Research Article

Infrared Target Detection and Location for Visual Surveillance Using Fusion Scheme of Visible and Infrared Images

Zi-Jun Feng,^{1,2} Xiao-Ling Zhang,¹ Li-Yong Yuan,³ and Jia-Nan Wang²

¹ School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

² School of Computer Science and Information Technology, Northeast Normal University, Changchun 130117, China

³ College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321000, China

Correspondence should be addressed to Xiao-Ling Zhang; cs@nenu.edu.cn

Received 21 May 2013; Accepted 16 July 2013

Academic Editor: William Guo

Copyright © 2013 Zi-Jun Feng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The main goal of image fusion is to combine substantial information from different images of the same scene into a single image that is suitable for human and machine perception or for further image-processing tasks. In this study, a simple and efficient image fusion approach based on the application of the histogram of infrared images is proposed. A fusion scheme to select adaptively weighted coefficients for preserving salient infrared targets from the infrared image and for obtaining most spatial detailed information from the visible image is presented. Moving and static infrared targets in the fused image are labeled with different colors. This technique enhances perception of the image for the human visual system. In view of the modalities of infrared images, low resolution, and low signal-to-noise ratio, an anisotropic diffusion equation model is adopted to remove noise and to effectively preserve edge information before the fusion stage. By using the proposed method, relevant spatial information is preserved and infrared targets are clearly identified in the resulting fused images.

1. Introduction

With the rapid improvement of sensor technology, many surveillance systems have been developed in recent years. Infrared sensors, which are useful tools because of their individual modalities, have been employed in fields that include military surveillance, medical imaging, and machine vision. Infrared sensors can detect relative differences in the amount of thermal energy reflected from objects in the scene. They are thus more effective than visible cameras under poor lighting conditions. Current studies have focused on object detection and tracking from infrared images, and infrared sensor-based methods have shown good performance [1–4]. However, sensors often differ in modalities, considering physical and technological limitations. The image data acquired using different sensors exhibit diverse modalities, such as degradation and thermal and visual characteristics. By combining two or more different sensor data sets, a surveillance system can perform better than a system that uses only a single sensor. This technology is called image fusion. Image fusion is defined as the process of combining substantial information from

several sensors using mathematical techniques to create a single composite image that is highly comprehensive and thus extremely useful for a human operator or in the execution of other computer vision tasks [5].

In this study, we focus on the fusion process of visible and infrared imagery. Visible imagery has high resolution and can provide spatial details of the scene. Infrared imagery aids in the detection and recognition of heat-based targets under poor lighting conditions or in cases in which the target and the background are of the same color [6, 7]. Relevant information in both visible and infrared images should be preserved in the resulting fused images. A surveillance system can benefit from an efficient fusion to improve targeting and identification performance. In recent years, image fusion has become an important topic. Many fusion methods for visible and infrared images have been proposed [4, 6, 8, 9]. However, these techniques often lead to spatial distortions in the visible image because of the mixing of irrelevant infrared information in the nontarget area of the infrared image. In [8], the experiments on visible and infrared images show the contamination of the background regions by the

infrared information. Considering the modalities of visible and infrared images, the following three basic requirements must be met to achieve a good fusion method: (1) infrared targets must be perfectly preserved in the fused image; (2) spatial detail information from the visible image must not contaminate the nontarget regions of the infrared image; and (3) the fused image should be enhanced to be easily understood by the human visual system.

In the current work, a novel fusion algorithm is proposed to meet these requirements. The method is aimed at preserving high spatial detail information and showing infrared targets clearly in the resulting fused image. An adaptive threshold can be determined by the histogram of the infrared image. If the gray value (pixel by pixel) of the infrared image is greater than or equal to the threshold value, then the pixel value of the fused image is calculated by a weight determination function. Otherwise, the new value of the fused image would be directly obtained from the visible image. The proposed fusion algorithm is highly suitable for the image fusion of visible and infrared images because it displays salient infrared targets and avoids spatial distortion. Furthermore, the idea of labeling moving targets with red contours and marking static targets (or the highlighted nontarget areas of the infrared image) with green contours is very effective for the human visual system.

The rest of the paper is organized as follows. Section 2 gives a brief introduction of nonlinear diffusion filtering. Section 3 describes in detail the image fusion scheme. Section 4 provides the experimental results and comparisons. Section 5 discusses infrared target detection. Section 6 presents the conclusions.

2. Image Denoising

Given the physical limitations of sensors, many infrared imaging systems produce images with low signal-to-noise ratio, contrast, and resolution; these features reduce the detectability of targets and impede further investigation of infrared images [1, 10, 11]. Thus, image denoising has become an important issue in addressing the shortcomings of infrared images. Several methods, such as Gaussian filtering [12] and wavelet-based denoising, have been developed to address these issues [10, 13]. A disadvantage of some common methods is that they do not only smooth noise, but they also blur edges that contain significant image features. In the current study, an anisotropic diffusion [14–16] algorithm is adopted in infrared image denoising to filter noise and effectively preserve edge information.

After more than two decades of research, partial differential equation- (PDE-) based nonlinear diffusion processes have been widely applied in signal and image processing. PDE-based and improved PDE approaches have been successfully applied in image denoising, segmentation, enhancement, and restoration [15–18]. The first PDE-based nonlinear anisotropic diffusion technique was reported by Perona and Malik [14]. Different from the isotropic heat conduction equation, the anisotropic diffusion equation filters an image depending on its local properties. The anisotropic diffusion

equation removes image noise and simultaneously preserves its edges.

The diffusion equation in two dimensions can be expressed as

$$\frac{\partial I(x, y, t)}{\partial t} = \text{div}(g(|\nabla I|) \nabla I), \quad (1)$$

$$I(x, y, 0) = I_0(x, y),$$

where div is the divergence operator, ∇ is a gradient operator, and $g(|\nabla I|)$ denotes the diffusion conductance function chosen as a decreasing function to ensure high diffusion in a homogeneous region and weak diffusion near the edges. A typical choice for the conductance function is

$$g(|\nabla I|) = \frac{1}{1 + (|\nabla I|/K)^2}, \quad (2)$$

where K is a gradient threshold that can be fixed by an empirical constant.

3. Image Fusion Scheme

First, a common weighted combination rule is presented

$$I_f(x, y) = \begin{cases} \omega_2 I_v(x, y) + \omega_1 I_i(x, y) & \text{if } I_i(x, y) \geq \text{th}, \\ I_v(x, y) & \text{otherwise,} \end{cases} \quad (3)$$

where $I_v(x, y)$ is the intensity value of the pixel coordinate (x, y) in the visible image; $I_i(x, y)$ and $I_f(x, y)$ are the intensity values of the corresponding pixel (x, y) in the infrared image and the fused image, respectively; and th is used to distinguish between the dark background and the bright objects determined by image histogram-based method. The fused image is produced pixel by pixel according to the gray value of the infrared image. If the gray value is greater than or equal to th , then the new gray value of the fused image is obtained at the corresponding pixel location using (3). Otherwise, the gray value of the fused image is directly obtained from the visible image. The next step of the fusion algorithm is the production of the weighted coefficients.

Second, a weight determination function is defined as

$$\omega_1 = \frac{1}{1 + \exp(-k((x - \text{sp}) / (G_{\max} - \text{sp}) - \sigma))}, \quad (4)$$

$$\omega_2 = 1 - \omega_1,$$

where x represents the intensity value of the infrared image, $G_{\max}$ is equal to 255, k controls the slope of the function, σ determines the translation of the curve, and sp is the starting position of the curve. Figure 1 depicts the function defined in (4), considering the conditions wherein $k = 10$, $\sigma = 0.5$, and $\text{sp} = 120$. An increase in k results in a steep curve. Therefore, k is usually limited to the range in [8, 12], depending on the intensity levels of each image. The quantity σ is a constant equal to 0.5; this constant specifies that the values of ω_1 and ω_2 must be 0.5 when x is the middle point in the range $[\text{sp}, G_{\max}]$.

FIGURE 1: Example of a weight determination function.

In this study, the parameters $k = 10$ and $\sigma = 0.5$ are chosen, but the sp value is an adaptive selection based on the intensity histogram of the infrared image.

Third, the sp value is defined as

$$sp = G_{\max} - 2(G_{\max} - th), \quad (5)$$

where $G_{\max}$ is equal to 255 and th is the same threshold described in (3) that distinguishes between the targets and the background according to the histogram of the infrared image. The value of th is in the range [40, 80], as determined by considering the intensity level distribution of the infrared image in the experiments.

4. Experimental Results

4.1. Image Data Sets. Two sets of visible and infrared images were chosen. They are publicly available through the ImageFusion.org website [19]. These image sets were chosen because (1) the grayscale levels of the visible image are abundant and the light objects of the scene are clearly distinguishable in the infrared image (Figures 2(a) and 2(b)); and (2) the infrared image is corrupted by noise and the intensities of targets are low (Figure 3(b)). Figures 3(a) and 3(b), respectively, show a visible frame and the corresponding infrared frame from the sequences extracted from MPEG files. The image in Figure 3(b) was warped using planar homography to align it with the visible spectrum image [20]. All input images were assumed to be registered, and each pair of images contains exactly the same scene.

The first image set (UN Camp, frame 1815) comprises a terrain scene characterized by a path, trees, fences, and a house roof (Figure 2(a)). A person standing behind the trees and closing the fence is shown in the infrared image (thermal $3\text{--}5\mu\text{m}$, Figure 2(b)). Although the two images exhibit the same scene, the visible image clearly provides spatial details of the scene, but the human figure is invisible.

On the contrary, the human figure is highlighted in the infrared image, but other objects are hard to recognize correctly. The second data set (AIC) was chosen from the AIC thermal and visible nighttime data set sequence [19, 20]. Frame 3853 (Figure 3(a)) was extracted from the visible MPEG format file, and the corresponding frame (Figure 3(b)) was obtained from the infrared (thermal: $7\mu\text{m}$ to $14\mu\text{m}$) sequence. To ensure consistency in the procedure, the color image was converted to a grayscale image. The white margin of the infrared image was manually filled with black color to achieve good image fusion. The scene contains buildings, bright windows, roads, and pedestrians. Further details and descriptions of the data acquisition procedure can be found in [20].

4.2. Infrared Image Denoising. The method for reducing noise in this case is nonlinear diffusion filtering using techniques similar to those discussed in Section 2. An additive operator splitting (AOS) algorithm [21], a type of efficient nonlinear diffusion filtering, was applied to the infrared images. The AOS algorithm performs anisotropic diffusion to remove image noise while preserving its edges.

4.3. Fusion Results and Comparisons. To illustrate the performance of the proposed image fusion approach, image fusion was performed using conventional methods, namely, weighted averaging (WAV-based), discrete wavelet transform (DWT-based), and nonsubsampling contourlet transform (NSCT-based). The DWT-based and NSCT-based methods are performed by simply merging the low-pass and high-pass subband coefficients using the averaging scheme and the choose-max selection scheme, respectively. The DWT-based fusion algorithm is performed using five-level decomposition. The NSCT-based method is performed by using db3 wavelets in scale decomposition; 9–7 wavelets are used in a nonsubsampling directional filter bank, in which the number of directions is 4.

Figures 2 and 3 illustrate the fusion results using the aforementioned methods. The proposed scheme captures most of the salient areas of the infrared images and preserves the spatial detail information from the visible images. The infrared targets are obvious, and the nontarget regions are seldom contaminated in the fused images. For clear comparisons, the difference images between the fused images and the source visible images are given in Figure 4. The fused images obtained using the proposed method have the best visual quality.

Objective evaluation criteria, namely, fusion root-mean-square error (FRMSE) and correlation coefficient (CC), were used to evaluate the fused images. Considering that the size of the target in the infrared image is small compared with that in the scene, we suppose that I_v is the ideal reference image. $I_v(x, y)$ and $I_f(x, y)$ denote the pixel value of the visible image and the fused image at points (x, y) , respectively. The size of the images is calculated as $M \times N$.

- (1) FRMSE can effectively reflect the similarity between two images. Small FRMSE values provide satisfactory

FIGURE 2: Source images and fusion results of different fusion algorithms from the UN Camp sequence set. (a) Visible image; (b) denoised infrared image; and (c)–(f) images obtained by the WAV-based, DWT-based, NSCT-based, and proposed methods, respectively.

FIGURE 3: Source images and fusion results of different fusion algorithms from the AIC sequence set. (a) Visible image; (b) denoised infrared image; and (c)–(f) images obtained by the WAV-based, DWT-based, NSCT-based, and proposed methods, respectively.

fusion results:

$$\text{FRMSE} = \sqrt{\frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N (I_v(x, y) - I_f(x, y))^2}. \quad (6)$$

(2) CC can be evaluated to compare $I_v(x, y)$ with $I_f(x, y)$:

$$\text{CC} = \left(\sum_{x=1}^M \sum_{y=1}^N [(I_v(x, y) - \bar{I}_v) \times (I_f(x, y) - \bar{I}_f)] \right)$$

FIGURE 4: Difference images between the fused images and the corresponding visible images. (a)–(d) For the UN Camp sequence, fusion was performed using the WAV-based, DWT-based, NSCT-based, and proposed methods, respectively. (e)–(h) For the AIC sequence, fusion was performed using the WAV-based, DWT-based, NSCT-based, and proposed methods, respectively.

$$\times \left(\sum_{x=1}^M \sum_{y=1}^N \left[(I_v(x, y) - \bar{I}_v)^2 \right] \right. \\ \left. \times \sum_{x=1}^M \sum_{y=1}^N \left[(I_f(x, y) - \bar{I}_f)^2 \right] \right)^{-1/2}, \quad (7)$$

where $\bar{I}_v$ and $\bar{I}_f$ are the means of the visible image and the fused image, respectively. The maximum coefficient corresponds to the optimum fusion.

Table 1 shows the results of the quantitative evaluation using the two evaluation methods. The proposed method achieves superior results. The values of the fusion results of the first image set are not entirely consistent with those of the second image set because the visible image from the UN Camp sequence contains abundant texture information. By contrast, the AIC visible sequence contains structural details of buildings, such as edges. These objective criteria prove that the images fused using the proposed method are strongly correlated with the corresponding visible images; that is, the proposed scheme ensures that useful spatial detailed information of the visible images is preserved in the fused images.

5. Infrared Target Detection

Automatic object detection remains difficult to be undertaken. An object detector needs to cope with the diversity of visual imagery that exists in the world at large. Different detection methods can be used for different environmental conditions, so numerous approaches for automatic object detection have been investigated [22–26]. A common method is to extract targets from the image sequence through background subtraction when the video is captured by a stationary

TABLE 1: Comparison of fusion results.

Method	UN Camp set		AIC set	
	FRMSE	CC	FRMSE	CC
WAV-based	26.0935	0.7913	26.1056	0.9512
DWT-based	26.9110	0.7394	14.1720	0.9605
NSCT-based	25.4268	0.7813	18.4112	0.9439
Proposed	13.0508	0.9443	7.3073	0.9892

camera. The simplest way is to calculate an average image of the scene or to smoothen the pixels of the background with a Kalman filter [23] under the assumption that the background consists of stationary objects. A preferable way to tolerate background variation in a video is to employ a Gaussian function that describes the distribution of each pixel belonging to a stable background object. Among these background subtraction techniques, the mixture of Gaussians (MoG) has been widely utilized to model scene backgrounds at the pixel level [24–26].

In MoG, the distribution of recently observed values of each pixel in the scene is characterized. A new pixel value is represented by one of the major components of the mixture model and is used to update the model. One of the significant advantages of this method is that when something is allowed to become part of the background the existing model of the background is not destroyed. The original background color remains in the mixture until it becomes the most probable distribution and a new color is observed [26]. Good results of foreground object detection by applying MoG to outdoor scenes have been reported.

In surveillance applications, moving and static targets need to be correctly detected (for military surveillance systems, static targets must not be arbitrarily ignored), and the locations of the targets need to be identified. In Section 4, the

FIGURE 5: Moving and static objects were detected and labeled with different colors. (a)–(c) UN Camp sequence: frames 1813, 1815, and 1830 are shown. (d)–(f) AIC sequence: frames 3796, 3853, and 3890 are shown.

resulting fused image is obtained according to the proposed fusion rule of visible and infrared images. However, the fused image in the implementation was not computed to determine whether a moving object is present in the scene.

To separate moving targets from the static targets, infrared images easily address the issues of background modeling based on the MoG method. The chromatic contours of infrared targets in the fused images are highly suitable for the human visual system. Visual effect is enhanced in the natural scene when moving, and static targets are labeled with different colors. The data sets are still UN Camp and AIC images, as described in Section 4. In Figure 5, the moving and static targets are marked in red and green, respectively. These colors help humans clearly distinguish targets from the background in the fused images. All AIC infrared frames were warped to align them with the corresponding visible spectrum images. Green-dotted borders can be observed on top of the frames of the AIC sequence (Figure 5).

6. Conclusions

An efficient fusion method for the image fusion of infrared and visible images is proposed. To obtain good results using the proposed fusion scheme, a weight determination function by a suitable coefficient that enhances infrared targets and preserves spatial detail information from the infrared and visible images, respectively, is described. The method is appropriate for the image fusion of visible and infrared images. The infrared targets in the natural scene can be clearly distinguished in the resulting fused images. The infrared targets are highlighted by marking them in red (or green) in the fused image. This technique is useful for visual

surveillance. Moving targets are also detected and marked in red in this study. Future work will be focused on determining ways to track certain moving targets effectively.

Acknowledgments

The authors would like to acknowledge TNO Human Factors in The Netherlands and Dr. C. Ó Conaire for providing source images and video sequences. This work is supported by the Science Foundation for Research Fund for the Doctoral Program of Higher Education of China (no. 20100043120012), National Natural Science Foundation of China for Young Scholars (no. 41101434), and Opening Fund of Top Key Discipline of Computer Software and Theory in Zhejiang Provincial Colleges at Zhejiang Normal University under Grant no. ZSDZZZXK37.

References

- [1] J. W. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Computer Vision and Image Understanding*, vol. 106, no. 2-3, pp. 162–182, 2007.
- [2] Y. Li, X. Mao, D. Feng, and Y. Zhang, "Fast and accuracy extraction of infrared target based on Markov random field," *Signal Processing*, vol. 91, no. 5, pp. 1216–1223, 2011.
- [3] J. Shaik and K. M. Iftikharuddin, "Detection and tracking of targets in infrared images using Bayesian techniques," *Optics and Laser Technology*, vol. 41, no. 6, pp. 832–842, 2009.
- [4] S. Colantonio, M. Benvenuti, M. G. di Bono, G. Pieri, and O. Salvetti, "Object tracking in a stereo and infrared vision system," *Infrared Physics and Technology*, vol. 49, no. 3, pp. 266–271, 2007.

- [5] T. Stathaki, *Image Fusion: Algorithms and Applications*, Academic Press, 2008.
- [6] A. Toet, M. A. Hogervorst, S. G. Nikolov et al., "Towards cognitive image fusion," *Information Fusion*, vol. 11, no. 2, pp. 95–113, 2010.
- [7] M. Leviner and M. Maltz, "A new multi-spectral feature level image fusion method for human interpretation," *Infrared Physics and Technology*, vol. 52, no. 2-3, pp. 79–88, 2009.
- [8] J. J. Lewis, R. J. O' Callaghan, S. G. Nikolov, D. R. Bull, and N. Canagarajah, "Pixel- and region-based image fusion with complex wavelets," *Information Fusion*, vol. 8, no. 2, pp. 119–130, 2007.
- [9] P. G. Wang, H. Tian, and W. Zheng, "A novel image fusion method based on FRFT-NSCT," *Mathematical Problems in Engineering*, vol. 2013, Article ID 408232, 9 pages, 2013.
- [10] C. Ni, Q. Li, and L. Z. Xia, "A novel method of infrared image denoising and edge enhancement," *IEEE Transactions on Signal Processing*, vol. 88, no. 6, pp. 1606–1614, 2008.
- [11] V. S. Petrović and C. S. Xydeas, "Sensor noise effects on signal-level image fusion performance," *Information Fusion*, vol. 4, no. 3, pp. 167–183, 2003.
- [12] A. P. Witkin, "Scale-space filtering: a new approach to multi-scale description," in *Proceedings of the IEEE International conference on Acousticw, Speech, and Signal Processing (ICASSP '84)*, vol. 9, pp. 150–153, March 1984.
- [13] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.
- [14] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.
- [15] J. Monteil and A. Beghdadi, "A new interpretation and improvement of the nonlinear anisotropic diffusion for image enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 940–946, 1999.
- [16] H. Singh, V. Kumar, and S. Bhooshan, "Anisotropic diffusion for details enhancement in multiexposure image fusion," *ISRN Signal Processing*, vol. 2013, Article ID 928971, 18 pages, 2013.
- [17] X. Li and M. Yin, "An opposition-based differential evolution algorithm for permutation flow shop scheduling based on diversity measure," *Advances in Engineering Software*, vol. 55, pp. 10–31, 2013.
- [18] Y. Hu, M. Yin, and X. Li, "A novel objective function for job-shop scheduling problem with fuzzy processing time and fuzzy due date using differential evolution algorithm," *International Journal of Advanced Manufacturing Technology*, vol. 56, no. 9–12, pp. 1125–1138, 2011.
- [19] ImageFusion.Org, The Online Resource for Research in Image Fusion, <http://www.imagefusion.org/>.
- [20] C. Ó. Conaire, N. E. O'Connor, E. Cooke, and A. F. Smeaton, "Comparison of fusion methods for thermo-visual surveillance tracking," in *Proceedings of the 9th International Conference on Information Fusion (FUSION '06)*, pp. 1–7, July 2006.
- [21] J. Weickert, B. M. Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 398–410, 1998.
- [22] L. Li, W. Huang, I. Y. H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in *Proceedings of the 11th ACM International Conference on Multimedia (MM '03)*, pp. 2–10, November 2003.
- [23] D. Koller, J. Weber, T. Huang et al., "Towards robust automatic traffic scene analysis in real-time," in *Proceedings of the 12th IAPR International conference on Pattern Recognition*, pp. 126–131, October 1994.
- [24] M. Harville, "A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models," in *Proceedings of the 7th European conference on Computer Vision (ECCV '02)*, pp. 543–560, May 2002.
- [25] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in *Proceedings of the 6th European conference on Computer Vision (ECCV '00)*, pp. 751–767, June 2000.
- [26] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.