

Fusion of visible and thermal imagery improves situational awareness

A. Toet^{a,*}, J.K. IJspeert^a, A.M. Waxman^b, M. Aguilar^b

^a*TNO Human Factors Research Institute, Kampweg 5, 3769 DE Soesterberg, The Netherlands*

^b*MIT, Lincoln Laboratory, Lexington, MA 02173, USA*

Received 28 April 1997; accepted 18 August 1997

Abstract

A new color image fusion scheme is applied to visible and thermal images of military relevant scenarios. An observer experiment is performed to test if the increased amount of detail in the fused images can improve the accuracy of observers performing a detection and localization task. The results show that observers can localize a target in a scene (1) with a significantly higher accuracy, and (2) with a greater amount of confidence when they perform with fused images (either gray or color fused), compared with the individual image modalities (visible and thermal). © 1997 Elsevier Science B.V.

Keywords: Image fusion; Situational awareness; Thermal imagery; Visual imagery

1. Introduction

Scene analysis by a human operator may benefit from a combined or fused representation of images of the same scene taken in different spectral bands. For instance, after a period of extensive cooling (e.g. after a long period of rain or early in the morning) the visible bands may represent the background in great detail (vegetation or soil areas, texture), while the infrared bands are less detailed due to low thermal contrast in the scene. In this situation a target that is camouflaged for visual detection cannot be detected in the visible bands, but may be clearly represented in the infrared bands when it is warmer or cooler than its environment. The fusion of visible and thermal imagery on a single display may then allow both the detection and the unambiguous localization of the target (provided by the thermal image) with respect to the context (provided by the visible image). The above-mentioned line of reasoning is frequently adopted to promote image fusion, and has resulted in an increased interest in image fusion methods, as is reflected in a steadily growing number of publications on this topic [1–7]. A large effort has been spent on the development of these new image fusion methods. However, until now there are no validation studies that investigate the applicability domain and the practical use of these techniques. Ultimately the performance of a fusion process must be measured as the

degree to which it enhances a viewer's ability to perform certain practical tasks. Preliminary results on enhanced detection of targets embedded into real scenes [9] have shown potential benefits of fusion. The present study is performed (a) to investigate the conditions for which the fusion of visible and thermal images may result in a single composite image with extended information content, and (b) to test the capability of a recently developed color image fusion scheme [3,8–13] to enhance the situational awareness of observers operating under these specific conditions.

2. Methods

2.1. Image capture

2.1.1. Apparatus

The visible-light camera was a Siemens K235 Charge Coupled Device (CCD) video camera, with a 756×581 CCD chip, and equipped with a remotely controlled COSMICAR C10ZAME-2 (Asahi Precision Co. Ltd., Japan) zoom lens ($f = 10.5\text{--}105$ mm; 1:1.4). The infrared (IR) camera was an Amber Radiance 1 (Goleta, CA, USA) Focal Plane Array (FPA) camera, with an array of 256×256 pixels, operating in the $3\text{--}5\text{ }\mu\text{m}$ (mid-range) band, and equipped with a 100 mm $f/2.3$ Si-Ge lens. Each pixel corresponds to a square 1.3 min of arc wide instantaneous field of view. The entire array of 256×256 pixels therefore corresponds to a field of view about 5.6 degrees wide.

* Corresponding author. Tel.: 0031 3463 56237; fax: 0031 3463 53977; e-mail: toet@tm.tno.nl.

The CCD and IR images must be aligned before they can be fused. The signals of the thermal and visual cameras are therefore spatially registered as closely as possible. A second order affine warping transformation is applied to map corresponding points in the scene to corresponding pixel locations in the image plane.

2.1.2. Conditions

The recording period is just before sunrise, and the atmosphere was slightly hazy. The visual contrast is therefore low. The thermal contrast is low because most of the objects in the scene have about the same temperature after having lost their excess heat by radiation during the night.

2.2. Image fusion

The computational image fusion methodology is developed at the MIT Lincoln Laboratory [8–13] and derives from biological models of color vision and fusion of visible light and infrared (IR) radiation.

In the case of color vision in monkeys and man, retinal cone sensitivities are broad and overlapping, but the images are contrast enhanced within bands by spatial opponent processing (via cone–horizontal–bipolar cell interactions) creating both ON and OFF center–surround response channels [14]. These signals are then color-contrast enhanced between bands via interactions among bipolar, sustained amacrine, and single-opponent color ganglion cells [15,16], all within the retina. Further color processing in the form of double-opponent color cells is found in the primary visual cortex of primates (and the retinas of some fish). Opponent processing interactions form the basis of such percepts as color opponency, color constancy, and color contrast, though the exact mechanisms are not fully understood. (See section 4 of [3,17] for development of double-opponent color processing applied to multispectral IR target enhancement.)

Fusion of visible and thermal IR imagery has been observed in several classes of neurons in the optic tectum (evolutionary progenitor of the superior colliculus) of rattlesnakes (pit vipers), and pythons (boid snakes), as described by [18,19]. These neurons display interactions in which one sensing modality (e.g. IR) can enhance or depress the response to the other sensing modality (e.g. visible) in a strongly nonlinear fashion. These tectum cell responses relate to (and perhaps control) the attentional focus of the snake as observed by its striking behavior. This discovery predates the observation of bimodal visual/auditory fusion cells observed in the superior colliculus [20]. Moreover, these visible/IR fusion cells are suggestive of ON and OFF channels feeding single-opponent color-contrast cells. This strategy forms the basis of our computational model.

There are also physical motivations for our approach to fusing visible and IR imagery, revealed by comparing and contrasting the rather different needs of a vision system that

processes reflected visible light (in order to deduce reflectivity ρ) versus one that processes emitted thermal IR light (in order to deduce emissivity ϵ). For opaque surfaces in thermodynamic equilibrium, spectral reflectivity ρ and emissivity ϵ are linearly related at each wavelength λ : $\rho(\lambda) = 1 - \epsilon(\lambda)$. This provides a rationale for the use of both on-center and off-center channels when treating infrared imagery as characterized by thermal emissivity. Thus, it is not surprising that often FLIR imagery looks more ‘natural’ when viewed with reverse polarity (black hot as opposed to white hot, suggestive of off-channel processing [14]). This simple relation strongly suggests that processing anatomies designed to determine reflectivity may also be well suited for determining emissivity; and so therefore will be computational models of these anatomies.

Fig. 1 represents the multiple stages of processing in the visible/IR fusion architecture. They mimic both the structure and function of the layers in the retina (from the rod and cone photodetectors through the single-opponent color ganglion cells) which begin the parvocellular stream of form and color processing. The computational model that underlies all the opponent processing stages utilized here, is the feedforward center–surround shunting neural network of Grossberg [21,22]. This type of processing serves (i) to enhance spatial contrast within the separate visible and IR bands, (ii) to create both positive (ON-IR) and negative (OFF-IR) polarity IR contrast images, and (iii) to create two types of single-opponent color-contrast images. These opponent-color images already represent fusion of visible and IR imagery in the form of grayscale image products. However, the two opponent-color images together with the enhanced visible image form a triple which can be presented as a fused color image product.

The neurodynamics of our center–surround receptive fields at pixel with integer coordinates ij is described by the following equations:

$$\frac{dE_{ij}}{dt} = -AE_{ij} + (1 - E_{ij})[CIC]_{ij} - (1 + E_{ij})[G_S * I^S]_{ij} \quad (1)$$

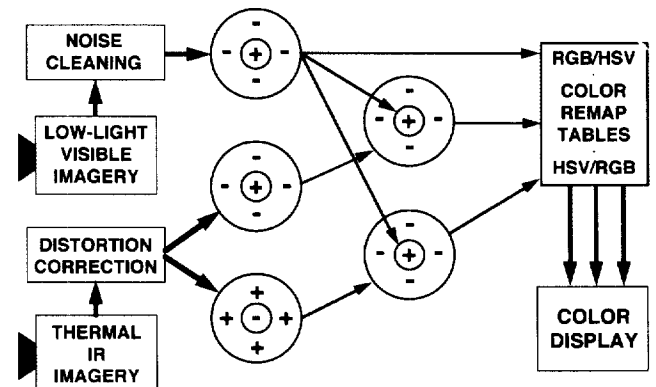


Fig. 1. Schematic representation of the visible/IR fusion architecture based on principles of opponent processing within and between the different image modalities.

$$E_{ij} = \frac{[CI^C - G_S * I^S]_{ij}}{A + [CI^C + G_S * I^S]_{ij}} \quad (2)$$

where E represents the opponent processed enhanced image, I^C is the input image that excites the single pixel em center of the receptive field (a single pixel center is used to preserve resolution of the processed images), and I^S is the input image that inhibits the gaussian surround G_S of the receptive field (see Fig. 1 for a schematical representation of the computational image fusion process). Eq. (1) describes the temporal dynamics of a charging neural membrane (cp. capacitor) which leaks charge at rate A , and has excitatory and inhibitory input ion currents determined by Ohm's law (the shunting coefficients $(1 \pm E)$ act as potential differences across the membrane, and the input image signals modulate the ion selective membrane conductances). Eq. (2) describes the equilibrium that is rapidly established at each pixel (i.e. at frame rate), and defines a type of nonlinear image processing with parameters A , C , and the size of the gaussian surround. The shunting coefficients of Eq. (1) clearly imply that the dynamic range of the enhanced image E is bounded, $-1 < E < 1$, regardless of the dynamic range of the input imagery. When the imagery which feeds the center and surround is taken from the same input image (visible or IR), the numerator of Eq. (2) is the familiar difference-of-gaussians filtering which, for $C > 1$, acts to boost high spatial frequencies superimposed on the background. The denominator of Eq. (2) acts to adaptively normalize this contrast enhanced imagery based on the local mean. In fact, Eq. (2) displays a smooth transition between linear filtering (when A exceeds the local mean brightness, such as in dark regions) and ratio processing (when A can be neglected as in bright regions of the imagery). This is particularly useful for processing the wide dynamic range visible imagery obtained with low-light CCDs. Eq. (2) is used to process separately the input visible and IR imagery. These enhanced visible and ON-IR images are reminiscent of the lightness images postulated in Land's retinex theory [23] (see also [21] on discounting the illuminant).

A modified version of Eqs. (1) and (2), with an inhibitory center and excitatory surround, is also used to create an enhanced OFF-IR image (i.e. a reverse polarity enhanced IR image). Following noise cleaning of the imagery (both realtime median filtering and non-realtime Boundary Contour/Feature Contour System processing [21,17] have been explored), and distortion correction to ensure image registration, two grayscale fused single-opponent color-contrast images are formed using Eq. (2) with the enhanced visible feeding the excitatory center and the enhanced IR (ON-IR and OFF-IR, respectively) feeding the inhibitory surround. In analogy to the primate opponent-color cells [16], we label these two single-opponent images $+Vis - IR$ and $+Vis + IR$. In all cases, we retain only positive responses for these various contrast images. Additional application of Eq. (2) to these two single-opponent images serves to sharpen their appearance, restoring their resolution to the higher of the

two (usually visible) input images. These images then represent a simple form of double-opponent color-contrast between visible and ON/OFF-IR.

The two opponent-color contrast images are analogous to the IR-depressed-visual and IR-enhanced-visual cells, respectively, of the rattlesnake [18,19]; they even display similar nonlinear behavior. In fact, with the IR image being of lower resolution than the visible image (in the snake, and for man-made uncooled IR imagers), a single IR pixel may sometimes be treated as a small surround for its corresponding visible pixel. In this context, the opponent-color contrast images can also be interpreted as coordinate rotations in the color space of visible versus IR, along with local adaptive scalings of the new color axes. Such color space transformations are fundamental to Land's [23–25] analyses of his dual-band red and white colorful imagery.

To achieve a natural color presentation of these opponent images (each being an 8-bit grayscale image), the enhanced visible is assigned to the green channel, the difference signal of the enhanced visible and IR images is assigned to the blue channel, and the sum of the visible and IR images is assigned to the red channel of an RGB display [3,8–12,17]. Finally, these three channels can be interpreted as R, G, B inputs to a color remapping stage in which, following conversion to H, S, V (hue, saturation, value) color space, hues can be remapped to alternative 'more natural' hues, colors can be desaturated, and then reconverted to R, G, B signals to drive a color display. The result is a fused color presentation of visible/IR imagery.

At MIT-LL this approach to image fusion has been implemented on a dual-C80 hardware platform capable of processing 30 fps at 640×480 resolution.

The grayscale fused images are produced by taking the luminance component of the corresponding color fused images.

2.3. Stimuli

The stimuli used in this experiment are five different types of images:

- graylevel images representing the signal of the video (CCD) camera,
- graylevel images representing the signal of the infrared (IR) camera,
- color images representing the result of the fusion of corresponding CCD and IR image pairs (i.e. the combination of CCD and IR images of the same scene and registered at the same instant),
- graylevel images representing luminance component of the abovementioned color fused images, and
- schematic graylevel images, representing segmented versions of the original visual (CCD) images.

The graylevel images are quantized to 8 bits. The color images are quantized to 24 bits (8 bits for each of the RGB channels), and bitmapped to a 256 color map, adding some color dither.

The individual images correspond to successive frames of a time sequence. The time sequences represent 3 different scenarios. These scenarios were developed by the Royal Dutch Army [26]. They simulate typical surveillance tasks and were chosen because of their military relevance.

The corresponding schematic images are constructed from the visual images by

- applying standard image processing techniques like histogram equalization and contrast stretching to enhance the representation of the reference contours in the original visual images,
- drawing the contours of the reference features (judged by eye) on a graphical overlay on the contrast enhanced visual images, and
- filling the contours with a homogeneous graylevel value.

The images thus created represent segmented versions of the visual images.

Scenario I corresponds to the guarding of a UN camp [26], and involves monitoring a fence that encloses a military asset. To distinguish innocent bypassers from individuals planning to perform subversive actions the guard must be able to determine the exact position of a person in the scene at any time. During the image acquisition period the fence is clearly visible in the CCD image. In the IR image however, the fence is merely represented by a vague haze. A person (walking along the fence) is clearly visible in the IR image but can hardly be distinguished in the CCD image. In the fused images both the fence and the person are clearly visible. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the fence.

Scenario II corresponds to guarding a temporary base [26]. Only a small section of the dune like terrain is visible, the rest is occluded by trees. The assignment of the guard is to detect and counter infiltration attempts in a very early stage. During the registration period the trees appear larger in the IR image than they really are because they have nearly the same temperature as their local background. In the CCD image however, the contours of the trees are correctly represented. A person (crossing the interval between the trees) is clearly visible in the IR image but is represented with low contrast in the CCD image. In the fused images both the outlines of the trees and the person are clearly visible. As a result it is difficult to determine the position of the person relative to the trees using either the CCD or the IR images. The fused images correctly represent both the contours of the trees and the person. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the midpoint of the interval delineated by the contours of the trees that are positioned on both sides of the person.

Scenario III corresponds to the surveillance of a large area [26]. The scene represents a dune landscape, covered with semi-shrubs and sandy paths. The assignment of the guard is to detect any attempt to infiltrate a certain area.

During the registration period the sandy paths in the dune area have nearly the same temperature as their local background, and are therefore represented with very low contrast in the IR image. In the CCD image however, the paths are depicted with high contrast. A person (walking along a trajectory that intersects the sandy path) is clearly visible in the IR image but is represented with less contrast in the CCD image. In the fused images both the outlines of the paths and the person are clearly visible. It is difficult (or even impossible) to determine the position of the person relative to the sandy path he is crossing from either the IR or the CCD images. An observer's situational awareness can therefore be tested by asking the subject to report the position of the person relative to the sandy path.

2.4. Apparatus

A Pentium 100 MHz computer, equipped with a Diamond SVGA board, is used to present the stimuli, measure the response times and collect the observer responses. The stimuli are presented on a 17 inch Vision Master (Iiyama Electric Co., Ltd) color monitor, using the 640×480 pixels mode and a 100 Hz refresh rate.

2.5. Procedure

The subject's task is to assess from each presented image the position of the person in the scene relative to the reference features.

In Scenario I the reference features are the poles that support the fence. These poles are clearly visible in the CCD images (Fig. 2(a)), but not represented in the IR images (Fig. 2(b)) because they have almost the same temperature as the surrounding terrain. Fig. 2 includes the contrast enhanced versions of the CCD and IR images (Fig. 2(c) and Fig. 2(d), respectively) to illustrate this fact. In the (graylevel and color) fused images (Fig. 2(e) and Fig. 2(f)) the poles are again clearly visible.

In Scenario II the outlines of the trees serve to delineate the reference interval. The contours of the trees are correctly represented in the CCD images (Fig. 3(a)). However, in the IR images (Fig. 3(b)) the trees appear larger than their physical size because they almost have the same temperature as the surrounding soil. This effect is illustrated in Fig. 3(c) and Fig. 3(d), which are the contrast enhanced versions of respectively the CCD and IR images from Fig. 3(a) and Fig. 3(b). As a result, the scene is incorrectly segmented after quantization and it is not possible to perceive the correct borders of the area between the trees. In the (graylevel and color) fused images (Fig. 3(e) and Fig. 3(f)) the outlines of the trees are again correctly represented and clearly visible.

In Scenario III the area of the small and winding sandy path provides a reference contour for the task at hand. This path is represented at high contrast in the CCD images (Fig. 4(a)), but it is not represented in the IR images



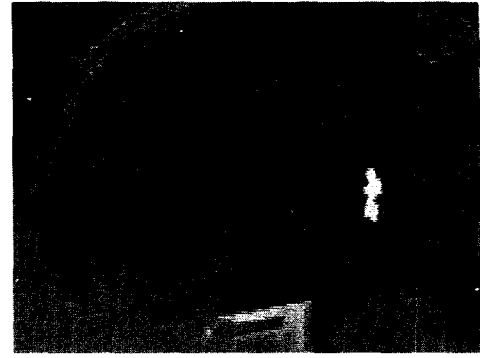
(a)



(b)



(c)



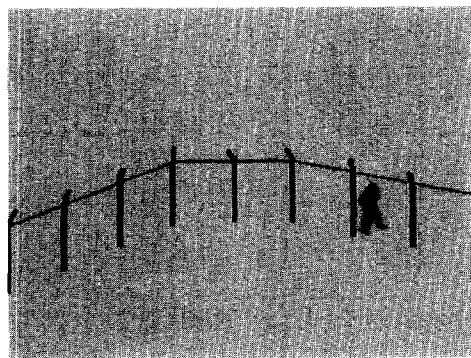
(d)



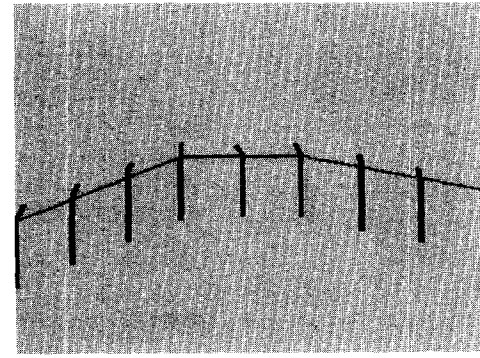
(e)



(f)



(g)



(h)

Fig. 2. (a) Original CCD, (b) original IR, (c) contrast enhanced CCD, (d) contrast enhanced IR, (e) graylevel fused, (f) color fused, (g) baseline test, and (h) reference images of Scenario I.

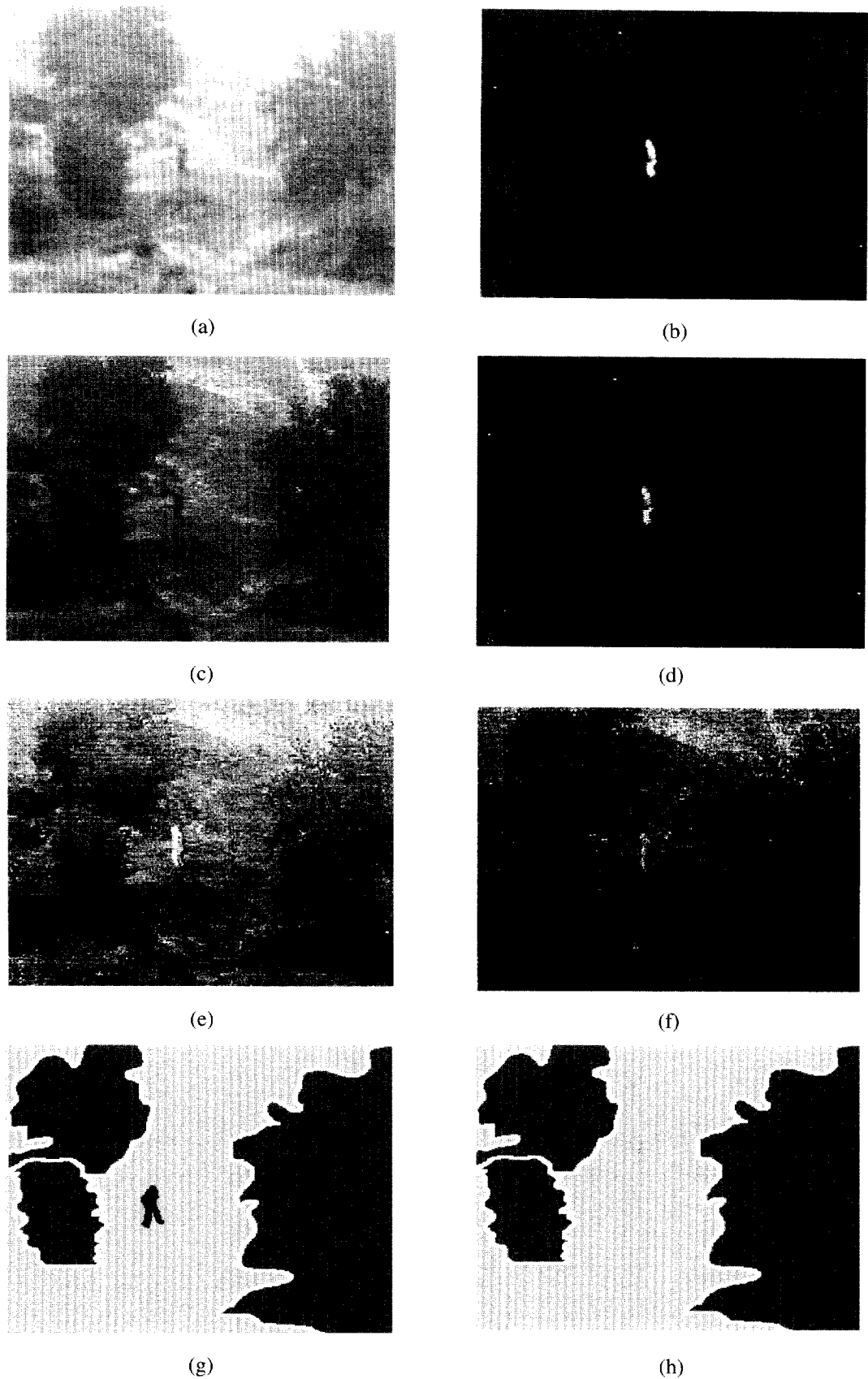
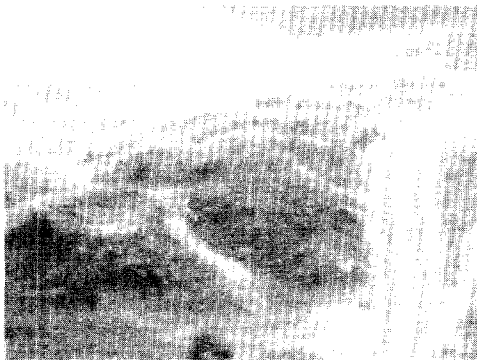
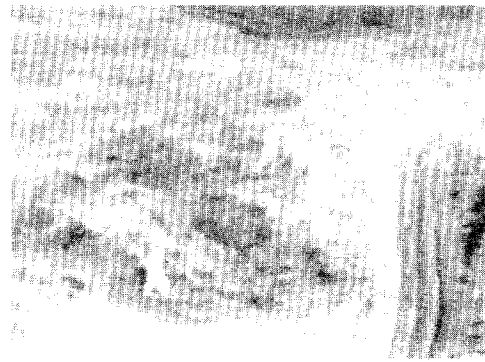


Fig. 3. (a) Original CCD, (b) original IR, (c) contrast enhanced CCD, (d) contrast enhanced IR, (e) graylevel fused, (f) color fused, (g) baseline test, and (h) reference images of Scenario II.



(a)



(b)



(c)



(d)



(e)



(f)



(g)



(h)

Fig. 4. (a) Original CCD, (b) original IR, (c) contrast enhanced CCD, (d) contrast enhanced IR, (e) graylevel fused, (f) color fused, (g) baseline test, and (h) reference images of Scenario III.

(Fig. 4(b)) because it has the same temperature as the surrounding soil. Again, this effect can be seen more easily in Fig. 4(c) and Fig. 4(d), which are the contrast enhanced versions of respectively the CCD and IR images from Fig. 4(a) and Fig. 4(b). In the (graylevel and color) fused images (Fig. 4(e) and Fig. 4(f)) the path is again clearly visible.

For each scenario a total of nine frames is used in the experiment. In each frame the person is at a different location relative to the reference features. These different locations are equally distributed relative to the reference features. Each stimulus is presented for 1 s, centered on the midpoint of the screen, preceded by a blank screen with an awareness message, which is presented for 1 s. A schematical representation of the reference features is shown immediately after each stimulus presentation (Figs. 2–4(h)). The position of the center of the reference image is randomly displaced around the center of the screen between presentations to ensure that subjects can not use prior presentations as a frame of reference for detection and localization.

The subject's task is to indicate the perceived location of the person in the scene by placing a mouse controlled cursor at the corresponding location in this schematical drawing. The subject has in principle unlimited time to reach a decision. When the left mouse button is pressed the computer registers the coordinates corresponding to the indicated image location (the mouse coordinates) and computes the distance in the image plane between the actual position of the person and the indicated location. The subject presses the right mouse button if the person in the displayed scene has not been detected. The subject can only perform the localization task by memorising the perceived position of the person relative to the reference features.

The schematic reference images are also used to determine the optimal (baseline) localization accuracy of the observers. For each of the three scenarios a total of 9 baseline test images (Figs. 2–4(g)) are created by placing a binary (dark) image of a walking person at different locations in the reference scene. The different locations of the person in these images are equally distributed over the entire reference interval. The image of the walking person was extracted from a thresholded and inverted thermal image. In the resulting set of schematic images both the reference features and the person are highly visible. Also, there are no distracting features in these images that may degrade localization performance. Therefore, observer performance for these schematic test images should be optimal and may serve as a baseline to compare performance obtained with the other image modalities.

A complete run consists of 135 presentations (5 image modalities \times 3 scenarios \times 9 frames per scenario), and typically lasts about one hour.

2.6. Subjects

A total of 6 subjects, aged between 20 and 30 years, serve

in the experiments reported below. All subjects have normal (or corrected to normal) vision, and no known color deficiencies.

2.7. Viewing conditions

Viewing is binocular. The experiments are performed in a dimly lit room. The images are projected onto the screen of a CRT display. This screen subtends a viewing angle of 25.5×19.5 degrees at a viewing distance of 0.60 m.

3. Results

Fig. 5 shows that subjects are uncertain about the location of the person in the scene for about 26% of the visual image presentations and 22% of the thermal image presentations. The (graylevel and color) fused images result in a smaller fraction of about 13% 'not sure' replies. The lowest number of 'not sure' replies is obtained for the baseline reference images: only about 4%. This indicates that the increased amount of detail in fused imagery does indeed improve an observer's subjective situational awareness.

Fig. 6 shows the mean weighted distance between the actual position of the person in each scene and the position indicated by the subjects (the perceived position), for the visual (CCD) and thermal (IR) images, and for the gray-level and color fusion schemes. This Figure also shows the optimal (baseline) performance obtained for the schematic test images representing only the segmented reference features and the walking person. A low value of this mean weighted distance measure corresponds to a high observer accuracy and a correctly perceived position of the person in the displayed scenes relative to the main reference features. High values correspond to a large

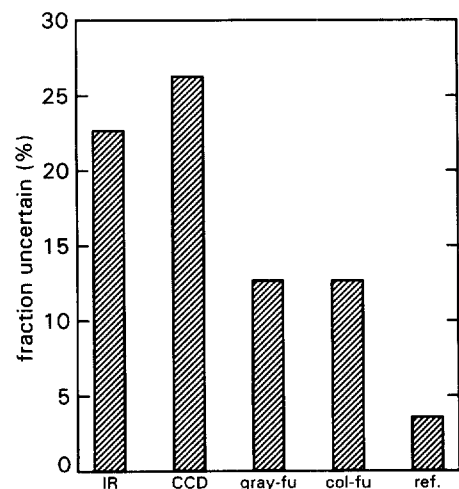


Fig. 5. The percentage of image presentations in which observers are uncertain about the relative position of the person in the scene, for each of the 5 image modalities tested (IR, CCD, graylevel fused, color fused, and schematical reference images).

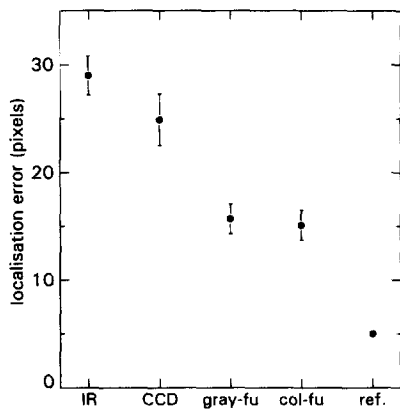


Fig. 6. The mean weighted distance between the actual position of the person in each scene and the perceived position for each of the 5 image modalities tested (IR, CCD, graylevel fused, color fused, and schematic reference images). The error bars indicate the size of the standard error in the perceived location.

discrepancy between the perceived position and the actual position of the person. In all scenarios the person was at approximately 300 m distance from the viewing location. At this distance one pixel corresponds to 11.4 cm in the field.

Fig. 6 shows that the localization error obtained with the fused images is significantly lower than the error obtained with the individual thermal and visual image modalities ($p = 0.0021$). The smallest errors in the relative spatial localization task are obtained for the schematic images. This result represents the baseline performance, since the images are optimal in the sense that they do not contain any distracting details and all the features that are essential to perform the task (i.e. the outlines of the reference features) are represented at high visual contrast. The lowest overall accuracy is achieved for the thermal images. The visual images appear to yield a slightly higher accuracy. However, this accuracy is misleading since observers are not sure about the person in a large percentage of the visual images, as shown by Fig. 5. The difference between the results for the graylevel fused and the color fused images is not significant ($p = 0.134$), suggesting that spatial localization of targets (following detection) does not exploit color contrast as long as there exists sufficient brightness contrast in the gray fused imagery.

4. Discussion

This study investigates (a) for which conditions the fusion of visual and thermal images results in a single composite image with extended information content, and (b) whether a recently developed color image fusion scheme [3,8–11,17] can enhance the situational awareness of observers operating under these specific conditions and using visual and thermal images.

Conditions in which fusion of visual and thermal imagery are most likely to result in images with increased information content occur around sunrise. At this time the contrast of both the visual and the thermal images is very low. One can construct other scenarios involving night operations in which both modalities are lacking in contrast. The visual contrast is low around sunrise because of the low luminance of the sky. However, contours of extended objects are still visible. After some image enhancement (like center-surround shunting, histogram equalization or contrast stretching) even an appreciable amount of detail can be perceived. Small objects with low reflectance, like a person wearing a dark suit or camouflage clothing, or objects that are partly obscured, are not represented in the visual image under these conditions, and can therefore not be detected. The thermal contrast is low around sunrise because most of the objects in the scene have about the same temperature after losing their excess heat by radiation during the night. As a result the contours of extended objects are not at all or incorrectly represented in the thermal image. The fusion of images registered around sunrise should therefore result in images that represent both the context (the outlines of extended objects) and the details with a large thermal contrast (like people) in a single composite image. To test this hypothesis a large set of image sequences is captured around sunrise on different days. The scenes used in this study represent 3 different scenarios that were developed by the Royal Dutch Army [26]. The images are fused using the recently developed MIT color fusion scheme [3,8–11,17]. Graylevel fused images are also produced by taking the luminance component of the color fused images. Visual inspection of the results shows that the fusion of thermal and visual images indeed results in composite images with an increased amount of information.

An observer experiment is performed to test if the increased amount of detail in the fused images can yield an improved observer performance in a task that requires a certain amount of situational awareness. The task that is devised involves the detection and localization of a person in the displayed scene relative to some characteristic details that provide the spatial context. The person is optimally represented in the thermal imagery and the reference features are better represented in the visual imagery. The hypothesis is therefore that the fused images provide a better representation of the overall spatial structure of the depicted scene. To test this hypothesis subjects perform a relative spatial localization task with a selection of thermal, visual, and (both graylevel- and color-) fused images representing the abovementioned military scenarios. The results show that observers can indeed determine the relative location of a person in a scene with a significantly higher accuracy when they perform with fused images, compared with the individual image modalities.

This study shows no significant difference between the localization performance with color fused images and with

their luminance components (the derived graylevel fused images). However, in some conditions color fused images are easier to visually segment than graylevel fused images. As a result, color coding may greatly improve the speed and accuracy of information uptake [27], and fewer fixations may be required to locate color coded targets [28]. Therefore, dynamic tasks like navigation and orienting, that probably depend on a quick and correct scene segmentation, may benefit from a color fused image representation.

It is likely that the fusion of thermal and low-light level imagery may yield an even better observer performance over extended exposure times which often lead to exhaustion or distraction. Further research, preferably involving dynamic scenarios, is needed to test the hypothesis that color image fusion schemes can boost observer performance in these tasks. Also, it needs to be investigated whether there are circumstances under which degradation in image interpretation performance may occur (e.g. situations where the IR imagery introduces clutter).

5. Conclusions

The fusion of thermal and visual images registered around sunrise results in composite images with an increased amount of detail that clearly represent all details in their correct spatial context.

Observers can localize a target in a scene (1) with a significantly higher accuracy, and (2) with a greater amount of confidence when they perform with fused images (either gray or color fused), compared with the individual image modalities (visible and thermal).

Acknowledgements

TNO Human Factor Research Institute was sponsored in part by the Royal Netherlands Air Force and by the Royal Netherlands Army. MIT Lincoln Laboratory was sponsored in part by the US Office of Naval Research and the US Air Force, under Air Force Contract F19628-95-C-0002.

References

- [1] A. Toet, J. Walraven, New false colour mapping for image fusion, *Optical Engineering* 35 (3) (1996) 650–658.
- [2] P.J. Burt, R.J. Kolczynski, Enhanced image capture through fusion, in *Proceedings of the Fourth International Conference on Computer Vision*, IEEE Computer Society Press, Washington, USA, 1993, pp. 173–182.
- [3] A.N. Gove, Cunningham, R.K., A.M. Waxman, Opponent-color visual processing applied to multispectral infrared imagery, in *Proceedings of 1996 Meeting of the IRIS Specialty Group on Passive Sensors II, Infrared Information Analysis Center, ERIM, Ann Arbor, US*, 1996, pp. 247–262.
- [4] H. Li, B.S. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, *Graphical Models and Image Processing* 57 (1995) 235–245.
- [5] A. Toet, Hierarchical Image Fusion, *Machine Vision and Applications* 3 (1990) 1–11.
- [6] A. Toet, L.J. van Ruyven, J.M. Valetton, Merging thermal and visual images by a contrast pyramid, *Optical Engineering* 28 (1989) 789–792.
- [7] T.A. Wilson, S.K. Rogers, L.R. Myers, Perceptual-based hyperspectral image fusion using multiresolution analysis, *Optical Engineering* 34 (1995) 3154–3164.
- [8] A.M. Waxman, D.A. Fay, A.N. Gove, M. Seibert, J.P. Racamoto, J.E. Carrick, E.D. Savoye, Color night vision: fusion of intensified visible and thermal IR imagery, in *Proceedings of the SPIE Conference on Synthetic Vision for Vehicle Guidance and Control*, Vol. SPIE-2463, 1995, pp. 58–68.
- [9] A.M. Waxman, A.N. Gove, M. Seibert, D.A. Fay, J.E. Carrick, J.P. Racamoto, E.D. Savoye, B.E. Burke, R.K. Reich, W.H. McGonagle, D.M. Craig, Progress on color night vision: visible/IR fusion, perception and search, and low-light CCD imaging, in *Proceedings of the SPIE Conference on Enhanced and Synthetic Vision*, Vol. SPIE-2736, 1996, pp. 96–107.
- [10] A.M. Waxman, A.N. Gove, D.A. Fay, J.P. Racamoto, J.E. Carrick, M. Seibert, E.D. Savoye, B.E. Burke, R.K. Reich, W.H. McGonagle, D.M. Craig, Solid state color night vision: fusion of low-light visible and thermal IR imagery, in *Proceedings of the 1996 Meeting of the IRIS Specialty Group on Passive Sensors II, Infrared Information Analysis Center, ERIM, Ann Arbor*, 1996, pp. 263–280.
- [11] A.M. Waxman, J.E. Carrick, D.A. Fay, J.P. Racamoto, M. Aguilar, E.D. Savoye, Electronic imaging aids for night driving: low-light CCD, thermal IR, and color fused visible/IR, in *Proceedings of the SPIE Conference on Transportation Sensors and Controls*, Vol. SPIE-2902 (1996).
- [12] A.M. Waxman, A.N. Gove, D.A. Fay, J.P. Racamoto, J.E. Carrick, M. Seibert, E.D. Savoye, Color night vision: opponent processing in the fusion of visible and IR imagery *Neural Networks* 10 (1) (1997) 1–6.
- [13] A.M. Waxman, J.E. Carrick, J.P. Racamoto, D.A. Fay, M. Aguilar, E.D. Savoye, Color night vision—3rd update: Realtime fusion of low-light CCD visible and thermal IR imagery, in *Proceedings of the SPIE Conference on Enhanced and Synthetic Vision*, Vol. SPIE-3088 (1997).
- [14] P. Schiller, The ON and OFF channels of the visual system, *Trends in Neuroscience* 15 (1992) 86–92.
- [15] P. Schiller, N.K. Logothetis, The color-opponent and broad-band channels of the primate visual system, *Trends in Neuroscience* 13 (1990) 392–398.
- [16] P. Gouras, Color vision, in: E.R. Kandel, J.H. Schwartz, T.M. Jessell (Eds.), *Principles of Neural Science*, 3rd ed., Elsevier, Oxford, UK, 1991, pp. 467–480.
- [17] A.M. Waxman, M.C. Seibert, A.N. Gove, D.A. Fay, A.M. Bernardon, W.R. Steele, R.K. Cunningham, Neural processing of targets in visible, multispectral IR and SAR imagery *Neural Networks* 8 (1995) 1029–1051.
- [18] E.A. Newman, P.H. Hartline, Integration of visual and infrared information in bimodal neurons of the rattlesnake optic tectum *Science* 213 (1981) 789–791.
- [19] E.A. Newman, P.H. Hartline, The infrared vision of snakes, *Scientific American* 246 (1982) 116–127.
- [20] A.J. King, The integration of visual and auditory spatial information in the brain, in: D.M. Guthrie (Ed.), *Higher Order Sensory Processing*, Manchester University Press, Manchester, UK, 1990, pp. 75–113.
- [21] S. Grossberg, *Neural Networks and Natural Intelligence*, MIT Press, Cambridge, MA, 1988.
- [22] S.A. Elias, S. Grossberg, Pattern formation, contrast control, and oscillations in the short memory of shunting on-center off-surround networks, *Biological Cybernetics* 20 (1975) 69–98.
- [23] E.H. Land, Recent advances in retinex theory and some implications

- for cortical computations: Color vision and the natural image Proceedings of the National Academy of Sciences of the USA 80 (1983) 5163–5169.
- [24] E.H. Land, Color vision and the natural image, Part I Proceedings of the National Academy of Sciences 45 (1959) 115–129.
- [25] E.H. Land, Experiments in color vision, Scientific American 200 (5) (1959) 84–99.
- [26] A.R. Buimer, Scenarios for multi-spectral image fusion, Memo 21 December 1993, Section Planning, Training Center Infantry (OCI), Harderwijk, The Netherlands, 1993 (in Dutch).
- [27] R.E. Christ, Review and analysis of colour coding research for visual displays Human Factors 17 (1975) 542–570.
- [28] P.K. Hughes, D.J. Creed, Eye movement behaviour viewing colour-coded and monochrome avionic displays, Ergonomics 37 (1994) 1871–1884.