displayed at point 32A, emerging as ray 1D as well. Thus, nearby subject matter 23N will appear as shown in the dotted line denoted 23F, with the top point appearing as 23FA even though the actual point should appear as 23NA (that is, it would appear as point 23NA in the absence of the apparatus). Thus, a camcorder cannot properly function as a true EyeTap device.

to a particular viewpoint is not a problem (because the sunglasses could be anchored to a fixed viewpoint with respect to at least one eye of a user), the other important limitation—that such systems work for only subject matter in the same depth plane—remains.

This problem exists whether the camera is right in front of the display or off to one side. Some real-world examples, having the camera to the left of the display, are shown in figure 6. In these setups, subject matter moved closer to the apparatus will show as being not properly aligned. Consider a person standing right in front of the camera but not in front of the TV in figure 6. Clearly, this person will not be behind the television but yet will appear on the television. Likewise, a person standing directly behind the television will not necessarily be seen by the camera, which is located to the left of

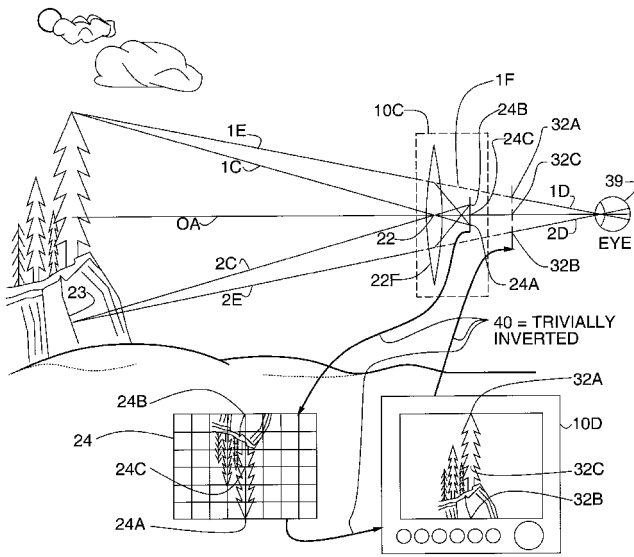


Figure 5. Suppose the camera portion of the camcorder, denoted by reference numeral 10C, were fitted with a very large objective lens (22F). This lens would collect eyeward-bound rays of light 1E and 2E. It would also collect rays of light coming toward the center of projection of lens 22. Rays of light coming toward this camera center of projection are denoted 1C and 2C. Lens 22 converges rays 1E and 1C to point 24A on the camera sensor element. Likewise, rays of light 2C and 2E are focused to point 24B. Ordinarily, the image (denoted by reference numeral 24) is upside down in a camera, but cameras and displays are designed so that, when the signal from a camera is fed to a display (such as a TV set), it shows rightside up. Thus, the image appears with point 32A of the display creating rays of light such as the one denoted 1D. Ray 1D is collinear with eyeward-bound ray 1E. Ray 1D is responsive to, and collinear with, ray 1E that would have entered the eye in the absence of the apparatus. Likewise, by similar reasoning, ray 2D is responsive to, and collinear with, eyeward-bound ray 2E. It should be noted, however, that the large lens (22F) is just an element of fiction. Thus, lens 22F is a fictional lens, because a true lens should be represented by its center of projection; that is, its behavior should not change, other than by depth of focus, diffraction, and amount of light passed, when its iris is opened or closed. Therefore, we could replace lens 22F with a pinhole lens and simply imagine lens 22 to have captured rays 1E and 2E, when it actually, in fact, captures only rays 1C and 2C.

the television. Thus, subject matter that exists at a variety of different depths and is not confined to a plane may be impossible to align in all areas with its image on the screen.

3 VideoOrbits Head Tracking and Motion Estimation for EyeTap Reality Mediation

Because the device absorbs, quantifies, processes, and reconstructs light passing through it, there are extensive applications in creating a mediated version of reality. The computer-generated information or virtual light as seen through the display must be properly registered and aligned with the real-world objects within the user's field of view. To achieve this, a method of camera-based head tracking is now described.

3.1 Why Camera-Based Head Tracking?

A goal of personal imaging (Mann, 1997b) is to facilitate the use of the Reality Mediator in ordinary everyday situations, not just on a factory assembly line "workcell" or other restricted space. Thus, it is desired that the apparatus have a head tracker that need not rely on any special apparatus being installed in the environment.

Therefore, we need a new method of head tracking based on the use of the camera capability of the apparatus (Mann, 1997b) and on the VideoOrbits algorithm (Mann & Picard, 1995).

3.2 Algebraic Projective Geometry

The VideoOrbits algorithm performs head tracking, visually, based on a natural environment, and it works without the need for object recognition. Instead, it is based on algebraic projective geometry and a direct featureless means of estimating the change in spatial coordinates of successive frames of EyeTap video arising from movement of the wearer's head, as illustrated in figure 7. This change in spatial coordinates is characterized by eight parameters of an "exact" projective (homographic) coordinate transformation that registers pairs of images or scene content. These eight parameters are "exact" for two cases of static scenes: (i) images taken from the same location of an arbitrary 3-D scene, with a camera that is free to pan, tilt, rotate about its optical axis, and zoom (such as when the user stands still and moves their head); or (ii) images of a flat scene

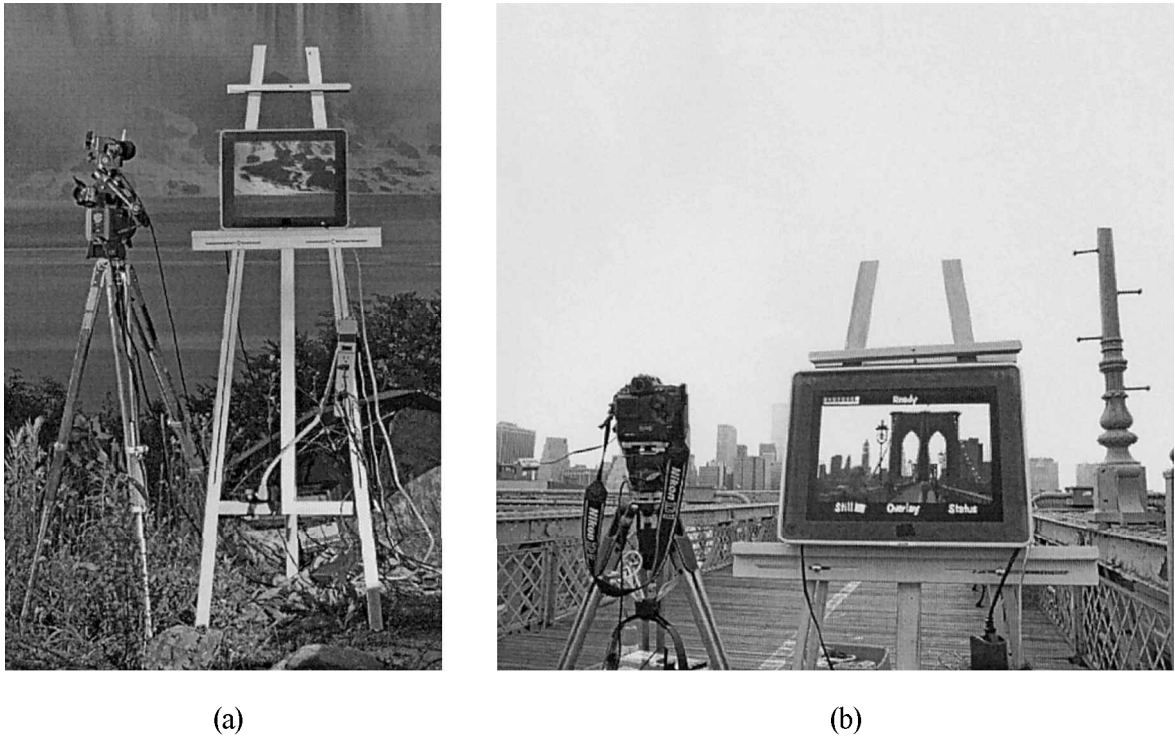


Figure 6. Illusory transparency: Examples of a camera supplying a television with an image of subject matter blocked by the television. (a) A television camera on a tripod at left supplies an Apple "Studio" television display with an image of the lower portion of Niagara Falls blocked by the television display (resting on an easel to the right of the camera tripod). The camera and display were carefully arranged, along with a second camera to capture this picture of the apparatus. Only when viewed from the special location of the second camera does the illusion of transparency exist. (b) Various cameras with television outputs were set up on the walkway but none of them can re-create the subject matter behind the television display in a manner that creates a perfect illusion of transparency, because the subject matter does not exist in one single depth plane. There exists no choice of camera orientation, zoom setting, and viewer location that creates an exact illusion of transparency for the portion of the Brooklyn Bridge blocked by the television screen. Notice how the railings don't quite line up correctly because they vary in depth with respect to the first support tower of the bridge.

taken from arbitrary locations (such as when it is desired to track a planar patch, viewed by a user who is free to move about). Thus, it is well suited for tracking planar patches with arbitrary view motion, a situation that commonly arises, for instance, when a sign or billboard is to be tracked.

It is stressed here that the algorithm presented is used to track image motion arising from arbitrary relative motion of the user's head with respect to rigid planar patches. Initial placement of computer-generated information (for instance, the four corners of the rigid planar rectangular patch) is assumed to have been completed

by another method. Once placed, however, no further model or knowledge of the scene is required to track the location of computer-generated information.

The featureless projective approach generalizes inter-frame camera motion estimation methods that have previously used an affine model (which lacks the degrees of freedom to "exactly" characterize such phenomena as camera pan and tilt) and/or that have relied upon finding points of correspondence between the image frames. The featureless projective approach, which operates directly on the image pixels, is shown to be superior in accuracy and ability to enhance resolution. The pro-

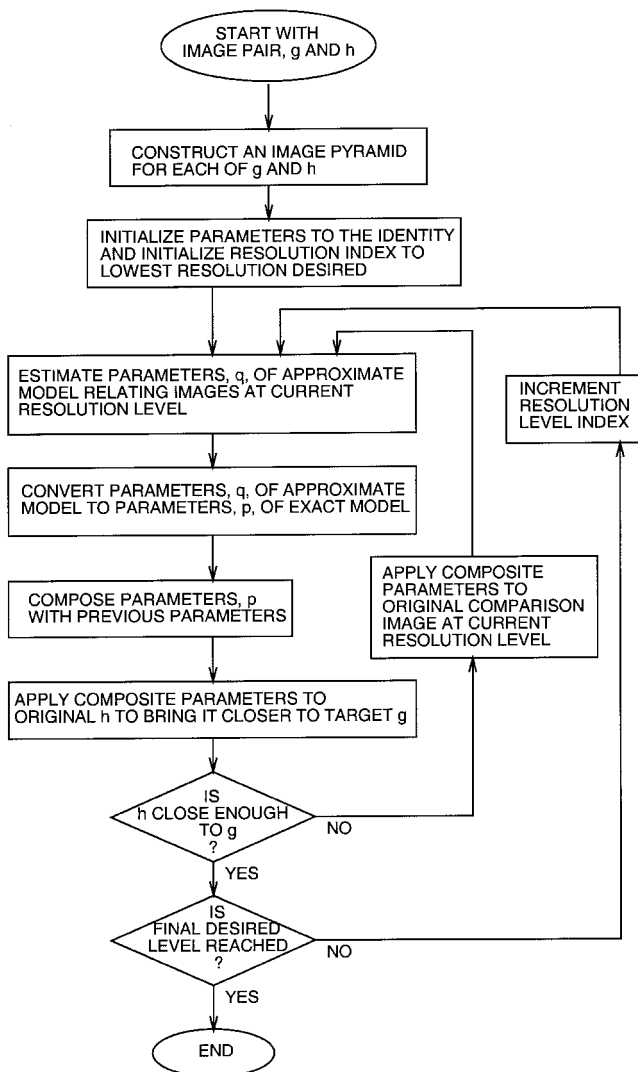


Figure 7. The VideoOrbits head-tracking algorithm: The new head-tracking algorithm requires no special devices installed in the environment. The camera in the personal imaging system simply tracks itself based on its view of objects in the environment. The algorithm is based on algebraic projective geometry and provides an estimate of the true projective coordinate transformation, which, for successive image pairs is composed using the projective group (Mann & Picard, 1995). Successive pairs of images may be estimated in the neighborhood of the identity coordinate transformation of the group, whereas absolute head tracking is done using the exact group by relating the approximate parameters q to the exact parameters p in the innermost loop of the process. The algorithm typically runs at five to ten frames per second on a general purpose computer, but the simple structure of the algorithm makes it easy to implement in hardware for the higher frame rates needed for full-motion video.

posed methods work well on image data collected from both good-quality and poor-quality video under a wide variety of conditions (sunny, cloudy, day, night). These fully automatic methods are also shown to be robust to deviations from the assumptions of static scene and no parallax. The primary application here is in filtering out or replacing subject matter appearing on flat surfaces within a scene (for example, rigid planar patches such as advertising billboards).

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, Ozkan, and Sezan (1992) have applied this assumption to high-resolution image reconstruction. Although translation is the least constraining and simplest to implement of the seven coordinate transformations in table 1, it is poor at handling large changes due to camera zoom, rotation, pan, and tilt.

Zheng and Chellappa (1993) considered the image registration problem using a subset of the affine model: translation, rotation and scale. Other researchers (Irani & Peleg, 1991; Teodosio & Bender, 1993) have assumed affine motion (six parameters) between frames. Behringer (1998) considered features of a silhouette. For the assumptions of static scene and no parallax, the affine model exactly describes rotation about the optical axis of the camera, zoom of the camera, and pure shear, which the camera does not do, except in the limit as the lens focal length approaches infinity. The affine model cannot capture camera pan and tilt and therefore cannot properly express the “keystoning” and “chirping” we see in the real world. (*Chirping* refers to the effect of increasing or decreasing spatial frequency with respect to spatial location, as illustrated in figure 8.)

This chirping phenomenon is implicit in the proposed system, whether or not there is periodicity in the subject matter. The only requirement is that there be some distinct texture upon a flat surface in the scene.

3.3 Video Orbits

Tsai and Huang (1981) pointed out that the elements of the projective group give the true camera motions with respect to a planar surface. They explored the

Table I. Image Coordinate Transformations Discussed in this Paper

Model	Coordinate transformation from \mathbf{x} to \mathbf{x}'	Parameters
Translation	$\mathbf{x}' = \mathbf{x} + \mathbf{b}$	$\mathbf{b} \in \mathbb{R}^2$
Affine	$\mathbf{x}' = \mathbf{A}\mathbf{x} + \mathbf{b}$	$\mathbf{A} \in \mathbb{R}^{2 \times 2}, \mathbf{b} \in \mathbb{R}^2$
Bilinear	$x' = q_{x'xy}xy + q_{x'xx}x + q_{x'y}y + q_{x'}$ $y' = q_{y'xy}xy + q_{y'xx}x + q_{y'y}y + q_{y'}$	$q_* \in \mathbb{R}$
Projective	$\mathbf{x}' = \frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^T\mathbf{x} + 1}$	$\mathbf{A} \in \mathbb{R}^{2 \times 2}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$
Relative-projective	$\mathbf{x}' = \frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^T\mathbf{x} + 1} + \mathbf{x}$	$\mathbf{A} \in \mathbb{R}^{2 \times 2}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$
Pseudo-perspective	$x' = q_{x'xx}x + q_{x'y}y + q_{x'} + q_{\alpha}x^2 + q_{\beta}xy$ $y' = q_{y'xx}x + q_{y'y}y + q_{y'} + q_{\alpha}xy + q_{\beta}y^2$	$q_* \in \mathbb{R}$
Biquadratic	$x' = q_{x'x^2}x^2 + q_{x'xy}xy + q_{x'y^2}y^2 + q_{x'x}x + q_{x'y}y + q_{x'}$ $y' = q_{y'x^2}x^2 + q_{y'xy}xy + q_{y'y^2}y^2 + q_{y'x}x + q_{y'y}y + q_{y'}$	$q_* \in \mathbb{R}$

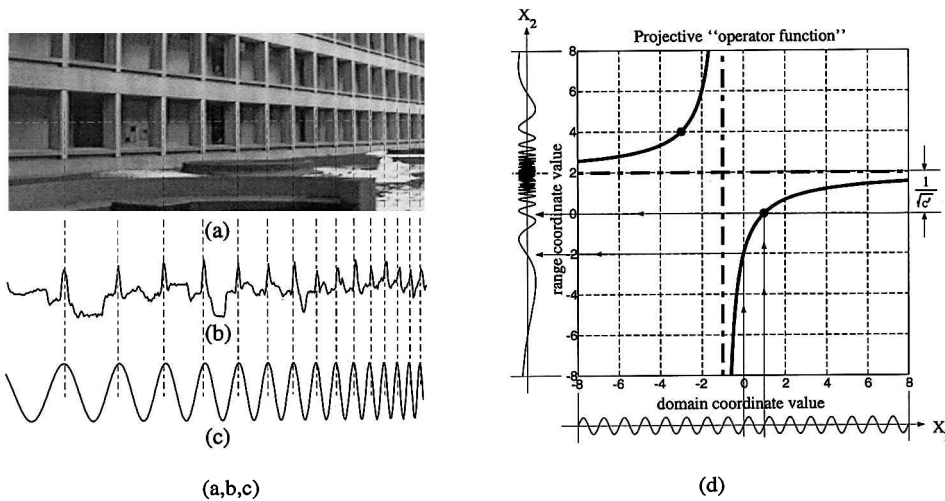


Figure 8. The projective chirping phenomenon. (a) A real-world object that exhibits periodicity generates a projection (image) with “chirping” (“periodicity in perspective”). (b) Center raster of image. (c) Best-fit projective chirp of form $\sin(2\pi((ax + b)/(cx + 1)))$. (d) Graphical depiction of exemplar 1-D projective coordinate transformation of $\sin(2\pi x_1)$ into a “projective chirp” function, $\sin(2\pi x_2) = \sin(2\pi((2x_1 - 2)/(x_1 + 1)))$. The range coordinate as a function of the domain coordinate forms a rectangular hyperbola with asymptotes shifted to center at the vanishing point $x_1 = -1/c = -1$ and exploding point, $x_2 = a/c = 2$, and with chirpiness $c' = c^2/(bc - a) = -1/4$

group structure associated with images of a 3-D rigid planar patch, as well as the associated Lie algebra, although they assume that the correspondence problem

has been solved. The solution presented in this paper (which does not require prior solution of correspondence) also relies on projective group theory.

3.3.1 Projective Flow—A New Technique for Tracking a Rigid Planar Patch. A method for tracking a rigid planar patch is now presented. Consider first one-dimensional systems because they are easier to explain and understand. For a 1-D affine coordinate transformation, the graph of the range coordinate as a function of the domain coordinate is a straight line; for the projective coordinate transformation, the graph of the range coordinate as a function of the domain coordinate is a rectangular hyperbola (figure 8(d)).

Whether or not there is periodicity in the scene, the method still works, in the sense that it is based on the projective flow across the texture or pattern, at all various spatial frequency components of a rigid planar patch. The method is called *projective-flow* (*p-flow*), which we will now describe in 1-D.

We begin with the well-known Horn and Schunk brightness change constraint equation (Horn & Schunk, 1981):

$$u_f E_x + E_t \approx 0, \quad (1)$$

where E_x and E_t are the spatial and temporal derivatives respectively of the image $E(x)$, and u_f is the optical flow velocity, assuming pure translation. Typically, we determine u_m which minimizes the error equation (1) as

$$\varepsilon_{flow} = \sum_x (u_m E_x + E_t)^2 \quad (2)$$

Projective-flow (*p-flow*), and arises from substitution of $u_m = ((ax + b)/(cx + 1)) - x$ in place of u_f in equation (1).

A judicious weighting by $(cx + 1)$ simplifies the calculation, giving

$$\varepsilon_w = \sum (axE_x + bE_x + c(xE_t - x^2E_x) + E_t - xE_x)^2. \quad (3)$$

Differentiating and setting the derivative to zero (the subscript w denotes weighting has taken place) results in a linear system of equations for the parameters, which can be written compactly as

$$\left(\sum \phi_w \phi_w^T \right) [a, b, c]^T = \sum (xE_x - E_t) \phi_w \quad (4)$$

where the regressor is $\phi_w = [xE_x, E_x, xE_t - x^2E_x]^T$.

The notation and derivations used in this paper are as

described by Mann (1998, p. 2139). The reader is invited to refer to that work for a more in-depth treatment of the matter.

3.3.2 The Unweighted Projectivity Estimator.

If we do not wish to apply the ad hoc weighting scheme, we may still estimate the parameters of projectivity in a simple manner still based on solving a linear system of equations. To do this, we write the Taylor series of u_m

$$u_m + x = b + (a - bc)x + (bc - a)cx^2 + (a - bc)c^2x^3 + \dots \quad (5)$$

and use only the first three terms, obtaining enough degrees of freedom to account for the three parameters being estimated. Letting $\varepsilon = \sum ((b + (a - bc - 1)x + (bc - a)cx^2)E_x + E_t)^2$, $\mathbf{q}_2 = (bc - a)c$, $\mathbf{q}_1 = a - bc - 1$, and $\mathbf{q}_0 = b$, and differentiating with respect to each of the three parameters of \mathbf{q} , setting the derivatives equal to zero, and verifying with the second derivatives gives the linear system of equations for unweighted projective flow:

$$\begin{bmatrix} \sum x^4 E_x^2 & \sum x^3 E_x^2 & \sum x^2 E_x^2 \\ \sum x^3 E_x^2 & \sum x^2 E_x^2 & \sum x E_x^2 \\ \sum x^2 E_x^2 & \sum x E_x^2 & \sum E_x^2 \end{bmatrix} \begin{bmatrix} q_2 \\ q_1 \\ q_0 \end{bmatrix} = - \begin{bmatrix} \sum x^2 E_x E_t \\ \sum x E_x E_t \\ \sum E_x E_t \end{bmatrix} \quad (6)$$

3.4 Planetracker in 2-D

We now discuss the 2-D formulation. We begin again with the brightness constancy constraint equation, this time for 2-D images (Horn & Schunk, 1981), which gives the flow velocity components in both the x and y directions:

$$\mathbf{u}_f^T \mathbf{E}_x + E_t \approx 0 \quad (7)$$

As is well known, the optical flow field in 2-D is underconstrained.¹ The model of pure translation at every point has two parameters, but there is only one equation (7) to solve, thus it is common practice to

1. Optical flow in 1-D did not suffer from this problem.

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 (for example, the entire patch of subject matter to be filtered out, such as a billboard or sign).

Our task is not to deal with the 2-D translational flow, but with the 2-D projective flow, estimating the eight parameters in the coordinate transformation:

$$\mathbf{x}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{\mathbf{A}[x, y]^T + \mathbf{b}}{\mathbf{c}^T[x, y]^T + 1} = \frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^T\mathbf{x} + 1} \quad (8)$$

The desired eight scalar parameters are denoted by $\mathbf{p} = [\mathbf{A}, \mathbf{b}; \mathbf{c}, 1]$, $\mathbf{A} \in \mathbf{R}^{2 \times 2}$, $\mathbf{b} \in \mathbf{R}^{2 \times 1}$, and $\mathbf{c} \in \mathbf{R}^{2 \times 1}$.

We have, in the 2-D case:

$$\varepsilon_{\text{flow}} = \sum (\mathbf{u}_m^T \mathbf{E}_x + E_t)^2 = \sum \left(\left(\frac{\mathbf{A}\mathbf{x} + \mathbf{b}}{\mathbf{c}^T\mathbf{x} + 1} - \mathbf{x} \right)^T \mathbf{E}_x + E_t \right)^2, \quad (9)$$

where the sum can be weighted as it was in the 1-D case:

$$\varepsilon_w = \sum \left((\mathbf{A}\mathbf{x} + \mathbf{b} - (\mathbf{c}^T\mathbf{x} + 1)\mathbf{x})^T \mathbf{E}_x + (\mathbf{c}^T\mathbf{x} + 1)E_t \right)^2. \quad (10)$$

Differentiating with respect to the free parameters \mathbf{A} , \mathbf{b} , and \mathbf{c} , and setting the result to zero gives a linear solution:

$$\left(\sum \phi \phi^T \right) [a_{11}, a_{12}, b_1, a_{21}, a_{22}, b_2, c_1, c_2]^T = \sum (\mathbf{x}^T \mathbf{E}_x - E_t) \phi \quad (11)$$

where

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

For a more in-depth treatment of projective flow, the reader is invited to refer to Mann (1998).

3.5 Unweighted Projective Flows

As with the 1-D images, we make similar assumptions in expanding equation (8) in its own Taylor series, analogous to equation (5). By appropriately constraining the twelve parameters of the biquadratic model, we obtain a variety of eight-parameter approximate models. In estimating the “exact unweighted” projective group parameters, one of these approximate models is used in an intermediate step.²

The Taylor series for the bilinear case gives

$$\begin{aligned} u_m + x &= q_{x'xy}xy + (q_{x'x} + 1)x + q_{x'y}y + q_{x'} \\ v_m + y &= q_{y'xy}xy + q_{y'x}x + (q_{y'y} + 1)y + q_{y'} \end{aligned} \quad (12)$$

Incorporating these into the flow criteria yields a simple set of eight linear equations in eight unknowns:

$$\left(\sum_{x,y} (\phi(x, y) \phi^T(x, y)) \right) \mathbf{q} = - \sum_{x,y} E_t \phi(x, y) \quad (13)$$

where $\phi^T = [E_x(xy, x, y, 1), E_y(xy, x, y, 1)]$.

For the relative-projective model, ϕ is given by

$$\phi^T = [E_x(x, y, 1), E_y(x, y, 1), E_t(x, y)], \quad (14)$$

and, for the pseudo-perspective model, ϕ is given by

$$\begin{aligned} \phi^T &= [E_x(x, y, 1), E_y(x, y, 1), \\ & (x^2E_x + xyE_y, xyE_x + y^2E_y)]. \end{aligned} \quad (15)$$

3.5.1 Four-Point Method for Relating Approximate Model to Exact Model. Any of the preceding approximations, after being related to the exact projective model, tend to behave well in the neighborhood of the identity, $\mathbf{A} = \mathbf{I}$, $\mathbf{b} = \mathbf{0}$, $\mathbf{c} = \mathbf{0}$. In 1-D, the model Taylor series about the identity was explicitly expanded; here, although this is not done explicitly, it is assumed that the terms of the Taylor series of the model correspond to those taken about the identity. In the 1-D case, we solve the three linear equations in three

2. Use of an approximate model that doesn't capture chirping or preserve straight lines can still lead to the true projective parameters as long as the model captures at least eight meaningful degrees of freedom.

unknowns to estimate the parameters of the approximate motion model, and then relate the terms in this Taylor series to the exact parameters, a , b , and c (which involves solving another set of three equations in three unknowns, the second set being nonlinear, although very easy to solve).

In the extension to 2-D, the estimate step is straightforward, but the relate step is more difficult, because we now have eight nonlinear equations in eight unknowns, relating the terms in the Taylor series of the approximate model to the desired exact model parameters. Instead of solving these equations directly, a simple procedure is used for relating the parameters of the approximate model to those of the exact model, which is called the “four-point method”:

1. Select four ordered pairs (for example, the four corners of the bounding box containing the region under analysis, or the four corners of the image if the whole image is under analysis). Here, suppose, for simplicity, that these points are the corners of the unit square: $\mathbf{s} = [s_1, s_2, s_3, s_4] = [(0, 0)^T, (0, 1)^T, (1, 0)^T, (1, 1)^T]$.
2. Apply the coordinate transformation using the Taylor series for the approximate model (such as equation (12)) to these points: $\mathbf{r} = \mathbf{u}_m(\mathbf{s})$.
3. Finally, the correspondences between \mathbf{r} and \mathbf{s} are treated just like features. This results in four easy-to-solve equations:

$$\begin{bmatrix} x'_k \\ y'_k \end{bmatrix} = \begin{bmatrix} x_k, y_k, 1, 0, 0, 0, -x_k x'_k, -y_k x'_k \\ 0, 0, 0, x_k, y_k, 1, -x_k y'_k, -y_k y'_k \end{bmatrix} \quad (16)$$

$$[a_{x'_k}, a_{y'_k}, b_{x'_k}, a_{y'_k}, a_{y'_k}, b_{y'_k}, c_x, c_y]^T$$

where $1 \leq k \leq 4$. This results in the exact eight parameters, \mathbf{p} .

We remind the reader that the four corners are not feature correspondences as used in the feature-based methods, but, rather, are used so that the two featureless models (approximate and exact) can be related to one another.

It is important to realize the full benefit of finding the exact parameters. Although the approximate model is sufficient for small deviations from the identity, it is not

adequate to describe large changes in perspective. However, if we use it to track small changes incrementally, and each time relate these small changes to the exact model (8), then we can accumulate these small changes using the law of composition afforded by the group structure. This is an especially favorable contribution of the group framework. For example, with a video sequence, we can accommodate very large accumulated changes in perspective in this manner. The problems with cumulative error can be eliminated, for the most part, by constantly propagating forward the true values, computing the residual using the approximate model, and each time relating this to the exact model to obtain a goodness-of-fit estimate.

3.5.2 Overview of the Algorithm for Unweighted Projective Flow. Frames from an image sequence are compared pairwise to test whether or not they lie in the same orbit:

1. A Gaussian pyramid of three or four levels is constructed for each frame in the sequence.
2. The parameters \mathbf{p} are estimated at the top of the pyramid, between the two lowest-resolution images of a frame pair, g and h , using the repetitive method depicted in figure 7.
3. The estimated \mathbf{p} is applied to the next-higher-resolution (finer) image in the pyramid, $\mathbf{p} \circ g$, to make the two images at that level of the pyramid nearly congruent before estimating the \mathbf{p} between them.
4. The process continues down the pyramid until the highest-resolution image in the pyramid is reached.

4 Reality Mediation in Variable-Gain Image Sequences

Until now, we have assumed fixed-gain image sequences. In practice, however, camera gain varies to compensate for varying quantity of light, by way of automatic gain control (AGC), automatic level control, or some similar form of automatic exposure.

In fact, almost all modern cameras incorporate some form of automatic exposure control. Moreover, next-

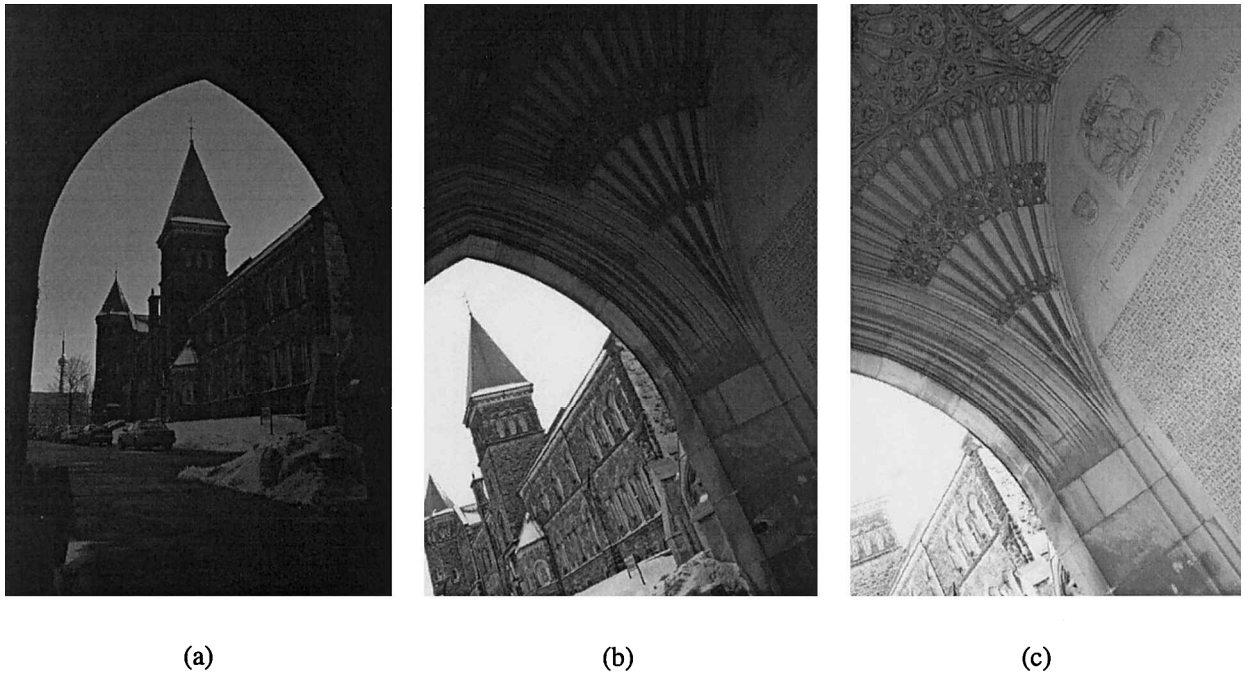


Figure 9. Automatic exposure is the cause of differently exposed pictures of the same (overlapping) subject matter, creating the need for comparametric imaging in intelligent vision systems (Mann, 2001a). (a) Looking from inside Hart House Soldier's Tower, out through an open doorway, when the sky is dominant in the picture, the exposure is automatically reduced, and the wearer of the apparatus can see the texture (such as clouds) in the sky. He can also see University College and the CN Tower to the left. (b) As he looks up and to the right to take in subject matter not so well illuminated, the exposure automatically increases somewhat. The wearer can no longer see detail in the sky, but new architectural details inside the doorway start to become visible. (c) As he looks further up and to the right, the dimly lit interior dominates the scene, and the exposure is automatically increased dramatically. He can no longer see any detail in the sky, and even the University College building, outside, is washed out (overexposed). However, the inscriptions on the wall (names of soldiers killed in the war) now become visible.

generation imaging systems such as the EyeTap eye-glasses also feature an automatic exposure control system to make possible a hands-free, gaze-activated wearable system that is operable without conscious thought or effort. Indeed, the human eye itself incorporates many features akin to the automatic exposure or AGC of modern cameras.

Figure 9 illustrates how the Reality Mediator (or nearly any camera for that matter) takes in a typical scene.

As the wearer looks straight ahead, he sees mostly sky, and the exposure is quite small. Looking to the right at darker subject matter, the exposure is automatically increased. Because the differently exposed pictures depict

overlapping subject matter, we have (once the images are registered, in regions of overlap) differently exposed pictures of identical subject matter. In this example, we have three very differently exposed pictures depicting parts of the University College building and surroundings.

4.1 Variable-Gain Problem Formulation

Differently exposed images (such as individual frames of video) of the same subject matter are denoted as vectors: $f_0, f_1, \dots, f_i, \dots, f_{I-1}, \forall i, 0 \leq i < I$.

Each video frame is some unknown function, $f(\cdot)$, of

the actual quantity of light, $q(\mathbf{x})$ falling on the image sensor:

$$f_i = f\left(k_i q \left(\frac{\mathbf{A}_i \mathbf{x} + \mathbf{b}_i}{c_i \mathbf{x} + d_i} \right)\right), \quad (17)$$

where $\mathbf{x} = (x, y)$ denotes the spatial coordinates of the image, k_i is a single unknown scalar exposure constant, and parameters \mathbf{A}_i , \mathbf{b}_i , c_i and d_i denote the projective coordinate transformation between successive pairs of images.

For simplicity, this coordinate transformation is assumed to be able to be independently recovered (for example, using the methods of the previous section). Therefore, without loss of generality, images considered in this section will be taken as having the identity coordinate transformation, which corresponds to the special case of images differing only in exposure.

Without loss of generality, k_0 will be called the *reference exposure* and will be set to unity, and frame zero will be called the *reference frame*, so that $f_0 = f(q)$. Thus, we have

$$\frac{1}{k_i} f^{-1}(f_i) = f^{-1}(f_0), \quad \forall i, 0 < i < I. \quad (18)$$

The existence of an inverse for f follows from a semimonotonicity assumption. Semimonotonicity follows from the fact that we expect pixel values to either increase or stay the same with increasing quantity of illumination, q .

Photographic film is traditionally characterized by the so-called ‘‘density versus log exposure’’ characteristic curve (Wyckoff, 1961, 1962). 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:

$$\log(f^{-1}(f_1(\mathbf{x}))) = Q = \log(f^{-1}(f_2(\mathbf{x}))) - K. \quad (19)$$

Because the logarithm function is also monotonic, the problem comes down to estimating the semimonotonic function $F(\cdot) = \log(f^{-1}(\cdot))$ and the scalar constant K . There are a variety of techniques for solving for F and K directly (Mann, 2000). In this paper, we choose to use a method involving comparometric equations.

4.1.1 Using Comparometric Equations. Variable-gain image sequences, f_b , are created by the response, f , of the imaging device to light, q . Each of these images provides us with an estimate of f differing only by exposure, k . Pairs of images can be compared by plotting $(f(q), f(kq))$, and the resulting relationship can be expressed as the monotonic function $g(f(q)) = f(kq)$ not involving q . Equations of this form are called *comparometric equations* (Mann, 2000). Comparometric equations are a special case of a more general class of equations called *functional equations* (Aczél, 1966).

A comparometric equation that is particularly useful for mediated-reality applications will now be introduced, first by its solution (from which the comparometric equation itself will be derived). (It is generally easier to construct comparometric equations from their solutions than it is to solve comparometric equations.) The solution is

$$f(q) = \left(e^b q^a / (e^b q^a + 1) \right)^c, \quad (20)$$

which has only three parameters (of which only two are meaningful parameters because b is indeterminable and may be fixed to $b = 0$ without loss of generality). Equation (20) is useful because it describes the shape of the curve that characterizes the response of many cameras to light, $f(q)$, called the *response curve*. The constants a and c are specific to the camera.

This model accurately captures the essence of the so-called toe and shoulder regions of the response curve. In traditional photography, these regions are ignored; all that is of interest is the linear mid-portion of the density versus log exposure curve. This interest in only the midtones arises because, in traditional photography, areas outside this region are considered to be incorrectly exposed. However, in practice, in input images to the reality mediator, many of the objects we look at will be massively underexposed and overexposed because not everything in life is necessarily a well-composed picture. Therefore, these qualities of the model (20) are of great value in capturing the essence of these extreme exposures, in which exposure into both the toe and shoulder regions are often the norm rather than an aberration.

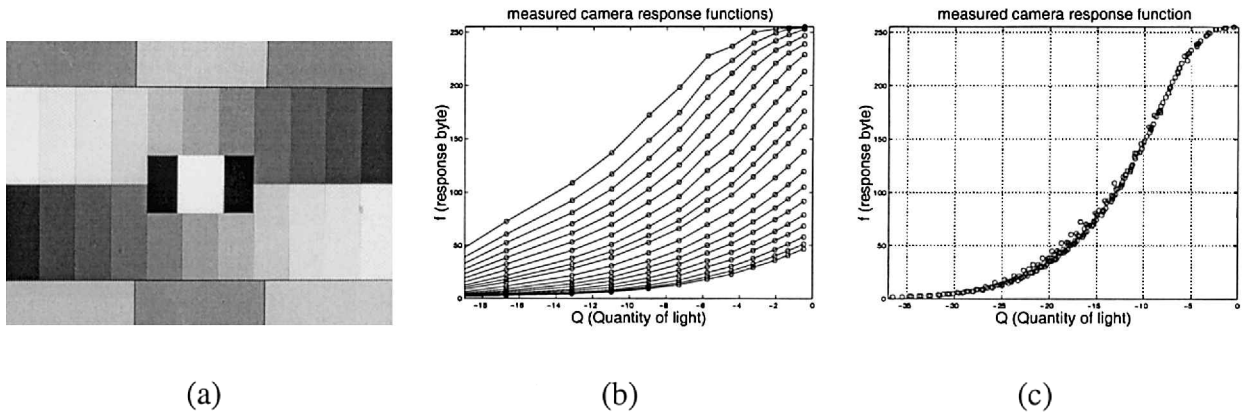


Figure 10. (a) One of nineteen differently exposed pictures of a test pattern. (b) Each of the nineteen exposures produced eleven ordered pairs in a plot of $f(Q)$ as a function of Q . (c) Shifting these nineteen plots left or right by the appropriate K_i , allowed them all to align to produce the ground-truth known-response function $f(Q)$.

Furthermore, equation (20) has the advantage of being bounded in normalized units between 0 and 1.

The comparometric equation of which the proposed photographic response function (20) is a solution, is given by

$$g(f) = \frac{\sqrt[c]{f}}{(\sqrt[c]{f} + e^{-aK})^c}, \quad (21)$$

where $K = \log_2(k_2/k_1)$ is the ratio of the two exposures.

To validate this model, we:

- estimate the parameters a , c , and k of $g(f(q))$ that best fit a plot $(f(q), f(kq))$ derived from differently exposed pictures (as, for example, shown in figure 9), and
- verify our estimate of f by using lab instruments.

Although there exist methods for automatically determining the a and c and relative gain k from pairs of differently exposed images by using comparometric equations, these methods are beyond the scope of this paper, and the reader is invited to refer to Mann (2000) for a full discussion of these methods and of comparometric equations.

A CamAlign-CGH test chart from DSC Laboratories, Toronto, Canada (Serial No. S009494), as shown in

figure 10(a), was used to verify the response function recovered using the method.

The individual bars were segmented automatically by differentiation to find the transition regions, and then robust statistics were used to determine an estimate of $f(q)$ for each of the eleven steps, as well as the black regions of the test pattern. Using the known reflectivity of each of these twelve regions, a set of twelve ordered pairs $(q, f(q))$ was determined for each of the nineteen exposures, as shown in figure 10(b). Shifting these results appropriately (by the K_i values) to line them up, gives the ground-truth, known response function, f , shown in figure 10(c).

Thus, equation (21) gives us a recipe for lightening or darkening an image in a way that looks natural and is also based on this proven theoretical framework. For instance, given a pair of images taken with a camera with a known response function (which is to say that the a and c are known for the camera), the relative gain between images is estimated, and either of the pair is lightened or darkened to bring it into the same exposure as the other image. Similarly, any computer-generated information in a mediated or augmented scene is brought into the appropriate exposure of the scene.

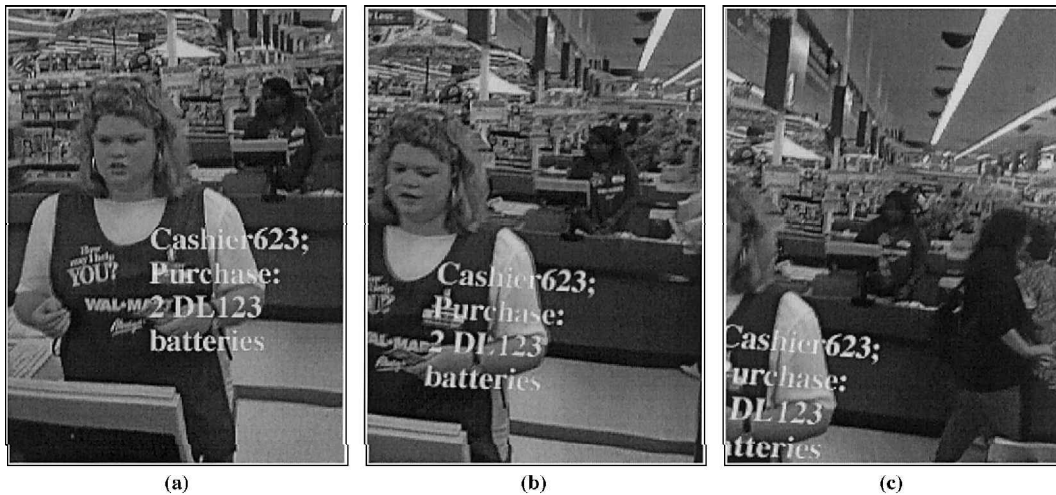


Figure 11. Mediated reality as a photographic/videographic memory prosthesis: (a) Wearable face recognizer with virtual “name tag” (and grocery list) appears to stay attached to the cashier (b), even when the cashier is no longer within the field of view of the tapped eye and transmitter (c).

4.2 Mediated Reality as a Form of Communication

The mathematical framework for mediated reality arose through the process of marking a reference frame (Mann & Picard, 1995) with text or simple graphics in which it was noted that, by calculating and matching homographies of the plane, an illusory rigid planar patch appeared to hover upon objects in the real world, giving rise to a form of computer-mediated collaboration (Mann, 1997b). Figure 11 shows images processed in real time by VideoOrbits.

4.3 Diminished Reality

Diminished reality deliberately removes parts of a real-world scene or replaces them with computer-generated information (Mann & Fung, 2001). For instance, deliberately diminished reality has application in construction. Klinker, Stricker and Reiners (2001) discuss a number of techniques for interpolating the pixels behind a diminished object: “Many construction projects require that existing structures be removed before new ones are built. Thus, just as important as augmenting reality is technology to diminish it” (p. 416).

Real-world “spam” (unwanted and unsolicited advertising) typically occurs on planar surfaces, such as billboards. The VideoOrbits algorithm presented here is well suited toward diminishing these unwanted and intrusive real-world planar objects.

Because the camera response function and exposure values can be computed automatically in a self-calibrating system, the computer-mediated reality can take form by combining the results estimation of the gain using the camera response function and estimation of the coordinate transformation between frames with the VideoOrbits methodology for computer-mediated reality.

Figure 12(a, b) shows a nice view of the Empire State Building spoiled by an offensive jeans advertisement (a billboard depicting a man pulling off a women’s clothes). The computer-mediated reality environment allows the billboard to be automatically replaced with a picture of vintage (original 1985) mediated-reality sunglasses. (See figure 12(c, d).) By removing the billboard, a deliberately diminished version of the scene is created. The information of the advertisement is now removed, and computer-generated information is inserted in its place, helping to avoid information overload.

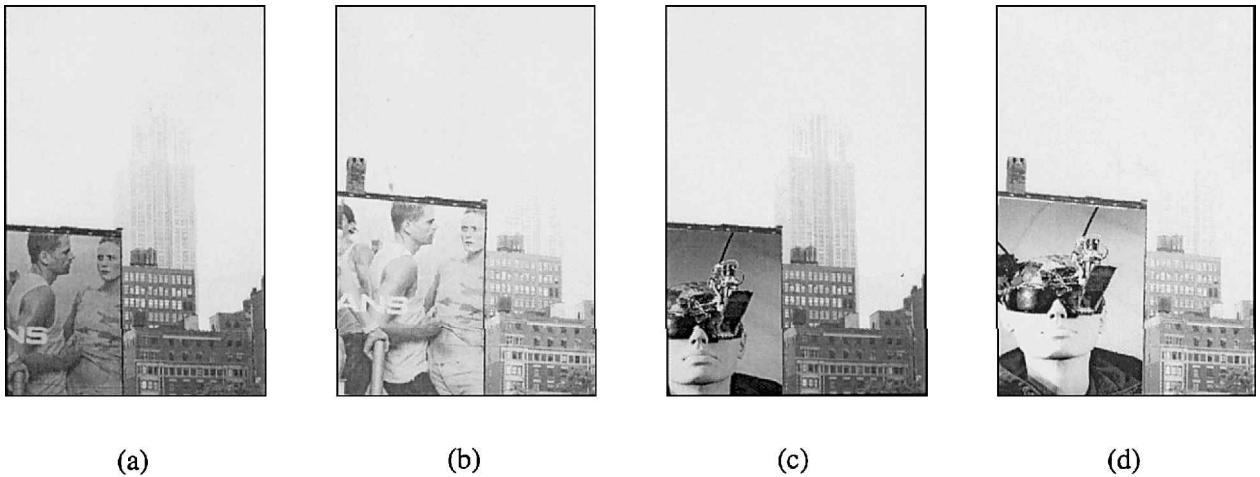


Figure 12. (a, b) Two frames from a video sequence in New York City, showing how a nice view of the Empire State Building is spoiled by an offensive jeans advertisement (a billboard depicting a man pulling off a women's clothes). Notice the effect of AGC being similar to that depicted in figure 9. (a) Because a large proportion of sky is included in the image, the overall exposure is quite low, so the image is darker. (b) Because the darker billboard begins to enter the center portion of view, the gain increases and the entire image is lighter. (c, d) Two frames from a video sequence in New York City, showing how the same visual reality can be diminished. Our ability to see the offensive advertisement is reduced. The diminished reality is then augmented with a view of the vintage 1985 smart sunglasses. Now, the resulting image sequence is an example of mediated reality. Notice how the exposure of the new matter, introduced into the visual field of view, tracks the exposure of the offensive advertising material originally present. The result is a visually pleasing, mediated-reality experience extending over a very wide dynamic range. (c) Because a large proportion of sky is included in the image, the overall exposure is quite low, and so the image is darker. The additional material inserted into the image is thus automatically made darker, comparably, to match. (d) Because the original image was lighter, the new matter introduced into the visual reality stream is also made lighter, comparably, to match.

5 Conclusion

Because wearable computers and EyeTap encapsulate users, the technologies mediate a user's experience with the world. Having designed, built, worn, and tested dozens of different embodiments of these devices for use in ordinary day-to-day life provided us with much in the way of valuable insight into the concepts of mediated reality. The resulting Reality Mediators alter the user's visual perception of their environment. The user's head motion is tracked by the VideoOrbits algorithm, and the camera gain is tracked using comparative equations. This allows for computer-generated information to be registered both spatially and tonally with the real world. An extension of the concept of mediated reality is the replacement of unwanted information, such as advertising, with computer-generated infor-

mation, giving rise to the notion of a deliberately diminished reality.

Acknowledgments

This work was funded in part by Xilinx and Altera.

References

- Aczél, J. (1966). *Lectures on functional equations and their applications* (Vol. 19). New York and London: Academic Press.
- Azuma, R. T. (2001). Augmented reality: Approaches and technical challenges. In W. Barfield & T. Caudell (Eds.), *Fundamentals of wearable computers and augmented reality* (pp. 27–63). New Jersey: Lawrence Erlbaum Press.

- Behringer, R. (1998). Improving the precision of registration for augmented reality in an outdoor scenario by visual horizon silhouette matching. *Proceedings of first IEEE workshop on augmented reality (IWAR98)*, 225–230.
- Caudell, T., & Mizell, D. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing processes. *Proc. Hawaii International Conf. on Systems Science*, 2, 659–669.
- Drascic, D., & Milgram, P. (1996). Perceptual issues in augmented reality. *SPIE Volume 2653: Stereoscopic Displays and Virtual Reality Systems III*, 123–134.
- Earnshaw, R. A., Gigante, M. A., & Jones, H. (1993). *Virtual reality systems*. London: Academic Press.
- Ellis, S. R., Bucher, U. J., & Menges, B. M. (1995). The relationship of binocular convergence and errors in judged distance to virtual objects. *Proceedings of the International Federation of Automatic Control*, 297–301.
- Feiner, S., MacIntyre, B., & Seligmann, D. (1993a). *Karma (knowledge-based augmented reality for maintenance assistance)*. Available online at: <http://www.cs.columbia.edu/graphics/projects/karma/karma.html>.
- . (1993b). Knowledge-based augmented reality. *Communications of the ACM*, 36(7), 52–62.
- Fuchs, H., Bajura, M., & Ohbuchi, R. *Teaming ultrasound data with virtual reality in obstetrics*. Available online at: <http://www.ncsa.uiuc.edu/Pubs/MetaCenter/SciHi93/1c.Highlights-BiologyC.html>.
- Klinker, G., Stricker, D., & Reinert, D. (2001). Augmented reality for exterior construction applications. In W. Barfield & T. Caudell (Eds.), *Fundamentals of wearable computers and augmented reality* (pp. 397–427). New Jersey: Lawrence Erlbaum Press.
- Horn, B., & Schunk, B. (1981). Determining optical flow. *Artificial Intelligence*, 17, 185–203.
- Irani, M., & Peleg, S. (1991). Improving resolution by image registration. *CVGIP*, 53, 231–239.
- Mann, S. (1997a). Humanistic intelligence. *Proceedings of Ars Electronica*, 217–231. (Available online at: <http://wearcam.org/ars/> and <http://www.aec.at/fleshfactor>.)
- . (1997b). Wearable computing: A first step toward personal imaging. *IEEE Computer*, 30(2), 25–32.
- . (1997c). An historical account of the ‘WearComp’ and ‘WearCam’ projects developed for ‘personal imaging.’ *International symposium on wearable computing*, 66–73.
- . (1998). Humanistic intelligence/humanistic computing: ‘Wearcomp’ as a new framework for intelligent signal processing. *Proceedings of the IEEE*, 86(11), 2123–2151.
- . (2000). Comparometric equations with practical applications in quantigraphic image processing. *IEEE Trans. Image Proc.*, 9(8), 1389–1406.
- . (2001a). *Intelligent image processing*. New York: John Wiley and Sons.
- . (2001b). Wearable computing: Toward humanistic intelligence. *IEEE Intelligent Systems*, 16(3), 10–15.
- Mann, S., & Fung, J. (2001). Videoorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. *International Symposium on Mixed Reality (ISMR2001)*, 48–55.
- Mann, S., & Picard, R. W. (1995). *Video orbits of the projective group; a simple approach to featureless estimation of parameters* (Tech. Rep. No. 338). Cambridge, MA: Massachusetts Institute of Technology. (Also appears in *IEEE Trans. Image Proc.*, (1997), 6(9), 1281–1295.)
- Sutherland, I. (1968). A head-mounted three dimensional display. *Proc. Fall Joint Computer Conference*, 757–764.
- Tekalp, A., Ozkan, M., & Sezan, M. (1992). High-resolution image reconstruction from lower-resolution image sequences and space-varying image restoration. *Proc. of the Int. Conf. on Acoust., Speech and Sig. Proc.*, III-169.
- Teodosio, L., & Bender, W. (1993). Salient video stills: Content and context preserved. *Proc. ACM Multimedia Conf.*, 39–46.
- Tsai, R. Y., & Huang, T. S. (1981). Estimating three-dimensional motion parameters of a rigid planar patch. *IEEE Trans. Acoust., Speech, and Sig. Proc.*, ASSP(29), 1147–1152.
- Wyckoff, C. W. (1961). *An experimental extended response film* (Tech. Rep. No. NO. B-321). Boston, Massachusetts: Edgerton, Germeshausen & Grier, Inc.
- Wyckoff, C. W. (1962, June–July). An experimental extended response film. *S.P.I.E. NEWSLETTER*, 16–20.
- You, S., Neumann, U., & Azuma, R. (1999). Hybrid inertial and vision tracking for augmented reality registration. *Proceedings of IEEE VR*, 260–267.
- Zheng, Q., & Chellappa, R. (1993). A computational vision approach to image registration. *IEEE Transactions Image Processing*, 2(3), 311–325.