Perceptual evaluation of different image fusion schemes

Alexander Toet[a],*, Eric M. Franken[b],1

[a]TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands
[b]TNO Physics and Electronics Laboratory, Oude Waalsdorperweg 63, 2509 JG The Hague, The Netherlands

Abstract

Human scene recognition performance was tested with images of night-time outdoor scenes. The scenes were registered both with a dual band (visual and near infrared) image intensified low-light CCD camera (DII) and with a thermal middle wavelength band ($3-5$ μm) infrared (IR) camera. Fused imagery was produced through a grayscale pyramid image merging scheme, in combination with two different colour mappings. Observer performance was tested for each of the (individual and fused) image modalities. The results show that DII imagery contributes most to global scene recognition (situational awareness), whereas IR imagery serves best for the detection and recognition of targets like humans and vehicles. Grayscale fused imagery yields appreciable performance levels in most conditions. With an appropriate colour mapping, colour fused imagery yields the best overall scene recognition performance. However, an inappropriate colour mapping significantly decreases observer performance compared to grayscale image fusion. The deployment of a DII system in addition to a $3-5$ μm IR system through image fusion can increase the performance of human observers when the colour mapping relates to the nature of the visual task and the conditions (scene content) at hand.
© 2003 Elsevier Science B.V. All rights reserved.

Keywords: Image fusion; Infrared; Intensified imagery; Scene recognition; Situational awareness

1. Introduction

Modern night-time cameras are designed to expand the conditions under which humans can operate. A functional piece of equipment must therefore provide an image that leads to good perceptual awareness in most environmental and operational conditions (to 'Own the weather' or 'Own the night'). The two most common night-time imaging systems display either emitted infrared (IR) radiation or reflected light, and thus provide complimentary information of the inspected scene. IR cameras have a history of decades of development. Although modern IR cameras function very well under most circumstances, they still have some inherent limitations. For instance, after a period of extensive cooling (e.g. after a long period of rain) the infrared bands provide less detailed information due to low thermal contrast in the scene, whereas the visual bands may represent the background in great detail (vegetation or soil areas, texture). In this situation it can be hard or even impossible to distinguish the background of a target in the scene, using only the infrared bands, whereas at the same time, the target itself may be highly detectable (when its temperature differs sufficiently from the mean temperature of its local background). On the other hand, a stationary target that is well camouflaged for visual detection will be hard (or even impossible) to detect in the visual bands, whereas it can still be detectable in the thermal bands. A combination of visible and thermal imagery may then allow both the detection and the unambiguous localization of the target (represented in the thermal image) with respect to its background (represented in the visual image). A human operator using a suitably combined or fused representation of IR and (intensified) visual imagery may therefore be able to construct a more complete mental representation of the perceived scene, resulting in a larger degree of situational awareness [12].

This study was performed to test the complementarity of information, obtained from different types of night vision systems (IR and image intensifiers), and the capability of several greylevel and colour image fusion schemes applied to these images to combine and convey

* Corresponding author. Tel.: +31-3463-56237; fax: +31-3463-53977.
E-mail addresses: toet@tm.tno.nl (A. Toet), e.m.franken@fel.tno.nl (E.M. Franken).
1 Tel.: +31-70-3740473; fax: +31-70-3740654.

information, about both the global structure and the fine detail of scenes, to human observers. The image modalities used were conventional single band as well as dual band (visual and near infrared) intensified low-light CCD images (II and DII, respectively) and thermal middle wavelength band (3–5 μm) infrared images. Three different image fusion schemes were investigated. Colour and grayscale fused imagery was produced through a conventional pyramid image merging scheme, in combination with two different colour mappings. This fusion approach is representative for a number of methods that have been suggested in Refs. [1,7,10,13–15], and may serve as a starting point for further developments. Observer performance with the individual image modalities serves as a baseline for the performance that should be obtained with fused imagery. The results of these tests indicates to what extent DII and IR images are complementary, and can be used to identify the characteristic features of each image modality that determine human visual performance. The goal of image fusion is to combine and preserve in a single output image all the perceptually important signal information that is present in the individual input images. Hence, for a given observation task, performance with fused imagery should at least be as good (and preferably better) as performance with the individual image modality that yields the optimal performance for that task. Knowledge of the nature of the features in each of the input images that determine observer performance can be used to develop new multimodal image visualization techniques, based on improved image fusion schemes that optimally exploit and combine the perceptually relevant information from each of the individual night-time image modalities.

2. Methods

2.1. Scene recording

A variety of stationary outdoor scenes, displaying several kinds of vegetation (grass, heather, semi shrubs, trees), sky, water, sand, vehicles, roads, and persons, were registered at night with a recently developed dual-band visual intensified (DII) camera (see below), and with a state-of-the-art thermal middle wavelength band (3–5 μm) infrared (IR) camera (Radiance HS). Both cameras had a field of view (FOV) of about 6 × 6 degrees. Some image examples are shown in Figs. 1–5.

The DII camera was developed by Thales Optronics and facilitated a two-colour registration of the scene, applying two bands covering the part of the electromagnetic spectrum ranging from visual to near infrared (400–900 nm), The crossover point between the bands of the DII camera lies approximately at 700 nm. The short (visual) wavelength part of the incoming spectrum is mapped to the R channel of an RGB colour composite image. The long (near infrared)

wavelength band corresponds primarily to the spectral reflection characteristics of vegetation, and is therefore mapped to the G channel of an RGB colour composite image. This approach utilizes the fact that the spectral reflection characteristics of plants are distinctly different from other (natural and artificial) materials in the visual and near IR range [5]. The spectral response of the long-wavelength channel ('G') roughly matches that of a Generation III image intensifier system. This channel is stored separately and used as an individual image modality (II).

Images were recorded at various times of the diurnal cycle under various atmospheric conditions (clear, rain, fog, …) and for various illumination levels (1 lux–0.1 mlux). Object ranges up to several hundreds of meters were applied. The images were digitized on-site (using a Matrox Genesis frame grabber, using at least 1.8 times over-sampling).

2.2. Stimuli

First, the recorded images were registered through an affine warping procedure, using fiducial registration points that were recorded at the beginning of each session. After warping, corresponding pixels in images taken with different cameras represent the same location in the recorded scene. Then, patches displaying different types of scenic elements were selected and cut out from corresponding images (i.e. images representing the same scene at the same instant in time, but taken with different cameras). These patches were deployed as stimuli in the pyschophysical tests. The signature of the target items (i.e. buildings, persons, vehicles etc.) in the image test sets varied from highly distinct to hardly visible.

To test the *perception of detail*, patches were selected that display either buildings, vehicles, water, roads, or humans. These patches are 280 × 280 pixels, corresponding to a FOV of 1.95 × 1.95°.

To investigate the *perception of global scene structure*, larger patches were selected, that represent either the horizon (to perform a horizon perception task), or a large amount of different terrain features (to enable the distinction between an image that is presented upright and one that is shown upside down). These patches are 575 × 475 pixels, corresponding to a FOV of 4.0 × 3.3°.

To test if the combined display of information from the individual image modalities may enhance the perception of detail (target recognition) and situational awareness, corresponding stimulus pairs (i.e. patches representing the same part of a scene at the same instant in time, but taken with different cameras) were fused.

Grayscale fused (GF) images were produced by combining the IR and II images though a pyramidal image fusion scheme [1,10,13]. A 7-level Laplacian pyramid [1] was used, in combination with a maximum absolute contrast node (i.e. pattern element) selection rule.
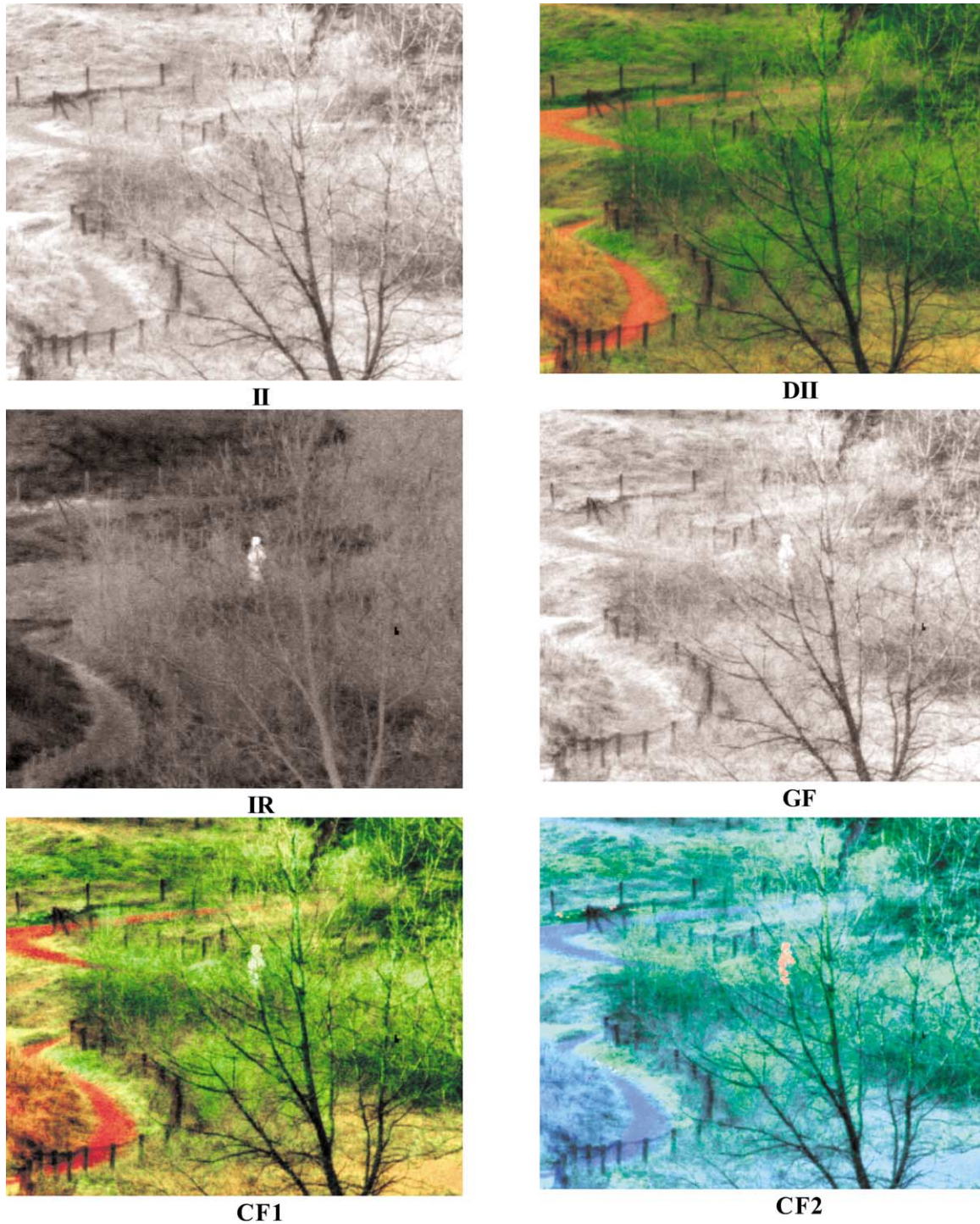
Fig. 1. The different image modalities used in this study. II and DII: the long wavelength band and both bands of the false colour intensified CCD image. IR: the thermal 3−5 μm IR image. GF: the greylevel fused image and CF1(2) and colour fused images produced with Method 1(2). This image shows a scene of a person in terrain, behind a tree.

Colour fused imagery was produced by the following two methods.

- *Colour Fusion Method 1* (CF1): The short and long wavelength bands of the DII camera were, respectively, mapped to the R and G channels of an RGB colour image. The resulting RGB colour image was then converted to the YIQ (NTSC) colour space. The luminance (Y) component was replaced by the corresponding aforementioned grayscale (II and IR) fused image, and the result was transformed back to the RGB colour space (note that the input Y from combining the R

Fig. 2. As Fig. 1, for a scene displaying a road, a house, and a vehicle.

Fig. 3. As Fig. 1, for a scene displaying a person along a riverside. Notice the reflection of the person's silhouette on the water surface in the thermal image.

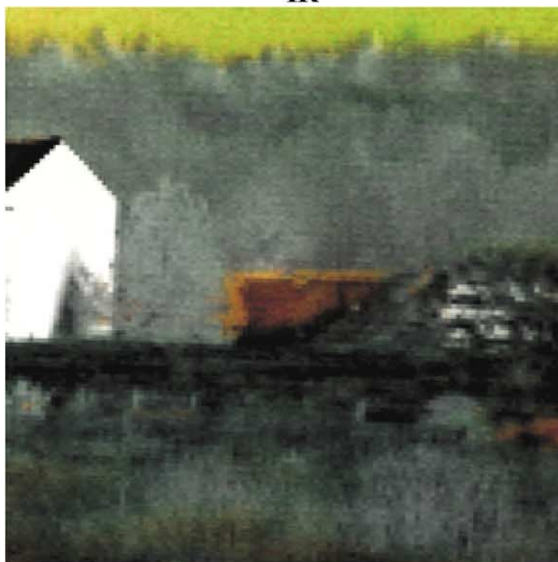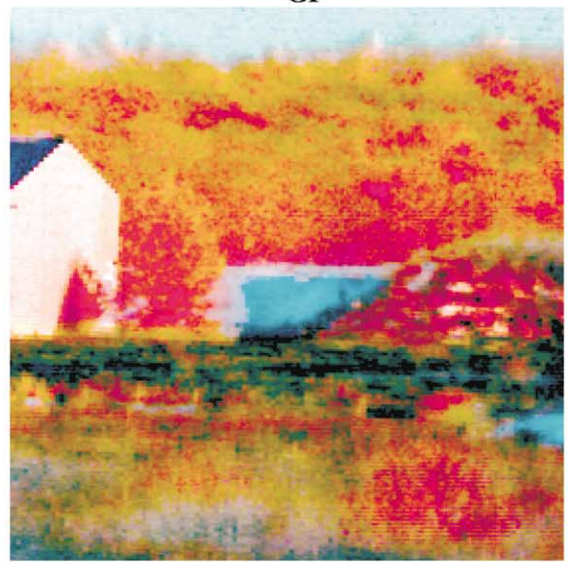Fig. 4. As Fig. 1, for a scene displaying people on a road through the woods.

Fig. 5. As Fig. 1, for a scene displaying a house and trees.

and G channel is replaced by a *Y* which is created by fusing the G channel with the IR image). This colour fusion method results in images in which grass, trees and persons are displayed as greenish, and roads, buildings, and vehicles are brownish.

- *Colour Fusion Method 2* (CF2): First, an RGB colour image was produced by assigning the IR image to the R channel, the long wavelength band of the DII image to the green channel (as in Method 1), and the short wavelength band of the DII image to the blue channel (instead of the red channel, as in Method 1). This colour fusion method results in images in which vegetation is displayed as greenish, persons are reddish, buildings are red-brownish, vehicles are whitish/bluish, and the sky and roads are most often bluish.

The multiresolution grayscale image fusion scheme employed here, selects the perceptually most salient contrast details from both of the individual input image modalities, and fluently combines these pattern elements into a resulting (fused) image. As a side effect of this method, details in the resulting fused images can be displayed at higher contrast than they appear in the images from which they originate, i.e. their contrast may be enhanced [9,11]. To distinguish the perceptual effects from contrast enhancement from those of the fusion process, observer performance was also tested with contrast enhanced versions of the individual image modalities. The contrast in these images was enhanced by a multiresolution local contrast enhancement scheme. This scheme enhances the contrast of perceptually relevant details for a range of spatial scales, in a way that is similar to the approach used in the hierarchical fusion scheme. A detailed description of this enhancement method is given elsewhere [9,11].

2.3. Apparatus used for stimuli display

A Pentium II 400 MHz computer was used to present the stimuli, measure the response times and collect the observer responses. The stimuli were presented on a 17 inch Vision Master 400 (Iiyama Electric Co., Ltd) colour monitor, using the 1152×864 true colour (32 bit) mode (corresponding to a resolution of 36.2 pixels/cm), with a colour temperature of 6500 K, and a 100 Hz refresh rate.

2.4. Tasks

The perception of the global structure of a depicted scene was tested in two different ways. In the first test, scenes were presented that had been randomly mirrored along the horizontal, and the subjects were asked to distinguish the orientation of the displayed scenes (i.e. whether a scene was displayed right side up or upside down). In this test, each scene was presented twice: once upright and once upside down. In the second test, horizon views were presented together with short markers (55×4 pixels) on the left and right side of the image and on a virtual horizontal line. In this test, each scene was presented twice: once with the markers located at the true position (height) of the horizon, and once when the markers coincided with a horizontal structure that was opportunistically available (like a band of clouds) and that may be mistaken for the horizon. The task of the subjects was to judge whether the markers indicated the true position of the horizon. The perception of the global structure of a scene is likely to determine situational awareness.

The capability to discriminate fine detail was tested by asking the subjects to judge whether or not a presented scene contained an exemplar of a particular category of objects. The following categories were investigated: buildings, vehicles, water, roads, and humans. The perception of detail is relevant for tasks involving visual search, detection and recognition.

The tests were blocked with respect to both (1) the imaging modality and (2) the task. This was done to minimize observer uncertainty, both with respect to the characteristics of the different image modalities, and with respect to the type of target.

Blocking by image modality yielded the following six classes of stimuli:

1. Grayscale images representing the thermal $3-5\ \mu m$ IR camera signal.
2. Grayscale images representing the long-wavelength band (G-channel) of the DII images.
3. Colour (R and G) images representing the two channels of the DII.
4. Grayscale images representing the IR and II signals fused by GF.
5. Colour images representing the IR and DII signals fused by CF1.
6. Colour images representing the IR and DII signals fused by CF2.

Blocking by task resulted in trial runs that tested the perception of global scene structure by asking the observers to judge whether

- the horizon was veridically indicated
- the image was presented right side up

and the recognition of detail by asking the observers to judge whether the image contained an exemplar of one of the following categories:

- building
- person
- road or path
- fluid water (e.g. a ditch, a lake, a pond, or a puddle)
- vehicle (e.g. a truck, car or van)

The entire experiment consisted of 42 different trial runs (6 different image modalities × 7 different tasks). Each task was tested on 18 different scenes. The experiment therefore involved the presentation of 756 images in total. The order in which the image modalities and the tasks were tested was randomly distributed over the observers. This was done to eliminate any possible learning effects.

2.5. Procedure

Before starting the actual experiment, the observers were shown examples of the different image modalities that were tested. They received verbal information, describing the characteristics of the particular image modality. It was explained how different types of targets are displayed in the different image modalities. This was done to familiarize the observers with the appearance of the scene content in the different images modalities, thereby minimizing their uncertainty.

Next, subjects were instructed that they were going to watch a sequence of briefly flashed images, and that they had to judge each image with respect to the task at hand. For a block of trials, testing the perception of detail, the task was to judge whether or not the image showed an exemplar of a particular category of targets (e.g. a building). For a block of trials, testing the perception of the overall structure of the scene, the task was to judge whether the scene was presented right side up, or whether the position of the horizon was indicated correctly. The subjects were instructed to respond as quickly as possible after the onset of a stimulus presentation, by pressing the appropriate one of two response keys.

Each stimulus was presented for 400 ms. This brief presentation duration, in combination with the small stimulus size, served to prevent scanning eye movements (which may differ among image modalities and target types), and to force subjects to make a decision based solely on the instantaneous percept aroused by the stimulus presentation. Immediately after the stimulus presentation interval, a random noise image was shown. This noise image remained visible for at least 500 ms. It served to erase any possible afterimages (reversed contrast images induced by, and lingering on after, the presentation of the stimulus, that may differ in quality for different image modalities and target types), thereby equating the processing time subjects can use to make their judgement. Upon each presentation, the random noise image was randomly left/right and up/down reversed. The noise images had the same dimensions as the preceding stimulus image, and consisted of randomly distributed sub-blocks of 5 × 5 pixels. For trial blocks testing the monochrome IR and II imaging modalities and grayscale fused imagery, the noise image sub-blocks were either black or mean grey. For trial blocks testing DII and colour fused imagery, the noise image sub-blocks were randomly coloured, using a color palette similar to that of the modality being tested. In all tests, subjects were asked to quickly indicate their visual judgement by pressing one of two response keys (corresponding to a YES/NO response), immediately after the onset of a stimulus image presentation. Both the accuracy and the reaction time were registered.

2.6. Subjects

A total of 12 subjects, aged between 20 and 55 years, served in the experiments reported below. All subjects have corrected to normal vision, and reported to have no colour deficiencies.

2.7. Viewing conditions

The experiments were performed in a dimly lit room. The images were projected onto the screen of the CRT display. Viewing was binocular, at a distance of 60 cm. At this distance, the images subtended a viewing angle of either 14.8×12.3 or $7.3 \times 7.3°$, corresponding to a scene magnification of 3.8.

3. Results

This section reports the results of the observer experiments for the different tasks and for each of the aforementioned image modalities. The first two tasks measure the degree to which the scene structure is correctly perceived. The remaining five tasks measure the perception of detail.

For each visual discrimination task the numbers of hits (correct detections) and false alarms (fa) were recorded to calculate $d' = Z_{\text{hits}} - Z_{\text{fa}}$, an unbiased estimate of sensitivity [4].

The effects of contrast enhancement on human visual performance is found to be similar for all tasks. Fig. 6 shows that contrast enhancement significantly improves the sensitivity of human observers performing with II and DII imagery. However, for IR imagery, the average sensitivity decreases as a result of contrast enhancement. This is probably a result of the fact that the contrast enhancement method employed in this study increases the visibility of irrelevant detail and clutter in the scene. Note that this result does not indicate that (local) contrast enhancement in general should not be applied to IR images.

Fig. 7 shows the results of all scene recognition and target detection tasks investigated here. As stated before, the ultimate goal of image fusion is to produce a combined image that displays more information than either of the original images. Fig. 7 shows that this aim is only achieved for the following perceptual tasks and conditions:

- the detection of roads, where CF1 outperforms each of the input image modalities,
- the recognition of water, where CF1 yields the highest observer sensitivity, and
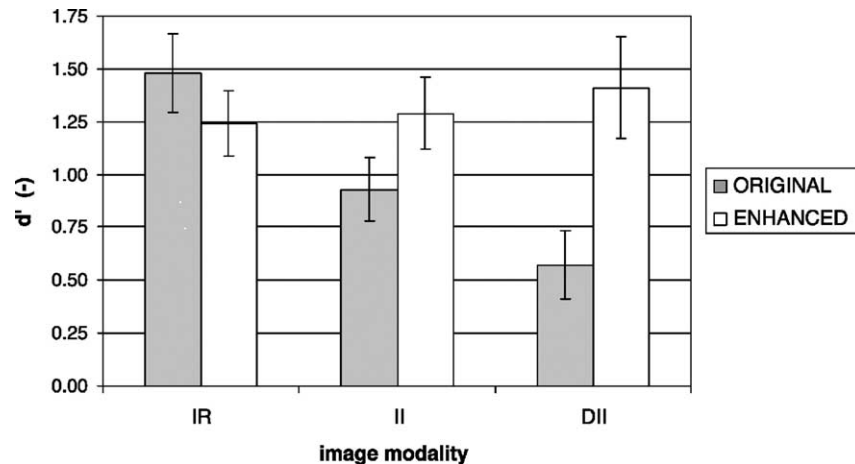
Fig. 6. The effect of contrast enhancement on observer sensitivity d'. d' represents the capability of an observer to distinguish a target in the image.

- the detection of vehicles, where three fusion methods tested perform significantly better than the original imagery.

These tasks are also the only ones in which CF1 performs better than CF2. An image fusion method that always performs at least as good as the best of the individual image modalities can be of great ergonomic value, since the observer can perform using only a single image. This result is obtained for the recognition of scene orientation from colour fused imagery produced with CF2, where performance is similar to that with II and DII imagery. For the detection of buildings and humans in a scene, all three

fusion methods perform equally well and slightly less than IR. CF1 significantly outperforms grayscale fusion for the detection of the horizon and the recognition of roads and water. CF2 outperforms grayscale fusion for both global scene recognition tasks (orientation and horizon detection). However, for CF2 observer sensitivity approaches zero for the recognition of roads and water.

Rather surprisingly, the response times (not shown here) did not differ significantly between all different image modalities. The shortest reaction times were obtained for the detection of humans (about 650 ms), and the longest response times were found for the detection of the position of the horizon (about 1000 ms).
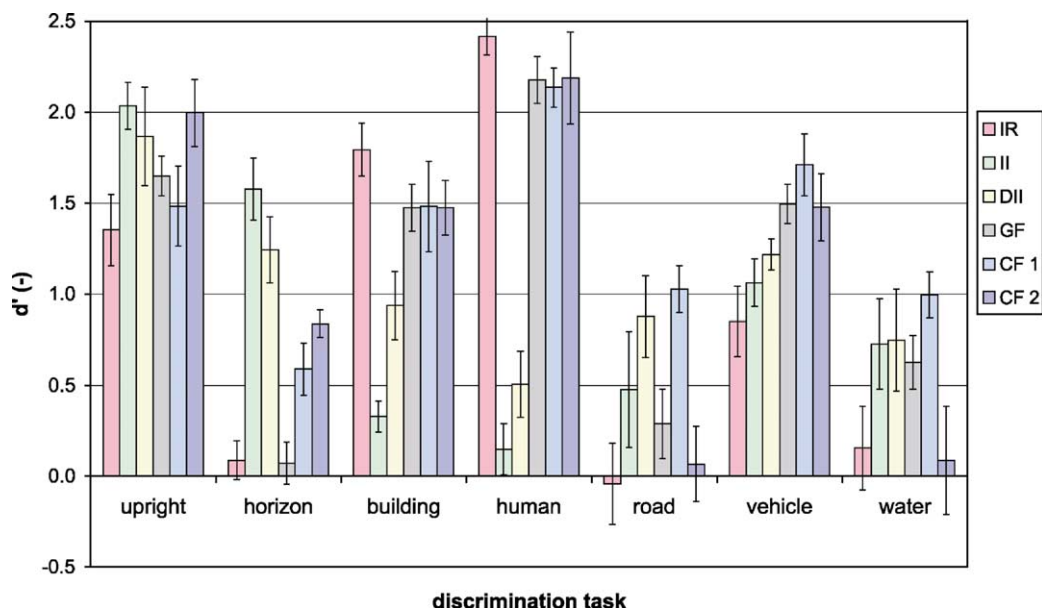

Fig. 7. Observer sensitivity d' for discrimination of global layout (orientation and horizon) and local detail (buildings, humans, roads, vehicles, and water), for six different image modalities. These modalities are (in the order in which they appear in the labeled clusters above): infrared (IR), single-band or grayscale (II) and double-band or colour (DII) intensified visual, grayscale (GF) and colour fused (CF1, CF2) imagery. d' represents the capability of an observer to distinguish a target in the image.

Section 3.1 discusses the results in detail for each of the seven different perception tasks.

3.1. Perception of global structure

The perception of the scene layout was tested by measuring the accuracy with which observers were able to distinguish a scene that was presented right side up from one that was presented upside down, and perceive the position of the horizon.

The first group of bars in Fig. 7 (labelled 'upright') represents the results for the scene orientation perception task. For the original image modalities, the best results are obtained with the intensified imagery (the II performed slightly better than the DII). The IR imagery performs significantly worse. CF2 performs just as well as II, whereas CF1 performs similar to IR. Graylevel fusion is in between both colour fusion methods. Observers remarked that they based their judgement mainly on the perceived orientation of trees and branches in the scene. CF2 displays trees with a larger colour contrast (red-brown on a light greenish or bluish background) than CF1 (dark green trees on a somewhat lighter green background), resulting in a better orientation detection performance. Also, CF2 produces bright blue skies most of the time, which makes the task more intuitive.

The perception of the true position of the horizon, represented by the second group of bars in Fig. 2, is best performed with II imagery, followed by the DII modality. Both intensified visual image modalities perform significantly better than IR or any kind of fused imagery. The low performance with the IR imagery is probably a result of the fact that a tree line and a band of clouds frequently have a similar appearance in this modality. The transposition of these 'false horizons' into the fused image modalities significantly reduces observer performance. For greylevel fused imagery, the observer sensitivity is even reduced to a near-zero level, just as found for IR. Colour fused imagery restores some of the information required to perform the task, especially CF2 that produces blue skies. However, the edges of the cloud bands are so strongly represented in the fused imagery that observer performance never attains the sensitivity level obtained for the intensified visual modalities alone (II and DII).

In both the orientation and horizon perception tasks subjects tend to confuse large bright areas (e.g. snow on the ground) with the sky.

3.2. Perception of detail

The best score for the recognition of *buildings* is found for IR imagery. In this task, IR performs significantly better than II or DII. DII imagery performs significantly better than II, probably because of the colour contrast between the buildings and the surrounding vegetation (red-brown walls on a green background, compared to grey walls on a grey background in case of the II imagery). The performance with fused imagery is slightly less than with IR, and independent of the fusion method.

The detection of *humans* is best performed with IR imagery, in which they are represented as white hot objects on a dark background. II imagery yields a very low sensitivity for this task; i.e. humans are hardly ever noticed in intensified visual imagery. The sensitivity for the detection of humans in DII imagery is somewhat higher, but remains far below that found for IR. In this case, there is almost no additional information in the second wavelength band of the DII modality, and therefore almost no additional colour contrast. As a result, most types of clothing are displayed as greenish, and are therefore hard to distinguish from vegetation. Performance with fused imagery is only slightly below that with IR. There is no significant difference between the different grayscale and colour fusion types.

Roads cannot reliably be recognized from IR imagery (d' becomes even negative, meaning that more false alarms than correct detections are scored). DII performs best of the individual image modalities, and significantly higher than II because of the additional colour contrast (DII displays roads as red-brown, on a green background). Grayscale fused imagery results in a performance that is significantly below that found for DII, and somewhat lower than that obtained for II imagery. This is probably a result of (1) the introduction of irrelevant luminance details from the IR imagery, and (2) the loss of colour contrast as seen in the DII imagery. CF1 produces colour fused imagery that yields a higher sensitivity than each of the original image modalities, although observer performance is not significantly better than with DII imagery. The additional improvement obtained with this combination scheme is probably caused by the contrast enhancement inherent in the fusion process. The sensitivity obtained for imagery produced by CF2 is near zero. This is probably a result of the fact that this method displays roads with a light blue colour. These can therefore easily be mistaken for water or snow. This result demonstrates that the inappropriate use of colour in image fusion severely degrades observer performance.

Image fusion clearly helps to recognize *vehicles* in a scene. They are best discriminated in colour fused images produced with CF1, that displays vehicles in brown-yellow on a green background. CF2 (that shows vehicles as blue on a brown and green background) and grayscale fusion both result in an equal and somewhat lower observer sensitivity. Fused imagery of all types performs significantly better than each of the original image modalities. The lowest recognition performance is obtained with IR imagery.

Water is best recognized in colour fused imagery produced with CF1. This method displays water sometimes as brown-reddish, and sometimes as greyish. The II, DII and greylevel fusion scheme all yield a similar and slightly lower performance. CF2 results on a near zero observer sensitivity for this task. This method displays water

sometimes as purple-reddish, thus giving it a very unnatural appearance, and sometimes as bluish, which may cause confusion with roads, that have the same colour. These results again demonstrate that it is preferable not to use any colour at all (grayscale), than to use an inappropriate colour mapping scheme.

3.3. Summary

Table 1 summarizes the main findings of this study. IR has the lowest overall performance of all modalities tested. This results from a low performance for both large scale orientation tasks, and for the detection and recognition of roads, water, and vehicles. In contrast, intensified visual imagery performs best in both orientation tasks. The perception of the horizon is significantly better with II and DII imagery. IR imagery performs best for the perception and recognition of buildings and humans—DII has the best overall performance of the individual image modalities. Thus, IR on one hand and (D)II images on the other hand contain *complementary* information, which makes each of these image modalities suited for performing different perception tasks.

CF1 has the best overall performance of the image fusion schemes tested here. The application of an appropriate colour mapping scheme in the image fusion process can indeed significantly improve observer performance compared to grayscale fusion. In contrast, the use of an inappropriate colour scheme can severely degrade observer sensitivity. Although the performance of CF1 for specific observation tasks is below that of the optimal individual sensor, for a combination of observation tasks (as will often be the case in operational scenarios) the CF1 fused images can be of great ergonomic value, since the observer can perform using only a single image.

Table 1
The relative performance of the different image modalities for the seven perceptual recognition tasks. Rank orders—1,1, and 2 indicate, respectively, the worst, second best, and best performing image modality for a given task. The tasks involve the perception of the global layout (orientation and horizon) of a scene, and the recognition of local detail (buildings, humans, roads, vehicles, and water). The different image modalities are: infrared (IR), grayscale (II) and dual band false-colour (DII) intensified visual, grayscale fused images (GF) and two different colour fusion (CF1, CF2) schemes. The sum of the rank orders (although it has no strict meaning in a statistical sense) summarizes the overall performance of the modalities

	IR	II	DII	GF	CF1	CF2
Upright	−1	2	1			2
Horizon	−1	2	1			
Building	2	−1		1	1	1
Human	2	−1		1	1	1
Road	−1		1		2	
Vehicle	−1			2	2	1
Water	−1				2	
Overall	−1	2	3	4	8	5

4. Conclusions

Night-time images recorded using an image intensified low-light CCD camera and a thermal middle wavelength band $(3–5\ \mu m)$ infrared camera contain *complementary* information. This makes each of the individual image modalities only suited for specific observation task. However, the complementarity of the information of the image modalities can be exploited using image fusion, which would enable multiple observation tasks using a single night-time image representation.

Since there evidently exists no one-to-one mapping between the temperature contrast and the spectral reflectance of a material, the goal of producing a night-time image, incorporating information from IR imagery, with an appearance similar to a colour daytime image can never be fully achieved. The options are therefore (1) to settle for a single mapping that works satisfactory in a large number of conditions, or (2) to adapt (optimize) the colour mapping to the situation at hand. However, the last option is not very attractive since a different colour mapping for each task and situation tends to confuse observers [3,8].

Multimodal image fusion schemes based on local contrast decomposition do not distinguish between material edges and temperature edges. For many tasks, material edges are the most important ones. Fused images frequently contain an abundance of contours that are irrelevant for the task that is to be performed. Fusion schemes incorporating some kind of contrast stretching enhance the visibility of all details in the scene, irrespective of their visual significance. The introduction of spurious or irrelevant contrast elements in a fused image may clutter the scene, distract the observer and lead to misinterpretation of perceived detail. As a result, observer performance may degrade significantly. A useful image fusion scheme should therefore take into account the visual information content (meaning) of the edges in each of the individual image modalities, and combine them accordingly in the resulting image.

For most perceptual tasks investigated here (except for horizon and road detection), grayscale image fusion yields appreciable performance levels. When an appropriate colour mapping scheme is applied, the addition of colour to grayscale fused imagery can significantly increase observer sensitivity for a given condition and a certain task (e.g. CF2 for orientation detection, both colour fusion methods for horizon detection, CF1 for road and water detection). However, inappropriate use of colour can significantly decrease observer performance compared to straightforward grayscale image fusion (e.g. CF2 for the detection of roads and water).

For the observation tasks and image examples tested here, optimal overall performance was obtained for images fused using CF1. The overall performance was higher than for either of the individual image modalities. Note that in this fusion method, no colour mapping is applied to the IR

information. Instead, the IR information is blended into the image without changing the colour.

The present findings agree with those from previous studies [2,3,6,8]. The present results will be analysed further

1. to distinguish perceptually relevant features from noise and distracting elements, and
2. to find out if there are features that are consistently mistaken by subjects for another type of scenic detail.
3. to further optimize image fusion techniques (with or without using colour-mapping).

Acknowledgements

References

[1] P.J. Burt, E.H. Adelson, Merging images through pattern decomposition, in: A.G. Tescher (Ed.), Applications of Digital Image Processing VIII, The International Society for Optical Engineering, Bellingham, WA, 1985, pp. 173–181.

[2] E.A. Essock, M.J. Sinai, J.S. McCarley, W.K. Krebs, J.K. DeFord, Perceptual ability with real-world nighttime scenes: image-intensified, infrared, and fused-color imagery, Human Factors 41 (3) (1999) 438–452.

[3] W.K. Krebs, D.A. Scribner, G.M. Miller, J.S. Ogawa, J. Schuler, Beyond third generation: a sensor-fusion targeting FLIR pod for the F/A-18, in: B.V. Dasarathy (Ed.), Sensor Fusion: Architectures, Algorithms, and Applications II, International Society for Optical Engineering, Bellingham, WA, 1998, pp. 129–140.

[4] N.A. Macmillan, C.D. Creelman, Detection Theory: A User's Guide, Cambridge University Press, Cambridge, MA, 1991.

[5] C.M. Onyango, J.A. Marchant, Physics-based colour image segmentation for scenes containing vegetation and soil, Image and Vision Computing 19 (8) (2001) 523–538.

[6] D. Ryan, R. Tinkler, Night pilotage assessment of image fusion, in: R.J. Lewandowski, W. Stephens, L.A. Haworth (Eds.), SPIE Proceedings on Helmet and Head Mounted Displays and Symbology Design Requirements II, The International Society for Optical Engineering, Bellingham, WA, 1995, pp. 50–67.

[7] J. Schuler, J.G. Howard, P. Warren, D.A. Scribner, R. Klien, M. Satyshur, M.R. Kruer, Multiband E/O color fusion with consideration of noise and registration, in: W.R. Watkins, D. Clement, W.R. Reynolds (Eds.), Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process, The International Society for Optical Engineering, Bellingham, WA, 2000, pp. 32–40.

[8] P.M. Steele, P. Perconti, Part task investigation of multispectral image fusion using gray scale and synthetic color night vision sensor imagery for helicopter pilotage, in: W. Watkins, D. Clement (Eds.), Proceedings of the SPIE Conference on Targets and Backgrounds, Characterization and Representation III, The International Society for Optical Engineering, Bellingham, WA, 1997, pp. 88–100.

[9] A. Toet, Adaptive multi-scale contrast enhancement through nonlinear pyramid recombination, Pattern Recognition Letters 11 (1990) 735–742.

[10] A. Toet, Hierarchical image fusion, Machine Vision and Applications 3 (1990) 1–11.

[11] A. Toet, Multi-scale contrast enhancement with applications to image fusion, Optical Engineering 31 (5) (1992) 1026–1031.

[12] A. Toet, J.K. IJspeert, A.M. Waxman, M. Aguilar, Fusion of visible and thermal imagery improves situational awareness, Displays 18 (1998) 85–95.

[13] A. Toet, J.J. Ruyven, J.M. Valeton, Merging thermal and visual images by a contrast pyramid, Optical Engineering 28 (1989) 789–792.

[14] A. Toet, J. Walraven, New false colour mapping for image fusion, Optical Engineering 35 (3) (1996) 650–658.

[15] A.M. Waxman, A.N. Gove, D.A. Fay, J.P. Racamato, J.E. Carrick, M.C. Seibert, E.D. Savoye, Color night vision: opponent processing in the fusion of visible and IR imagery, Neural Networks 9 (6) (1996).