THE PSYCHOPHYSICS OF SENSOR FUSION: A MULTIDIMENSIONAL SIGNAL DETECTION ANALYSIS

Jason S. McCarley
University of Illinois at Urbana-Champaign, Human Factors Division
Savoy, Illinois

William K. Krebs
United States Navy, ONR/NRL SCI & TECH 101

Image fusion techniques take input from multiple single-band sensors and combine it to create a single multi-band image. While such processing offers to create imagery that is more information rich than that produced by single-band sensors, it may sometimes degrade the perceptual quality of the input content. Working within the context of Ashby and Townsend's (1986) General Recognition Theory, a multidimensional signal detection model of perceptual interactions between stimulus features, the current study measured the perceptibility of single-band content within fused images. Data indicate that the perceptibility of information from one single-band input channel can be degraded as the contrast level of the alternative single-band input is manipulated. False-color rendering of fused images, likewise, can sometimes improve and sometimes degrade perceptibility of single-band content. Implications for the design and testing of sensor-fused displays are discussed.

INTRODUCTION

Image fusion techniques accept imagery from multiple single-band input sensors and combine it to form a single image. The potential benefits of such technology to human perceptual performance are twofold. First, by melding information from multiple sensors, fused imagery can potentially mitigate the need for operators to switch attention between and mentally integrate the contents of different displays (Irwin, 1996). Second, by exploiting the contrast between input images, fusion algorithms can potentially enhance the perceptibility of stimulus information. For example, spatial filtering based on the luminance differences between input stimuli can be used to enhance stimulus contrast in the fused output image (Waxman et al., 1997). Luminance contrast between input image also allows for color rendering of fused imagery (Scribner et al., 1998; Waxman et al., 1997). In fact, rendering within a three-dimensional color space allows fused presentation of up to three input images with no mathematical loss of information.

Unfortunately, there is no guarantee that a fused image will be of equal or higher quality—that is, of equal or greater value to a human operator—than any of the input images from which is it is created (Essock et al., 2001). Information is necessarily lost, for example, when multiple images are combined within a monochromatic display. The product image will therefore be improved relative to the input only if the information sacrificed is unimportant to the human operator. Even within a chromatic rendering with no loss of information, moreover, the perceptibility of valuable information may be compromised. For instance, information from one input sensor may mask or degrade that that from a second. While objectively present within an image, task relevant information may thus be difficult to perceive.

While a number of authors have described metrics for assessing the quality of fused images (see Dixon et al., in press, for review), these may not always correlate well with measures of human perceptual performance. In judging the value of fused imagery, it therefore remains important to directly assess the effects of fusion on human perception. Attempts at this, however, have produced inconsistent results, with some studies suggesting that sensor fusion improves visual performance and others offering less promising results (e.g., Krebs & Sinai, 2002; McCarley & Krebs, 2000; Toet et al., 1997). Some of this inconsistency is due to differences between fusion algorithms (Dixon et al., in press; McCarley & Krebs, 2000). Indeed, one point of psychophysical testing is to compare the quality of different image fusion techniques. Comparisons between techniques, though, are complicated by variability in the psychophysical tasks that have been used for testing. These have included target detection (e.g., Dixon et al, in press; McCarley & Krebs, 2000), scene recognition and orientation judgment (Krebs & Sinai, 2002), and texture discrimination (Essock et al., 2001). A finding that image fusion improves performance on one task does not ensure that similar benefits will obtain in a different task, making broad conclusions about the value of image fusion difficult.

The goal of the present work was to seek a generalized and theoretically-motivated procedure for assessing the perceptual effects of image fusion. More specifically, we sought a broadly applicable methodology for assessing the perceptibility of information within multi-band fused images. Given some knowledge of the relationship between single-band perceptual quality and the quality of human performance on various tasks (e.g., target detection, spatial orientation judgments), such a generalized measure of information perceptibility, perhaps in conjunction with objective image quality metrics, could make it possible to predict the effects of sensor fusion on performance across a range of tasks without need for psychophysical testing on each of those tasks individually. The theoretical framework employed here was Ashby and Townsend's (1986) *General Recognition Theory* (*GRT*), which uses signal detection theory (Green & Swets, 1966) to characterize perceptual interactions between features in multidimensional stimuli. Within a typical GRT experiment, a set of four stimuli is created through the factorial manipulation of two different stimulus features, each tested at two different values. Perceptual effects of the four stimuli can be represented with probability density functions within a bivariate space, where each axis denotes the strength of the percept produced by one stimulus component (Figure 1). Each trial the subject is presented a single stimulus from the set of four and asked to identify it. Information about the underlying distributions can be inferred from the subject's performance.

Within this framework, Ashby and Townsend (1986) identify several potential forms of dimensional interaction. Here, we consider the phenomenon of *perceptual separability*. Perceptual separability obtains when the perceptual effects of one stimulus dimension are the same across all levels of the second dimension. In Figure 1, separability is violated for feature A but not B: the probability distributions for Feature A have greater variance at level 2 of Feature B than at level 1, while the distributions of Feature B are equivalent across levels of Feature A. Perceptual separability for a given feature can be tested by comparing *d'* for discriminations of that feature across values of the second feature. If *d'* for discrimination of A_1 from A_2 is different at level 2 of feature B than it is at level 1, then feature A is not perceptually separable from feature B (Kadlec & Townsend, 1992).

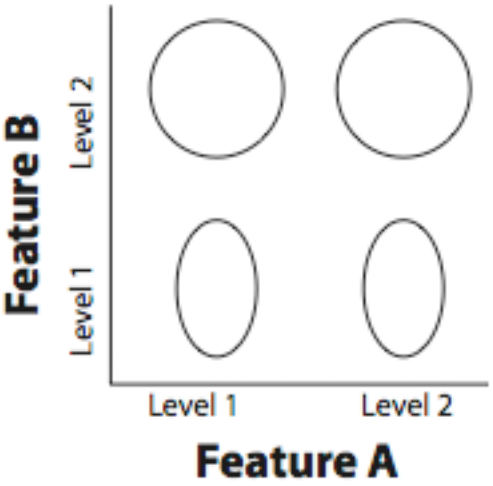

Figure 1. Equal-density contours representing hypothetical distributions of perceptual responses to four different stimuli within a two-dimensional stimulus space.

The current study employed GRT to test the perceptual separability of single-band image information within dual-band sensor-fused images. Stimuli were fused long-wave (LW) and mid-wave (MW) images, with a set of four items created by the factorial manipulation of LW and MW image contrast (either low or high) (Figure 2). Thus, in one image ($LW_{Low}MW_{Low}$), input images from both sensors were low in contrast; in another ($LW_{High}MW_{Low}$), the LW input was high in contrast and the MW input was low; in a third, the LW input was low in contrast and the MW input was high ($LW_{Low}MW_{High}$); in the last, both input images were high in contrast ($LW_{High}MW_{High}$). Each trial, the observer saw one of four stimuli and was asked to report which of the four it was. To respond correctly, therefore, the observer was required to discriminate the contrast level of the LW and MW content within the fused image. Data were tested for perceptual separability of LW and MW information. More specifically, data were analyzed to determine whether *d'* for judgments of LW component contrast (i.e., judgments discriminating stimuli in the top row of Figure 2 from stimuli in the bottom row) was affected by contrast of the MW component, and likewise to determine whether *d'* for the ability to discriminate contrast of the MW component (i.e., judgments discriminating stimuli in the left column of Figure 2 from stimuli in the right column) was affected by contrast of the LW component. Both chromatic and grayscale versions of the fused stimuli in order to determine if perceptual separability within fused images is either improved or compromised by color rendering.

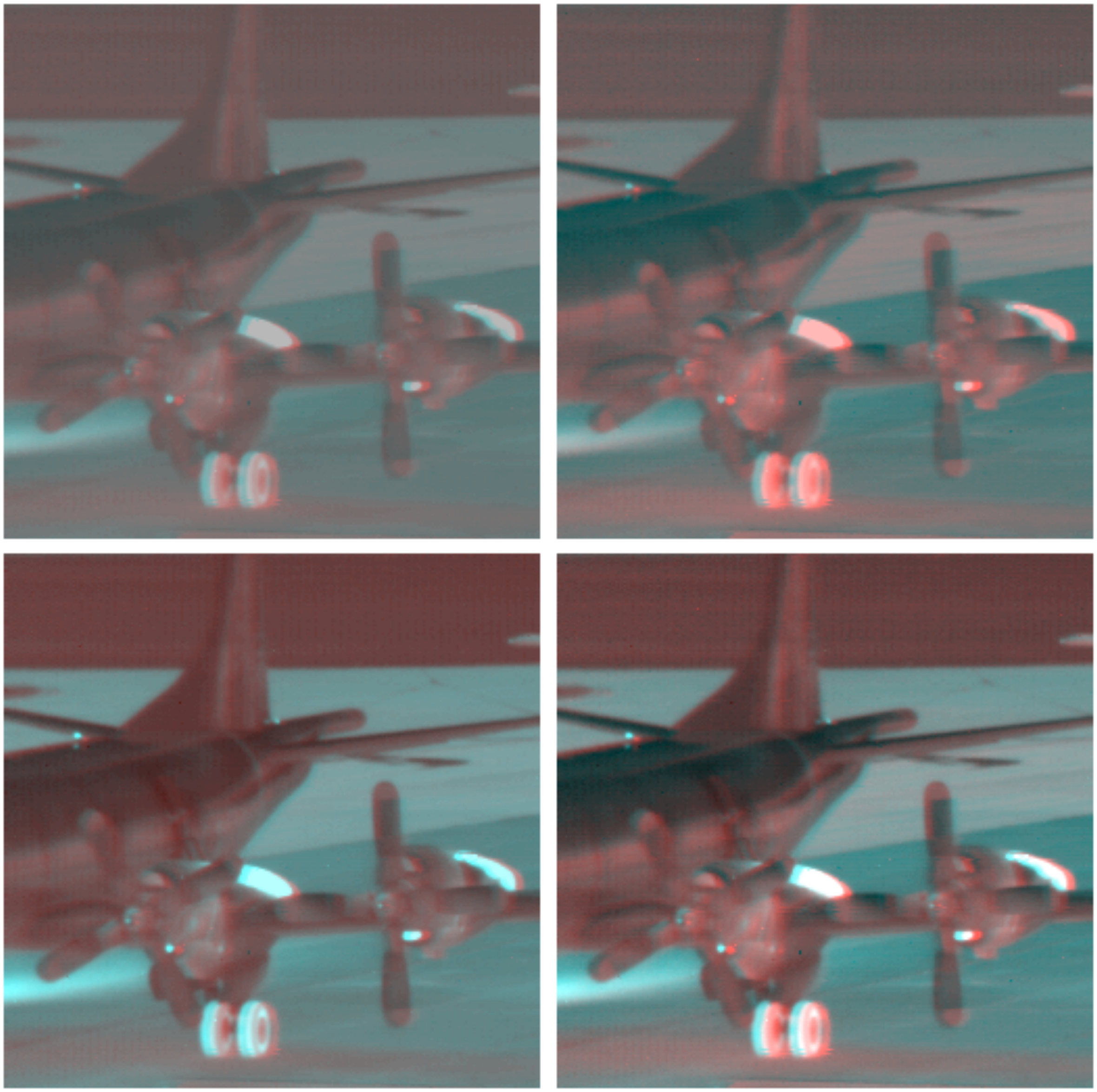

Figure 2. Fused images used as stimuli. Top left: $LW_{Low}MW_{Low}$; Top right: $LW_{High}MW_{Low}$; Bottom left: $LW_{Low}MW_{High}$; Bottom right: $LW_{High}MW_{High}$. Testing was done using both false-colored and achromatic versions of the stimulus set.

METHOD

Observers

Observers were 12 adults, including the first author. All had normal or corrected-to-normal visual acuity, and normal color vision.

Stimuli

Stimuli were dual-band images created by fusing a pair of spatially aligned LW and MW images. A set of four fused stimuli were created by factorial manipulation of the contrast of the LW and MW input images, with each being presented at a contrast of either 100% or 50% relative to its original level. Chromatic fused images were created using the simple fusion technique of Scribner et al. (1998), which produces red/cyan dual-band imagery by mapping one input image to red pixel values and the alternate input image to blue and green pixel values. Here, the LW image was mapped to red, and the MW image to cyan. Achromatic fused images were created by grayscale rendering of the chromatic images. All images were 256 x 256 pixels in size. Stimuli were presented on a 17” monitor with resolution of 1024 x 768 pixels and a refresh rate of 75 Hz.

Procedure

The observer initiated each trial with a mouse click. After a 150 ms delay, the imperative stimulus image appeared briefly then was overwritten by a pattern mask. To avoid floor and ceiling effects on response accuracy, exposure duration was controlled by a staircase procedure that maintained overall performance level at 50% correct, twice the level of chance. After stimulus presentation, a response display appeared containing all four of the potential stimulus items, one located in each corner of the screen. The observer’s task was to click on the image that matched the stimulus for that trial. Each observer completed 10 blocks of 40 trials each. Chromatic and achromatic stimuli were presented in alternating blocks. The first two blocks were treated as warm-up and excluded from analysis.

RESULTS

d’ was calculated for discrimination of LW contrast at each level of MW contrast, i.e., *d’* for was calculated for discrimination of low contrast LW images from high contrast LW images for conditions in which MW contrast was low ($LW_{Low}MW_{Low}$ vs. $LW_{High}MW_{Low}$), and for conditions in which MW contrast was high ($LW_{Low}MW_{High}$ vs. $LW_{High}MW_{High}$). Likewise, *d’* for discrimination of MW contrast was calculated at each level of LW contrast ($LW_{Low}MW_{Low}$ vs. $LW_{Low}MW_{High}$ and $LW_{High}MW_{Low}$ vs. $LW_{High}MW_{High}$). Values of *d’* were calculated by treating the low-contrast image from within a pair of stimulus conditions as noise and the high-contrast image as signal. Thus, for example, in calculating *d’* for discrimination of the $LW_{Low}MW_{High}$ from $LW_{High}MW_{High}$, a hit occurred when the stimulus $LW_{High}MW_{High}$ was correctly identified, and a false alarm occurred when the stimulus $LW_{Low}MW_{High}$ was incorrectly identified as $LW_{High}MW_{High}$.

For statistical analysis, *d’* scores for discrimination of LW contrast and MW contrast were submitted to separate two-way ANOVAs with contrast of the alternative component (high vs. low) and rendering (chromatic vs. achromatic) as within-subjects factors. Data are presented in Figure 3. Analysis of *d’* for discrimination of LW contrast revealed that sensitivity was higher when the MW content was presented at low contrast than when it was presented at high contrast [$p = .006$]. In other words, contrast discrimination of LW information was degraded by an increase in the contrast of MW information. Sensitivity was also higher when fused images were rendered in color than when they were achromatic [$p < .001$]. The interaction of MW contrast by rendering was non-significant [$F < 1$]. Analysis of *d’* for discrimination of MW contrast level showed no main effect of LW contrast level [$p = .14$], but indicated that sensitivity was decreased when fused images were rendered in color [$p = .001$]. The interaction of LW contrast by rendering was non-significant [$p = .11$], though trends in the data suggested that high LW contrast may have improved perceptibility of MW information within color but not within grayscale renderings.

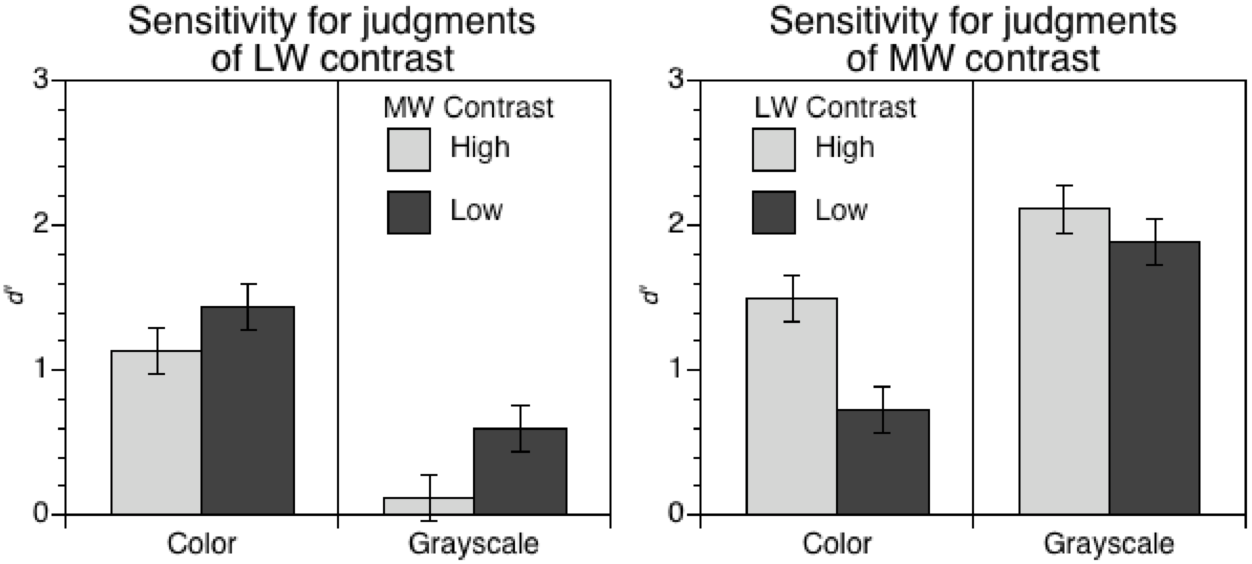

Figure 3. Sensitivity (*d'*) for judgments of LW contrast (left panel) and MW contrast (right panel).

DISCUSSION

Results indicate violations of perceptual separability in judgments of sensor-fused information content, demonstrating that sensitivity for contrast judgments of one component image can be affected by the contrast of the other component image. The discriminability of the information provided by one sensor, that is, can be modulated by the contrast level of input from the second sensor. This finding implies that image fusion may entail tradeoffs between the perceptibility of single-band input content. Interestingly, data also indicate that judgments of LW and MW content were both affected by chromatic rendering of fused images, though in opposite directions; sensitivity for judgments of LW contrast was increased by color rendering, while sensitivity for judgments of MW contrast was decreased. This pattern of effects echoes earlier results indicating that chromatic rendering of fused images may sometimes aid and sometimes hinder performance in a target detection task (McCarley & Krebs, 2000). Such findings indicate that color rendering does not always improve sensor-fused image quality, and suggest that designers should potentially allow operators to toggle between color and grayscale displays.

It is important to note that the specific pattern of effects seen here is of limited generalizability, at least for the moment. Results were obtained using a specific fusion technique, with input from a specific set of input bands, and with imagery depicting only a single scene. To draw broader conclusions it will be necessary to make psychophysical comparisons employing a larger range of fusion techniques, input bands, and forms of scene content. Additional research will be needed, moreover, to explore the relationship between information perceptibility as measured here and objective image quality metrics, and to determine how well each of those classes of measure predict human performance on specific real-world tasks (e.g., target detection, spatial judgments). However, the present data suggest that GRT may be valuable as a theoretical framework to help guide and unify the psychophysical study of sensor fusion.

REFERENCES

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154-179.

Dixon, T. D., Canga, E. F., Noyes, J. M., Troscianko, T., Nikolov, S. G., Bull, D. R., & Canagarajah, C. N. (in press). Methods for the assessment of fused images. *ACM Transactions on Applied Perception*.

Essock, E. A., McCarley, J. S., Sinai, M. J., & DeFord, J. K. (2001). Human perception of sensor-fused imagery. In R. R. Hoffman & A. B. Markman (Eds.), *Interpreting remote sensing imagery* (pp. 137-182). Boca Raton, FL: CRC Press.

Green, D. M., & Swets, J. A., (1966). *Signal detection theory and psychophysics*. Oxford, England: John Wiley.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science, 5*, 94-100.

Kadlec, H., & Townsend, J. T. (1992). Signal detection analyses of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181-227). Hillsdale, NJ: Erlbaum.

Krebs, W. K., & Sinai, M. J. (2002). Psychophysical assessments of image-sensor fused imagery. *Human Factors, 44*, 257-271.

McCarley, J. S., & Krebs, W. K. (2000). Visibility of road hazards in thermal, visible, and sensor-fused night-time imagery. *Applied Ergonomics, 31*, 523-530.

Scribner, D. A., Warren, P., Schuler, J., Satyshur, M., & Kruer, M. R. (1998). Infrared color vision: An approach to sensor fusion. *Optics and Photonics News, 8*, 27-32.

Toet, A., Ijspeert, J. K., Waxman, A. M., & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. In J. G. Verly (Ed.), *Proceedings of the SPIE-Enhanced and Synthetic Vision* (Vol. 3088, pp. 177-188). Bellingham, WA: SPIE-International Society for Optical Engineering.

Waxman, A. M., Gove, A. N., Fay, D. A., Racamoto, J. P., Carrick, J. E., Seibert, M. C., et al. (1997). Color night vision: Opponent processing in the fusion of visible and IR imagery. *Neural Networks, 10*, 1-6.