TNO report
FEL-97-B046

Fusion of IR and visual images

TNO Physics and Electronics Laboratory

Oude Waalsdorperweg 63
PO Box 96864
2509 JG The Hague
The Netherlands

Phone +31 70 374 00 00
Fax +31 70 328 09 61

Date
February 1997

Author(s)
Dr K. Schutte

Classification

Classified by	:	Drs. C.W. Lamberts
Classification date	:	4 February 1997

Title	:	Ongerubriceerd
Managementuittreksel	:	Ongerubriceerd
Abstract	:	Ongerubriceerd
Report text	:	Ongerubriceerd
Appendix A	:	Ongerubriceerd

Copy no	:	7	
No of copies	:	38	
No of pages	:	27	(incl appendix, excl RDP & distribution list)
No of appendices	:	1	

All information which is classified according to Dutch regulations shall be treated by the recipient in the same way as classified information of corresponding value in his own country. No part of this information will be disclosed to any party.

The classification designation Ongerubriceerd is equivalent to Unclassified, Stg. Confidentieel is equivalent to Confidential and Stg. Geheim is equivalent to Secret.

The TNO Physics and Electronics Laboratory is part of TNO Defence Research which further consists of:

TNO Prins Maurits Laboratory
TNO Human Factors Research Institute

Netherlands Organization for Applied Scientific Research (TNO)

Managementuittreksel

Titel	:	Fusion of IR and visual images
Auteur(s)	:	Dr. K. Schutte
Datum	:	februari 1997
Opdrachtnr.	:	-
IWP-nr.	:	766.2
Rapportnr.	:	FEL-97-B046

Beeldfusie (image fusion) is het proces waarbij beelden opgenomen met meerdere sensoren samengevoegd worden tot één enkel beeld. Het is één van de vele mogelijkheden binnen het bredere veld van sensorfusie. Dit rapport beantwoordt vragen die horen bij het fuseren van twee soorten beelden, m.n. visueel licht en thermisch infrarood, tot één enkel grijswaardebeeld. Hiertoe zijn verschillende beeldseries opgenomen en verwerkt met verschillende methodieken.

Voor het fuseren van de beelden is een optimaliteitscriterium gedefinieerd. Afhankelijk van de aanname of de ingangsbeelden afhankelijk dan wel onafhankelijk zijn blijkt dat de optimale oplossing voor het fusieprobleem een gewogen optelling is of een piramidale oplossing zoals het algoritme van Burt. Voor beide methodes is afgeleid hoe de ingangsbeelden gewogen dienen te worden om een optimaal resultaat te verkrijgen.

De resultaten van de beeldfusie zoals gepresenteerd in dit rapport geven aan dat (voor de gegeven beelden) het gefuseerde beeld meer informatie oplevert dan de afzonderlijke beelden. Hierbij dient het voorbehoud gemaakt te worden dat de gepresenteerde beelden speciaal uitgezocht zijn; in de praktijk levert vaak één van de sensoren beduidend meer informatie dan de andere, waardoor het gebruik van enkel die sensor eigenlijk alle relevante informatie levert.
Daarnaast biedt beeldfusie de mogelijkheid om ook informatie welke zich in het infraroodbeeld bevindt af te beelden op een manier zoals die aansluit bij de perceptie van normale visuele beelden.

Hiernaast is bekeken met wat voor hardware dergelijke systemen tegenwoordig gerealiseerd kunnen worden. Het blijkt dat dit mogelijk moet zijn met de nieuwste DSP-boards welke als plug-in in een standaard-PC geplaatst kunnen worden.

Een eerste aanzet is gegeven naar het verrichten van Automatic Target Recognition (ATR) op basis van meerdere input-beeldseries. De resultaten hiervan zien er veelbelovend uit.

Contents

1. Introduction

Many systems are, or could be, equipped with both a visual CCD camera and a thermal infrared (IR) camera. The motivation for this is that both cameras can provide complementary information. When a system is equipped with such a dual camera system a choice must be made how the data acquired should be presented for the operator. Two basically different methods can be used: separate presentation and combined presentation. This report is about the latter.

1.1 Separate presentation

With separate presentation the data acquired with both sensors are independently presented to the operator. In our case of two image sensors this means that two separate images are presented, or both are subsequently imaged using the same display device.
Both presentation methods mentioned above have their problems. Displaying two images simultaneously means that the amount of hardware needed for that task must be larger. Subsequent imaging on the same device means that intervention of an operator is needed to switch. Both methods share the problem that the operator needs to divide his attention between both images.
These problems make that separate presentation of both sensor inputs in not ideal.

1.2 Combined presentation

With combined presentation *sensor fusion* is applied to the sensor input before they are presented to the operator. Within this report we will concentrate upon *image fusion*. This is a special kind of sensor fusion where the output of the fusion operation also is an image similar to the inputs.
Ideal, the combined presentation should exhibit the following properties:

1. All relevant information present in the raw sensor data is present in the combination.
2. The combined presentation looks natural, such that the operator does not have to convert the output of the system to understandable items.
3. No exotic hardware is needed to calculate or visualise the combined presentation.

As shown in the remainder of this report, these properties are not easily to achieve.

1.3 Overview

Chapter 2 describes the mathematical foundation of the image fusion processes. Chapter 3 describes a few ideas how to solve the same problem using multi colour representation. Experimental results are given in chapter 4. Chapter 5 describes the computational complexity of the algorithms used in image fusion. Finally, conclusions and recommendations based upon the presented material are given in chapter 6.

2. Fusion methods

This chapter describes what methods are available to perform image fusion. Image fusion is the process where two input images of similar modality are transformed to one output image of the same modality. In this chapter we will first discuss image alignment en warping. After that we set an optimality criterion for image fusion. Based upon that criterion, image blending and pyramid based image fusion algorithms are presented.

2.1 Image alignment

When two images are taken from a three-dimensional world, those images can only be overlapping when the optical centres of the two cameras are co-located. When this is not the case, different distances to the cameras will result in different disparities. When optical centres cannot be co-located (due to the fact that two sensors cannot be present at the same location), a similar solution can be achieved by using dichroic mirrors. In practice this might prove difficult; partial compensation can be achieved by applying image warping.

Besides the need to co-locate the optical centres of the cameras, the sensors should also be aligned such that the optical axes are parallel. The pixel grid of both imagers should also be aligned such that each pixel corresponds to exactly the same position in both images. This can be achieved with very similar sensors such as RGB cameras; but this will prove very hard with dissimilar apparatus such as a visual and a IR imager. However, these errors can be completely corrected using image warping.

2.2 Image warping

As shown above, prerequisite to the image fusion algorithms is that the images are *warped*. Warping is a geometric process where the image pixels are moved such that after the warping operation in both images pixels at the same image co-ordinates refer to the same location in the real world. In this report we use an affine transformation combined with bilinear interpolation. Image warping, also known as geometrical transformation, is not in scope of this report. For a detailed description of image warping see textbooks such as [1] chapter 17.3-17.4, and [2] chapter 5.9.

For the ease of understanding, in the rest of the report it is assumed that the images are warped appropriately unless noted otherwise.

2.3 Optimal fusion algorithms

In this section, conditions for optimal fusion are derived. With optimal we mean the notion that we optimise the signal to noise ratio. The outcome of the derivation depends on the assumptions made; for the following assumptions the derivation is shown:

- Identical signal in both sensors, independent, additive noise
- Independent signal in the sensors with independent, additive noise

2.3.1 Optimal fusion with identical signal

We model the signal derived from sensor one as:

$$\underline{v}_1 = s_1 + o_1 + \underline{n}_1 \tag{2.1}$$

with s_1 the signal of interest, o_1 a constant offset, and $\underline{n}_1$ stochastic noise. Similar, we have for sensor 2:

$$\underline{v}_2 = s_2 + o_2 + \underline{n}_2 \tag{2.2}$$

As we assume that the two signals s_1 and s_2 are identical, we can write:

$$s_2 = a.s_1 \tag{2.3}$$

We assume that the noise is zero mean white noise; then we can write:

$$\sigma_{n_2} = b.\sigma_{n_1} \tag{2.4}$$

with σ the standard deviation of the noise. We look for a solution of the type:

$$\underline{v} = \underline{v}_1 + c.\underline{v}_2 \tag{2.5}$$

which under the assumption of independent noise leads to:

$$\underline{v} = s_1(1+c.a) + \sqrt{1+c^2b^2}\,\underline{n}_1 + o_1 + c.o_2 \tag{2.6}$$

Using formula (2.6), we can give the signal to noise ratio (SNR):

$$\mathrm{SNR} = \frac{s_1}{\underline{n}_1}\frac{1+ca}{\sqrt{1+c^2b^2}} \tag{2.7}$$

Our optimality criterion is to optimise this SNR; this means finding the optimal c given a and b:

$$c_{opt} = \max_c \frac{s_1}{\underline{n}_1}\frac{1+ca}{\sqrt{1+c^2b^2}} = \max_c \frac{1+ca}{\sqrt{1+c^2b^2}} = \max_c \frac{1+2ca+c^2a^2}{1+c^2b^2} \tag{2.8}$$

We can find the extrema of the formula by calculating the derivative to c and setting that to zero; this leads to:

$$c = -\frac{1}{a} \vee c = \frac{a}{b^2} \tag{2.9}$$

The first solution is a minimum, which is not of interest. The second solution is the maximum we are looking for. The meaning of this solution is that you should compensate for multiplication factor a and a noise weigh factor $1/b^2$. Application of this derivation can be found in section 2.4.

2.3.2 Optimal fusion with independent signal

Using similar definitions as in the previous section, under the assumption of independent signals for both sensors we have a different relation between the two signals. Here we model the signal by its standard deviation; the relation between the signals of the two sensors is:

$$\sigma_{s_2} = \sqrt{a}.\sigma_{s_1} \tag{2.10}$$

For the noise in both sensors we use:

$$\sigma_{n_2} = \sqrt{b}.\sigma_{n_1} \tag{2.11}$$

Looking for a solution using to formula (2.5), we get:

$$\underline{v} = s_1\sqrt{1+ac} + \underline{n}_1\sqrt{1+bc} + o_1 + co_2 \tag{2.12}$$

Again, we choose c such that we optimise the SNR:

$$c_{opt} = \max_c \frac{s_1}{\underline{n}_1}\frac{\sqrt{1+ac}}{\sqrt{1+bc}} = \max_c \frac{\sqrt{1+ac}}{\sqrt{1+bc}} = \max_c \frac{1+ac}{1+bc} \tag{2.13}$$

Using the constraint that $a,b,c \geq 0$, we obtain:

$$c_{opt} = \begin{cases} c=0 & | \quad a<b \\ c=\infty & | \quad a>b \end{cases} \tag{2.14}$$

which essentially means: use the image with the best signal to noise ratio.

2.4 Blending

Blending is the process where we use a weighted summation of the two sensors input to obtain the output:

$$v_{out} = f.v_{in1} + (1-f).v_{in2} \tag{2.15}$$

which is similar to formula (2.5) using $f = 1/(c+1)$. Under the assumption of identical signal, we should use the second term of formula (2.9), which leads to:

$$f = \frac{b^2}{a+b^2} \tag{2.16}$$

with a the signal ratio between the two inputs and b^2 the ratio of the variance of the noise between the two inputs.

2.5 Burt method

With the assumption of independent signals, the image with the best SNR should be chosen. However, suppose we can decompose both images into independent parts. In that case, the result derived in section 2.3.2 can be applied to these independent parts. The algorithm developed by Burt [3] performs this with a pyramid decomposition of the image as depicted in Figure 2.1. At each pyramid level, the Burt algorithm separates the low and the high frequency components:

$$L_k = \mathrm{L}[I_k] \tag{2.17}$$

$$H_k = I_k - L_k \tag{2.18}$$

with L[] the convolution operator with a separated kernel and as kernel elements $[\frac{1}{4}-\frac{a}{2},\frac{1}{4},a,\frac{1}{4},\frac{1}{4}-\frac{a}{2}]$ where we used $a=\frac{3}{8}$. The next level is a factor 2 reduced version of the previous low frequency component:

$$I_{k+1} = \mathrm{R}[L_k] \tag{2.19}$$

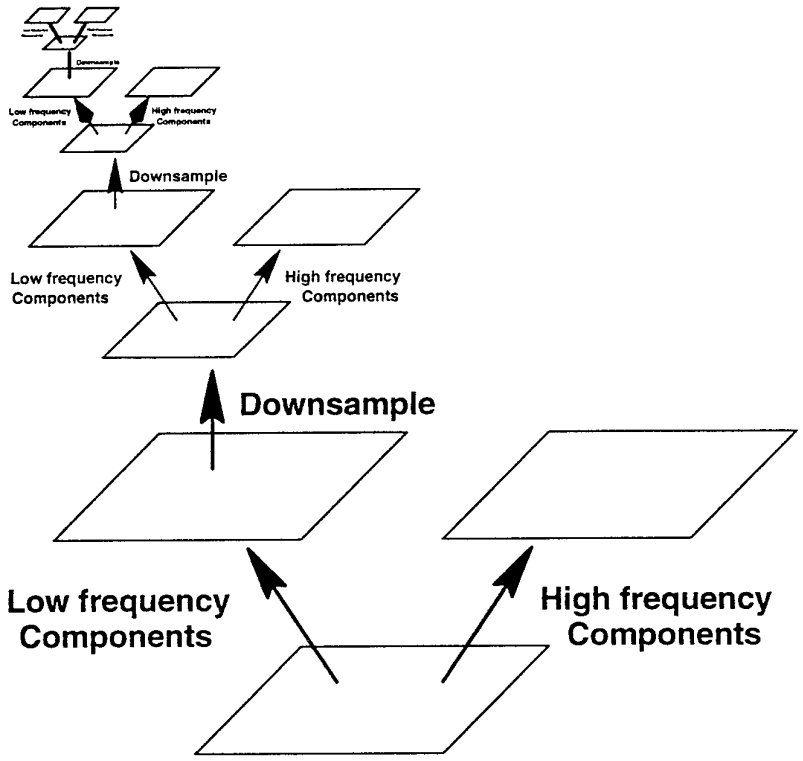


Figure 2.1: Pyramid decomposition.

with R[] the reduction operator. As the new image level is at a different scale, the definitions of high and low frequencies have changed; this means that the separation in different frequency components can again be applied. The result of the pyramid decomposition is a set of high frequency component images plus the highest level low frequency component:

$$P = (H_0, H_1, \ldots, H_k, L_k) \tag{2.20}$$

The total amount of memory needed to store the pyramid decomposition of an image is slightly (≈ 1.3 times) larger then the original image. The meaning of each high frequency component H_k is that it represents the information present in the image at scale k; every pixel $h_k(i,j)$ within this image represents the local structure of scale k at its location. Under the (not completely realistic) assumptions that the low pass filtering applied in formula (2.17) is an ideal low pass filter at Nyquist rate for the reduced next level, and the input image is band delimited at Nyquist rate, the components H_k can be seen as frequency bands of the original image.

Using the pyramid decomposition the Burt algorithm constructs the output image as follows: first, combine the highest level by blending:

$$L_{f,n} = f.L_{1,n} + (1-f).L_{2,n} \tag{2.21}$$

Then, for each level, construct the fused high frequency component:

$$h_{f,k}(i,j) = \begin{cases} h_{1,k}(i,j) & | \; |h_{1,k}(i,j)| > |h_{2,k}(i,j)| \\ h_{2,k}(i,j) & | \; |h_{1,k}(i,j)| < |h_{2,k}(i,j)| \end{cases} \tag{2.22}$$

which is performed for each pixel. The fused image level is reconstructed by:

$$I_{f,k} = L_{f,k} + H_{f,k} \tag{2.23}$$

The previous level low frequency image is constructed by:

$$L_{f,k-1} = \mathrm{E}[I_{f,k}] \tag{2.24}$$

with E[] the expansion operator. This is repeated until the fused image level 0 is formed.

2.5.1 Reduce-expand operation

Actually, the operation described above is not completely accurate. It is under the assumption that the result of E[R[I]] equals I. The problem is that the expansion operator is implemented single pixel replication with zero filling followed by low pass filtering with the L[] operator. This is a non-ideal low pass filter which results in artefacts. The solution to this problem is to replace formula (2.18) with:

$$H_k = I_k - \mathrm{E}[\mathrm{R}[L_k]] \tag{2.25}$$

2.5.2 The optimality criterion applied to the Burt algorithm

In the Burt algorithm as described above, no use has been made of the optimality constraint as obtained in section 2.3.2 for independent sensor signals. If we assume that the pixel value $h_{n,k}(i,j)$ represents the local signal near (i,j) for scale k, we should change formula (2.22) to:

$$h_{f,k}(i,j) = \begin{cases} h_{1,k}(i,j) & | \; |h_{1,k}(i,j)| > \dfrac{|h_{2,k}(i,j)|}{\sqrt{b}} \\ \dfrac{|h_{2,k}(i,j)|}{\sqrt{b}} & | \; |h_{1,k}(i,j)| < \dfrac{|h_{2,k}(i,j)|}{\sqrt{b}} \end{cases} \tag{2.26}$$

with b as used in formula (2.11). This can easily be realised by scaling the images first to their expected noise levels; this is done by dividing image two by $\sqrt{b}$ before starting with the Burt algorithm.

2.6 Toet method

The Toet algorithm [4] is very similar to the Burt algorithm. The difference is that the ratio of the low pass images is used instead of the difference. In formulas, this means to replace formula (2.25) by:

$$C_k = \frac{I_k}{\mathrm{E}[\mathrm{R}[L_k]]} \tag{2.27}$$

and formula (2.23) by:

$$I_{f,k} = C_{f,k} \cdot L_{f,k} \tag{2.28}$$

C_k is related to the local contrast. It can be argued that using this is appropriate when signals are not additive but multiplicative.

2.7 Wavelet methods

Wavelet methods are also a way to decompose image into localised scale specific signals [5][6][7]. Within the family of wavelet methods many different decompositions of an image are possible. In fact, the Burt algorithm can be seen as a Wavelet method.

An initial study into this subject [7] showed that, within the wavelets examined, no better performance is found than achieved with the Burt algorithm. It is likely that the best choice of the exact wavelet to be used is dependent on the application at hand. Based on current knowledge it is probably safe to state that for general applications no wavelet can perform significantly better than the Burt algorithm.

3. Multi-colour image presentation

The image fusion methods described in the previous chapter worked under the condition that the output image is of the same modality as the input images. In this chapter we will discuss methods where the result of two monochrome input images is a colour image.

3.1 False colour presentation

The idea with false colour presentation is that the two input images both are assigned to an image band. For this, we should define three functions:

$$\begin{aligned} r(i,j) &= f_r(v_1(i,j), v_2(i,j)) \\ g(i,j) &= f_g(v_1(i,j), v_2(i,j)) \\ b(i,j) &= f_b(v_1(i,j), v_2(i,j)) \end{aligned} \tag{3.1}$$

where the triplet r,g,b is used as the red, green and blue values of the pixel involved. Ideally, the colour mapping functions should be chosen such that:

1. Interesting objects (humans, vehicles, missiles, ...) should be contrasting to their background.
2. Objects are represented in the same colour as percepted in real life: the sky should be blue, vegetation green, etc.

The first item has as a consequence that the mapping scheme used is application dependent. Also, it means that some knowledge about the interesting objects should be used to be able to let these objects stand out.
The second item is quite easy to accomplish for one weather condition at a certain time of the day. However, it is not easy to satisfy this requirement for all possible weather conditions.

The results shown in the next chapter used a simplified approach of the MIT colour image fusion scheme as described in [8]. The simplification is that the image enhancement used in the original is not applied. The specific mapping function used is:

$$\begin{aligned} r(i,j) &= \frac{v_1(i,j) + v_2(i,j)}{2} \\ g(i,j) &= v_1(i,j) \\ b(i,j) &= \mathrm{CLIP}[v_1(i,j) - v_2(i,j)] \end{aligned} \tag{3.2}$$

where v_1 is the visual image, v_2 is the IR image, and CLIP[] an operator which sets negative values to zero.

3.2 Target cueing

In the previous section the remark was made that for good contrast between foreground and background objects object knowledge is needed. With target cueing this is used even more; target cueing is a technique where first autonomously by the system the interesting objects (further called targets) are recognised, after which they are projected or exaggerated in the image. An example of this technique is to visualise hotspots in the IR image as red dots in the visual image.

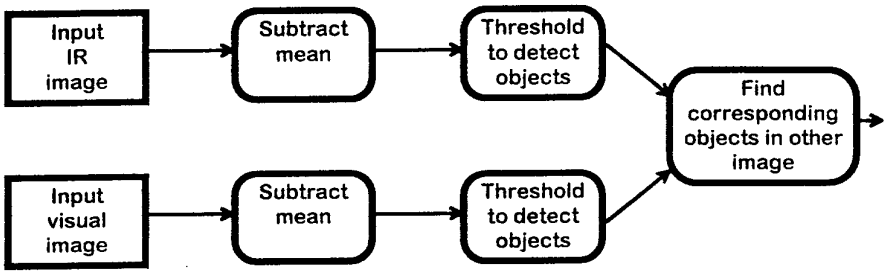


Figure 3.1: Target cueing algorithm used.

Figure 3.1 shows the block diagram of the target cueing algorithm used for the results presented in the next chapter. First, from both images the temporal mean is subtracted. For real-time results, this should be a moving average; for the presented results the mean over the whole sequence is used. Second, both images are thresholded with an image specific threshold. For those pixels which are candidate objects, we search in the spatial neighbourhood for a corresponding object in the other image. If those are found, this blob is labelled as ***BOTH***. Otherwise, depending on whether the blob occurred in the visual or the IR image, the blob is labelled as ***VISUAL*** or ***IR***. Each found blob is depicted as a cross in the results given in the next chapter. ***BOTH, VISUAL*** and ***IR*** are colour coded as red, blue and green respectively.

4. Results

In this chapter we show results for the fus13, ms01 and UNcamp sequences. The fus13 sequence is recorded near TNO-FEL looking to a road with passing (civil) vehicles. The ms01 sequence, provided by Defense Research Establishment Valcartier [9], shows a vehicle and a helicopter over a battlefield with humans and a smoke field blocking visual imaging. The UNcamp sequence, provided by TNO-TM [8], shows a scene representative for situations found when guarding a UN camp. Actually, this sequence is recorded from the TNO-FEL tower, just as the fus13 sequence.

4.1 fus13 sequence

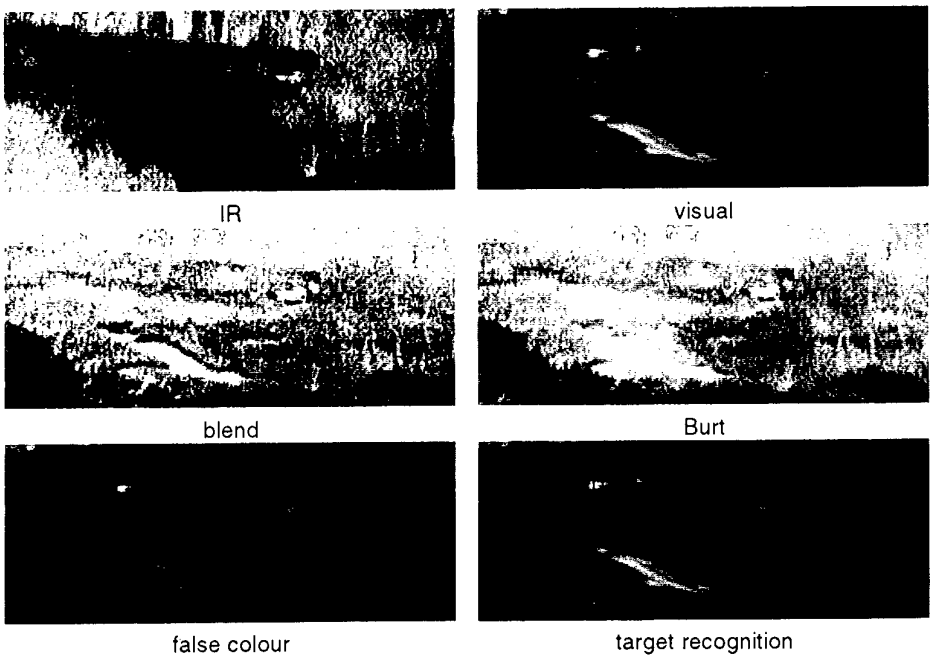


Figure 4.1: Image 21 from the fus13 sequence.

Figure 4.1 depicts the results obtained for image 21 of the fus13 scene. Shown is only a smaller part of the full frame. The IR image is strong in showing the left vehicle, and the visual image is strong in showing the right vehicle. All fusing methods are capable of emphasising both vehicles. The target recognition image shown right bottom shows objects found in both channels in red, and IR and visual alarms only in green respectively blue. For the full sequence, using sensor fusion reduced the false alarm rate from 88% (IR) respectively 161% (visual) to 10% for the combination. Full details for the target recognition experiment can be found in appendix A. The false colour image shows the trees as red; the fact that they are not green is probably related to the time of the day the image is taken, early in the morning.

4.2 ms01 sequence

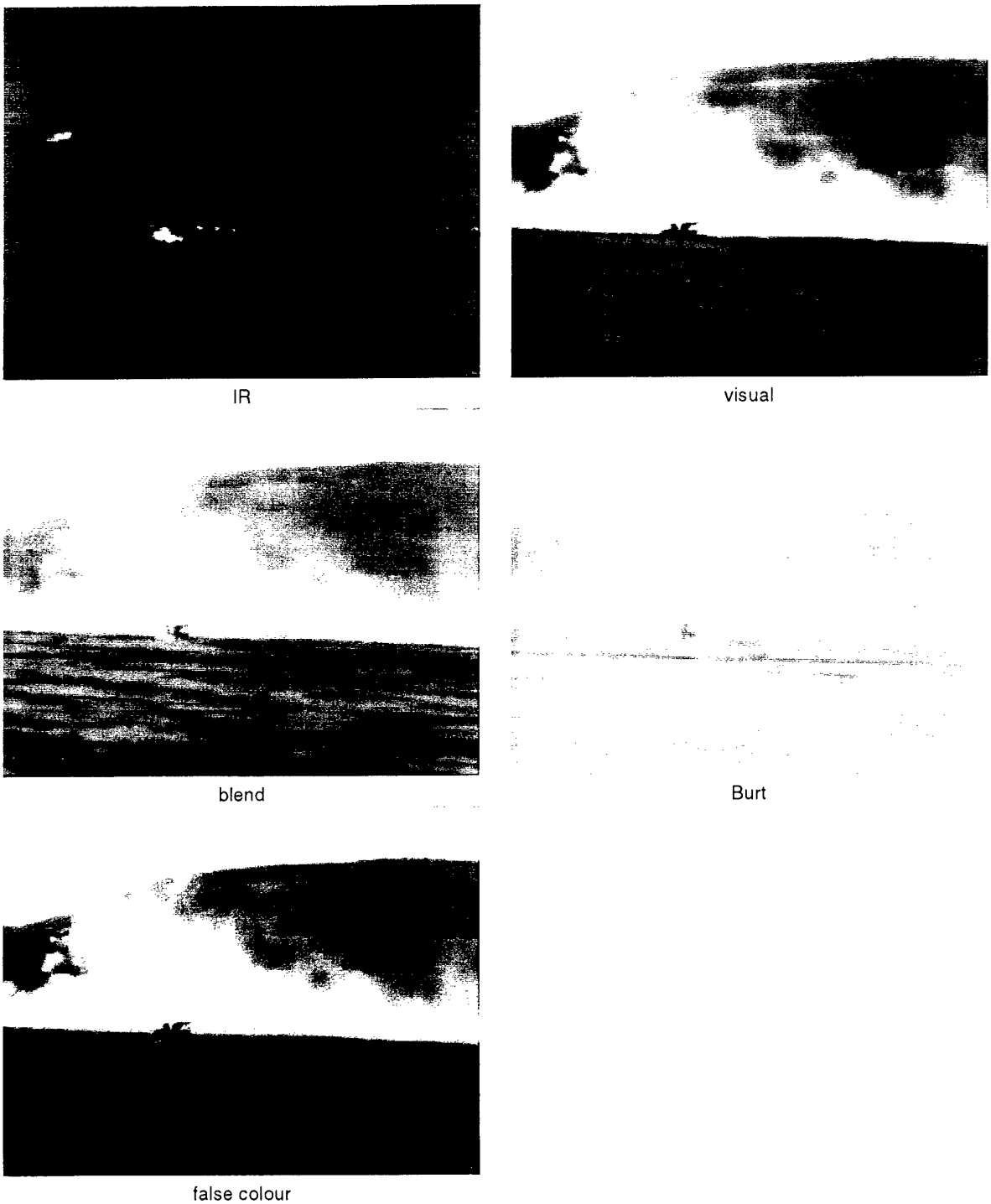


Figure 4.2: Image 73 from the ms01 sequence.

Figure 4.2 shown image 73 from the ms01 sequence. The IR image shows the helicopter, vehicle and men, where the visual image depicts the helicopter, vehicle and the smoke. Also, the mountain in the background is much clearer in the visual image. The fused images show all interesting details. The exact shape of the vehicle is within the fused images the best recognisable in the false colour image.

4.3 UNcamp sequence

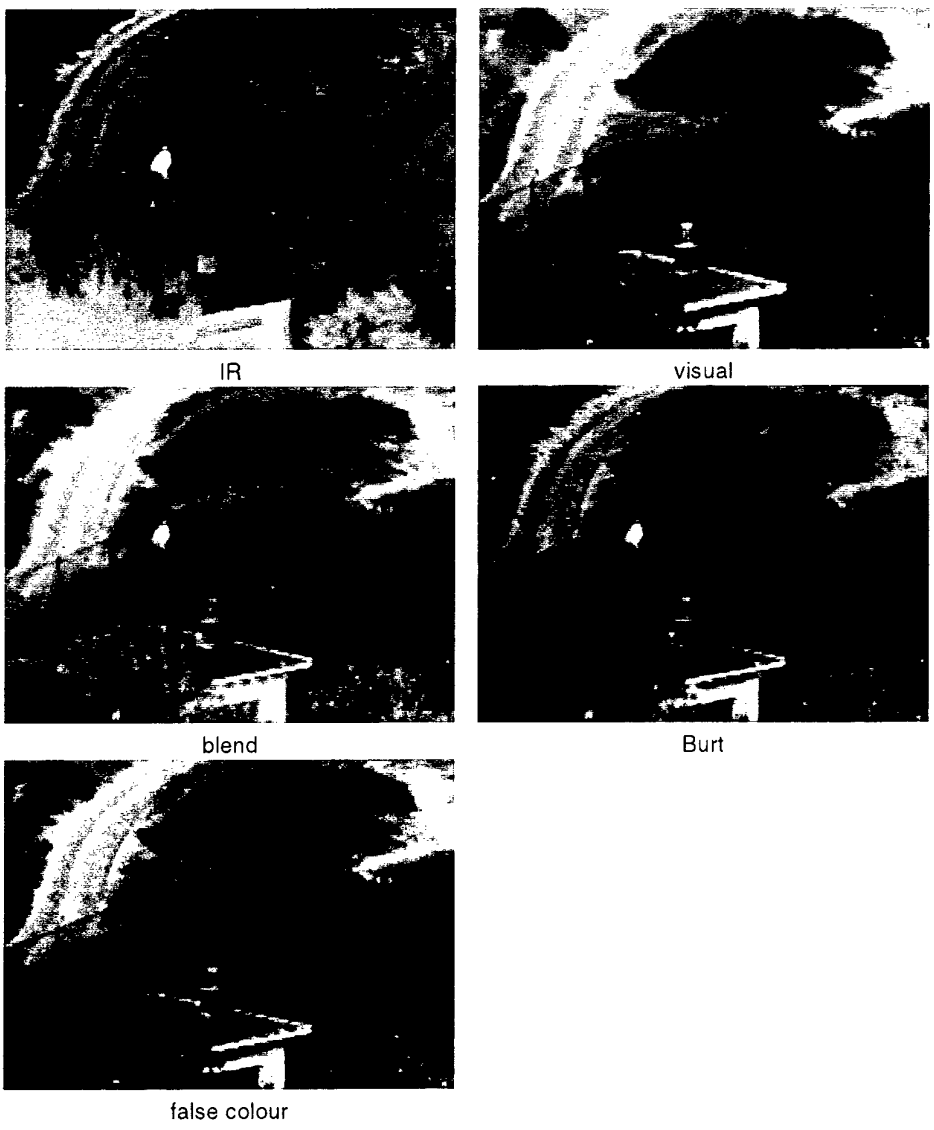


Figure 4.3: Image 1 from the UNcamp sequence.

Figure 4.3 shows the first image of the UNcamp sequence. The IR image depict the man, and the visual image the gate. On the fused images both the gate and the man are visible. This image is also used in a recent report of TNO Human Factors

Research Institute by Toet, IJspeert and van Dorresteijn [8]. The 'blend' and 'false colour' images in this report are quite similar to the images obtained with the MIT method given in that report, although the contrast enhancement utilised by the MIT method results in a brighter results.

4.4 Temporal noise

Not visible on the separate images, and not easy to visualise in a report like this, is the nature of the temporal noise present in the sequences. The noise behaviour of both the blend images and the Burt images has been studied. When characterised by standard deviation, there is little difference between the temporal noise observed. However, with visual inspection it can clearly be seen that there is a difference in temporal noise for the blend method and the Burt method. On some parts of the image the sequence produced by the Burt method seems to flicker, which is annoying to human perception.

The reason for this flicker lies in the nature of the Burt method. In equation (2.22) and equation (2.26) we select the IR or the visual signal depending on which has the highest absolute value. The absolute value is dependent on the temporal noise in the channel, and thus it is possible that based upon small temporal noise changes in a temporal sequence sequentially the IR and the visual signal can be selected. When, for the spatial frequency band under consideration, the IR and visual signals have an opposite sign, the selection change between the two bands results in flicker, visible when viewing the sequence.

Separate from this report a CD ROM will be produced which contains the complete sequences mentioned in this report and the processing results. This will give an idea of the temporal behaviour of the methods discussed in this report.

4.5 Scenario matching

In this chapter some image sequences are given in which image fusion improves the perception of the scene imaged. However, it must be kept in mind that these image sequences are specially selected for this report. Under normal operational circumstances one of the two sensors (often the IR camera) will give a superior image over the other sensor. If this sensor's image contains all information about the scene relevant to the current scenario, including data from the other sensor will not add information to the operator. This leads to the question which sensor or image fusion scheme must be used with a given scenario and weather conditions.

4.6 Visual appearance

In the preceding part of this chapter, the evaluation of the image was based on information content only. However, under operational circumstances also essential are:

- Present relevant information only, or at least emphasized.
- The resulting image should be easily interpretable for the user.

In general, the first point needs information about the current task to be performed to separate relevant information and irrelevant information. In the results presented this has been performed with the target recognition approach, where the possible targets are emphasized with a red marker. For a user of the images who is not well trained in the recognition of IR images, the fused images are more easily to interpret. Especially the false colour images are quite close to normal perception.

5. Computational aspects of image fusion

For the research described in this report in 1994 the Sensar VFE-100 image processor was obtained by TNO-FEL. It turned out that the VFE-100 was very useful in the process of getting feeling with image fusion. The image processor enables real-time processing for this kind of problems, and as such proved to be very valuable. Although the VFE-100 is capable of most of the algorithms described in this report, it is not utilised in this research for the following reasons:

- The VFE-100 is designed to do all the processing from input to display. Due to the fact that digital I/O is not (reliable) possible, it is very hard to use such a machine in a test-and-improve environment.
- Programming a machine specially designed for a specific task is much harder then a general purpose computer such as a current workstation. Since research on these topics include a lot of trail and error, the additional effort needed to program the VFE-100 seemed not worth the effort.
- The display of the VFE-100 is not capable of displaying images in colour.
- Many algorithms applied need floating point numbers. The VFE-100 can only handle 8 bit integers. This can result in unwanted rounding errors.

Similar problems will also yield in the future. To overcome these problems a fast implementation on a general purpose processor is preferably. With the increase of processor speed, realisation on generic computer architectures should be possible in the near future. The remainder of this chapter will show that current plug-in DSP boards are capable for task such as real-time image fusion.

The computational aspects of image fusion can be divided into two parts:

- computations needed for image warping
- computations needed for image fusion

These parts are discussed separately. In these discussions, we will focus on the floating point operations needed, based upon the assumption that floating point operations are much more expensive then integer operations.

5.1 Computing image warping

The computations needed for image warping consist of two pieces:
computing the location of the output pixel in the input image co-ordinates, and computing the interpolated value at this input image co-ordinate. Bilinear interpolating seems adequate; however, to save computations one might consider zero order interpolation.
Image warping cannot be implemented in a standard image pipeline: beforehand one cannot tell in which order the input pixels are needed, and thus the pixels

cannot be processed sequentially from input to output. This means that a frame store must be used to store an intermediate image.

5.1.1 Affine transformation

The computation of the location of the output pixel in input co-ordinate space is done by an affine transformation; the formula for an affine transformation is:

$$\begin{pmatrix} i_x \\ i_y \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} o_x \\ o_y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \tag{5.1}$$

which comes to four multiplications and six additions for each pixel. Alternatively, one could use the fact that every pixel on each image row must be processed; in that case the equation becomes:

$$\begin{pmatrix} i_x[n+1] \\ i_y[n+1] \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} + \begin{pmatrix} i_x[n] \\ i_y[n] \end{pmatrix} \tag{5.2}$$

which are 2 additions for each pixel.

5.1.2 Bilinear interpolation

Bilinear interpolation is specified by the formulas:

$$ix = \text{floor}[x] \tag{5.3}$$

$$iy = \text{floor}[y] \tag{5.4}$$

$$fx = x - ix \tag{5.5}$$

$$fxinv = 1 - fx \tag{5.6}$$

$$fy = y - iy \tag{5.7}$$

$$t1 = im[ix, iy] \cdot fxinv + im[ix+1, iy] \cdot fx \tag{5.8}$$

$$t2 = im[ix, iy+1] \cdot fxinv + im[ix+1, iy+1] \cdot fx \tag{5.9}$$

$$o = t1 \cdot (1 - fy) + t2 \cdot fy \tag{5.10}$$

This comes down to first doing a linear interpolation in the pixel row above and below the point, followed by a linear interpolation between the values of those rows.

The computational complexity of this process is six multiplications and seven additions/subtractions for each pixel, plus four indirection in a double array to access the image data used.

5.2 Computations needed for image fusion

This section will give the computational complexity of the Burt algorithm. The complexity of the Burt algorithm is typical for image pyramid approaches. Within the pyramids, most calculations (about 75%) is needed in the first pyramid level. We will estimate the number of calculations needed for that level, and estimate the total number of calculations for the total pyramid from that number.

The most expensive part of the Burt algorithm is the low pass filtering. This is performed both in the low pass filtering itself, given in formula (2.17), but also in the expansion. As reduce-expand has to be used, we have for each level three expand operations: two in pyramid formation, and one in image formation from the two pyramids. Also, we have two low pass filtering operations for the low pass filtering itself. This adds up to 5 low pass filtering operations.
Each low pass filtering consists of a separable convolution with a symmetric filter of width 5. This gives per pixel 3 multiplications and 5 additions. So, for all the low pass filtering operations in the first pyramid level we need 15 multiplications and 25 additions; for all pyramid levels this adds up to 20 multiplications and 33 additions per pixel.

5.3 Total computational complexity

The three separate contributions in the computational complexity of affine transformation, bilinear interpolation and pyramid fusion add up to 42 additions, 26 multiplications and 4 image data accesses. For fusion of 768x512 image streams at 25 frames/second, we have an incoming pixel rate from each sensor of about 10 Mpixel/second. This brings the complete computational budget near 1 Gflops. This is just within reach of the newest DSP boards, such as the Matrox Genesis with multiple Texas Instruments TMS320C80 DSPs.

6. Conclusions and recommendations

In this chapter we present some conclusions drawn from material presented in the previous chapters. At the end of this chapter we will present recommendations for application of this knowledge as well as new research directions will be given.

In chapter two a theoretical discussion is presented about optimal fusion. As a result it is demonstrated that for dependent images blending is the optimal solution, and that for independent images a select best strategy should be followed. In great detail the Burt method is given as an example of a select best strategy; also a deduction is presented how images with different signal to noise ratios should be combined.

In chapter three the basics of two methods leading to colour output images are presented.

Chapter four presents some results for three image sequences with the methods introduced earlier. From these images, one can deduce that no deciding advantage can be demonstrated between the blend, Burt and false colour images. It must be noted that the images presented in this report are selected for inclusion; in realistic scenario's quite often one of the two inputs will be superior to the other.

Chapter five gives a theoretical overview of the computational aspects for image fusion. The main conclusion drawn in that chapter is that real-time application of image fusion with a pyramidal approach such as the Burt method is just within the capabilities of the newest DSP boards.

6.1 Recommendations for using image fusion

- When a platform is equipped with multiple imaging sensors, make the possibility to use image fusion available. As shown, there are scenarios where fused images show more information than each of the input images separately.
- When a platform is equipped with multiple imaging sensors, and these sensors are sequentially displayed on the same screen, apply image warping to the images even when no explicit image fusion is applied. This will guarantee that when switching between images the location of the objects remains the same.

6.2 Recommendations for further research

- The examples indicate that colour image output is valuable. Further research is needed to determine the improvement which can be reached with colour displays over monochrome displays, and whether the improvement is enough to compensate for less sharpness in colour displays due to colour masks.
- The flicker of the Burt images is caused by different temporal noise realisations for separate frames in the sequence. Further research could be carried out to investigate into the possibility to reduce this flicker by using knowledge how the decisions where made for previous frames.
- The target recognition example indicates that the use of multiple imaging sensors improves target recognition rate. With target recognition, it is not necessary that all sensors used have the same modality and/or input rate. This opens the possibility to integrate radar and laser range sensors with imaging sensors. Usage of target tracking might lead to higher signal to noise ratios due to suppresses temporal noise.
- Both image fusion and image enhancement are promising techniques to improve the perception of the scene by the operator. A combination of those techniques seems promising. Future research should provide insight in the relation between image fusion and image enhancement, and under what operational circumstances what technique, or combination of techniques, should be utilised.
- As it is unlikely that for all operational circumstances an optimal algorithm exists, a scheme must be designed to select the best algorithm given the known operational variables. This should include detection of sensor jamming and evading its artefacts.

7. References

[1] 'Computer Graphics: Principles and Practice', *James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes*, 2nd edition in C, ISBN 0-201-84840-6, Addison Wesley, 1996

[2] 'Digital Image Processing', *Rafael C. Gonzalez and Richards E. Woods*, ISBN 0-201-50803-6, Addison Wesley, 1992

[3] 'Merging images through pattern decomposition', *P.J. Burt and E.H. Adelson*, Applications of Digital Image Processing VIII, SPIE vol. 575, pp. 173-181, 1985

[4] 'Image fusion by a ratio of low-pass pyramid', *A. Toet*, Pattern Recognition Letters 9, pp. 245-253, 1996

[5] 'Wavelets for a Vision', *Stéphane Mallat*, Proceedings of the IEEE, Vol. 84, pp. 604-614, 1996

[6] 'Gaussian Pyramid Wavelet Transform for Multiresolution Analysis of Images', *H. Olkkonen and P. Pesola*, Graphical Models and Image Processing, vol. 58, pp. 394-398, 1996

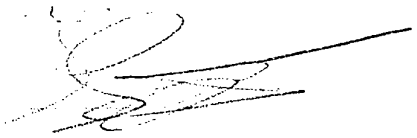
[7] 'Image Registration and Pyramid Decomposition for Multisensor Image Fusion', *P.J.L. van Beek*, to appear as report TNO-FEL.

[8] 'Image fusion improves situational awareness', *A. Toet, J.K. IJspeert, M.J. van Dorresteijn*, TM-96-A051, TNO Human Factors Research Institute, 1996

[9] 'Multisensor Image Fusion for Detection of Targets in the Battlefield of the Future', *L. Sévigny*, NATO AC/243, Panel 3, RSG9 38th meeting Progress Report - Canada, Defence Research Establishment Valcartier, 1996

8. Signature

C.W. Lamberts
Group leader



Dr K. Schutte
Author

Appendix A Target recognition experiment

An evaluation is made from the sequence with target cueing; an example of this sequence is depicted in figure 4.1. For each of the twenty-six images in the sequence, a visual evaluation is made. Within that visual evaluation it was decided whether the object found was a true object, or a false alarm. Also the type of alarm was noted: in the IR image only, in the visual image only, or in both channels. The results of this evaluation are shown in table A.1.

Table A.1: Recognition rates for all frames.

	Correct			Wrong		
	IR	Vis.	Comb.	IR	Vis.	Comb.
0	1		1			
1			1		2	
2			1		4	
3			1	3		
4			1	1	2	
5	1		1		2	1
6			2	1	1	
7			2	3	1	
8			2	2	3	
9			2	4	2	
10			2	3	2	
11		1	1	1	4	1
12		1	1		6	
13		1	1		4	
14			2	2	5	
15			2	1	1	
16			2	2	3	
17			2	3	1	
18		1	2	1	5	
19			2	2	5	
20			2	1	5	
21			2	2	3	
22			2	1	6	1
23			2	3	3	1
24		1	1	3	2	
25		1	1		4	
26	1	1		2	2	
Totals:	3	7	41	41	78	4

The totals given in table A.1 can be used to determine the detection quality of using IR only, visual only, or the combination of both. First, we need to determine the cumulative number of objects present in the sequence; this is 3+7+41=51 objects. Of these 51 objects, in the IR images 3+41=44 objects can be seen, which is 86%. In the visual images 7+41=48 objects can be seen, which is 94%. In the combination 41 objects are recognised, which is 80%. Second, we determine the

false alarm rates. For IR, this is (41+4)/51=88%. For visual images, this is (78+4)/51=161%. For the combination, this is 4/41=10%.

Table A.2: Detection and false alarm rates.

	infrared	visual	combination
detection rate	86%	94%	80%
false alarm rate	88%	161%	10%

For this single experiment this means that when using both sensor signals accepting a slight decrease of detection rate results in a very big decrease of the number of false alarms. More realistic, one strives for a fixed false alarm rate for all methods, which can be achieved by varying thresholds used. Although it cannot be concluded from the results given here, experiments have shown that with such a fixed false alarm rate the combination of both sensors results in the highest detection rate.

REPORT DOCUMENTATION PAGE
(MOD-NL)

1. DEFENCE REPORT NO (MOD-NL)	2. RECIPIENT'S ACCESSION NO	3. PERFORMING ORGANIZATION REPORT NO
TD97-0087		FEL-97-B046
4. PROJECT/TASK/WORK UNIT NO	**5. CONTRACT NO**	**6. REPORT DATE**
6025958	-	February 1997
7. NUMBER OF PAGES	**8. NUMBER OF REFERENCES**	**9. TYPE OF REPORT AND DATES COVERED**
27 (incl 1 appendix, excl RDP & distribution list)	9	

10. TITLE AND SUBTITLE

Fusion of IR and visual images

11. AUTHOR(S)

Dr K. Schutte

12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TNO Physics and Electronics Laboratory, PO Box 96864, 2509 JG The Hague, The Netherlands
Oude Waalsdorperweg 63, The Hague, The Netherlands

13. SPONSORING AGENCY NAME(S) AND ADDRESS(ES)

TNO Physics and Electronics Laboratory, PO Box 96864, 2509 JG The Hague, The Netherlands
Oude Waalsdorperweg 63, The Hague, The Netherlands

14. SUPPLEMENTARY NOTES

The classification designation Ongerubriceerd is equivalent to Unclassified, Stg. Confidentieel is equivalent to Confidential and Stg. Geheim is equivalent to Secret.

15. ABSTRACT (MAXIMUM 200 WORDS (1044 BYTE))

Image fusion is the process where several independently recorded images are combined into a single image. This report descibes the fusion of thermal IR and visual imagery.
For image fusion an optimality criterion is defined. If the input images are assumed to be dependent the optimal fusion process is image blending. For independent input images pyramidal solutions such as Burt's algorithm are shown to be optimal.
Several image sequences are processed with the techniques described in the report.
Results presented in the report indicate that, for the images used, image fusion results in images with a higher information content. Hardware requirements for the algorithms presented are given. Real-time implementation on current plug-in DSP boards seems possible.
An initial study is performed into Automatic Target Recognition based upon multiple input images. The results look promising.

16. DESCRIPTORS	IDENTIFIERS
Data fusion Infrared images Automatic Target Recognition	Visual images Wavelet decomposition

17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
Ongerubriceerd	Ongerubriceerd	Ongerubriceerd

18. DISTRIBUTION AVAILABILITY STATEMENT	17d. SECURITY CLASSIFICATION (OF TITLES)
Unlimited Distribution	Ongerubriceerd

Distributielijst

1.	Bureau TNO Defensieonderzoek
2.	Directeur Wetenschappelijk Onderzoek en Ontwikkeling*)
3.	HWO-KL
4.	HWO-KLu
5.	HWO-KM
6.	HWO-CO
7 t/m 9.	KMA, Bibliotheek
10.	KMA, t.a.v. Ir. J. Rogge
11.	DMKM/WCS/COSPON, t.a.v. Drs. W. Pelt
12.	OCMAN/KCEN/OCC/EXTPL, t.a.v. Maj. D.W. Dull tot Backenhagen
13.	DMKLu/MXS, t.a.v. Ir. S.J.J. de Bruin
14 t/m 22.	AC/243 (Panel 3/RSG.9)
23.	TNO Technische Menskunde, t.a.v. Dr. A. Toet
24.	CC-TNO
25.	Directie TNO-FEL, t.a.v. Dr. J.W. Maas
26.	Directie TNO-FEL, t.a.v. Ir. J.A. Vogel, daarna reserve
27.	Archief TNO-FEL, in bruikleen aan M&P*)
28.	Archief TNO-FEL, in bruikleen aan Ir. C. Eberwijn
29.	Archief TNO-FEL, in bruikleen aan Drs. C.W. Lamberts
30.	Archief TNO-FEL, in bruikleen aan Dr. J.S. de Vries
31.	Archief TNO-FEL, in bruikleen aan Dr. P. Schwering
32.	Archief TNO-FEL, in bruikleen aan Dr. K. Schutte
33.	Archief TNO-FEL, in bruikleen aan Dr. A.J. van der Wal
34.	Archief TNO-FEL, in bruikleen aan Drs. R.J.L. Lerou
35.	Documentatie TNO-FEL
36 t/m 38.	Reserve
	TNO-PML, Bibliotheek**)
	TNO-TM, Bibliotheek**)
	TNO-FEL, Bibliotheek**)

Indien binnen de krijgsmacht extra exemplaren van dit rapport worden gewenst door personen of instanties die niet op de verzendlijst voorkomen, dan dienen deze aangevraagd te worden bij het betreffende Hoofd Wetenschappelijk Onderzoek of, indien het een K-opdracht betreft, bij de Directeur Wetenschappelijk Onderzoek en Ontwikkeling.

*) Beperkt rapport (titelblad, managementuittreksel, RDP en distributielijst).
**) RDP.