SIViP (2007) 1:293–301
DOI 10.1007/s11760-007-0025-4

ORIGINAL PAPER

Context enhancement through infrared vision: a modified fusion scheme

Zheng Liu · Robert Laganière

Received: 12 January 2007 / Revised: 25 May 2007 / Accepted: 26 May 2007 / Published online: 29 June 2007

Abstract In the night vision applications, visual and infrared images are often fused for an improved awareness of situation or environment. The fusion algorithms can generate a composite image that retains most important information from source images for human perception. The state of the art includes manipulating in the color spaces and implementing pixel-level fusion with multiresolution algorithms. In this paper, a modified scheme based on multiresolution fusion is proposed to process monochrome visual and infrared images. The visual image is first enhanced based on corresponding infrared image. The final result is obtained by fusing the enhanced image with the visual image. The process highlights the features from visual image, which is most suitable for human perception.

Keywords Night vision · Infrared image · Context enhancement · Image fusion

1 Introduction

The use of thermal vision technique can be found in numerous applications such as military, law enforcement, surveillance, navigation, security and wildlife observation. Thermal imaging is a type of infrared (IR) imaging, which detects radiation in the infrared range of the electromagnetic spectrum. The IR image provides an enhanced spectral range that is imperceptible to human beings and contributes to the contrast between objects of high temperature variance and environment. Compared with a visual image, the IR image is represented with a different intensity map. The same scene exhibits different features existing in different electromagnetic spectrum band.

Some automotive manufactures, like BMW, offer the night vision system as a high-tech option for their flagship vehicles. After dark, the chances of being in fatal car crash go up sharply, though traffic is way down [3]. The night vision system can help the drivers see as much as three or four times farther ahead and quickly distinguish objects [3]. Figure 1 gives a snapshot of the night vision system on a BMW vehicle. In a surveillance system, the objective is to detect human body in the environment with inadequate illumination.

A well-adopted method to process visual and IR images is the pixel-level fusion [1]. The visual and IR images are first transformed to the wavelet domain. Through combining the coefficients in the transform domain, a new composite image can be obtained by applying the inverse transform. The purpose of the fusion operation is to highlight the objects with high temperature variance and preserve the details of the background. Any enhancement may be applied to the visual image before the fusion operation as proposed by Tao et al. [11]. However, there are always questions to the "direct" fusion of visual and IR images at pixel level. In a low thermal contrast environment, the background details like vegetation or soil areas should be represented well in the visual bands [13]. Does the fusion with an IR image contribute to the fidelity of the background objects in such a scenario? The answer might be "no". The fusion may degrade the original information contained in the visual image when they are not complementary. The features presented in the visual band is most suitable for human perception. The details of color-based fusion methods proposed by Toet and Xue can be found in [13] and [14], respectively. In this work, we focus on the intensity images.

We proposed a modified scheme for the fusion of IR and visual images in this paper. The contrast of the visual image is

Z. Liu · R. Laganière (✉)
VIVA Laboratory, STE 5023 School of Information Technology and Engineering, University of Ottawa, 800 King Edward Avenue, Ottawa, ON K1N 6N5, Canada
e-mail: laganer@site.uottawa.ca

Fig. 1 The BMW night vision system on a vehicle (courtesy of BMW)

first enhanced through using the pixel value from IR image as an exponential factor. The result is then fused with the visual image again to emphasize the features obtained in the visual band. The objects with hight temperature variance are highlighted in the final result. As a result, an enhanced version of the visual image is achieved and can be presented to the end users.

The rest of the paper is organized as follows. The feasibility of using adaptive enhancement and image fusion technique for night vision application is investigated in Sect. 2. The modify fusion scheme is proposed in Sect. 3. Section 4 presents more experimental results obtained with the proposed method. Discussion and conclusion can be found in Sects. 5 and 6, respectively.

2 Enhancement and fusion

2.1 Histogram-based operations

Histogram-based operations like histogram equalization and matching provide a basic tool to enhance images. The description of such operations is available in most of the text books on image processing, for example, reference [2], and will not be repeated herein again. We use two images from two video sequences from (http://www.imagefusion.org) to demonstrate the processing methods. One is a visual image as shown in Fig. 2a and the other one in Fig. 2c is from an infrared camera. Due to the inadequate of illumination, the human body appearing in the IR image cannot be identified from the visual image, but visible in the IR image. In Fig. 2, the histograms of the two images are plotted.

First, the histogram equalization is applied to the visual image. The result is presented in Fig. 3a and its histogram can be found in Fig. 3b. However, the human body is still hard to identified. The histogram of the visual image is then manipulated to match the histogram of the IR image. The results are given in Fig. 3c, d, respectively. Similarly, no significant improvement is achieved.

2.2 Adaptive enhancement

As the real-world scenes exhibit with high dynamic range radiance spanning more than six orders of magnitude, the processing of the image is implemented through compression of

Fig. 2 The visual image and infrared image

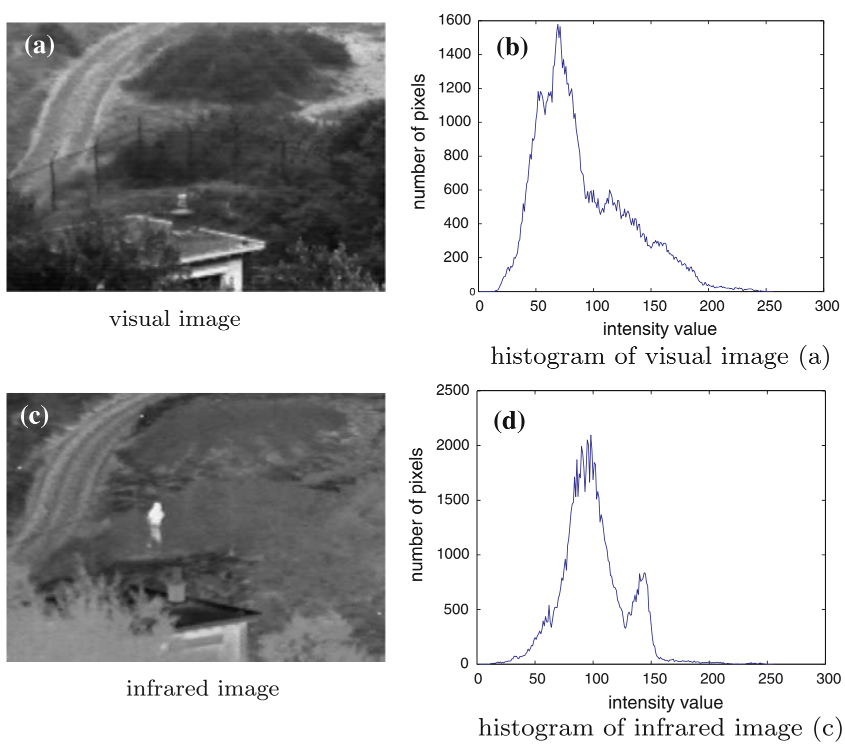
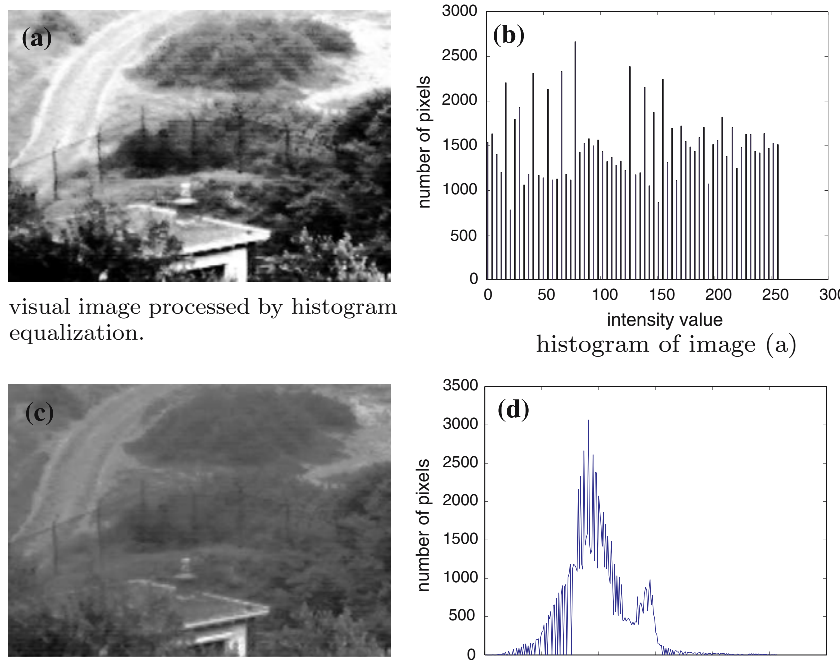


Fig. 3 The histogram-based processing of visual image



visual image processed by histogram equalization.

histogram of image (a)

visual image processed by histogram matching

histogram of image (c)

the dynamic range [9]. Herein, we test two methods proposed by Tao et al. in [9–12]. The two approaches are the adaptive and integrated neighborhood dependent approach for nonlinear enhancement (AINDANE) and an nonlinear enhancement method based on an illuminance–reflectance model. The basic idea of the two approaches is the same, i.e. applying a nonlinear transfer function to compress the dynamic range. The implementation procedures of the two algorithms are illustrated with the flowcharts shown in Fig. 4. The major difference between these two approaches is the nonlinear function and is highlighted with a dashed square. The final enhancement is implemented with equation [9]:

$$S(x, y) = 255 I_n'(x, y)^{E(x,y)} \tag{1}$$

where $S(x, y)$ is the enhanced image and $E(x, y)$ is obtained by the following two equations:

$$E(x, y) = r(x, y)^P = \left[\frac{I_G(x, y)}{I(x, y)}\right]^P \tag{2}$$

$$I_G(x, y) = G(m, n) \times I(x, y). \tag{3}$$

Herein, $G(m, n)$ is a $m \times n$ Gaussian kernel and parameter P is an empirical parameter [9]. The original image is represented as $I(x, y)$. In the ANIDANE algorithm, parameter P is given by:

$$P = \begin{cases} 3 & \sigma \leq 3 \\ \frac{27-2\sigma}{7} & 3 < \sigma < 10 \\ 1 & \sigma \geq 10 \end{cases} \tag{4}$$

where σ is the global standard deviation of the image. Image $I_n(x, y)$ is obtained by normalizing $I(x, y)$ to the range of [0, 1]. The nonlinear transfer function is given by:

$$I_n' = f(I_n, z) = \frac{I_n^{(0.75z+0.25)} + (1 - I_n)\,0.4\,(1 - z) + I_n^{(2-z)}}{2} \tag{5}$$

where the parameter z is determined by:

$$z = \begin{cases} 0 & L \leq 50 \\ \frac{L-50}{100} & 50 < L \leq 150 \\ 1 & L > 150. \end{cases} \tag{6}$$

Herein, the intensity level L corresponds a value of 0.1 in the cumulative distribution function of the image. In the method using illuminance-reflectance model, the P value becomes [10]:

$$P = \begin{cases} 2 & \sigma \leq 30 \\ -0.03\sigma + 2.9 & 30 < \sigma \leq 80 \\ 1/2 & \sigma > 80. \end{cases} \tag{7}$$

An inverse sigmoid function is used to obtain I_n'. This function can be expressed with the following equation:

$$I_n' = \frac{-\frac{1}{a}\ln\left[\frac{1}{I_n\left(\frac{1}{1+e^{-av_{\max}}} - \frac{1}{1+e^{-av_{\min}}}\right) + \frac{1}{1+e^{-av_{\min}}}} - 1\right] - v_{\min}}{v_{\max} - v_{\min}} \tag{8}$$

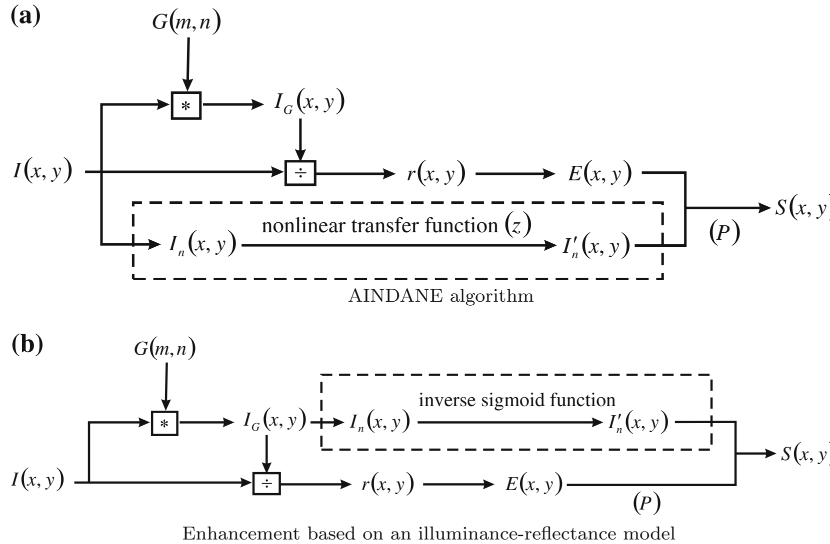


Fig. 4 The adaptive image enhancement algorithms

The parameter $v_{\max}$, $v_{\min}$, and a can be tuned manually. In the experiment, $v_{\max}$ and a are selected as three and one, respectively. The value of v_{min} is determined by the global mean I_m of the image I:

$$v_{\min} = \begin{cases} -6 & I_m \leq 70 \\ \frac{I_m - 70}{80} \times 3 - 6 & 70 < I_m < 150 \\ -3 & I_m \geq 150. \end{cases} \tag{9}$$

Figure 5 shows the results achieved by applying the two adaptive enhancement approaches. These images do not show the details of the poorly illuminated regions, although they are enhanced to some extent. To achieve an optimal enhancement, user needs to adjust the parameters used in the adaptive enhancment algorithms and this may vary with images.

2.3 Pixel-level image fusion

Multiresolution pixel-level fusion is often employed to combine visual and IR images. The implementation of this approach is twofold. One is the multiresolution analysis (MRA) algorithm, which represents the image in the transform domain. The other is the so-called fusion rule, which combines the coefficients or sub-images in the transform domain. Different combinations of these two basic elements have been investigated and published. An excellent review of image fusion techniques can be found in [1] and [6]. The basic operation for the fusion consists of three steps as illustrated in Fig. 6. The first step is to represent the input images in the transform domain by applying the multiresolution algorithm. A input image $I(x, y)$ can be represented as a sum over a collection of functions $g_i(x, y)$:

$$I(x, y) = \sum_i y_i(x, y)\, g_i(x, y) \tag{10}$$

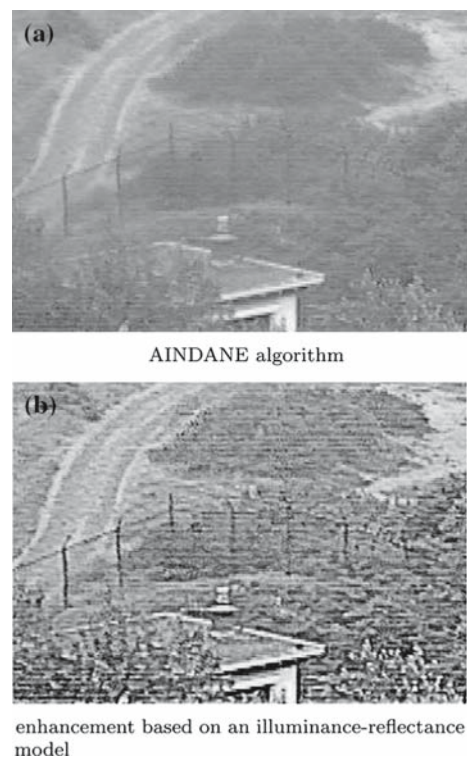


Fig. 5 The enhancement results of visual image achieved by adaptive algorithms

where $y_i(x, y)$ are the transform coefficients or sub-images. These coefficients are obtained by projecting the image onto a set of projection functions, $h_i(x, y)$:

$$y_i(x, y) = \sum_{x,y} h_i(x, y)\, I(x, y). \tag{11}$$

The projection function $h_i(x, y)$ are translated and dilated copies of one another. The sub-sampling factor is increased

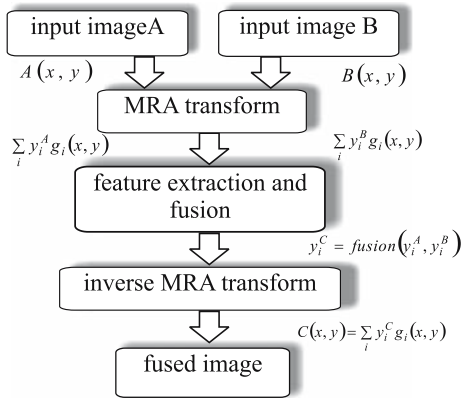


Fig. 6 The pixel-level multiresolution image fusion

by a factor of 2, namely, the image at lower level are down-sampled to the $1/4$ size of the adjacent hight level sub-images.

Although there are a number of multiresolution fusion algorithms available, a benchmark study of those fusion methods is beyond the scope of this paper. Herein, we want to present the fusion result of the visual and IR image obtained by a specific multiresolution analysis approach, namely, the steerable pyramid [4]. Readers are referred to references [7,8] for the details of steerable transform theory. The fusion rule consists of the absolute maximum value selection (AMVS) for the high-pass sub-bands and the average for the low-pass band. Written in mathematic formulas, there are:

$$y_{i,\text{high}}^{\text{new}}(x, y) = \begin{cases} y_{i,\text{high}}^A(x, y), & F_i^A(x, y) \geq F_i^B(x, y) \\ y_{i,\text{high}}^B(x, y), & \text{Others} \end{cases} \quad (12)$$

$$y_{i,\text{low}}^{\text{new}}(x, y) = \frac{1}{2}\left(y_{i,\text{low}}^A(x, y) + y_{i,\text{low}}^B(x, y)\right). \quad (13)$$

In this case, $F_i(x, y)$ is the absolute value of $y_{i,\text{high}}(x, y)$. The pixel-level fusion result is given in Fig. 7. Although the human body is perceptible in the fused image, this image does not present all the features appearing in the visual band. This image is not the scene that we see with our eyes.

3 A modified scheme

In the multiresolution fusion, the averaging of low-pass sub-image incorporates the contrast of the IR image while the AMVS scheme retains the details from both the IR and visual images. There is a question to such operation. If the purpose is only to find the object of high temperature variance, the IR image should be enough and a visual image is not needed. However, the IR image is not what people percept with their eyes. A visual image is most appropriate for human perception. The details from IR does destroy the contents of the fused image. Thus, the advantage of fusing IR and visual image at pixel level is ambiguous.

A modified fusion method is proposed as follow. The visual image is first enhanced by incorporating the information from IR image. To retain the details from the visual image, the enhanced image is fused with the visual image by applying the multiresolution fusion algorithm. This method implements the enhancement and fusion, but it is totally different from the approach in [11], where visual image is enhanced first and then fused with the IR image.

To enhance the visual image, the visual image, $I_{\text{visual}}(x, y)$, and IR image, $I_{IR}(x, y)$ are first normalized to the range [0, 1] and we get $I'_{\text{visual}}(x, y)$ and $I'_{IR}(x, y)$. Then, the enhanced image is:

$$I_{\text{visual}}^{\text{en}}(x, y) = 255 I'_{\text{visual}}(x, y)^{I'_{IR}(x,y)}. \quad (14)$$

Such enhancement can be illustrated with Fig. 7a. If we have two pixel values from IR and visual image, the enhanced pixel value can be easily located from the mesh surface in Fig. 8.

Then, the enhanced image $I_{\text{visual}}^{\text{en}}(x, y)$ is fused with the visual image $I_{\text{visual}}(x, y)$ by using the steerable pyramid based fusion algorithm [4]:

$$S(x, y) = \text{steer\_fuse}\left(I_{\text{visual}}(x, y), I_{\text{visual}}^{\text{en}}(x, y)\right). \quad (15)$$

In the implementation, three oriented band-pass filters are employed in the steerable pyramid algorithm. The decomposition level is four. The same setup is used throughout the experiment. The final result is shown in Fig. 9c. The

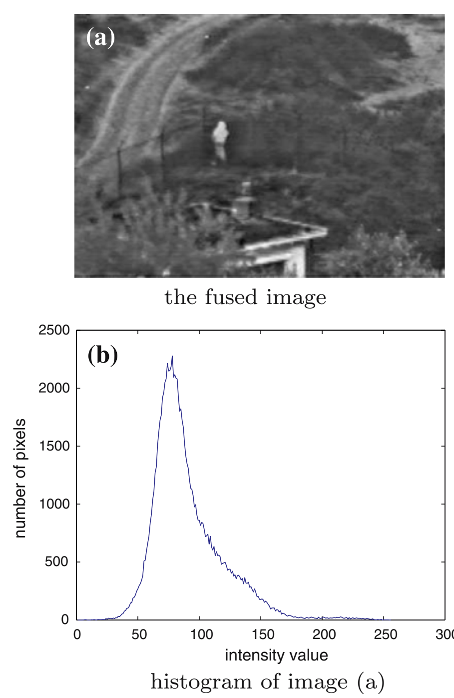


Fig. 7 The pixel-level fusion of visual and IR images

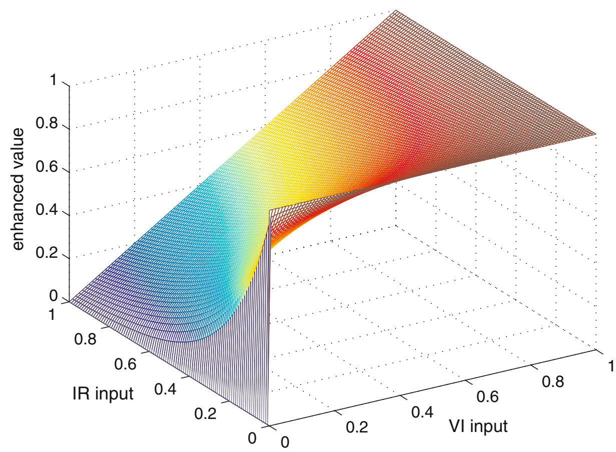


Fig. 8 The enhancement function

background information is retained well while the human body is successfully highlighted.

4 More results

More experimental results (nine groups) are presented in this section. For each group there are four images: (a) is the visual image; (b) is the infrared one; (c) is the result of MRA-based pixel-level fusion result; and (d) is obtained with the proposed method. We assume the visual and the infrared images in the experiments are perfectly registered. Although the registration of visual and IR image raises another important issue in image fusion, that topic is beyond the scope of this paper. The nine groups of images are retrieved from (http://www.imagefusion.org).

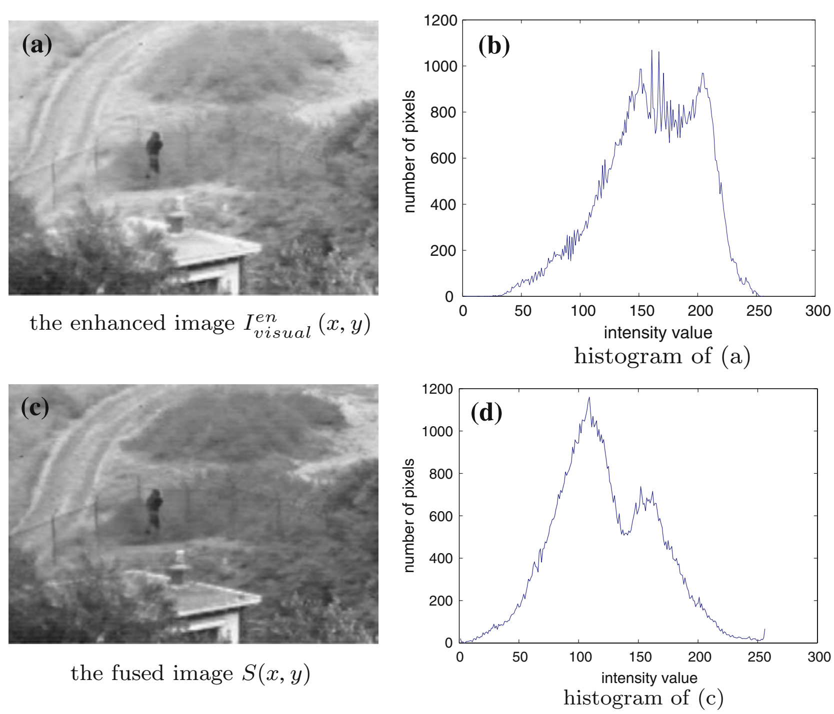
Compared to the MRA-based pixel-level fusion, the proposed method presents the results in the visual spectrum band, which are more suitable for human perception. Hidden features in the visual image are highlighted in the fused result. The only exception is the example shown in Fig. 13. The human body is not perceptible in the fused results obtained by either of the two methods. This will be discussed in next section.

Another observation is that the MRA-based pixel-level fusion mixed the two intensity tables of the visual and IR image. The human body in the fused image exhibits a quite similar appearance to that in the IR image. In contrast, our proposed method achieves a modification of the pixel value in the visual image. The details depend on the visual image only (Figs. 10, 11, 12).

5 Discussion

In the results shown in Fig. 13, neither the MRA-based pixel-level fusion nor the modified method can identify the human body, which is detectable in the IR image. In the corresponding visual image, the pixels in the region of human body reach an almost uniform value of high intensity and the fusion process cannot reveal the insignificant difference in intensity. In this case, the image mosaic technique can be used to generate a composite image as described in [5]. The object (human

Fig. 9 The result achieved by modified fusion method



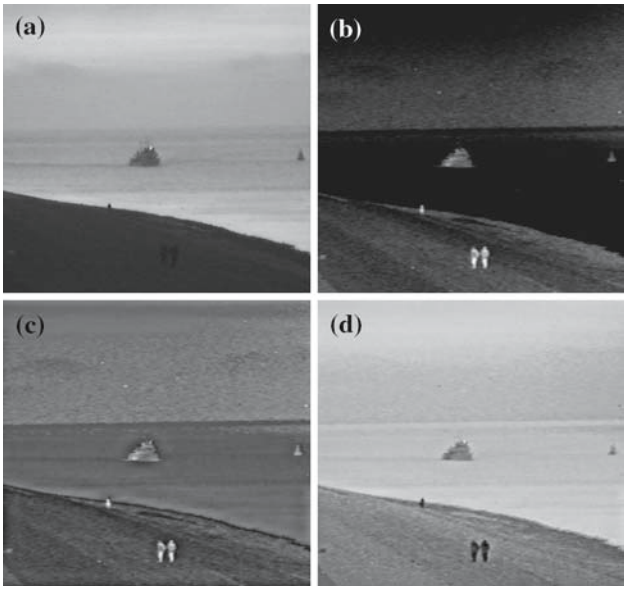


Fig. 10 TNO Kayak (frame 7118a)

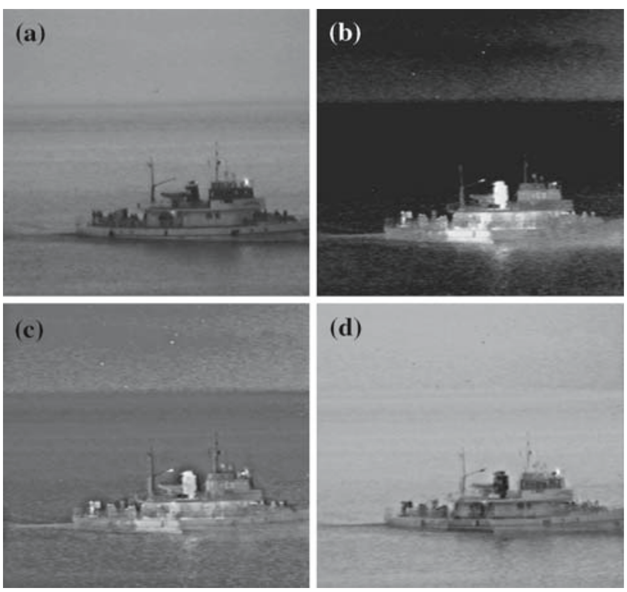


Fig. 11 TNO Kayak (frame 7436a)

body) needs to be identified from the IR image first. Then, the detected object can be embedded in the visual image. However, this situation will not be a problem for the applications like surveillance or transportation, because the high-intensity pixels already aggregate a bright and distinct spot in the visual image and if the task is just to identify there is an object. Image mosaic is a kind of seamless cut-and-paste operation. Therefore, the composite image will consist of regions from IR and visual images, which are presented in infrared and visual band, respectively. Herein, we want to highlight the details and present in the visual band.

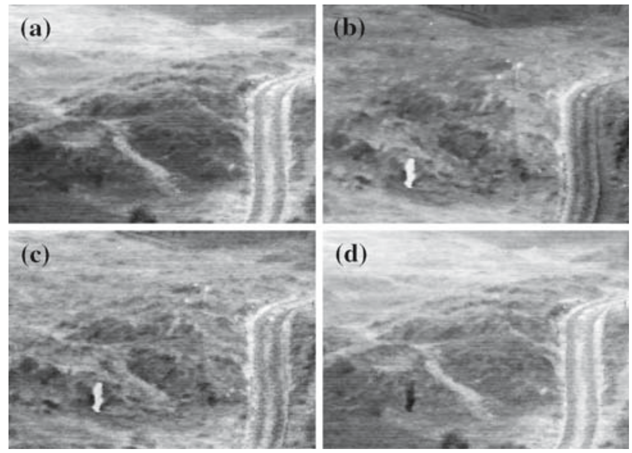


Fig. 12 TNO Dune (frame 7404)

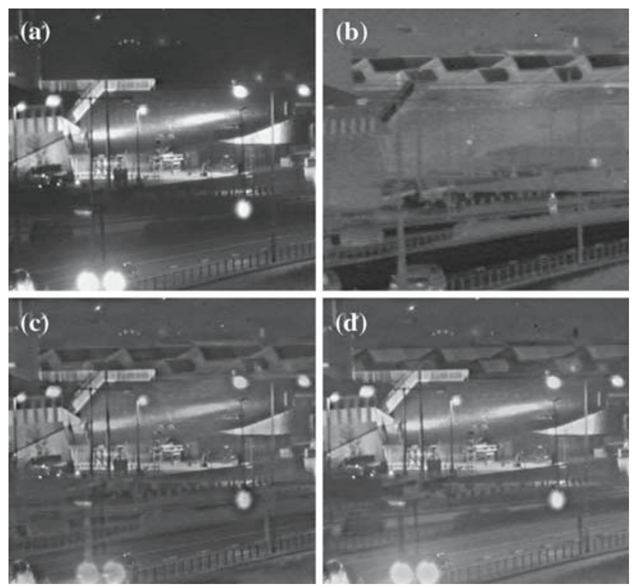


Fig. 13 TNO Kayak (frame e518a)

The IR image should be appropriate and good enough for the (human) object detection. The fusion with the visual image will not contribute to such purpose. However, when the task is to provide an observation of the scene, the fusion of the two image modalities as proposed in this paper could be a solution. Compared to the "direct" MRA-based pixel-level fusion, the results obtained with the proposed method are much closer to the nature scene and therefore they are appropriate to human perception. In other words, the fusion results are still presented in the visual band. Future study may investigate if the fused image is useful for post-processing like segmentation.

Another important issue is the objective assessment of the efficiency of the fusion algorithms. This still remains a challenge for the research of image fusion, especially when no perfect reference image is available for comparison and this is the most case in a practical application. Current research focuses on the measurement of information transferred to the fused image from the source images. Such measurement is valid for most cases. However, is a composite image, which contains the edges and boundaries from both IR and visual images, always the most optimal one for *human perception*, like the images in our experiments? Again, the answer is "no". The fusion operation should be able to convert the features from one spectrum band to another spectrum band rather than simply transfer those features. This could be the requirement for the applications where the fused images are presented for human perception instead of an automated processing or analysis procedure.

In our method, the IR intensity value is simply used as the exponential function to enhance the visual image. More sophisticate function may be considered in the future work. The IR image can be segmented first and each segment can be applied with different functions upon the requirements (Figs. 14, 15, 16, 17, 18).

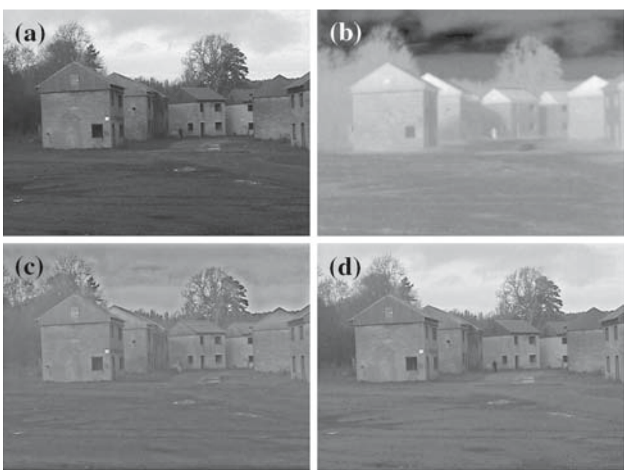


Fig. 14 Octec (frame 2)

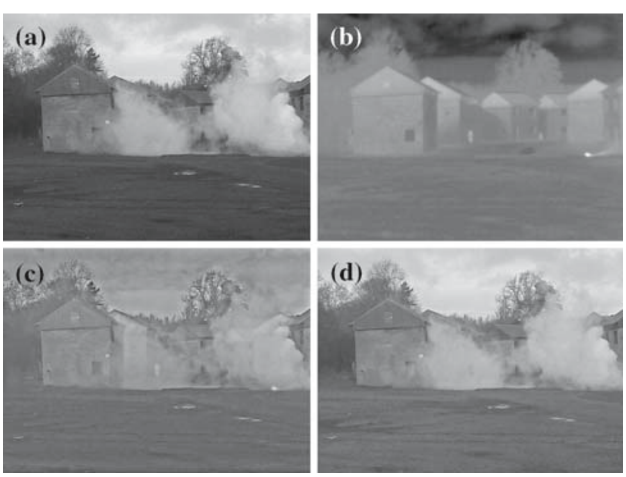


Fig. 15 Octec (frame 21)

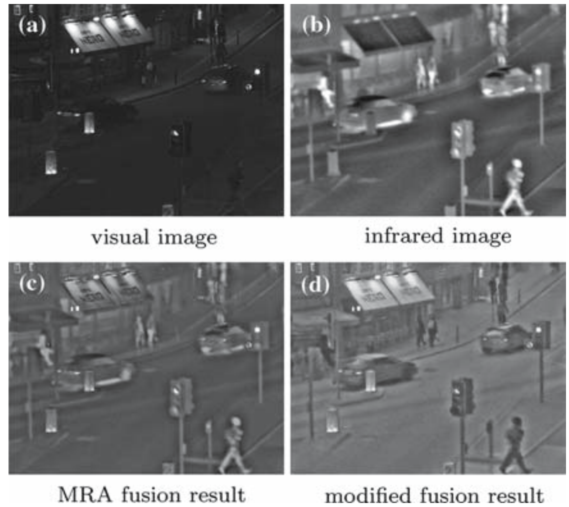


Fig. 16 Bristol Queen's road

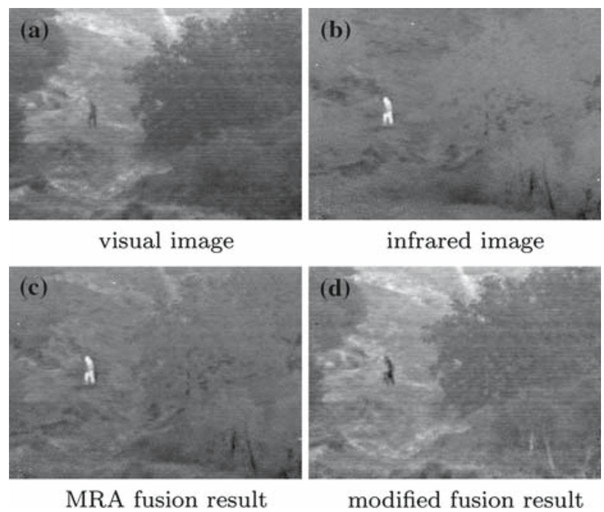


Fig. 17 TNO trees (frame 4906)

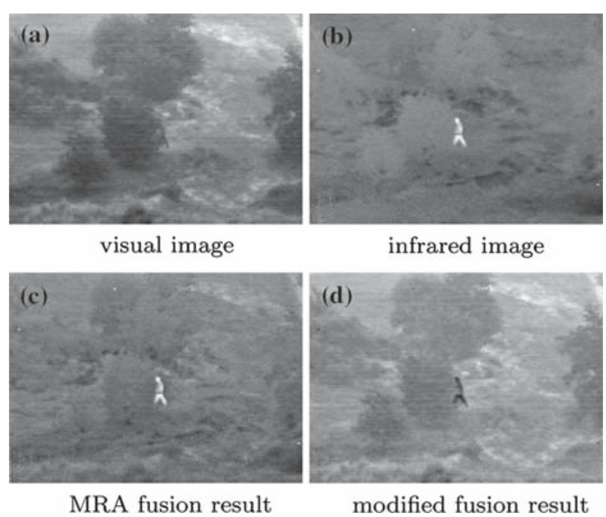


Fig. 18 TNO trees (frame 4917)

6 Conclusion

In this paper, a modified fusion process for night vision applications is presented. The method is straightforward and easy to implement. No empirical parameters need to be estimated. Usually, a visual image is fused with the corresponding infrared image at the pixel level with MRA-based algorithms. However, the presentation of information is not optimal, because the features falling in the range of visual band is more suitable for human perception. In the proposed method, the visual image obtained in an environment of poor visibility or inadequate illumination, is first enhanced by using the corresponding infrared image as the exponential factor. The enhanced result is then fused with the visual image to highlight the features in the visual spectrum band. In the fused result, the objects with higher radiation of heat are highlighted while the features from visual image are enhanced as well. This will help a driver to identify the obstacles on the road or improve the awareness of the environment in a surveillance application.

Acknowledgments The method described in this paper was validated with images retrieved from the website: http://www.imagefusion.org.

References

1. Blum, R.S., Xue, Z., Zhang, Z.: An overview of image fusion. In: Blum, R.S., Liu, Z. (eds.) Multi-Sensor Image Fusion and Its Applications, chap. 1, pp. 1–35. Taylor & Francis, London (2005)
2. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. 2nd edn. Prentice–Hall, New Jersey (2002)
3. Jones, W.D.: Safer driving in the dead night. IEEE Spect. 20–21 (2006)
4. Liu, Z., Tsukada, K., Hanasaki, K., Ho, Y.K., Dai, Y.P.: Image fusion by using steerable pyramid. Pattern Recognit. Lett. **22**, 929–939 (2001)
5. Liu, Z., Xue, Z., Blum, R.S., Laganiere, R.: Concealed weapon detection and visualization in a synthesized image. Pattern Anal. Appl. **8**(4), 375–389 (2006)
6. Piella, G.: A general framework for multiresolution image fusion: from pixels to regions. Inf. Fusion **4**(4), 259–280 (2003)
7. Siomoncelli, E., Freeman, W.: The steerable pyramid: a flexible architecture for multi-scale derivative computation. In: Proceedings of 2nd IEEE International Conference on Image Processing, pp. 444–447, Washington DC (1995)
8. Siomoncelli, E.P., Freeman, W.T., Adelson, E.H., Heege, D.: Shiftable multiscale transform. IEEE Trans. Inf. Theory **38**(2), 587–607 (1992)
9. Tao, L., Asari, V.K.: Adaptive and integrated neighborhood-dependent approach for nonlinear enhancement of color images. J. Electronic Imaging **14**(4), 1–14 (2005)
10. Tao, L., Asari, V.K.: An efficient illuminance-reflectance nonlinear video stream enhancement model. In: N. Kehtarnavaz (ed.) Proceedings of SPIE, Real-Time Image Processing, vol. 6063 (2006)
11. Tao, L., Ngo, H., Zhang, M., Livingston, A., Asari, V.: A multi-sensor image fusion and enhancement system for assisting drivers in poor lighting conditions. In: Proceedings of the 34th Applied Imagery and Pattern Recognition Workshop (2005)
12. Tao, L., Tompkins, R., Asari, V.K.: An illuminance-reflectance nonlinear video enhancement model for homeland security applications. In: Proceedings of 34th Applied Imagery and Pattern Recognition Workshop (2005)
13. Toet, A.: Fusion of Image from Different Electro-Optical Sensing Modalities for Surveillance and Navigation Tasks, chap. 7. pp. 237–264. Taylor & Francis, London (2005)
14. Xue, Z., Blum, R.S.: Concealed weapon detection using color image fusion. In: Proceedings of 6th International Conference of Information Fusion, vol. 1, pp. 622–627 (2003)