

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Cognitive Image Fusion and Assessment

Alexander Toet  
*TNO Human Factors*  
*The Netherlands*

## 1. Introduction

The increasing availability and deployment of imaging sensors operating in multiple spectral bands has led to a requirement for methods that combine the signals from these sensors in an effective and ergonomic way for presentation to the human operator. Effective combinations of complementary and partially redundant multispectral imagery can provide information that is not directly evident from the individual input images.

Image fusion for human inspection should combine information from two or more images of a scene into a single composite image that is more informative than each of the input images alone, and that requires minimal cognitive effort to understand. The fusion process should therefore maximize the amount of relevant information in the fused image, while minimizing the amount of irrelevant details, uncertainty and redundancy in the output. Thus, image fusion should preserve task relevant information from the source images, prevent the occurrence of artifacts or inconsistencies in the fused image, and suppress irrelevant features (e.g. noise) from the source images (Smith & Heather, 2005). The representation of fused imagery should optimally agree with human cognition, so that humans can quickly grasp the gist and meaning of the displayed scenes. For instance, the representation of spatial details should effortlessly elicit the recognition of known Gestalts, and the color schemes used should be natural (ecologically correct) and thus agree with human intuition. Irrelevant details (clutter) should be suppressed to minimize cognitive workload and to maximize recognition speed.

Some potential benefits of image fusion are: wider spatial and temporal coverage, decreased uncertainty, improved reliability, and increased robustness of the system. Image fusion has applications in defense for situation awareness (Toet et al., 1997b), surveillance (Riley & Smith, 2006), target tracking (Zou & Bhanu, 2005), intelligence gathering (O'Brien & Irvine, 2004), and person authentication (Kong et al., 2007). Other important applications are found in industry and medicine (for a recent survey of different applications of image fusion techniques see Blum & Liu, 2006).

The way images are combined depends on the specific application and on the type of information that is relevant in the given context (Smith & Heather, 2005). By examining the effects of several image fusion methods on different cognitive tasks, Krebs et al. (Krebs & Ahumada, 2002) showed that the benefits of sensor fusion are task dependent. However, until now the human end user has not been involved in the design process and the development of image fusion algorithms to any great extent. Mostly, image fusion algorithms are developed in isolation, and the human end-user is little more than an

afterthought, so that separate follow-up evaluation studies are usually required to assess to what extent humans benefit from these methods (Aguilar et al., 1999; Dixon et al., 2005; Dixon et al., 2006a; Dixon et al., 2006b; Essock et al., 1999; Essock et al., 2005; Krebs & Sinai, 2002; Smith et al., 2002; Toet & Franken, 2003; Waxman et al., 2006). Recently has it been realized that the only way to guarantee the ultimate effectiveness of image fusion methods for human observers is to include human evaluation as an integral part of the design process (Muller & Narayanan, 2009).

In this chapter we present some image fusion techniques and assessment methods that are based on the principles of cognitive engineering. Cognitive image fusion is based on concepts derived from neural models of visual perception and pattern recognition. Here we focus on the intuitive representation of spatial structures (outlines) and image color. We will argue that cognitive fusion leads to fused image representations that are optimally tuned to the human information processing capabilities.

### **1.1 The representation of spatial detail in fused imagery**

Human visual image recognition performance depends on the amount of informative spatial features (like edges, corners, and lines) that are available in the image (Ullman, 2007). Hence, for optimal interpretation a fusion scheme should maximize the number of meaningful details in the resulting fused image. However, there is still a large semantic gap between computer image representations and human image understanding (Vogel & Schiele, 2007). This is a significant obstacle for the development of effective image fusion schemes. For instance, image segmentation and decomposition schemes still lead to undesirable over- and under- segmentation of semantically contiguous boundaries. Edge representations of images still yield incomplete object boundaries or numerous spurious (noise related) edges. As a result most image representation schemes do not correspond to human perception. It has been suggested to use cognitive principles to bridge the gap between human and computer image understanding (Jakobson et al., 2004). Some first attempts to apply concepts derived from neural models of visual processing and pattern recognition to image fusion and interpretation have been quite successful (Chiarella et al., 2004; Fay et al., 2004; Waxman et al., 2003).

### **1.2 Color representation of fused imagery**

Fused imagery has traditionally been represented in graytones. However, the increasing availability of fused and multi-band vision systems has led to a growing interest in color representations of fused imagery (Li & Wang, 2007; Shi et al., 2005a; Shi et al., 2005b; Tsagiris & Anastassopoulos, 2005; Zheng et al., 2005). In principle, color imagery has several benefits over monochrome imagery for human inspection. While the human eye can only distinguish about 100 shades of gray at any instant, it can discriminate several thousands of colors. As a result, color may improve feature contrast, thus enabling better scene segmentation and object detection (Walls, 2006). Color imagery may yield a more complete mental representation of the perceived scene, resulting in better situational awareness. Scene understanding and recognition, reaction time, and object identification are indeed faster and more accurate with color imagery than with monochrome imagery (Cavanillas, 1999; Gegenfurtner & Rieger, 2000; Goffaux et al., 2005; Oliva & Schyns, 2000; Rousselet et al., 2005; Sampson, 1996; Spence et al., 2006; Wichmann et al., 2002). Also, observers are able to selectively attend to task-relevant color targets and to ignore non-targets with a task-

irrelevant color (Ansorge et al., 2005; Folk & Remington, 1998; Green & Anderson, 1956). As a result, simply producing a fused image by mapping multiple spectral bands into a three dimensional color space already generates an immediate benefit, and provides a method to increase the dynamic range of a sensor system (Driggers et al., 2001).

However, the color mapping should be chosen with care and should be adapted to the task at hand. Although general design rules can be used to assure that the information available in the sensor image is optimally conveyed to the observer (Jacobson & Gupta, 2005), it is not trivial to derive a mapping from the various sensor bands to the three independent color channels, especially when the number of sensor bands exceeds three (e.g. with hyperspectral imagers; Jacobson et al., 2007). In practice, many tasks may benefit from a representation that renders fused imagery in natural colors. Natural colors facilitate object recognition by allowing access to stored color knowledge (Joseph & Proffitt, 1996). Experimental evidence indicates that object recognition depends on stored knowledge of the object's chromatic characteristics (Joseph & Proffitt, 1996). In natural scene recognition paradigms, optimal reaction times and accuracy are obtained for normal natural (or diagnostically) colored images, followed by their grayscale version, and lastly by their (nondiagnostically) false colored version (Goffaux et al., 2005; Oliva, 2005; Oliva & Schyns, 2000; Rousselet et al., 2005; Wichmann et al., 2002).

When sensors operate outside the visible waveband, artificial color mappings inherently yield false color images whose chromatic characteristics do not correspond in any intuitive or obvious way to those of a scene viewed under natural photopic illumination (e.g. Fredembach & Süssstrunk, 2008). As a result, this type of false color imagery may disrupt the recognition process by denying access to stored knowledge. In that case, observers need to rely on color contrast to segment a scene and recognize the objects therein. This may lead to a performance that is even worse compared to single band imagery alone (Sinai et al., 1999a). Experiments have indeed convincingly demonstrated that a false color rendering of fused night-time imagery which resembles natural color imagery significantly improves observer performance and reaction times in tasks that involve scene segmentation and classification (Essock et al., 1999; Sinai et al., 1999b; Toet et al., 1997a; Toet & IJspeert, 2001; Vargo, 1999; White, 1998), whereas color mappings that produce counterintuitive (unnaturally looking) results are detrimental to human performance (Krebs et al., 1998; Toet & IJspeert, 2001; Vargo, 1999). One of the reasons often cited for inconsistent color mapping is a lack of physical color constancy (Vargo, 1999). Thus, the challenge is to give nightvision imagery not merely an intuitively meaningful ("naturalistic") color appearance, but also one that is stable for camera motion and changes in scene composition and lighting conditions. A natural and stable color representation serves to improve the viewer's scene comprehension and enhance object recognition and discrimination (Scribner et al., 1999). Several techniques have been proposed to render night-time imagery in color (e.g. Sun et al., 2005; Toet, 2003; Tsagiris & Anastassopoulos, 2005; Wang et al., 2002; Zheng et al., 2005). Simply mapping the signals from different nighttime sensors (sensitive in different spectral wavebands) to the individual channels of a standard color display or to the individual components of perceptually decorrelated color spaces, sometimes preceded by principal component transforms or followed by a linear transformation of the color pixels to enhance color contrast, usually results in imagery with an unnatural color appearance (e.g. Howard et al., 2000; Krebs et al., 1998; Li et al., 2004; Schuler et al., 2000; Scribner et al., 2003). More intuitive color schemes may be obtained through opponent processing through feedforward center-surround shunting neural networks similar to those found in vertebrate

color vision (Aguilar et al., 1998; Aguilar et al., 1999; Fay et al., 2000a; Fay et al., 2000b; Huang et al., 2007; Warren et al., 1999; Waxman et al., 1995; Waxman et al., 1997; Waxman et al., 1998). Although this approach produces fused nighttime images with appreciable color contrast, the resulting color schemes remain rather arbitrary and are usually not strictly related to the actual daytime color scheme of the scene that is registered. We recently developed a color transform that can give fused multisensor imagery a natural color appearance (Hogervorst & Toet, 2008a; Hogervorst & Toet, 2008b; Hogervorst & Toet, 2010). The method derives an optimal color mapping by optimizing the match between a set of corresponding samples taken from a daytime color reference image and a multi-band nighttime image. Once the mapping has been determined, it can be implemented as a color lookup-table transform. As a result, the color transform is extremely simple and fast, and can easily be applied in real-time with standard hardware. Moreover, it yields fused images with a natural color appearance and provides object color constancy, since the relation between sensor output and colors is fixed. Since the mapping is sample-based, it is highly specific for different types of materials in the scene and can therefore easily be adapted for the task at hand, such as optimizing the visibility of camouflaged objects.

### 1.3 The need for image fusion quality metrics

Because the number of image fusion techniques and systems available is steadily increasing, there is a growing need for metrics to evaluate and compare the quality of fused imagery. Clearly, the ultimate image fusion scheme should use semantically meaningful image representations, and should use fusion rules that give higher priority (weights) to regions with semantically higher importance to the operator. Generally the ideal fused image (reference) is not available. In applications where the fused images are intended for human observation, the performance of fusion algorithms can be measured in terms of improvement in user performance in tasks like detection, recognition, tracking, or classification. This approach requires a well defined task that allow quantification of human performance (e.g. Toet et al., 1997b; Toet & Franken, 2003). However, this usually means time consuming and often expensive experiments involving a large number of human subjects. In recent years, a number of computational image fusion quality assessment metrics have therefore been proposed (e.g. Angell, 2005; Blum, 2006; Chari et al., 2005; Chen & Varshney, 2005; Chen & Varshney, 2007; Corsini et al., 2006; Cvejic et al., 2005a; Cvejic et al., 2005b; Piella & Heijmans, 2003; Toet & Hogervorst, 2003; Tsagiris & Anastassopoulos, 2004; Ulug & Claire, 2000; Wang & Shen, 2006; Xydeas & Petrovic, 2000; Yang et al., 2007; Zheng et al., 2007; Zhu & Jia, 2005). Although some of these metrics agree with human visual perception to some extent, most of them cannot predict observer performance for different input imagery and scenarios. Metrics that accurately describe human performance are of great value, since they can be used to optimize image fusion systems and to predict human observer performance for different scenarios. However, since reliable human performance related metrics are extremely difficult to design, they are not yet available at present.

### 1.4 Overview of this chapter

In the rest of this chapter we investigate how different grayscale and color image fusion methods affect the perception of scene layout, object recognition, and the detection of camouflaged objects. We assessed the different fusion techniques by quantifying the performance of human observers using the fused imagery.

## 2. Scene layout recognition

In this section we investigate the effects of grayscale and color image fusion on a spatial localization task. We assess the different fusion techniques by quantifying the (objective) localization accuracy and the (subjective) confidence of human observers performing the task using the fused imagery.

### 2.1 Imagery

We recorded spatially registered visible light and mid-wave (3-5  $\mu\text{m}$ ) thermal motion sequences representing three military surveillance scenarios (for details see Toet et al., 1997b). The individual images used in this study correspond to successive frames from these time sequences. Corresponding visual and thermal frames were fused using an opponent color fusion technique developed by the MIT Lincoln Laboratory (Waxman et al., 1995; Waxman et al., 1996a; Waxman et al., 1996b; Waxman et al., 1996c; Waxman et al., 1997; Waxman et al., 1999). Grayscale fused images were obtained by taking the luminance component of the corresponding color fused images. The MIT color fusion method provides images with a semi natural color appearance, and enhances image contrast by filtering the input images with a feedforward center-surround shunting neural network (Grossberg, 1988).

In all three scenarios, the thermal images provide a poor representation of the scene layout, whereas they clearly show the presence of a person in the scene (Fig. 1). In contrast, the visible images clearly show the scene structure, whereas they poorly represent the person. In the fused images, both the background details and the person are clearly visible. Situational awareness is tested by asking observers to report the perceived position of the person relative to characteristic details in the scene. Because the relevant information is distributed over the individual image modalities (the images are complementary), this task cannot be performed with any of the individual image modalities. We used schematic (cartoon-like) representations of the actual scenes to obtain a baseline performance and to register the observer responses. Fig. 1 shows an example of a scenario in which the reference features are the poles that support the fence. These poles are clearly visible in the CCD images but not represented in the IR images because they have nearly the same temperature as the surrounding terrain. In the (graylevel and color) fused images the poles are again clearly visible.

### 2.2 Experiment

Each image was briefly (2s) shown on a CRT display, followed by the presentation of a corresponding schematic reference image. The subject's task was to indicate the perceived location of the person in the scene by placing a mouse controlled cursor at the corresponding location in this reference image. When the left mouse button was pressed the computer registered the coordinates corresponding to the indicated image location (the mouse coordinates) and computed the distance in the image plane between the actual position of the person and the indicated location. The subject pressed the right mouse button if the person in the displayed scene could not be detected. The subject could only perform the localization task by memorizing the perceived position of the person relative to the reference features in the scene.

The schematic reference images were also used to determine the optimal (baseline) localization accuracy of the observers. Baseline test images (Fig. 1) were created by placing a binary (dark) image of a walking person at different locations in the reference scene. 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 total of 6 subjects, aged between 20 and 30 years, served in the experiments reported below (for details see Toet et al., 1997b).

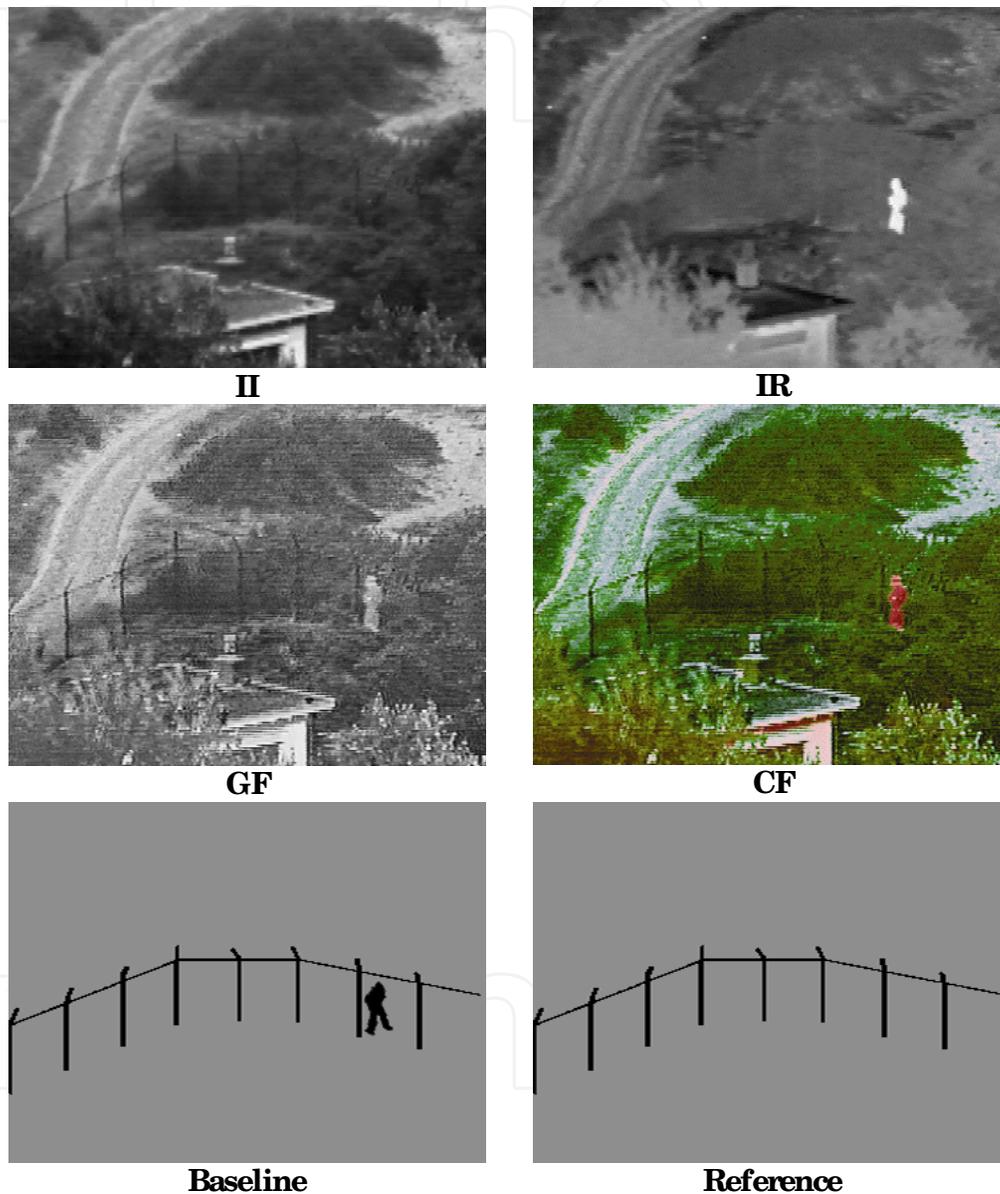


Fig. 1. Original intensified visual image (II), original thermal image (IR), graylevel fused (GF) image, color fused (CF) image, baseline test image (Baseline), and reference (Reference) image.

### 2.3 Results and discussion

Fig. 2 shows that subjects are uncertain about the location of the person in the scene for about 20% 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. 3 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 graylevel 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 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 discrepancy between the perceived position and the actual position of the person.

Fig. 3 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. 2. 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.

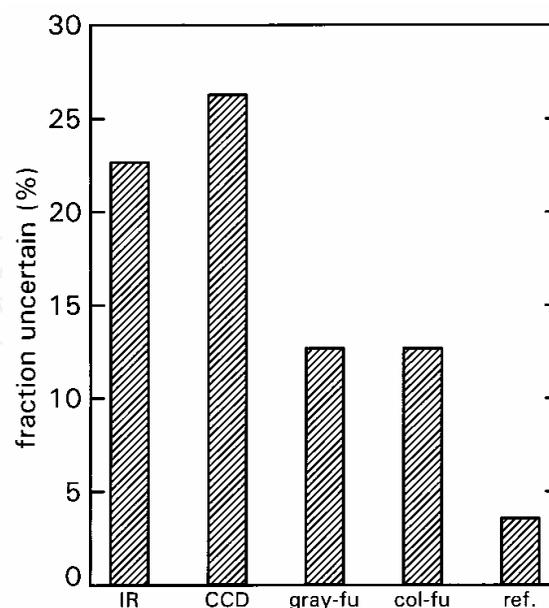


Fig. 2. Percentage of image presentations in which observers are uncertain about the relative position of the person in the scene, for each of the 5 image modalities tested (IR, intensified CCD, graylevel fused, color fused, and schematic reference images).

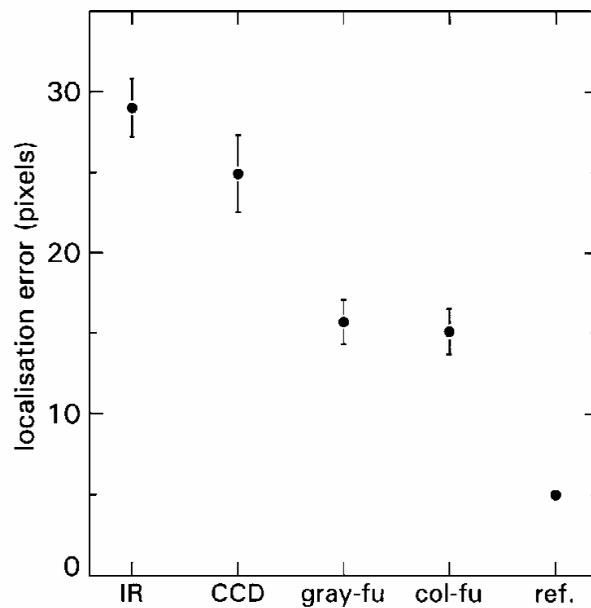


Fig. 3. The mean weighted distance between the actual position of the person in the scene and the perceived position for each of the 5 image modalities tested (IR, intensified CCD, graylevel fused, color fused, and schematic reference images). The error bars indicate the size of the standard error in the perceived location.

Summarizing, for the scenarios investigated here, we conclude that fused images provide a better representation of the layout of the scene, but color does not help to localize the targets.

### 3. Scene gist recognition

In this section we investigate the effects of grayscale and color image fusion on the perception of detail and the global structure of scenes. We assess the different fusion techniques by quantifying the sensitivity of human observers performing the task using the fused imagery.

#### 3.1 Imagery

A variety of 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 dual-band visual intensified (DII) camera (see below), and with a middle wavelength band (3-5  $\mu\text{m}$ ) infrared (IR) camera (Radiance HS). An example is shown in Fig. 4.

The DII camera provided a two-color 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 color 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. 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 infrared range (Onyango & Marchant, 2001). The spectral response of the

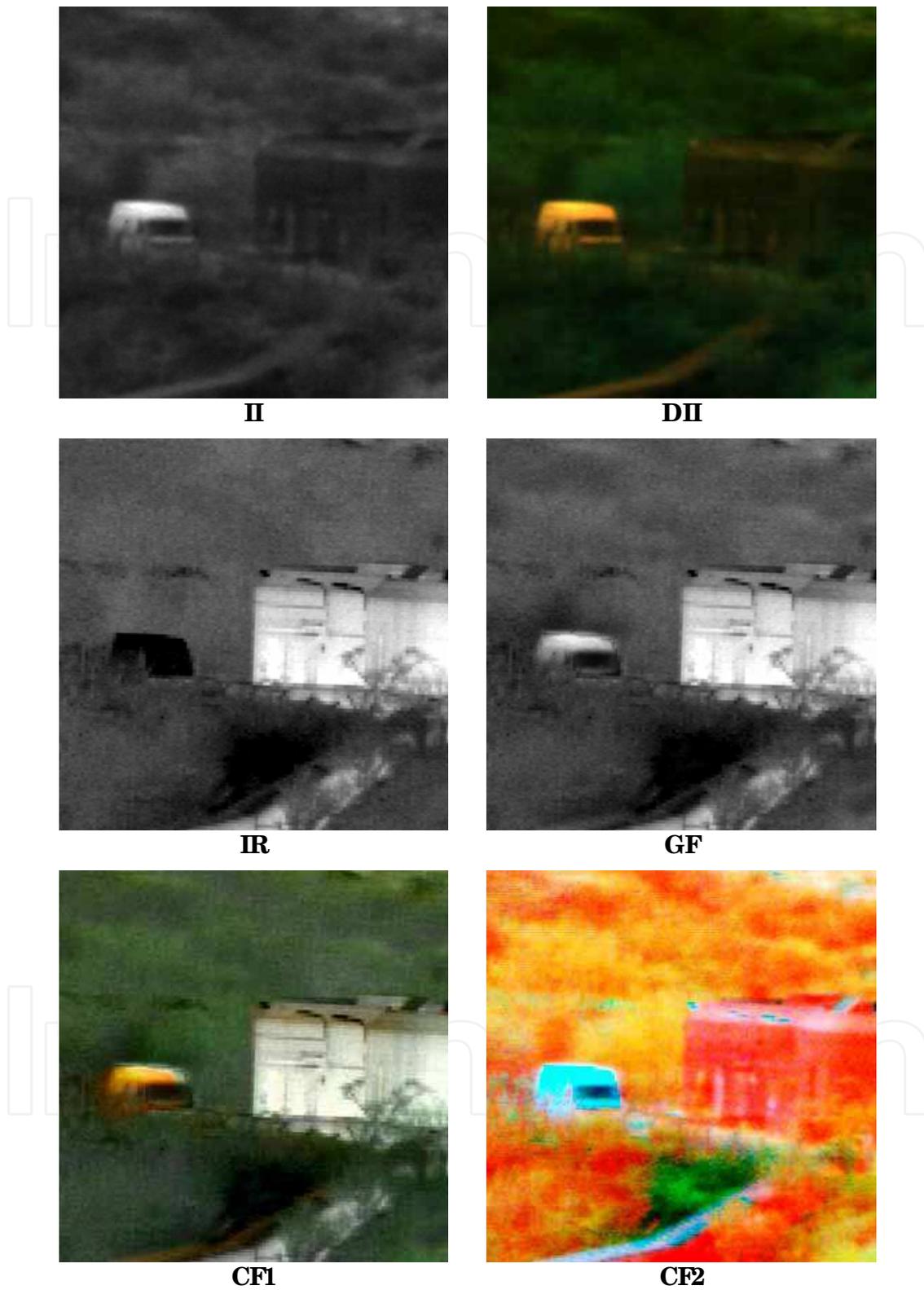


Fig. 4. Example of the different image modalities used in this study. II and DII: the long wavelength band and both bands of the false color intensified CCD image. IR: the thermal 3-5  $\mu\text{m}$  IR image. GF: the greylevel fused image and CF1(2) and color fused images produced with Method 1(2). This image shows a scene with a road, a house, and a vehicle.

long-wavelength channel ('G') roughly matches that of a Generation III image intensifier (II) system, and is stored separately.

The images were registered, and patches displaying different types of scenic elements were selected and cut out from corresponding images in the different spectral bands. These patches were deployed as stimuli in the psychophysical 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*, small patches were selected that display either buildings, vehicles, water, roads, or humans. 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).

Grayscale fused (GF) images were produced by combining the IR and II images through a pyramidal image fusion scheme (Burt & Adelson, 1985; Toet et al., 1989; Toet, 1990b). Color fused imagery was produced by the following two methods.

- *Color 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 color image. The resulting RGB color image was then converted to the YIQ (NTSC) color 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 color space (note that the input Y from combining the R and G channel is replaced by a Y which is created by fusing the G channel with the IR image). This color fusion method results in images in which grass, trees and persons are displayed as greenish, and roads, buildings, and vehicles are brownish.
- *Color Fusion Method 2* (CF2): First, an RGB color 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 color 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 (Toet, 1990a; Toet, 1992). 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, using 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 (for details see Toet, 1990a; Toet, 1992).

### 3.2 Experiment

A computer was used to briefly (400ms) present the images on a CRT display, measure the response times and collect the observer responses. 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 no known color deficiencies.

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 two horizontally aligned short markers on the left and right side of the image. 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 could 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 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.

### 3.3 Results and discussion

For each visual discrimination task the numbers of hits (correct detections) and false alarms ( $f_a$ ) were recorded to calculate  $d' = Z_{\text{hits}} - Z_{f_a}$ , an unbiased estimate of sensitivity (Macmillan & Creelman, 1991).

The effects of contrast enhancement on human visual performance is similar for all tasks. Fig. 5 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. 6 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. 6 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
- 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 color 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.

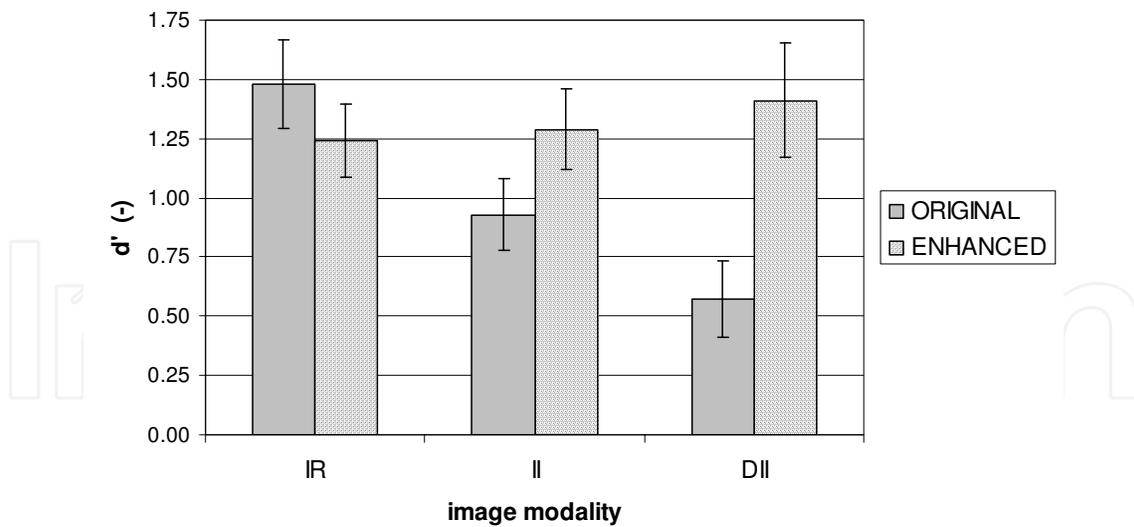


Fig. 5. The effect of contrast enhancement on observer sensitivity  $d'$ .

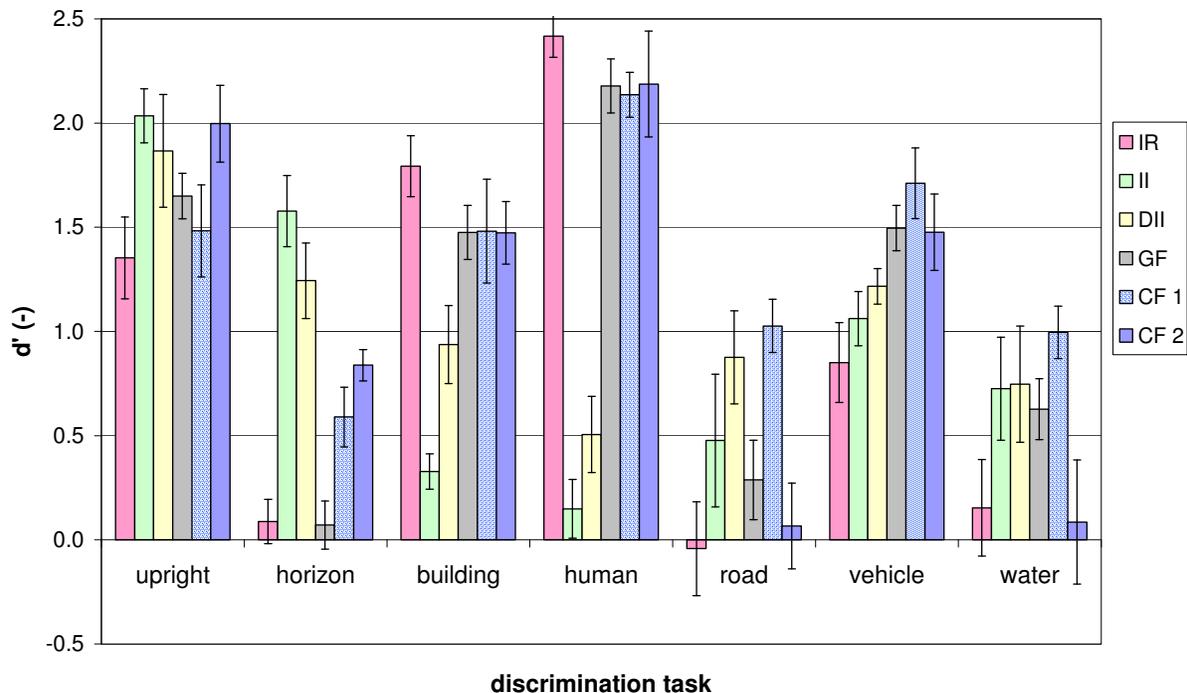


Fig. 6. Observer sensitivity  $d'$  for discrimination of global layout (orientation and horizon) and local detail (buildings, humans, roads, vehicles, and water), for six different image modalities. These modalities are (in the order in which they appear in the labeled clusters above): infrared (IR), single-band or grayscale (II) and double-band or color (DII) intensified visual, grayscale (GF) and color fused (CF1, CF2) imagery.

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.

|          | 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   |

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), greyscale (II) and dual band false-color (DII) intensified visual, grayscale fused images (GF) and two different color fusion (CF1, CF2) schemes. The sum of the rank orders indicates the overall performance of the modalities.

CF1 has the best overall performance of the image fusion schemes tested here. The application of an appropriate color mapping scheme in the image fusion process can indeed significantly improve observer performance compared to grayscale fusion. In contrast, the use of an inappropriate color 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.

#### 4. Object recognition

In this section we will show how manual segmentations of a set of corresponding input and fused images can be used to evaluate the perceptual quality of image fusion schemes. Human visual perception is mostly concerned with object detection and boundary discrimination. The method is therefore based on the hypothesis that fused imagery should provide an optimal representation of the object boundaries that can be determined from the individual input image modalities. To compute the quality of the different image fusion schemes we formulate boundary-detection as a classification problem of discriminating non-boundary from boundary pixels, and apply the precision-recall framework, using reference contour images derived from the human-marked boundaries as a reference standard.

##### 4.1 Imagery

Seven sets of IR and visible images, including noisy, clean, cluttered and uncluttered images, were used in this study (Fig. 7). These multi-sensor images are part of the Multi-

Sensor Image Segmentation Data Set (Lewis et al., 2006), and are publicly available through the ImageFusion.org website (ImageFusion.Org, 2007). These images were fused with three different pixel-based fusion algorithms: Contrast Pyramid (PYR); Discrete Wavelet Transform (DWT); and the Dual-Tree Complex Wavelet Transform (CWT; see Lewis et al., 2007).

## 4.2 Experiment

A group of 63 subjects with normal or corrected to normal vision manually segmented both the individual and the fused images. The average subject's age was 21.3 years (standard deviation = 2.7 years). A mixture of CRT (37) and TFT (26) screens were used. The segmentation instructions quite general, in order not to bias the subject to produce a specific type of segmentation. Thus, variations in segmentations were due to differences in perception and not to some other aspect of the experimental set up.

## 4.3 Results and discussion

Fig. 8 shows the annotated union of the human segmentations of each of the 7 scenes used in this study. In general the manual segmentations represent the actual scene layout quite well. Typical examples of human segmentations are shown in Fig. 12.

To compare the performance of subjects with the different individual (visual and infrared) and (CWT, DWT, and Pyramid) fused image modalities we adopted the following approach. For all features marked by the human subjects, we computed the percentage of subjects that completely delineated them. Then we defined the relevant features in the different scenes as those features that were fully segmented in either the visual or infrared images by more than half of the number of subjects. In previous studies we found a clear distinction in the performance of human observers using the different individual and fused image modalities for the detection of respectively terrain features, persons and man-made objects like buildings and cars (Toet et al., 1997b; Toet & Franken, 2003). In this study, we therefore classify the relevant features in three categories: terrain features, living creatures, and man-made objects. Typical terrain features are roads, trees, hills, and clouds. Typical man-made objects are houses, fences, poles, chimneys, boats, and buoys. Living creatures are for instance people and dogs. Then we computed the average percentage of subjects that fully segmented image features, for each feature category and for all image modalities.

The results are shown in Fig. 9. It appears that terrain features are best detected in the visual image modality, which yields the worst performance for the detection of living creatures. For the set of images tested in this study, human performance for the detection of man-made objects is quite similar for all image modalities. The CWT fusion scheme appears to yield the best overall performance. Each of the fused image modalities performs similar to the infrared image modality for the detection of living creatures, indicating that these schemes correctly include details from the infrared images in the resulting fused images. However, the performance of the fused image modalities for the detection of terrain features is below the performance with the visual image modality. This suggests that the representation of the visual details is not optimal in the fused images. Finally, for each of the fused image modalities we computed the mean percentage of objects that were segmented by a percentage of the human subjects that was larger than the percentage that segmented the same objects in either of the input image modalities. This number represents the percentage of cases in which a fused image is more than the sum of its parts: subjects can perceive details better in the fused image than in each of the individual input images.

| Scene      | Visual  | Infrared   | Features   |
|------------|---|--|--|
| UNCamp     |    |    | Man-made:<br>fence<br>poles<br>roof<br>chimney<br>Terrain :<br>road<br>hill with shrubs<br>trees left<br>trees right<br>Living creatures:<br>man |
| Dune_7404  |   |   | Terrain :<br>crater<br>hill<br>small path<br>road<br>Living creatures:<br>man  |
| Octec02    |  |  | Man-made:<br>house 1-6<br>Terrain :<br>trees left<br>trees right<br>road<br>Living creatures:<br>man   |
| Octec21    |  |  | Man-made:<br>house 1-6<br>Terrain :<br>trees left<br>trees right<br>smoke cloud<br>Living creatures:<br>man                                      |
| Trees_4906 |  |  | Terrain :<br>trees left<br>trees upper right<br>trees lower right<br>border between trees<br>Living creatures:<br>man                            |

Fig. 7. The visual (2<sup>nd</sup> column) and infrared (3<sup>rd</sup> column) images of each of the 7 different scenes used in this study, with a list of the characteristic man-made and terrain features, as well as people or animals that were used to score subject performance.

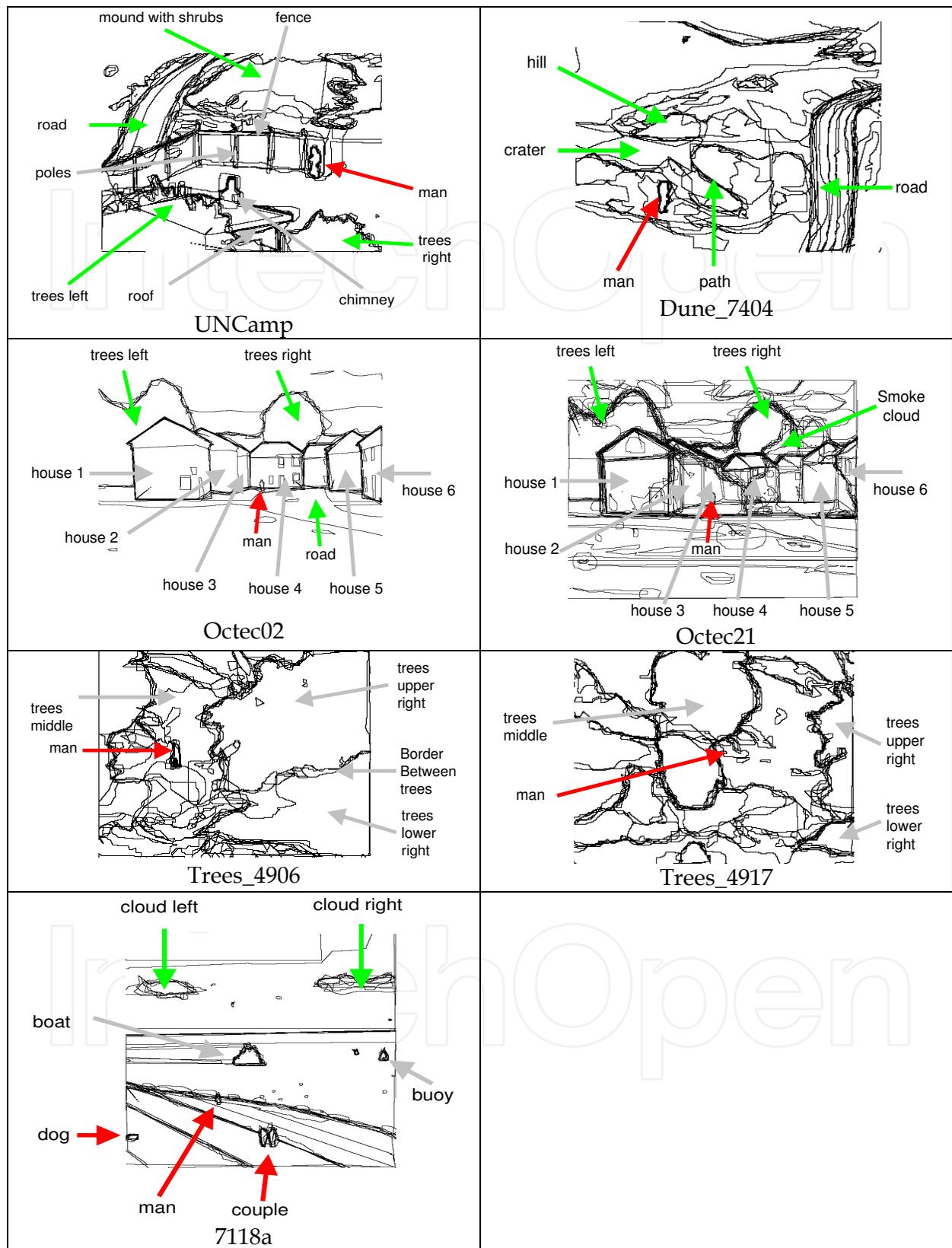


Fig. 8. Annotated union of all human segmentations of each of the 7 scenes used in this study. Indicated and labeled are the terrain features (green), man-made objects (gray) and living creatures (red) used to score the subject performance.

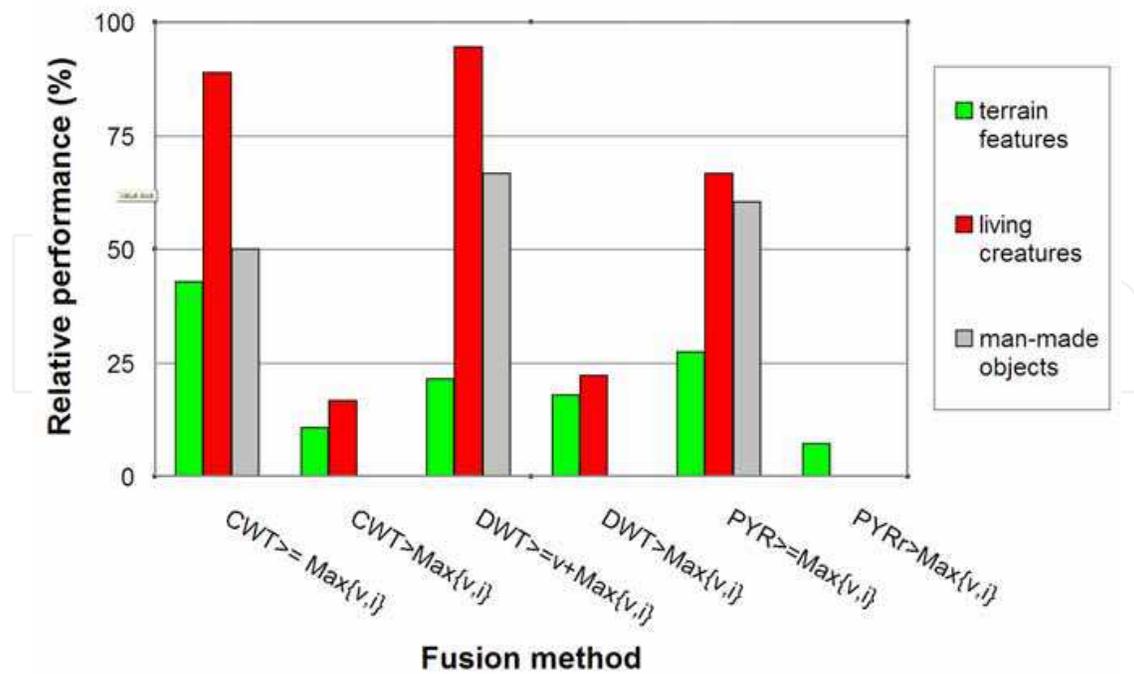


Fig. 9. Performance of subjects with the (CWT, DWT, and Pyramid) fused image modalities, expressed as the mean percentage of objects segmented by a fraction of subjects that was equal to or larger than the fraction of subjects that segmented these objects in either the visual (v) or infrared (i) image modalities.

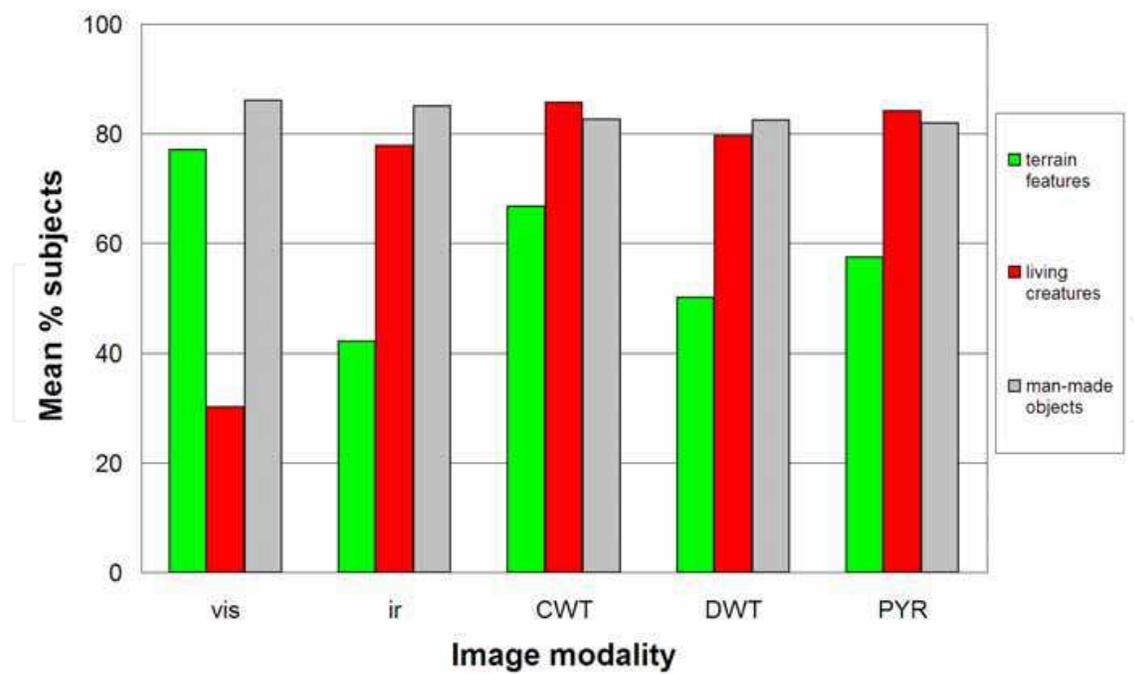


Fig. 10. The mean percentage of subjects that detected relevant features, for each class of objects (terrain, living creatures and man-made objects) and for each of the individual (visual and infrared) and (CWT, DWT, and Pyramid) fused image modalities.

Fig. 10 shows the percentage of cases in which the performance with fused imagery was both equivalent to (equal or larger than) or better (larger) than the performance with each of the input image modalities. This figure confirms the results in Fig. 9 by showing that most image modalities yield a performance for the detection of living creatures that is equivalent to that obtained with the input image modalities. The performance for the detection of man-made objects is below the performance with the input images. The performance for the detection of terrain features is considerably reduced, suggesting that the details from the visual images are not optimally represented in the fused images. There are only a few cases in which the performance with fused imagery exceeds the performance with the individual image modalities. This only occurs for the detection of living creatures and of terrain features.

Currently no well-established methods for objective image segmentation quality evaluation are available (Correia & Pereira, 2002; Correia & Pereira, 2006; Correia & Pereira, 2003). We will therefore use the boundary precision-recall measure, which has become a standard evaluation procedure in the information retrieval community (van Rijsbergen, 1979), as the evaluation criterion in the comparison of the human segmentation boundaries for the different image modalities. Precision is the fraction of detections that are true positives rather than false positives, while recall is the fraction of true positives that are detected rather than missed. Precision and recall are traditionally used to measure the performance of information extraction and information retrieval systems (van Rijsbergen, 1979), and have more recently also been applied to measure the performance of edge detection and image segmentation schemes (Martin et al., 2004), and the efficacy of multimodal image fusion schemes (Davis & Sharma, 2007). In the context of boundary detection, two types of errors arise. Type-I errors occur if a true object boundary has not been detected by the segmenter (boundary deletion). Type-II errors occur if a detected object boundary does not correspond to a segment boundary in the reference (false alarm, or boundary insertion). Precision and recall can then be expressed by the Type-I and Type-II error rates as follows:

$$R = \frac{\text{number of correctly detected reference boundary pixels}}{\text{total number of reference boundary pixels}} \quad (1)$$

and

$$P = \frac{\text{number of correctly detected reference boundary pixels}}{\text{total number of detected boundary pixels}} \quad (2)$$

Thus, precision is the fraction of detected boundaries that are indeed true boundaries, while recall is the fraction of true boundaries in the image that are actually detected. Note that both precision and recall are bounded between 0 and 1.

In the context of boundary detection, the precision and recall measures are particularly meaningful in applications that make use of boundary maps, such as stereo or object recognition. It is reasonable to characterize this type of higher level processing in terms of how much true signal is required to succeed (recall), and how much noise can be tolerated (precision). A particular application can define a relative cost  $\alpha$  between these quantities. The  $F$ -measure (van Rijsbergen, 1979), defined as

$$F = PR / (\alpha R + (1 - \alpha)P) \quad (3)$$

captures this trade off as the weighted harmonic mean of the precision  $P$  and the recall  $R$ . Like  $R$  and  $P$  the  $F$ -measure is bounded between 0 and 1. In a precision-recall graph, higher  $F$ -measures correspond to points closer to  $(P,R) = (1,1)$ , representing maximal precision and recall for a given  $\alpha$ . In our present experiments we choose the neutral parameterization and set  $\alpha$  equal to 0.5, so that precision and recall are weighted equally, and (3) becomes

$$F = 2PR / (R + P) \quad (4)$$

Here we propose to use a combination of the manual segmentations of each of the individual image modalities to construct a reference contour image that can be used to evaluate the different fusion schemes. The segmentation data set provides multiple human segmentations for each image. Simply constructing a reference contour image by taking the union of individual manual boundary maps is not effective because of the localization errors present in the data set itself. Localization errors are inherent in a human image segmentation task, since human subjects are limited in the accuracy with which they can draw the edges they observe in the images. Evidently, some objects simply have no well defined boundaries (grass, trees, clouds). Moreover, for the type of imagery used in this study, the object representations are often not sharp. As a result, there is an inherent positional uncertainty in the manually drawn boundaries for the imagery used in this study. In the rest of this study we will use procedures to match different boundary representations. Simply matching corresponding coincident boundary pixels and declaring all unmatched pixels either false positives or misses would not tolerate any localization error. Therefore, we permit a controlled amount of localization error, by adopting a distance tolerance region with a radius of 20 pixels (this value was found to yield appreciable results throughout the entire procedures presented in this study). Any boundary pixel detected within this tolerance region around the location of a true (reference) boundary pixel is regarded as a correct detection.

Now we will discuss the steps taken in the construction of a reference contour image. First, the individual boundary maps resulting from the human segmentations are converted into binary mask images. For a given object boundary, a boundary mask image represents all pixels that are within a given tolerance distance of this boundary. These boundary masks are introduced to allow for small localization errors in the human segmentation data. The binary boundary mask image is obtained by first computing the exact squared two-dimensional Euclidean distance transform of the binary contour image (Figure 6a) using a square 3x3 structuring element (Lotufo & Zampirolli, 2001). The result of this transform is a graylevel image in which the value of each background pixel represents the Euclidian distance to the nearest boundary pixel (Fig. 11b; e.g.

[http://en.wikipedia.org/wiki/Distance\\_transform](http://en.wikipedia.org/wiki/Distance_transform)). Thresholding this distance image at the aforementioned distance threshold level of 20 pixels gives the binary mask image (Fig. 11).

Next, for each image modality, the corresponding binary boundary mask images from all subjects are summed. The summed mask image represents the number of subjects that have marked each individual pixel as a boundary pixel. The summed mask image is then thresholded at a level corresponding to half the number of subjects that contributed to the sum. Thus we obtain a binary mask image that represents the consensus among at least half of the subjects about the boundary status of each pixel (i.e. a pixel has value 1 if at least half of the subjects have marked this pixel as a boundary pixel; Fig. 12 lower left).

Then, we compute the morphological skeleton (Maragos & Schafer, 1986) of the binary consensus mask image (Fig. 12 lower right). This is done by a morphological thinning operation (Serra, 1982) that successively erodes away pixels from the boundary (while preserving the end points of line segments) until no more thinning is possible, at which point what is left represents the skeleton.

Finally, a joined binary mask for the combination of visual and infrared boundaries is produced as the logical union of the corresponding individual binary consensus mask images (Fig. 13 lower left). From the resulting binary mask image a joined skeleton image is then constructed (Fig. 13 lower right). In the following we will refer to the joined binary consensus mask and its morphological skeleton as respectively the *reference mask* and the *reference contour image*. Note that the reference contour image represents the combination of the maximal amount of object boundary information that was extracted by human visual inspection from each of the individual image modalities.

For each of the fused image modalities precision and recall measures are then computed as follows. First, we count the number of non-zero (object) pixels in the reference contour image ( $n_{ref}$ ) and those in the corresponding boundary image manually drawn by a human subject ( $n_{subject}$ ). To compute the number of pixels in the boundary image drawn by the subject that are accounted for by the reference mask (the number of hits:  $na_{subject}$ ) we take the intersection of the subject's boundary image and the reference mask, and count the number of non-zero pixels. Similarly, to compute the number of pixels in the reference contour image that are accounted for by the subject's boundary drawing ( $na_{reference}$ ) we

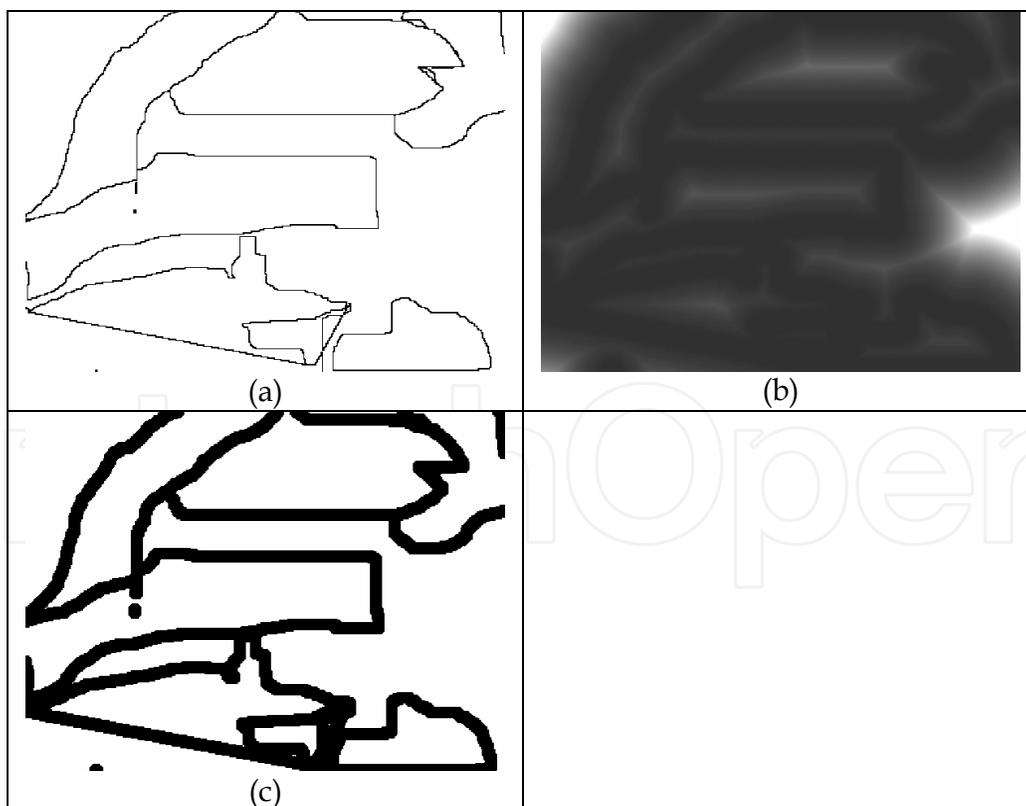


Fig. 11. (a) Boundary drawn for the visual image of the UNCamp scene (see Fig. 7) by a human subject. (b) Distance transform of (a). (c) Mask image obtained by thresholding (b) at distance level 20.

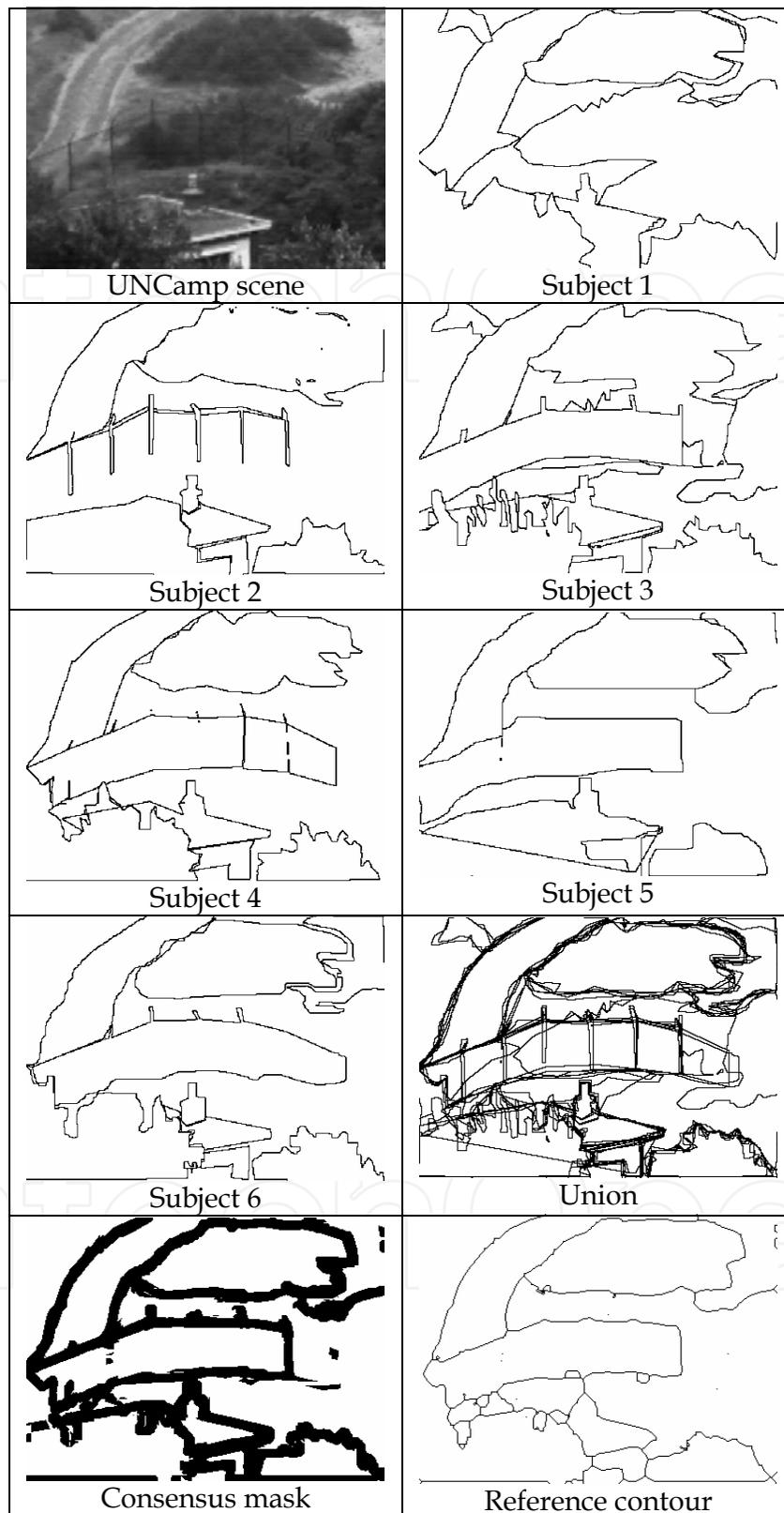


Fig. 12. Boundaries drawn by 6 human subjects for the visual image of the UNCamp scene, the union of all these boundaries, the consensus mask image (lower left) representing the thresholded sum of all boundary masks (i.e. the dilated boundary images; not shown here), and the resulting reference contour (lower right).

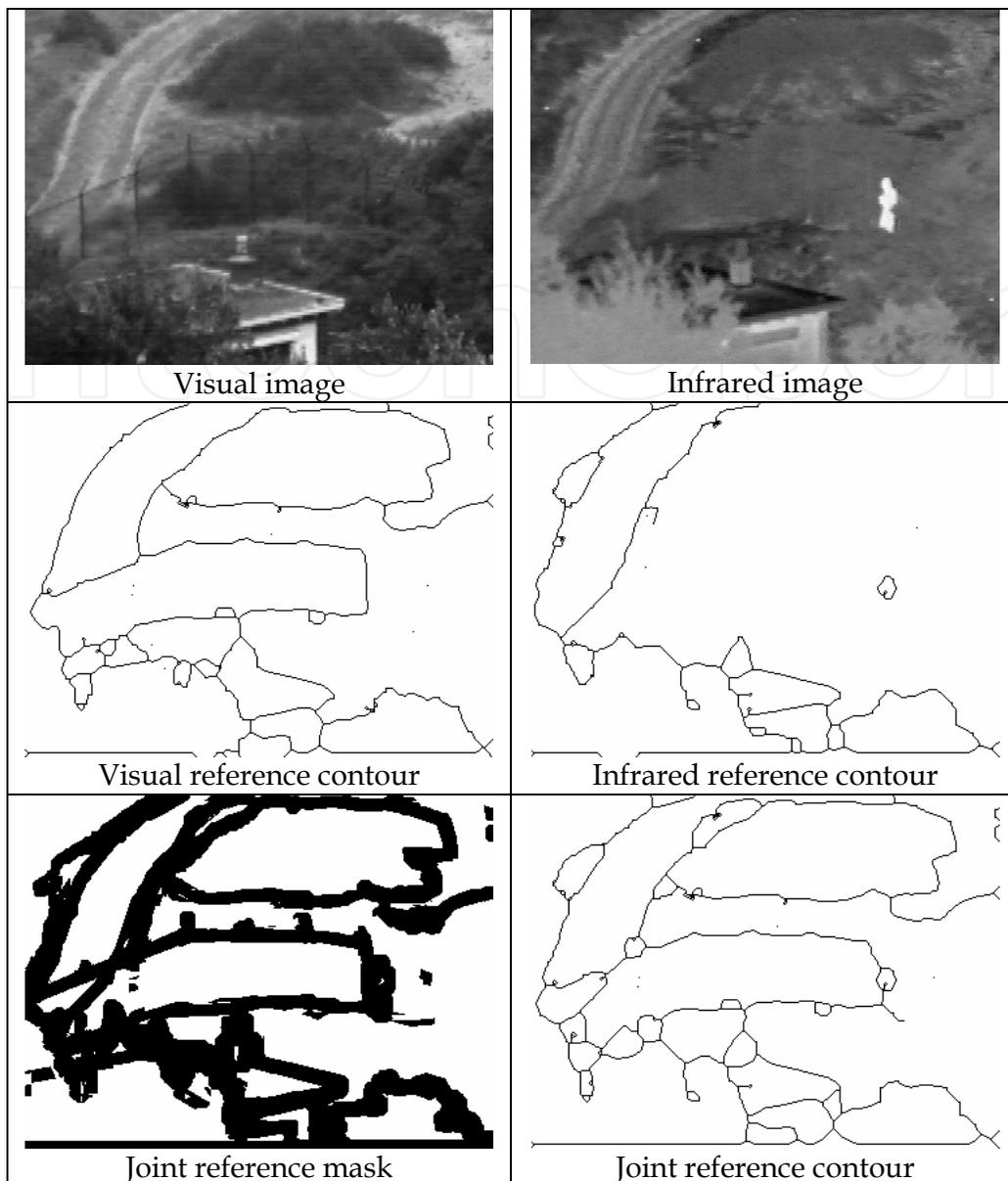


Fig. 13. The visual and infrared input images of the UNCamp scene (top row), their reference contour representations (middle row), and the joint reference contour mask and contour image (lower row).

take the intersection of the reference contour image and the subject's boundary mask image, and count the number of non-zero pixels. For each individual human subject  $i$  the precision ( $P_i$ ) and recall ( $R_i$ ) measures are then computed as the fraction of accounted pixels in both the subject's drawing and the reference contour image:

$$P_i = na_{\text{subject}} / n_{\text{subject}} \quad ; \quad R_i = na_{\text{reference}} / n_{\text{reference}} \quad (5)$$

The F-measure for each human subject  $i$  is then be computed from (5) as:

$$F_i = \frac{2 \cdot na_{\text{subject}} \cdot na_{\text{reference}}}{n_{\text{reference}} \cdot na_{\text{subject}} + na_{\text{reference}} \cdot n_{\text{subject}}} \quad (6)$$

The overall precision, recall and F-measures are then computed as the mean over all  $N$  subjects:

$$P = \frac{1}{N} \sum_{i=1}^N P_i ; R = \frac{1}{N} \sum_{i=1}^N R_i ; F = \frac{1}{N} \sum_{i=1}^N F_i \quad (7)$$

Fig. 14 shows the precision and recall measures computed for each of the individual skeletons of the visual, infrared, CWT, DWT and Pyramid fused image modalities. This figure shows that the individual manual segmentations agree to a large extent with their overall skeleton representation (median value of  $F=0.72$ ). A collection of manual image segmentations can therefore be represented by a single overall skeleton.

Fig. 15 shows the precision and recall measures computed for the unified skeleton representation of the visual and infrared human boundary data, and the human boundary data for each of the (CWT, DWT and Pyramid-) fused image modalities. This result shows that the precision of the boundaries drawn by the subjects is actually quite high, meaning that the fusion schemes do not seem to introduce any spurious details. However, the fraction of recalled details is around 0.5, which is rather low. This reflects the effect that terrain details are not well perceived by the subjects in the fused images.

Summarizing, we conclude that reference contour images are a useful tool to evaluate the performance of image fusion schemes.

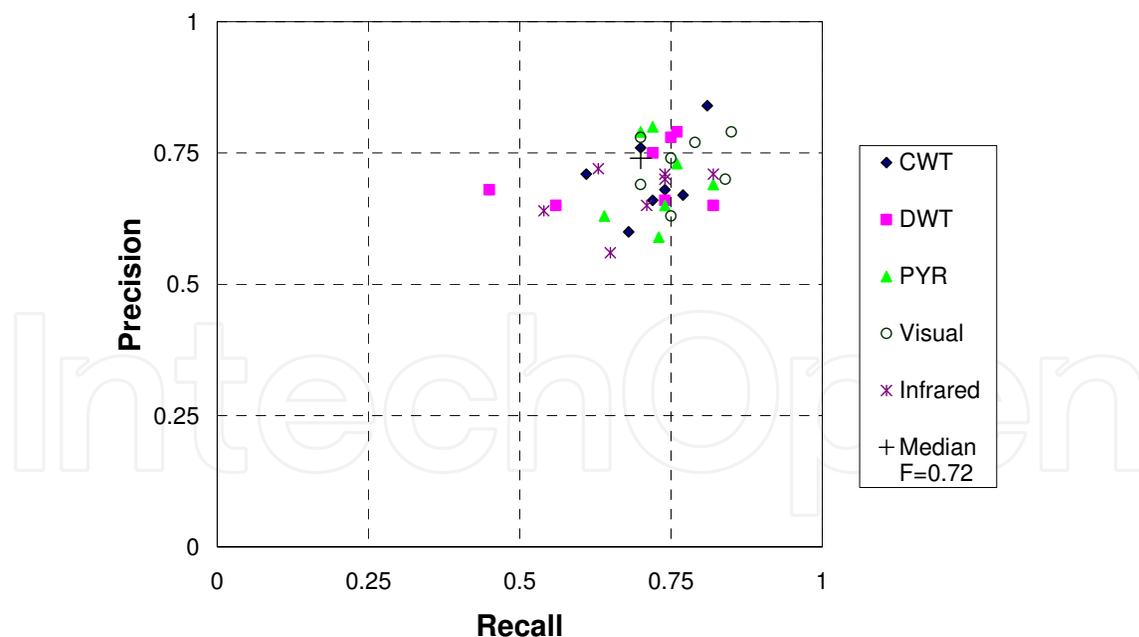


Fig. 14. Consistency between the skeleton representation of each of the individual (visual, infrared) and each of the fused (CWT, DWT, PYR) image modalities and the subject data. This figure shows that the skeleton is a reliable representation of the data.

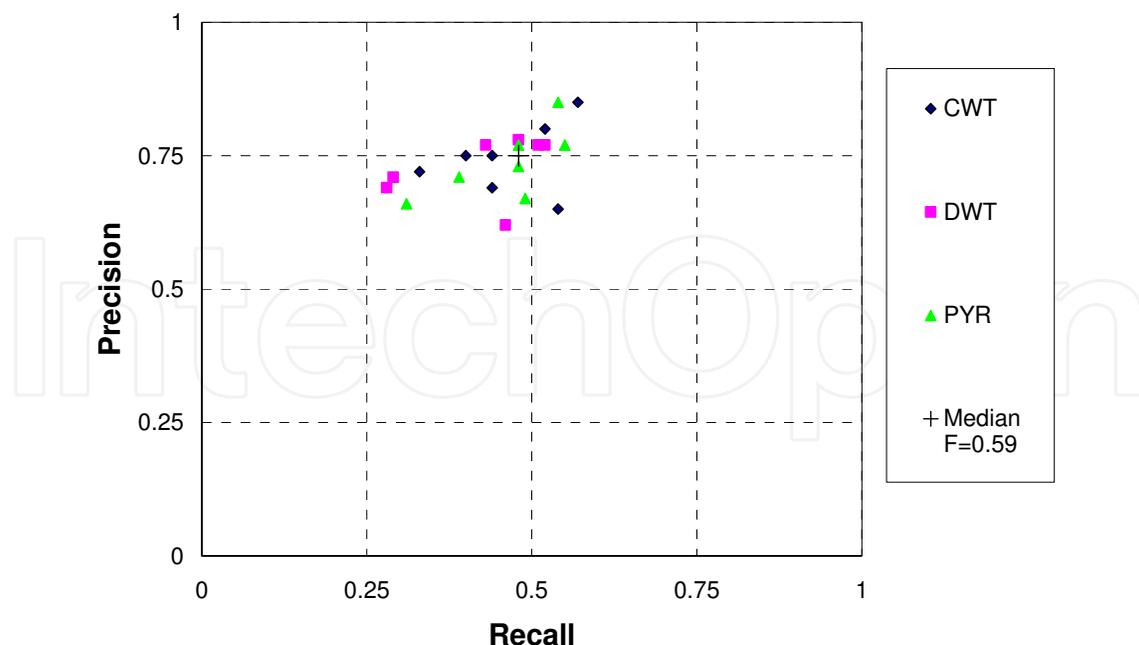


Fig. 15. Consistency between the unified skeleton representation of the visual and infrared human boundary data, and the human boundary data for each of the (CWT, DWT and Pyramid-) fused image modalities.

## 5. Camouflage detection

Although natural color mapping schemes provide many perceptual benefits, they are not suitable for all purposes. A typical example is the task of detecting soldiers wearing camouflage suits in a rural setting, using a two-band nightvision system sensitive to the visual and thermal part of the electromagnetic spectrum. When the false color representation of the fused nightvision image optimally agrees with the daytime appearance of the scene, the soldiers will blend in with their environment (will be camouflaged), which makes it nearly impossible to perform the task. In such cases a color mapping scheme should be used which displays the objects of interest with higher color contrast while retaining an intuitive (natural) color setting for the rest of the scene.

As an example we present the results of a color mapping which optimizes the detection of man-made camouflaged targets in a rural setting, while retaining a natural color representation of the environment.

### 5.1 Imagery

We registered optically aligned visual (wavelengths shorter than 700 nm) and near-infrared (NIR; wavelengths longer than 700 nm) nighttime images of a rural scene containing grass and trees, with and without targets in the scene. The targets were blue and green foam tubes (Fig. 16). For comparison we also created a standard intensified image of each scene containing both bands, since this is the type of image typically provided by standard night vision goggles. First, a red-green false color representation of the fused dual-band sensor image was obtained by mapping the visual band to the Red channel and the NIR band to the Green channel of an RGB-image (Fig. 17d). Next, for each combination of sensor outputs



Fig. 16. Images showing the two target types, the green target (a) and the blue target (b).

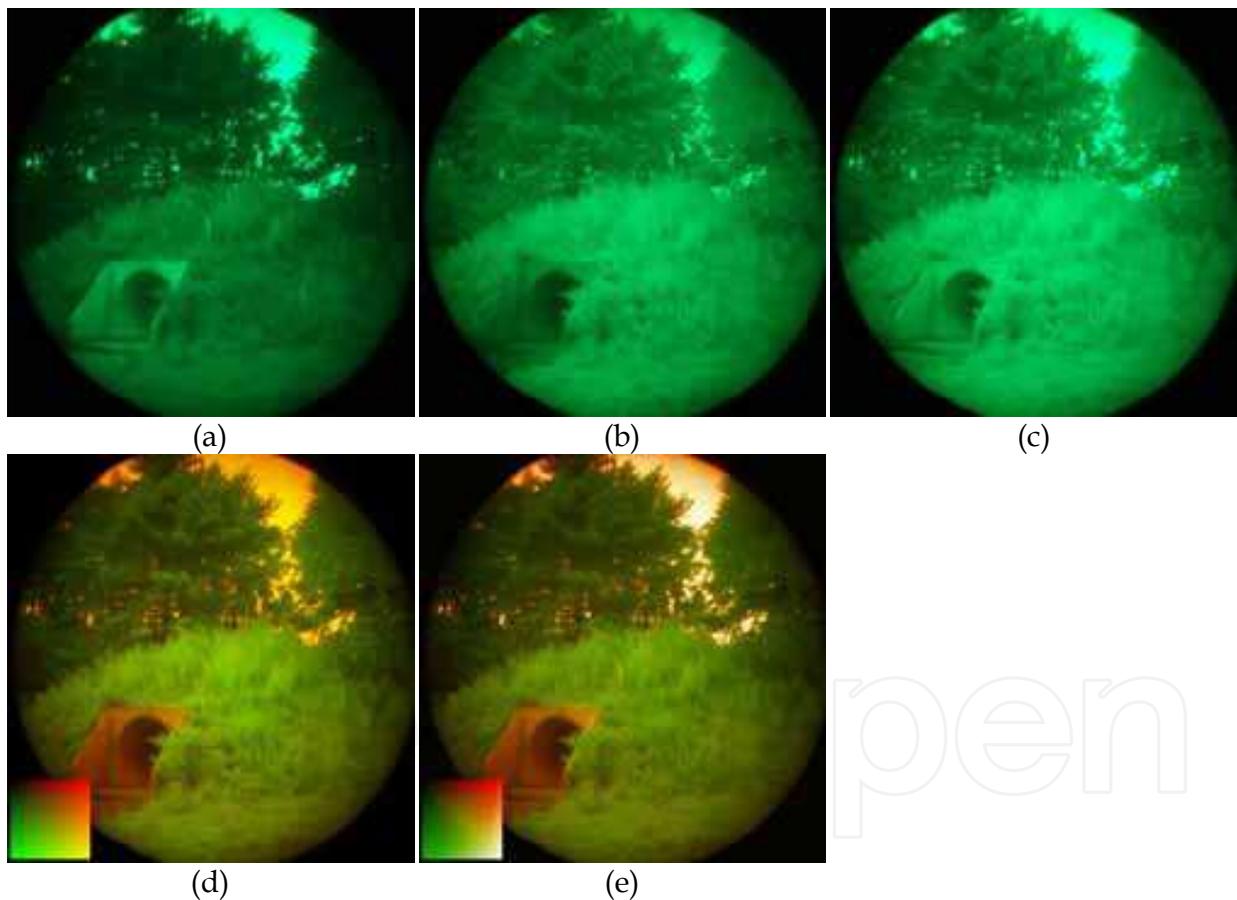


Fig. 17. Lookup table based color remapping applied to a dual-band visual (a) and NIR (b) image. (c) A regular intensified image representation for comparison (e.g. a standard night vision goggle image). (d) A red-green false color representation of the dual-band image with the visual band assigned to the Red and NIR band assigned to the Green channel of an RGB display. The inset in (d) shows all possible dual-band outputs as shades of red (large response in band 1, small in band 2), green (small response in band 1, large in band 2) and yellow (large responses in both bands). (e) The result of the color transformation. The inset shows how the colors in the inset of (d) are transformed.

(represented by a shade of red, green, yellow; see inset of Fig. 17d) a color was selected to display this sensor output. This process was implemented by transforming the red-green image (Fig. 17d) into an indexed image in which each pixel value refers to the entry of a color lookup table. When a different color lookup table is used, the colors in the indexed image are automatically transformed, such that all pixels with the same index are displayed in the same color. The method is described in detail elsewhere (Hogervorst & Toet, 2008a; Hogervorst & Toet, 2010). We found that the color transformation which maximizes the visibility of the targets while preserving the natural appearance of the scene is quite similar to the red-green representation, with a few modifications that specifically address the target colors.

The inset of Fig. 17e shows the colors assigned to all dual-band outputs (the inset of Fig. 17d) by the chosen color scheme. This color scheme emphasizes the distinction between objects containing chlorophyll (the background plants) and objects containing no chlorophyll (e.g. the foam tube targets; notable from the sharp transition between green and red at the diagonal). The dual band sensor system separates the incoming light in a part with wavelengths below 700nm and one with wavelengths above 700 nm. Since chlorophyll shows a steep rise around 700nm, this dual-band system is especially suited for discriminating materials containing chlorophyll from materials containing no chlorophyll. Elements containing chlorophyll (e.g. plants) are displayed in green (i.e. in their natural color), while objects without chlorophyll are displayed in the perceptually opposite color red. To further increase the naturalness, elements with high output in both channels are displayed in white (bottom right corner of the inset of Fig. 17e). The result of our color mapping is shown in Fig. 17e.

## 5.2 Experiment

We evaluated the abovementioned color mapping in a target detection paradigm. We registered both nighttime dual-band (visual and NIR) images and daytime full color digital photographs of a scene containing grass and trees, with and without targets present. Performance for detecting targets was established for imagery of the dual-band fusion system, each of the individual sensor bands (visual and NIR), standard NVG, and daytime images (taken with a regular digital photo camera). The visual angle and display area of the daytime images were matched to those of the nighttime images.

The targets were green (Fig. 18a) and blue (Fig. 18b) foam insulation tubes. The reflectance of the tubes was such the green tubes were mostly undetectable in a standard intensified



Fig. 18. The green target (a) and the blue target (b) situated in a background with grass and trees.

image representation and in the NIR band (see Fig. 17), but quite distinct (as bright objects) in the visible band (see Fig. 17). In contrast, the blue tubes were mostly undetectable in the visual band, but clearly visible (as dark objects) in the NIR band and in regular intensified images (Fig. 19).

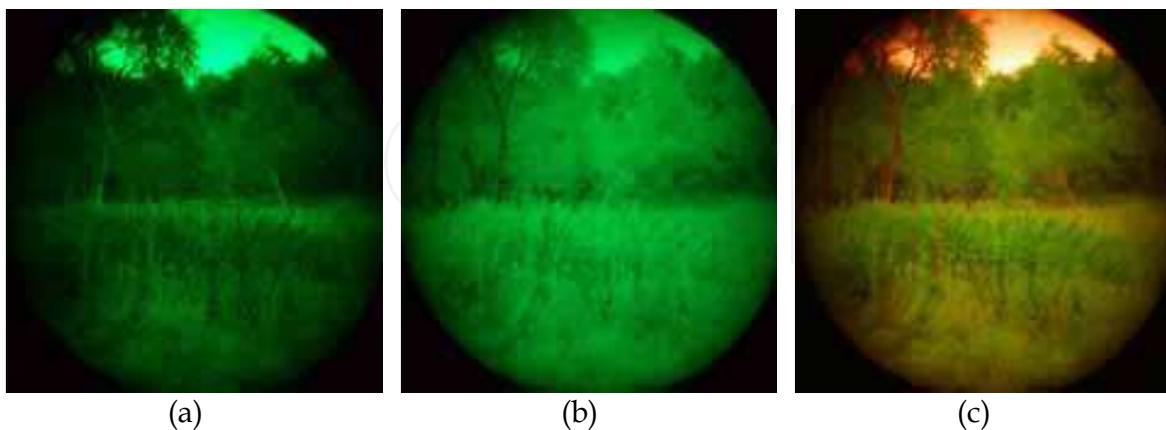


Fig. 19. Visual (a), NIR (b) and the color fused dual-band image (c) for a scene including a blue target. The target is visible in the NIR band as a dark tube. The dual-band image shows the target as a reddish object.

We recorded whether subjects detected the targets when present (Hits and Misses), and whether they judged there to be a target when no target was present (False Alarms and Correct Rejections). We also recorded the response times. Since no False Alarms occurred in this experiment (i.e. the False-Alarm rate was zero), observer performance is fully characterized by the Hit-rate, i.e. the fraction of targets that was detected ( $ph = \#Hits / (\#Hits + \#Misses)$ ). Observer performance was measured for 5 different image modalities:

1. Daytime: full color daylight images (taken with a standard digital daytime camera),
2. II: grayscale intensified images, combining both visual and NIR part of the spectrum,
3. VIS: grayscale intensified images representing only the visual part of the spectrum,
4. NIR: grayscale intensified images representing only the NIR part of the spectrum,
5. FC: false color images resulting from the natural color remapping method.

Seven subjects participated in the experiment. The images were shown on a CRT. The subjects indicated as quickly as possible whether a target was present or not, by clicking the appropriate mouse button. Next, the image disappeared and was replaced by a low resolution equivalent of the image, consisting of 20x15 uniformly colored squares (to prevent subjects from continuing their search after responding). We registered the time between onset of the stimulus and detection (the response time). The subject then indicated the perceived target location or clicked on an area outside the image labeled "no target found". Responses outside an ellipse with horizontal diameter of 162 and vertical diameter of 386 pixels centered on the vertically elongated target were considered as incorrect.

### 5.3 Results and discussion

Fig. 20 shows the fraction of hits (hit-rate) for the various sensor conditions and target colors. Shown are the average hit-rates over subjects. Not surprisingly, performance is highest in the Daytime condition. As expected (see Fig. 17 and Fig. 19), performance for detecting the green targets is high in the visual (VIS) condition and low in the image intensified (II) and NIR sensor conditions. Performance for detecting the blue targets is

somewhat poorer in the single-band conditions. These targets can be detected in the NIR condition (reasonably well) and in the II condition (poorly), while they are hardly detected in the VIS condition. Detection performance for both targets is high with the false-color dual-band sensor. Optimal fusion results in performance that equals maximum performance in the individual bands. The hit-rate for the green targets is somewhat lower for the dual-band than for the visual condition. But the hit-rate for the blue targets is somewhat higher for dual-band than for NIR condition. The average hit-rate of the false color dual band sensor (0.75) is not significantly different from the average of the hit-rate for green in VIS and the hit-rate for blue in NIR (0.78). This means that this fusion scheme is near optimal. The results also show that the performance with the standard intensified imagery is clearly much worse than with the false-color dual-band NVG system.

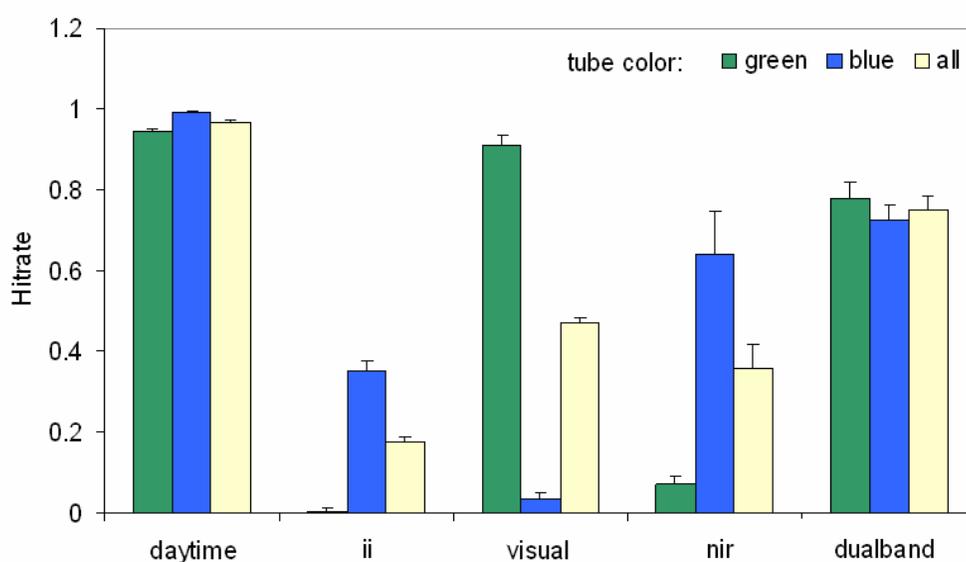
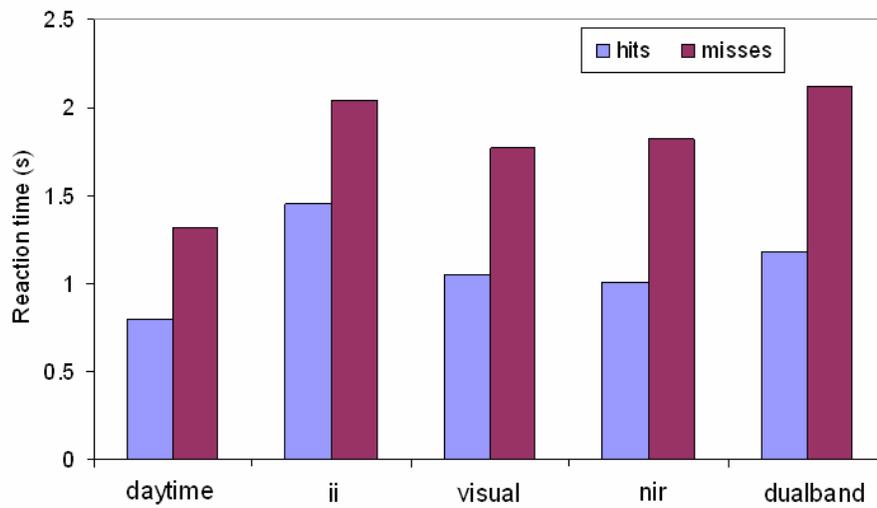


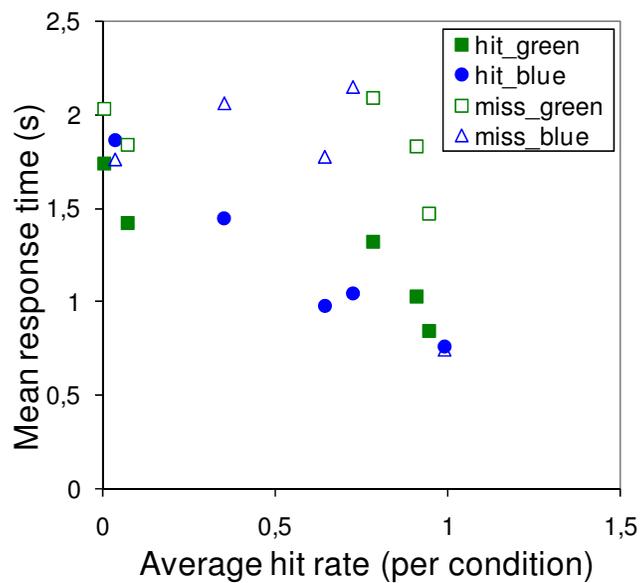
Fig. 20. Average (over all subjects) hit-rate (fraction of hits) for each of the 5 different image modalities and the 2 target colors, including the overall hit-rate (“all”). The error bars represent standard errors in the mean derived from the variance between subjects.

Fig. 21 shows the response times of the trials containing a target (shown are the geometric means over the response times, i.e. the exponent of the average log response times) for all conditions for the hits and misses. Note that the hits for the NIR and II modalities correspond primarily to the trials containing blue targets; the hits for the Visual modality correspond primarily to the trials containing green targets. The response times for the false color dual-band condition are comparable, but slightly larger than in the single-band Visual and NIR conditions. This may be due to the fact that in this condition subjects had to attend to two types of targets, while in the single band conditions only one of the target colors was apparent.

It turns out that the response times for missed targets are comparable to the response times for stimuli in which no target is present. The average response times for missed targets do not correlate with the hit-rates (see Fig. 21b). In contrast, the average response times for hits is highly correlated with the hit-rate ( $r = -0.90$ ,  $p < 0.01$ , see Fig. 21b). This indicates that when targets are more easily detected, the hit-rate goes up and the response time goes down.



(a)



(b)

Fig. 21. (a) The geometric mean (i.e. averaged in log) response times for the various image modalities, separated for hits and misses. (b) Relationship between the hit-rate for each image modality and the (geometric) mean response times for hits and misses for the two target colors.

The results show that performance of the false color dual-band system is just as good as the maximum performance that can be attained using either of its individual bands (visual and NIR). While the green targets can be detected with the visual band of the system alone, the blue targets are mostly missed when subjects have to rely on this band alone. In contrast, the blue targets can be detected with the NIR band, but the green targets are then largely missed in this modality. With the false color dual-band image modality both targets can be detected. The total number of targets detected in the dual band image modality is the same as the total number of targets detected in the visual and modality plus the number of targets detected in the NIR image modality. This indicates that the fused color representation of the two bands is (nearly) optimal from a perceptual standpoint.

## 6. Conclusions

We find that observers can localize persons in a scene more accurately using fused intensified visual and thermal imagery, than with each of the individual image modalities. The addition of color does not improve this accuracy. A spatial localization task is useful tool to assess the information content of fused imagery intended for surveillance and navigation tasks.

IR and intensified visual imagery contain complementary information. IR imagery mainly contributes to the recognition of buildings and living creatures, whereas intensified visual imagery predominantly shows natural terrain features and efficiently provides the gist of the scene. Experiments testing scene recognition and situational awareness can be used to investigate the perceptual quality of images fusion and color mapping schemes.

The fusion methods used in this study degrade the perception of terrain features. Our finding that the fraction of recalled boundary contours is rather low suggests that details from the visual images are not fully transferred to the fused images. The detection of living creatures is similar in all fused images, indicating that these high-contrast details from the IR images are correctly represented in the fused images. Reference contour images obtained from human segmentations are a useful tool to systematically evaluate the quality of the representation of object boundaries in fused imagery.

The application of appropriate color mapping schemes in the image fusion process can significantly improve observer performance compared to grayscale fusion. In contrast, the use of inappropriate color schemes can severely degrade observer sensitivity. However, color mappings which are perceptually suboptimal may still have ergonomic value and lead to an overall improvement of observer performance, because they eliminate the need to switch attention between different image modalities, thereby reducing the user's cognitive workload. Color mapping schemes can also be tuned to optimize the visibility of camouflaged targets in fused imagery, thus providing larger hit rates and faster detection times. Detection and recognition experiments can be used to assess and optimize the perceptual quality of color mapping schemes.

## 7. References

- Aguilar, M.; Fay, D.A.; Ireland, D.B.; Racamoto, J.P.; Ross, W.D. & Waxman, A.M. (1999). Field evaluations of dual-band fusion for color night vision, In: *Enhanced and Synthetic Vision 1999*, Verly, J.G. (Eds.), Vol. SPIE-3691, pp. 168-175, The International Society for Optical Engineering, Bellingham, WA.
- Aguilar, M.; Fay, D.A.; Ross, W.D.; Waxman, A.M.; Ireland, D.B. & Racamoto, J.P. (1998). Real-time fusion of low-light CCD and uncooled IR imagery for color night vision, In: *Enhanced and Synthetic Vision 1998*, Verly, J.G. (Eds.), Vol. SPIE-3364, pp. 124-135, The International Society for Optical Engineering, Bellingham, WA.
- Angell, C. (2005). Fusion performance using a validation approach, In: *Information Fusion 2005*.
- Ansorge, U., Horstmann, G. & Carbone, E. (2005). Top-down contingent capture by color: evidence from RT distribution analyses in a manual choice reaction task. *Acta Psychologica*, Vol.120, No.3, 243-266.
- Blum, R.S. (2006). On multisensor image fusion performance limits from an estimation theory perspective. *Information Fusion*, Vol.7, No.3, 250-263.

- Blum, R.S. & Liu, Z. (2006). *Multi-sensor image fusion and its applications*. CRC Press, Taylor & Francis Group, ISBN , Boca Raton, Florida, USA.
- Burt, P.J. & Adelson, E.H. (1985). Merging images through pattern decomposition, In: *Applications of Digital Image Processing VIII*, Tescher, A.G. (Eds.), Vol. SPIE-575, pp. 173-181, The International Society for Optical Engineering, Bellingham, WA.
- Cavanillas, J.A. (1999). *The role of color and false color in object recognition with degraded and non-degraded images*. (Master's thesis) Monterey, CA: Naval Postgraduate School.
- Chari, S.K.; Fanning, J.D.; Salem, S.M.; Robinson, A.L. & Haford, C.E. (2005). LWIR and MWIR fusion algorithm comparison using image metrics, In: *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XVI*, Holst, G.C. (Eds.), Vol. SPIE-5784, pp. 16-26, The International Society for Optical Engineering, Bellingham, WA.
- Chen, H. & Varshney, P.K. (2005). A Perceptual Quality Metric For Image Fusion Based on Regional Information, In: Vol. SPIE-, The International Society for Optical Engineering, Bellingham, WA.
- Chen, H. & Varshney, P.K. (2007). A human perception inspired quality metric for image fusion based on regional information. *Information Fusion*, Vol.8, No.2, 193-207.
- Chiarella, M.; Fay, D.A.; Ivey, R.T.; Bomberger, N.A. & Waxman, A.M. (2004). Multisensor image fusion, mining and reasoning: rule sets for higher-level AFE in a COTS environment, In: *Proceedings of the Seventh International Conference on Information Fusion*, Svenson, P. & Schubert, J. (Eds.), pp. 983-990, International Society of Information Fusion, Mountain View, CA.
- Correia, P. & Pereira, F. (2002). Standalone objective segmentation quality evaluation. *EURASIP Journal on Applied Signal Processing*, Vol.4, No., 389-400.
- Correia, P. & Pereira, F. (2006). Video object relevance metrics for overall segmentation quality evaluation. *EURASIP Journal on Applied Signal Processing*, Vol.Article ID 82195, No., 1-11.
- Correia, P.L. & Pereira, F.M. (2003). Methodologies for objective evaluation of video segmentation quality, In: *Visual Communications and Image Processing 2003*, Ebrahimi, T. & Sikora, T. (Eds.), Vol. SPIE-5150, pp. 1594-1600, The International Society for Optical Engineering, Bellingham, WA., USA.
- Corsini, G.; Diani, M.; Masini, A. & Cavallini, M. (2006). Enhancement of Sight Effectiveness by Dual Infrared System: Evaluation of Image Fusion Strategies, In: *Proceedings of the 5th International Conference on Technology and Automation (ICTA'05)*, pp. 376-381.
- Cvejic, N., Loza, A., Bull, D. & Canagarajah, N. (2005a). A novel metric for performance evaluation of image fusion algorithms. *Transactions on Engineering, Computing and Technology*, Vol.V7, No., 80-85.
- Cvejic, N., Loza, A., Bull, D. & Canagarajah, N. (2005b). A similarity metric for assessment of image fusion algorithms. *International Journal of Signal Processing*, Vol.2, No.2, 178-182.
- Davis, J.W. & Sharma, V. (2007). Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer Vision and Image Understanding*, Vol.106, No.2-3, 162-182.
- Dixon, T.D., Canga, E.F., Troscianko, T., Noyes, J.M., Nikolov, S.G., Bull, D.R. & Canagarajah, C.N. (2006a). Assessment of images fused using false colouring. *Journal of Vision*, Vol.6, No.6, 459-a.

- Dixon, T.D.; Li, J.; Noyes, J.M.; Troscianko, T.; Nikolov, S.G.; Lewis, J.; Canga, E.F.; Bull, D.R. & Canagarajah, C.N. (2006b). Scanpath analysis of fused multi-sensor images with luminance change: a pilot study, In: *Special Session on Image Fusion Assessment. Proceedings of the 9th International Conference on Information Fusion*, Nikolov, S. & Toet, A. (Eds.), International Society of Information Fusion, Mountain View, CA.
- Dixon, T.D.; Noyes, J.; Troscianko, T.; Canga, E.F.; Bull, D. & Canagarajah, N. (2005). Psychophysical and metric assessment of fused images, In: *Proceedings of the 2nd symposium on Applied perception in graphics and visualization*, Bülthoff, H.B. & Troscianko, T. (Eds.), Vol. ACM International Conference Proceeding Series; Vol. 95, pp. 43-50, ACM Press, New York, USA.
- Driggers, R.G.; Krapels, K.A.; Vollmerhausen, R.H.; Warren, P.R.; Scribner, D.A.; Howard, J.G.; Tsou, B.H. & Krebs, W.K. (2001). Target detection threshold in noisy color imagery, In: *Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XII*, Holst, G.C. (Eds.), Vol. SPIE-4372, pp. 162-169, The International Society for Optical Engineering, Bellingham, WA.
- Essock, E.A., Sinai, M.J., DeFord, J.K., Hansen, B.C. & Srinivasan, N. (2005). Human perceptual performance with nonliteral imagery: region recognition and texture-based segmentation. *Journal of Experimental Psychology: Applied*, Vol.10, No.2, 97-110.
- Essock, E.A., Sinai, M.J., McCarley, J.S., Krebs, W.K. & DeFord, J.K. (1999). Perceptual ability with real-world nighttime scenes: image-intensified, infrared, and fused-color imagery. *Human Factors*, Vol.41, No.3, 438-452.
- Fay, D.A.; Waxman, A.M.; Aguilar, M.; Ireland, D.B.; Racamato, J.P.; Ross, W.D.; Streilein, W. & Braun, M.I. (2000a). Fusion of 2- /3- /4-sensor imagery for visualization, target learning, and search, In: *Enhanced and Synthetic Vision 2000*, Verly, J.G. (Eds.), Vol. SPIE-4023, pp. 106-115, SPIE -The International Society for Optical Engineering, Bellingham, WA, USA.
- Fay, D.A.; Waxman, A.M.; Aguilar, M.; Ireland, D.B.; Racamato, J.P.; Ross, W.D.; Streilein, W. & Braun, M.I. (2000b). Fusion of multi-sensor imagery for night vision: color visualization, target learning and search, In: *Proceedings of the 3rd International Conference on Information Fusion*, Vol. I, pp. TuD3-3-TuD3-10, ONERA, Paris, France.
- Fay, D.A.; Waxman, A.M.; Ivey, R.T.; Bomberger, N.A. & Chiarella, M. (2004). Multisensor image fusion and mining: learning targets across extended operating conditions, In: *Enhanced and Synthetic Vision 2004*, Verly, J.G. (Eds.), Vol. SPIE-5424, pp. 148-162, The International Society for Optical Engineering, Bellingham, WA., USA.
- Folk, C.L. & Remington, R. (1998). Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, Vol.24, No.3, 847-858.
- Fredembach, C. & Süssstrunk, S. (2008). Colouring the near-infrared, In: *Proceedings of the IS&T/SID 16th Color Imaging Conference*, pp. 176-182.
- Gegenfurtner, K.R. & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, Vol.10, No.13, 805-808.
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, Vol.12, No.6, 878-892.

- Green, B.F. & Anderson, L.K. (1956). Colour coding in a visual search task. *Journal of Experimental Psychology*, Vol.51, No., 19-24.
- Grossberg, S. (1988). *Neural networks and natural intelligence*. MIT Press, ISBN , Cambridge, MA.
- Hogervorst, M.A. & Toet, A. (2008a). Method for applying daytime colors to nighttime imagery in realtime, In: *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications 2008*, Dasarathy, B.V. (Eds.), Vol. SPIE-6974, pp. 697403-1-697403-9, The International Society for Optical Engineering, Bellingham, WA, USA.
- Hogervorst, M.A. & Toet, A. (2008b). Presenting nighttime imagery in daytime colours, In: *Proceedings of the 11th International Conference on Information Fusion*, pp. 706-713, International Society of Information Fusion, Cologne, Germany.
- Hogervorst, M.A. & Toet, A. (2010). Fast natural color mapping for night-time imagery. *Information Fusion*, Vol.11, No.2, 69-77.
- Howard, J.G.; Warren, P.; Klien, R.; Schuler, J.; Satyshur, M.; Scribner, D. & Kruer, M.R. (2000). Real-time color fusion of E/O sensors with PC-based COTS hardware, In: *Targets and Backgrounds VI: Characterization, Visualization, and the Detection Process*, Watkins, W.R. et al. (Eds.), Vol. SPIE-4029, pp. 41-48, The International Society for Optical Engineering, Bellingham, WA.
- Huang, G., Ni, G. & Zhang, B. (2007). Visual and infrared dual-band false color image fusion method motivated by Land's experiment. *Optical Engineering*, Vol.46, No.2, 027001-1-027001-10.
- ImageFusion.Org (2007). The Online Resource for Research in Image Fusion, In: <http://www.imagefusion.org/>, Last viewed March 2007.
- Jacobson, N.P. & Gupta, M.R. (2005). Design goals and solutions for display of hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.43, No.11, 2684-2692.
- Jacobson, N.P., Gupta, M.R. & Cole, J.B. (2007). Linear fusion of image sets for display. *IEEE Transactions on Geoscience and Remote Sensing*, Vol.45, No.10, 3277-3288.
- Jakobson, G.; Lewis, L. & Buford, J. (2004). An approach to integrated cognitive fusion, In: *Proceedings of the Seventh International Conference on Information Fusion*, Svenson, P. & Schubert, J. (Eds.), pp. 1210-1217, International Society of Information Fusion, Chatillon, France.
- Joseph, J.E. & Proffitt, D.R. (1996). Semantic versus perceptual influences of color in object recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol.22, No.2, 407-429.
- Kong, S.G., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B.R., Koschan, A., Yi, M. & Abidi, M.A. (2007). Multiscale Fusion of Visible and Thermal IR Images for Illumination-Invariant Face Recognition. *International Journal of Computer Vision*, Vol.71, No.2, 215-233.
- Krebs, W.K. & Ahumada, A.J. (2002). Using an image discrimination model to predict the detectability of targets in color scenes, In: *Proceedings of the Combating Uncertainty with Fusion - An Office of Naval Research and NASA conference , April 22-24, 2002.*, Office of Naval Research and NASA, Woods Hole, MA.
- Krebs, W.K.; Scribner, D.A.; Miller, G.M.; Ogawa, J.S. & Schuler, J. (1998). Beyond third generation: a sensor-fusion targeting FLIR pod for the F/A-18, In: *Sensor Fusion:*

- Architectures, Algorithms, and Applications II*, Dasarathy, B.V. (Eds.), Vol. SPIE-3376, pp. 129-140, International Society for Optical Engineering, Bellingham, WA, USA.
- Krebs, W.K. & Sinai, M.J. (2002). Psychophysical assessments of image-sensor fused imagery. *Human Factors*, Vol.44, No.2, 257-271.
- Lewis, J.J.; Nikolov, S.G.; Canagarajah, C.N.; Bull, D.R. & Toet, A. (2006). Uni-Modal versus Joint Segmentation for Region-Based Image Fusion, In: *Proceedings of the 9th International Conference on Information Fusion*, International Society of Information Fusion, Mountain View, CA.
- Lewis, J.J., O'Callaghan, R.J., Nikolov, S.G., Bull, D.R. & Canagarajah, N. (2007). Pixel- and region-based image fusion with complex wavelets. *Information Fusion*, Vol.8, No.2, 119-130.
- Li, G. & Wang, K. (2007). Applying daytime colors to nighttime imagery with an efficient color transfer method, In: *Enhanced and Synthetic Vision 2007*, Verly, J.G. & Guell, J.J. (Eds.), Vol. SPIE-6559, pp. 65590L-1-65590L-12, The International Society for Optical Engineering, Bellingham, MA.
- Li, J.; Pan, Q.; Yang, T. & Cheng, Y. (2004). Color based grayscale-fused image enhancement algorithm for video surveillance, In: *Proceedings of the Third International Conference on Image and Graphics (ICIG' 04)*, pp. 47-50, IEEE Press, Washington, USA.
- Lotufo, R. & Zampiroli, F. (2001). Fast multidimensional parallel euclidean distance transform based on mathematical morphology, In: *Proceedings of the XIVth Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2001)*, Wu, T. & Borges, D. (Eds.), pp. 100-105, IEEE Computer Society, Washington, USA.
- Macmillan, N.A. & Creelman, C.D. (1991). *Detection theory: a user's guide*. Cambridge University Press, ISBN , Cambridge, MA.
- Maragos, P. & Schafer, R. (1986). Morphological skeleton representation and coding of binary images. *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol.34, No.5, 1228-1244.
- Martin, D.R., Fowlkes, C.C. & Malik, J. (2004). Learning to Detect Natural Image Boundaries Using Local Brightness, Color, and Texture Cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI*, Vol.26, No.1, 1-20.
- Muller, A.C. & Narayanan, S. (2009). Cognitively-engineered multisensor image fusion for military applications. *Information Fusion*, Vol.10, No.2, 137-149.
- O'Brien, M.A. & Irvine, J.M. (2004). Information fusion for feature extraction and the development of geospatial information, In: *Proceedings of the 7th International Conference on Information Fusion (FUSION 2004)*, pp. 976-982, International Society of Information Fusion, Mountain View, CA.
- Oliva, A. (2005). Gist of a scene, In: *Neurobiology of Attention*, Itti, L. et al. (Eds.), pp. 251-256, Academic Press.
- Oliva, A. & Schyns, P.G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, Vol.41, No., 176-210.
- Onyango, C.M. & Marchant, J.A. (2001). Physics-based colour image segmentation for scenes containing vegetation and soil. *Image and Vision Computing*, Vol.19, No.8, 523-538.
- Piella, G. & Heijmans, H.J.A.M. (2003). A new quality metric for image fusion, In: *Proceedings of the IEEE International Conference on Image Processing*, Vol. III, pp. III-209-III-212, IEEE Press, Washington, USA.

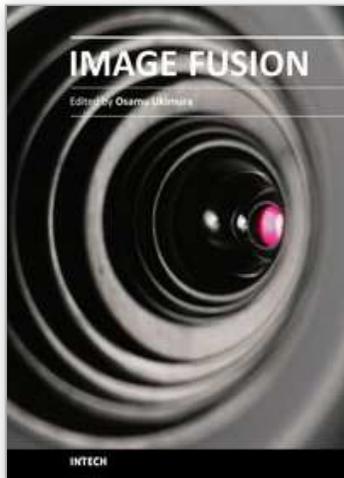
- Riley, P. & Smith, M. (2006). Image fusion technology for security and surveillance applications, In: *Optics and Photonics for Counterterrorism and Crime Fighting II*, Lewis, C. & Owen, G.P. (Eds.), Vol. SPIE-6402, pp. 640204-640204, The International Society for Optical Engineering, Bellingham, WA.
- Rousselet, G.A., Joubert, O.R. & Fabre-Thorpe, M. (2005). How long to get the "gist" of real-world natural scenes? *Visual Cognition*, Vol.12, No.6, 852-877.
- Sampson, M.T. (1996). *An assessment of the impact of fused monochrome and fused color night vision displays on reaction time and accuracy in target detection* (Report AD-A321226). Monterey, CA: Naval Postgraduate School.
- Schuler, J.; Howard, J.G.; Warren, P.; Scribner, D.A.; Klien, R.; Satyshur, M. & Kruer, M.R. (2000). Multiband E/O color fusion with consideration of noise and registration, In: *Targets and Backgrounds VI: Characterization, Visualziation, and the Detection Process*, Watkins, W.R. et al. (Eds.), Vol. SPIE-4029, pp. 32-40, The International Society for Optical Engineering, Bellingham, WA, USA.
- Scribner, D.; Schuler, J.M.; Warren, P.; Klein, R. & Howard, J.G. (2003). Sensor and image fusion, In: *Encyclopedia of optical engineering*, Driggers, R.G. (Eds.), pp. 2577-2582, Marcel Dekker Inc., New York, USA.
- Scribner, D.; Warren, P. & Schuler, J. (1999). Extending color vision methods to bands beyond the visible, In: *Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp. 33-40, Institute of Electrical and Electronics Engineers.
- Serra, J. (1982). *Image analysis and mathematical morphology*. Academic Press, ISBN , London, UK.
- Shi, J.; Jin, W.; Wang, L. & Chen, H. (2005a). Objective evaluation of color fusion of visual and IR imagery by measuring image contrast, In: *Infrared Components and Their Applications*, Gong, H. et al. (Eds.), Vol. SPIE-5640, pp. 594-601, The International Society for Optical Engineering, Bellingham, MA.
- Shi, J.-S., Jin, W.-Q. & Wang, L.-X. (2005b). Study on perceptual evaluation of fused image quality for color night vision. *Jurnal of Infrared and Millimeter Waves*, Vol.24, No.3, 236-240.
- Sinai, M.J.; McCarley, J.S. & Krebs, W.K. (1999a). Scene recognition with infra-red, low-light, and sensor fused imagery, In: *Proceedings of the IRIS Specialty Groups on Passive Sensors*, pp. 1-9, IRIS, Monterey, CA.
- Sinai, M.J.; McCarley, J.S.; Krebs, W.K. & Essock, E.A. (1999b). Psychophysical comparisons of single- and dual-band fused imagery, In: *Enhanced and Synthetic Vision 1999*, Verly, J.G. (Eds.), Vol. SPIE-3691, pp. 176-183, The International Society for Optical Engineering, Bellingham, WA.
- Smith, M.I.; Ball, A.N. & Hooper, D. (2002). Real-time image fusion: a vision aid for helicopter pilotage, In: *Real-Time Imaging VI*, Kehtarnavaz, N. (Eds.), Vol. SPIE-4666, pp. 83-94, The International Society for Optical Engineering, Bellingham, WA., USA.
- Smith, M.I. & Heather, J.P. (2005). Review of image fusion technology in 2005, In: *Thermosense XXVII*, Peacock, G.R. et al. (Eds.), Vol. SPIE-5782, pp. 6-1-6-17, The International Society for Optical Engineering, Bellingham, WA.
- Spence, I., Wong, P., Rusan, M. & Rastegar, N. (2006). How color enhances visual memory for natural scenes. *Psychological Science*, Vol.17, No.1, 1-6.

- Sun, S., Jing, Z., Li, Z. & Liu, G. (2005). Color fusion of SAR and FLIR images using a natural color transfer technique. *Chinese Optics Letters*, Vol.3, No.4, 202-204.
- Toet, A. (1990a). Adaptive multi-scale contrast enhancement through non-linear pyramid recombination. *Pattern Recognition Letters*, Vol.11, No.11, 735-742.
- Toet, A. (1990b). Hierarchical image fusion. *Machine Vision and Applications*, Vol.3, No.1, 1-11.
- Toet, A. (1992). Multi-scale contrast enhancement with applications to image fusion. *Optical Engineering*, Vol.31, No.5, 1026-1031.
- Toet, A. (2003). Natural colour mapping for multiband nightvision imagery. *Information Fusion*, Vol.4, No.3, 155-166.
- Toet, A. & Franken, E.M. (2003). Perceptual evaluation of different image fusion schemes. *Displays*, Vol.24, No.1, 25-37.
- Toet, A. & Hogervorst, M.A. (2003). Performance comparison of different graylevel image fusion schemes through a universal image quality index, In: *Signal Processing, Sensor Fusion, and Target Recognition XII*, Kadar, I. (Eds.), Vol. SPIE-5096, pp. 552-561, The International Society for Optical Engineering, Bellingham, WA., USA.
- Toet, A. & Ijspeert, J.K. (2001). Perceptual evaluation of different image fusion schemes, In: *Signal Processing, Sensor Fusion, and Target Recognition X*, Kadar, I. (Eds.), Vol. SPIE-4380, pp. 436-441, The International Society for Optical Engineering, Bellingham, WA.
- Toet, A., Ijspeert, J.K., Waxman, A.M. & Aguilar, M. (1997b). Fusion of visible and thermal imagery improves situational awareness. *Displays*, Vol.18, No.2, 85-95.
- Toet, A.; Ijspeert, J.K.; Waxman, A.M. & Aguilar, M. (1997a). Fusion of visible and thermal imagery improves situational awareness, In: *Enhanced and Synthetic Vision 1997*, Verly, J.G. (Eds.), Vol. SPIE-3088, pp. 177-188, International Society for Optical Engineering, Bellingham, WA, USA.
- Toet, A., van Ruyven, J.J. & Valetton, J.M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering*, Vol.28, No.7, 789-792.
- Tsagiris, V. & Anastassopoulos, V. (2004). Information measure for assessing pixel-level fusion methods, In: *Image and Signal Processing for Remote Sensing X*, Bruzzone, L. (Eds.), Vol. SPIE-5573, pp. 64-71, The International Society for Optical Engineering, Bellingham, WA.
- Tsagiris, V. & Anastassopoulos, V. (2005). Fusion of visible and infrared imagery for night color vision. *Displays*, Vol.26, No.4-5, 191-196.
- Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, Vol.11, No.2, 58-64.
- Ulug, M.E. & Claire, L. (2000). A quantitative metric for comparison of night vision fusion algorithms, In: *Sensor Fusion: Architectures, Algorithms, and Applications IV*, Dasarathy, B.V. (Eds.), Vol. SPIE-4051, pp. 80-88, The International Society for Optical Engineering, Bellingham, WA.
- van Rijsbergen, C.J. (1979). *Information retrieval. 2nd Edition*. Butterworth-Heinemann, ISBN , Newton, MA, USA.
- Vargo, J.T. (1999). *Evaluation of operator performance using true color and artificial color in natural scene perception* (Report AD-A363036). Monterey, CA: Naval Postgraduate School.
- Vogel, J. & Schiele, B. (2007). Semantic modeling of natural scenes for content-based image retrieval. *International Journal of Computer Vision*, Vol.72, No.2, 133-157.

- Walls, G.L. (2006). *The vertebrate eye and its adaptive radiation*. Cranbrook Institute of Science, ISBN , Bloomfield Hills, Michigan.
- Wang, L.; Jin, W.; Gao, Z. & Liu, G. (2002). Color fusion schemes for low-light CCD and infrared images of different properties, In: *Electronic Imaging and Multimedia Technology III*, Zhou, L. et al. (Eds.), Vol. SPIE-4925, pp. 459-466, The International Society for Optical Engineering, Bellingham, WA.
- Wang, Q. & Shen, Y. (2006). Performance assessment of image fusion, In: *Advances in Image and Video Technology*, Vol. Lecture Notes in Computer Science Volume 4319, pp. 373-382, Springer Verlag, Heidelberg/Berlin, Germany.
- Warren, P., Howard, J.G., Waterman, J., Scribner, D.A. & Schuler, J. (1999). *Real-time, PC-based color fusion displays* (Report A073093). Washington, DC: Naval Research Lab.
- Waxman, A.M.; Aguilar, M.; Baxter, R.A.; Fay, D.A.; Ireland, D.B.; Racamoto, J.P. & Ross, W.D. (1998). Opponent-color fusion of multi-sensor imagery: visible, IR and SAR, In: *Proceedings of the 1998 Conference of the IRIS Specialty Group on Passive Sensors*, Vol. I, pp. 43-61.
- Waxman, A.M., et al. (1999). Solid-state color night vision: fusion of low-light visible and thermal infrared imagery. *MIT Lincoln Laboratory Journal*, Vol.11, No., 41-60.
- Waxman, A.M.; Carrick, J.E.; Fay, D.A.; Racamoto, J.P.; Augilar, M. & Savoye, E.D. (1996a). 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, The International Society for Optical Engineering, Bellingham, WA.
- Waxman, A.M.; Fay, D.A.; Gove, A.N.; Seibert, M.C.; Racamoto, J.P.; Carrick, J.E. & Savoye, E.D. (1995). Color night vision: fusion of intensified visible and thermal IR imagery, In: *Synthetic Vision for Vehicle Guidance and Control*, Verly, J.G. (Eds.), Vol. SPIE-2463, pp. 58-68, The International Society for Optical Engineering, Bellingham, WA.
- Waxman, A.M.; Fay, D.A.; Hardi, P.; Savoye, D.; Biehl, R. & Grau, D. (2006). Sensor Fused Night Vision : Assessing Image Quality in the Lab and in the Field, In: *Special Session on Image Fusion Assessment. Proceedings of the 9th International Conference on Information Fusion*, Nikolov, S. & Toet, A. (Eds.), International Society of Information Fusion, Mountain View, CA.
- Waxman, A.M.; Fay, D.A.; Ivey, R.T. & Bomberger, N. (2003). Multisensor image fusion & mining: from neural systems to COTS software, In: *Proceedings of the International Conference on Integration of Knowledge Intensive Multi-Agent Systems 2003*, pp. 355-362, IEEE Press, Washington, MA.
- Waxman, A.M.; Gove, A.N. & Cunningham, R.K. (1996b). Opponent-color visual processing applied to multispectral infrared imagery, In: *Proceedings of 1996 Meeting of the IRIS Specialty Group on Passive Sensors*, Vol. II, pp. 247-262, Infrared Information Analysis Center, ERIM, Ann Arbor, US.
- Waxman, A.M., Gove, A.N., Fay, D.A., Racamoto, J.P., Carrick, J.E., Seibert, M.C. & Savoye, E.D. (1997). Color night vision: opponent processing in the fusion of visible and IR imagery. *Neural Networks*, Vol.10, No.1, 1-6.
- Waxman, A.M.; Gove, A.N.; Seibert, M.C.; Fay, D.A.; Carrick, J.E.; Racamoto, J.P.; Savoye, E.D.; Burke, B.E.; Reich, R.K. et al. (1996c). Progress on color night vision: visible/IR fusion, perception and search, and low-light CCD imaging, In: *Enhanced and*

- Synthetic Vision 1996*, Verly, J.G. (Eds.), Vol. SPIE-2736, pp. 96-107, The International Society for Optical Engineering, Bellingham, WA.
- White, B.L. (1998). *Evaluation of the impact of multispectral image fusion on human performance in global scene processing*. (M.Sc.) Monterey, CA: Naval Postgraduate School.
- Wichmann, F.A., Sharpe, L.T. & Gegenfurtner, K.R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol.28, No.3, 509-520.
- Xydeas, C.S. & Petrovic, V.S. (2000). Objective pixel-level image fusion performance measure, In: *Sensor Fusion: Architectures, Algorithms, and Applications IV*, Dasarathy, B.V. (Eds.), Vol. SPIE-4051, pp. 89-98, The International Society for Optical Engineering, Bellingham, WA.
- Yang, C., Zhang, J., Wang, X. & Liu, X. (2007). A novel similarity based quality metric for image fusion. *Information Fusion*, Vol.9, No.2, 156-160.
- Zheng, Y., Essock, E.A., Hansen, B.C. & Haun, A.M. (2007). A new metric based on extended spatial frequency and its application to DWT based fusion algorithms. *Information Fusion*, Vol.8, No.2, 177-192.
- Zheng, Y.; Hansen, B.C.; Haun, A.M. & Essock, E.A. (2005). Coloring night-vision imagery with statistical properties of natural colors by using image segmentation and histogram matching, In: *Color imaging X: processing, hardcopy and applications*, Eschbach, R. & Marcu, G.G. (Eds.), Vol. SPIE-5667, pp. 107-117, The International Society for Optical Engineering, Bellingham, WA.
- Zhu, X. & Jia, Y. (2005). A method based on IHS cylindrical transform model for quality assessment of image fusion, In: *MIPPR 2005: Image Analysis Techniques*, Li, D. & Ma, H. (Eds.), Vol. SPIE-6044, pp. 607-615, The International Society for Optical Engineering, Bellingham, MA.
- Zou, X. & Bhanu, B. (2005). Tracking humans using multi-modal fusion, In: *2nd Joint IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (OTCBVS'05)*, pp. W01-30-1-W01-30-8, IEEE Press, Washington, USA.

IntechOpen



## **Image Fusion**

Edited by Osamu Ukimura

ISBN 978-953-307-679-9

Hard cover, 428 pages

**Publisher** InTech

**Published online** 12, January, 2011

**Published in print edition** January, 2011

Image fusion technology has successfully contributed to various fields such as medical diagnosis and navigation, surveillance systems, remote sensing, digital cameras, military applications, computer vision, etc. Image fusion aims to generate a fused single image which contains more precise reliable visualization of the objects than any source image of them. This book presents various recent advances in research and development in the field of image fusion. It has been created through the diligence and creativity of some of the most accomplished experts in various fields.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Alexander Toet (2011). Cognitive Image Fusion and Assessment, Image Fusion, Osamu Ukimura (Ed.), ISBN: 978-953-307-679-9, InTech, Available from: <http://www.intechopen.com/books/image-fusion/cognitive-image-fusion-and-assessment>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen