chapter six

Human Perception of Sensor-Fused Imagery

Edward A. Essock, Jason S. McCarley, Michael J. Sinai, and J. Kevin DeFord

Contents

Keywords: *sensor fusion, image fusion, infrared, image intensification, color, texture, segmentation, detection, recognition, natural scenes, perceptual organization, psychophysics, nighttime, night vision*

6.1 Introduction

Remote sensing provides information about external stimuli that humans would not otherwise be able to perceive via their biological sensory systems. The electronic sensors used in remote sensing obtain a spatial mapping of external stimuli that is then converted to a type of stimulation that the human biological sensors *can* detect. Typically, the electronic sensor output is converted into a spatial mapping rendered in variations of light energy emitted by a cathode ray tube (CRT) visual display. A crucial yet often overlooked factor that determines the usefulness of a remote sensing system is the ability of the human perceptual system to extract useful information from the transformed rendering of the alternative sensor's information that has been recoded into the visible-light display that the human views.[21] In other words, such sensors provide ways to present spatial information to the human visual system for analysis even though that information is not originally perceivable by the human visual system. Human viewing of such imagery, often termed "nonliteral imagery," implicitly subjects the spatial information to the powerful processing of the human visual system just as if the image were originally an optical mapping on the retina of reflected / emitted visible light conveying spatial and dynamic relations in a natural visual scene. The image produced by remapping the alternative spatial information into a visually accessible format must be analyzed and perceptually organized by the visual system, just as if it were a conventional light image, in order for the human viewer to find the remapped information useful.

In this chapter we focus on one type of nonliteral imagery; nighttime imagery obtained by electronic sensors. Typically, these sensors are either infrared detectors or image intensifiers. (The latter sensor, typified by traditional night vision goggles, serves to electronically amplify the effect of the few photons in the nighttime outdoor environment.) In our review, we focus on imagery produced by combining or fusing imagery from multiple electronic sensors, a procedure performed in an attempt to capitalize on the advantages of each type of sensor. We present an overview of the research findings to date that bear on the question of the extent to which the sensors and fusion methods effectively convey spatial information that humans can perceptually organize and use functionally. Our goals in this chapter are to:

1. Explain the need for behavioral assessment of perceptual performance with artificial imagery
2. Summarize the findings of performance assessment to date with nighttime imagery

3. Emphasize the need for much more research that draws upon the expertise of the sensor engineers and scientists, on the one hand, and human factors psychologists and psychophysicists, on the other

In this introduction we also need to emphasize that this is an evolving and rapidly changing field with sensors and fusion techniques continually being altered and improved. As a consequence, reports of rigorous psychophysical testing will often lag behind advances in technology, reporting results with imagery of a method that has since been modified. The advantage to this however, is that it is beneficial to build up a body of psychophysical results with a variety of both sensors and methods that reflect different monetary and/or computational costs as these methods may find different applications. The data and images presented in this chapter should be viewed in this light. That is, they reflect the fusion method and imagers tried at a particular time and may not be necessarily the best end-products that a given lab can produce. In this regard, we thank all individuals who provided the images that we present in this chapter or who helped in other ways.

The chapter is organized into the following sections. In the remainder of this introduction, the types of nonliteral imagery and the processing of that imagery by the human visual system are characterized in general terms. The second main section of the chapter considers the nature of multiple-sensor systems. It includes the display of multiple-sensor images (both fused and non-fused), and addresses the complexities created by sensor fusion—including the potential loss of image information—and constraints placed on the effectiveness of fusion by the nature of human visual processing. The third main section provides an overview of some of the diverse approaches to fusing this type of imagery by various researchers. The fourth section addresses some of the special difficulties of testing perceptual performance with natural outdoor scenes. The fifth section catalogs the findings of studies conducting behavioral testing of perceptual ability with nighttime imagery and fused imagery, with the results grouped in subsections by type of perceptual ability evaluated: object recognition texture-based region segmentation, object/target detection tasks, preferences of human viewers, and miscellaneous tasks. The final section of the chapter offers a table summarizing the behavioral results, conclusions, and a look to the future.

6.1.1 Types of nonliteral imagery

Visual images from electronic sensors are encountered in a number of situations in contemporary life. They are seen routinely on television in the form of Doppler radar images on daily weather reports, and they are also seen in specialized situations such as MRI, ultrasound, or other types of medical imagery. One prominent use of nonliteral imagery is to allow visual perception when there is not adequate light to support human biological vision.

Two main types of alternative sensors have been used to assist human nighttime vision. One of these detects longwave infrared (IR) radiation, providing essentially a thermal image based on spatial heat differentials in the scene. The other, termed an image intensifier (I^2), detects primarily short-wavelength infrared light ("near IR") and relatively long-wavelength visible light (i.e., red, and to a lesser extent, orange), then amplifies the detected signal to create a visible display. Thus, the IR and I^2 devices rely largely on emitted and reflected electromagnetic radiation, respectively.

Images from IR or I^2 sensors, or images from their combination (i.e., fused images) are presently used in a number of applications to enhance night vision ability. Indeed, some commercial automobiles offer IR cameras and associated displays to help drivers look beyond the headlights when driving at night.[4,30,70] More common has been the use of IR or I^2 nighttime imagery by pilots or crews of helicopters, fixed-wing aircraft, and boats in military, Coast Guard, and police tasks to increase success of nighttime piloting, targeting, search and rescue, and surveillance.[9,13,47,58] In short, these images are used to increase situational awareness during night operations. In addition, IR sensors are also used by firefighters to perform visual tasks in conditions of dense smoke, and image intensifiers are used as visual aids in vision disorders involving rod dysfunction.[6]

6.1.2 Perceptual organization of nighttime imagery

Although the IR or I^2 images (see Figure 6.1) that are presented on the visual display offer a substantial improvement over the unaided human visual system in nighttime, these images are clearly inferior to unaided daylight vision or even to daylight artificial imagery presented on a visual display. When a human views any image, good visual performance is dependent upon the ability to make perceptual sense out of the image. This "perceptual organization" of a scene entails complex processing of the information that is conveyed to the retina by the scene (e.g.,[14,18]). For example, in the perceptual organization of any scene, the visual system must make use of correspondences of local properties in order to segment the image into regions and then organize the regions into meaningful objects.[5,14,25] In particular, image points with similar values of low-level image properties such as brightness/color and multipixel relations (texture) need to be linked together as an early step in perceptually forming regions and objects. The perceptual processes that form boundaries and regions are highly dependent upon the richness of low-level information in an image (e.g.,[5,14,18,40]).

When a sensor presents imagery that is degraded relative to normal daylight viewing, less low-level information is available in the scene, and satisfactory perceptual organization becomes an even more difficult task for the visual system. For example, when a scene is displayed in monochrome, as opposed to color, a strong cue for perceptual organization and preattentive processing is unavailable.[39] Similarly, information such as microtexture at

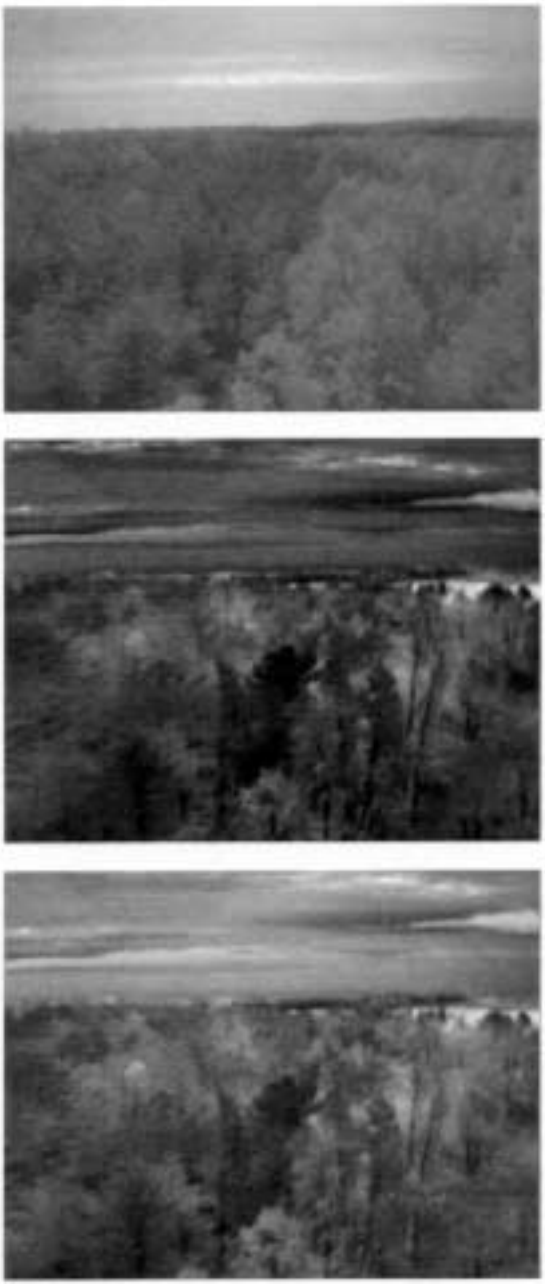

Figure 6.1 An example of a nighttime scene that has been imaged using a low-light, image-intensified CCD camera (top), an IR sensor (middle), and a gray-scale–fused version (bottom) where the two single sensor images have been combined (see text). (From Steele, P. M. and Perconti, P., *Proc. SPIE 11th Annu. Inter. Symp. Aerosp./Defense Sensing, Simulation and Controls*, 3062, 88–100, 1997. With permission.)

various spatial scales can also be lost in images produced via particular types of artificial sensors due to diminished contrast at various spatial scales. For example, small-scale spatial information can be lost due to the resolution limit of a given device and can also be lost in low-light images simply because of a low level of signal relative to the noise of the detectors. Thus, the imagery presented to the human visual system that displays the information from electronic sensors is not as rich as daylight imagery of natural scenes and therefore is expected to adversely affect the perceptual organization of the scene.

The ability of the human user to make good perceptual sense out of the resultant visual image from an artificial sensor is particularly critical in the case of sensor systems intended to improve human vision in real-world nighttime viewing conditions. Here, not only is a realistic rendering of a very complex natural environment desired, but the rendering must support the user in an interaction with the dynamic environment in which he is immersed. In comparison, medical imagery or imagery from satellites representing geologic or vegetative spatial information need not be as rich and realistic for successful use. With standard night vision technology, providing a crude spatial mapping on a display monitor is actually a simple matter. However, it is presently impossible to present a nighttime image on the monitor that allows the fidelity and richness of human visual perception in daylight conditions.

In an attempt to improve upon the perceptual utility of nighttime imagery, efforts have been made recently to combine imagery from multiple sensors. For several years, our group has been investigating the perceptual performance that is allowed by the display of spatial mappings of the natural environment obtained by single or multiple electronic sensors. Because human perception is very complex and multifaceted, defining the specific perceptual abilities to be evaluated in a naturalistic setting is a challenging endeavor. Furthermore, natural outdoor scenes are themselves extremely complex, making the design of rigorous psychophysical tests of perceptual ability even more challenging. In the following sections these issues are considered.

6.2 Multiple sensor systems

Because different sensors are sensitive to different wavebands of electromagnetic energy, they convey different information about distal objects. An imaging sensor that is optimal for one perceptual task, or under one set of environmental conditions, can be inadequate under different circumstances. For example, discrimination of details is difficult on IR imagery obtained near morning when different objects may have cooled to near the temperature of the background (“thermal cross-over”). Similarly, the reflectivity of two objects may be quite similar with respect to an image-intensifier sensor, or there simply may be little starlight / moonlight available to reveal their different reflectances. Because of the advantages and disadvantages of these sin-

gle sensors, it may sometimes be beneficial to provide human operators with the output of multiple sensor systems.

In an attempt to improve the perceptual effectiveness offered by images from a single sensor, researchers have presented images from more than one type of sensor to a user. This has taken the form of multiple displays viewed successively or simultaneously, or images from multiple sensors fused into a single display.

6.2.1 Presentation on multiple displays

One obvious way of presenting views of a scene from different sensors is to simply present the output from each sensor within its own display, and to allow observers to alternate between displays electronically or by shifting their gaze. Unfortunately, the usefulness of such a set-up might be mitigated in several ways by demands created by the need for observers to alternate fixation and attention between displays. Because renderings obtained through different sensors are likely to differ in gross characteristics such as mean luminance and contrast, shifts from one display to another can demand changes of the adaptive state of the visual system and entail transient but substantial decrements in visual performance.[47] Furthermore, because multiple displays must be spatially or temporally separated, their combined utility might be limited by the difficulty of remembering and mentally integrating their contents.[24,46]

An alternative, dichoptic presentation (simultaneous presentation of different stimuli to the two eyes), could avoid these problems by allowing observers to view two distinct displays concurrently, but this too could be problematic. Though observers under some circumstances can process and combine information from both of two dichoptically presented images,[67] dichoptic displays of longer than 150 msec duration allow the possibility of binocular rivalry.[3] This phenomenon, in which portions of each dichoptically viewed stimulus perceptually mask portions of the other,[23] can lead to the perception of a patchwork stimulus in which portions of each image are alternately visible and hidden. Worse, observers have no volitional or conscious control over which portions of the images are visible at a given instant and therefore have no way of ensuring that information from either source will be available when needed.

6.2.2 Presentation via fused images

A more effective method of providing users with information gathered through two or more sensors might be through the process of sensor fusion. Sensor fusion seeks to obviate shortcomings of individual imaging sensors by combining the output of multiple single-band sensors to form a single multiband image. Fusion algorithms take as input two or more single-band views of a scene obtained from a common vantage point, and from these produce a single composite image.

This manipulation offers users two general types of potential benefits. First, it could allow observers to view and access useful information from multiple images simultaneously, without needing to alternate gaze and attention between displays and without the complications of dichoptic presentation (which can be thought of as biological fusion). Second, fusion algorithms can exploit differences between single-band images of the same scene to provide a fused rendering enhanced with new, *emergent*, information that is not present within either of the component images singly. That is, objects in the world will register differently with respect to the properties measured by different sensors and comparison of these spatial mappings will create different signatures across the spectral bands, or contrast between sensors (i.e., any differencing operation between sensors). Information derived from such intersensor contrast might augment the spatial information conveyed by individual component images.[66,73,76] Furthermore, intersensor contrast can be taken as the basis for color rendering of an image, much as differences in output of the retinal cones provide the basis for human color vision. To the extent that sharp changes in contrast between single-band images tend to occur at objects' edges, the emergent information derived through fusion would then appear in the image as color contrast between image regions that correspond to different distal objects.

For several reasons, however, many of these potential benefits might not be easily realized. Ultimately, an image is useful to a human operator only to the extent that it is interpretable. An observer's visual system, presented with a sensor-fused rendering, must parse, or segment, the image into a representation of distal surfaces and objects, extracting structure and meaning just as it would from a natural image. The potential utility of a sensor-fused image, like that of a natural image or an electronically sensed single-band image, will therefore be constrained by limitations in the observer's perceptual system. Information present within a sensor-fused image, and thus theoretically available to mathematical pattern analysis, could in actuality be of little value to a human observer. Emergent information, for example, might be of low salience and therefore of little perceptual utility even for trained observers. Worse, single-band information that is salient within a component image might not be easily perceptible within a fused image. Features of one input image might instead obscure details of the other, leaving the information conveyed by one single-band source degraded or even imperceptible within a dual-band rendering. Paradoxically, fusion could then impair visual performance. Finally, there is nothing to assure that sensor fusion, even if helpful under some circumstances, will be equally beneficial to visual performance in other circumstances. Visual perception is not the result of a single, monolithic process or mechanism, but represents the working of multiple channels and processing modules operating for various purposes on various aspects of a stimulus.[12,19,31,35] Thus, it is possible that sensor fusion might aid some aspects of visual performance while hindering others. Clearly, important issues are raised for human factors engineers in this realm.

6.2.3 Complications raised by sensor fusion

The aim of sensor fusion is to receive as input two or more images of a common scene and from these create a unitary, composite rendering. In principle, a fusion algorithm might receive input from any variety of sensors. For several reasons, however, research into sensor fusion has largely focused on the possibility of melding nighttime I^2 and IR imagery.[50,66,73,76] First, because I^2 and IR night vision devices are now in widespread use, algorithms for fusing their output could find diverse and ready application. Second, and more important, I^2 and IR sensors together provide information which might be optimal for exploiting sensor fusion.[73,76] As noted above, image intensified sensors respond to long-wave visible and short-wave IR light, typically collecting energy that has been reflected off the surfaces within a scene. Long-wave IR sensors, conversely, respond to thermal radiation, collecting energy emitted by objects within a scene. Figure 6.1 shows a nighttime outdoor scene imaged by both an I^2 sensor (see Figure 6.1, top) and an IR sensor (see Figure 6.1, middle). As seen in the figure, the information provided by these two classes of sensor is largely complementary, indeed formally so, related by Kirchhoff's Law, $\rho(\lambda) = 1 - (\lambda)$, where ρ is the spectral reflectance and is the spectral emissivity. Thus, an algorithm that effectively combines the sensors' output could create imagery that is of considerably enhanced utility by capitalizing on the differences of objects in terms of their emissivity / reflectivity ratios.

6.2.3.1 Potential problems with fused imagery

There are many ways in which sensor fusion could fail as an aid to human perception, however. Some of these are simple technical shortcomings which might eventually be overcome by developments in the construction of imaging systems. Sensor-fusion researchers, for example, have typically faced the difficulty of aligning or registering pixels in the single-band images to be fused. Because of various optical distortions, even sensors with matched fields-of-view can generate images that are not precisely registered spatially (see description in ref. [53]). Before image fusion is possible, therefore, preprocessing is generally necessary to map pixels in one image onto those pixels in the companion image that represent the same distal points. That is, one image must be stretched ("rubber-sheeted") into alignment with the other. This distortion can cause noticeable degradations, and can produce fused imagery that is aesthetically displeasing and disruptive of visual performance.[29] The problem of misregistration between component images, however, is being mitigated by the development of methods that allow sensors of different spectral sensitivities to be arranged vertically with perfect pixel registration[38] or directly by matching field of view, pixel number, and optical path.[1,74]

Other difficulties, however, are more fundamental and will not be so easily conquered. It is impossible to achromatically fuse nonidentical images without loss of physical image information. Within an achromatic image,

pixels are constrained to vary along a single dimension. Achromatic fusion thus entails mapping multiple one-dimensional spaces of single-band pixel values onto a single one-dimensional space of multiband pixel values, and ensures that information within a fused image will be insufficient for full recovery of individual input images (thus, the task in achromatic fusion is to ensure that it is the perceptually less-important information that is lost; a task requiring considerable empirical psychophysical testing, as described in later sections). This loss of physical information can only be avoided through the use of a chromatic display rendered in a color space whose dimensionality is equal to or greater than the number of images being fused, such that one dimension can be dedicated to represent each input image (or transformations). Furthermore, such a display, would augment single-band stimulus information by making differences between component images explicit as chromatic contrast within the composite display.

6.2.3.2 Constraints due to the human perceptual system

It is at this point however, that characteristics of human visual perception become constraints on the utility of sensor fusion. One potential impediment to the use of color-fused imagery as an aid to human performance arises from the difficulty of achieving seminaturalistic color renderings from sensor-fusion. Because single-band images to be fused will generally be collected with sensors whose spectral sensitivities differ from those of the human photoreceptors, color-fused scenes will typically be rendered in colors different from those in which they appear during unaided daylight vision. To the extent that visual perception exploits stored color knowledge for recognition or other purposes, fused-color imagery might therefore be disruptive of visual performance. Fortunately, evidence indicates that the role of stored color knowledge in vision is secondary to the role of color in image segmentation, i.e., in the process of delineating objects' edges within the image [7,80], suggesting that the benefits of false-color rendering relative to achromatic rendering may outweigh its costs. Resolution of this specific issue requires future research specifically weighing costs and benefits of color rendering.

Another limit on the scope of sensor fusion, however, is clearer and more obviously insurmountable. Because the human visual system is trichromatic, the color space in which images are fused can be of no greater useful dimensionality than three. Thus, no more than three images can ever be fused directly and presented for analysis by a human observer without loss of physical information. Furthermore, the chromatic discrimination ability of the human visual system imposes an even more subtle limit, dictating that fusion of even three or fewer images will generally entail functional loss (i.e., perceptual, as opposed to physical, loss) of single-band information (see also ref. [22]). This is because equal physical steps in a physical "color" space are not equally discriminable to a human observer. Thus, the mapping of physical differences onto perceptual color space is nonlinear, and if performed blindly,

could serve to make important physical differences *less* discriminable when rendered in false color.

Ultimately, the quality of a sensor-fused rendering is not measured by the information contained within the image, but rather by the *useful* information, in the sense of human factors constraints, that is truly conveyed by the image to an observer. Limits of visual acuity, of contrast and chromatic sensitivity, and of higher-level perceptual organization ensure that not all of the information within an image is perceptible. Indeed, it is unlikely that observers could perceptually recover component images from a composite rendering of a single scene[8], even if it were mathematically possible to do so. Thus, sensor fusion algorithms face the challenge of overcoming limits of human perceptual processing by selectively preserving and enhancing the single-band information that is necessary for accurate visual perception.

Fortunately, for a sensor-fused image to be useful, it is not essential that its component images be fully recoverable. Rather, it is important that the task-relevant distal structure perceptible within the component images remain visible, or become more visible, in the fused image. An observer typically has no need of reconstructing component images pixel-by-pixel from the information within a fused image, but must only recover the texture and contours necessary for perceptual organization of the depicted layout of surfaces and objects. For this purpose, much of the information between and within single-band renderings of a common scene will be redundant and thus, at least to some extent, expendable (i.e., in a physical sense). The goal of a sensor-fusion algorithm must therefore be to create a multiband image that *at worst* preserves information necessary for human perception of the distal structure visible within the single-band renderings, and that *at best* derives new information to improve veridical perception of the distal stimuli depicted in the imagery. In addition to this physical-level definition of information, perceptual/functional considerations abound[26] (e.g., issues of user training and experience with particular display formats, the use of particular colors or redundant information coding schemes, and the familiarity with nighttime imagery as well as certain types of scenes).

There have been several approaches developed to attempt to meet this goal of creating improved sensor-fused imagery. To give the reader a sense of the different approaches to fusion of night vision images employed to date, this literature is characterized in the next section.

6.3 An overview of fusion approaches

A variety of fusion algorithms have been developed (e.g.,[45,50,53,59,64–66,71–77]), and the diversity of the methods being developed indicates that no single ideal fusion method has been established. In this section, several fusion methods employing different basic approaches are described to provide an overview of the types of methods being pursued.

6.3.1 *Monochrome fusion*

The most direct method of fusion involves combining an image from each of two different sensors (e.g., IR and I^2) into a single composite grayscale image. For example, Toet and associates[60,61,65] developed a monochrome fusion method that uses a hierarchical image-merging scheme based on a spatial decomposition of the original imagery performed at several spatial scales. Thus, the original single sensor imagery is first decomposed into a contrast-based mapping at several spatial scales represented in a multiresolution pyramid. The amount of contrast present at each scale is compared for each single-sensor image, and the single-sensor image with the greatest contrast at that scale at a given location is chosen for the fused image. Thus, the fused image is a construction made from local patches of information in one or the other of the two sensors' images. Peli et al.[43] report a similar type of stratified fusion in which the fusion is based upon contrast calculated locally within spatial frequency bands[41]. An example of this type of monochrome fusion[43] is shown in Figure 6.2 (bottom) along with the initial IR (middle) and I^2 (top) images. Indeed, examples of image structure present only in one of each of the two single-sensor images can be seen in the fused image. Fusion stratified by spatial scale (i.e., spatial frequency bands) also occurs to some extent in the algorithm of Therrien, Scrofani, and Krebs[59], using two scales. Numerous other methods of monochrome fusion have been developed (e.g., [44,48,68,73]; also see ref. [33], for a review).

The approach of Waxman, et al.[1,2,71–76] differs, emphasizing intersensor differences (intersensor contrast) instantiated by shunting center/surround filter mechanisms, although center/surround filters also can be conceived of as the application of a band-pass filter as is done in the multiscale filter methods just noted. An example of results with this type of monochrome fusion[73] is shown in Figure 6.1 (bottom) resulting from the combination of the IR and I^2 images shown in Figure 6.1. Again, notice how the fused image appears to retain much of the image structure present in both of the component single-sensor images. It should be noted that this particular fusion method is best viewed as an intermediate step in the creation of a type of color-fused imagery.[73]

6.3.2 *Color fusion*

In addition to achromatic fusion, several labs have developed methods to perform fusion in either two- or three-dimensional color space (e.g., [45,50,66,73,76,77]). An obvious way to perform color fusion would be to simply send one image (e.g., IR), as the input to one of the display monitor's guns (e.g., red), and a second image (e.g., I^2) to another of the monitor's guns (e.g., green, or even green and blue jointly), thereby allowing the monitor's array of RGB phosphor dots and the human visual system to fuse the two monochromatic images into one color image. In a sense, this is another example of biological fusion, but instead of using the binocular neural pathway to biologically fuse dichoptic

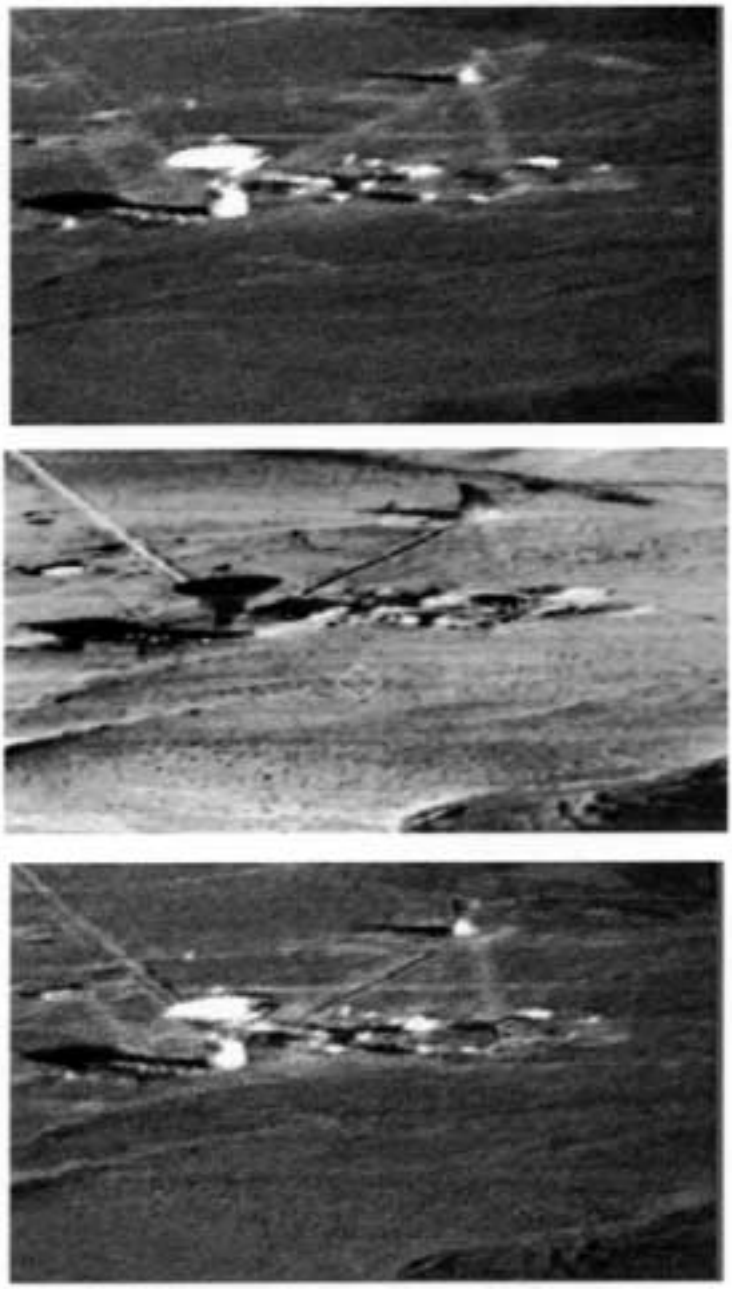

Figure 6.2 A nighttime scene of a large satellite dish, roads and various other structures. The normalized visible image (top), IR image (middle), and the gray-scale–fused version (bottom) are shown. (From Peli et al., *Proc. SPIE Conf. Sensor Fusion: Architecture, Algorithms, and Appli. III*, 3719, 1999. With permission.)

images, the neural color pathway is biologically fusing single-sensor maps in perceptual color-space. Several methods go beyond this type of direct (biological) color fusion to extract intersensor differences and to display them in a particular color format.

One method of color fusion, developed at the Naval Research Laboratory (NRL), is to create chromatic dual-band imagery by mapping pixel values from single-band I^2 and IR images into a two-dimensional space.[53] The output of this method applied to the I^2 and IR images shown in Figure 6.1 is shown in Figure 6.3 (top). Note that the fused image reveals a chromatically salient horizon while preserving other structure in the images. This method utilizes an intensity axis and a single color-axis (corresponding to red/cyan of various saturations) wherein intensity in the fused image at a given image point is determined by the sum of the two single-band intensities at that point, and the chromaticity is determined by the difference between the single-band intensities at that location. That is, saturation on a red/cyan color axis encodes intersensor contrast. This method uses a principal components procedure to enhance chromatic contrast by normalizing pixel values in the direction orthogonal to the principal component direction when plotting the raw I^2 value of each pixel against the raw IR value.

By using red and cyan as color primaries, the fusion algorithm implements a form of color cancellation, creating a flattened, or two-dimensional, color space (i.e., with chromaticity based on saturation of one pair of complementary colors). Rather than producing a variety of colors, colors in this type of imagery range from saturated red through gray to saturated cyan; pixels that are bright only in the IR component image appear red in the fused image, pixels that are bright in only the I^2 component image appear cyan, and pixels whose values are approximately the same in both component images appear achromatic.

Several researchers have taken two-band color fusion a step further utilizing a three-dimensional color space. For example, a method developed by Werblin and associates at the University of California, Berkeley (UCB)[77] combines single-band images through processing similar to that represented in their computational model of retinal processing, emphasizing the spatial and temporal diffusion of local neural responses (e.g., the time-course of retinal center/surround receptive-field spatial antagonism).[78] An example of this method's fusion is shown in Figure 6.3 (middle) in which it has been applied to the I^2 and IR images like those shown in Figure 6.1. Note that the resultant image shows good preservation of image structure from each single-sensor image. The algorithm first provides single-band image enhancement, and then fuses the enhanced single-band images within a three-dimensional false-color space. The initial filtering incorporates gradual diffusion of gray-scale values within homogeneous regions of the image, thereby "growing" regions in conjunction with region boundaries (Das, personal communication, 1997). A similar concept is seen in the method of Waxman et al.[73,76] incorporating region growing implemented by Grossberg's BCS/FCS model of human

Figure 6.3 Three color-fused versions of the nighttime scene shown in Figure 6.1. Fusion by the NRL method (top), the UCB method (middle), and the MIT-LL method (bottom). See text for details of fusion methods. The color images from all three methods appear to lose some of the detail apparent in the image-intensifier image of Figure 6.1. See color version of this figure in the color section following page 114. (The UCB image was provided by Frank Werblin and the NRL and MIT-LL versions are reprinted from Steele, P. M. and Perconti, P., *Proc. SPIE 11th Annu. Inter. Symp. Aerosp./Defense Sensing, Simulation and Controls*, 3062, 88–100, 1997. With permission.)

boundary and region formation.[20] In both methods this region growing serves to produce better segmented imagery with regions better filled in.

The color rendering in Werblin and associates' method also occurs through sending the output of the sensors differentially to the R, G, and B inputs of the color display monitor. In this method, the R, G and B values of each pixel are set to the corresponding pixel's intensity in the IR image (for the R input), the intensity of the corresponding pixel in the I^2 image (for the B input) and a combination of the intensity of the corresponding I^2 and IR pixels (for the G input). A considerable potential advantage of this method based on the analog processing within the retina is that it is being developed for implementation on an analog chip to support real-time processing (Kozek, personal communication, June 1999).

Very rich color rendering has been obtained by spreading the color mappings across a larger volume of three-dimensional perceptual color space. Although the earlier version of Werblin and associates' procedure produced a color mapping onto a rather flat three-dimensional color space, a more recent version produces a fuller three-dimensional color mapping. Figure 6.4 shows an example of this in which a scene, imaged by an IR sensor and a light (i.e., daylight) sensor, have been fused. The scene shows the use of daylight/IR sensor fusion for nighttime use when artificial lighting is present (e.g., automobile or building lights). Note how objects with different combinations (spectral signatures) of light and heat values (e.g., the person, window, trees and building) are distinguished in the color coding (bottom right) and to a lesser extent, in the monochrome fusion (top right) compared to the single-sensor images.

Another method of obtaining a richer color rendering is to employ three sensor bands of imagery, typically I^2, mid-wavelength infrared and long-wavelength infrared sensors types (e.g., ref. [52]). A method by Scribner and colleagues[52,53] uses images in these three bands as the input to the three color channels of the display, following conjoint application of the principal components scaling, to yield direct fusion of three-band imagery in three-dimensional color space with minimal processing. Figure 6.5 shows an example of this method in which long-wave IR, mid-wave IR and visible images have been fused. The scene contains a Jeep near shrubs before an agricultural field, with the Jeep, particularly the windshield of the jeep, most apparent on the fused image. Note that the prefusion imagery is of lower resolution than in some other examples (e.g., see Figure 6.4).

Two related color-fused methods that use this same general approach to fusion have been reported by Toet[64] which rely on splitting the output of a visible camera (400–900 nm) into a band of shorter wavelengths of the visible spectrum and a band of longer wavelengths. For one color-fusion method, the short wavelength image is used as the R input, the longer wavelength image as the G input, and the gray-scale image (from multiresolution pyramidal fusion) is taken as the luminance component in NTSC color space. The color-fused rendering is displayed after transformation to RGB color space.

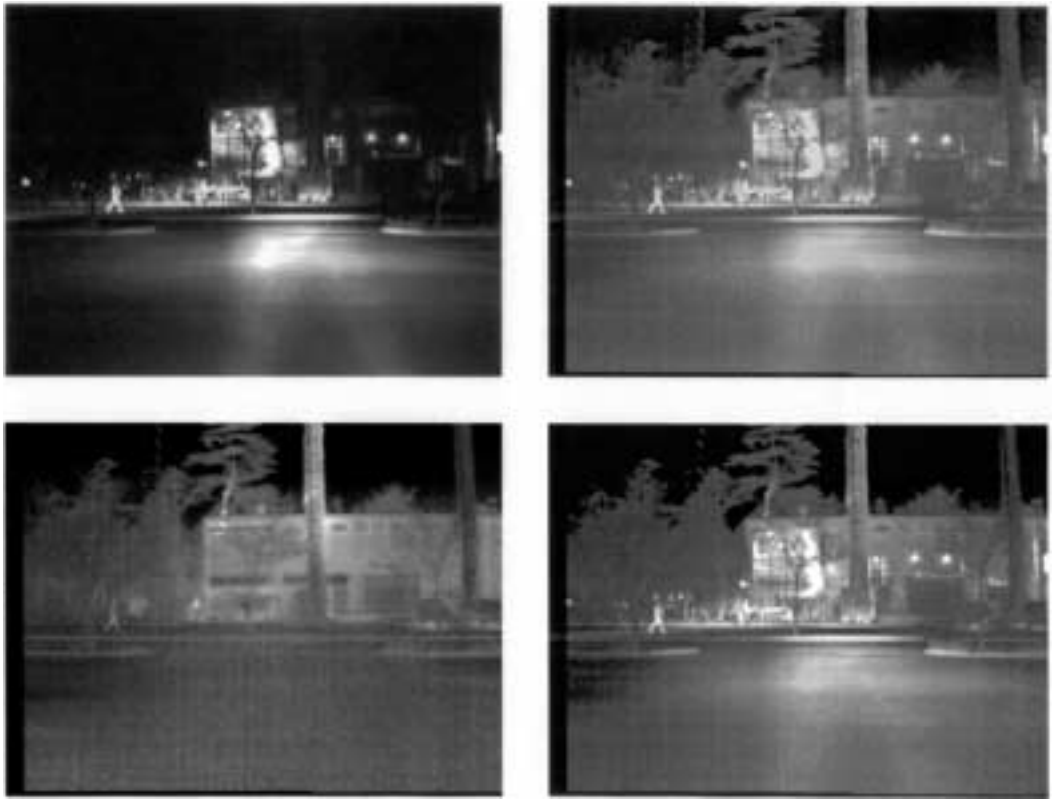

Figure 6.4 Nighttime imagery taken with a visible daylight camera (top left), an IR camera (bottom left), and the sensor-fused versions of the same scene (gray-scale fusion top right, color fusion bottom right). See color version of this figure in the color section following page 114. (Images courtesy of Frank Werblin.)

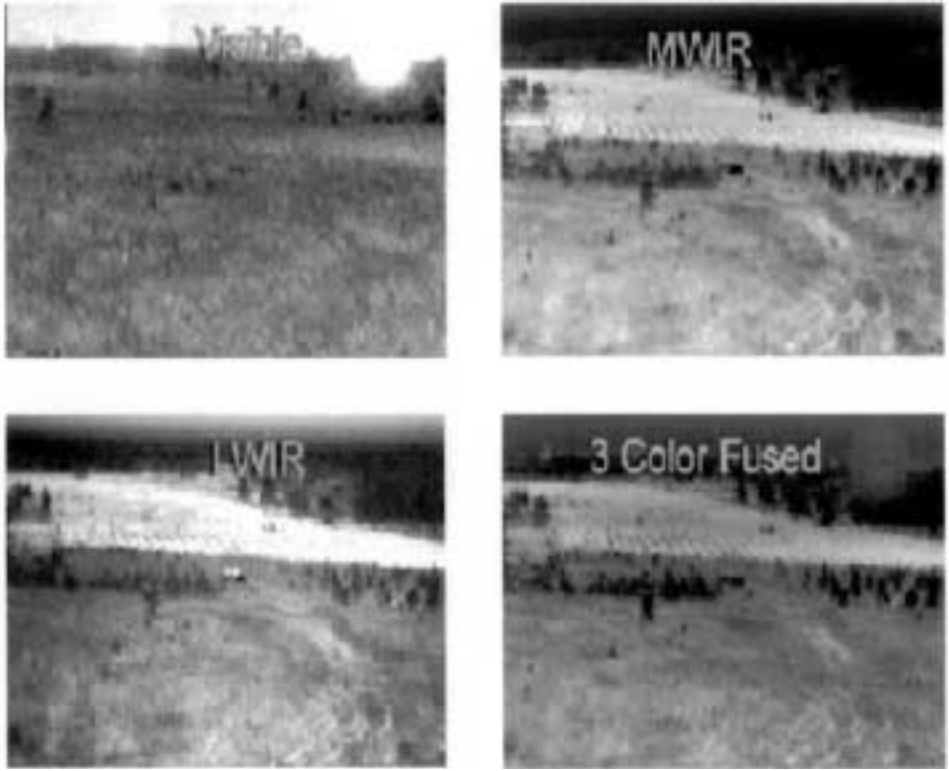

Figure 6.5 An example of color-fused imagery created from three input bands rather than two. The visible image is shown on the top left, the long-wave IR image on the bottom left, and the mid-wave IR image on the top right. The three-color fused image created from these three sensor bands is shown at the bottom right. See color version of this figure in the color section following page 114. (Images courtesy of Dean Scribner.)

The second method uses the IR image as the R input, the longer wavelength image as the G input and the shorter wavelength visible image as the B input and thus provides a type of fusion similar to three-band fusion.

Alternatively, one may use dual-band fusion with a processing method that more explicitly extracts new information from the comparison of the two images to provide the input for the third channel. For example, a method by Waxman and associates[1,2,72–76] developed at MIT Lincoln Laboratory (MIT-LL) provides rich false-color rendering (i.e., a less flattened three-dimensional perceptual color space) from dual-band imagery. An example of color-fused imagery from this method is shown in Figure 6.3 (bottom) in which the I^2 and IR imagery of Figure 6.1 has been fused. The fused image is particularly notable for the realistic colors appropriate for an outdoor scene. In this method, a center/surround shunting neural network is first used to enhance and to normalize image contrast and also to form the three center/surround spatial filters that provide the input to the three color channels (R, G, and B). In their initial reports, Waxman and associates produced the false-color images by using as R, G, and B input the output of the center/surround filtered IR image, the center/surround filtered I^2 image, and a third channel produced by contrasting I^2 input to the center with IR input to the surround. This third channel conveys intersensor contrast and also provides for a perceptually rich (i.e., not flattened) three-dimensional color space by providing a third input channel and it also can serve as a gray-scale–fused image (shown previously in Figure 6.1, bottom).

In more recent developments (e.g., ref. [1]), this group bases two of the three channels on intersensor contrast and one on the image-intensifier image; with each of the center mechanisms receiving I^2 input (from a single pixel in size) and the antagonistic surrounds contrasting this center input with either the I^2, the positive-polarity IR, or the negative-polarity IR signal. The resultant images are passed to the G, R, and B monitor inputs, respectively, to complete the color fusion after color remapping is performed to make the colorations look more realistic (i.e., vegetation colored green). A variation of this has been reported[62,63,72] for use when the IR image is of a resolution comparable to the I^2 image, in which case three differenced inputs are used. The authors[71,73,76] relate this method based on intersensor center/surround opponency to the spatial center/surround color-opponency shown in the primate visual system and the IR/visible mutual inhibition shown in the pit viper sensory system.

Other methods of obtaining dual-image color fusion based on intersensor contrast have also been developed.[62,66] An example of output from one of these methods[62] is shown in Figure 6.6 illustrating fusion of images from a visible-light CCD camera and a mid-wave IR camera. The TNO color-fusion algorithm of Toet et al.[62,66] is based on differencing operations and is in this sense, of the same genera as the method of Waxman et al.[71,72,74,75]. In this method, first the minimum intensity value of the two images (CCD and IR) at each pixel is selected, resulting in a map termed the “common component”

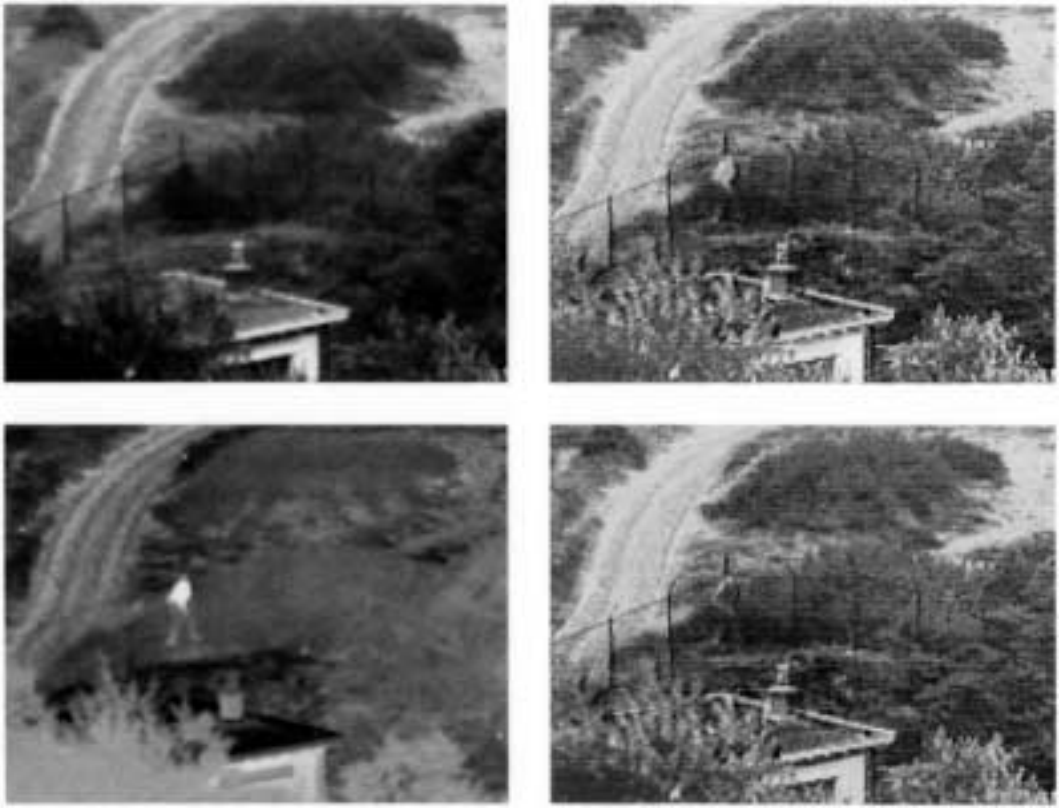

Figure 6.6 A nighttime scene taken with a visible camera (top left), an IR camera (bottom left), and the sensor-fused versions of the same scene (gray-scale fusion, top right; color fusion, bottom right). See text for details. See color version of this figure in the color section following page 114. (Images courtesy of Lex Toet.)

of the two images. Next, this common component is subtracted from both the CCD image and the IR image at each pixel to obtain the "unique component" of each sensor's image. Finally, the unique component of each sensor is then subtracted from the *other* sensor image and sent to the R input (IR − unique CCD) and the G input (CCD − unique IR). The B channel gets the differenced original images (CCD − IR)[63] or differenced unique images (unique CCD − unique IR).[66] (Gray-scale fusion is taken as a weighted sum of the three monitor inputs resulting from the color-fusion procedure.)

It is noteworthy that the concept of intersensor contrast (or differencing) is quite common in the satellite remote sensing literature (e.g., the "normalized difference vegetation index"). Thus, the areas share similar issues and hopefully this literature will become better integrated in the future, leading to more cross-pollination of ideas.

Finally, an alternative approach for the utilization of color in fused imagery is offered by Peli et al.[43] in which color is used as a means of increasing the distinctiveness of certain specific regions thought likely to be targets, somewhat analogous to "painting" targets with a salient color in a monochrome-fused image.

We anticipate that efforts for future improvements in the field of color-fusion in the near-term are likely to focus on several key issues. Possibly the most important is to support full processing and fusion in real time. Currently, simpler fusion methods, for example mapping three sensor bands with minimal processing directly onto R, G, and B inputs, can be performed in real time, but extremely computationally intensive methods cannot. Great strides in real-time processing are being made.[2,52] It would seem that hardware developments such as specialized processors, dual-band IR focal plane array sensor packages that obviate registration and scaling, and algorithms that are optimized for speed will contribute to this in the near future. Real-time processing will also lead the researchers to a consideration of temporal noise reduction and temporal processing aspects. Other issues likely to be investigated are those surrounding the color remapping that determines the specific coloration of particular types of object. Issues include whether use of more natural, as opposed to "unnatural," scene colorations are more useful in a functional (perceptual) sense; whether unique coloration of a particular type of scenic characteristic can offer a performance advantage; or whether false-color mapping is useful at all (cf. ref. [22]). A related issue concerns how to minimize the change in coloration from moment to moment as an object changes its reflectivity or emissivity (e.g., as when a helicopter passes through mist or fog) and how to minimize the perceptual consequences of this. Investigation of experience and training with color-fused imagery will surely be an active area. Another area in which we anticipate considerable efforts is the use of a greater number of spectral bands and research into the conditions in which various combinations of narrow or hyperspectral bands are particularly useful (e.g., [52]). Finally, we anticipate an increase in the evaluation of all of these hardware and algorithm implementations in *functional*

terms, that is, the human factors assessment of perceptual performance as described in the next section.

6.4 *Psychophysical testing with natural scenes*

Typically in vision research, simple stimuli that can be exactly specified and varied on a single dimension are used for testing visual performance.[12,69] For example, a typical stimulus is a patch of a sinewave grating that might be varied in either orientation, spatial frequency, or contrast. Common stimuli for evaluating, visual search, for example, might consist of an array of identical bars with the target differing only in bar orientation or color. In the case of a texture segmentation test, a region containing bars of a second orientation or second color might be physically grouped within the field of bars of the first type.

Natural scenes, on the other hand, are incredibly complex compared to these simplified test patterns typically used for evaluating perceptual ability. Images of natural scenes contain regions such as grass or trees that are nominally or symbolically unitary, but are actually highly variable in terms of texture composition or local luminance/color composition from one part of the symbolically uniform region to another. That is, relative to the variations of stimuli typically tested in psychophysical experiments, regions of natural scenes that are nominally uniform (e.g., grass) are actually highly diverse. Specifying the texture present in only a small region of a natural scene is an extraordinarily difficult problem, approached in myriad ways in the computational vision literature.[14] (Even if viewed in the frequency domain where the power spectrum of natural scenes typically falls off linearly, exceedingly complex relations exist between the phase and power spectra as well as orientation.) Simply finding a suitable way to define contrast of a natural image is a challenging and controversial problem.[42] Due to this complexity, testing a particular perceptual ability with natural scenes may seem to be straightforward but is actually a quite complicated problem and one that is fraught with numerous potential pitfalls.

As an example, consider a test of whether the presence of an object is better detected by a human viewer when an IR or I^2 sensor is used to image the same scene. To have a reliable psychophysical measurement, the dependent measure must be based on a large number of trials, and hence requires a number of different stimuli so that the observer does not always know when and where the test object will appear. One approach is to manipulate the image using software in order to cut out a "target," say a truck, and paste it at different locations to make alternative stimuli. The problem that this introduces is that artificial edges or borders may be created between the cut-and-pasted target and its new background, with respect to some property (e.g., texture or even luminance). Just by chance, the target patch might be highly detectable or undetectable, depending on where in the image it was placed, and so the performance measured can actually be determined by

factors unrelated to the original quality of the image. Such edge-effect confoundings can be very pronounced perceptually. Indeed, even an apparently small texture difference makes a highly salient emergent edge around the patch, similar to a subjective contour. In this example, the extent to which the IR or I^2 imagery conveys the background property (e.g., variations of grass) determines how salient the emergent edge surrounding the target is, and hence governs performance rather than performance being governed by how well the target is conveyed by a given type of sensor.

Another problem associated with comparing the utility of alternative sensors is that even if one target (again, say, a particular truck in the scene) "pops out" perceptually and yields high detection performance on one type of image (say IR), this alone is no indication that that image type mediates better performance for other than artifactual reasons. For example, if a truck is cut from one image and pasted into an image from a second sensor to serve as a visual target, the truck could be highly detectable for the very reason that the quality of the imagery is so poor. As an extreme example, imagine an IR image of a grassy field that is so poor that all vegetation is a uniform value (say black) with no texture details apparent at all. On such an image, the pasted truck would be highly detectable, mediating superior test performance for the very reason that the imagery is so poor! However, it would require additional testing with a different target stimulus (e.g., grass), or test of a different perceptual ability, to reveal that the imagery supporting "good" performance is actually of extremely poor quality.

Evaluating perceptual performance with real-world scenes, or even naturalistic imagery,[79] requires special efforts to minimize confoundings (i.e., it is not even clear that confoundings can ever be avoided entirely). That special efforts to minimize confoundings are required is especially true when alternative imaging methods or sensors are to be compared because the experimenter must be concerned about equating several images.

In our work, we have tried a variety of methods to avoid problems such as these. To avoid the spurious interactions at the target's boundary, we have often cut patches from the images (i.e., from the I^2 and IR images at the same pixel coordinates) and presented the target patch on a uniform background field rather than on the imaged scene itself. We have presented patches to be identified in this way, singly on a uniform gray field[15,16], and we have also presented multiple small patches simultaneously on a uniform field to test texture-based grouping.[17,57]

To make sure that the target patch cut out at a particular *x-y* location from an image of one sensor type does not have any unique properties that make it stand out from other patches cut out at the same location on other sensor types (e.g., the mean brightness of a truck patch from different sensors), we have used target *categories* (e.g., vehicles or buildings) with the exemplars in the target categories and those in the nontarget stimulus set intentionally chosen to be heterogeneous.[15,16] That is, by having a set of target patches that vary considerably in terms of luminance and an equally broad overlapping

range of luminances for the nontarget patches, potential artifacts caused by differences from one image to the next in the portion of the natural scene selected to be a target can be minimized. Since it is essentially impossible to equate all properties of a natural image, this method seeks an approach that essentially randomizes the difference across *groups* of test stimuli and does not try to equate a single natural-scene target patch on all possible image properties. As summarized in the next section, several labs are currently working on other ways to improve psychophysical test methods for testing various aspects of perceptual performance with remotely sensed imagery of natural scenes.

6.5 Psychophysical experiments

The concerns discussed earlier raise at least two general questions that psychophysical study of sensor-fusion might address. First, what single-band and intersensor emergent information does a particular sensor-fusion algorithm convey? Specifically, does information provided by one component image perceptually dominate that provided by another and is the emergent information within fused images perceptible? Second, how useful is the information within a sensor-fused image for different aspects of human visual perceptual performance? Since sensor fusion can aid perceptual performance either by presenting multiple sources of single-band information concurrently, or by deriving information not available within component images singly, a single demonstration that fusion influences visual performance may provide ambiguous implications for development of fusion algorithms. A measure of both the perceptible single-band information and the perceptible emergent information conveyed by the sensor-fused imagery might eliminate this ambiguity, lending greater meaning to a finding that sensor fusion affects image quality in the sense of determining which aspects of the fusion algorithm are beneficial and what aspects are not. Because of the number of diverse fusion algorithms currently in use and because of the many aspects to visual performance, these are both important questions. With these questions in mind, we review the results of perceptual testing with night vision real-world imagery. In the next four sections we review studies of object recognition, texture-based region segmentation, object/region detection, and observer preferences.

6.5.1 Object recognition

We began our own psychophysical tests of human performance with nighttime imagery using a recognition task that we developed.[10,15,54] Following the reasoning described above, this task was designed to use a set of target patches (e.g., buildings) and a set of nontarget patches (shown in Figure 6.7(a) for the category “buildings”), making sure that both sets overlapped in terms of mean and maximum (local) luminance, contrast and spatial frequency power

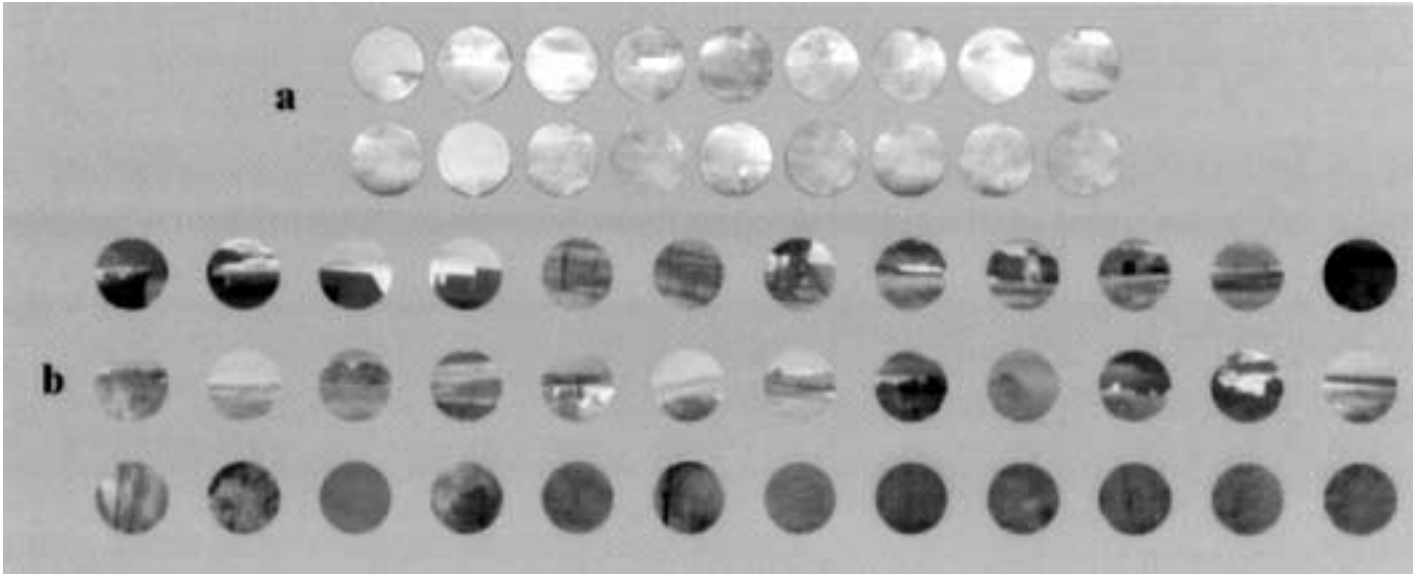

Figure 6.7 Patches of nighttime imagery used in the target category recognition task are shown. Part "a" shows patches used for the "buildings" target category for the I^2 imagery with target patches in the top row and nontarget distractor patches in the second row. Reprinted with permission from Essock et al., 1999, Human Factors Society. The three rows of panel "b" (from top to bottom) show the members of the target categories, "man-made objects," "sky / tree horizon" and "trees" for IR imagery from a similar category recognition task. (Part "b" imagery courtesy of Army NVESD and ONR.)

spectrum.[16] In one of our first studies [15,16] we compared the ability of human observers to perceptually organize low-level pixel and texture information in the 1.4° regions of the nighttime scene into recognizable patterns, specifically as examples of the categories of buildings, people, or tree / sky horizon, for IR, I^2 and, color-fused images. The imagery was obtained with a medium-resolution IR sensor and a Gen III I^2 sensor, and was color-fused by the earlier method of Waxman and associates.[73] Images were presented for a brief amount of time (100 msec) to preclude eye movements so that different patterns of eye movements and fixations could not be elicited by the different image types. Duration of visual processing was controlled and equated by use of a noise mask following the stimulus. Performance was measured by d', a criterion-free estimator of sensitivity calculated from hits and false alarms.[34] We found clear evidence of improved perceptual performance for color-fused imagery with this particular imagery and task.[15,16] The results, shown in Figure 6.8, indicate that the ability to recognize an image patch as containing an exemplar of the target category was significantly better with color fusion than for imagery from either of the single sensors.

Subsequently we used this task with other imagery that had been obtained with a higher-resolution IR sensor and other fusion methods: one gray-scale and three alternative color-fusion methods (see ref. [58] for sensor details and Figures 6.1 and 6.3 for examples of the imagery). These

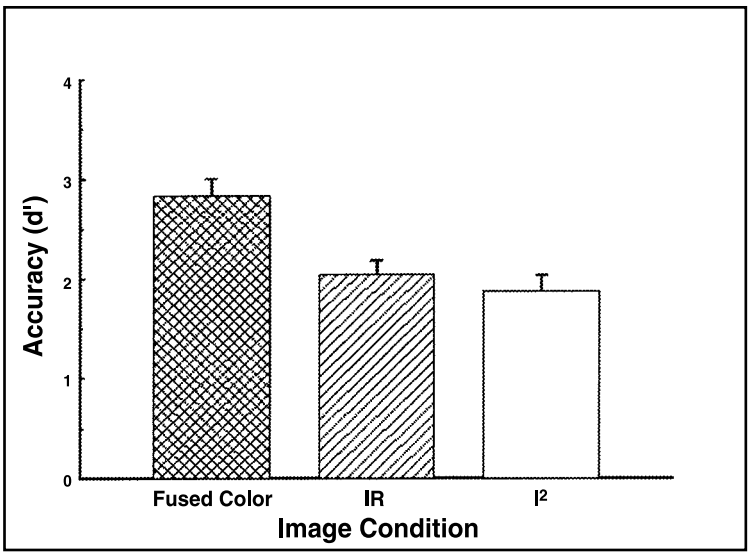

Figure 6.8 Performance for a region recognition task comparing overall performance with I^2, IR, and color-fused imagery. Results from various target categories were averaged together in this graph. (From Essock, E. A. et al., *Hum. Factors*, 41, 438–452, 1999. With permission.)

studies[10,54] used several additional target categories. They also increased the variability of the nontarget patches by drawing the distractors from the patches of the target categories not being used as target categories. Three target types, man-made, horizon, and trees, are shown in Figure 6.7(b). Again, we found that color fusion can improve human perceptual performance relative to monochrome fusion for certain types of targets, and that monochrome fusion can improve performance relative to performance with single-sensor images. The results are shown in Figure 6.9 for I^2, IR, gray-scale–fused, MIT-LL color-fused,[72–75] NRL color-fused[50–53] and UCB color-fused[77] image types.

These results were obtained for targets defined by contrasting regions, for example road next to grass or trees next to sky. However, while one method of color fusion may improve performance over that for single-sensor imagery, another method of color fusion may actually hurt percep-tual performance, underscoring the fact that sensor-fusion will not necessarily avoid the loss of perceptually important image structure. Using the same imagery, Steele and Perconti[58] also found this to be the case. Furthermore, they reported that while one color method (MIT-LL) mediated more accurate perceptual performance on one type of task (consisting of queries about objects and locations), the other color method tested (NRL) was better on another type of task (queries about geometrical relations). Using the same region recognition task developed previously,[10,15,54] Toet et al.[64] also report that performance is

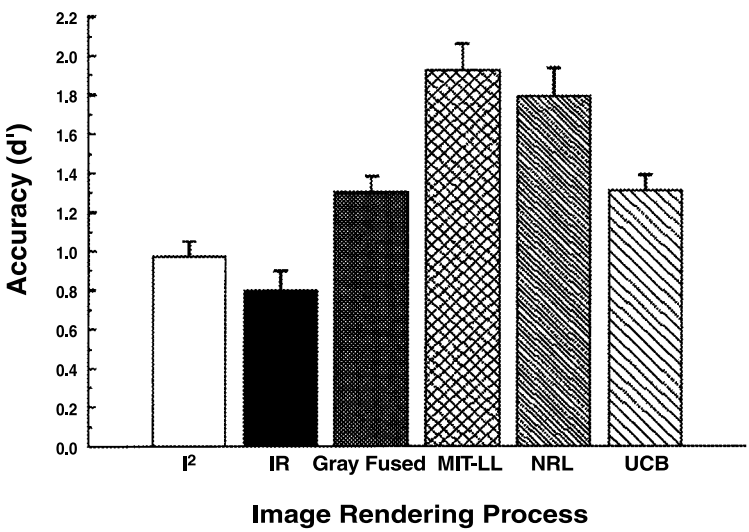

Figure 6.9 Region recognition performance for a second experiment using I^2, IR, and four types of fused imagery (see text). Data reflect average performance for target categories that consist of patches containing two contrasting regions within them, e.g., road next to grass or trees next to sky.

better with one color-fusion method for some region types, but better with another method for other region types. Indeed, another study[62] concluded that color rendering per se adds nothing beyond the information already contained in gray-scale–fused imagery for performing a spatial localization task.

Interpretation of the findings of the Steele and Perconti[58] study, like several of these studies of performance evaluation, however, is complicated by certain methodological problems. In their study, very few trials were run, a ceiling effect was apparent, and the results obtained were not consistent across the two dependent measures used. The long presentation time (i.e., allowing differential scan patterns and exposure) as well as the general notion of the test queries (differing single questions and varied scenes, as opposed to a specific and criterion-free psychophysical procedure with controlled stimuli and numerous trials), also make the results difficult to interpret for the reasons outlined in the other sections of this chapter. Rigorous psychophysical testing is difficult with natural scenes, and slow, but necessary, if the results are to be meaningful.

Finally, two other studies also bear on this issue. Although they did not evaluate region recognition per se, they involve similar aspects of perceptual organization. In a comparison of performance with imagery from single sensors and from a local contrast-based monochrome fusion technique, Ryan and Tinkler[48] found that pilots received higher ratings of piloting performance when using monochrome-fused imagery than when using imagery from either sensor alone. In another study, Sinai et al.[56] found that people were better able to recognize a scene as matching a previously shown scene if the second scene was rendered in fused imagery compared to I^2 or IR.

The ability to recognize what type of real-world scene (e.g., grass or buildings) is contained in an image region can be, of course, highly dependent upon the nature of the object / region in question. For example, some situations simply yield very high contrast in particular single-sensor images. People or recently run vehicles are much hotter than many outdoor environments and are therefore typically imaged with higher contrast on longwave IR imagery. It has been our experience that observers perform better with I^2 images for sky, trees, sky / tree horizon, and man-made structures. Typical data from the target-patch recognition paradigm are shown in Figure 6.10. Similarly, Steele and Perconti[58] report that I^2 is most useful for trees, terrain slopes, and brush. We have found that fusion in general, and color fusion in particular, strongly help some types of perceptual performance for image regions that consist of two contrasting region textures. Specifically, in one study[10] we found that the perception of roads, man-made objects, and horizon was helped dramatically by color fusion (see Figure 6.9), whereas perception of tree, sky, and grass regions was not aided when tested on a recognition task. Indeed, performance for the latter set was actually hurt significantly by color fusion relative to performance with single-sensor IR images. We suggested that this is because color-fusion methods improve the imaging of regions with contrasting textures (such as road next to grass or

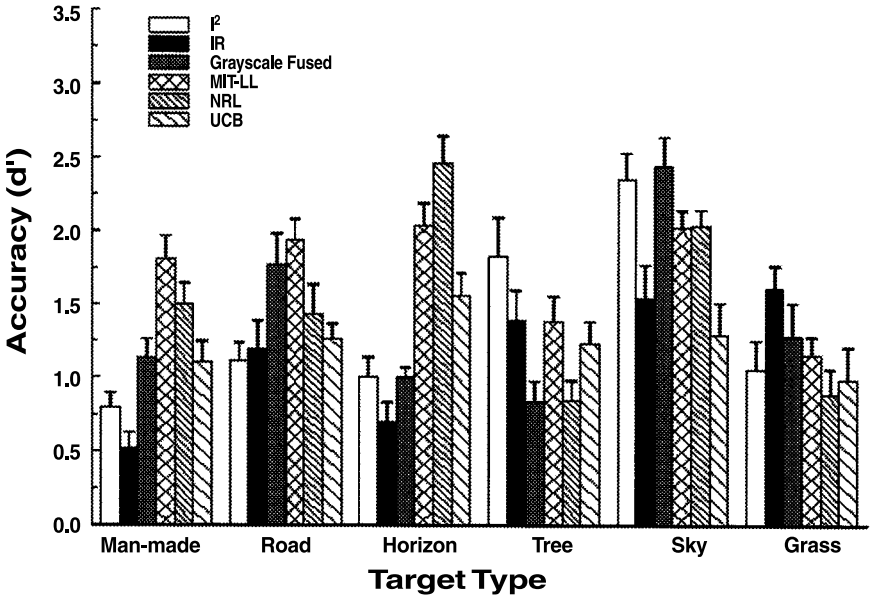

Figure 6.10 Region recognition performance shown for all target categories used, including both contrasting regions (man-made, road, sky / tree horizon, as shown in Figure 6.9), as well as homogeneous regions (sky, tree, and grass). A strong interaction between target category and image type can be seen where performance varies dramatically depending on the image format and the target category.

dirt, man-made objects against other structures or surrounding grass or vegetation, and regions of trees against sky), but hurt the imaging of regions of single homogeneous textures (regions consisting entirely of grass, trees, or sky). In other words, fusion helps most in making different types of scenic regions more differentiable in the perceptual organization of a scene.

6.5.2 Texture-based region segmentation

On a different type of task, the segmentation of regions based on texture, we have found that color fusion that would otherwise appear to be quite good upon informal inspection can hurt performance on some perceptual tasks.[17] In the early stages of perceptual organization, the human visual system links together locations with similar texture properties to begin to form perceptual regions and, eventually, surfaces and meaningful objects.[14,25] For example, locations (pixels) of a certain color or brightness need to be linked together to form a region. In "growing" these regions, the system must be able to account for (i.e., ignore) noise in the pixel values and also for actual image variations such as those due to shadows, gradients of texture size (change of texture-element visual angle with distance), and actual variations of the

structure within the region itself (e.g., wood grain on a desktop, or patterns on tree trunks in a forest). If successful, this segmentation process will produce bounded regions that can be readily imparted with meaning (e.g., a road through a field of grass).

Human texture segmentation, sometimes called "texture discrimination" or "grouping," is typically measured by forming an array of simple texture elements within which the elements in one region are perturbed in some way. For example, a region of "T" shapes may be imbedded in an array of "+" shapes. To test image segmentation with natural scenes we formed arrays that consisted of copies of a sample of an image texture rather than copies of the typical type of texture element. The image texture was a small (0.6°) patch sampled from an image of an outdoor scene, for example, a small patch of imaged grass or trees. The texture patch was duplicated many times to form an array of this texture. To test segmentation ability, we rotated the individual patches of natural texture within one region of the texture array. Just as people can discriminate a region of vertical bars from a region of horizontal bars in a typical texture-grouping test array (shown in Figure 6.11(a)), if a given sensor was effective at imaging texture information in a way effective for the human visual system to use, then people should be good at performing texture segmentation with such stimuli (shown in Figure 6.11(b)).

This paradigm specifically tests texture-based segmentation as opposed to segmentation based on color or intensity differences (which are, in another sense, "texture" features themselves[14]). In other words, to detect a difference in two texture fields whose only difference is a 90° rotation of elements located within one side of the array, the spatial texture within the patch must be well conveyed since the texture elements differ only in the orientation of the spatial structure. The luminance, contrast, and color relations are all identical since the spatial structure of the patch is not changed (it is only rotated). To the extent that one sensor or fusion type provides the human visual system with more of this type of spatial texture information than another, the better performance should be on this segmentation task.

We measured performance on this texture-grouping task in two ways, both using a single-interval forced-choice paradigm. In one method, the texture difference was present on every trial, but it was arranged horizontally on half the trials and vertically on the other half. Observers had to demonstrate region segmentation ability by correctly identifying the orientation of the texture boundary (or, equivalently, the orientation of the regions formed by the grouping). In the other method, a vertical texture boundary was present on half the trials and was absent on the other half. Ability to segment the natural scene texture arrays was measured as the observers' ability to correctly respond "edge" or "no edge" to these trials.

Several different texture arrays made from several representative patches each taken from regions consisting of grass and trees were used. Results, shown in Figure 6.12, indicated that people were better able to perform this texture-based segmentation on the basis of the spatial structure within the IR

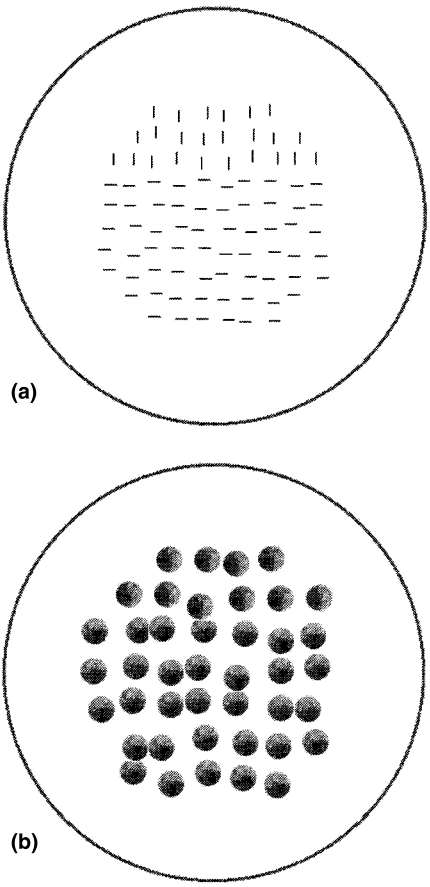

Figure 6.11 Two texture fields used to test region segmentation are shown. **(a)** An array of simple texture elements as commonly used in texture segmentation experiments (left side). **(b)** A similar array as in "a", but made with texture elements that consist of a small patch of a real-world scene rather than a line segment element (right side). The texture elements on the left side of both arrays (i.e., both "a" and "b") are rotated 90° with respect to the texture elements on the right side. If the real-world patch as imaged by a particular sensor or fusion method contains adequate texture information, the left and right sides will group readily into two distinct perceptual regions as in "a."

imagery than any other image type (I^2 or any of the four types of fused imagery). Indeed, we found that fusion actually *hurt* performance significantly for all types of natural texture tested relative to performance with IR imagery. Thus, again, we see that although gray-scale fusion or color fusion may very well help perceptual performance with one type of region or for one perceptual task, fusion most definitely can hurt performance for other perceptual tasks or for other particular region types. Presumably, this texture

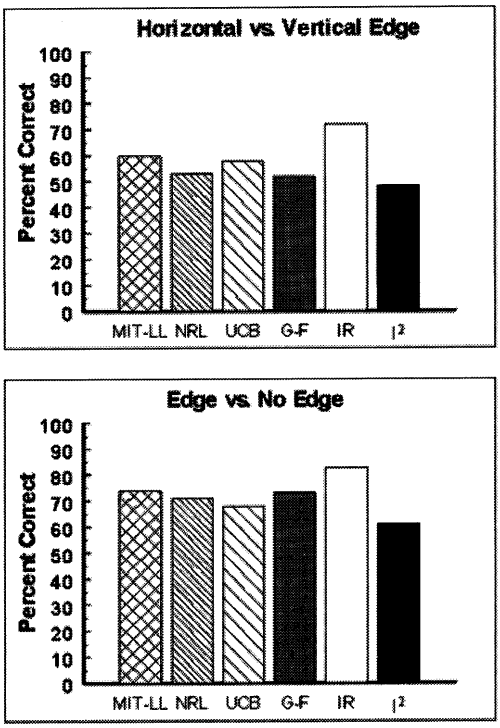

Figure 6.12 Results from two experiments on texture segmentation using six types of night-vision imagery (I^2, IR, MIT-LL gray-scale fused, MIT-LL color-fused method, UCB color-fused method, and NRL's color-fused method. Results are for (top) a task requiring discrimination of a horizontal texture boundary from a vertical boundary, and (bottom) a task requiring detection of the presence of a texture boundary as opposed to trials when no texture difference was present.

segmentation perceptual task relies upon image structure that is different from that required to perform the region recognition perceptual organization task[10] that found a different pattern of results (i.e., gray-scale fusion and color fusion helping performance for contrasting regions).

We pursued this issue further in a recent study in which we compared performance on three different types of texture-based image segmentation tasks.[11,57] The first task was the texture-grouping segmentation task just described. Thus this first task emphasized the grouping together of small regions of local texture into a larger region in order to distinguish this region from another region possessing identical structure, but with a different configuration (orientation) of the "emergent" texture contained in the small patches. In the second task, segmentation was performed on the basis of a difference in texture content in two relatively large regions. A texture boundary between the two regions was imposed within a scene (13.5° × 9°) by altering

the spatial content within one region (one side of the images was filtered in the frequency domain to modify spatial content). The third segmentation task was based on detecting naturally occurring region boundaries in a fairly large scene (3° × 3°). The results further suggested that a given type of fusion may improve one aspect of perceptual performance (here, one type of segmentation performance) while hurting performance on other aspects. Specifically, we found that fusion did not help performance on the first two of these segmentation tasks (texture grouping and filtered-region detection), but did improve performance on the third segmentation task (detection of natural region boundaries). Furthermore, we demonstrated that the tasks for which fusion did not help performance were tasks whose performance was governed by the prevalence (power) of the middle spatial frequencies in the scene and that performance on the task for which performance was improved by fusion was governed instead by power at low spatial frequencies.

In addition to these specific findings about segmentation, more generally, this approach provides a way to assess the image structure required by humans to perform a specific perceptual task and, by extension, what scale a fusion algorithm should emphasize in order to maximize user performance on a specific type of perceptual task.

6.5.3 Object/region detection tasks

Our region recognition work utilizing target categories was similar to a target detection task, but emphasized the determination of membership in a target *category* rather than the detection of a specific target. As we described above, this procedure controlled for some of the problems inherent in performing target detection measurements with "real-world" imagery. However, there have been a number of attempts to test target detection ability more directly, in spite of the difficulties inherent in testing it with real-world images. Some attempts have used objects that are present in the actual images. Typically these studies pose questions to the observer that are tested one time each, asking, for example, about objects ("Is there a bridge present?"), or about position ("Is the sensor platform above the tree line?"). With movies of image sequences, observers can be asked to report when they first detect a certain object in the image sequence.[29,58] These methods offer the advantage that they are highly realistic, but have the associated disadvantage of being unable to conduct numerous repeated trials of the same task due to practical constraints on obtaining fused imagery and given the fact that the same image cannot be used repeatedly because of biases due to learning extraneous cues associated with the presence of the target. For example, if only one image sequence is available, after a single trial of asking "Is there a bridge present?" the image should not be reused to ask a question about another object (due to learning / memory bias favoring different target objects). Furthermore, repeated trials asking the same question cannot be asked to increase reliability and statistical confidence (and relatedly, to more easily avoid ceiling effects).

This inherent trade-off between realism (i.e., external validity) and psychophysical rigor is often encountered when employing real-world scene stimuli in psychophysical studies. However, by keeping this issue in mind when obtaining imagery, imagery for more rigorous psychophysical testing can be obtained.[37,55,62] We have used imagery choreographed to contain a varied number of people or vehicles allowing us to have an appropriately large number of trials in an object detection task. We had observers report whether people, vehicles or neither were present in a scene presented to them[55] and similarly, Krebs et al.[27,37] have tested with imagery obtained with pedestrians located at a predetermined set of positions. Results with these methods of implementing a target detection task using actual embedded targets suggest that fusion can, but does not always, improve performance over single-band imagery.

An additional approach for dealing with the confoundings and complexities of psychophysical testing with natural scenes is to model the presumed behavioral ability of observers. That is, rather than testing human performance, one can make up a mathematical detection rule and calculate detectability as defined by that rule.[28,53] With certain detection rules, color fusion is found to make pixels more distinctive and hence objects are more detectable when colored than when displayed in single-sensor monochrome.[53] However, at the present time, replacing the human perceptual system by a single calculation does not come close to capturing the complexity of human perceptual processing, particularly with respect to the effects of context and type of subject matter, the effect of various false-color relations, and the effects of complex texture variations.

Other efforts to implement a target detection measure in human psychophysical studies have included cutting out an object occurring naturally in an image to serve as a target,[32,49] and also perturbing a region of the image to create a target object.[57,76] These methods can be performed either before processing and fusion, creating an artificial target in the image that is then processed, or after processing and fusion when a portion of the processed image is cut out from the processed images. The cutout target patch may be then either placed into a scene processed in the same way or placed in a neutral homogeneous background. As described above, the first method, cutting and then pasting an image patch into a scene, is hampered by the perceptually powerful texture edges that can occur at the interface border. It is also hampered by the possibility that a patch moved from one area of an image to another will be highly detectable precisely because the sensor format does such a poor job at imaging features in various regions of the scene. For this reason, although we have tried the cut-and-paste method,[49] we have argued that forcing observers to *recognize* what *type* of region or object is contained in the image area is preferable methodologically.[15,16,54] Similarly, we have argued that the edge effects introduced by pasting the target into an image should be avoided, for example by presenting the cut-out image patches on a non-image background such as a uniform monochrome

screen.[10,15,16,54] Results of our study with cut-and-paste targets[49] suggested that color fusion did not show a clear advantage over IR imagery.

The similar method of introducing an artificial perturbation into an image at various locations to serve as a target has also been tried. That is, rather than moving a real object into the scene, either by the cut-and-paste method or by actually moving the object at the time of acquiring the original imagery, the image has been manipulated. Waxman et al.[76] altered the image contrast in a small square region of the I^2 and IR images to serve as a target. They reported that target detection reaction times can be helped by color fusion, but performance can often be seen to be hurt at other contrast values (their Figure 5a, ref. [76]). Unfortunately, statistics to determine which trend predominates are not available.

Using objects placed in a scene Toet et al.,[62] found that the detection of the presence of a visual target (person) is increased with the two color-fusion methods tested (MIT-LL and TNO). As reported in the earlier discussion on region recognition, Toet et al. found that, the *position* of the person in the scene can also be best localized when using fused (color and gray-scale) images, but performance with gray-scale–fused images does not differ from that with color-fused images. They conclude that if enough achromatic spatial information is present, chromatic information is not used by observers to *localize* objects (rather, that color information aids in target *detection*). Our own target detection results with a similar method incorporating a physical translocation of objects during the recording of the imagery, also showed an improvement for the detection of the targets employed in that study in fused imagery,[37,55] but our results with the less rigorous cut-and-paste method indicated no advantage of fusion over IR.[49] Resolution of this conflict will require more testing to assess the limitations of the cut-and-paste method with the same types of targets and fusion methods.

Taken together, the results with these various approaches to testing target/object detection ability suggest that the ability of humans to detect specific region types or objects is often, but by no means always, improved by color fusion. It would appear that results depend upon the exact fusion method used, the type of target and background used, and the exact nature of the target detection task employed. More human factors perceptual testing is required to fully sort out these initial results.

6.5.4 Observer preferences for imagery

The crucial question in assessing color fusion is how well humans perform in methodologically rigorous perceptual tasks. Just as the image with the highest detail (contrast at high spatial frequencies) may not be the one that mediates the best performance on a given perceptual task, the observers' preference for one type of imagery or another may not correspond to best performance on a given task. With that important caveat in mind, it is still interesting to consider what type of imagery is preferred by viewers. In one such

study comparing IR, I^2 and contrast-based local monochrome fusion, both experienced pilots and nonpilots vastly preferred the monochrome-fused imagery across a variety of thermal and illumination conditions.[48] In their study, Steele and Perconti[58] found that although Army helicopter pilots rating segments of video imagery found the imagery barely acceptable, one color-fusion method (MIT-LL) was preferred to another (NRL), but neither color-fusion method was preferred more than the IR imagery. In a second task, using a set of 25 still scenes, no preferences for any format were shown. In a study using a more rigorous scaling and preference mapping procedure[36] achromatic fusion generally was preferred, but single-sensor imagery, this time I^2, was preferred most. Indeed, most strikingly, there was a negative preference for color-fused imagery; that is, static color images were disliked.

In sum, single-sensor imagery is sometimes, but not always, preferred over fused imagery and no study has yet shown a strong preference for color-fused imagery on these types of preference measurements. Not surprisingly, some color-fusion methods are preferred to others. In light of the improved performance shown on certain perceptual tasks with color-fused imagery, the lack of a preference for this imagery is perhaps surprising. On the other hand, the color rendering in the fused imagery is frequently very different from natural daylight scenes; it often looks very unnatural. Thus it may not be surprising that it is often subjectively rated very low. Possibly, the low preferences are due to the lack of familiarity with the false-color renderings, in which case this dislike might be overcome with training. Alternatively, the low preferences might be because observers tend to weigh the appearance of spatial details, which are not perceptually salient in most color-fused imagery, heavily when making preference judgements. The answer to this question awaits further research.

6.6 Conclusion

In this review we have seen repeatedly that, compared to performance with the unfused single-sensor imagery, fusion may create imagery that helps human perceptual performance, but it also may create imagery that hurts performance. Sometimes the same fused imagery will yield increased human performance on one aspect of human perception yet decreased performance on another. Furthermore, we have seen that a given fusion procedure can lead to increased performance on a particular aspect of perception for one type of scene content, yet hurt performance on the same task for another type of scene content (e.g., recognition of grass and road). In sum, the single most compelling conclusion to be drawn from the research to date is that in assessing the utility of sensor fusion, the role of the human viewer must always be considered—the perceptual organization performed by the human visual processing is complex and multifaceted, and the effects of fusion on perception are not presently predictable without empirical psychophysical testing. To be useful, sensor fusion must create imagery that not only combines component

images, but renders them in a format tailored for the multiple facets of human visual perception and for the demands of the perceptual task at hand.

From the studies in which human performance was carefully assessed by rigorous psychophysical methods, the inescapable conclusion is that sensor fusion in general, and color fusion in particular, *can* improve human visual perception and performance. This was seen numerous times in this literature review. In order to summarize studies that bear on this, we have tabulated the results of the studies that have assessed image fusion by behavioral means (see Table 6.1). We attempted to include as many studies as possible in the table and leave it to the reader to evaluate the methodological rigor and the import of the tasks utilized in the studies. For each perceptual task reported by each of these studies, a check mark is entered in the table to indicate whether fusion improved, was equal to, or hurt human performance, relative to the results from the unfused single-sensor imagery with the best perceptual performance of the tested single-sensor types. Where imagery from multiple types of fusion was tested in a single study, results from the imagery with best performance were used. In general, across all aspects of perception tested, fusion (either color or gray-scale fusion) was observed to help performance twice as often as it hurt performance. Fusion had no significant effect on performance about as often as it helped. In comparing performance with the various types of imagery (rightmost column of Table 6.1), performance with IR imagery beat performance with I^2 imagery as often as performance with I^2 imagery beat that with IR imagery (this finding alone explains the need for sensor fusion). In those studies that tested performance with both gray-scale and color-fused imagery, performance with color-fused imagery was almost uniformly superior; performance with gray-scale fusion only beat performance with color-fused imagery in a single case. Frequently, however, performance with color-fused and gray-scale–fused imagery was equivalent, with neither producing benefits relative to the other.

That performance with one type of imagery (say, color fused) does not always beat performance of another type indicates that the comparative utility of a particular type of imagery is dependent upon factors beyond image format. The conclusions drawn in the earlier sections of this chapter indicate that the utility of a type of imagery depends on the specific perceptual ability being considered, and the particular content of the scene (e.g., grass, sky, or man-made objects). Presumably it also depends upon sensor characteristics and quality for a given type of sensor and also upon the particular fusion method as well, although there is not enough data to indicate this yet.

What general conclusions can be drawn? From the literature it appears that (1) color fusion greatly helps to delineate image regions corresponding to contrasting regions of scenic content (e.g., road next to grass, or trees next to sky), (2) there exists strong agreement that color fusion can aid in target / object detection, and that color fusion specifically helps relative to gray-scale fusion, and (3) color fusion can aid in determining spatial relations in a scene and region classification, fundamental aspects of perceptual

Table 6.1 Summary of Results of the Psychophysical Testing to Fused Imagery (Fusion of Low-Light and IR Imagery)

Study	Task	Fusion helps	Fusion same	Fusion hurts	Complete results
Aguilar et al., 1999	Subjective rating of task difficulty for:				
	Tracking people		✓		cf = ir > i^2
	Identification of people			✓	i^2 > cf > ir
	Discrimination of people		✓		cf = ir > i^2
	Identify activity	✓			cf > ir > i^2
	Detect vehicles		✓		cf = ir > i^2
	Identify vehicles		✓		cf = i^2 > ir
	Identify uniforms		✓		cf = i^2 > ir
	Discriminate uniforms		✓		cf = i^2 > ir
	Identify weapons	✓			cf > i^2 = ir
	Discriminate weapons		✓		cf = i^2 > ir
	Detect camouflage	✓			cf > i^2 > ir
	Obscurants: Vegetation		✓		cf = ir > i^2
	Obscurants: Smoke screen		✓		cf = ir > i^2
DeFord et al., 1997	Region recognition (uniform regions)			✓	i^2 > cf = gf = ir
	Region recognition (contrasting regions)	✓			cf > gf > i^2 > ir
See also Sinai et al., 1996					
DeFord et al., 2000	Texture segmentation: Grouping			✓	ir > cf = gf > i^2
	Texture segmentation: Filtered regions		✓		cf = gf = ir > i^2
Essock et al., 1996	Region recognition	✓			cf > ir > i^2
Essock et al., 1997	Texture segmentation: Grouping			✓	ir > cf = gf > i^2
Essock et al., 1999	Region recognition	✓			cf > ir = i^2
Krebs et al., 1998	Target detection in video sequence			✓	ir > cf = gf > i^2
Krebs et al., 1999a	Target detection of person	✓			cf = gf > i^2 > ir
Krebs et al., 1999b	Target detection of embedded object		✓		cf = lwir > mwir > gf > swir

Ryan and Tinkler 1995	Subjective rating of image quality	✓			$gf > ir = i^2$
Sampson et al., 1996	Target detection of embedded object		✓		$cf = gf = ir = i^2$
Sinai et al., 1999a	Target identification (people and vehicles)	✓			$cf > gf = ir = i^2$
	Scene orientation		✓		$cf = gf = i^2 > ir$
Sinai et al., 1999b	Scene recognition	✓			$cf = gf > i^2 > ir$
Sinai et al., 2000	Texture segmentation: Grouping			✓	$ir > gf > i^2$
	Region discrimination	✓			$gf > ir = i^2$
Steele and Perconti 1997	Object identification / location queries		✓		$cf = gf = ir > i^2$
	Shape and orientation		✓		$cf = gf = i^2 > ir$
	Determination of level horizon		✓		$cf = gf = ir = i^2$
	Target detection in video sequence		✓		$cf = gf = ir = i^2$
	Subjective rating of image quality		✓		$cf = gf = ir = i^2$
Toet et al., 1997a	Spatial relations in scene	✓			$cf = gf > ir = i^2$
	Detection misses in spatial relations task	✓			$cf > gf > ir > i^2$
See also Aguilar et al., 1998 and Toet et al., 1997b					
Toet et al., 2000	Scene orientation		✓		$cf = i^2 > gf > ir$
	Horizon discrimination			✓	$i^2 > cf > gf = ir$
	Region recognition (buildings)			✓	$ir > cf = gf > i^2$
	Region recognition (humans)			✓	$ir > cf = gf > i^2$
	Region recognition (road)	✓			$cf > gf = i^2 > ir$
	Region recognition (vehicle)	✓			$cf = gf > i^2 = ir$
	Region recognition (water)	✓			$cf > gf = i^2 > ir$
Waxman et al., 1996	Detection of embedded square	✓			$cf > ir = i^2$

organization. On the other hand however, color fusion seems to be detrimental for perceptual analysis for certain types of scene content, such as regions of uniform scenic content (e.g., trees or grass).

These instances of detrimental effect are, at present, somewhat unpredictable, in that the extent to which they may vary with the imagery (i.e., different sensors, environmental conditions, and fusion method) is not yet clear. Several studies have shown clearly that fusion utility interacts with type of scene content (e.g., trees vs. horizon). This issue, that the image content affects performance of a particular aspect of perceptual organization, needs to be addressed. For example, to optimize image segmentation for target search, research must determine what image structural content is important and therefore should be emphasized, and what content is irrelevant and should not have processing time devoted to it. We have begun this type of analysis for image segmentation and have found that for one aspect or another of general image segmentation, one particular band of spatial frequencies or another is more important. For example, fusion has been shown to help performance on scene segmentation if the segmentation is based on low spatial frequency content, but is detrimental to segmentation performance if segmentation is based on middle spatial frequency content. This type of finding raises the possibility of altering fusion algorithms to emphasize the particular spatial frequency scales most relevant to the perceptual abilities being utilized under given circumstances.

This review indicates that fused imagery, particularly color-fused imagery, offers considerable improvement to human perceptual organization and situational awareness in some situations. What can be done to ensure that fusion helps (and not hurts) performance more of the time is not yet known. This review of the present literature has suggested two general research areas that should be pursued to address this. First, the image content required to perform various fundamental aspects of perceptual organization should be determined and this knowledge incorporated into image processing and fusion algorithms. As noted above, fusion methods might then be "tuned" to emphasize perceptually relevant information for various perceptual applications. We have already shown by using band-pass filtered imagery in behavioral testing that one can begin to assess what types of spatial content may be important for a certain human task, and what information may be deemphasized in fusion without loss of perceptual performance. Further work along these lines is needed.

The second area of needed research concerns the selection of color mappings in the false-color imagery (and associated aspects of training of users). Color mappings need to be examined in more detail with respect to human color vision and human performance. First of all, the color space of the human viewing the false-color display should be considered. Since human color discrimination and also color salience in "preattentive" vision are not uniform with respect to physical color space metrics, it would seem more profitable to map the physical color space produced by the sensor fusion

method linearly onto the appropriate perceptual color space rather than directly onto another physical (RGB) color space. In other words, something akin to histogram equalization needs to be performed in the three-dimensional space (rather than in one dimension) and it needs to be performed in *perceptual* space in order to spread out the false-color mapping such that the steps in this mapping are as perceptually useful as possible for the particular perceptual task facing the user. Secondly, the issues raised above related to whether it is more useful to use novel colorations of particular targets, use some particular unnatural color rendering of whole scenes, or use roughly natural colors for scenes, needs to be assessed. That is, whether natural coloration of outdoor scenes is optimal or desirable for a given perceptual task is not yet known. Finally, the benefits possible from the training of users with false-color imagery need to be investigated.

Future research on human perceptual ability with fused imagery also needs to investigate other issues. Performance with temporal sequences of imagery needs to be investigated, particularly as the quality of real-time fusion is ever increasing, and, of course, because of the dynamic nature of human visual processing and perceptual organization. Future research also needs to pursue issues concerning the use of alternative and additional bands of radiation including hyperspectral imagery. This research needs to address whether different sets of bands can be selected to create intersensor contrast, or "signatures" across bands, that are particularly effective for humans performing different perceptual tasks.

Color fusion is in essence biological fusion, computationally done by the human visual system. Like fusion of dichoptic images by the binocular neural pathway of the human visual system, human processing combines three one-dimensional displays into one three-dimensional (color) display. This is an incredibly powerful transformation leading to a much larger volume of realizable perceptual space. However there are problems: due to the processing of the human visual system, and due to the prior learning and experience in the visual world, color fusion does not presently necessarily improve visual performance of humans. Depending upon the characteristics of the particular images, the fusion and rendering method, the nature of the scenic content, and most importantly, the particular aspect of multifaceted human perceptual organization being considered, human performance may improve or may decrease due to color fusion. The challenge that now faces researchers is to determine whether fusion methods can be improved to be more broad-spectrum with respect to various aspects of human perceptual organization, or whether aspects of fusion and display need to be made selectable to emphasize the aspects of visual perception needed for a given type of task situation. Ideally, one fusion/rendering method will optimize numerous aspects of perception. Perhaps analysis of the image content required to mediate different human perceptual abilities can exploit knowledge of which types of image content to emphasize in fusion in different user situations. This exciting research is in its infancy and the answers to these questions await future research.

This work was supported by grant #N00014-99-1-0516 from the Office of Naval Research (ONR). We wish to thank ONR and Army Night Vision and Electronic Sensors Directorate (NVESD) for providing the imagery used in this research. We also thank ONR, Army NVESD, Tami Peli, Dean Scribner, Lex Toet, and Frank Werblin for providing the images reproduced in this chapter. Finally, we emphasize that all fusion was performed by the individuals/labs indicated in the text.

References

1. Aguilar, M., Fay, D. A., Ireland, D. B., Racamato, J. P., Ross, W. D., and Waxman, A. M., Field evaluations of dual-band fusion for color night vision, *Proc. SPIE Conf. Enhanced and Synthetic Vision*, SPIE-3691, 168–175, 1999.
2. Aguilar, M., Fay, D. A., Ross, W. D., Waxman, A. M., Ireland, D. B., and Racamato, J. P., Real-time fusion of low-light CCD and uncooled IR imagery for color night vision, *Proc. SPIE Conf. Enhanced and Synthetic Vision*, SPIE-3364, 124–135, 1998.
3. Anderson, J. D., Bechtoldt, H. P., and Dunlap, G. L., Binocular integration in line rivalry, *Bull. Psychonomic Soc.*, 11, 399–402, 1978.
4. Barham, P., Oxley, P., and Ayala, B., Evaluation of the human factors implications of Jaguar's first prototype near-infrared night vision system, in *Vision in Vehicles* 6, Gale, A. G., et al., Eds., Elsevier Science, Amsterdam, 1998.
5. Beck, J., Prazdny, K., and Rosenfeld, A., A theory of textural segmentation, in *Human and Machine Vision*, Beck, J., Hope, B., and Rosenfeld, A., Eds., Academic Press, New York, 1–38, 1983.
6. Berson, E. L., Night Blindness: some aspects of management, in *Clinical Low Vision*, Faye, E. E., Ed., Little, Brown and Co., Boston, 1976.
7. Biederman, I. and Ju, G., Surface versus edge-based determinants of visual recognition, *Cognit. Psychol.*, 20, 38–64, 1988.
8. Caelli, T. and Yuzyk, J., What is perceived when two images are combined? *Perception*, 14, 41–48, 1985.
9. Cameron, A. A., The development of the combiner eyepiece night vision goggle, *Proc. SPIE Conf. on Helmet-Mounted Displays II*, SPIE—The International Society for Optical Engineering, Bellingham, WA, 1290, 16–29, 1990.
10. DeFord, J. K., Sinai, M. J., Krebs, W. K., Srinivasan, N., and Essock, E. A., Perceptual organization of color and non-color nighttime real-world imagery, *Invest. Ophthal. Visual Sci. (Suppl.)*, 38, S641, 1997.
11. DeFord, J. K., Sinai, M. J., Purkiss, T. J., and Essock, E. A., Segmentation of nighttime real-world scenes, *Invest. Ophthal. and Visual Sci. (Suppl.)*, 41, S221, 2000.
12. DeValois, R. L. and DeValois, K. K., *Spatial Vision*. Oxford, New York, 1998.
13. Donderi, D. C., Visual acuity, color vision, and visual search performance at sea, *Hum. Factors*, 36, 129–144, 1994.
14. Essock, E. A., An essay on texture: the extraction of stimulus structure from the visual image, in Burns B., Ed., *Percepts, Concepts, and Categories*, Elsevier Science, Amsterdam, 1992, 3–35.
15. Essock, E. A., McCarley, J. S., Sinai, M. J., and Krebs, W. K., Functional assessment of night-vision enhancement of real-world scenes, *Invest. Ophthal. and Visual Sci. (Suppl.)*, 37, S517, 1996.
16. Essock, E. A., Sinai, M. J., McCarley, J. S., Krebs, W. K., and DeFord, J. K., Perceptual ability with real-world nighttime scenes: image-intensified, infrared, and fused-color imagery, *Hum. Factors*, 41, 438–452, 1999.

17. Essock, E. A., Sinai, M. J., Srinivasan, N., DeFord, J. K., and Krebs, W. K., Texture-based segmentation in real-world nighttime scenes, *Invest. Ophthal. and Visual Sci. (Suppl.)*, 38, S639, 1997.
18. Gibson, J. J., *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
19. Goodale, M. A., The cortical organization of visual perception and visuomotor control, in S. M. Kosslyn and Osherson, D. N., Eds., *An Invitation to Cognitive Science*, 2nd ed., Volume 2: *Visual Cognition*, MIT Press, Cambridge, MA, 167–213, 1995.
20. Grossberg, S., *Neural Networks and Natural Intelligence* (Chapters 1–4). MIT Press, Cambridge, MA, 1988.
21. Hoffman, R. R., Remote perceiving: a step toward a unified science of remote sensing, *Geocarto Int.*, 5, 3–13, 1990.
22. Hoffman, R. R., Detweiler, M. A., Lipton, K., and Conway, J. A., Considerations in the use of color in meteorological displays, *Weather Forecasting*, 8, 505–518, 1993.
23. Howard, I. P. and Rogers, B. J., *Binocular Vision and Stereopsis*, Oxford, New York, 1995.
24. Irwin, D. E., Information integration across saccadic eye movements, *Cognit. Psych.*, 23, 420–456, 1990.
25. Julesz, B., Toward an axiomatic theory of preattentive vision, in *Dynamic Aspects of Neocortical Function*. Edelman, G. M., Gall, W. E. and Cowan W. M., Eds., Wiley, New York, 585–612, 1984.
26. Klein, G. A. and Hoffman, R. R., Seeing the invisible: perceptual-cognitive aspects of expertise, in *Cognitive Science Foundations of Instruction*, Rabinowitz, M., Ed., Erlbaum, Mahwah, NJ, 203–226, 1992.
27. Krebs, W. K., McCarley, J. S., Kozek, T., Miller, G., Sinai, M. J., and Werblin, F. S., An evaluation of a sensor fusion system to improve drivers' nighttime detection of road hazards, *Proc. 43rd Annu. Meet. Hum. Factors Ergonomics Soc.*, 1999a.
28. Krebs, W. K., Scribner, D. A., McCarley, J. S., Ogawa, J. S., and Sinai, M. J. Comparing human target detection with multidimensional matched filtering methods, *Proc. NATO Res. Technol. Conf. Search and Target Acquisition*. Amsterdam, The Netherlands, 1999b.
29. Krebs, W. K., Scribner, D. A., Miller, G. M., Ogawa, J. S., and Schuler, J., Beyond third generation: a sensor fusion targeting FLIR pod for the F / A-18, *Proc. SPIE-Sensor Fusion: Architectures, Algorithms, and Appl. II*, 3376, 129–140, 1998.
30. Krebs, W. K., Scribner, D. A., Schuler, J., Miller, G., and Lobik, D., Human factor test and evaluation of a low light sensor fusion device for automobile applications, Automotive Night Vision / Enhanced Driving Conference, Detroit, MI, June 5, 1996.
31. Livingstone, M. S. and Hubel, D. H., Segregation of form, color, movement, and depth: anatomy, physiology, and perception, *Science*, 240, 740–749, 1988.
32. Lowe, R. K., Components of expertise in the perception and interpretation of meteorological charts, in *Interpreting Remote Sensing Imagery: Human Factors*, Hoffman, R. R. and Markman, A. B., Eds., Lewis Publishers, Boca Raton, FL, 2001.
33. Luo, R. and Kay, M., Data fusion and sensor integration: state of the art in the 1990s, in *Data Fusion in Robotics and Machine Intelligence*, Abidi, M. and Gonzalez, R., Eds., Academic Press, San Diego, 7–136, 1992.
34. Macmillan, N. A. and Creelman, C. D., *Detection Theory: A User's Guide*. Cambridge University Press, New York, 1991.
35. Marr, D., *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, San Francisco, 1982.

36. McCarley, J. S., Krebs, W. K., Essock, E. A., and Sinai, M. J., Multidimensional scaling of single-band and sensor-fused dual-band imagery, unpublished data.
37. McCarley, J. S., and Krebs, W. K., Detection of road hazards in thermal, visible, and sensor-fused nighttime imager, *Appl. Ergonomics*, 31, 523–530, 2000.
38. McDaniel, R., Scribner, D., Krebs, W., Warren, P., Ockman, N., and McCarley, J., Image fusion for tactical applications, *Proc. SPIE—Infrared Technology and Applications XXIV*, 3436, 685–695, 1998.
39. Mollon, J. D., Tho' she kneel'd in that place where they grew . . . : the uses and origins of primate color vision, *J. Exp. Biol.*, 146, 21–38, 1989.
40. Nothdurft, H.-C., The role of features in preattentive vision: comparison of orientation, motion and color cues, *Vision Res.*, 33, 1937–1958, 1993.
41. Peli, E., Contrast in complex images, *J. Optical Society Am.* A, *7*, 2030–2040, 1990.
42. Peli, E., in search of a contrast metric: matching the perceived contrast of Gabor patches at different phases and bandwidths, *Vision Res.*, 37, 3217–3224, 1997.
43. Peli, T., Ellis, K., Stahl, R., and Peli, E., Integrated color coding and monochrome multi-spectral fusion, *Proc. IRIS Specialty Group Sensor Data Fusion, Multi-Source Fusion Theory Applications*, 1999.
44. Peli, T. and Lim, J. S., Adaptive filtering for image enhancement, *J. Optical Eng.*, 21, 108–112, 1982.
45. Peli, T., Peli, E., Ellis, K., and Stahl, R., Multi-spectral image fusion for visual displays, *Proc. SPIE Conf. Sensor Fusion: Architectures, Algorithms, and Appli. III*, 3719, SPIE—The International Society for Optical Engineering, 1999.
46. Philips, W. A., On the distinction between sensory storage and short-term visual memory, *Percept. Psychophysics*, 16, 283–290, 1974.
47. Rabin, J. and Wiley, R., Switching from forward-looking infrared to night vision goggles: transitory effects on visual resolution, *Aviation Space, Environmental Medicine*, 65, 327–329, 1994.
48. Ryan, D. and Tinkler, R., Night pilotage assessment of image fusion, *Proc. SPIE Conf. on Helmet and Head-mounted Displays and Symbology Design Requirements II*, 2465, SPIE—The International Society for Optical Engineering, Bellingham, WA, 50–67, 1995.
49. Sampson, M. T., Krebs, W. K., Scribner, D. A., and Essock, E. A., Visual search in natural (visible, infrared, and fused visible and infrared) stimuli, *Invest. Ophthal. Visual Sci. (Suppl.)*, 37, S296, 1996.
50. Scribner, D. A., Satyshur, M. P., and Kruer, M. R., Composite infrared color images and related processing, IRIS Specialty Group on Targets, Backgrounds, and Discrimination, January, San Antonio, TX, 1993.
51. Scribner, D. A., Satyshur, M. P., Schuler, J., and Kruer, M. R., Infrared color vision, IRIS Specialty Group on Targets, Backgrounds, and Discrimination, January, Monterey, CA, 1996.
52. Scribner, D. A., Warren, P., and Schuler, J., Extending color vision methods to bands beyond the visible, *Proc. IEEE Workshop on Comput. Vision beyond the Visible Spectrum: Methods and Appl.*, June, Fort Collins, CO, 1999.
53. Scribner, D. A., Warren, P., Schuler, J., Satyshur, M., and Kruer, M. R., Infrared color vision, *Optics Photonics News*, 9, 27–32, 1998.
54. Sinai, M. J., Essock, E. A., and Krebs, W. K., Perceptual organization of real-world scenes, Annual Meeting of the Psychonomics Society, Chicago, IL, November, 1996.

55. Sinai, M. J., McCarley, J. S., Krebs, W. K., and Essock, E. A., Psychophysical comparisons of single- and dual-band fused imagery, *Proc. SPIE Conf. on Enhanced and Synthetic Vision*, Orlando, FL, 3691, 176–183, 1999a.
56. Sinai, M. J., McCarley, J. S., and Krebs, W. K., Scene recognition using infrared, low-light, and fused-color imagery, *Proc. IRIS Specialty Group on Passive Sensors*, February, Monterey, CA, 1999b.
57. Sinai, M. J., DeFord, J. K., Purkiss, T. J., and Essock, E. A., Relevant spatial frequency information in the texture segmentation of night-vision imagery, *Proc. SPIE Conf. on Enhanced and Synthetic Vision*, Orlando, FL, 4023, 2000.
58. Steele, P. M. and Perconti, P., Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage, *Proc. SPIE 11th Annu. Inter. Symp. Aerosp./Defense Sensing, Simulation and Controls*, SPIE—The International Society for Optical Engineering, Bellingham, WA, 3062, 88–100, 1997.
59. Therrien, C. W., Scrofani, J., and Krebs, W. K., An adaptive technique for the enhanced fusion of low-light with uncooled thermal infrared imagery, *IEEE Conference on Image Processing*, October, 1997.
60. Toet, A., Image fusion by a ratio of low-pass pyramid, *Pattern Recognition Letters*, 9, 245–253, 1989.
61. Toet, A., Multiscale contrast enhancement with applications to image fusion, *Optical Eng.*, 31, 1026–1031, 1992.
62. Toet, A., Ijspeert, J. K., Waxman, A. M., and Aguilar, M., Fusion of visible and thermal imagery improves situational awareness, *Proc. SPIE Conf. on Enhanced and Synthetic Vision*, 3088, 177–188, 1997a.
63. Toet, A., Ijspeert, J. K., Waxman, A. M., and Aguilar, M., Fusion of visible and thermal imagery improves situational awareness, *Displays*, 18, 85–95, 1997b.
64. Toet, A., Schoumans, N., and IJspeert, J. K., Perceptual evaluation of different nighttime modalities, *Fusion 2000: Proc. 3rd Inter. Conf. Information Fusion*, Paris, 2000.
65. Toet, A., van Ruyven, L. J., and Valeton, J. M., Merging thermal and visual images by a contrast pyramid, *Optical Eng.*, 28, 789–792, 1989.
66. Toet, A. and Walraven, J., New false color mapping for image fusion, *Optical Eng.*, 35, 650–658, 1996.
67. Uttal, W. R., Baruch, T., and Allen, L., Dichoptic and physical information combination: a comparison, *Perception*, 24, 351–362, 1995.
68. Vrahimis, G., Multisensor image fusion–a neural network based approach, *Progress in Connectionist-Based Information Systems: Proc. Inter. Conf. on Neural Information Processing and Intelligent Information Systems*, Springer-Verlag, Singapore, 1998.
69. Wandell, B. A., *Foundations of Vision*, Sinauer Associates, Inc., Massachusetts, 1995.
70. Ward, N. J., Stapleton, L., and Parkes, A., Behavioral and cognitive impact of night-time driving with HUD contact analogue infrared imaging, *Proc. 14th Inter. Technical Conf. on Enhanced Safety of Vehicles*, Munich Germany, May 23–26, 1994.
71. Waxman, A. M., Carrick, J. E., Fay, D. A., Racamato, J. P., Aguilar, M., and Savoye, E. D., Electronic imaging aids for night driving: low-light CCD, thermal IR, and color fused visible / IR, *Proc. SPIE Conf. on Transportation Sensors and Control*, 2902, 1996b.
72. Waxman, A. M., Carrick, J. E., Racamato, J. P., Fay, D. A., Aguilar, M., and Savoye, Color night vision—3rd update: realtime fusion of low-light CCD

visible and thermal IR imagery, *Proc. SPIE Conf. on Enhanced and Synthetic Vision*, 3088, 1997b.
73. Waxman, A. M., Fay, D. A., Gove, A., Siebert, M., Racamato, J. P., Carrick, J. E., Savoye, E. D., Color night vision: fusion of intensified visible and thermal imagery, *Pro. SPIE Conf. on Synthetic Vision for Vehicle Guidance and Control*, 2463, 58–63, 1995.
74. Waxman, A. M., Fay, D. A., Ireland, D. B., Racamato, J. P., Ross, W. D., Streilein, W. W., Braun, M., and Aguilar, M., Fusion of 2–/3–/4–sensor imagery for visualization, target learning, and search, *Proc. SPIE Conf. on Enhanced and Synthetic Vision*, 4023, 2000.
75. Waxman, A. M., Gove, A. N., Fay, D. A., Racamato, J. P., Carrick, J. E., Seibert, M. C., and Savoye, E. D., Color night vision: opponent processing in the fusion of visible and IR imagery, *Neural Networks*, 10, 1–6, 1997a.
76. Waxman, A. M., Gove, A., Siebert, M. C., Fay, D. A., Carrick, J. E., Racamato, J. P., Savoye, E. D., Burke, B. E., Reich, R. K., McGonagle, W. H., and Craig, D. M., Progress on color night vision: visible/ir fusion, perception and search, and low-light CCD imaging, *Proc. SPIE Conf. on Enhanced and Synthetic Vision*, 2736, 96–107, 1996a.
77. Werblin, F. S., Roska, T., Chua, L., Jacobs, A., Kozek, T., and Zarandy, A., Transfer of retinal technology to applications beyond biological vision, *Investigative Ophthalmol. Visual Sci.*, 38, S479, 1997.
78. Werblin, F. S., Roska, T., and Chua, L., The analogic cellular neural network as a bionic eye, *Int. J. Circuit Theory Appl.*, 23, 541–569, 1995.
79. Wolfe, J. M., Visual search in continuous, naturalistic stimuli, *Vision Res.* 34, 1187–1195, 1994.
80. Wurm, L. H., Legge, G. E., Isenberg, L. M., and Luebker, A., Color improves object recognition in normal and low vision, *J. Exp. Psychol. Hum. Percept. Performance*, 19, 899–911, 1993.