International Journal of Image and Data Fusion
Vol. 2, No. 4, December 2011, 287–308

Taylor & Francis
Taylor & Francis Group

Augmenting full colour-fused multi-band night vision imagery with synthetic imagery in real-time

Alexander Toet[a]*, Maarten A. Hogervorst[a], Rob van Son[b] and Judith Dijk[b]

[a]*TNO, Kampweg 5, Soesterberg 3769DE, The Netherlands;* [b]*TNO, Oude Waalsdorperweg 63, The Hague, 2597 AK, The Netherlands*

(*Received 13 March 2011; final version received 14 June 2011*)

We present the design and first field trial results of an all-day all-weather enhanced and synthetic-fused multi-band colour night vision surveillance and observation system. The system augments a fused and dynamic three-band natural-colour night vision image with synthetic 3D imagery in real-time. The night vision sensor suite consists of three cameras, sensitive in, respectively, the visual (400–700 nm), the near-infrared (NIR, 700–1000 nm) and the long-wave infrared (LWIR, 8–14 μm) bands of the electromagnetic spectrum. The optical axes of the three cameras are aligned. Image quality of the fused sensor signals is enhanced in real-time through dynamic noise reduction, super resolution and local adaptive contrast enhancement. The quality of the LWIR image is enhanced through scene-based non-uniformity correction. The visual and NIR signals are used to represent the fused multi-band night vision image in natural daytime colours, using the Colour-the-Night colour remapping technique. Colour remapping can also be deployed to enhance the visibility of thermal targets that are camouflaged in the visual and NIR range of the spectrum. The dynamic false-colour night-time images are augmented with corresponding synthetic 3D scene views, generated in real-time using a geometric 3D scene model in combination with position and orientation information supplied by the Global Positioning System and inertial sensors of the system. Initial field trials show that this system provides enhanced situational information in various low-visibility conditions.

Keywords: augmented reality; image fusion; false colour; natural colour mapping; real-time fusion; night vision

1. Introduction

Night vision cameras are a vital source of information for a wide-range of critical military and law enforcement applications related to surveillance, reconnaissance, intelligence gathering and security. The two most common night-time imaging systems cameras are low-light-level (e.g. image-intensified) cameras, which amplify reflected visible to near infrared (NIR) light, and thermal or long-wave infrared (LWIR) cameras, which convert thermal energy from the midwave (3–5 μm) or the long-wave (8–12 μm) part of the spectrum into a visible image. Until recently, monochrome display of night vision imagery

*Corresponding author. Email: Lex.Toet@tno.nl

ISSN 1947–9832 print/ISSN 1947–9824 online

http://dx.doi.org/10.1080/19479832.2011.598135
http://www.tandfonline.com

has been the standard. However, the increasing availability of multi-band infrared and visual night vision systems has led to a growing interest in the colour display of night vision imagery (Shi *et al*. 2005a, b, Tsagiris and Anastassopoulos 2005, Zheng *et al*. 2005, Li and Wang 2007). In principle, colour imagery has several benefits over monochrome imagery for surveillance, reconnaissance and security applications. For instance, a colour representation significantly increases the number of different materials that can be discriminated in a scene, and may enhance feature contrast, thus enabling better performance in scene recognition and object detection tasks. However, when sensors operate outside the visible waveband, artificial colour mappings generally produce false colour images whose chromatic characteristics do not correspond in any intuitive or obvious way to those of a scene viewed under natural photopic illumination. This type of representation may disrupt the recognition process, resulting in an observer performance that is degraded compared to that obtained with single-band imagery alone (Sinai *et al*. 1999). Several different techniques have been proposed to give multi-band night-time imagery a realistic colour appearance (Waxman *et al*. 1995, Wang *et al*. 2002, Toet 2003b, Sun *et al*. 2005, Tsagiris and Anastassopoulos 2005, Zheng *et al*. 2005), some of which were implemented in real-time night vision systems (Waxman *et al*. 1999, Fay *et al*. 2000, Howard *et al*. 2000, Wang *et al*. 2007, Yue and Topiwala 2007). However, most of these techniques are computationally expensive and/or do not achieve colour constancy. We recently introduced a new method to display multi-band night-time imagery in natural daytime colours (Hogervorst *et al*. 2006, Hogervorst and Toet 2008b). The method is computationally simple and can therefore be easily deployed in real-time. Moreover, it provides stable colourisation under variations in lighting conditions and scene content.

Here, we describe the implementation of our previously developed colour mapping scheme in a prototype portable and stand-alone all-day all-weather surveillance and navigation system, mounted on a mobile platform. The system provides co-aligned visual, NIR and thermal signals from two image intensifiers and an uncooled microbolometer, respectively. A fast lookup-table implementation of our new colour mapping scheme (Hogervorst and Toet 2008b) is deployed to achieve real-time display of the multi-band images in natural daylight colours (using information from the visual and NIR bands) and to maximise the detectability of thermal targets (using the LWIR signal). To provide optimal situational information for applications such as military command and control, facility security and catastrophe management, the system further augments the dynamic false-colour night-time images with corresponding synthetic 3D scene views. The synthetic scene is generated in real-time using a geometric 3D scene model of the operation theatre, in combination with the six degree-of-freedom position and orientation information supplied by the Global Positioning System (GPS) and inertial sensors of the system. The overall system is designed to serve as a mobile all-day all-weather navigation and surveillance tool.

The rest of this of this article is organised as follows. In Section 2, we present the hardware components of the surveillance and navigation system, and the layout of its sensor suite. Section 3 briefly reviews our new colour mapping procedure. Section 4 briefly describes the dynamic image enhancement techniques that were applied. In Section 5, we discuss the synthetic vision system and the way it is interfaced with the sensor suite. Section 6 presents some results of a recent field trial. Finally, conclusions are made in Section 7.

2. System hardware

2.1 *Sensor suite*

Figure 1 shows a schematic representation of the layout of the sensor suite and its optical components. The system contains two digital image intensifiers and a thermal camera.

The two image intensifiers are high-resolution (1280×960) Photonis PP3000U Intensified Camera Units (ICUs, www.photonis.com). The ICU is a low light level, intensified CMOS camera. It has a 2/3″ CMOS sensor with a spectral response range of 400–900 nm, and delivers both a PAL or NTSC composite video signal output (ITU-R BT.656-4, 640×480 pixels, 25 frames/s) and an SDI – LVDS 270 Mbits/s signal. Both ICUs are equipped with Pentax C2514M CCTV lenses, with a minimal focal length of 25 mm and a lens aperture of F/1.4, resulting in a field-of-view (FOV) of 30.7×24.8.

The thermal camera is a XenICs Gobi 384 uncooled a-Si infrared microbolometer (www.xenics.com). It has a 384×288 pixels focal plane array, and a spectral sensitivity range of 8–14 μm, which is the range of most interest for outdoor applications. It is equipped with an Ophir supIR18 mm F/1 lens (www.ophiropt.com) providing a $29.9^\circ \times 22.6^\circ$ wide angle view. The Gobi 384 has a 16-bit Ethernet and CameraLink interface, and a frame rate of 44 frames/s.

Two beam splitters are deployed to direct the appropriate band of the incoming radiation to each of the three individual sensors (Figure 1). The incoming radiation is first split into a long-wave (thermal) and a visual+NIR part by a heat reflecting (hot) mirror (a custom made Melles Griot dichroic beam splitter consisting of Schott N-BK7 Borosilicate Crown glass with an Indium Tin Oxide coating, with a reflection $R > 85\%$). The long-wave part of the spectrum is reflected into the lens of the thermal camera, while

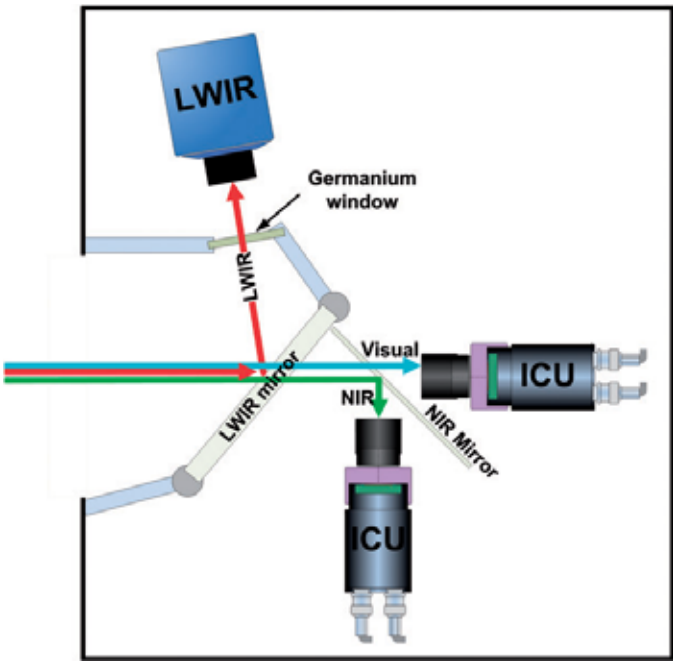

Figure 1. Schematic representation of the layout of the sensors and filters. The visual band of the incoming radiation goes straight into one of the ICU. The NIR band is reflected by a beam splitter into the second ICU. Long-wave radiation is reflected by a hot mirror into the aperture of a thermal camera.

the visual+NIR light is transmitted to a combination of two digital image intensifiers that are mounted under an angle of 90°. Next, an NIR reflecting mirror (45° angle of incidence, Borofloat glass, type Edmund Optics B43-958, $101 \times 127 \times 3.3\,\text{mm}^3$, see: www.edmundoptics.com) is used to separate the incoming light, by transmitting the visual (400–700 nm) and reflecting the NIR part (700–900 nm), such that one image intensifier registers the visual part and the other one only detects the NIR part of the incoming radiation. The sensor geometry is such that the optical axes of all cameras are aligned. The sensors and the mirrors are mounted on a common metal frame. The whole configuration is portable, and is contained in a dust and water resistant housing that can be easily mounted on mobile platforms (Figure 2). A Germanium window is used to cover the aperture of the thermal camera. The sensor suite delivers both analogue video and digital signal outputs.

2.2 *GPS receivers*

An internal U-blox EVK-5P Positioning Engine (www.u-blox.com) provides a position and orientation (i.e. sensor location and viewing direction) signal through the high-speed 7-port USB 2.0 hub (Figure 2a). The accuracy in position is less than 3 m. The accuracy in orientation is less than 5°. In local area operations, and when high accuracy is required, an external Trimble SPS751 GPS receiver set (www.trimble.com) is connected to the system, to achieve high position accuracy (<1 cm) through real-time kinematic GPS signal correction.

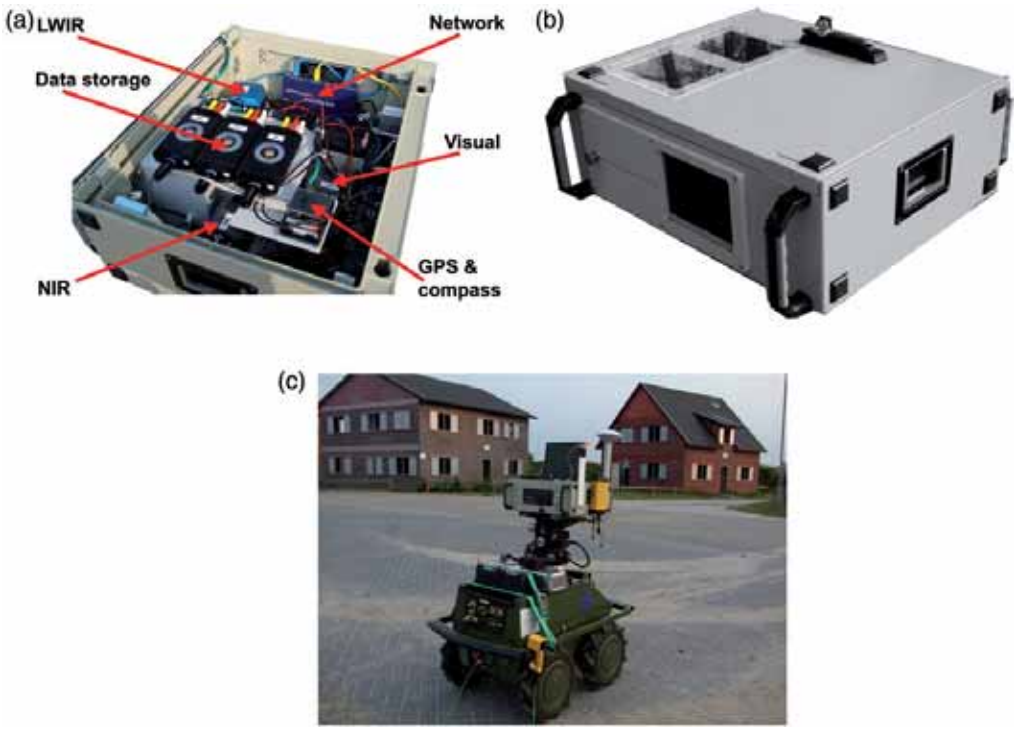

Figure 2. The sensor suite. (a) The interior showing the sensors and other components. (b) The system is contained in a water and dust resistant housing with a single aperture and built-in displays that enable signal monitoring. (c) The sensor suite mounted on an all-terrain platform.

2.3 *Electronic compasses*

An internal Silicon Labs F350-COMPASS-RD multi-axis electronic compass (www.silabs.com) provides the azimuth and tilt angles of the optical axis of the sensor suite, with an accuracy of a few degrees (Figure 2a). When the viewing direction needs to be known with higher accuracy, an external Xsens 3D inertial measurement unit motion sensor with accelerometer, magnetometer and gyroscope (www.xsens.com) is connected to the system to measure Yaw, Roll and Pitch with an accuracy less than 0.1°.

2.4 *Computer*

A Dell Precision M2400 Intel Core Duo P8600 2.4 GHz laptop with a solid-state hard disk is used to store, colourise and visualise the sensor signals and to generate and display the synthetic scene view. The current implementation achieves real-time (~25 Hz) visualisation, signal enhancement and data registration.

2.5 *Displays*

Two 6.4″ TFT video displays, embedded in the system casing, enable simultaneous monitoring of two of the three video signals (either Visual/NIR, Visual/LWIR or NIR/LWIR; Figure 2b). The laptop display (14 inch, 1440 × 900 pixels) is used to view the final fused, coloured and enhanced images, together with the synthetic 3D scene views.

2.6 *Data transfer and storage*

The Photonis ICUs are connected to a high-speed 7-port USB 2.0 hub. This enables the user to interface with the ICUs and to adjust their settings, or to download and install preferred settings.

A Pleora iPORT PT1000-ANL-2/6 frame grabber (www.pleora.com) is used to digitise the analogue video output signals of (1) both ICUs and (2) the Gobi 384. Digitisation is performed at a rate of 25 frames/s, with at a resolution of 640 × 480 pixels and 10 bits per pixel. The Pleora transmits these signals to a Netgear Gigabit Ethernet switch. The 16-bit TCP/IP Ethernet interface of the XenICs Gobi 384 is also directly connected to the Netgear Gigabit Ethernet switch.

Three Pinnacle Video Transfer Units (www.pinnaclesys.com/PVT) are provided to store (a) the analogue video signals of all three cameras, and (b) the audio signals of two (optional) external microphones, either on three internal 320 GB hard disks, or on USB memory sticks. The microphones can for instance be positioned on the front and back of the camera suite. The microphone on the front can then be used to register relevant audio information from the registered scene, and the second microphone can for instance be used to record spoken annotations.

2.7 *Image registration and matching*

The Visual and NIR images provided by the two ICU digital image intensifiers have a size of 640 × 480 pixels and represent an FOV of 30.7° × 24.8°. The LWIR image provided by the XenICs Gobi 384 thermal camera has a size of 384 × 288 pixels and represents an FOV

of $29.9^\circ \times 22.6^\circ$. As a result of the optical alignment of the camera systems, the FOV of the LWIR image corresponded to the central part of the FOV of the Visual and NIR images. The size of this common FOV area was 621×461 pixels in the Visual and NIR images. To enable the fusion of the LWIR image with the other two channels, the LWIR image (384×288 pixels) was therefore bilinearly interpolated and up-sampled (by a factor of about 1.6) to 621×461 pixels. The Visual and NIR images were cropped to their central part of 621×461 pixels, so that only the common FOV area remained. As a result, the size of the three-band colour images used in the rest of this study was 621×461 pixels, and an individual pixel represented about 2.9×2.9 min of arc of the visual field.

3. Colour fusion transform

The signals from the three night-time cameras are fused and represented in natural daytime colours using the Colour-the-Night colour remapping procedure (Toet 2003a, b), which can be efficiently implemented as a lookup table transform (Hogervorst *et al*. 2006, Hogervorst and Toet 2008a, 2010). Colour fusion can therefore be performed in real-time on a standard PC. For reasons of portability, we chose to use a standard laptop, connected to the sensor suite via an Ethernet connection.

For the sake of completeness we will now describe our new Colour-the-Night method for deriving a natural colour transformation using the example shown in Figure 3 (for an extensive review of the method see Hogervorst *et al*. 2006, Hogervorst and Toet 2008a, 2010). Figure 3(a) depicts a full colour daytime reference image, which is in this case a colour photograph taken with a standard digital camera. Figure 3(b) and (c), respectively, show a visible and NIR image of the same scene. Figure 3(f) shows the result of applying daytime colours to the two-band night-time sensor image using our new colour mapping technique.

The method works as follows. First, the multi-band sensor image is transformed to a false-colour image by taking the individual bands (Figure 3b and c) as input to the R and G channels (and B when the sensor contains three bands), referred to as the RG-image (Figure 3e). In practice, any other combination of two channels can also be used (i.e. one could just as well use the combinations R & B or B & R). Mapping the two bands to a false colour RGB-image allows us to use standard image conversion techniques, such as indexing. In the next step, the resulting false colour (RG-image; Figure 3e) 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 colour lookup table (the number of entries in this table can be chosen by the user). In the present example of a sensor image consisting of two bands (R and G; Figure 3e), the colour lookup table contains various combinations of R and G values (the B-values are zero when the sensor or sensor pair only provides two bands). For each index representing a given R, G combination (a given false colour), the corresponding natural colour equivalent is obtained by locating all pixels in the target image with this particular index and finding the corresponding pixels in the (natural colour) reference image (Figure 3a). First, the RGB-values of the reference image are converted to perceptually de-correlated $l\alpha\beta$ values (Ruderman *et al*. 1998). Next, we calculate the average $l\alpha\beta$-vector over this ensemble of pixels. This assures that the computed average colour reflects the perceptual average colour. Averaging automatically takes the distribution of the pixels into account: colours that appear more frequently are attributed a greater weight. Let us for instance assume that we would like to derive the natural colour associated with index 1.

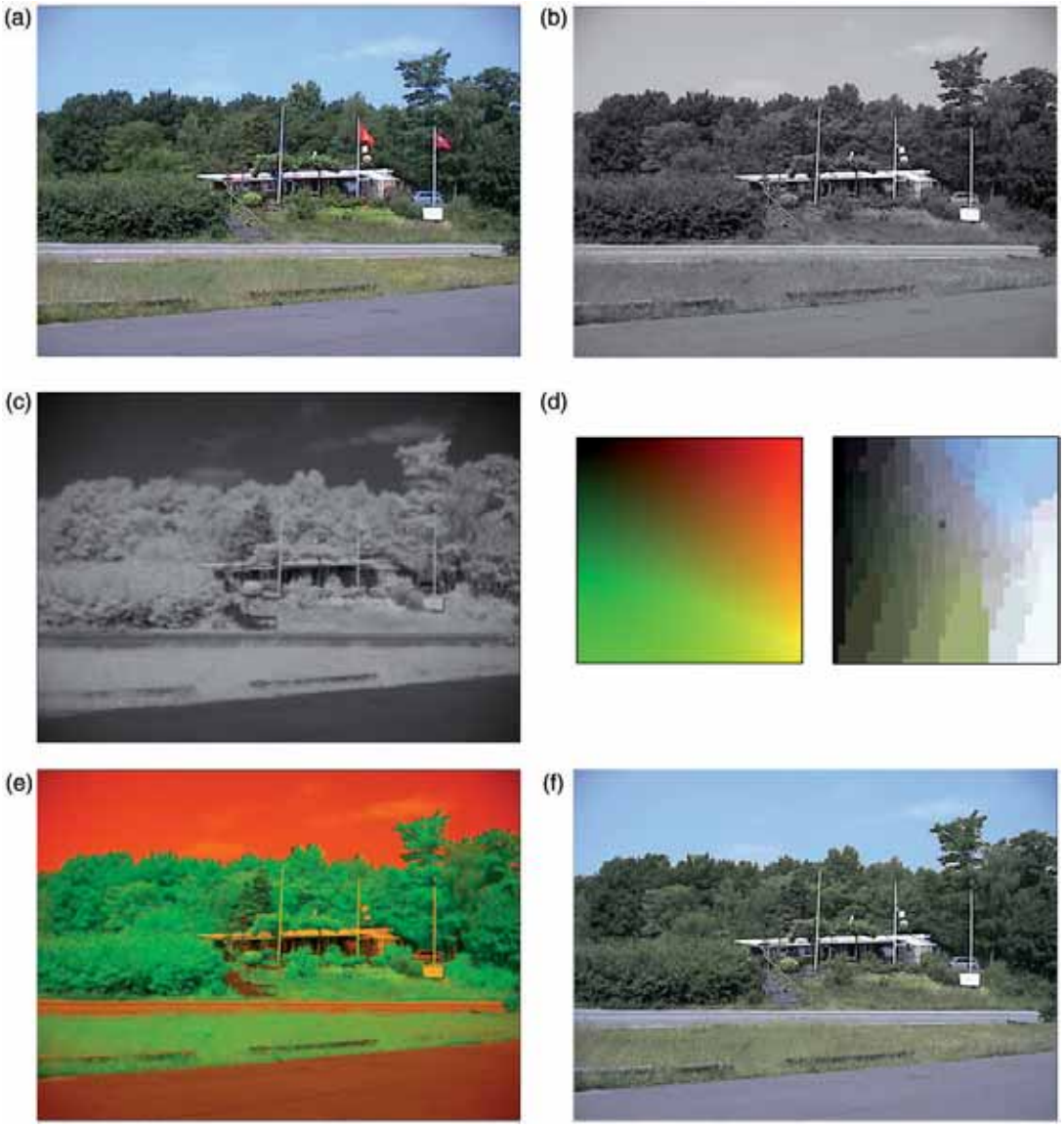

Figure 3. (a) Natural daylight colour reference image. Visible (b) and NIR (c) images of the same scene. (d) The colour mapping derived from corresponding pixel pairs in (a)–(c). (e) Combined RG false colour representation of (b) and (c), obtained by assigning (b) to the green and (c) to the red channel of an RGB colour image (the blue channel is set to zero). (f) Result of the application of the mapping scheme in (d) to the two band false colour image in (e).

In that case, we locate all pixels in the (indexed) false colour multi-band target image with index 1. We then collect all the corresponding pixels in the reference daytime colour 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 colour lookup table. These steps are successively carried out for all indices. This process yields a new colour lookup table containing the natural colours associated with the various multi-band combinations in the false colour (RG) colour lookup table.

The Colour-the-Night transform is then performed by simply replacing (swapping) the RG-colour lookup table (left image in Figure 3d) with the colour lookup table with natural

colours (right image in Figure 3d). The result is an image with a natural colour appearance, in which the colours are optimised for this particular sample set (Figure 3d). The relation between the sensor triples and natural colour triples can be derived beforehand from a set of corresponding samples from an image pair consisting of, respectively, (1) a multi-band sensor image of a given scene and (2) a registered daytime colour image (either a colour photograph or a synthetic colour image) of the same scene.

For an 8-bit three-band system, the 3D colour lookup table contains $256 \times 256 \times 256$ entries. When the colour lookup table contains fewer entries, the colour mapping is achieved by determining the closest match (in a nearest neighbour sense) of the table entries to the observed multi-band sensor values. Once the colour transformation has been derived and the colour lookup table pair that defines the mapping has been created, the table pair remains fixed and can be deployed in a real-time application. The lookup table transform requires minimal computing power and can therefore be performed in real-time. An additional advantage of the colour lookup transform method is that object colours depend only on the multi-band sensor values, and are independent of the actual image content. As a result, objects keep the same colour over time when registered with a moving camera (i.e. colour constancy is inherent in this procedure).

4. Image enhancement

4.1 *Dynamic super resolution*

Camera images suffer from temporal noise (particularly in low-visibility conditions), which may hinder target detection and comfortable observation. When the camera and scene are stationary, or when the frames can be aligned accurately, temporal noise reduction can be achieved through averaging multiple frames or through more complex techniques such as dynamic super resolution (DSR: Schutte *et al*. 2003). DSR combines noise reduction with resolution enhancement. For optimal performance, DSR relies on accurate motion estimation, which is provided by a scene-based motion estimation (SBME) algorithm. SBME estimates image motion from the image information itself.

For intensified imagery only noise reduction is applied, since these images are mostly degraded by noise. The thermal camera image contains more aliasing, so that resolution enhancement will produce better results.

4.2 *Scene-based non-uniformity correction*

Scene-based non-uniformity correction (SBNUC) serves to correct for the different response characteristics of the detectors in the thermal camera sensor (Schutte *et al*. 2003). SBNUC is based on the assumption that the scene content moves slowly relative to the frame rate, whereas the non-uniformities are semi-static. Accurate motion estimation is used to identify the actual apparent scene movement, from which a motion-corrected difference image can be calculated, containing the current residual non-uniformity with some additional temporal noise. Based upon an advanced image formation model, SBNUC provides optimal reduction of the non-uniformities over time.

4.3 *Local adaptive contrast enhancement*

Local Adaptive Contrast Enhancement (LACE) is a multi-resolution technique to compress the overall dynamic range of the images, such that small amplitude details are retained and an overall natural look is maintained (Schutte 1997). For an input image I, the enhanced output image O is given by

$$O = I + \sum_{i=1}^{k} c(\sigma_i) \cdot (I - \mu_i), \tag{1}$$

where μ_i and σ_i represent, respectively, the local mean and standard deviation of the input image at resolution level i, and the multiplication factor is given by

$$c(\sigma) = \alpha \frac{M}{\sigma} - 1, \tag{2}$$

where M is the global mean of I, and α is a free parameter.

LACE has three free parameters: the number of resolution levels (k), the weight of a processed level relative to the original image and the maximum allowed multiplication (Schutte 1997).

Since LACE alters the dynamic range considerably, it is applied after the image fusion and colour remapping procedures.

5. Synthetic scene generation

The system also includes a scene generator capable of generating and displaying synthetic 3D views of the environment in real-time (see Figure 4 for a schematic overview of the system). The scene generator combines *a priori* information about the area of operation with real-time input from the position information system, to produce a corresponding

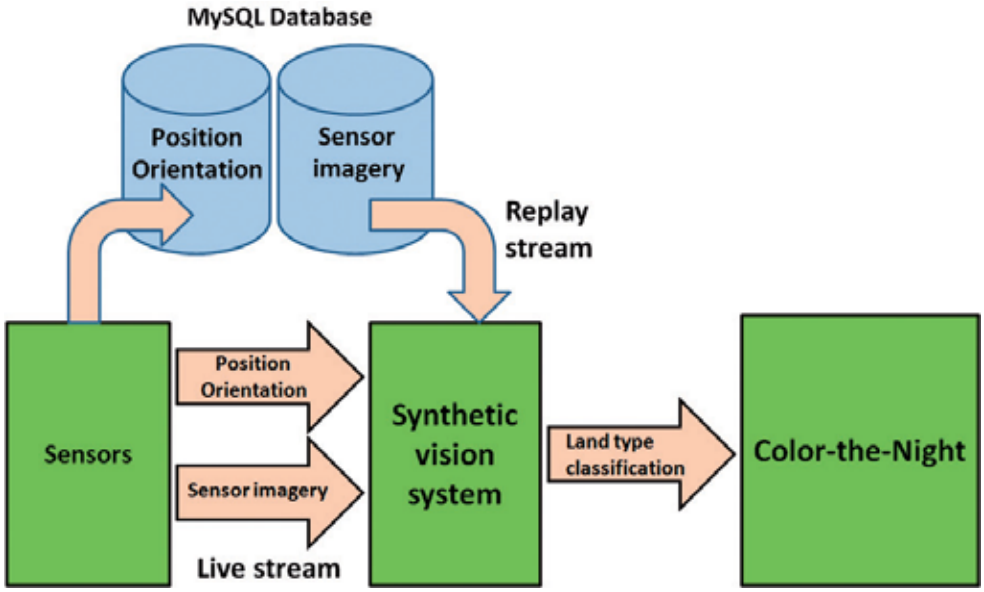

Figure 4. System overview. The arrows represent the interfaces to and from the synthetic vision system.

view of the environment. The synthetic scenes augment the sensor images and serve to increase and maintain the operator's situation awareness. The system is particularly beneficial in adverse real-world viewing conditions (e.g. due to bad atmospheric conditions or obscurants) or when the operator's view is obstructed by obstacles (e.g. houses or vegetation).

The environmental information used by the synthetic vision system consists of several components:

- A 3D visual representation of the area of operation, containing terrain elevation, buildings and their interiors, vegetation and other static or more or less permanent objects in the environment.
- A classification of the objects in the environment into a discrete number of types (i.e. buildings, building walls and doors, vegetation, transportation surfaces, etc.).
- Material classifications for all surfaces in the environment.

The synthetic vision system is built using TNO's enhanced virtual environment (VE), a modular data-driven and data-centric framework for 3D simulation and visualisation (de Kraker *et al*. 2009). The system consists of a 3D view and a number of other views that are capable of displaying sensor images. A graphical user interface allows the user to control the visual appearance of the 3D VE and to control the timing of a scenario, in the case of a replay of data that were recorded earlier.

A geo-specific 3D visual terrain database of the Dutch MOUT training village Marnehuizen has been used for the present experiments. This database has been modelled using construction blueprints and close-up photographs. Because of possible differences between the original blueprints and the final construction, the accuracy of the terrain database is estimated at 1 m.

The synthetic vision system offers a number of features that allow the operator to manipulate his view on the world in real-time and acquire insight in his situation. It can be used for mission planning and preparation (to explore an environment before actually entering it), to fill in missing details in the sensor images in low-visibility conditions, or to acquire additional situational awareness in the field.

5.1 *Play modes*

The system is capable of visualising the VE in two discrete play modes (Figure 4). The first mode is the live-play mode, in which the system uses up-to-date information to display the VE in real-time. The second mode is the replay mode, in which recorded data (both position and information data as well as sensor images) is retrieved from a database. In replay mode, the user is able to control the timing of the scenario by means of a time slider bar and several buttons (e.g. reverse, forward, fast-forward).

5.2 *View modes*

In the context of a synthetic vision system, a viewpoint represents the vantage point from which the virtual space is visualised. Synthetic visualisations allow the use of viewpoints that would not be normally feasible in reality. Typical display formats are egocentric or first-person perspective (1PP), exocentric or third-person perspective (3PP) and tethered viewpoints (e.g. Haskell and Wickens 1993, Wang and Milgram 2009).

It is well-known that the nature of the task dictates the best viewpoint on a geographic terrain. Unlike a 1PP, a 3PP provides a cognitive advantage, since a large part of the environment can be integrated into a single representation (St. John *et al*. 2001), thus eliminating the need to scan and redirect the physical viewing direction when constructing a mental representation of the spatial surrounding (Amorim *et al*. 2000). However, a 1PP is best for tasks judging relative position, due to the distortions associated with a 3PP (St. John *et al*. 2001, Wickens *et al*. 2000). As a result, tasks like navigation and wayfinding, which require global situational awareness, are best performed with a 3PP showing a large part of the surrounding terrain, whereas tasks like navigating and wayfinding are best done with a 1PP (Lamb and Hollands 2005a, Wickens and Hollands 2000). Hence, users may in principle optimise their performance in VEs by switching between these different viewing modes (Amorim *et al*. 2000). Viewpoint tethering has been proposed as a way of integrating information from both egocentric and exocentric frames of reference (Haskell and Wickens 1993). Viewpoint tethering reflects a particular type of viewpoint design that is characterised by coupling the viewpoint to the sensor through a virtual tether. The virtual tether may either be rigid (e.g. Haskell and Wickens 1993) or dynamic (Wang and Milgram 2009). Tethered viewing generally yields intermediate performance on both relative position and global situational awareness tasks (Lamb and Hollands 2005b).

The multi-sensory navigation and surveillance system presented in this article provides three viewing modes. These are views from a 1PP, a tethered 3PP and a free-roaming 3PP. The first-person view uses the system's position and orientation information to provide an accurate view of the synthetic environment from the sensor's perspective. From the first-person view, the operator is able to match the information visible on the sensory views directly with the information from the synthetic view (Figure 5e). This information provides cues for object recognition and classification. The synthetic scenes generated from this viewpoint can also be directly fused with or overlaid on the sensor images. Furthermore, they can also be used to derive a colour mapping in situations where no daytime colour reference images are available.

The tethered third-person view allows the operator to view the environment from a 3PP that is relative and attached to the sensor's actual position (Figure 5h). In this view, the sensor's viewing cone is represented visually in the environment to provide cues about the sensor's position and orientation in the area of operation. Using this view, it is relatively easy for the operator to gain additional insight into otherwise difficult-to-view parts of the area of operation, while at the same time keeping track of his own position and local surroundings.

The free-roaming third person view is similar to the tethered third-person view, in the sense that it is also external. The free-roaming view, however, is not attached to the sensor's actual position and allows the operator to explore the environment freely. In this view, the sensor's viewing cone is not represented. Compared to the tethered third-person view, this view allows the user complete freedom when exploring the area of operation, but may also cause the operator to lose track of his own position.

5.3 *Object class switches*

In any mode of the synthetic vision system, the operator is able to 'switch off' classes of objects. This feature is particularly effective when both sensory and synthetic views are

Figure 5. (a)–(c) Visual, NIR and LWIR input signals. (d) Image obtained by fusing the signals from (a)–(c) and applying Colour-the-Night remapping. (e) Corresponding synthetic scene in a 1PP. (f) As (e), after removing the walls of the building. (g) As (f), seen from a free-roaming 3PP. (h) Synthetic scene from a third person tethered point of view, showing the view cone of the sensor suite. (i) Material view in a free-roaming 3PP.

densely cluttered. Also, this feature allows the user to see through obstacles. This can be employed to inspect the area behind a particular object quickly, by removing it entirely from view, or to inspect an object's interior, by making it transparent. This feature can be employed to plan possible paths to traverse or to identify potential hiding places for opposing forces.

5.4 *Material view*

The material view switches texture mapping of the synthetic environment's surfaces. In this view, the original photorealistic textures are replaced by colour textures that represent material characteristics. This enables an operator to distinguish different aspects easily, such as material type and trafficability of ground surface elements, and to use these aspects for object classification and navigation purposes.

5.5 *Integration with the sensor suite*

The synthetic vision system is integrated with the sensor suite in several ways, both using information gathered by the system and providing information back (Figure 4). The following sections describe the interfaces between the synthetic vision system and the sensor suite.

5.6 *Position information system interface*

As described earlier, the synthetic scene generator uses information from system's position and orientation sensors to pinpoint the system's location within the synthetic environment (Figure 4). The system is currently capable to receive information both through a dedicated interface and through a standardised NMEA interface. In replay mode, these data are retrieved from an MySQL database.

5.7 *Night vision sensor interface*

Besides displaying synthetic views, the synthetic vision system is also capable of directly showing sensor images. These images can be displayed either next to the synthetic image in a 4×4 tiled window, or as image overlays on top of the synthetic image. The latter view mode allows for precise visual correlation between sensor images and the synthetic, first-person image. In replay mode, these data are retrieved from an MySQL database (Figure 4).

5.8 *Colour mapping interface*

The Colour-the-Night algorithm uses an environment-specific mapping (i.e. mappings previously derived from imagery representing e.g. typical urban, countryside or forest environments) to create synthetic daylight images from sensor images. Knowledge about the specific type of environment that the sensor is observing can therefore be used as a cue for choosing the correct mapping. By reasoning about the content of the visible

environment within the synthetic vision system's FOW, it is possible to deduce which type of environment is observed.

Based on a configurable mapping table, the system is capable of mapping the VE's textures to land type classifications. By counting all the pixels in screen that belong to a certain land type, a histogram of the distribution of land types is eventually calculated. These histograms are then matched to pre-calculated land type histograms that belong to specific Colour-the-Night mapping schemes. Using a least-square matching algorithm, the best match is selected as the Colour-the-Night mapping scheme used for the current frame.

6. Field trial

We tested the system during nocturnal data collection trials in the Dutch MOUT village Marnehuizen (Google Earth: Latitude 53.386293 and Longitude 6.262944). The ambient light levels were below 0.01 lux, making it very hard to distinguish details with the naked eye. Our geo-specific 3D visual terrain database of Marnehuizen had an accuracy of less than 1 m.

Figure 2 shows the sensor suite mounted on an all-terrain platform on the village square of Marnehuizen in daytime. Figure 5 shows the output of the (a) visual, (b) NIR and (c) LWIR, respectively, sensors signals at night. Notice that terrain details that contain chlorophyll (e.g. plants) strongly reflect NIR light and are therefore represented with a high intensity in the NIR channel (Figure 5b), and are ultimately displayed in shades of green in the final Colour-the-Night representation (Figure 5d). Figure 5(e) shows the corresponding synthetic scene viewed from the same location as the sensor suite. Figure 5(f) and (g) shows the result of removing the exterior walls from the synthetic representation, enabling the inspection of the building's interior. This type of view provides information that may be used to determine optimal entry and exit routes and to study the layout of buildings before actually entering them, or to determine possible hiding locations of opponents. Figure 5(h) shows how the synthetic representation can be used to inspect the viewing cone of the sensor suite. This information may be used to determine the optimal location and viewing direction when the sensor suite needs to be deployed in a surveillance scenario. Finally, Figure 5(i) shows a material view of the environment that can be used to determine the relative position of buildings and roads. This information may be used to assess the trafficability of the terrain and to determine optimal traversal routes.

Figure 6 illustrates the effects of the actual choice of the reference image (either a true colour photograph or a synthetic image) used in the Colour-the-Night colour transfer procedure, and the results of the image enhancement methods. Figure 6(a)–(c) shows the output of the visual, NIR and LWIR, respectively, sensors signals. Figure 6(d) shows the false colour image that is obtained by mapping these three bands to the R, G and B channels, respectively. Figure 6(e) shows a daytime colour photograph representing a different but similar scene in the neighbourhood. The Colour-the-Night mapping was derived from this photograph together with a corresponding three-band night-time image, and applied to Figure 6(d). The result is shown in Figure 6(f). Figure 6(g) shows the effect of dynamic noise reduction in (f). Figure 6(h) shows the result of LACE on Figure 6(g). The next four images in Figure 6 show that a synthetic colour image can also be used to derive the Colour-the-Night mapping when no true colour photograph is available. Figure 6(i) shows the synthetic scene corresponding to the actual scene that is represented

Figure 6. Top-down, left to right: (a)–(c) Visual, NIR and LWIR input signals. (d) False colour image obtained by mapping the three bands to the R, G and B channels, respectively. (e) Full colour daytime reference image. (f) Image obtained after applying Colour-the-Night remapping to the image shown in (d), using a mapping derived from (e). (g) The image shown in (f) after dynamic noise reduction. (h) Image from (g) after local contrast enhancement. (i) 1PP synthetic view of the scene depicted in (d). (j–l) As (f–h), after applying Colour-the-Night remapping to the image shown in (d), using a mapping derived from the synthetic scene (i).

by the system's sensors (Figure 6d). Application of the Colour-the-Night remapping that can be derived from a set of corresponding samples from the image pair (Figure 6d and i) to the false colour image (Figure 6d) yields Figure 6(j). Note that this image looks quite realistic, similar to Figure 6(f), although it is somewhat more bluish, due to the deep blue representation of the sky in (Figure 6i). Figure 6(k) shows the result of dynamic noise reduction on Figure 6(j). Finally, Figure 6(l) shows the result of LACE on Figure 6(k).

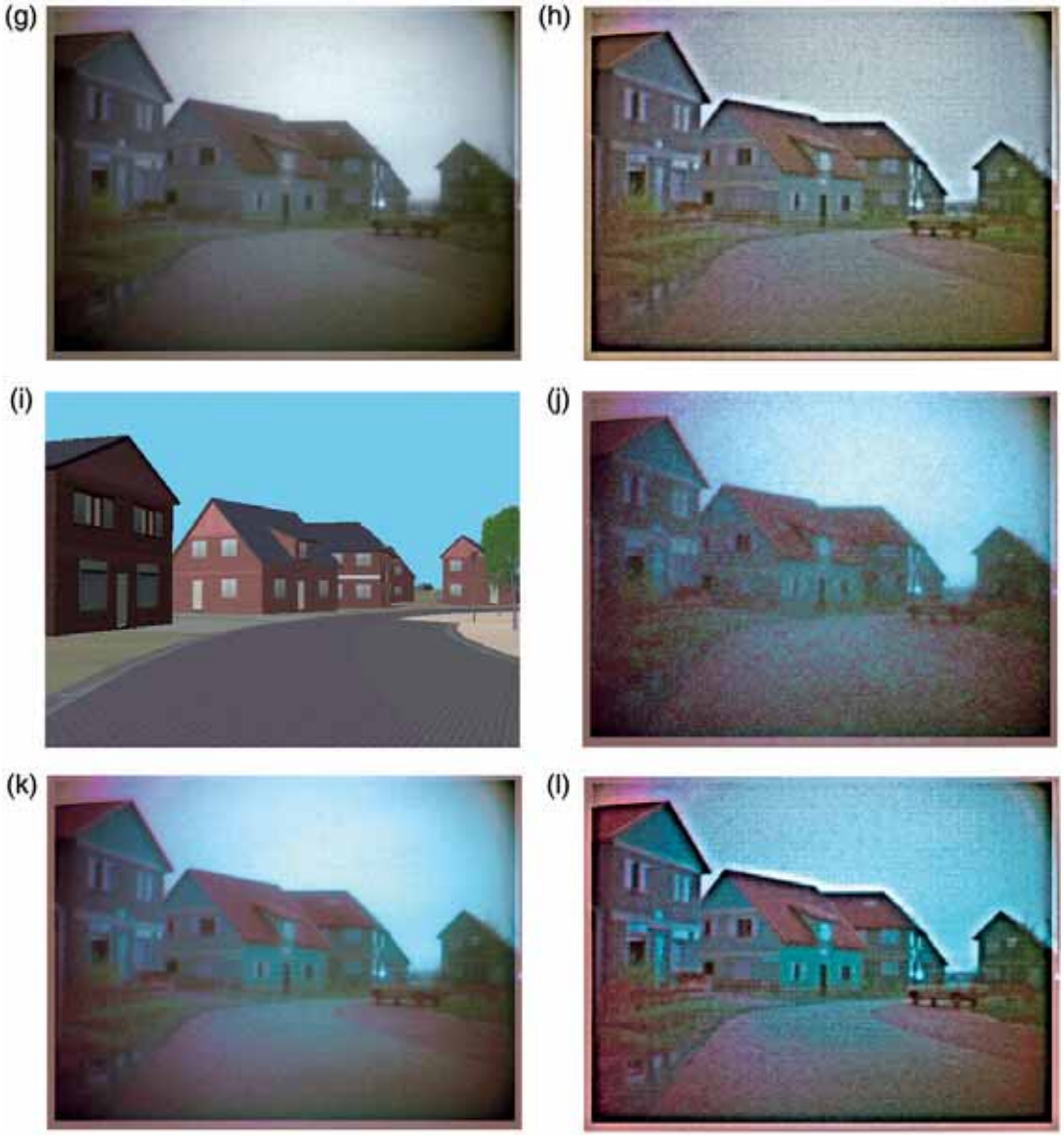

Figure 6. Continued.

Figure 7 shows a scene with barbed wire and a billboard in the foreground, and a running soldier in the background. Notice that the barbed wire is clearly visible in the fused Colour-the-Night image representation (Figure 7e) and the LWIR image (c), whereas it can be hardly distinguished in the visual (Figure 7a) and NIR (Figure 7b) image bands. This example also demonstrates that the synthetic scene represents only the static permanent scene elements like houses, streets and trees, whereas it does not represent objects that are only temporary present and therefore not included in the database, like the barbed wire, the billboard and the person.

Figure 8 illustrates that fusing LWIR with visual and NIR imagery adds the capability to see through smoke. The visual (Figure 8a) and NIR (Figure 8b) bands clearly represent the smoke cloud, which obscures the kneeling armed soldier in the scene. The thermal image (Figure 8c) clearly represents the soldier but shows no signs of the smoke cloud.

Figure 7. A scene with barbed wire and a billboard in the foreground, and a running soldier in the background. (a)–(c) Visual, NIR and LWIR signals. (d) False colour image obtained by mapping (a)–(c) to the R, G and B channels, respectively. (e) Image (d) after Colour-the-Night remapping. (f) The corresponding synthetic scene in a 1PP.

Fused in false colours (Figure 8d), the image shows both the smoke cloud and the man behind it, while the Colour-the-Night natural colour remapping (Figure 8e) makes it easy to distinguish the smoke cloud from the green grass and the blue sky. Hence, the Colour-the-Night image representation clearly provides enhanced situational information. The corresponding synthetic scene provides additional situational information by showing the presence of some additional buildings in the background that are not represented in the sensor images.

Figure 9 shows a village square, surrounded by low walls, a vehicle and three armed soldiers. The soldiers are wearing camouflage uniforms which make it hard to distinguish

Figure 8. A scene showing an armed soldier, kneeling and hiding behind a smoke cloud between two houses. (a)–(c) Visual, NIR and LWIR signals. (d) False colour image obtained by mapping (a)–(c) to the R, G and B channels, respectively. (e) Image (d) after Colour-the-Night remapping. (f) The corresponding synthetic scene in a 1PP.

them in the visual and NIR channels against the background of the military vehicle. They are represented at higher contrast in the LWIR image. Notice that the low vertical wall surrounding the village square is highly distinct in the LWIR image, whereas it cannot be distinguished from the floor of the square in the visual and NIR bands. Also, the texture of the pavement is well-articulated in the LWIR image but not represented in the other bands. The fused and re-coloured image Figure 9 has a quite natural colour appearance and clearly represents the low vertical wall around the square, the armed soldiers, the texture of the pavement and the vegetation in the background. The corresponding synthetic scene Figure 9(f) clearly shows that the small square in the foreground is surrounded by a low brick wall. This information is not evident from the original sensor images (Figure 9a–c), although it is

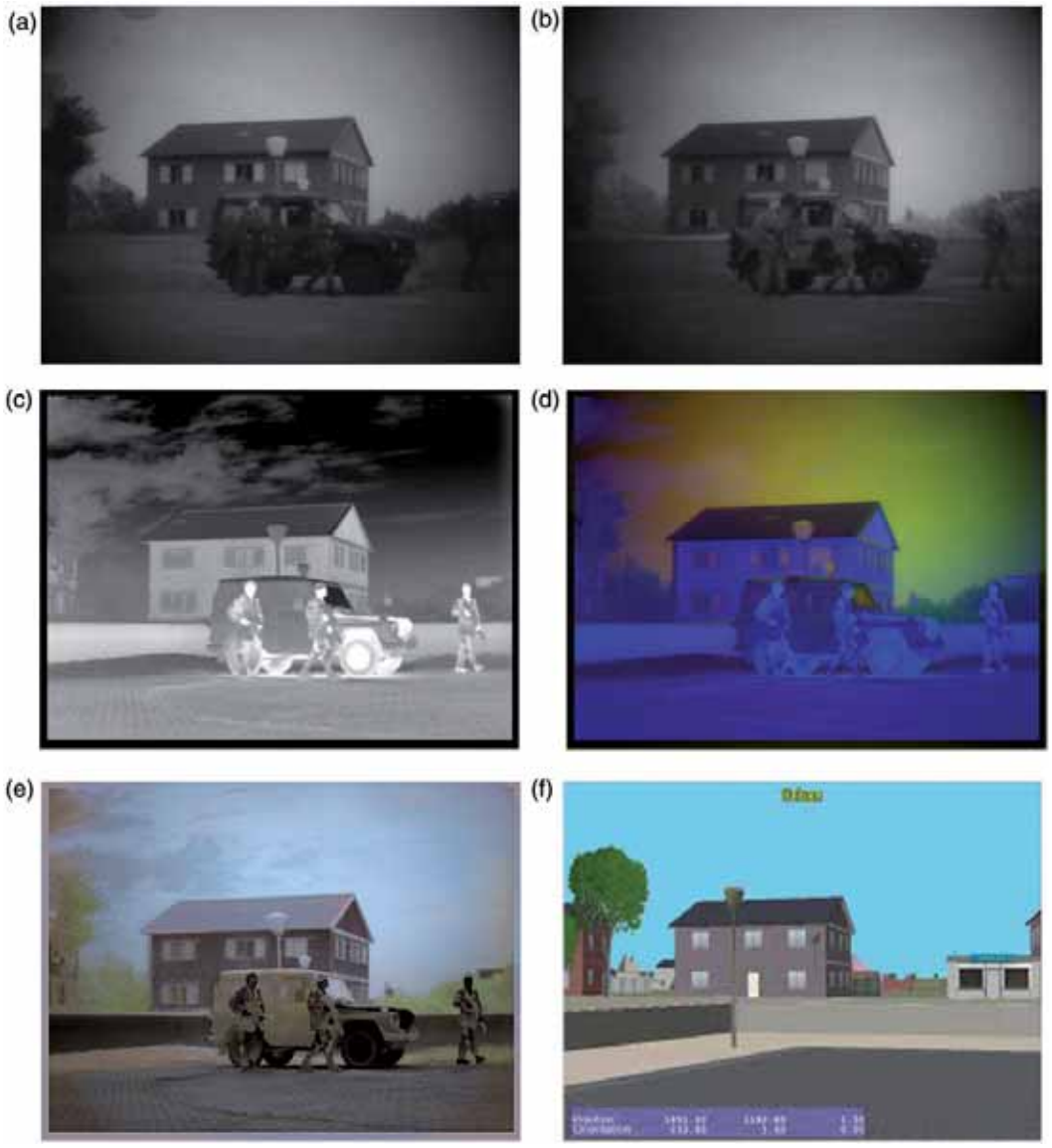

Figure 9. A scene showing a village square, surrounded by low walls, a vehicle and three armed soldiers. (a)–(c) Visual, NIR and LWIR signals. (d) False colour image obtained by mapping (a)–(c) to the R, G and B channels, respectively. (e) Image (d) after Colour-the-Night remapping. (f) The corresponding synthetic scene in a 1PP.

better visible in the fused natural colour representation (Figure 9e). The synthetic image (Figure 9f) also shows an entrance to the house in the background, which is obscured in the sensor images by the jeep in the foreground. Hence, in this example, the synthetic image provides additional information on the layout of the scene.

7. Conclusions

In this study, we present an all-day all-weather enhanced and synthetic fused multi-band natural-colour night vision surveillance and observation system. The system augments a

dynamic three-band natural-colour night vision image with synthetic 3D imagery in real-time. Initial field trials show that this system provides enhanced situational information in several different low-visibility conditions. The combination of natural colour-fused multi-band imagery with corresponding synthetic imagery provides a powerful surveillance and navigation tool for use in military or first-responder scenarios. A video presentation (www.scivee.tv/node/29094, doi: 10.4016/29094.01) and a video documentary (www.scivee.tv/node/29095, doi: 10.4016/29095.01) showing the system in action during a nocturnal field trial are provided elsewhere.

In this study, we used a geo-specific 3D visual terrain database of the environment where the system was deployed. Although such a database provides many additional capabilities (e.g. the inspection of the inside of buildings, inspection of parts of the scene that are blocked from view, on-site derivation of colour mapping schemes), its availability is not strictly required. When no terrain database is available, the system can also be synchronised with other types of geo-referenced imagery, like images from Google Earth or Street View, satellite images, or images from airborne Intelligence, Surveillance and Reconnaissance (ISR) Systems. The functionality of the synthetic viewing system will then be limited by the amount if information in the geo-referenced imagery that is provided (e.g. satellite imagery may not allow a 1PP representation).

The natural colour mapping scheme is not strictly dependent on the availability of daytime or synthetic imagery representing the current operating theatre. The only *a priori* information that is needed to produce a realistic colour setting of the fused multi-band night vision imagery is a general classification of the type of operating theatre in which the system will be deployed (e.g. urban, woodland, desert, maritime). In previous studies, we observed that the Colour-the-Night colour remapping procedure gives fused multi-band night vision imagery the general appearance of the reference imagery from which the colour scheme has been derived (Toet 2003b, Hogervorst and Toet 2010). Hence, general colour schemes can be derived and deployed for each of the typical operating theatres. The only requirement is that the imagery from which the colour schemes are derived contains the same kind of materials that also occur in the intended operating theatre (Toet 2003b). Although the colour setting of the resulting fused multi-band night vision images may not be identical to their actual daytime colours, the images will still have a realistic colour appearance, making it easy to grab the gist of the scene and perceive its content (Goffaux *et al*. 2005, Rousselet *et al*. 2005, Nijboer *et al*. 2008).

Acknowledgements

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-09-3095. The US Government is authorised to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

References

Amorim, M.-A., Trumbore, B., and Chogyen, P.L., 2000. Cognitive repositioning inside a desktop VE: the constraints introduced by first-versus third-person imagery and mental representation richness. *Presence: Tele-Operators and Virtual Environments*, 9 (2), 165–186.

de Kraker, K.J., Kerbusch, P., and Borgers, E., 2009. Re-usable behavior specifications for tactical doctrine. In: *Proceedings of the 18th conference on behavior representation in modeling and simulation (BRIMS 2009)*, 31 March–2 April. Orlando, FL: SISO, Inc., 15–22.

Fay, D.A., *et al*., 2000. Fusion of multi-sensor imagery for night vision: color visualization, target learning and search. *In: Proceedings of the 3rd international conference on information fusion*, I, 10–13 July. Paris, France: ONERA, TuD3-3–TuD3-10.

Goffaux, V., *et al*., 2005. Diagnostic colours contribute to the early stages of scene categorization: behavioural and neurophysiological evidence. *Visual Cognition*, 12 (6), 878–892.

Haskell, I.D. and Wickens, C.D., 1993. Two- and three-dimensional displays for aviation: a theoretical and empirical comparison. *International Journal of Aviation Psychology*, 3 (2), 87–109.

Hogervorst, M.A. and Toet, A., 2008a. Method for applying daytime colors to nighttime imagery in realtime. *In:* B.V. Dasarathy, ed., *Multisensor, multisource information fusion: architectures, algorithms, and applications 2008*, SPIE-6974. Bellingham, WA, USA: The International Society for Optical Engineering, 697403-1–697403-9.

Hogervorst, M.A. and Toet, A., 2008b. Nighttime imagery in natural daytime colors. *In:* B.V. Dasarathy, ed., *Multisensor, multisource information fusion: architectures, algorithms, and applications 2008*, SPIE-6974-2. Bellingham, WA, USA: The International Society for Optical Engineering.

Hogervorst, M.A. and Toet, A., 2010. Fast natural color mapping for night-time imagery. *Information Fusion*, 11 (2), 69–77.

Hogervorst, M.A., Toet, A., and Kooi, F.L., 2006. *TNO defense security and safety. Method and system for converting at least one first-spectrum image into a second-spectrum image*. Patent Number Patent Number PCT/NL2007050392. Patent Application Number Application Number 0780855.5-2202.

Howard, J.G., *et al*., 2000. Real-time color fusion of E/O sensors with PC-based COTS hardware. *In:* W.R. Watkins, D. Clement, and W.R. Reynolds, eds., *Targets and backgrounds VI: characterization, visualization, and the detection process*, SPIE-4029. Bellingham, WA: The International Society for Optical Engineering, 41–48.

Lamb, M. and Hollands, J.G., 2005. Viewpoint tethering in complex terrain navigation and awareness. *In: Human factors and ergonomics society annual meeting proceedings, perception and performance*. Santa Monica, CA: Human Factors and Ergonomics Society, 1573–1577.

Li, G. and Wang, K., 2007. Applying daytime colors to nighttime imagery with an efficient color transfer method. *In:* J.G. Verly and J.J. Guell, eds., *Enhanced and synthetic vision 2007*, SPIE-6559. Bellingham, MA: The International Society for Optical Engineering, 65590L-1–65590L-12.

Nijboer, T.C., *et al*., 2008. Recognising the forest, but not the trees: an effect of colour on scene perception and recognition. *Consciousness and Cognition*, 17 (3), 741–752.

Rousselet, G.A., Joubert, O.R., and Fabre-Thorpe, M., 2005. How long to get the 'gist' of real-world natural scenes? *Visual Cognition*, 12 (6), 852–877.

Ruderman, D.L., Cronin, T.W., and Chiao, C.-C., 1998. Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A*, 15 (8), 2036–2045.

Schutte, K., 1997. Multi-scale adaptive gain control of IR images. *In:* B.F. Andresen and M. Strojnik, eds., *Infrared technology and applications XXIII*, SPIE-3061. Bellingham, WA: The International Society for Optical Engineering, 906–914.

Schutte, K., de Lange, D.J., and van den Broek, S.P., 2003. Signal conditioning algorithms for enhanced tactical sensor imagery. *In:* G.C. Holst, ed., *Infrared imaging systems: design, analysis, modeling, and testing XIV*, SPIE-5076. Bellingham, WA: The International Society for Optical Engineering, 92–100.

Shi, J., *et al*., 2005a. Objective evaluation of color fusion of visual and IR imagery by measuring image contrast. *In:* H. Gong, Y. Cai, and J.-P. Chatard, eds., *Infrared components and their*

applications, SPIE-5640. Bellingham, MA: The International Society for Optical Engineering, 594–601.

Shi, J.-S., Jin, W.-Q., and Wang, L.-X., 2005b. Study on perceptual evaluation of fused image quality for color night vision. *Journal of Infrared and Millimeter Waves*, 24 (3), 236–240.

Sinai, M.J., McCarley, J.S., and Krebs, W.K., 1999. Scene recognition with infra-red, low-light, and sensor fused imagery. In: *Proceedings of the IRIS specialty groups on passive sensors*. Monterey, CA: IRIS, 1–9.

St. John, M., *et al*., 2001. The use of 2D and 3D displays for shape-understanding versus relative-position tasks. *Human Factors*, 43 (1), 79–98.

Sun, S., *et al*., 2005. Color fusion of SAR and FLIR images using a natural color transfer technique. *Chinese Optics Letters*, 3 (4), 202–204.

Toet, A., 2003a. Color the night: applying daytime colors to nighttime imagery. *In:* J.G. Verly, ed., *Enhanced and synthetic vision 2003*, SPIE-5081. Bellingham, WA, USA: The International Society for Optical Engineering, 168–178.

Toet, A., 2003b. Natural colour mapping for multiband nightvision imagery. *Information Fusion*, 4 (3), 155–166.

Tsagiris, V. and Anastassopoulos, V., 2005. Fusion of visible and infrared imagery for night color vision. *Displays*, 26 (4–5), 191–196.

Wang, L., *et al*., 2002. Color fusion schemes for low-light CCD and infrared images of different properties. *In:* L. Zhou, C.-S. Li, and Y. Suzuki, eds., *Electronic imaging and multimedia technology III*, SPIE-4925. Bellingham, WA: The International Society for Optical Engineering, 459–466.

Wang, L., *et al*., 2007. Real-time color transfer system for low-light level visible and infrared images in YUV color space. *In:* I. Kadar, ed., *Signal processing, sensor fusion, and target recognition XVI*, SPIE-6567-61. Bellingham, WA: The International Society for Optical Engineering, 1–8.

Wang, W. and Milgram, M., 2009. Viewpoint animation with a dynamic tether for supporting navigation in a virtual environment. *Human Factors*, 51 (3), 393–403.

Waxman, A.M., *et al*., 1995. Color night vision: fusion of intensified visible and thermal IR imagery. *In:* J.G. Verly, ed., *Synthetic vision for vehicle guidance and control*, SPIE-2463. Bellingham, WA: The International Society for Optical Engineering, 58–68.

Waxman, A.M., *et al*., 1999. Solid-state color night vision: fusion of low-light visible and thermal infrared imagery. *MIT Lincoln Laboratory Journal*, 11, 41–60.

Wickens, C.D. and Hollands, J.G., 2000. *Engineering psychology and human performance*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Wickens, C.D., Thomas, L.C., and Young, R., 2000. Frames of reference for the display of battlefield information: judgment-display dependencies. *Human Factors*, 42 (4), 660–675.

Yue, Z. and Topiwala, P., 2007. Real-time EO/IR sensor fusion on a portable computer and head-mounted display. *In:* B.V. Dasarathy, ed., *Multisensor, multisource information fusion: architectures, algorithms, and applications 2007*, SPIE-6571. Bellingham, WA: The International Society for Optical Engineering, 657106.

Zheng, Y., *et al*., 2005. Coloring night-vision imagery with statistical properties of natural colors by using image segmentation and histogram matching. *In:* R. Eschbach and G.G. Marcu, eds., *Color imaging X: processing, hardcopy and applications*, SPIE-5667. Bellingham, WA: The International Society for Optical Engineering, 107–117.