Reprinted from
*Journal of Imaging Science and Technology*® 61(1): 010401-1–010401-6, 2017.


# Gaze-Contingent Center-Surround Fusion of Infrared Images to Facilitate Visual Search for Human Targets

**Mackenzie G. Glaholt and Grace Sim**
*Defence Research and Development Canada, 1133 Sheppard Avenue W. Toronto, ON, M3K2C9, Canada*
*E-mail: Mackenzie.Glaholt@drdc-rddc.gc.ca*

**Abstract.** *We investigated gaze-contingent fusion of infrared imagery during visual search. Eye movements were monitored while subjects searched for and identified human targets in images captured simultaneously in the short-wave (SWIR) and long-wave (LWIR) infrared bands. Based on the subject's gaze position, the search display was updated such that imagery from one sensor was continuously presented to the subject's central visual field ("center") and another sensor was presented to the subject's non-central visual field ("surround"). Analysis of performance data indicated that, compared to the other combinations, the scheme featuring SWIR imagery in the center region and LWIR imagery in the surround region constituted an optimal combination of the SWIR and LWIR information: it inherited the superior target detection performance of LWIR imagery and the superior target identification performance of SWIR imagery. This demonstrates a novel method for efficiently combining imagery from two infrared sources as an alternative to conventional image fusion.* © *2017 Her Majesty the Queen in Right of Canada.*

## INTRODUCTION

Electro-optical sensor imaging technology has advanced greatly over the past three decades and infrared imaging is increasingly a feature in portable imaging devices. Devices that offer sensor imaging across multiple spectral bands (e.g., short-wave infrared, long-wave infrared, as well as visible spectrum images) are on the horizon. However, certain practical problems emerge with the presentation of information from multiple sensor imagers to the human viewer. Because the user can only view one image at a time, when images from multiple spectral bands are available the viewer must either toggle between them, or else the images must be merged in some way to create a composite. This latter approach is known as image fusion.

Image fusion seeks to combine the visual information from two or more images such that the unique task-relevant information from each of the source images is preserved and no artifacts or distortions are introduced.[1] A large literature is devoted to the development of algorithms that optimize the fusion of source imagery to achieve this outcome [Refs. 2–5; for reviews see Refs. 6–8]. One of the challenges for image fusion is that the visual information being combined from each source image competes for the same visual space in the fused image. Depending on the method employed, fusion can result in "destructive interference," where the information in each source image cancels the other, or task-relevant information is diluted or lost as a result of the fusion process.

In the present study we propose an alternative method for combining the information from two images in order to facilitate target detection and identification performance. Our method of combining information from two sensors represents a departure from conventional image fusion, for which visual information is combined across the entire image. Instead, we presented imagery from one source image to the viewer's central vision and imagery from a second source image to the viewer's non-central vision. In this way, the information from each source image is presented simultaneously, but to different areas of the viewer's visual field, and therefore the "fusion" occurs within the viewer's perceptual system.

More specifically, this method exploits functional characteristics of the human retina. Cone photoreceptors, which detect light under photopic conditions, are not uniformly distributed across the retina but are instead concentrated within an area of the retina known as the *fovea*, upon which the central few degrees of the visual field are projected.[9] The fovea has high visual acuity and is capable of resolving fine visual details.[10,11] Outside of the fovea, the density of cone photoreceptors drops off steeply, and consequently visual acuity decreases as retinal eccentricity increases.[12] However, non-central areas of the retina remain sensitive to visual information carried by lower spatial frequencies, as well as high contrast stimuli and motion [Ref. 13; for a review see Ref. 14]. Accordingly, during visual search for targets in natural scenes, salient areas in the non-central visual field are selected by the visual system, and eye movements serve to align the fovea with those areas for detailed inspection and target identification.[15,16]

We hypothesized that the presentation of sensor imagery supporting target detection and target identification to the viewer's non-central and central visual fields, respectively, would simultaneously facilitate the detection and identification of targets. To test this we employed a gaze-contingent display methodology [for reviews see Refs. 17–19], in which eye movements were monitored and the display was continuously updated according to the viewer's gaze position. Using this method we were able to present different imagery to the viewer's central and non-central visual fields, resulting in "center-surround fusion." Thus, in

Received July 4, 2016; accepted for publication Oct. 5, 2016; published online Dec. 8, 2016. Associate Editor: Henry Y. T. Ngan.

contrast to the conventional image fusion approach where visual information from single-band imagery is combined across the entire image, the center-surround fusion method presents single-band imagery simultaneously to different areas of the viewer's visual field.

In order to demonstrate the potential utility of this approach, we considered the fusion of two different infrared sensor imaging sources: long-wave infrared (8–12 μm; LWIR) and short-wave infrared (0.9–1.7 μm; SWIR) imagery. Pilot work in our laboratory confirmed that for LWIR imagery, human targets tend to exhibit high thermal contrast against a forested background and consequently LWIR is naturally optimized for the detection of human targets in this context. SWIR imagery, on the other hand, was found to produce poorer target detection performance than LWIR due to lower target-background contrast, though target identification performance in SWIR was far superior. This is likely because SWIR imagery tended to produce higher contrast in the spatial frequencies that are used for target identification. Based on these findings, we hypothesized that SWIR and LWIR imagery could be combined in a center-surround fusion scheme that would simultaneously optimize both target detection and identification. More specifically, based on these imagery characteristics and the physiological properties of the human retina, we predicted that the optimal center-surround fusion scheme would present SWIR imagery to central vision and LWIR imagery to non-central vision. Accordingly, we compared visual search performance in this condition with the reverse scheme (LWIR-center, SWIR-surround), and also the two conditions where the same imagery was presented in both areas of the visual field (SWIR-center, SWIR-surround; LWIR-center, LWIR-surround).

## METHOD

### *Subjects*

Sixteen male members of the Canadian Armed Forces participated in the experiment (mean age = 26 years, s.d. = 5.2, all right-handed, all normal or corrected to normal vision). Subjects provided informed consent and were remunerated according to Government of Canada Treasury Board guidelines for a total of $12.72 CAD for their 1 hour of experiment participation. The research protocol was reviewed and approved by the Human Research Ethics Committee at the DRDC Toronto Research Center.

### *Apparatus*

Eye movements were measured using an SR Research EyeLink 1000 system at 1000 Hz. Average calibrated gaze-position error was less than 0.5° and maximum error was less than 1°. The stimuli were presented on a BenQ 2420TX monitor (viewing distance = 685 mm; viewable area = 531 mm × 298 mm; 42.3° × 24.5°) with a refresh rate of 120 Hz and a screen resolution of 1920 × 1080 pixels. The experiment room was dimly lit and a chin rest with a head support was used to minimize head movement and ensure a consistent viewing distance. The experiment was implemented in SR Research Experiment Builder.

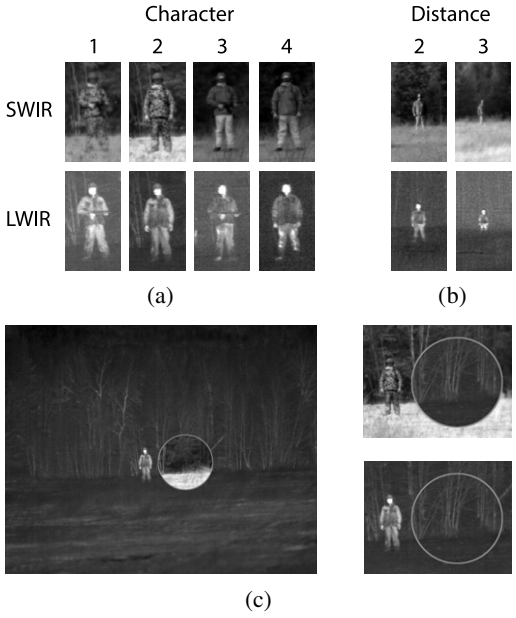


Figure 1. Examples of stimuli used in the experiment. Each of the Characters as they appeared in SWIR and LWIR imagery at Distance 1 (panel a), as well as the appearance of Character 4 at Distances 2 and 3 (panel b). An example of a visual search scene (slightly truncated in width and height) demonstrating the most efficient center-surround fusion scheme: SWIR-center, LWIR-surround (panel c, left). An example of the LWIR-center, SWIR-surround scheme (panel c, upper right) and the LWIR-center, LWIR-surround scheme (panel c, lower right).

### *Materials*

The stimuli used for the visual search task were photographed in rural Quebec, Canada. Photos were taken on a partly cloudy day in November, between the 1000 and 1500 hours, with an ambient temperature of 2–3°C. Each scene contained a single human target standing against a forested background. The target character was a male model dressed in one of four configurations (Figure 1, panel a): military uniform holding a weapon (Char. 1), military uniform without weapon (Char. 2), civilian clothing holding a weapon (Char. 3), civilian clothing without weapon (Char. 4). The target character was approximately 1.75 m tall and stood at one of three distances (100 m, 200 m, 300 m) away from the camera set-up.

The search scene was photographed in two spectral bands using a camera mounting system that housed a SWIR (spectral band = 0.9–1.7 μm; resolution = 1280 × 1024; field of view = 9.17° × 7.33°) and a LWIR (spectral band = 8–12 μm; resolution = 1024 × 768; field of view = 9.91° × 7.45°) camera. The cameras were oriented to capture, as closely as possible, the same field of view of the distal scene. The images were subjected to several pre-processing steps in Adobe Photoshop CS4. The LWIR images were up-sampled to the resolution of the SWIR images multiplied by the ratio of their fields of view (1381 × 1106). They were then aligned with the SWIR

images (target overlap matched by hand) and cropped down to the SWIR resolution (1280 × 1024). All images were collected as 16-bit grayscale images and were down-sampled to 8-bits and then contrast-adjusted manually in order to maximize the contrast of the target character. Accordingly, this produced two pixel-aligned versions of each scene: one captured in the SWIR band, and one captured in the LWIR band. At the time of presentation, images were centered within a gray background (R = G = B = 70) for a final image size of 1920 × 1080 pixels.

Based on the screen size and the subject's viewing distance, the 100 m, 200 m, and 300 m targets occupied 3.18°, 1.41°, and 0.88° of vertical visual angle for respectively, corresponding to apparent target distances of 31 m, 71 m, and 114 m (henceforth Distances 1, 2, 3). Each Character × Distance pairing was photographed against a different scene background (12 total), and for each of these unique scene backgrounds the target was photographed in three randomly chosen positions. This resulted in a total of 36 original scenes. In order to increase experimental power, each image was duplicated by mirroring in the vertical axis, to produce a total of six target positions (three photographed; three mirrored) for each character at each distance yielding 72 images in total.

### *Design*

Subjects carried out visual search in a gaze-contingent viewing mode. On each trial, two images were displayed: one image was drawn in the background and the other in the foreground (i.e., overtop). A mask was drawn that contained a 5° diameter (220 pixels) circular aperture, within which the foreground image remained visible and outside of which the background image was visible. The subject's eye movements were monitored and the position of the mask was updated during each display refresh cycle such that the position of the 5° aperture was centered at the subject's point of gaze within the display. In this way, the foreground image was continuously presented to the subject's central visual field ("center") and the background image was continuously presented to the subject's non-central ("surround") visual field. Note that with a 5° diameter the center display area encompasses the fovea (central 3°) and part of the parafovea (central 9° excluding fovea). This was done to accommodate eye tracking gaze-position error which could be as large as 1°; with a 5° diameter center area we ensured that imagery from the surround area was not cast upon the fovea. In all conditions the boundary between the foreground and background images was marked by a 2-pixel gray border, and consequently even for the cases where the center and surround images were the same, a 5° gray circle tracked the subject's gaze position on the screen. This was done to control for the potential distracting influence of a gaze-contingent visible edge during search.

In a 2 × 2 × 3 factorial design we crossed the sensor image in the Center display area [SWIR, LWIR] with the sensor image in the Surround display area [SWIR, LWIR], and the Distance [1, 2, 3] to the target. Each subject saw each scene four times, once in each sensor condition and in a random order for a total of 288 experimental trials. The order of scene presentation was counterbalanced across every four subjects such that each scene was equally likely to appear in each sensor condition first, second, third, and fourth.

### *Procedure*

The experimenter explained the general procedure of the experiment to the subject and the subject provided informed consent to participate. A 9-point eye-tracker calibration and validation test was completed for each subject to ensure an average gaze-prediction error of less than 0.5° and a maximum error of less than 1°. The subject was instructed to search for and identify the character present in the scenes as quickly and accurately as possible. Before each trial, subjects were presented with a familiarization screen showing all of the characters in each of the two sensor conditions at Distance 1 (e.g., Fig. 1, panel a) along with the mapping of each of the four characters to one of the buttons on the gamepad. Subjects pressed the "space" bar to proceed and a uniform gray screen appeared with an oval containing the words "look here" in one of the four corners of the image. Once the subject had fixated this start position for 300 ms the search scene was displayed. The start position for each scene was held constant across presentations, ensuring that the distance between the start position and the target was equated across sensor conditions. The trial was terminated by the subject's response on a gamepad, or else after 10 seconds. There were 8 practice trials followed by the 288 experimental trials, broken up into 8 blocks of 36 trials. At the end of each block, the subject was offered a break and an eye-tracker calibration was carried out if necessary. The entire procedure lasted approximately 1 hour.

## RESULTS

When analyzing the eye movement record for each visual search trial we considered only eye fixations (identified by the Eyelink parser) that began after the appearance of the visual search scene and prior to participant's button response. Trials in which the subject did not respond were excluded from analysis (<1%). In order to characterize the efficiency with which subjects searched for and identified the targets within each trial, we computed three measures of performance: detection interval, gaze duration on target, and response accuracy. For each measure we applied a 2 × 2 × 3 repeated measures ANOVA crossing Center [SWIR, LWIR], Surround [SWIR, LWIR], and Distance [1, 2, 3]. Of critical importance were effects or interactions involving Center and Surround.

The detection interval was defined as the latency following the onset of the stimulus display and until the target first entered the center area, which corresponds to the time at which the center pixel of the target first entered within 2.5° of the subject's point of fixation. As can be seen in Figure 2 (panel a), the detection interval depended primarily on the imagery presented in the surround area. In particular, the detection interval was shorter for the sensor conditions with LWIR imagery in the surround region ($F(1, 15) = 97.84$, $MSE = 4.26 \times 10^3$, $p < 0.001$).

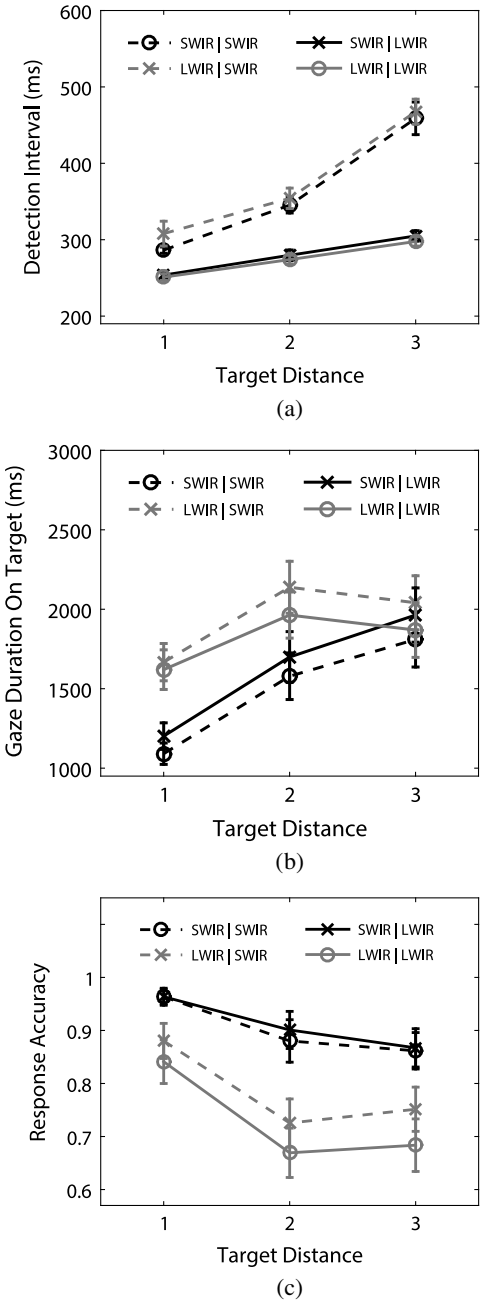


Figure 2. Measures of visual search performance: detection interval (panel a), cumulative gaze duration on target (panel b) and response accuracy (panel c). Legend entries are coded as Center|Surround. For convenient interpretation, dark lines represent SWIR-center, gray lines represent LWIR-center, solid lines represent LWIR-surround, dotted lines represent SWIR-surround, circles indicate center-surround schemes with the same imagery, and x's indicate schemes with different imagery.

This confirms our expectation that LWIR imagery would support faster target detection in this context. There was also a Surround × Distance interaction ($F(2, 30) = 46.02$, $MSE = 1.30 \times 10^3$, $p < 0.001$) where the difference between LWIR- and SWIR-surround increased as a function of target distance. This was primarily driven by a lengthening of the detection interval for the conditions with SWIR imagery in the surround region.

The effect of Center was not significant ($F(1, 15) = 1.23$, $MSE = 6.06 \times 10^2$, *n.s.*). There was, however, a Center × Surround interaction ($F(1, 15) = 8.45$, $MSE = 4.75 \times 10^2$, $p < 0.05$): there was a small penalty in performance for having different images in the center and surround display areas. Follow-up t-tests on Center at each level of Surround, collapsing across Distance, confirmed that detection occurred slightly later (13 ms averaged over Distances) for the LWIR-center, SWIR-surround condition than the SWIR-center, SWIR-surround condition ($t(15) = 2.06$, $SE = 6.35$, $p = 0.05$), and also slightly later (5 ms averaged over Distances) in the SWIR-center, LWIR-surround condition than the LWIR-center, LWIR-surround condition ($t(15) = 2.42$, $SE = 2.15$, $p < 0.05$). Despite these "mismatch costs," the pattern of detection intervals is clear: SWIR-center, LWIR-surround condition produced nearly identical performance to LWIR-center, LWIR-surround condition. The three-way interaction did not approach significance ($F < 1$).

Gaze duration on target was computed by summing the durations of all fixations for which the target was within the center area. As can be seen in Fig. 2 (panel b), gaze duration depended on the Center sensor content ($F(1, 15) = 12.42$, $MSE = 4.10 \times 10^5$, $p < 0.01$), where having SWIR sensor content in the center produced shorter gaze duration on target. This confirms that target identification occurred more rapidly when the target was viewed in SWIR imagery. This effect decreased over target distance, as was indicated by a Center × Distance interaction ($F(2, 30) = 6.98$, $MSE = 1.18 \times 10^5$, $p < 0.01$). The Surround sensor content did not have a significant effect ($F < 1$), though there was a significant Center × Surround interaction ($F(1, 15) = 12.94$, $MSE = 6.26 \times 10^4$, $p < 0.01$). Follow-up t-tests on Surround at each level of Center, collapsing across Distance, confirmed that gaze duration on target was longer (128 ms averaged over Distances) for SWIR-center, LWIR-surround than SWIR-center, SWIR-surround ($t(15) = 5.23$, $SE = 87.10$, $p < 0.01$), and also longer (131 ms averaged over Distances) for LWIR-center, SWIR-surround compared to LWIR-center, LWIR-surround ($t(15) = 2.63$, $SE = 49.87$, $p < 0.05$). As was the case for detection interval, this interaction reflected a mismatch penalty where gaze duration tended to be longer when the center and surround were different sensors compared to when they were the same. The three-way interaction was not significant ($F(2, 30) = 1.30$, $MSE = 2.46 \times 10^4$, *n.s.*).

Response accuracy was computed as the proportion of responses that correctly identified the target character. As can be seen in Fig. 2 (panel c) and consistent with the pattern of findings in gaze duration, response accuracy was higher overall when the Center was SWIR imagery versus LWIR imagery ($F(1, 15) = 16.32$, $MSE = 0.64$, $p < 0.01$), confirming that SWIR was a superior sensor for target identification. There was also a marginal effect of Surround ($F(1, 15) = 4.36$, $MSE = 0.006$, $p = 0.05$), where conditions with SWIR in the surround area produced higher response accuracy than those with LWIR in the surround area.

In addition, there was a Center × Surround interaction ($F(1, 15) = 11.92, MSE = 0.004, p < 0.01$). To interpret this interaction, we conducted follow-up t-tests on Surround at each level of Center, collapsing across Distance, and found that when LWIR imagery was presented in the center, there was significantly higher response accuracy (0.05 averaging over Distance) when SWIR was in the surround area compared to LWIR ($t(15) = 3.09$, $SE = 0.078$, $p < 0.01$). This suggests that in the LWIR-center, SWIR-surround condition, target information was extracted from the SWIR-surround area. Importantly, there was no effect of Surround when SWIR imagery was presented in the center ($t < 1$), confirming that the SWIR-center, LWIR-surround fusion scheme did not detrimentally affect response accuracy. The three-way interaction did not approach significance ($F < 1$).

## DISCUSSION

Presently we investigated gaze-contingent center-surround fusion of SWIR and LWIR imagery during visual search for human targets in natural scenes. Consistent with our initial hypotheses, we found that LWIR imagery produced the most efficient target detection while SWIR imagery produced superior target identification. Importantly, we found that the center-surround fusion scheme with SWIR in the center region and LWIR in the surrounding region supported both efficient target detection and identification, indicating that this combination successfully captured the functional benefits of each source image. Conversely, the fusion scheme with LWIR in the center and SWIR in the surrounding region produced poor target detection and identification performance.

This finding demonstrates a novel application of gaze-contingent displays in the context of image fusion. Whereas traditional image fusion techniques combine image information across the entire image and consequently are confronted with the problem of competing information at each spatial location in the fused image, the center-surround fusion scheme employed in the present study bypasses this issue by dividing the visual field and presenting single-band imagery to areas that are suited to particular types of information processing. In particular, the central visual field has high spatial acuity and is therefore suited to the processing of detailed visual information in support of target identification. The non-central visual field, while less suited to the processing of high-detail visual information, remains sensitive to salient luminance contrast and is therefore well suited to the detection of targets with high luminance contrast. We found that LWIR imagery produced superior target detection while SWIR imagery produced superior target identification, and consequently the presentation of these imagery sources to non-central and central vision, respectively, resulted in their effective combination in support of visual search.

Interestingly, we also found evidence of a performance cost associated with presenting different imagery to the central and non-central visual fields. In particular, for detection performance and gaze duration on target, there was a tendency for performance in the mismatched imagery conditions (i.e., different imagery in center and surround) to be slightly worse than the matched imagery conditions. The reason for this is not immediately clear, but it might stem from a delay in visual processing that occurs when the appearance of a target changes between its initial processing in non-central vision and subsequent processing in central vision. These mismatch effects were very small in detection interval (i.e., ∼10 ms), but were more substantial in gaze duration on target (i.e., ∼130 ms). In terms of response accuracy we observed a different pattern: the only mismatch effect was found for the LWIR-center conditions, where the SWIR-surround actually produced improved performance over the LWIR-surround. This indicates that processing for target identification can occur to some extent while the target is in the non-central visual field, and this appears to be more effective when the image information being processed is optimized for identification (e.g., SWIR). Importantly, response accuracy for the SWIR-center, LWIR-surround condition was not different from the SWIR-center, SWIR-surround condition, indicating the SWIR-center, LWIR-surround fusion scheme produces accurate (though slightly slower) target identification.

These findings point to several areas for further research. For example, there are multiple sources of imagery that could be explored in the context of center-surround fusion (e.g., visible spectrum imagery, as well as near-infrared imaging, and medium-wave infrared). These image sources might provide further optimization to visual search if deployed in a center-surround fusion scheme. In addition, the present stimulus set was limited to the context of visual search for human targets against a forested background. While this is an operationally relevant context for security, law-enforcement, and military, there are a wide variety of conceivable targets (e.g., vehicles, objects) and backgrounds that could be considered (e.g., urban, interiors), and the optimal center-surround sensor combinations might differ according to context. Finally, it might be possible to minimize mismatch effects by modulating the imagery in the center and surround fields. For example, one might present fused SWIR/LWIR imagery that, via weighted fusion, is biased toward SWIR in the center and LWIR in the surround. Such a scheme might be expected to minimize the discontinuity in perceptual processing of target information between the center and surround fields, while preserving the detection and identification performance benefits that were documented here.

In conclusion, we demonstrated a novel application of the gaze-contingent display technique to present visual information from two infrared sensors to a human viewer. Departing from conventional image fusion methods where visual information from two images is combined across the whole visual field, we presented visual information that facilitates target detection and identification to the viewer's non-central and central visual fields, respectively, and under these conditions we observed optimal visual search performance. In principle this center-surround fusion

approach could be used to simultaneously present any pair of sensor image sources that are optimized for the detection and identification of targets during visual search.


## ACKNOWLEDGMENTS

The authors would like to thank David Alain, Louis Durand, and Philips Laou for their assistance in collecting the imagery used for this study.


## REFERENCES


[1] A. Toet, M. A. Hogervorst, S. G. Nikolov, J. J. Lewis, T. D. Dixon, D. R. Bull, and C. N. Canagarajah, "Towards cognitive image fusion," Inf. Fusion **11**, 95–113 (2010).

[2] A. Apatean, A. Rogozan, and A. Bensrhair, "Visible-infrared fusion schemes for road obstacle classification," Transp. Res. C **35**, 180–192 (2013).

[3] W. Gan, X. Wu, W. Wu, X. Yang, C. Ren, X. He, and K. Liu, "Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter," Infrared Phys. Technol. **72**, 37–51 (2015).

[4] S. G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B. R. Abidi, A. Koschan, M. Yi, and M. A. Abidi, "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," Int. J. Comput. Vis. **71**, 215–233 (2007).

[5] V. Tsagaris and V. Anastassopoulos, "Fusion of visible and infrared imagery for night colour vision," Displays **26**, 191–196 (2005).

[6] Z. Omar and T. Stathaki, "Image fusion: An overview," *Proc. 5th Int'l. Conf. on Intelligent Systems, Modelling and Simulation* (IEE Computer Society, Langkawi, Malaysia, 2014), pp. 306–310.

[7] G. Piella, "A general framework for multiresolution image fusion: from pixels to regions," Inf. Fusion **4**, 259–280 (2003).

[8] M. I. Smith and J. P. Heather, "A review of image fusion technology in 2005," Proc. SPIE **5782**, 29–45 (2005).

[9] F. W. Weymouth, "Visual sensory units and the minimal angle of resolution," Am. J. Ophthalmology **46**, 102–113 (1958).

[10] L. Loschky, G. McConkie, J. Yang, and M. Miller, "The limits of visual resolution in natural scene viewing," Vis. Cogn. **12**, 1057–1092 (2005).

[11] L. A. Remington, *Clinical Anatomy of the Visual System*, 3rd ed. (Elsevier: St. Louis, Butterworth Heinemann, Ed., 2012).

[12] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling," J. Opt. Soc. Am. A **8**, 1175–1787 (1991).

[13] D. Finlay, "Motion perception in the peripheral visual field," Perception **11**, 457–462 (1982).

[14] H. Strasburger, I. Rentschler, and M. Jüttner, "Peripheral vision and pattern recognition: a review," J. Vis. **11**, 1–82 (2011).

[15] G. L. Malcolm and J. M. Henderson, "Combining top-down processes to guide eye movements during real-world scene search," J. Vis. **10**, 1–11 (2010).

[16] J. M. Henderson and A. Hollingworth, "High-level scene percetion," Ann. Rev. Psychol. **50**, 243–271 (1999).

[17] A. Duchowski, N. Cournia, and H. Murphy, "Gaze-contingent displays: A review," Cybern. Psychol. Behav. **7**, 621–634 (2004).

[18] E. M. Reingold, L. Loschky, G. W. McConkie, and D. M. Stampe, "Gaze-contingent multiresolutional displays: an integrative review," Hum. Factors **45**, 307–328 (2003).

[19] A. Toet, "Gaze directed displays as an enabling technology for attention aware systems," Comput. Hum. Behav. **22**, 615–647 (2006).