Just Enough Reality: Comfortable 3-D Viewing via Microstereopsis

Mel Siegel and Shojiro Nagata

***Abstract*—We address human factors and technology issues for the design of stereoscopic display systems that are natural and comfortable to view. Our title "just enough reality" hints at the contrast between the popularly perceived requirements for strict "virtual reality" and the expert's pragmatic acceptance of "sufficient reality" to satisfy the human interface requirements of real-world applications. We first review how numerous perceptions and illusions of depth can be exploited to synergistically complement binocular stereopsis. Then we report the results of our experimental studies of stereoscopy with very small interocular separations and correspondingly small on-screen disparities, which we call "microstereopsis." We outline the implications of microstereopsis for the design of future stereoscopic camera and display systems, especially the possibility of achieving zoneless autostereoscopic displays. We describe a possible class of implementations based on a nonlambertian filter element, and a particular implementation that would use an electronically switched louver filter to realize it.**

***Index Terms*—3-D, autostereoscopic, microstereoscopic, stereoscopic, zoneless displays.**

I. Introduction

HOW TO BUILD a 3-D-stereoscopic camera and display system that reproduces exactly—at least geometrically—the retinal images of the original scene has been understood since the early days of photography [1], [2]. Almost as old and well known are a host of heuristic rules for deviating from this perfect geometry for the sake of mitigating its negative side effects [2], [3]. While it may seem strange that it is necessary to mitigate perfection, it is in fact necessary because the perceptual synthesis is perfect only in the domain of geometrical optics. To understand the 3-D world, the human eye-brain system integrates many cues at many cognitive levels. The perceptual conflict between the geometrical cues that are synthesized correctly and other cues that are not synthesized correctly causes physiological and psychological stresses that have recently come to be known as "virtual reality sickness," "simulator sickness," etc. [4], [5]. The most important and well-known conflict is between convergence and accommodation: the eyes converge to a virtual world-point in front of or behind the screen, but they focus on the screen *per se*. This discrepancy is physically and mentally uncomfortable. And despite the availability, and routine application, of nominally mitigating heuristics, we often notice that people using computer workstations equipped with 3-D-stereoscopic capability avoid using the stereo except when it becomes impossible for them to do the task at hand without it.

These observations stimulate us to seek a "kinder gentler stereo" paradigm: a natural and unobtrusive approach to 3-D-stereoscopic display of still and moving camera images and computer graphics that is free of physical and mental stress, as is naked-eye viewing of the real world. We seek an approach to stereo image-pair capture (and, for computer graphics, stereo image-pair generation) without cue conflicts, eyewear, and viewing zones, and with negligible "lock-in" time to perceive the virtual scene comfortably in full depth.

In Section II, we review depth perception and depth illusion modalities, emphasizing the synergy between binocular perspective parallax and other modalities. In Section III, we introduce the hypothesis of microstereopsis, describe our experiments toward demonstrating and quantifying it, and introduce the concept it stimulates for a new class of zoneless autostereoscopic[1] displays. In Section IV, we describe possible implementations of this idea. In Section V, we summarize our conclusions and suggest future work.

II. Depth Perception and Illusion

A. Depth Perception

Depth perception is stimulated by binocular perspective parallax between left and right eye views, and by motion parallax (monocularly and binocularly), even when the picture contains no recognizable objects. This was elegantly illustrated for perspective parallax by the classic static random-dot stereogram experiments of Julesz [6], and was recently reiterated for motion parallax in random dot video experiments by Nagata [7].

Binocular parallax depends on the separation of the centers of projection of the left and right lenses, but properly makes no reference to "gaze direction."[2] The difference in gaze directions between left and right eyes is a measure of convergence, which is perceived via the state of the eye-pointing muscles and constitutes, along with accommodation—the state of the eye-focusing muscles—the two low-level cues. By "low level," we

Manuscript received March 15, 1999; revised September 30, 1999. This paper was recommended by Guest Editor M. Tanimoto.

M. Siegel is with the the Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA.

S. Nagata is with NHK Science and Technical Research Labs, Human Science and Technical Research Labs, Human Science Research Divsion, Setagaya-ku, Tokyo 157, Japan.

Publisher Item Identifier S 1051-8215(00)02794-4.

[1]An autostereoscopic display is a CRT, LCD, etc., with on-screen optics that steer the right eye's image to the right eye, etc. They typically use lenticular arrays or barrier grids which generate azimuthal viewing zones outside of which steroscopic viewing is absent or incorrect. A zoneless autosteoscopic display would deliver correct stereo from any viewing azimuth.

[2]A regrettably superfluous parameter in some computer graphics models.

Fig. 1. "Garfield" photographed with telephoto, normal, and wide-angle lenses, from corresponding (a) far, (b) intermediate, and (c) near distances. [View in color at http://www.cs.cmu.edu/~mws/ieee3dvt99.]

mean that they depend on proprioception of the state of particular muscles rather than any cognitive processing of scene content. As we already mentioned, the conflict between convergence and accommodation has long been recognized as a source of stereo-viewer discomfort and stress. It is the result of the eyes focusing on the physical screen, where the 2-D projections of real world points are drawn, but converging elsewhere in space, where 3-D virtual world points are synthesized in front of or behind the screen.

When the picture, be it either still or video, contains recognizable objects, depth perception is also stimulated, apparently independently and synergistically, by about ten[3] "understanding-related" effects, which we can think of being fundamentally high level or cognitive in contrast with the fundamentally low level or sensory character of convergence and accommodation. These cognitive cues [8]–[11] include the following.

1) Interposition and partial occlusion: we understand that nearer objects can block the visibility of farther objects, but not vice versa.
2) Size and scale: the relative apparent sizes of known objects in a picture provides a depth scale for these objects, and also a relative size scale for unknown objects in the same scene whose relative depth can be inferred from adjacency, partial occlusion, etc.
3) Convergence of parallel lines ("linear perspective"): the local apparent distance between understood-to-be-parallel lines elucidates the depth of local objects.
4) Foreshortening due to perspective: a picture taken up close with a wide angle lens appears to have more depth than a picture of the same scene with the same field of view taken from farther away with a telephoto lens,[4] as illustrated in Fig. 1.
5) Vertical position in field: more distant objects are higher in the field of view if they are below the horizon and lower in the field if they are above the horizon.
6) Familiarity: we perceive more depth in pictures of familiar objects and scenes than in pictures of unfamiliar objects and scenes.
7) Distribution of light and shadow: this is the basis of "shape from shading" in computer vision.
8) Aerial perspective: due to atmospheric attenuation and scattering by dust, distant objects are bluer and less sharp than nearby objects.

[3]Different authors group and enumerate them differently.

[4]This can be understood as a consequence of an approximately Gaussian optical system's longitudinal magnification being approximately $-m^2$, where m is the transverse magnification.

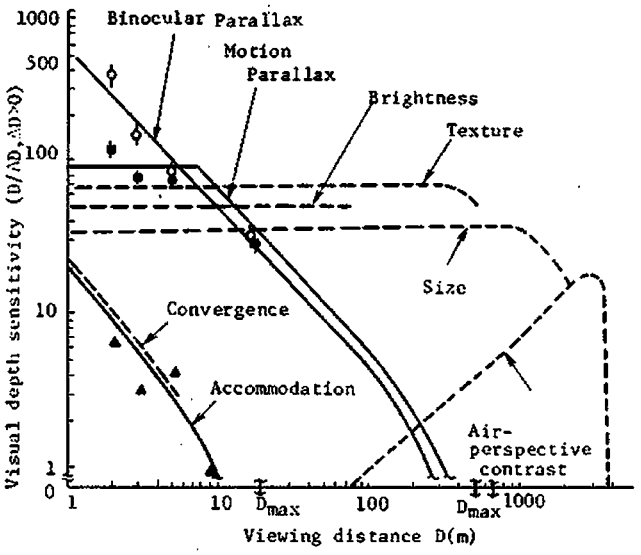


Fig. 2. Depth sensitivities of various depth cues as a function of viewing distance. For details see [11].

Cue 6) in a general way encompasses 1)–5), but most authors enumerate it separately. These cues are sometimes grouped in categories, e.g., "pictorial," "geometrical," "physiological," "psychological," etc. Note that 7) and 8) relate to the visual environment—illumination and transmission, respectively—rather than to the world objects.

Many of these modalities are depicted in Fig. 2 as contour lines on a log-log map of visual depth sensitivity ($D/\Delta D$) versus viewing distance (D) [11].

B. Depth Illusions

In addition to these genuine depth cues, there are about another ten illusions [11]–[15] wherein the viewing conditions or environmental factors affect the viewer's perception of apparent depth in actually flat pictures.

1) Looking at a picture with only one eye. This is called "the Claparade effect" [16].
2) Looking at a picture through an Iconoscope,[5] a viewing device that optically reduces perspective parallax.[6]
3) Viewing a picture from a greater distance. It is observed that there is a greater illusion of depth in large pictures observed from a large distance (as a movie screen) than in small pictures observed from a small distance (as a TV or computer monitor).
4) Changing the convergence of the eyes from that normally required by the distance from which the picture is viewed. Prisms can be used to alter the convergence *either inward or outward*.

[5]"Iconoscope" is coincidentally the name given the TV camera sensor ("tube") invented by Zworykin at RCA in 1933. Prior to this usage, the term was used exclusively for optical viewers of the same principle as the "vue d'optique," called "peeking machine" (Nozoki–Karakuri) in late Edo era, Japan [11] (same book, different article) and [17].

[6]Although reducing perspective parallax is the usual explanation there may be several illusions associated with viewing through a lens, their relative strengths depending on the lens diameter, focal length, and aberrations.

5) Looking at a picture through a small hole held close to the eye. This may be done monocularly or, using both eyes, with the aperture in front of only one eye. The monocular case gives a stronger illusion.
6) Changing the accommodation of the eyes from that normally required by the distance from which the picture is viewed. Lenses can be used to alter the accommodation *either inward or outward*.
7) Looking at a picture binocularly, with one eye receiving a sharp image and the other a blurred one.[7]
8) Looking at the reflection of a picture in a mirror.[8]
9) Looking at a picture with abnormal rotation of the visual images about the axes of vision.
10) Illuminance disparity (in taking a pair of otherwise identical pictures), and luminance disparity (between actually identical pictures), when viewing the pair with a stereoscope.[9]

Some of these effects are exclusively monocular, some are exclusively binocular, and some may be either monocular or binocular. The common element among all of them is that viewing conditions that reduce the perception of depth in solid (3-D) scenes are observed paradoxically to increase the illusion of depth in flat (2-D) pictures. Authors describing these illusions speculate that interfering with normal binocular stereopsis, with which the picture would be flat, frees the brain to synthesize the 3-D world from the high level cues embedded in the flat image content.

All of these phenomena are effective with both still and moving imagery. With video, motion parallax comes into play as a source of depth perception on a par with binocular perspective parallax. Analogously, in a general sense with the static pair binocular parallax and convergence, there are a pair of motion parallax cues, one relating to angular velocity *per se* (ω) and the second relating to difference of angular velocities $(\Delta\omega)$. This is illustrated in Fig. 3, where contours in the $(\omega, \Delta\omega)$-plane demarcate regions of depth perception and regions of motion difference perception superimposed on data from a particular subject.

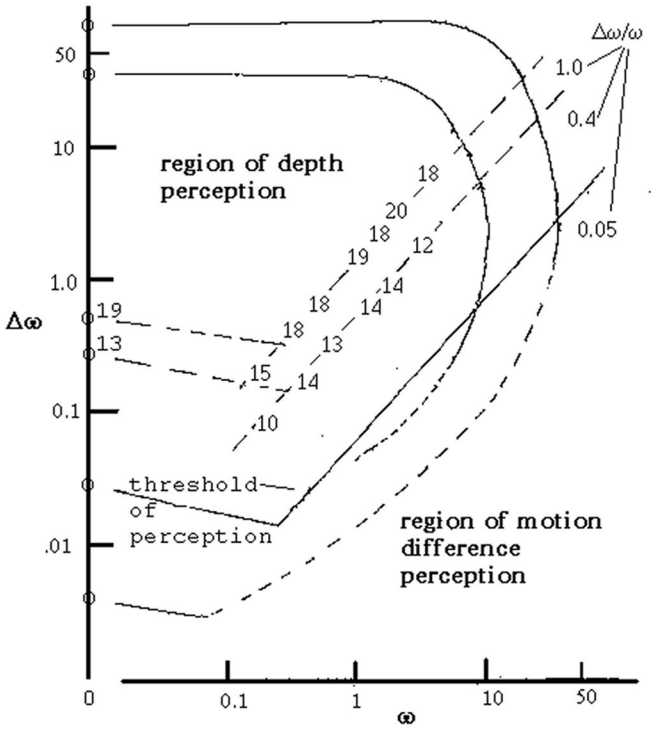


Fig. 3. Depth perception and motion difference perception regions as a function of angular velocity ω degrees/s and difference of angular velocities $\Delta\omega$ degrees/s for left–right motion at constant velocity ω of random dot patterns with spatially sinusoidal motion difference $\Delta\omega$. Outside arc is threshold of perception of motion difference. Inside arc and solid lines are thresholds of perception of depth. Two dashed lines indicate iso-depth perception; numbers on these lines are depth estimates reported by the subject. Notice approximate constancy of $\Delta\omega$ for small ω, approximate constancy of $\Delta\omega/\omega$ for large ω. Small negative slope at small ω is attributed to perception of rotation at $\omega = 0$. See [10] and [11] for details.

C. Integration of Depth Perception Modalities

Depth perception and cognition stimulating modalities that are not in conflict seem to be synergistic [7], [11], i.e., "the whole is greater than the sum of the parts." Thus, we expect that when consistent stimuli from several modalities are presented simultaneously, some can be increased and others correspondingly reduced while keeping the overall perception of depth and cognition of scene content unchanged. In Section III, on microstereopsis, we pursue the possibility that adequate depth perception and cognition can be stimulated by dramatically reducing binocular parallax disparity and on-screen disparity, while correspondingly and consistently increasing complementary depth perception modalities, e.g., motion parallax, perspective distortion, light-and-shadow effects, etc. In Section IV, we combine an unexpected result of the microstereopsis work with a depth illusion phenomena to generate and develop a concept and a suggested implementation for a zoneless autostereoscopic display.

III. Microstereopsis

Developers and users of stereoscopic applications recognize that there is an inverse relationship between disparity and the viewer's ease of stereopair fusion. This ease influences the user's perception of comfort. For example, in [20], we read: "for close viewing [meaning for tasks requiring serious concentration, versus for example, entertainment] the disparity should be only as big as requested." In Section III, we report initial experiments that demonstrate that surprisingly small disparities are adequate to stimulate binocular stereopsis. We also report qualitatively that "microstereoscopic" imagery is more comfortable to view than conventional stereoscopic imagery.

[7]This is the perceptual basis of what we would today call "mixed resolution coding" of stereo images.

[8]This is impossible to understand if the mirror is an ideal optical device. Presumably it is either a framing effect, or it is due to the inevitable visibility of (dirt on) the mirror's physical surface.

[9]When identical movies are viewed with luminance disparity, e.g., with a neutral density filter over one eye, and the scene content contains a particular kind and direction of motion, the resulting "Pulfrich effect" [18], [19] is particulary strong. However, the Pulfrich effect is physiologically distinct from the illusion described here.

A. Initial Hypothesis

Our hypothesis is that binocular stereopsis can be adequately stimulated by disparities generated by real or virtual camera interocular separations that are very much smaller than the nominal 65-mm human interocular separation, that this imagery is easier for viewers to fuse than is conventional stereo imagery, and that it is more comfortable (less stressful) to view than is conventional stereo imagery. We call this paradigm "microstereopsis." The hypothesis and the definition of microstereopsis will be progressively refined and quantified.

B. Motivation

Our thinking is motivated by an analogy with color vision. The Helmholtz tricolor model of color vision continues to serve well for all practical applications of color perception synthesis, e.g., photography, printing, and cathode ray tube displays. But, in parallel with the straightforward and essentially physical Helmholtz model, there is Land's partly psychological "retinex" [12], [22] color differential theory, and the experiments that bear it out. Land showed, among other things, that minutely disparate color separations can be displayed so as to stimulate perception of the full visible spectrum present in a complex real-world original scene.

This encourages us to suggest the possibility that, in an analogous fashion, minute perspective disparities might be adequate to stimulate perception of the full depth range in a complex real-world scene.

C. Refined Hypothesis

Study of the "illusions of depth in flat pictures" described in Section II teaches us that if a single picture contains enough familiar detail that its depth structure is partly discernable via high level understanding of the scene content (versus, e.g., low-level triangulation based on the binocular perspective disparity between corresponding points in a stereopair), then under appropriate viewing conditions, the picture will stimulate a correct and adequate illusion of depth. This observation leads to the following line of reasoning.

1) If a scene contains enough familiar detail that its depth structure can be deduced by high-level reasoning, then adequate binocular stereopsis can be stimulated by perspective disparity that is substantially smaller than the disparity demanded by the "geometrical correctness" defined explicitly in [1] and implicitly in [2] and other well-known optics texts.[10]
2) Smaller-than-"correct" disparities stimulate smaller-than-"typical" portions of the physical and mental discomforts attributable to conflicts between depth-sensing modalities.
3) Disparity reduction by left–right shift of the members of a stereo pair to make the disparity around the center-of-interest approximately zero[11] is effective.
4) Disparity reduction by reducing the interocular separation to a value smaller than the human interocular separation (which, according to footnote 10, is the required camera separation for geometrical correctness) is also effective.
5) Left–right shift and reduced interocular separation are synergistic, and are especially effective in combination.
6) The combination of reduced interocular separation and other depth perception stimulating factors, especially perspective distortion (shown visibly increasing in Fig. 1 from left to right, corresponding to telephoto, normal, and wide-angle lenses), and motion parallax (Fig. 3) are also complementary.

D. Experiments with Microstereopsis

We have tested this hypothesis in a series of informal experiments in which the authors, their colleagues, and many visitors to their labs have been shown various implementations and their perceptions queried and noted. Formal human factors experiments with rigorous controls, comprehensive data recording, and complete statistical analysis have not yet been conducted, so these results must be considered anecdotal. We believe that our tentative results—demonstrated by the figures reproduced herein—will be validated by future experiments.

1) Test-Image Generation: We collected test images of the still-life "Garfield" (the toy cat) scene, from which three typical frames are shown in Fig. 1. We used three camera-to-scene ranges (132, 48, and 30 cm) and three corresponding lens focal lengths (50, 20, and 12.5 mm). Fig. 1 illustrates that the fields of view are approximately the same in these three cases. For each range/focal length we collected 41 frames at 1-mm intervals between -20 and $+20$ mm from the centerline. This was accomplished by mounting the camera on a sturdy left–right ruled translation rail. Thus the camera axis was always moved precisely parallel to itself, i.e., there was no camera convergence. The camera had been previously modified so that its CCD could be shifted left–right with respect to the optical axis of the lens. Continuous "center-of-interest compensation" was achieved by moving the CCD, after each camera move, so as to return the

[10]We summarize, briefly, for readers who do not have immediate access to [1], [2], or equivalent references. The goal of a "virtual reality" display system is to write on the retinas exactly what the real scene would write on them. Straightforward geometrical considerations dictate that, for the class of display systems that multiplex both perspectives onto one flat screen: 1) the camera lens separation must be the same as the viewer's interocular separation; 2) the viewer's position, with respect to the screen, must be the same as the camera's position with respect to the region of overlap between the camera fields-of-view; and 3) in most cases, the camera lens axes must be parallel (the exception is for dual projector display systems with the projectors converged by the same angle as the cameras were originally converged).

[11]There are three basic ways to do this. The first is to converge the cameras so as to overlap the fields-of-view and to zero the disparity in the vicinity of the scene's center-of-interest. This method is discouraged because the resulting keystone distortion causes vertical disparities that conflict with comfortable viewing. The second is to use parallel camera axes and shift the images left–right to zero the disparity at the center-of-interest. This results in pleasant to view stereo, but the usable image width is smaller that the original image width by the size of the shift, making the useable aspect ratio of the viewable image a function of the distance to the center-of-interest. The third is to use parallel camera axes and to shift the camera sensors (e.g., CCDs) outward to overlap the fields-of-view at the distance to the center-of-interest. This is the perfect solution; its only drawback is that the cameras have to be physically modified. Given a "center-of-interest finding algorithm," any of the three would be easy to automate. The three methods can be combined.

Fig. 4. Approximately conventional interocular separation, no center-of-interest compensation, i.e., 1-m camera-to-scene, 50-mm lens, parallel axes, 40-mm interocular separation. This pair can be comfortably fused by "free viewing" (with gaze directions converged) with 65-mm on-page interocular separation, but the disparities are uncomfortably large on a CRT or a distant screen. [View in color at http://www.cs.cmu.edu/~mws/ieee3dvt99.]

bridge of Garfield's nose to a fixed mark on the monitor screen.[12] All images were digitized to 640 pixels × 480 lines × RGB and saved using a lossless compression algorithm.

2) Test-Image Stereo Animation: From each of the three test-image sets we can create one stereo pair with 40-mm interocular separation, two pairs with 39-mm interocular separation, and so forth, down to 40 pairs with 1-mm interocular separation. To display stereo and motion with any desired interocular, stereo pairs are assembled on-the-fly in a format appropriate for the display technology that is employed, e.g., above/below, side-by-side, interlace, or anaglyph.[13] Each stereo pair set at a given interocular is animated to create a movie; the movie is 40 frames long for the 1-mm interocular, and correspondingly shorter for smaller interoculars, down to only a single still frame pair for the 40-mm interocular. We have demonstrated all four of the display alternatives mentioned; however the experiments reported in this paper were all done using the above/below format. The full 640 × 480 resolution is displayed by this format, whereas some other formats achieve stereopair multiplexing only at an undesirable price in resolution. The display technology uses a StereoGraphics Z-Screen on a Silicon Graphics Indy computer monitor. The Z-Screen produces alternate left–right circular polarization switched synchronously with the left–right stereo frame alternation. Viewers wear passive left–right circularly polarized glasses that look and feel like inexpensive plastic sunglasses. With center-of-interest compensation, the perceived motion is approximately rotational about a vertical axis through the center-of-interest. It is not exactly rotational because the camera moves on a tangent line rather than on a circular arc. Without center-of-interest compensation the perceived motion is the complement of the camera's actual motion.

3) Test-Image Stereo Pair Examples: Fig. 4 shows a Garfield pair with 40-mm interocular separation without center-of-interest compensation.

Fig. 5 uses the same optics and geometry as Fig. 4, but the CCD was shifted from frame-to-frame to provide center-of-interest compensation as described above.

Fig. 6 shows gray-level representations of the disparities between left and right views in Figs. 4 and 5.

[12]For reference and demonstration, some data were also collected without center-of-interest compensation, i.e., with the CCD always centered on the lens axis.

[13]Our anaglyphs use NASA's *de facto* web page standard: red channel for the left image, green and blue channels for the right image.

Fig. 5. Same as Fig. 4, but with center-of-interest compensation by shifting CCDs. This pair can be free-viewed comfortably with gaze directions parallel. [View in color at http://www.cs.cmu.edu/~mws/ieee3dvt99.]

Fig. 6. Gray-level representations of disparities (differences) between the image pairs of Figs.. 4 and 5. When corresponding pixels are identical, this representation is midrange gray; negative (left–right) differences are darker, positive differences lighter.

Fig. 7. Microstereopsis illustrated by interocular separation 3% of normal. Binocular stereopsis is adequately stimulated with the normal lens used in the previous figures, but is enhanced here by using a 12.5-mm wide-angle lens and a correspondingly closer camera-to-subject distance. Easily free viewed. [View in color at http://www.cs.cmu.edu/~mws/ieee3dvt99.]

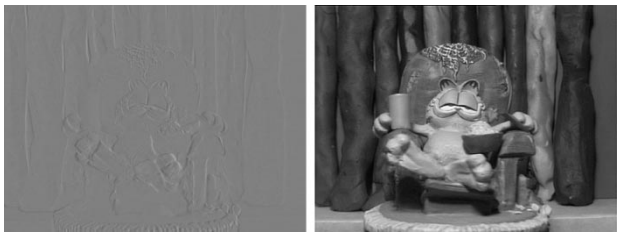


Fig. 8. (a) Disparity (difference) between left and right views in Fig. 7(b). (b) The same two frames formatted as an anaglyph. View with red lens on left, green/blue lens on right. [View in color at http://www.cs.cmu.edu/~mws/ieee3dvt99.]

Fig. 7 illustrates microstereopsis. The effect is made easier to see on the printed page by using a wide-angle lens and correspondingly decreasing the camera-to-scene distance. The coupled high-level understanding effect of perspective distortion and low-level geometrical effect of increased relative disparity (ratio of interocular to range) contribute to enhancing depth perception in this pair.

Fig. 8 shows a gray-level representation of the disparity (difference) between the left and right views of Fig. 7. Notice the long run lengths of zero difference (represented by gray

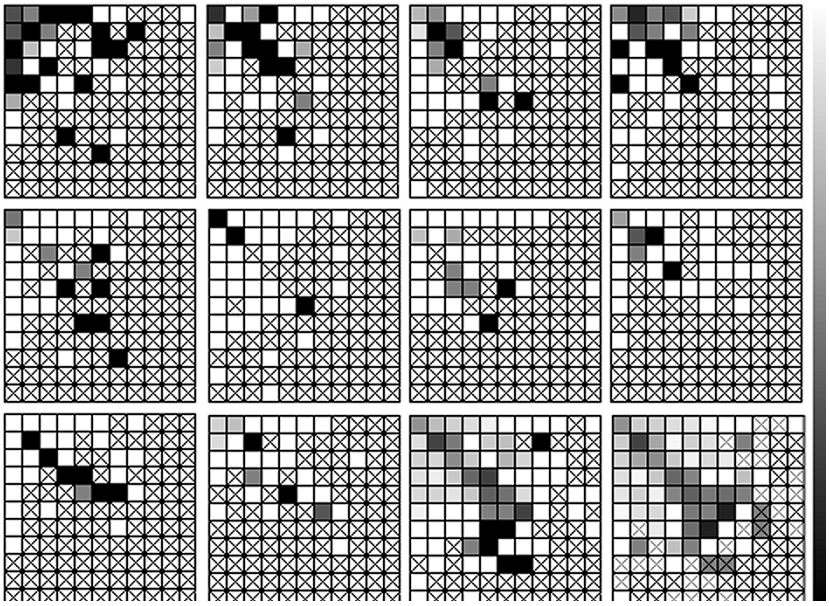


Fig. 9. Depth pair ordering summary. Numbered left-to-right, top-to-bottom. Frames 1–10: individual subjects. Frame 11: average of the 10 individuals (800 pairs of frames). Frame 12: average over the 10 individuals, symmetrized across the diagonal. Origins are at upper left of each frame, where both pictures have zero disparity. From left to right, the left picture disparity increases from 0 to 10-mm interocular separation. From top to bottom, the right picture disparity increases the same way. Grey-level scale indicates 0%–100% correct identification.

level 128) and the apparently narrow distribution of gray levels; these are indicative of the high compressibility of the difference image. The image shown is a 640 $\times$ 480 $\times$ 256 gray-level JPEG file that occupies only 5275 bytes. An anaglyph representation of the stereo pair is also shown (right side); notice the color ghosting and fringing usually seen in anaglyphs is practically absent in this picture. Interocular separation between left and right views is 2 mm.

4) Applicable Range: With display methods that preserve the full image resolution (e.g., the above/below format on a Silicon Graphics Indy workstation), binocular stereopsis is perceived in all sequences by viewers who self-report normal stereo vision. However binocular depth perception is weak at 132-cm range and interocular separation less than 3 mm. In fact, pending definitive experiments, it is arguably absent at the largest range and smallest interocular separation when observing an individual still pair versus observing an animation in which binocular stereopsis is complemented by motion parallax.

The parameter range that the Garfield image set covered is thus seen to have been appropriate, spanning the spectrum of possibilities from "geometrically correct" virtual reality to disparity that is almost too small to adequately stimulate binocular stereopsis. The lower limits are impressively small, 1-mm interocular at the shortest (30 mm) working distance, but are still large compared to the 1/500 color differential that is adequate for retinex color vision. Mechanical limitations and the unavailability of higher resolution displays make it impractical for the present apparatus to investigate the regime below 1 mm. However, it is clear from the observed trend that with lower-than-HDTV image and display resolutions it will not be useful to go much below 1 mm anyway. It is plausible that with HDTV-standard higher resolution (pixel count) image sensors and higher resolution (dot count) displays even smaller disparities will be effective.

5) Depth-Pair Ordering Experiment: In a pilot experiment, we evaluated ten viewers' ability to perceive microstereopsis based on their pair-ordering of the "depth" or "3-D-ness" that they perceive in pairs of simultaneously viewed stereoscopic pictures. We presented each viewer with two 640 pixel $\times$ 320 line $\times$ 2 camera position frames side-by-side and asked him or her to indicate (by pressing the left, center, or right mouse button) whether the left picture, neither picture, or the right picture seemed to "have more depth." The pictures were randomly chosen by a computer program selecting images on-the-fly from the Garfield wide-angle set with center-of-interest compensation. The program's random number generator was reseeded with each viewer's start time. Each viewer was shown 80 pairs of stereo pictures in two sets of 40. In the first 40, one picture was always flat, i.e., the left and right eye subimages were identical. In the second 40, both were nominally stereo, although sometimes one or both (rarely) happened randomly to have zero disparity. Within each 40, there were 5 sequences of 8 $\times$ 2 pictures. Each of the 8 was drawn from a random Poisson-weighted disparity distribution. The first 8 had a mean camera interocular separation of 4.5 mm, the second 8 had a mean of 3.5 mm, and so on down to the fifth 8, which had a mean of 0.5 mm. (The mean is the only parameter of a Poisson distribution; the standard deviation is the square root of the mean.) The flat images in the first 40 are taken from perspectives randomly displaced from the midpoint of the data set by the same statistical distribution that randomizes the disparities. The viewers are told in advance that the task "starts easy and gets more difficult as it progresses," but they are not told that in the first 40 sets one of the pictures is always flat.

The results of this experiment are summarized in Fig. 9. In each of its 12 frames, gray level indicates the fraction (0%–100%) of correct responses for left and right pictures having the disparities that correspond to each square's position

in the matrix. Left picture disparities increase from left to right, and right picture disparities increase from top to bottom. Each square ticks off 1 mm of camera interocular separation, from 0 to 10 mm. Along the top edge, the disparity of the left picture increases as the camera interocular separation increases from 0 to 10 mm, whereas the right picture remains flat. Along the left edge, the disparity of the right picture increases as the camera interocular separation increases from 0 to 10 mm, whereas the left picture remains flat. In the interior of the matrix, left and right pictures are both stereo, but with generally different disparities corresponding to different camera interoculars. Along the top-left to bottom-right diagonal, the left and right pictures have the same disparity. However they are not usually the same picture: each is typically taken from a different pair of camera perspectives, but with the same separation. The gray levels indicate the fraction of correct identifications for each disparity pair, black to white corresponding to 0%–100% as shown on the scales at the right of each frame. X's indicate disparity pairs that did not appear in the random set.

It is clear from inspecting these frames, especially the symmetrized summary frame at the lower right, that even at 1-mm interocular separation, versus a flat picture there is a significantly greater-than-chance probability ($\sim$2/3) that viewers will correctly identify the microstereoscopic picture. At 2-mm interocular separation, it is clear that they will usually make the identification correctly. As disparity increases, especially as it increases versus a flat picture, the microstereoscopic picture is conclusively identified. Even the first viewer (upper left frame), who self-identifies as stereodeficient, did well with disparities corresponding to camera interocular separations more than about 5 mm.

Along the diagonal, where disparity difference is nil even when absolute disparity is high, viewers are seen to have great difficulty recognizing that the depths are identical. This error is possibly due to the coming into play of depth sensing modalities or judgement criteria other than binocular perspective. One viewer spontaneously mentioned "sharpness" as strongly influencing his decisions in close cases. Sharpness inevitably varies a little from picture to picture.

We reiterate that these data were collected informally, using as subjects only colleagues who happened to be available at the time of the pilot test, and without employing a rigidly-controlled protocol. However, the random on-the-fly generation of the test pictures ought to completely preclude the possibility of any tester-bias effects.

In addition to the data shown in Fig. 9, data records include a code that identifies the viewer, the time the session began, the sequence of individual Garfield pictures that were assembled into stereopairs and stereo picture pairs, and the times between mouse clicks, i.e., the time each viewer took to reach each decision. Additional information could later be extracted from the saved data, but it is beyond the scope of this paper.

6) Time to Achieve Fusion: Another informal observation is that with decreasing interocular separation, viewing comfort increases and perceived time to fuse left and right images decreases. Perceived time to fuse becomes effectively zero when interocular separation is reduced below about 10 mm. A key requirement of "kinder gentler stereo" is that this time be effectively zero; if the fusion time (perhaps coupled with accommodation time) is perceptible, then it interferes with normal work habits, for example, turning away from the screen briefly for a face-to-face exchange with a colleague becomes a significant disruption if stereo fusion has to be reacquired afterwards.

7) Adjustable Degree of Reality: With microstereopsis, the depth order of scene objects is disambiguated, but of course the perceived depth is not absolutely calibrated: microstereopsis apparently delivers "just enough reality" for computer graphics, video news and entertainment, and enough for most eye–hand coordinated tasks, including teleoperation of mobile robots. The degree of reality is adjustable to match the content, the task, and the stereo ability of the viewer. However, if strict geometrically correct virtual reality is required then microstereopsis cannot be used. On the other hand, we have failed to identify a real world task (in contrast with an academic task, e.g., to test stereoacuity) that actually requires strict geometrically correct virtual reality.

E. New Hypothesis Regarding Future Displays

A remarkable and unexpected outcome of these experiments, is that *at the smallest interocular separations, when the screen is viewed without stereo demultiplexing eyewear, the on-screen disparity is imperceptible*. Instead of the usual and expected ghosting (offset of overlaid images), viewers without stereo eyewear see only a slight blurring in the background and foreground. This is apparent in the anaglyph on the right side of Fig. 8: to the eye unaided by color filters it is almost indistinguishable from a normal flat photograph.

This observation suggests a new hypothesis with possibly important practical consequences: with microstereopsis, it should be possible to stimulate binocular stereopsis even in the presence of substantial left–right channel crosstalk. Crosstalk is, of course, the bane of stereoscopic display systems. In current systems, displayed disparities are large and crosstalk is perceived as ghosting, but when disparity becomes small enough that it is perceived as blur rather than as ghosting, the perceptual manifestation of crosstalk becomes as natural and as unobjectionable as depth-of-field.

This means that we can consider new display system concepts that use left–right multiplexing technologies that do not completely separate the channels, but that rather only weight the left–right mix in favor of the right eye when the right eye's picture is on the screen, and vice versa. This, in turn, suggests the possibility of zoneless autostereoscopic displays. An implementation is discussed in Section IV.

F. Crosstalk Experiment

We conducted a simple test of this hypothesis by modifying a conventional LCD shutter-glasses controller to give it adjustable crosstalk, i.e., variable imperfection. The experimental protocol is dictated by the delicate (nonlinear and hysteretic) nature of the adjustment: open loop, it is practically impossible to return to any previous "crosstalk setting." We circumvent his delicacy as follows. We first decrease the control voltage until crosstalk is perceived and the perception of depth is lost when an animated sequence with 30-cm range, center-of-interest compensation, and 20-mm interocular separation is presented on the

screen. Then, without making any adjustments, the animation is replaced with another, also with 30-cm range and center-of-interest compensation, but now with 2-mm interocular separation. This experiment has been tried with three subjects. They all reported comfortable perception of depth and no perception of ghosting with the 2-mm interocular separation animation. As with the experiment that quantifies the lower limits of microstereopsis (Fig. 9), absent formal human subjects protocol, this experiment must be regarded as anecdotal.

G. Microstereopsis Summary

Preliminary experiments demonstrate that there is a range of interocular separations for which: 1) disparity is big enough to stimulate binocular stereopsis and 2) disparity is small enough that left–right channel crosstalk is perceived as blur instead of as ghosting. The range of interocular separations at which this happens corresponds to only a few percent of the normal human interocular separation, i.e., 1–3 mm, out of the nominal 60–65 mm.

If we also apply a left–right image shift to make the disparity zero in the vicinity of the "center-of-interest" of the scene, then the blur is removed from the midground and transferred to the foreground and background. The blur then looks like normal depth-of-focus rather than either an out-of-focus condition or a ghosting imperfection of the stereo demultiplexer.

We call the combination of small interocular separation and center-of-interest compensation "microstereopsis."

A monitor displaying microstereoscopic imagery looks normal even when viewed without stereo-viewing eyewear. When viewed through stereo demultiplexing eyewear, the depth appears immediately and naturally, with no observable time delay to fuse the left and right views.

The natural appearance of the microstereo display when viewed with or without stereo demultiplexing eyewear suggests that it will fail "softly" for stereo-deficient users.

Because the difference between left and right eye views is extremely small, zero after quantization in most pixels, the microstereoscopic imagery can be deeply compressed. In experiments in which difference images were JPEG encoded, the compressed difference file size was just a few percent of the initial size of either of the views. A tailored coding approach could presumably do even better.

H. Camera Requirement

In the laboratory it is fine, when the scene is a still life, to capture microstereoscopic image pairs by moving the camera a few millimeters between two exposures made a few seconds apart. Even the time required to move the CCD to accomplish center-of-interest compensation is acceptable in the laboratory. But an important question we must ask is how to build practical microstereoscopic camera pairs; particularly, how to build practical microstereoscopic video camera pairs? Practical microstereoscopic cameras will have to capture left and right perspectives simultaneously, but with an interocular separation of only a few millimeters. We expect that the solution to this constraint conflict will lie in the area of "single-lens stereo" [23], in which left and right perspectives are obtained by employing left and right off-center subapertures added as a modification to an ordinary camera lens. This approach has the added advantage of automatically performing center-of-interest compensation, since the prismatic effect of the lens with off-center aperture produces an image shift that is exactly equivalent to the effect of proper CCD movement.

IV. Zoneless Autostereoscopic Displays

In Section III, we described apparatus and experiments through which we demonstrated that perception of microstereopsis is robust against crosstalk between the right and left eye channels. In this section, we propose that freeing the 3-D-display designer from the canonical requirement to achieve the lowest possible crosstalk gives him or her opportunities to propose new approaches to the engineering of autostereoscopic displays. In particular, we suggest the possibility of achieving a zoneless (or at least a zone-boundary free) autostereoscopic display.

To illustrate the concept, consider a flat-panel light source whose luminance is lambertian, i.e., independent of viewing direction. Overlay it with a filter that attenuates the luminance monotonically with horizontal viewing direction. It then looks somewhat brighter in one eye than the other, say the right eye. Now overlay an LCD video panel displaying the right image of a stereo pair. The last "illusion of depth" in Section II is "luminance disparity;" thus we expect the brain to conjure up an illusion of depth based on its understanding of the picture. Now let the sign of the luminance gradient be reversed at 50–60 Hz and the right and left image synchronously toggled. The result would ordinarily be considered a very badly engineered frame-sequential stereoscopic display. It would be bad in the sense that the crosstalk between left and right channels is close to 100%, and crosstalk is typically perceived as ghosting, and ghosting is bad. But in Section III-F, we showed that when the disparities are small and the center-of-interest is compensated, crosstalk can be perceived not as an unacceptably large degree of ghosting, but rather as an acceptably small degree of foreground and background blur. If an appropriate match can be tailored between the engineering (physical) and the perceptual (psychophysical) parameters, then a microstereoscopic display can be based on this arrangement. This display would be zoneless (correctly ordering the perspectives from any viewing angle), autostereoscopic (requiring no eyewear, headtracking, etc.), and "kind and gentle," in that it harmoniously combines complementary modalities of stereo perception and depth illusion.

A. Implementation: Non-Lambertian Screen or Source

Based on the model outlined above, we can describe in a general way (without yet being able to give numerical values for physical or psychophysical parameters) as to how we would go about engineering a zoneless autostereoscopic display that takes advantage of the robustness of microstereopsis against crosstalk. A reasonable initial approach is simply to illuminate the display screen (if it is transmissive, e.g., an LCD) or filter it (if it is emissive, e.g., a CRT) in a nonlambertian angular pattern.

Fig. 10. (R, L) states of louver filter. In R state, bias favors the right eye, in the L state it favors the left eye.

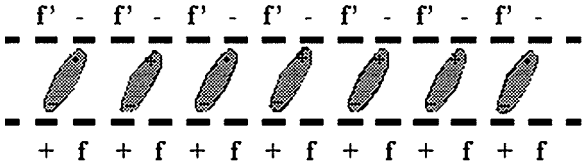


Fig. 11. Electrode pattern and polarization required to produce the L state of Fig. 10. Electrode sets f and f' are allowed to float during the phase shown here.

Passive screens with the required nonlambertian property (but stronger gradients that we probably want) are actually commercially available, e.g., 3M's "Privacy Shield" material for bank ATM's, laptop computers, and some automobile instrument-panel applications. This product is a microfabricated "venetian blind" or louver filter. The concept is illustrated in Fig. 10. Note that it is not supposed to be a binary barrier filter, as in some zoned autostereoscopic displays, but rather it has only a gentle angular gradient.

Two engineering challenges remain to be overcome to turn this idea into a practical microstereoscopic display: 1) we need an electronically switchable louver filter[14] and 2) the gradient needs to be strong enough between the eyes that sufficient bias is achieved, but not so strong over the full range of viewing azimuth that the illumination difference between the two states is annoying. Depending on the outcome of measurements of the psychophysical factors, this tradeoff may limit the display's useful range of viewing angles. On the other hand, even in the worst case, the idea should be workable for viewing the display approximately head-on. Even in that worst case, it should still be far less restrictive about head position, in both azimuth and in distance from the screen, than are any of the lenticular and barrier displays currently on the market.

B. Realization: Electronic Louver Filter

Electronic louver filters could be implemented using several emerging display technologies, e.g., suspended particles [24], reverse emulsions [25], or polymer encapsulated liquid crystals [26]. To illustrate briefly, we describe a suspended particle display technology approach schematically in Fig. 11. The method uses elongated opaque dielectric particles with permanent dipole moments suspended in a transparent dielectric liquid. The particles are oriented as desired by an electric field produced by electrodes patterned on the windows. This technology is currently in pilot production for "smart windows" for automatic control of indoor sunlight. It seems it will require only more complex electrode patterning and driver electronics to make it operate as an electronic louver filter.

V. Conclusion

In pursuit of "kinder gentler stereo," a depth-display paradigm that would overcome the negative aspects of conventional left–right multiplexed binocular image pair display systems, we have hypothesized and demonstrated experimentally that extremely small disparities are adequate to stimulate binocular stereopsis. We use center-of-interest compensation and small interocular separations (a few percent of the nominal 60–65-mm human interocular separation) to achieve this effect, which we call "microstereopsis." Our experiments—still informal—support our hypothesis that microstereopsis provides a low-stress alternative to 3-D-displays that rely on conventionally large left–right disparities. Synergistic combination of microstereopsis with motion parallax, perspective distortion, and other "single image" depth-perception modalities, as well as several depth-illusion stimulating modalities, provides the sought after "kinder gentler stereo." The approach fails "softly" for stereo-deficient users, and it is extremely synergistic with deep compression algorithms. We suggest and demonstrate that in a 3-D-display system based on microstereopsis, crosstalk between the multiplexed left and right eye channels is negligibly objectionable, in contrast with its being extremely objectionable in conventional 3-D-display systems. This suggests the possibility of engineering zoneless autostereoscopic 3-D-displays. We describe a class of implementations based on a nonlambertian light source behind the display, or a nonlambertian screen in front of it, and a particular implementation that uses an electronically switched louver filter for the nonlambertian element.

[14]To the best of our knowledge, only static louver filters, like the 3M "Privacy Shield" we described, are commercially available. We have briefly discussed the possibility with holders of several technologies, e.g., suspended particle displays, electroholographic devices, etc., and their response to the technical challenge is always favorable. The economic challenge remains unanswered.

Acknowledgment

One of the authors, M. Siegel, would like to thank Kansai Research Institute (KRI) for funding the study of monocular depth perception that stimulated the key new ideas discussed in this paper; StereoGraphics Corporation, for donated hardware and engineering support; G. Podnar, for design and construction of cameras for geometrically correct stereo; and A. Guisewite, for data collection and other support.

References

[1] V. S. Grinberg, G. W. Podnar, and M. W. Siegel, "Geometry of binocular imaging," in *Proc. SPIE/IS&T Stereoscopic Displays and Applications V*, San Jose, CA, Feb. 1994, pp. 56–65.

[2] A. C. Hardy and F. H. Perrin, *The Principles of Optics*, 1st ed. New York: McGraw-Hill, 1932, ch. XXV (Stereoscopy), pp. 517–533.

[3] L. Lipton. (1991.) *The Crystal Eyes Handbook*. Stereographics Corporation, San Rafael, CA. [Online]. Available: http://www.stereographics.com/documents.html

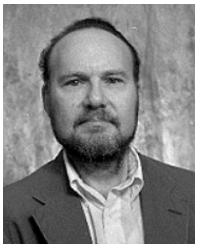
[4] P. A. Howarth. (2000). Visual Ergonomics Research Group , Loughborough University, Loughborough, U.K. [Online]. Available: http://www.lboro.ac.uk/departments/hu/groups/viserg/virtrel1.htm

[5] E. M. Kolasinsky. (2000) *Simulator sickness in virtual environments.* [Online]. Available: http://www.cyberedge.com/4a7a.html

[6] B. Julesz, "Foundations of cyclopean perception," *J. Opt. Soc. Amer.*, vol. 59, p. 1544, 1969.

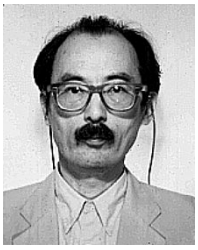
[7] S. Nagata, "Visual effects in multi-directional stereoscopic images," in *Proc. Int. Symp. 3-Dimensional Image Technology and Arts, Seikien Symp.*, vol. 8, sec. 4, 1992, pp. 137–145.

[8] T. Okoshi, *Three-Dimensional Imaging Techniques* (in English, a translation and extension of the Japanese "Sanjigen-Gazo Kogaku," Sangyo-Tosho, Tokyo, 1972). New York: Academic, 1976.
[9] N. A. Valyus, *Stereoscopy*. London, U.K.: The Focal Library, 1966.
[10] S. Nagata, "Interactions between binocular parallax and motion parallax in stereoscopic images" (in Japanese), *J. 3D Images*, vol. 5-2, pp. 73–82, April 1991.
[11] S. Nagata, "How to reinforce perception of depth in single 2D pictures," in *Pictorial Communication in Virtual and Real Environments*, S. Ellis, Ed. London, U.K.: Taylor and Francis, 1991, pp. 527–545.
[12] A. Ames, "The illusion of depth from single pictures," *J. Opt. Soc. Amer.*, vol. 10, pp. 137–148, 1925.
[13] A. W. Judge, *Stereoscopic Photography*. Boston, MA: American Photographic, 1926, sec. Pseudo-stereoscopic effects, pp. 137–139.
[14] H. Schlosberg, "Stereoscopic depth from single pictures," *Amer. J. Psych.*, vol. 54, pp. 601–605, 1941.
[15] A. H. Schwartz, "Stereoscopic perception with single pictures," *Opt. Spectra*, pp. 25–27, September 1971.
[16] A. H. Strief, "Die binokulare verflachung von bildern, ein vielseitig bedeutsames sehproblem," *Klinische Monatsblatter fur Augenheilkunde*, vol. 70, p. 1, 1923.
[17] T. Screech, "The western scientific gaze and popular imagery in later Edo Japan: The lens within the heart," in *Cambridge Studies in New Art History and Criticism*. Cambridge: Cambridge Univ. Press, 1996.
[18] C. Pulfrich, "Die stereoskopie in dienste der isochromen und heterochromen photometric," *Naturwissenromenschuften 10*, p. 533, 1922.
[19] M. J. Morgan and P. G. Thompson, "Apparent movement and the Pulfrich effect," *Perception*, vol. 4, pp. 3–18, 1975.
[20] N. Hiruma, "Accommodation response to binocular stereoscopic TV images," in *Human Factors in Organizational Design and Management III*. Amsterdam, The Netherlands: North Holland, 1990, p. 233.
[21] E. H. Land, "The retinex theory of color vision," *Amer. Scientist 52*, pp. 247–264, 1964.
[22] ——, "Recent advances in retinex theory . . .," in *Proc. Nat. Acad. Sci.*, vol. 80, 1983, pp. 5163–5169.
[23] W. J. Carter and M. A. Weissman, "Single-lens stereoscopy: A historical and technical overview," in *Proc. SPIE/IS&T Conf.*, vol. 2653, San Jose, CA, 1996, pp. 76–79.
[24] R. L. Saxe and R. I. Thompson. (2000) *Suspended-particle displays* . [Online]. Available: http://www.refr-spd.com/article.html
[25] Z. Bryning, R. Cromer, S. Beilin, and M. Bryning. (2000) *The REED Display*. Zikon Corporation, San Jose, CA. [Online]. Available: http://www.zikon.com/reed.html
[26] R. L. Sutherland and L. V. Natarajan, "Electrically switchable holograms: Novel PDLC structures," *Liquid Crystals Today (Newsletter of the International Liquid Crystal Society)*, vol. 7, no. 1, pp. 1–4, Mar. 1997.
[27] B. Julesz, *Foundations of Cyclopean Perception*. Chicago, IL: Univ. Chicago Press, 1971.
[28] T. Okoshi, *3D Image Engineering* (in Japanese,), ser. Advanced Science and Technology in Electronics. Tokyo, Japan: Asakura Shoten, 1991, vol. B4.



Mel Siegel received the B.A. degree from Cornell University, Ithaca, NY, and the M.S. and Ph.D. degrees from University of Colorado at Boulder, all in physics.

He is a Senior Research Scientist with the Robotics Institute, School of Computer Science, Carnegie Mellon University (CMU), Pittsburgh, PA, and Director of its Sensors, Measurement, and Control Laboratory. He did a post-doctorate in aerospace engineering and engineering physics at the University of Virginia, an an Assistant Professorship in physics at the State University of New York at Buffalo, and worked as Director of Research and Development for the mass spectrometer manufacturer Extranuclear Labs (now ABB Extrel), before joining the CMU Robotics Institute. His background is in physical and measurement sciences, analytical instrument development, and the application of computer science to measurement, diagnosis, and control. His interests also include 3-D-stereoscopic video and computer graphics display systems. The dominant themes in his recent research have been "difficult measurements in difficult environments," exemplified by his use of mobile robots to inspect aging aircraft, and high definition technology as enabling means for a new generation of 3-D-stereoscopic technologies and applications.



Shojiro Nagata received the B.S. degree in electrical communication engineering from Tokyo Denki University, Tokyo, Japan, in 1965.

He joined NHK Science and Technical Research Labs in 1960, where he has been engaged in research on television picture quality and visual information processing, including the neural networks of the human brain. His current interests are human 3-D spatial vision and advanced dispays, including stereoscopic/virtual reality displays and the ergonomic issues and human factors involved.

Mr. Nagata is a member of the International Society for Virtual Reality and Multi-Media Systems (VRMS) and the Virtual Reality Society of Japan, as well as other internal Japanese societies. Since 1994, he has been a Committee Member of the International Society for Optical Engineering (SPIE) Conference on Stereoscopic Displays and Applications.