Color Image Fusion for Concealed Weapon Detection

Alexander Toet*
TNO Human Factors, Kampweg 5, 3769 DE Soesterberg, The Netherlands

ABSTRACT

Recent advances in passive and active imaging sensor technology offer the potential to detect weapons that are concealed underneath a person's clothing or carried along in bags. Although the concealed weapons can sometimes easily be detected, it can be difficult to perceive their context, due to the non-literal nature of these images. Especially for dynamic crowd surveillance purposes it may be impossible to rapidly asses with certainty which individual in the crowd is the one carrying the observed weapon. Sensor fusion is an enabling technology that may be used to solve this problem. Through fusion the signal of the sensor that depicts the weapon can be displayed in the context provided by a sensor of a different modality. We propose an image fusion scheme in which non-literal imagery can be fused with standard color images such that the result clearly displays the observed weapons in the context of the original color image. The procedure is such that the relevant contrast details from the non-literal image are transferred to the color image without altering the original color distribution of this image. The result is a natural looking color image that fluently combines all details from both input sources. When an observer who performs a dynamic crowd surveillance task, detects a weapon in the scene, he will also be able to quickly determine which person in the crowd is actually carrying the observed weapon (e.g. "the man with the red T-shirt and blue jeans"). The method is illustrated by the fusion of thermal 8-12 μm imagery with standard RGB color images.

Keywords: Image fusion, infrared, color, concealed weapons

1. INTRODUCTION

Concealed weapon detection is needed anywhere large concentrations of the public can be targeted for terrorist actions including airports, public office buildings, subways, schools, banks and malls. Unobtrusive detection of concealed weapons on persons or in (abandoned) bags would provide law enforcement a powerful tool to focus resources and increase traffic throughput in high- risk situations. A wide range of sensor systems is currently available, each with its own strength and weaknesses. No single sensor will completely satisfy the concealed weapon detection mission. Sensor fusion is an enabling technology that may increase the sensitivity, and reduce the number of false alarms and clutter by combining the signals of two or more sensors of different and complementary modalities[3-5,8,11,18]. Here we propose a new image fusion scheme that produces fused color images that appear similar to standard color images. The fused images are therefore easy to interpret for human observers, and are therefore highly suitable for surveillance purposes.

The fusion process proposed here involves the fusion of perceptually relevant contrast details from one or more non-literal images to the achromatic channel of a standard color image of the same scene. Since we don't want undesirable cross-channel artifacts in the resulting fused image (i.e. as a result we want a fused image that has the same colors as the original color image) we perform our computations in a perceptually decorrelated color space. This color space was recently presented as a useful tool for manipulating color images[9,15]. Since it is logarithmic, uniform changes are to a first approximation equally detectable.

In the next sections we will explain the method and show some results of the application of the method to the fusion of thermal 8-12 μm imagery with standard color imagery.

* E-mail: toet@tm.tno.nl

2. METHODS

The aim of the present study is to fuse thermal and color images such that the structural information from the greylevel thermal image is transferred to the color image, while the appearance of the resulting fused color image is identical to that of the original color image. We achieve this goal by applying a perceptually decorrelated color space that was recently introduced to enhance the color representation of synthetic imagery [9,15].

The pairs of input images used in this study consist of respectively a normal daylight RGB color photograph and a 8-12 μm thermal greylevel image of the same scene. In the following we assume that both images are registered. In practice, registration is easily achieved by first performing an affine transformation that warps one image onto the other, using fiducial registration points that are visible in both images[16].

The color fusion procedure is as follows. First, the source and target image are both transformed to a ***LMS*** cone response space. The different bands of daytime color images are usually correlated. Since we want to be able to transfer details from the thermal image to the color image, without changing the overall color distribution of the color image, we first need to transform the input imagery to a space which minimizes the correlation between channels. Therefore, through principal component analysis, we rotate the axes in the ***LMS*** cone space to achieve maximal decorrelation between the data points. The first principal component represents an achromatic channel, while the other two components correspond to color opponent channels[9,15]. Then, the achromatic component of respectively the daylight color image and the thermal image are merged though a pyramidal image fusion scheme. The mean and standard deviation of the resulting fused luminance image are set equal to those of the original color image, to preserve the look and feel of the original color image[9]. Finally, the fused image is transformed back to RGB space for display. The result is similar to the original daylight image but also includes perceptually relevant details from the thermal image.

In the following sections we first discuss the RGB to ***LMS*** transform. Second, we describe the principal component transform in ***LMS*** cone space that is applied to decorrelate the different channels. Third, we briefly discuss the pyramidal luminance image merging scheme. Finally, we will show that the inverse transform of the fused color image back to RGB space results in a color image that fluently combines all relevant details of both input images while it looks similar to the original color image.

2.1. RGB to *LMS* transform

First the RGB tristimulus values are converted to device independent XYZ tristimulus values. This conversion depends on the characteristics of the display on which the image was originally intended to be displayed. Because that information is rarely available, it is common practice to use a device-independent conversion that maps white in the chromaticity diagram to white in RGB space and vice versa[6].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

The device independent XYZ values are then converted to LMS space by

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2}$$

Combination of (1) and (2) results in

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3811 & 0.5783 & 0.0402 \\ 0.1967 & 0.7244 & 0.0782 \\ 0.0241 & 0.1288 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3}$$

The data in this color space shows a great deal of skew, which is largely eliminated by taking a logarithmic transform:

$$\begin{aligned} \boldsymbol{L} &= \log L \\ \boldsymbol{M} &= \log M \\ \boldsymbol{S} &= \log S \end{aligned} \tag{4}$$

The inverse transform from ***LMS*** cone space back to RGB space is as follows. First, the ***LMS*** pixel values are raised to the power ten to go back to linear LMS space. Then, the data can be converted from LMS to RGB using the inverse transform of Equation (3):

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.4679 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \tag{5}$$

2.2. Principal component transform

The principal component transform[7,10,12] effectively rotates the ***LMS*** coordinate axes such that the pixel components are maximally decorrelated. The set of normalized eigenvectors of the covariance matrix of the set of pixel values, arranged in order of increasing eigenvalues, constitute the column vectors of the corresponding rotation matrix. Let R be the rotation matrix that decorrelates the color image pixels. The pixel values of the color image and the thermal image in this new coordinate system are then respectively given by

$$\begin{bmatrix} \boldsymbol{L}_c' \\ \boldsymbol{M}_c' \\ \boldsymbol{S}_c' \end{bmatrix} = R \begin{bmatrix} \boldsymbol{L}_c \\ \boldsymbol{M}_c \\ \boldsymbol{S}_c \end{bmatrix} \tag{6}$$

and

$$\begin{bmatrix} \boldsymbol{L}_t' \\ \boldsymbol{M}_t' \\ \boldsymbol{S}_t' \end{bmatrix} = R \begin{bmatrix} \boldsymbol{L}_t \\ \boldsymbol{M}_t \\ \boldsymbol{S}_t \end{bmatrix} \tag{7}$$

where the indices c and t refer to the color and thermal images respectively.

2.3. Fusion of the achromatic image components

When combining information from the thermal image with the color image into a single display it is essential that the relevant contrast details of the individual images are preserved in the final image, and that no spurious pattern elements (that may interfere with subsequent analysis) are introduced by the merging process. We therefore applied a pyramidal image fusion scheme to merge the achromatic component of both input images[2,13,14,17]. A 7-level Laplacian pyramid[2] was used, in combination with a maximum absolute contrast node (i.e. pattern element) selection rule. This procedure ensures that perceptually relevant image details from all individual bands are represented in the final grayscale fused image.

An image pyramid is a collection of images at different spatial scales that together represent the original source image. Such a multi-resolution image representation can be obtained through a recursive *reduction* of the input image, i.e. a combination of low-pass or band-pass filtering and decimation.

The well-known Laplacian pyramid, introduced by Burt and Adelson[2], is a sequence of images in which each image is Laplacian filtered and subsampled copy of its predecessor. The construction of this pyramid is as follows. First, a Gaussian or low-pass pyramid is constructed. The original image is adopted as the bottom or zero-level G_0 of the Gaussian pyramid. Each node of pyramid level i ($1 \le i \le N$, where N is the index of the top level of the pyramid, i.e. the lowest resolution level) is obtained as a (Gaussian) weighted average of the nodes at level $i-1$ that are positioned within a 5×5 window centered on that node. Because of the reduction in spatial frequency content each image in the sequence can be represented by an array that has only half the dimensions of its predecessor. The process which generates each image in the sequence from its predecessor is called a REDUCE operation, since both the sample density and the resolution are decreased. Thus for $1 \le l \le N$ we have $G_l = \mathrm{REDUCE}\left[G_{l-1}\right]$, meaning

$$G_l(i,j) \;=\; \sum_{m,n=-2}^{2} w(m,n)\, G_{l-1}(2i+m, 2j+n) \tag{8}$$

where w represents a standard binomial Gaussian filter of 5×5 pixels extent [1]. A set of band-pass filtered images $L_0, L_1, \ldots, L_{N-1}$, that correspond to Laplacian of difference of low-pass filtered images, can be obtained by taking the difference of successive levels of the Gaussian pyramid. Since these levels differ in sample density it is necessary to interpolate new values between the given values of the lower frequency image before it can be subtracted from the higher frequency image. Interpolation is achieved simply by defining the EXPAND operation as the inverse of the REDUCE operation. Let $G_{l,k}$ be the image obtained by applying EXPAND to G_l k -times. Then

$$G_{l,0} = G_l \tag{9}$$

and

$$G_{l,k} = \mathrm{EXPAND}\left[G_{l,k-1}\right] \tag{10}$$

meaning

$$G_{l,k}(i,j) = 4 \sum_{m,n=-2}^{2} w(m,n)\, G_{l,k-1}\left(\frac{i+m}{2}, \frac{j+n}{2}\right) \tag{11}$$

where only integer coordinates contribute to the sum. The sequence L_i is then defined by

$$L_i = G_i - \mathrm{EXPAND}\left[G_{i+1}\right] \quad \text{for} \quad 0 \le i \le N-1 \tag{12}$$

and

$$L_N = G_N \tag{13}$$

Thus, every level is a difference of two levels in the Gaussian pyramid, making it equivalent to a convolution with a Laplacian-like band-pass filter. The Laplacian pyramid is a complete representation of the input image. G_0 can be recovered exactly by reversing the steps used in the construction of the pyramid:

$$G_N = L_N \tag{14}$$

and

$$G_i = L_i + \mathrm{EXPAND}\left[G_{i+1}\right] \quad \text{for} \quad 0 \le i \le N-1 \tag{15}$$

The Laplacian image fusion scheme is a three step procedure[2,13,14]. First, a Laplacian pyramid is constructed for each of the source images. Next, a Laplacian pyramid is constructed for the composite image by selecting or combining values from corresponding nodes in the component pyramids. Finally, the composite or fused image is recovered from its pyramid representation through the EXPAND and add reconstruction procedure.

Here we apply the Laplacian image fusion scheme to the achromatic components of the color image and the thermal image:

$$\boldsymbol{L}'_f \;=\; \mathrm{Fusion}\,[\,\boldsymbol{L}'_c\,,\boldsymbol{L}'_t\,] \tag{16}$$

where Fusion represents the Laplacian image fusion procedure.

2.4. Color image fusion

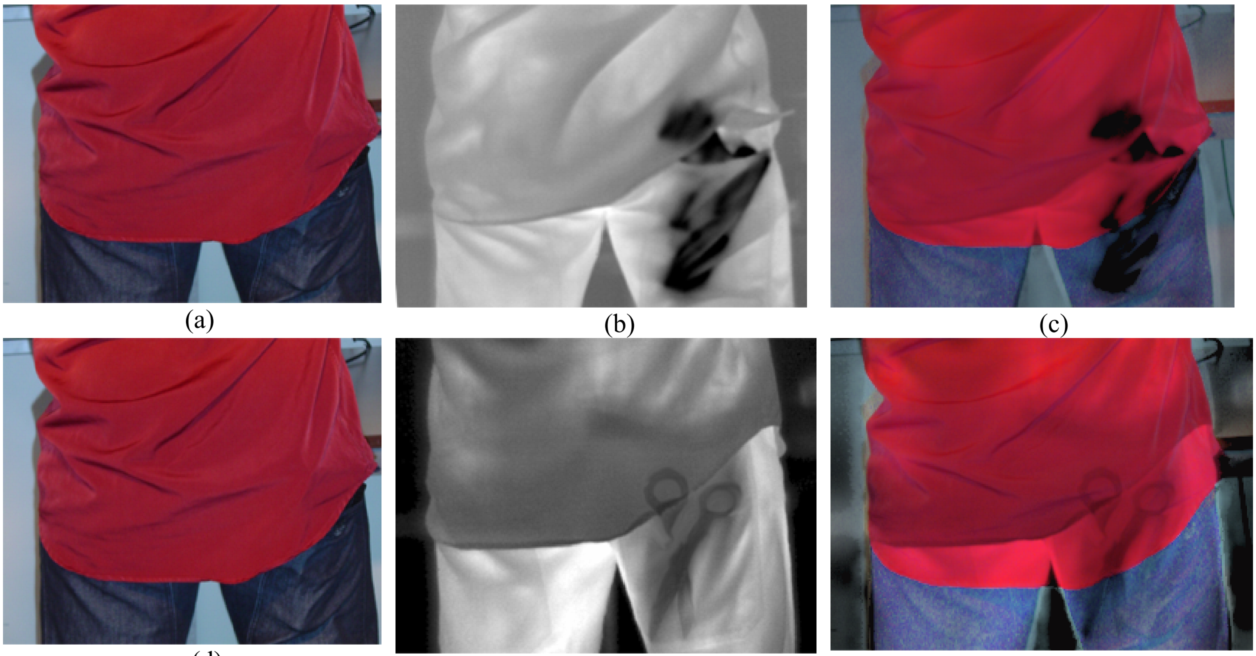
First we determine the first order statistics of the achromatic component of the original color input image. Let $\left\langle \boldsymbol{L}'_c \right\rangle$ and $\sigma_{\boldsymbol{L}'_c}$ represent respectively the mean and standard deviation of the achromatic band of this image. A fused color image is obtained by replacing the achromatic component of the original color input image by the result of the Laplacian image fusion procedure. The components of this fused image are given by $\left[\boldsymbol{L}'_f, \boldsymbol{M}'_c, \boldsymbol{S}'_c\right]$. To assure that the final color fused image will have the same appearance as the original color image[9,15], the first order statistics of $\boldsymbol{L}'_f$ are set equal to those of $\boldsymbol{L}'_c$:

$$\boldsymbol{L}''_f \;=\; \left[\boldsymbol{L}'_f - \left\langle \boldsymbol{L}'_f \right\rangle\right] \frac{\sigma_{\boldsymbol{L}'_c}}{\sigma_{\boldsymbol{L}'_f}} + \left\langle \boldsymbol{L}'_c \right\rangle \tag{17}$$

Finally, the resulting fused image $\left[\boldsymbol{L}''_f, \boldsymbol{M}'_c, \boldsymbol{S}'_c\right]$ is transformed back to RGB space for display, via the inverse rotation R^{-1}, logLMS, LMS, and XYZ color space using Equation (5).

3. EXAMPLES

Some examples of the fusion of thermal 8-12 μm images with standard daylight colour images are shown in Figures 1 and 2. Note that the concealed weapon is completely invisible in the visual images, whereas it is displayed at high contrast in the thermal imagery. The outline of the grip and the end of the barrel are clearly visible in Figures 1b and 2b,e,h. Features corresponding to the trigger assembly and the curve between the rear of the grip and the back of the barrel are also visible. Also, the contours of the pair of scissors can easily be perceived in Figure 1e. The combined or fused display of both image modalities yields images (Figs. 1c and f, and Figs. 2c,f and I) in which the concealed weapon is clearly depicted in the context of its local background.



(a) (b) (c)

(d) (e) (f)

Figure 1 Left column (a,d): original color photographs. Middle column (b,e): thermal 8-12 μm images. Right column (c,f): color fused results. Upper row: a Glock 17 9mm gun hidden in the backpocket of a pair of jeans, also partly covered with a polo shirt. Lower row: a hidden pair of scissors.

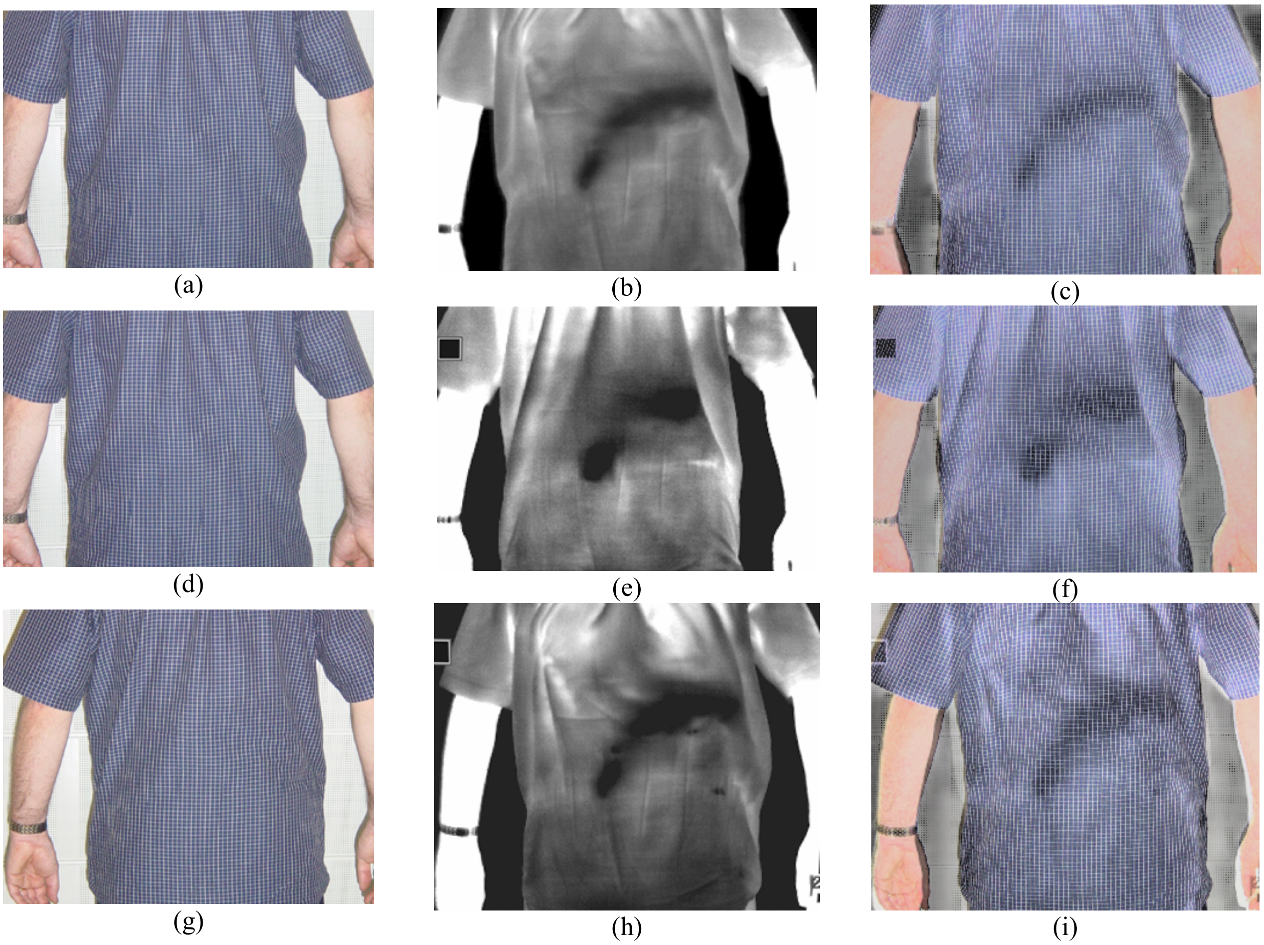


Figure 2 Left column (a,d,g): original color photographs. Middle column (b,e,h): thermal 8-12 µm images. Right column (c,f,i): color fused results. First row: Smith & Wesson 66 357 Magnum. Second row: Walther pp 7.65. Third row: Sigsauer p 226 9mm.

4. CONCLUSIONS

We presented a method to fuse non-literal (e.g. thermal) and color images such that the structural information from the non-literal greylevel image is transferred to the color image, while the appearance of the resulting fused color image is identical to that of the original color image. We achieve this goal by applying the transform in a perceptually decorrelated color space[9,15]. The method can be applied in standoff surveillance systems, since it is capable to clearly depict concealed weapons in the context of a full color image of the surveillance scene. The procedure is such that the relevant contrast details from the non-literal image are transferred to the color image without altering the original color distribution of this image. The result is a natural looking color image that fluently combines all details from both input sources. When an observer who performs a dynamic crowd surveillance task, detects a weapon in the scene, he will also be able to quickly tag the individual in the crowd that is actually carrying the observed weapon.

REFERENCES

1. Burt, P.J., The pyramid as a structure for efficient computation, In: A. Rosenfeld (Ed.), *Multiresolution image processing and analysis*, pp. 6-35, Springer, Berlin, GE, 1984.
2. Burt, P.J. and Adelson, E.H., Merging images through pattern decomposition, In: A.G. Tescher (Ed.), *Applications of Digital Image Processing VIII*, pp. 173-181, The International Society for Optical Engineering, Bellingham, WA, 1985.
3. Chen, H.-M., Varshney, P.K., Pramod, K., Uner, M.K. & Ramac, L.C. *Sensor fusion algorithms and performance limits,* (Report A539193), Electrical Engineering and Computer Science Department, Universtity of Syracuse, New York, (2001).
4. Currie, N.C., Demma, F.J., Ferris, D.D., McMillan, R.W. and Wicks, M.C., ARPA/NIJ/Rome Laboratory concealed weapon detection program: an overview, In: I. Kadar & V. Libby (Ed.), *Signal processing, sensor fusion, and target recognition*, pp. 492-502, The International Society for Optical Engineering, Bellingham, WA., USA, 1996.
5. Currie, N.C., Demma, F.J., Ferris, D.D., McMillan, R.W., Wicks, M.C. and Zyga, K., Imaging sensor fusion for concealed weapon detection, In: I.K. Rudin & S.K. Bramble (Ed.), *Investigative image processing*, pp. 71-81, The International Society for Optical Engineering, Bellingham, WA., USA, 1997.
6. Fairchild, M.D., *Color appearance models* , Addison Wesley Longman, Inc., Reading, MA, 1998.
7. Hall, E.L., *Computer Image Processing* , Academic Press, New York, USA, 1979.
8. Ramac, L.C., Uner, M.K., Varshney, P.K., Alford, M.G. and Ferris, D.D., Morphological filters and wavelet-based image fusion for concealed weapons detection, In: B.V. Dasarathy (Ed.), *Sensor Fusion: Architectures, Algorithms, and Applications II*, pp. 110-119, The International Society for Optical Engineering, Bellingham, WA, 1998.
9. Reinhard, E., Ashikhmin, M., Gooch, B. and Shirley, P., Color transfer between images, *IEEE Computer Graphics and Applications,* 21*(5) ,*pp. 34-41, 2001.
10. Richards, J.A., *Remote sensing digital image analysis* , Springer Verlag, Berlin, 1986.
11. Slamani, M.A., Rmac, L., Uner, M.K., Varshney, P.K., Weiner, D.D., Alford, M., Ferris, D.D. and Vannicola, V., Enhancement and fusion of data for concealed weapon detection, In: I. Kadar (Ed.), *Signal processing, sensor fusion, and target recognition VI*, pp. 8-19, The International Society for Optical Engineering, Bellingham,WA, 1997.
12. Taylor,P. (1999). Statistical methods. In: M.Berthold & D.J.Hand (Eds.), *Intelligent data analysis.* (pp. 67-127). Berlin, GE: Springer Verlag.
13. Toet, A., Image fusion by a ratio of low-pass pyramid, *Pattern Recognition Letters,* 9 ,pp. 245-253, 1989.
14. Toet, A., Hierarchical image fusion, *Machine Vision and Applications,* 3 ,pp. 1-11, 1990.
15. Toet, A. *Paint the night: applying daylight colours to nighttime imagery,* (Report TM-02-B006), TNO Human Factors, Soesterberg, The Netherlands, (2002).
16. Toet, A. and IJspeert, J.K., Perceptual evaluation of different image fusion schemes, In: I. Kadar (Ed.), *Signal Processing, Sensor Fusion, and Target Recognition X*, pp. 436-441, The International Society for Optical Engineering, Bellingham, WA, 2001.
17. Toet, A., Ruyven, J.J. and Valeton, J.M., Merging thermal and visual images by a contrast pyramid, *Optical Engineering,* 28 ,pp. 789-792, 1989.
18. Uner, M.K., Ramac, L.C., Varshney, P.K. and Alford, M.G., Concealed weapon detection: an image fusion approach, In: I.K. Rudin & S.K. Bramble (Ed.), *Investigative Image Processing*, pp. 123-132, The International Society for Optical Engineering, Bellingham, WA., USA, 1997.