6

Real-Time Full Color Multiband Night Vision

Alexander Toet and Maarten A. Hogervorst
TNO Human Factors
The Netherlands

1. Introduction

Night vision cameras are widely used for military and law enforcement applications related to surveillance, reconnaissance, intelligence gathering, and security. The two most common night-time imaging systems are low-light-level (e.g., image-intensified) cameras, which amplify the reflected visible to near infrared (VNIR) light, and thermal infrared (IR) cameras, which convert thermal energy from the midwave (3 to 5 microns) or the long wave (8 to 12 microns) part of the spectrum into a visible image. These systems create images with a single (one-dimensional) output per pixel. As a result their ability to discriminate different materials is limited. This can be improved by combining systems that are sensitive to different parts of the electromagnetic spectrum, resulting in multiband or hyperspectral imagers. The number of different outputs increases dramatically by combining multiple sensors (e.g. up to N^2 for two sensors, when the number of different outputs for each sensor is N), which in turn leads to a significant increase in the number of materials that can be discriminated. The combination of multiple bands allows for meaningful color representation of the system output. It is therefore not surprising that the increasing availability of fused and multiband infrared and visual nightvision systems (e.g. Bandara et al., 2003; Breiter et al., 2002; Cho et al., 2003; Cohen et al., 2005; Goldberg et al., 2003) has led to a growing interest in the (false) color display of night vision imagery (Li & Wang, 2007; Shi et al., 2005a; Shi et al., 2005b; Tsagaris & Anastasopoulos, 2006; Zheng et al., 2005).
In principle, color imagery has several benefits over monochrome imagery for surveillance, reconnaissance, and security applications. The human eye can only distinguish about 100 shades of gray at any instant. As a result, grayscale nightvision images are sometimes hard to interpret and may give rise to visual illusions and loss of situational awareness. Since people can discriminate several thousands of colors defined by varying hue, saturation, and brightness, a false color representation may facilitate nightvision image recognition and interpretation. For instance, color may improve feature contrast, thus enabling better scene segmentation and object detection (Walls, 2006). This may allow an observer to construct a more complete mental representation of the perceived scene, resulting in better situational awareness. It has indeed been found that scene understanding and recognition, reaction time, and object identification are 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 false color nightvision 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 bands exceeds three (e.g. with hyperspectral imagers; Jacobson et al., 2007). In practice, many tasks may benefit from a representation that renders a nighttime scene in daytime colors. Jacobson & Gupta (Jacobson et al., 2007; Jacobson & Gupta, 2005) therefore advise to use a consistent color mapping according to a natural palette. The use of natural colors facilitates 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 generally produce 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 & Süsstrunk, 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 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 & 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 an intuitively meaningful ("naturalistic") and stable color appearance, 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., 1995a; Waxman et al., 1997). 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. In the next section we give an overview of some recently developed color mapping schemes that can give false color multiband nightvision imagery a natural color appearance. First we present a simple false color mapping scheme that is inspired by previous color opponent processing schemes. Although this scheme produces fused false color images with large color contrast and preserves the identity of the input signals (thus making the images perceptually intuitive), the resulting color representation is not strictly natural looking (Toet & Walraven, 1996). We therefore developed a statistical extension of this coloring method which produces colorized multiband nightvision imagery with a regular daylight color appearance (Toet, 2003). This mapping transfers the first order statistics of the color distribution of a given color reference image to the multiband nighttime images, thereby giving them a similar color appearance as the reference image. In its original form this method is computationally expensive. However, computational simplicity (enabling real-time implementation) can be achieved by applying the statistical mapping approach in a lookup-table framework. Although the statistical mapping approach yields a natural color rendering, it achieves no color constancy, since the mapping depends on the relative amounts of the different materials in the scene (and will therefore change when the camera pans over or zooms in on a scene). We therefore developed a sample-based color mapping scheme that yields both color constancy and computational efficiency (Hogervorst & Toet, 2008a; Hogervorst & Toet, 2008b; Hogervorst & Toet, 2010). In contrast to the statistical color mapping method, the sample based color transfer method (for which a patent application is currently pending: Hogevorst et al., 2006) is highly specific for different types of materials in the scene and can easily be adapted for the task at hand, such as the detection of camouflaged objects. After explaining how the sample based color transformation can be derived from the combination of a given multi-band sensor image and a corresponding daytime reference image, we will dicuss how it can be deployed at night and implemented in real-time.

2. Color mapping

2.1 Center-surround opponent-color fusion

Opponent color image fusion was originally developed at the MIT Lincoln Laboratory (Gove et al., 1996; Waxman et al., 1995a; Waxman et al., 1996a; Waxman et al., 1996b; Waxman et al., 1997; Waxman et al., 1999) and derives from biological models of color vision (Schiller, 1982; Schiller, 1984; Schiller et al., 1986; Schiller, 1992) and fusion of visible light and infrared (IR) radiation (Newman & Hartline, 1981; Newman & Hartline, 1982).

In the case of color vision in monkeys and man, retinal cone sensitivities are broad and overlapping, but the images are contrast enhanced *within* bands by spatial opponent processing (via cone-horizontal-bipolar cell interactions) creating both ON and OFF center-surround response channels (Schiller, 1992). These signals are then contrast enhanced *between* bands via interactions among bipolar, sustained amacrine, and single-opponent-color ganglion cells (Gouras, 1991; Schiller & Logothetis, 1990), all within the retina. Further

color processing in the form of double-\-opponent-color cells is found in the primary visual cortex of primates (and the retinas of some fish). Opponent processing interactions form the basis of such percepts as color opponency, color constancy, and color contrast, though the exact mechanisms are not fully understood. Double-opponent-color processing has been applied to multispectral IR target enhancement (Gove et al., 1996; Waxman et al., 1995b).
Fusion of visible and thermal imagery has been observed in several classes of neurons in the optic tectum (evolutionary progenitor of the superior colliculus) of rattlesnakes (pit vipers), and pythons (boid snakes), as described by (Newman & Hartline, 1981; Newman & Hartline, 1982). These neurons display interactions in which one modality (e.g. IR) can enhance or depress the response to the other sensing modality (e.g. visible) in a strongly nonlinear fashion. Such interactions resemble opponent-processing between bands as observed in primate retina.
For opaque surfaces in thermodynamic equilibrium, spectral reflectivity ρ and emissivity ε are linearly related at each wavelength λ: $\rho(\lambda) = 1 - \varepsilon(\lambda)$. This provides a rationale for the use of both on-center and off-center channels when treating infrared imagery as characterized by thermal emissivity (Toet et al., 1997).
In the opponent-color image fusion methodology the individual input images are first enhanced by filtering them with a feedforward center-surround shunting neural network (Grossberg, 1988). This operation serves

1. to enhance spatial contrast in the individual visible and IR bands,
2. to create both positive and negative polarity IR contrast images, and
3. to create two types of single-opponent-color contrast images.

The resulting single-opponent-color contrast images represent grayscale fused images that are analogous to the IR-depressed visual and IR-enhanced visual cells of the rattlesnake (Newman & Hartline, 1981; Newman & Hartline, 1982).

2.2 Pixel based opponent-color fusion

Inspired by the opponent-color fusion approach (Waxman et al., 1995a; Waxman et al., 1996a; Waxman et al., 1996b; Waxman et al., 1997; Waxman et al., 1999), we derived a simplified (pixel based) version of this method, which fuses visible and thermal images into false color images with a relatively natural or intuitive appearance.
Let I_1 and I_2 be two input images with the same spatial resolution and dynamic range. The common component of both signals is computed as the morphological intersection:

$$I_1 \cap I_2 = \mathrm{Min}\{I_1, I_2\} \tag{1}$$

The unique or characteristic component I^* of each image modality remains after subtraction of the common component:

$$I_1^* = I_1 - I_1 \cap I_2 \quad ; \quad I_2^* = I_2 - I_1 \cap I_2 \tag{2}$$

The characteristic components are emphasized in the fused image by subtracting them from the opposite image modalities. The color fused image is then obtained by mapping these differences to respectively the red and green bands of a RGB false color image. The characteristic components of both image modalities can be further emphasized by mapping their difference to the blue band of the fused false color image, so that the final mapping is given by (Toet & Walraven, 1996):

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} I_2 - I_1^* \\ I_1 - I_2^* \\ I_2^* - I_1^* \end{pmatrix} \tag{3}$$

In case of visual and thermal input images, I_1 = Vis and I_2 = *IR*. Because the method is computationally simple it can implemented in hardware or even be applied in real-time using standard processing equipment (Aguilar et al., 1998; Aguilar et al., 1999; Waxman et al., 1999). The resulting color rendering enhances the visibility of certain details and preserves the specificity of the sensor information. In addition, it has a fairly natural color appearance (Fig. 1 and Fig. 2). The resulting images agree with our natural associations of warm (red) and cool (blue). To further enhance the appearance of the fused results, the R, G, B channels can be input to a color remapping stage in which, following conversion to H, S, V (hue, saturation, value) color space, hues can be remapped to alternative "more natural'" hues, colors can be desaturated, and then reconverted back to R, G, B signals to drive a color display. Because of the enhanced color contrast and its intuitive appearance this color fused image representation is expected to improve both visual target detection and recognition performance are expected to benefit in terms of both speed and precision. Two observer studies were performed to test this hypothesis.

In the first observer study we used grayscale intensified visual and thermal images, and color fused motion sequences, depicting scenes in which a person walked across a rural scene with man-made objects (Toet et al., 1997). The reference (terrain) features were represented with high contrast in the intensified visual images (Fig. 2a) and low contrast in

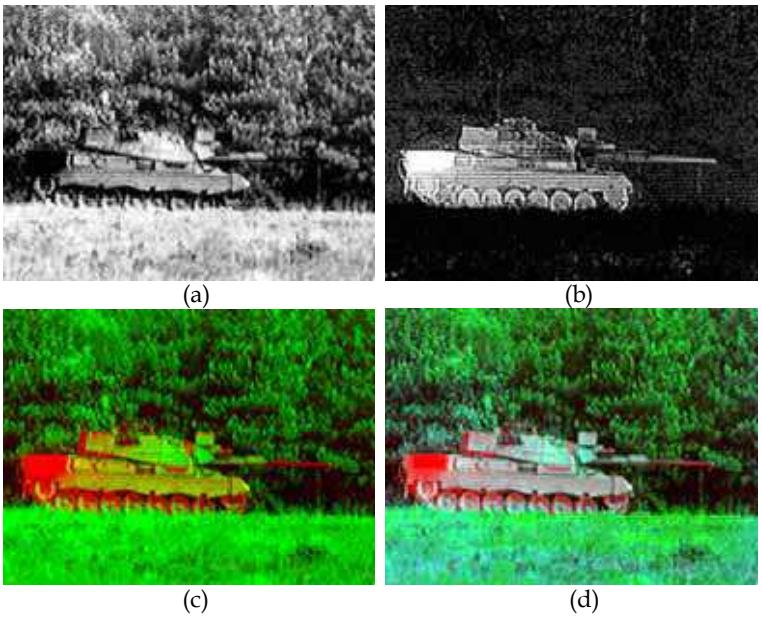
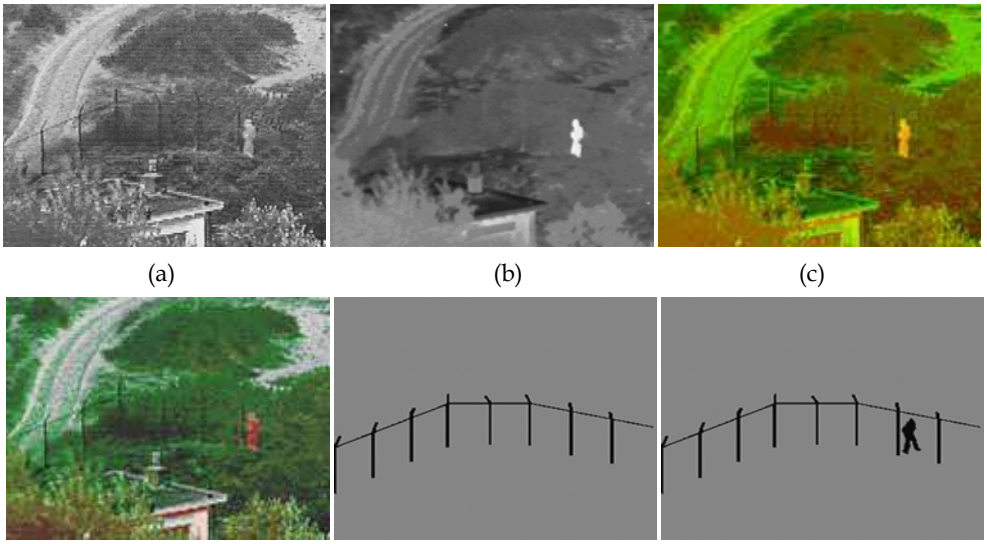


Fig. 1. Visual (a) and thermal (b) input images. (c) Result of mapping (a) and (b) to respectively the green and red channels of an RGB display. (d) Result of the mapping defined by equation.



(a) (b) (c)

(d) (e) (f)

Fig. 2. Scene representing a person walking along the outside of a fence. Visual (a) and thermal (b) input images. (c) Result of mapping (a) and (b) to respectively the green and red channels of an RGB display. (d) Result of the mapping defined in equation . (e) Reference image used in the spatial localization task. (f) Image used to assess baseline localization performance.

the thermal images (Fig. 2b), while the opposite was the case for the image of the person. All details were represented in the color fused images (Fig. 2d). During the localization experiments, individual frames from the motion sequences and for each of the three image modalities (visual, thermal and color fused) were briefly (1 s) and in random order presented to human observers. After the presentation of each frame a schematic grayscale image was shown representing only the reference features on a homogeneous background (e.g. Fig. 2e). Observers were asked to indicate the perceived position of the person in the scene by placing a mouse controlled cursor at the corresponding location in the schematic reference image. The position of the reference image on the display screen was given a small random variation to prevent participants from using cues from afterimages. Baseline performance was assessed using schematic images, similar to the reference image, but with a binary image of the person at his actual location in the corresponding frames (e.g. Fig. 2f). The results show that observers can localize a person in a scene with a significantly higher accuracy and with greater confidence when they perform with color fused images, compared to the individual image modalities (visible and thermal; Toet et al., 1997).

In the second observer study we used grayscale visual and thermal (8-12 μm) motion sequences, and color fused motion sequences, depicting a mountain range in the background and grasslands in the foreground, with infantry soldiers, vehicles, and a smoke screen (Fig. 3). The visual and thermal motion sequences are a subset (images 37--93) of the MS01 Test Sequence that consists of 110 corresponding image pairs, registered at CFB Valcartier (Sévigny, 1996). Both a tow truck and a helicopter move across the scene during the registration period. The infantry soldiers are not visible in the visual images (Fig. 3a-c),

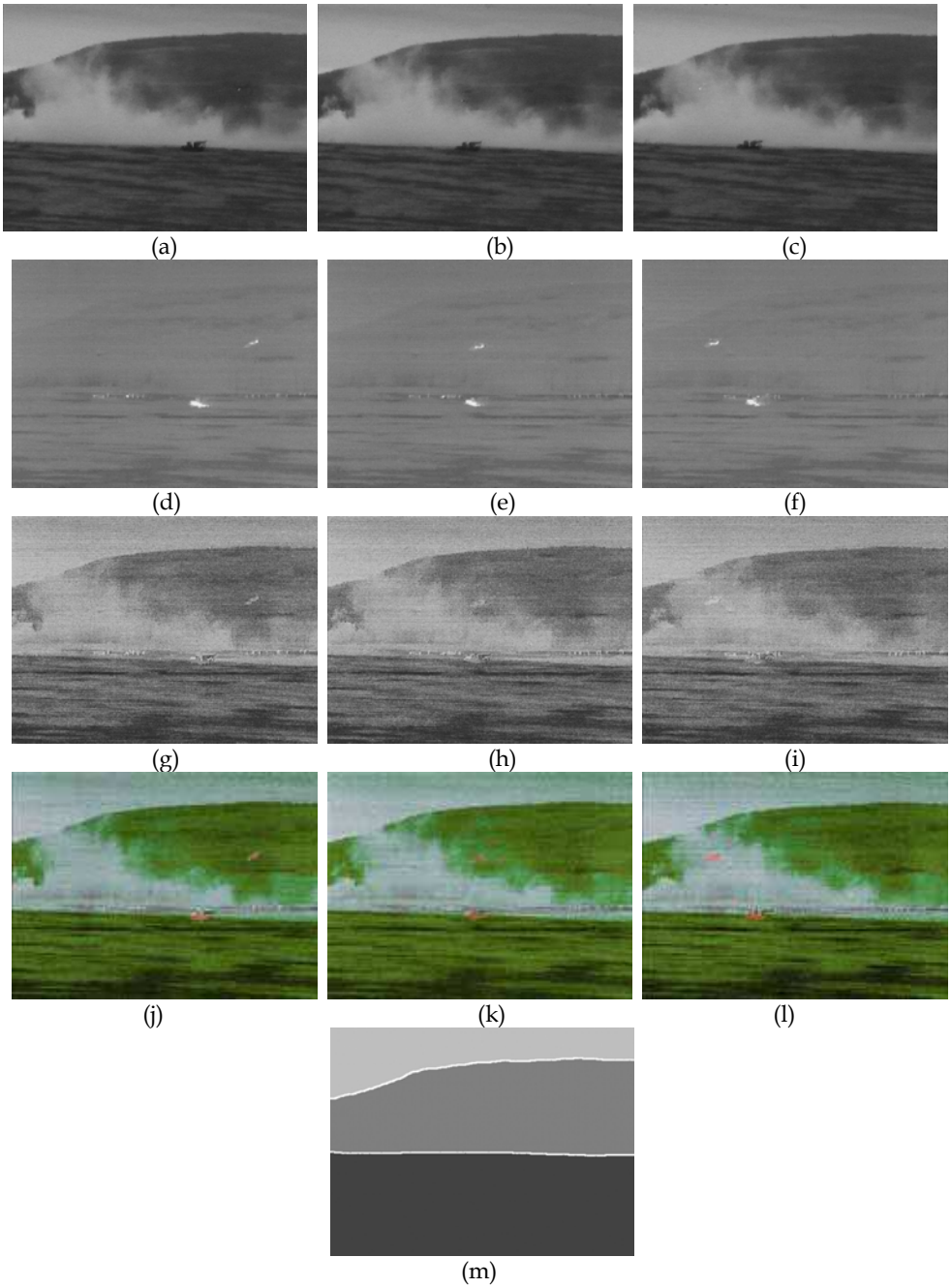


Fig. 3. Left to right: successive frames of a time sequence. Top-down: video (a-c), thermal (d-f), grayscale fused (g-i), and color fused (j-l) images. (m) Schematic reference image.

because they are obscured by the smoke screen. However, they can easily be perceived in the thermal images (Fig. 3d-f), where they appear as small hot spots. In the visual images, the visibility of the helicopter ranges from barely visible to not visible (when it flies behind the smoke screen: Fig. 3f). In the thermal images there is almost no contrast between the foreground (grassland) and the background (the mountain range). Also, the mountain range and the sky have little contrast in the thermal images (the skyline of the mountain range is barely visible). The contrast between the sky and the top of the mountain range is much larger in the CCD images. The visual and thermal images were color fused using equation 3 (see Fig. 3j-l). Fig. 3g-i shows the luminance component of Fig. 3j-l. Both the graylevel fused (Fig. 3g-i) and the color fused (Fig. 3j-l) images represent all details of interest. However, in the grayfused images, where details are represented by different shades of gray, it is sometimes hard to visually segment the scene, because there are no visible boundaries (edges) between different physical objects with the same mean luminance. For instance, in the visual images the smokescreen has a high luminance (is very bright). In the thermal images, the warmer (barren) parts of the grassland and the helicopter are represented as bright regions. As a result, there is sometimes very little contrast in the grayfused images between the smokescreen and respectively the helicopter and the grassland. In the color fused images, the additional color contrast leads to an effortless perceptual segmentation of the scene. We measured the accuracy with which observers can determine the position of the helicopter in a briefly (600 ms) presented motion sequence, for visual, thermal, and color fused image sequences. Only a limited portion (a restricted field of view) of the entire scene was displayed during each test interval. The field of view was randomly positioned on the dynamic battlefield scene. This experimental paradigm simulates a field of view search of the display of a moving camera scanning over a larger field of regard. For each individual frame a corresponding reference image was constructed, representing a segmented version of the background of the original scene (mountain range, grassland, and sky; see Fig. 3m). After watching each movie fragment, observers were asked to indicate the location where the helicopter was last seen, by placing a mouse-controlled cursor over the schematic reference image. This task requires observers to quickly determine *(a)* the location of reference contours, and *(b)* the location of the target at each stimulus presentation. This involves rapid visual scene segmentation. The performance in this relative spatial localization task depends on the accuracy with which the position of the helicopter can be perceived relative to the contours of the mountain range. The results of this experiment show that the accuracy with which observers can determine the position of a helicopter in a briefly presented and randomly positioned window on a dynamic battlefield scene is significantly higher for color fused images than for the individual visual and thermal images (Toet et al., 1997). The color fused images probably represent all relevant features at a sufficiently large perceptual contrast to allow rapid visual identification of the spatial layout of the scene, thereby enabling subjects to perform the task. We also observed that a restriction of the field of view results in a significant increase in the localization error for the visual and thermal image modalities, but not for the fused image modalities.

The false color mapping defined by equation 3 has also been successfully applied in other domains, like the fusion of retinal images (Kolar et al., 2008; Laliberté et al., 2002; Laliberté et al., 2003) and the fusion of infrared and synthetic images (Simard et al., 1999; Simard et al., 2000).

In ophthalmology, visual fundus images are often used in combination with fluorescein angiogram images. Visual images of the retina clearly represent hard exudates. Fluorescein

angiogram images represent the macula, the arteries and veins at high contrast, thus allowing the detection of occluded and leaking capillaries, microaneurims, macular edema, and neovascularization. Using the mapping defined by equation 3 to fuse visual fundus images with fluorescein angiogram images provides better color contrast rendering than other opponent-color fusion methods, thus enhancing diagnostic performance and reducing visual workload (Laliberté et al., 2002; Laliberté et al., 2003). It was for instance found that this mapping clearly represents neovessels and depicts the macula at high contrast (Laliberté & Gagnon, 2006) Fig. 4. shows two examples of the fusion of grayscale visual fundus images (Fig. 4a and d) with corresponding fluorescein angiogram images (Fig. 4b and e). The fused images (Fig. 4c and f) represent the interesting details like the vascular network (purple veins) and the exudates (yellow lesions) with large color contrast. When using equation 3 to fuse thermal and autofluorescent images of the retina (Kolar et al., 2008), the resulting false color images provide higher contrast for the hyperfluent areas of the autofluorescent images (which are symptoms for glaucoma in its early stages) and clearly represent the position of the optic nerve head from the infrared image.
Simulated flight tests with fused infrared and synthetic imagery showed that the fusion technique defined by equation 3 preserves all features relevant for obstacle avoidance, and significantly improves detection distances for all simulated visibility conditions (Simard et al., 1999; Simard et al., 2000).

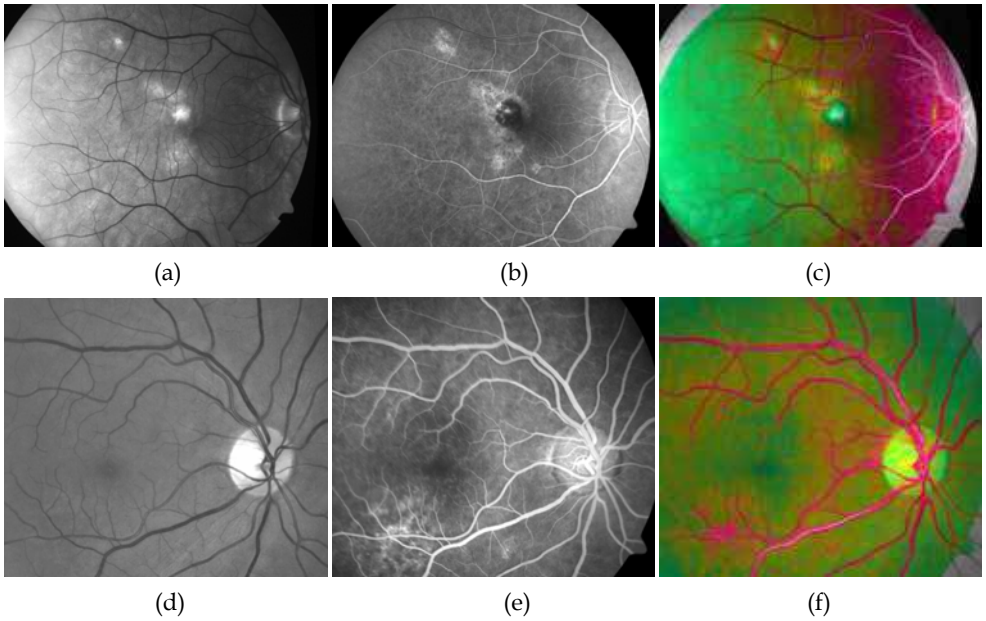


Fig. 4. Photographs (a,d) and fluorescein angiogram images (b,e) and their color fused representation (c,f).

2.3 Statistical color transform

Although the overall color appearance of images produced with the opponent-color fusion scheme is fairly intuitive, some details may still be depicted with unnatural colors. In this

section we present a method that gives multiband nighttime images the appearance of regular daylight color images by transferring the first order color statistics from full color daylight imagery to the false color multiband nightvision imagery. The method is based on a technique that was developed to enhance the color representation of synthetic imagery (Reinhard et al., 2001). The outline of the method is as follows. As input the method requires a false color RGB image. This can be produced by mapping the 2 or 3 individual bands (or the first 2 or 3 principal components when the sensor system delivers more than 3 bands) of a multiband nightvision system to the respective channels of an RGB image. Next, the false color RGB nightvision image and a regular full color daylight reference image are both transformed into the perceptually decorrelated *lαβ* opponent color space (Ruderman et al., 1998). Then, the mean and standard deviation of each of the 3 color channels of the multiband nightvision image are set equal to those of the reference image:

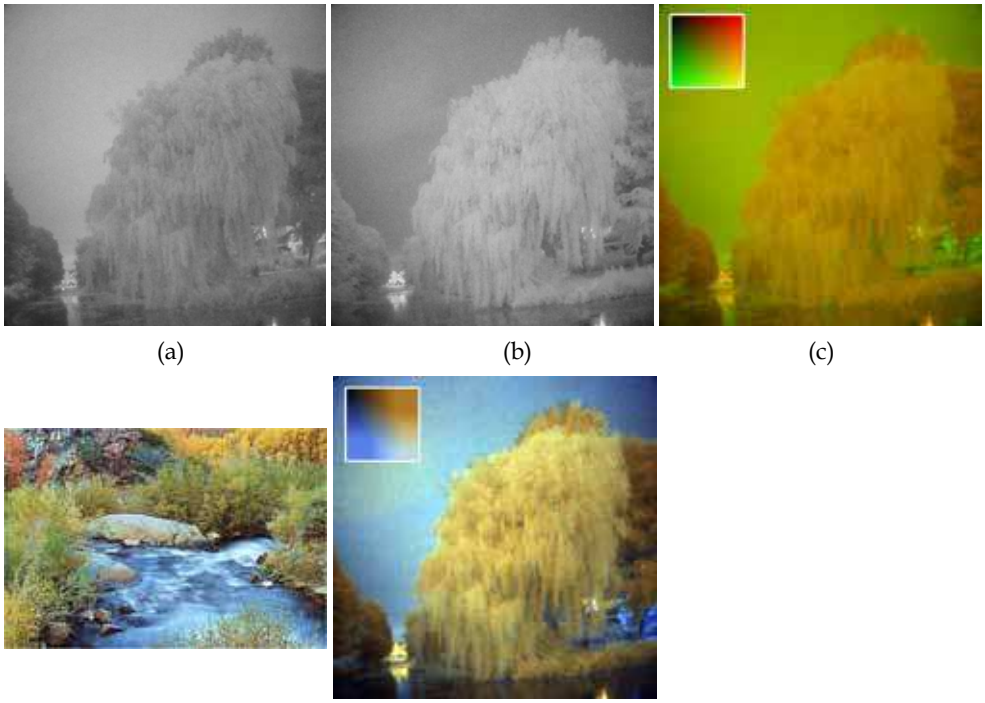
$$\begin{aligned} l^* &= \frac{\sigma_r^l}{\sigma^l}\left(l - \langle l \rangle\right) \\ \alpha^* &= \frac{\sigma_r^\alpha}{\sigma^\alpha}\left(\alpha - \langle \alpha \rangle\right) \\ \beta^* &= \frac{\sigma_r^\beta}{\sigma^\beta}\left(\beta - \langle \beta \rangle\right) \end{aligned} \tag{4}$$

where $\langle\ \rangle$ denotes the mean, σ the standard deviation, and the index *r* refers to the reference image.

Finally, the multiband nightvision image is transformed back to RGB space for display. The result is a full color representation of the multiband nightvision image with a color appearance that closely resembles the color appearance of the daylight reference image. The daylight reference image should display a scene which is similar (but not necessarily identical to) the one displayed by the multiband nightvision image. The order of the mapping is irrelevant, since the following procedure effectively rotates the color coordinate axes of the false color multiband nightvision images such that these will be aligned with the axes of the referenced daylight color image in the final result.

The statistical color transform is computationally expensive and therefore not suitable for real-time implementation. Moreover, although it can give multiband nighttime imagery a natural daylight color appearance, it can not achieve color constancy for dynamic imagery (Zheng & Essock, 2008), because the actual mapping depends on the relative amounts of different materials in (i.e. the composition or statistics of) the scene. Large objects in the scene will dominate the color mapping. As a result, the color of objects and materials may change over time when the sensor system pans over (or zooms in on) a given scene. We therefore developed a fixed lookup table based version of this statistical color mapping which is *(1)* computationally efficient, so that it can easily be deployed in real time, and which *(2)* yields constant object colors.

The new lookup table based statistical color transfer approach is illustrated in Fig. 5. for a multi-band image consisting of two channels. First, the two sensor images are mapped on two of the three channels of an RGB image. In this particular example (Fig. 5c) the visual band is (arbitrarily) mapped to R (red channel) and the near-infrared band is mapped to G (green channel). The result is the red-green false-color representation of the multi-band



(a) (b) (c)

(d) (e)

Fig. 5. Visible (a) and near infrared (b) image. (c) False color image obtained by assigning (a) to the green and (b) to the red channel of an RGB false color image (the blue channel is set to zero). The inset in this figure represents all possible combinations of signal values that can occur in the two sensor bands (upper left: both sensors give zero output; lower right: both sensors give maximal output). (d) Arbitrary reference daylight color image. (e) Result of mapping the first order statistics of the reference image (d) to the false color nighttime image (c). The inset in this figure represents the result of applying the same transform to the inset of (c), and shows all possible colors that can occur in the recolored sensor image (i.e. after applying the color mapping).

image shown in Fig. 5c. The statistical color transform can then be derived from the first order statistics of respectively *(a)* Fig. 5c and *(b)* a given daylight color reference image, like the one shown in square inset in Fig. 5d. The application of this statistical color transform to an input table of 2-tuples representing all possible sensor output values yields an output table containing all possible color values that can occur in the colorized nighttime image. The in- and output table pair defines the statistical color mapping and can therefore be deployed in a color lookup table transform procedure. The square inset in Fig. 5c represents the table of all possible two-band signal values as different shades of red, green and yellow. Application of the statistical color transform to the inset in Fig. 5c yields the inset shown in Fig. 5e. In a lookup table paradigm the insets in Fig. 5c and Fig. 5e together define the statistical color mapping. For any color in the false-color representation of Fig. 5c the corresponding color after application of the statistical color transform can easily be found by

(1) locating the color in the inset of Fig. 5c, and *(2)* finding the color at the corresponding location in the transformed inset in Fig. 5e. For instance, a pixel representing a high response in the visual channel and low response in the near infrared channel is represented with a red color (high value of red, low value of green) in the inset in Fig. 5c. In the inset of Fig. 5e the same pixel appears in a brownish color. The color transformation can thus be implemented by using the inset pair of Fig. 5c and Fig. 5e as color lookup tables. Then, the false color image Fig. 5c can be transformed into an indexed image using the red-green color lookup table (the inset of Fig. 5c). Replacing the color lookup table of the indexed image Fig. 5c by the transformed color lookup table (the inset of Fig. 5e) then transforms Fig. 5c into Fig. 5e. Note that the color mapping scheme is fully defined by the two color lookup tables. When all possible combinations of an 8-bit multi-band system are represented, these color lookup tables contain 256x256 entries. When a color lookup table with less entries is used (e.g. only 256), the color mapping can be achieved by determining the closest match of the table entries to the observed multi-band sensor values.

Once the color mapping has been derived from a multi-band nighttime image and its corresponding reference image, and once it has been defined as a lookup table transform, it can be applied to different and dynamic multi-band images. The advantage of this method is that the color of objects only depends on the multi-band sensor values and is independent of the image content. Therefore, objects keep the same color over time when registered with a moving camera. Another advantage of this implementation is that it requires minimal computing power. Once the color transformation has been derived and the pair of color lookup tables that defines the mapping has been created, the new color lookup table transform can be used in a (real-time) application.

2.4 Sample based color transform

In spite of all the afore-mentioned advantages of the lookup table based statistical color transform, there is still room for improvement. For instance, in this paradigm there is no strict relationship between sensor output values and object color, since the statistical approach inherently only addresses the global color characteristics of the depicted scene. In this section we will describe an alternative lookup table based method for applying natural colors to multi-band images, which alleviates this problem since it does not rely on image statistics. The color transformation is derived from a corresponding set of samples for which both the multi-band sensor values and the corresponding natural color (RGB-value) are known (Hogervorst et al., 2006). We will show that this method results in rendered multi-band images with colors that match the daytime colors more closely than the result of the statistical approach. In contrast to the statistical method, the derivation of the color mapping requires a registered image pair, consisting of a multi-band image and a daytime reference image of the same scene, sine the pixels serve as samples in this approach. Once the color mapping has been derived it can be applied to different multi-band nighttime images. Again, we will implement the color transformation using a color lookup table transformation, thus enabling real-time implementation.

The method is as follows. Given a set of samples (pixels) for which both the multi-band sensor output and the corresponding daytime colors are known, the problem of deriving the optimal color transformation is to find a transformation that optimally maps the N-dimensional (e.g. in our examples $N = 2$) multi-band sensor output vectors (one for each sample) to the 3-D vectors corresponding to the daytime colors (RGB). The mapping should

minimize the difference between the modeled colors and the measured colors. Moreover, the transformation should predict the mapping of untrained samples. Several methods exist to derive a suitable mapping, such as neural networks and support vector machines. What constitutes a suitable mapping is determined by the function that is minimized. Also the statement that the difference between the modeled colors and the measured colors is minimized should be formalized. We will minimize the average perceptual color difference between the modeled color and the measured color. More precisely, we will minimize the average squared distance between the perceptual color vectors $l\alpha\beta$ (see (Ruderman et al., 1998)). We will describe a (relatively) simple implementation that is not focused towards finding the theoretical optimum mapping, but that will lead to robust and good results and can be understood intuitively. We will now describe our new method for deriving a natural color transformation using the example shown in Fig. 6. Fig. 6a depicts the full color daytime reference image, which is in this case a color photograph taken with a standard digital camera. Figs. 6b and c respectively show a visible and near-infrared image of the same scene. Fig. 6f shows the result of applying daytime colors to the two-band night-time sensor image using our new color mapping technique.

The method works as follows. First, the multi-band sensor image is transformed to a false-color image by taking the individual bands (Fig. 6b and c) as input to the R and G channels (and B when the sensor contains three bands), referred to as the RG-image (Fig. 6e). In practice any other combination of 2 channels can also be used (one could just as well use the combinations R & B or B & R). Mapping the two bands to a false color RGB-image allows us to use standard image conversion techniques, such as indexing. In the next step the resulting false color (RG-image) Fig. 6e is converted to an indexed image. Each pixel in such an image contains a single index. The index refers to an RGB-value in a color lookup table (the number of entries can be chosen by the user). In the present example of a sensor image consisting of two bands (R and G; Fig. 6e) the color lookup table contains various combinations of R and G values (the B-values are zero when the sensor or sensor pair provides only two bands). For each index representing a given R,G combination (a given false color) the corresponding natural color equivalent is obtained by locating the pixels in the target image with this index and finding the corresponding pixels in the (natural color) reference image (Fig. 6a). First, the RGB-values are converted to perceptually de-correlated $l\alpha\beta$ values (see (Ruderman et al., 1998)). Next, we calculate the average $l\alpha\beta$-vector over this ensemble of pixels. This assures that the computed average color reflects the perceptual average color. Averaging automatically takes the distribution of the pixels into account: colors that appear more frequently are attributed a greater weight. Let us for instance assume that we would like to derive the natural color associated with index 1. In that case we locate all pixels in the (indexed) false color multi-band target image with index 1. We then take all corresponding pixels in the reference daytime color image, convert them to $l\alpha\beta$, and calculate the average $l\alpha\beta$-value. Next, we transform the resulting average $l\alpha\beta$-value back to RGB. Finally, we assign this RGB-value to index 1 of the new color lookup table. These steps are successively carried out for all indices. This process yields a new color lookup table containing the natural colors associated with the various multi-band combinations in the false color (RG) color lookup table. Replacing the RG-color lookup table (left side of Fig. 6d) by the color lookup table with natural colors (right side of Fig. 6d) yields an image with a natural color appearance, in which the colors are optimized for this particular sample set (Fig. 6d). Note that the red parts in the scene in Fig. 6a do not turn out red again in the

rendered night-time image Fig. 6f. This is due to the fact that other materials which occupy a larger area of the scene (and which therefore dominate the color mapping) give the same sensor output in the two bands. Also, the red flags are not apparent in the visible band (Fig. 6b). This has only a minor effect on the overall appearance of the scene as long as the parts that change between the different band recordings occupy only a relatively small area.

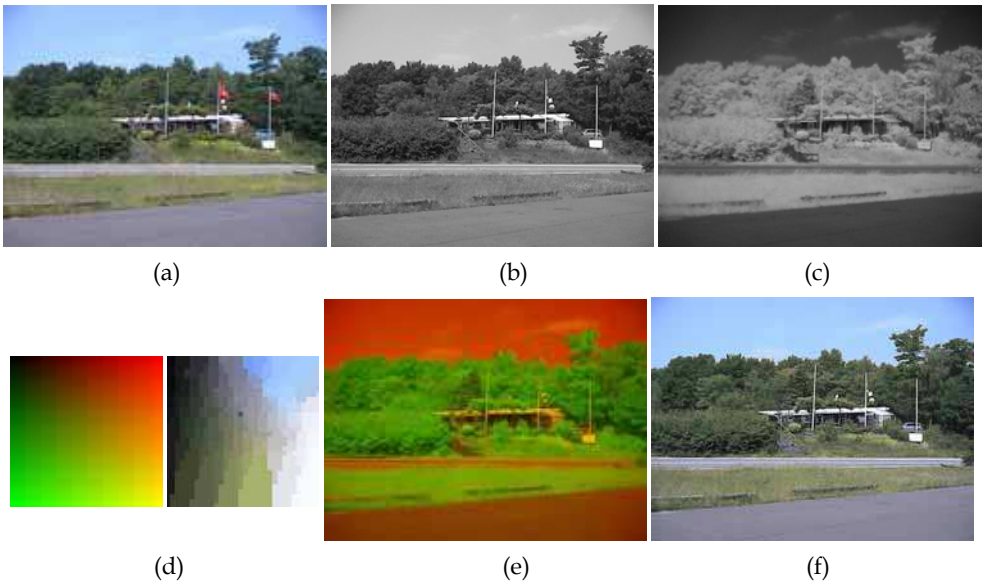


Fig. 6. (a) Natural daylight color reference image. Visible (b) and near-infrared (c) images of the same scene. (d) The color mapping derived from corresponding pixel pairs in a and b-c. (e) Combined RG false color representation of (b) and (c), obtained by assigning (b) to the green and (c) to the red channel of an RGB color image (the blue channel is set to zero). (f) Result of the application of the mapping scheme in (d) to the two band false color image in (e).

Fig. 7 illustrates the difference between the statistical and the sample based color transforms. In this example we determined a color mapping lookup table from a pair of images consisting of (Fig. 7a) the original version of a full color daylight photograph and (Fig. 7b) the same image from which the blue channel has been removed (B=0). Note that the sample-based color remapping (Fig. 7c), using the sample based color lookup table (inset) determined from the image pair Fig. 7a and Fig. 7b, nicely restores most of the blue hues in the scene, while the statistical color remapping procedure (Fig. 7d) is not capable to restore the missing information. This is due to the fact that the sample-based method allows for nonlinear transformations while the statistical method is a linear (affine) transformation (in CIELAB-color space).

Fig. 8a and b show respectively visual and near-infrared images of the same scene. Fig. 8c shows the red-green false color representation of Fig. 8a and b. Fig. 8d shows the daytime reference image corresponding to the multi-band sensor image. Straightforward application of the sample-base color transform results in Fig. 8e. Note that the colors of this figure closely match the daytime colors as shown in Fig. 8d (e.g. the sky is blue). However, the

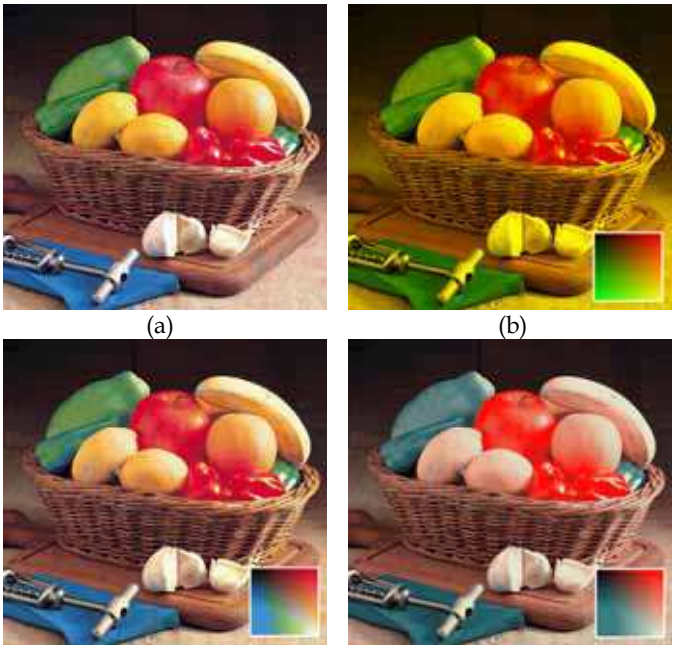


Fig. 7. (a) Full color RGB image. (b) Image from (a) after removal of the blue band (B=0). (c) Result of sample-based color remapping, using the color lookup table (inset) determined from the image pair (a,b). (d) Result of the statistical color remapping procedure.

image looks noisy and certain objects appear in the wrong color (e.g. the bench and parts of the roof). This is due to the fact that the luminance in the colorized image does not increase continuously with increasing sensor output (the luminance value in Fig. 8c). This gives an undesirable "solarizing" effect. We therefore derived from this color map (inset in Fig. 8e) another color map (inset in Fig. 8f) in which the luminance increases linearly with the distance to the top-left corner. Fig. 8f shows the result of this new color mapping. The colors in Fig. 8f closely match the daytime colors. The sky is dark instead of light-blue. This corresponds to the intuition that the sky should look dark at night, and does not affect the situational awareness. Also important is the fact that the color transformation shown in Fig. 8f is smooth, in contrast to the one shown in Fig. 8e. Intuitively a small variation in sensor output should lead to a small color change, i.e. a smooth color transformation is expected to lead to better matching colors when applied to other multi-band sensor images. Also, with smooth color transformations noise leads to less clutter. Furthermore, the color fused result provides a better impression of the depth in the scene (compare e.g. Fig. 8b and 8f).

Fig. 9a-c shows the result of applying the same lookup-table transform to multi-band sensor images of different scenes, together with the corresponding daytime full color images (Fig. 9d-f). Although the colors do not always fully match the daytime colors, they are still characteristic for the different materials displayed in the scene. Thus, the colorized fused image facilitates interpretation of the scene (situational awareness).

Dedicated color mapping schemes can be derived that are optimally tuned for different environments. When deploying the color transfer method at night an appropriate color

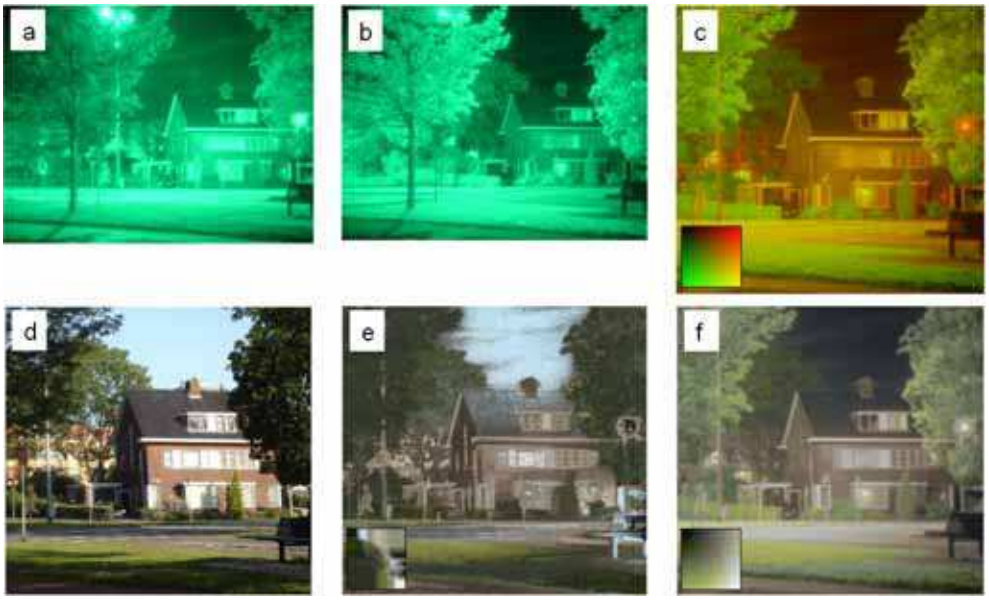


Fig. 8. Visual (a) and near-infrared (b) images of the same scene. (c) Combined red-green false color representation of (a) and (b). (d) Daytime reference color image. (e) Result of straightforward application of the sample-based transform method. (f) Result after smoothing and linearising the color lookup table.

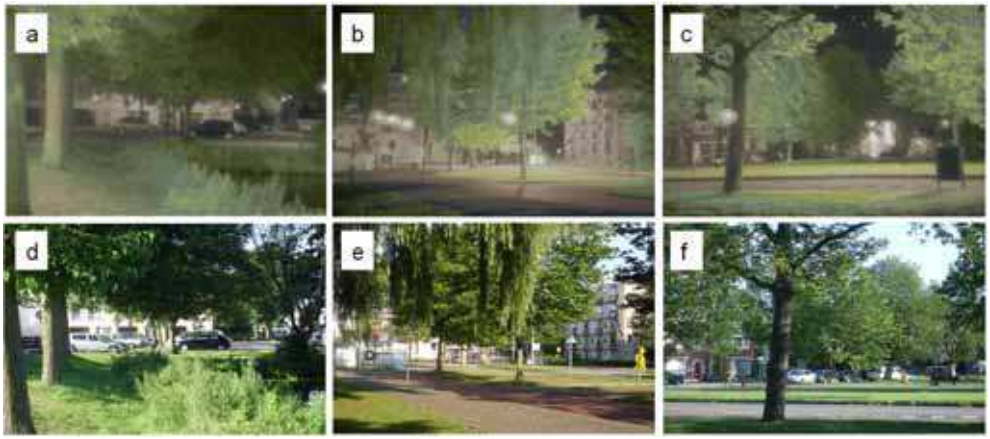


Fig. 9. (a-c) Results of applying the coloring scheme from Fig. 8f to different multi-band sensor images. (d-f) Daylight color images corresponding to (a-c).

mapping scheme should then be selected to colorize the night-time images. The color transformation consists of *(1)* creating a false-color image (e.g. an RG-image, see Fig. 6e), *(2)* converting this image into an index image using the RG-color table, and *(3)* replacing the

color lookup table with its daytime equivalent (RGB-color table, see Fig. 6d, right). The whole transformation is defined by the two color lookup tables (the RG-color table and the RGB color table). The software implementation can be very fast.

Different environments require different color mapping schemes to obtain a correct color representation. The sample images used to derive the color mapping scheme should reflect the statistics of the environment in which the mapping will be deployed.

In practice the ratio between the sensor outputs is characteristic for different materials. This becomes apparent when one inspects the color map images (e.g. Fig. 6d, right side) corresponding to the optimal color mapping of different reference and test images. In those images the hue varies little along straight lines through the top-left corner (lines for which the ratio between the two sensor outputs has a constant value). This feature can be used when deriving a color mapping from a limited number of samples. Also, the color mapping (e.g. Fig. 6d, right side) can be expected to be smooth, i.e. from point to point the color variations will be smooth. When a smooth color mapping scheme is used more subtle differences between sensor outputs will become visible.

Because the sample-based color mapping is highly specific it can effectively be used to highlight interesting image details which may otherwise go unnoticed. Camouflaged targets (e.g. persons or vehicles in military colors) are usually indistinguishable from their local background in naturally colored images. As a result, they will also have low color contrast when a natural color mapping is applied to multi-band nighttime imagery. An example of this effect is shown in Fig. 10, which presents an intensified image (Fig. 10a) and a thermal infrared (8-12 μm) image (Fig. 10b) of a person wearing a military battle dress with a camouflage pattern in a rural background. Fig. 10c shows the two-band false color image that is constructed by mapping the images from Fig. 10 (a) and (b) to respectively the R and G channels of an RGB color image (the B channel was set to zero). Fig. 10d shows the full color daytime reference image (a standard digital color photograph). Using the sample-based method we derived the color transformation that results in an optimal match between the colors in the reference image (Fig. 10d) and the multi-band sensor values (Fig. 10c). The color transformation for all sensor combinations is represented by the insets of Fig. 10c and Fig. 10e, while Fig. 10e shows the result of the color mapping. Note that the colors in the colorized multi-band image (Fig. 10e) closely match the colors in the reference image (Fig. 10d). However, since the person wears clothing in camouflage colors, he is also camouflaged in the colorized nighttime image. Although the person is clearly visible in the thermal image (Fig. 10b) he can hardly be detected in the colorized nighttime image. To make the person more salient in the colorized night-time image a color transformation can be used that depicts hot items (which usually are potential targets like vehicles or living beings) in red. For this purpose we created an alternative color mapping by manipulating the inset of Fig. 10e. The resulting lookup table is depicted in the inset of Fig. 10f. Fig. 10f shows the result of the application of this color transformation to the false color two-band nighttime image in Fig. 10c. Hot items now appear in red while the (relatively cold) background is still depicted in its natural colors. In this image representation, the naturally colored background provides the context and potential targets are highlighted by their color contrast. This color mapping may be useful for applications like surveillance and navigation, since these tasks require a correct perception of the background (terrain) for situational awareness in combination with optimal detection of targets and obstacles.