applied optics

Multiscale contrast direction adaptive image fusion technique for MWIR-LWIR image pairs and LWIR multifocus infrared images

A. Onur Karali, Serdar Cakir,* and Tayfun Aytaç

TÜBİTAK BİLGEM İLTAREN, Şehit Yzb. İlhan Tan Kşlas, 2432. cad., 2489. sok., TR-06800 Ümitköy, Ankara, Turkey
*Corresponding author: serdar.cakir@tubitak.gov.tr

Received 2 March 2015; revised 1 April 2015; accepted 5 April 2015; posted 8 April 2015 (Doc. ID 235519); published 29 April 2015

Infrared (IR) cameras are widely used in the latest surveillance systems because spectral characteristics of objects provide valuable information for object detection and identification. To assist the surveillance system operator and automatic image processing tasks, fusing images in the IR band was performed as a solution to increase situational awareness and different fusion techniques were developed for this purpose. Proposed techniques are generally developed for specific scenarios because image content may vary dramatically depending on the spectral range, the optical properties of the cameras, the spectral characteristics of the scene, and the spatial resolution of the interested targets in the scene. In this study, a general purpose IR image fusion technique that is suitable for real-time applications is proposed. The proposed technique can support different scenarios by applying a multiscale detail detection and can be applied to images captured from different spectral regions of the spectrum by adaptively adjusting the contrast direction through cross-checking between the source images. The feasibility of the proposed algorithm is demonstrated on registered multispectral [mid-wave IR (MWIR), long-wave IR (LWIR)] and LWIR multifocus images. Fusion results are presented and the performance of the proposed technique is compared with the baseline fusion methods through objective and subjective tests. The technique outperforms baseline methods in the subjective tests and provide promising results in objective quality metrics with an acceptable computational load. In addition, the proposed technique preserves object details and prevents undesired artifacts better than the baseline techniques in the image fusion scenario that contains four source images.

OCIS codes: (100.0100) Image processing; (100.2000) Digital image processing; (100.2960) Image analysis; (100.2980) Image enhancement; (110.3080) Infrared imaging.

http://dx.doi.org/10.1364/AO.54.004172

1. INTRODUCTION

Image fusion is the process of generating a highly descriptive single image from the input images captured at different sources, sensors, or wavelengths [1]. Image fusion frameworks can be divided into three main categories as "pixel (signal) level," "feature level," and "decision (symbol) level" fusion frameworks [2]. Pixel-level fusion is the lowest level of image fusion techniques dealing with the pixels obtained at the sensor directly and tries to improve the visual enhancement. One of the main advantages of the pixel-level schemes is their low computational complexity and easy implementation. But some of the pixel-based techniques are very sensitive to misregistration and suffer from blurring artifacts. Region-based methods are proposed to overcome this problem but region-based schemes are affected by the similar artifacts occurring especially near the edges of image subblocks [3]. In the feature-level fusion techniques, the fusion process is carried out in the form of feature descriptors. Feature-level algorithms typically perform fusion by extracting various types of features, such as regions or edges from the source images. Although feature-level fusion is usually robust to noise and misregistration, the pixel-level fusion retains more information from the source images compared to feature-level techniques [4]. The decision-level fusion scheme operates on the highest level, and refers to blending different discriminated results in the decision-making stage in accordance with the given fusion rules [5]. Decision-level fusion also loses some source image information in the fusion process. Therefore, pixel-level fusion schemes are more advantageous due to their low computational complexity, easy implementation, and less loss of information in the fusion process compared to feature- and decision-level fusion frameworks.

Fusion techniques are applied in various application areas, such as aided surveillance, multispectral imaging, medical

imaging, and target detection and tracking. Especially in a surveillance system, an efficient fusion technique can increase the operator awareness and the success rate of the target detection and tracking algorithms. Infrared (IR) imaging systems in a surveillance system have the capability of separating the foreground objects from the background with respect to their temperature and emissivity characteristics. However, due to low-resolution sensor arrays and possible absence of autofocus lens capabilities, high-frequency content of the objects is mostly missing. Rapid developments in the imaging technology and the availability of affordable devices in the market enable to use more than one imaging system for the applications mentioned herein. Consequently, in the modern surveillance systems, multiple cameras operating at different spectral ranges [midwave IR (MWIR) and long-wave IR (LWIR)] are used as a solution to overcome these difficulties [6,7].

In this study, a pixel-level image fusion technique, foretold to be computationally efficient and less lossy in terms of information, is proposed for the IR image fusion problem. The proposed technique is not only suitable for IR images but also applicable to a wide range of image types that are single channel and registered. In order to eliminate the undesired artifacts and obtain a descriptive fused image, a Laplacian of Gaussian (LoG) pyramid is constructed for multiscale edge detection and detected edges are fused by weighted averaging scheme. The weights for edge fusion are computed using a local discrimination metric by considering both interimage and intraimage region characteristics, and the direction of the weights are adapted to avoid suppression. Finally, adaptive unsharp masking is applied to enhance the visibility of the edges in the fused image. The proposed technique is suitable for real-time applications and also robust to content changes that may arise in the scene.

This paper is organized as follows: Section 2 provides a brief review about image fusion techniques. In Section 3, the proposed algorithm is explained. Experimental results are provided in Section 4. Comparisons including subjective test and quantitative fusion metrics, are presented in the same section. Concluding remarks are made and directions for future research are provided in the last section.

2. RELATED WORK ON IMAGE FUSION

The simplest method applied in image fusion is the averaging technique where pixel by pixel averages of the images are combined into a single image. This technique can result in suppression of edges and scene components due to the object emissivity characteristics. Direction of the contrast between foreground and background objects can vary between IR and visual band images, and fusion of low-pass image components may result in suppressed foreground objects. In addition, the human visual system is more sensitive to high-frequency components such as edges [8]. Frequency content of the scene components may vary with respect to spatial resolution of the imaging system, distance of the object from the camera system, level of background clutter, and included sensor and read-out circuit noises. Because of these variances, image pyramids, which are multiscale representations of the images, are proposed for decomposing the image into components that consist of different spatial frequencies [9]. Image pyramids-based techniques are frequently used in fusion [10–12]. In the pyramid approach, a set of bandpass filters is used to decompose scene components into groups that have different frequency characteristics. Burt [10] proposed an image fusion schema based on difference of low-pass pyramid. The Laplacian pyramid used in Burt's technique is obtained as the difference of the successive images in the Gaussian image pyramid. A pixel by pixel fusion is applied at Laplacian images of different levels with respect to pixel absolute values. A similar approach is adapted in [11] where pixel-level fusion is applied on Laplacian values of the image by choosing the directional absolute maximum values. Resultant images are reconstructed from fused Laplacian pyramids. Toet [12] proposed a similar technique where the image pyramid is constructed using the ratio of low-pass pyramid and fusion is achieved by comparing the ratio of successive Gaussian pyramid images because the human visual system is sensitive to local luminance contrast.

Discrete wavelet transform (DWT) is another technique frequently used in image fusion [13,14]. Traditionally, DWT can be used to create multiscale image representation and fusion can be accomplished by selection of salient wavelet coefficients and inverse DWT operation is applied on the fused wavelet coefficients. In the case of fusion of image sequences, to avoid flicker, Rockinger [15] proposed a method based on shift-invariant wavelet sequences. Ramac *et al.* [16] proposed a composite wavelet coefficient fusion technique that depends on the similarity of the image pairs at that location. Morphological filters are used to clean the image before wavelet decomposition and the averaging technique is used for similar locations whereas coefficients from the more salient region are copied when the regions are not similar. Petrovic *et al.* [17] proposed a wavelet-based fusion technique that is based on a "fuse-then-decompose" approach. Instead of traditional intensity-based multiscale fusion algorithms, Zheng *et al.* [18] proposed an undecimated transform-based approach that computes the support value transform of the input images to better define salient features as support values for each pixel and applied multiscale fusion procedure over these values. Zhenfeng *et al.* [19] performed a discrete curvelet transform-based approach to obtain the coefficient subbands in different scales and orientations. Then, they extracted focus measure values to perform a weighting strategy to fuse the low- and high-frequency components of the visible and infrared images.

Segmentation operation can also be used before multiscale representation for region-based parameter detection. Drajic and Cvejic [20] proposed a modified version of region-based dual-tree complex wavelet transform to create both region segments and multiscale representation of the images. Then, an algorithm is applied to determine which image region will be used for fusion operation based on their structural similarity measure [21]. To evaluate the performance of the region-based image fusion techniques, Cvejic and Seppanen *et al.* [22] defined a metric based on the similarities of the regions in the input images and the fused image by using contrast, size, and shape information.

Nonlinear mapping of salient features other than averaging is also used for image fusion. Xu *et al.* [23] proposed an

algorithm that uses Markov random field as a decision tool for fusion which can be applied both in multiscale and traditional single-scale approaches. We also proposed a pixel-level image fusion technique based on score assignment depending on local intensity variation in [24].

After fusion operation, artifacts such as destructive superposition and halo effects due to spectral operations appear in the resulting image. Ben-Shoshan and Yitzhaky [25] proposed a mask-based polarity adaptation of image pairs to avoid destructive superposition. Image fusion can also be used for enhancement. Xia and Kamel [26] proposed a cooperative neural fusion algorithm to enhance the restored images under blind Gaussian noise environment. They fused images as a linear combination of input images where scaling coefficients are obtained through the proposed neural fusion algorithm.

Kumar and Dass [27] proposed a pixel-level fusion technique using the total variation-based approach together with a principal component analysis technique.

3. CONTRAST DIRECTION ADAPTIVE IMAGE FUSION

In the proposed contrast direction adaptive fusion technique, to handle different scene object and background characteristics, an LoG pyramid is constructed to create a multiscale image representation. After the pyramid is constructed, for each level of the LoG pyramid, pixel-wise weighting coefficients and the pixel polarity reversion masks are computed using intraimage and interimage features. Finally, using these weighting coefficients, LoG image fusion, multiscale reconstruction, and adaptive unsharp masking operations are carried out. Details of each step are explained in the following subsections.

A. LoG Pyramid Construction

In the construction of the image pyramid, Laplacian pyramid-based multiscale decomposition is used since shift invariance property of Laplacian-based techniques is more advantageous than traditional wavelet-based techniques, as indicated by Rockinger *et al.* [15]. First, n levels of a Gaussian image pyramid is constructed using Burt's approach described in [9]. A sample image pyramid for a registered visual and LWIR band image pair can be seen in Fig. 1. After the Gaussian pyramid is constructed, an LoG pyramid is obtained by computing each pyramid layer as the difference of consecutive Gaussian pyramid layers, as given in Eq. (1):

$$L_i = G_i - [2\Uparrow](G_{i+1}), \qquad i = 0, \ldots, n-2, \tag{1}$$

where n is the level of the Gaussian pyramid, G_0 is the original input image, and $[2\Uparrow]$ denotes resize operator. During $[2\Uparrow]$ operation, bilinear interpolation is used in order to resize G_{i+1} to the size of the lower pyramid level G_i. A sample LoG pyramid constructed from the Gaussian pyramid in Fig. 1 is given in Fig. 2. LoG pyramid images are filtered to remove noisy responses as given in Eq. (2):

$$L_i^{\text{thr}}(x,y) = \begin{cases} 0 & \text{if } |L_i(x,y)| \le \tau_{\text{thres}} \\ L_i(x,y) & \text{if } |L_i(x,y)| > \tau_{\text{thres}} \end{cases}. \tag{2}$$

Here, (x, y) is the notation of element-wise operation and x and y correspond to column and row indices, respectively. τ_{thresh} is selected to be above the maximum noise level at the flat regions of the image, which depends on the camera parameters and scene brightness.

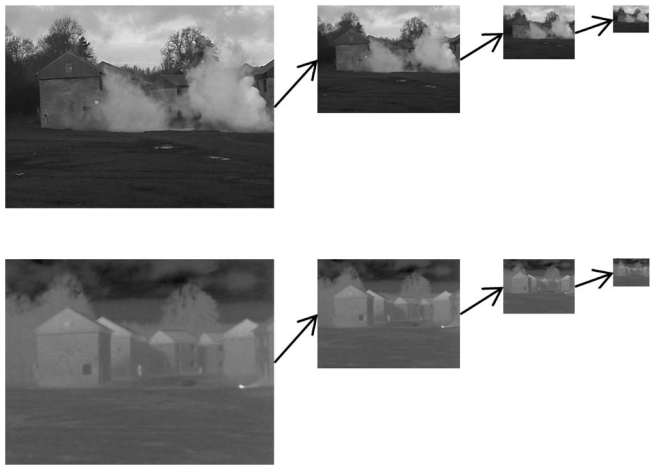


Fig. 1. Gaussian pyramid construction for $n = 3$.

B. Weight Calculation Scheme

LoG image data can be fused by salient point selection, summation, or averaging techniques. We propose an algorithm based on the weighted average of gradients values that are obtained from the LoG image pyramid. Weighting function is defined as the ratio of the LoG image pixel differences to the mean value of the surrounding neighborhood region. L_i is filtered with an averaging square filter of size $m/2^i$ to obtain M_i, which contains local average values for LoG image layers. The value of m is determined with respect to the size of the input images. Then, absolute difference images (D_i) are calculated by subtracting M_i from L_i and taking the absolute value of the results for each LoG image, as given in Eq. (3):

$$D_i(x,y) = |L_i(x,y) - M_i(x,y)|, \qquad i = 0, \ldots, n-2. \tag{3}$$

Deviation from the mean is used as a simple metric to measure how distinguishable a pixel is from its neighboring region.

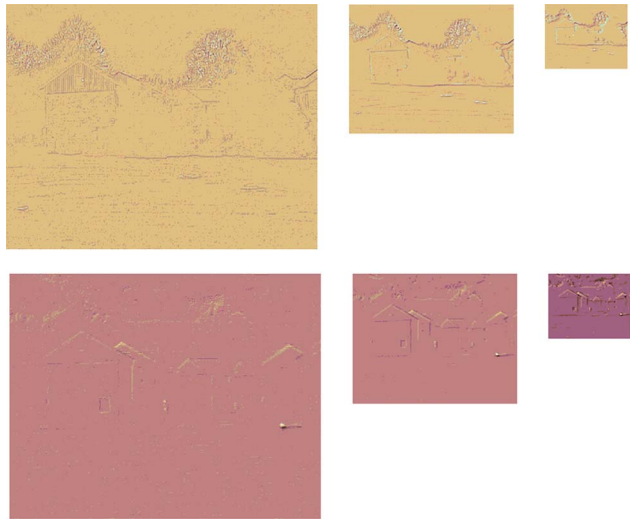


Fig. 2. LoG pyramid for $n = 3$.

Difference values calculated in this step are used to determine the weighting coefficients for pixel-wise LoG image fusion. Let D_i^1 and D_i^2 denote the difference values for the first and the second source images, respectively. The desired condition for fusion is to define a weighting factor near to one for regions that have a high difference value ratio across the fusion bands and, if the difference values D_i^1 and D_i^2 are similar, weighted sum of the coefficients becomes simply the average of the bands. Consequently, the ratio of the absolute difference values is defined to generate weighting coefficients to combine LoG pixel values. In the calculation scheme, a negligibly small constant α term is used to prevent division by zero. The weighting coefficients are computed for all of the pixels of each level LoG images, as given in Eq. (4):

$$W_i(x,y) = \frac{D_i^1(x,y)}{D_i^1(x,y) + D_i^2(x,y) + \alpha}, \quad i = 0, \ldots, n-2. \tag{4}$$

C. Contrast Direction Adaptation

As previously mentioned, to avoid suppression, direction of the edges for LoG images are adapted before applying weighted averaging operation. For this purpose, the sign of the values of the LoG images are extracted which are denoted as S_i and used to create contrast conversion matrix C_i. These operations are shown in Eq. (5):

$$S_i(x,y) = \begin{cases} 1 & \text{if } L_i^{\text{thr}}(x,y) \geq 0 \\ -1 & \text{otherwise} \end{cases}$$
$$C_i(x,y) = S_i^1(x,y) \cdot S_i^2(x,y). \tag{5}$$

S_i^1 and S_i^2 denote the sign matrices for the ith level LoG pyramid of the base and second images. When LoG images have different signed values in a pixel index, C_i is used to convert the gradient direction of one of the LoG pyramid images before accumulation operation. By this way, destructive superposition is avoided and high-frequency details extracted from both of the images are combined into the resultant fused image as described in detail in the next section.

D. Multiscale Reconstruction

A reconstruction process is applied after calculation of signature conversion and weighting matrices. Because this is a multiscale process, reconstruction is done step-by-step and at each step, a resized low-pass image from a lower level, weighting coefficients, and signal conversion matrices are used as input. The low-pass component of the first image is used as a base image and signal conversion operation is applied to the second input LoG images. After signal conversion, the weighted average of the LoG images is computed and added to the resized base image. Overall operations are shown in Eq. (6) and a single step of a reconstruction process is displayed in Fig. 3:

$$L_i^{\text{fused}}(x,y) = W_i(x,y) \cdot L_i^1(x,y) + C_i(x,y) \cdot (1 - W_i(x,y)) \cdot L_i^2(x,y)$$
$$I_i = L_i^{\text{fused}} + [2\Uparrow](I_{i+1}), \qquad i = 0, \ldots, n-2. \tag{6}$$

Here, I_i denotes the ith level fused image and hence, I_0 denotes the resultant fused image. For the sake of simplicity I_0 will be denoted as I, in the following sections.

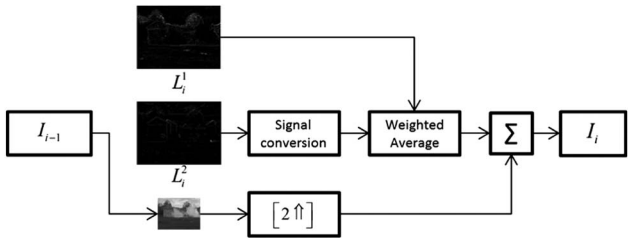


Fig. 3. Image reconstruction step.

E. Adaptive Unsharp Masking

Unsharp masking is a technique used to enhance the visibility of edges and corners by multiplying the high-frequency components of the images with a gain coefficient. Traditional unsharp masking technique is applied by using a fixed gain coefficient. However, there is a trade-off between the increased visibility of the details and the increased artificiality due to saturated regions. To avoid saturation at the regions near dynamic range limits, an adaptive unsharp masking technique is used [24]. In this technique, direction of the edge and the distance to the corresponding saturation limit is used to calculate pixel-wise gain coefficients.

First, a 3×3 Gaussian filter with standard deviation equal to one along horizontal and vertical directions is used to obtain low-frequency components (I_{LP}) of the fused image (I) calculated in the previous step. High-frequency components (I_{HP}) are obtained by subtracting I_{LP} from I. Directions of the edges are assigned using I_{HP} and the distance to the saturation point is obtained from the I_{LP}. Range R is calculated using these values, as given in Eq. (7):

$$R(x,y) = \begin{cases} I_{\text{LP}}(x,y), & I_{\text{HP}}(x,y) \leq 0 \\ 255 - I_{\text{LP}}(x,y), & I_{\text{HP}}(x,y) > 0 \end{cases}. \tag{7}$$

By using the R matrix values, a gain matrix G is calculated, as given in Eq. (8):

$$G(x,y) = \begin{cases} 1, & \frac{R(x,y)}{|I_{\text{HP}}(x,y)| \times 4} < 1 \\ \frac{R(x,y)}{|I_{\text{HP}}(x,y)| \times \beta}, & 1 \leq \frac{R(x,y)}{|I_{\text{HP}}(x,y)| \times \beta} < \beta \\ \beta, & \frac{R(x,y)}{|I_{\text{HP}}(x,y)| \times 4} \geq \beta \end{cases}. \tag{8}$$

Here, β denotes the gain coefficient of the technique. The fused image is computed using G matrix, as given in Eq. (9):

$$I_{\text{result}}(x,y) = I_{\text{LP}}(x,y) + G(x,y) * I_{\text{HP}}(x,y). \tag{9}$$

A general flow of the proposed fusion technique is provided in Fig. 4.

4. EXPERIMENTAL RESULTS

The proposed image fusion technique is applied to both multispectral, and multifocus image pairs. Six example image pairs selected from our fusion dataset are presented in Fig. 5. In Fig. 5, all image pairs except Fig. 5(a) have 320×256 spatial resolution. In the experiments, fusion results are obtained using three levels of a Gaussian pyramid where selection of the levels depends on the application specific constraints, such as background clutter and scene object size in pixel. Averaging square

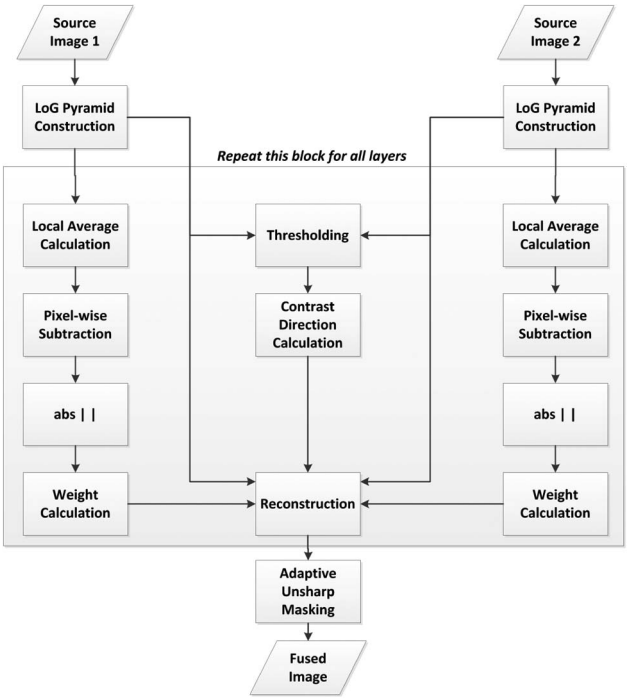


Fig. 4. Complete flow chart of the proposed image fusion technique.

filter of size 64, 32, and 16 is used for the calculation of M matrices. The threshold value τ_{thresh} for LoG pyramid to avoid noisy edge responses to change edge direction is selected to be 3 and the α coefficient is selected as 0.1, which is negligibly small to avoid division by zero condition. In the adaptive unsharp masking step, gain coefficient β is selected to be 4, which gives satisfying edge enhancement for the test images used in this work. Sample fusion results obtained using the proposed algorithm are presented in Fig. 6.

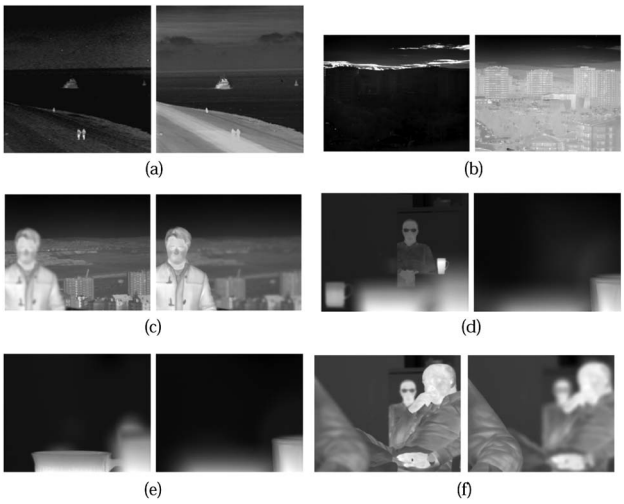


Fig. 5. Six example image pairs of our fusion dataset. Image pairs (a),(b): multispectral (MWIR-LWIR). Image pairs (c)–(f): LWIR multifocus.

Fig. 6. Results of the proposed fusion technique for image pairs presented in Fig. 5.

A. Performance Measures

Performance of the proposed fusion technique is evaluated using the objective measures including fusion metric based on mutual information (fast-FMI) [28], edge preservation coefficient ($Q^{AB/F}$) [29], fusion artifacts metric (NABF) [30], and visual information fidelity (VIFF) [31]. The fast-FMI technique calculates the regional mutual information between corresponding windows in fused and two source images. Higher values of fast-FMI measure are the indicator of better image fusion and the fast-FMI method is observed as consistent with the subjective results [28]. The $Q^{AB/F}$ measure is an objective pixel-level image fusion assessment framework based on normalized edge preservation measure. Lower values of the $Q^{AB/F}$ measure correspond to the more losses on the edge information [29]. The fusion artifacts (NABF) technique measures the level of noise or artifacts added to the fused image due to the fusion process [30]. Lower values of the NABF measure correspond to a better image fusion scheme. The VIFF method tries to measure how much "effective visual information (EFI)" in the fused image is extracted from source images, while EFI is defined as the maximum visual information of all the source-fused image pairs. The VIFF measure uses the visual information fidelity model to extract visual information from source-fused image pairs and it was experimentally

Table 1. Performance Results Comparison of Baseline and Proposed Image Fusion Techniques on the Objective Tests

	Fast-FMI [28]	$Q^{AB/F}$ [29]	NABF [30]	VIFF [31]
Zeeuw [14]	0.9453	0.8291	7.31e – 6	0.6141
Rockinger [15]	0.9343	0.7608	0.0019	0.6136
Burt [2]	0.9418	0.8661	0.0068	0.6559
Karali [24]	0.9450	0.8678	0.0135	0.6712
Mitanoudis [32]	0.9405	0.0088	0.0071	0.1718
Toet [12]	0.9405	0.8179	0.0206	0.7557
Proposed	0.9434	0.8732	0.0081	0.7180

observed that this measure is consistent with the subjective scores [31].

Performance of the proposed algorithm is also compared with the well-known image fusion techniques by evaluating the performance measures given above. The comparison of the average fusion results for the test images is provided in Table 1. The objective tests were carried out by using the image dataset mentioned in Section 4.B. By looking at the result presented in Table 1, one can say that the proposed technique gives promising results for the objective fast-FMI, $Q^{AB/F}$, and VIFF performance measures. The fast-FMI measure calculates the regional mutual information between the source and fused images, but the method responds positively in the case of blur type of artifact. In contrast, object/background separability is the main essence of the total image quality term. The proposed technique preserves and even enhances object/background separability while still providing satisfactory performance on the fast-FMI measure. The proposed method outperforms baseline methods on the $Q^{AB/F}$ measure because $Q^{AB/F}$ is based on the preservation of edges in source images. The proposed technique not only preserves but also enhances object edges in the source images. Therefore, the proposed technique is evaluated to be the most successful method in terms of the $Q^{AB/F}$ measure. The NABF technique measures the level of noise or artifacts in the fused image due to fusion process, but one of the main drawbacks of the NABF measure is that the method does not take into account the loss of information and contrast. Therefore, the fusion techniques that introduce enhancement on edges are evaluated as more artificial by the NABF measure. Therefore, the NABF measure results should be used with other performance measures instead of an individual performance evaluation. The VIFF measure tries to extract visual information from source-fused image pairs. The proposed technique obtains the second highest score on the VIFF measure but the fusion technique [12] that achieves the highest score on VIFF introduces shading type of artifacts.

The objective test showed that a general image quality term should be defined instead of an individual performance evaluation based on objective measures. Besides, the objective tests brought forward the necessity of performing subjective tests since objective measures fail to provide reliable results, especially the techniques introduces blurring, shading, and other types of artifacts. Moreover, subjective evaluation is a more accurate and reliable way to determine fusion performance in terms of a perceived level of quality by the observers.

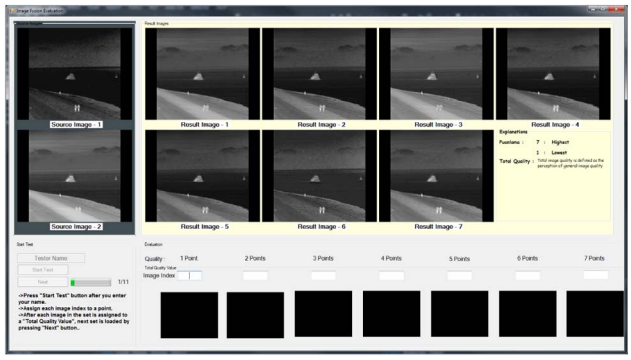


Fig. 7. Graphical user interface developed for subjective tests.

B. Subjective Tests

Since the statement of quality is generally considered as a subjective term, the subjective evaluation is formally defined to be the most accurate and reliable tool in the assessment of visual quality satisfying that the number of subjects is sufficiently large. In order to determine the performance of the image fusion techniques in a subjective manner, a subjective evaluation procedure is established. In the subjective tests, 24 people were asked to score the fusion techniques depending on their success on "total image quality." Total image quality is defined as the perception of general image quality in terms of the existence of source image scene components in the fused image, background/foreground separability, scene object details, and artificiality. The images containing higher background/foreground separability and object details are more likely to be scored with higher grades by the subjects. The subjects who participated in the tests consisted of mostly males (19 males, five females) and nearly all of the subjects have their Bachelor of Science Degree in engineering fields. In performance sorting, the subjects were asked to provide scores between 1 and 7 for the total image quality. Higher values of total image quality metric stands for better fusion performance. In the graphical user interface displayed in Fig. 7 designed for the subjective tests, the source images used in the fusion and the fused images corresponding to each fusion technique are visualized. In this manner, the subject has the opportunity to compare the results of the different image fusion frameworks. The subjects were asked to score totally 11 image pairs and the average scores of each fusion method are provided in Table 2. The average scores of the subjects corresponding to each image pair is also provided in Fig. 8.

Table 2. Performance Results Comparison of Proposed and Baseline Image Fusion Techniques on the Subjective Tests

Fusion Methods	Subjective Scores
Zeeuw [14]	3.276
Rockinger [15]	2.602
Burt [2]	4.102
Karali [24]	5.681
Mitanoudis [32]	2.958
Toet [12]	3.515
Proposed	5.863

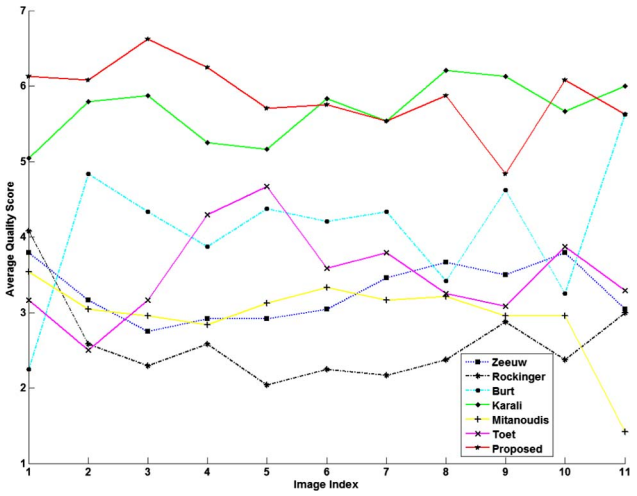
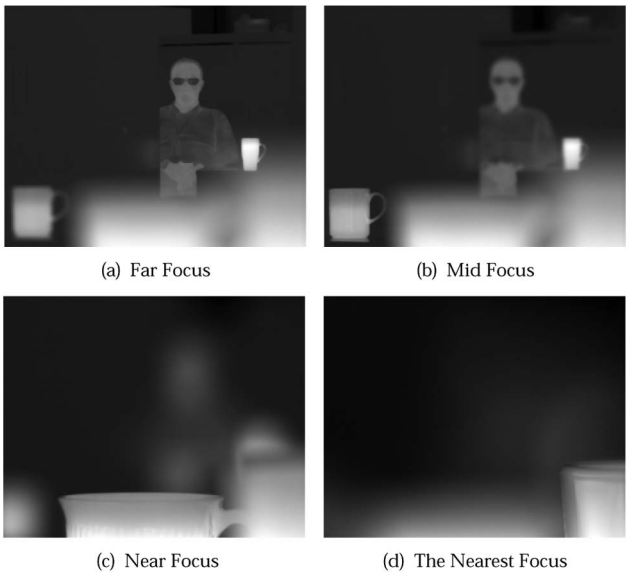


Fig. 8. Average quality scores for the test image sequences.

By looking at the results provided in Table 2, one can say that the proposed technique outperforms the baseline techniques in total image quality criterion. In subjective tests, our previous method [24] provided comparable performance with the method proposed herein. In the tests, it is observed that the subjects are more likely to give higher scores to the methods that emphasize object edges without introducing undesired artifacts. The foreground/background separation based on contrast is another important factor in the evaluation of image fusion performance. However, it was discovered that the preservation of contrast on the fused image was not taken into account by some of the subjects. Still, the proposed technique outperforms the baseline techniques in subjective tests.

As an additional experiment, the images captured at four different focus instances are fused using the proposed and baseline fusion schemes. In this framework, the method is evaluated subjectively to determine whether it preserves the object details



(a) Far Focus (b) Mid Focus

(c) Near Focus (d) The Nearest Focus

Fig. 9. Objects captured at different focus instances.

(a) Zeeuw [14] (b) Rockinger [15]

(c) Burt [2] (d) Karali [24]

(e) Mitanoudis [32] (f) Toet [12]

(g) Proposed

Fig. 10. Fusion results obtained using proposed and baseline fusion frameworks for the images shown in Fig. 9.

captured at different focus instances. In Fig. 9, the objects located at different distances and captured using different focus instances are illustrated within the multifocus fusion framework. The images presented in Fig. 9 are fused by baseline and proposed image fusion frameworks and the fusion results are provided in Fig. 10. By looking at the results provided in Fig. 10, one can say that the proposed scheme preserves the object details better than the baseline fusion techniques. The proposed technique not only preserves object details and object/background separability but also it does not introduce undesired artifacts such as shading, griding, etc. The stationary wavelet technique [15] introduces diagonal artifacts on the fused image and it reduces the object/background separability. Similarly, the mean wavelet fusion technique [14] reduces the object/background separability while smoothing the objects in the image. Fusion by selection scheme [2] preserves object

details and object/background separability but the method introduces undesired artifacts (griding) to the objects in the fused image. The Laplacian-based fusion scheme [12] loses some object/background separation and object details (handle of the mugs, glasses, etc.) due to smoothing and the method introduces shading type of artifacts on the outer contour of the objects. The statistical fusion method [24] enhances the edges around objects but it introduces noisy textures and weakens object/background separation. The principal component analysis (PCA)-based scheme [32] worked quite well on image pairs but, the method could not preserve object details [Fig. 10(e)] in a multifocus fusion framework.

The proposed algorithm is implemented using MATLAB on a computer with Intel i7-3770 processor with 8.0 Gb of RAM. Average processing time of an image pair of resolution 320×256 is measured 59 ms, which is suitable for real-time implementations.

5. CONCLUSIONS AND FUTURE WORK

In this paper, a multiscale contrast direction adaptive image fusion technique that is both applicable to multispectral and multifocus image pairs is proposed. In order to evaluate the performance of the proposed technique, well-known image performance measures, such as fast-FMI, $Q^{AB/F}$, NABF, and VIFF, are used and the performance measure values are listed for the proposed and baseline fusion schemes. The proposed fusion technique provides comparable results with other baseline techniques in objective performance measures. In order to obtain a more reliable performance evaluation, subjective tests are carried out. According to the subjective tests, the proposed fusion framework outperforms baseline techniques in total quality measure. The proposed method and baseline techniques were also used in the fusion scheme that contains more than two source images and it was observed that the proposed framework preserves object details and prevents undesired artifacts better than the baseline fusion methods. The computational complexity of the proposed fusion scheme was also evaluated and the method is found to be suitable for the real-time applications.

The authors would like to thank the observers for joining in the subjective evaluations.

REFERENCES

1. Z. Zhang and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," Proc. IEEE **87**, 1315–1326 (1999).
2. P. J. Burt and J. Kolczynski, "Enhanced image capture through fusion," in *IEEE 4th International Conference on Computer Vision* (IEEE, 1993), pp. 173–182.
3. G. Piella, "Adaptive wavelets and their applications to image fusion and compression," Ph.D. thesis (University of Amsterdam, 2003).
4. A. Galande and R. Patil, "The art of medical image fusion: A survey," in *International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2013), pp. 400–405.
5. J. Byeungwoo and D. Landgrebe, "Decision fusion approach for multitemporal classification," IEEE Trans. Geosci. Remote Sens. **37**, 1227–1233 (1999).
6. S. K. Chari, J. D. Fanning, S. M. Salem, A. L. Robinson, and C. E. Halford, "LWIR and MWIR fusion algorithm comparison using image metrics," Proc. SPIE **5784**, 16–26 (2005).
7. S.-l. Han, P. Zhang, and W.-l. Hu, "Research on dual-band image fusion algorithms and simulation based on infrared radiation characteristics," Proc. SPIE **8907**, 89070Q (2013).
8. B. Girod, "What's wrong with mean-squared error?" in *Digital Images and Human Vision* (MIT, 1993), pp. 207–220.
9. P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," IEEE Trans. Commun. **31**, 532–540 (1983).
10. P. J. Burt and E. H. Adelson, "Merging images through pattern decomposition," Proc. SPIE **0575**, 173 (1985).
11. J. Scott and M. A. Pusateri, "Laplacian based image fusion," in *IEEE Applied Imagery Pattern Recognition Workshop* (IEEE, 2010), pp. 1–7.
12. A. Toet, "Image fusion by a ratio of low-pass pyramid," Pattern Recogn. Lett. **9**, 245–253 (1996).
13. L. J. Chipman, T. M. Orr, and L. N. Graham, "Wavelets and image fusion," Proc. SPIE **2569**, 208–219 (1995).
14. P. M. Zeeuw, "Wavelet and image fusion," CWI, Amsterdam, http://www.cwi.nl/~pauldz/ (1998).
15. O. Rockinger, "Image sequence fusion using a shift invariant wavelet transform," in *IEEE Proceedings of the International Conference on Image Fusion* (IEEE, 1997), pp. 288–291.
16. L. C. Ramac, M. K. Uner, P. K. Varshney, M. Alford, and D. Ferris, "Morphological filters and wavelet based image fusion for concealed weapons detection," Proc. SPIE **3376**, 110–119 (1998).
17. V. S. Petrovic and C. S. Xydeas, "Gradient-based multiresolution image fusion," IEEE Trans. Image Process. **13**, 228–237 (2004).
18. S. Zheng, W. Shi, J. Liu, G. Zhu, and J. Tian, "Multisource image fusion method using support value transform," IEEE Trans. Image Process. **16**, 1831–1839 (2007).
19. S. Zhenfeng, L. Jun, and C. Qimin, "Fusion of infrared and visible images based on focus measure operators in the curvelet domain," Appl. Opt. **51**, 1910–1921 (2012).
20. D. Drajic and N. Cvejic, "Adaptive fusion of multimodal surveillance image sequences in visual sensor networks," IEEE Trans. Consum. Electron. **53**, 1456–1462 (2007).
21. Z. Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Process. Lett. **9**, 81–84 (2002).
22. N. Cvejic and S. J. G. T. Seppanen, "A nonreference image fusion metric based on the regional importance measure," IEEE J. Sel. Top. Signal Process. **3**, 212–221 (2009).
23. M. Xu, H. Chen, and P. K. Varshney, "An image fusion approach based on Markov random fields," IEEE Trans. Geosci. Remote Sens. **49**, 5116–5127 (2011).
24. A. O. Karali, S. Cakir, O. E. Okman, and T. Aytac, "Image fusion using statistical intensity fusion and adaptive unsharp masking," in *IEEE Signal Processing and Communications Applications Conference (SIU)* (IEEE, 2012), pp. 1–4.
25. Y. Ben-Shoshan and Y. Yitzhaky, "Enhancement of image fusion methods," Proc. SPIE **2569**, 208–219 (1995).
26. Y. Xia and M. S. Kamel, "Novel cooperative neural fusion algorithms for image restoration and image fusion," IEEE Trans. Image Process. **16**, 367–381 (2007).
27. M. Kumar and S. Dass, "A total variation-based algorithm for pixel-level image fusion," IEEE Trans. Image Process. **18**, 2137–2143 (2009).
28. M. Haghighat, A. Aghagolzadeh, and H. Seyedarabi, "A non-reference image fusion metric based on mutual information of image features," Comput. Electr. Eng. **37**, 744–756 (2011).
29. C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure," Electron. Lett. **36**, 308–309 (2000).
30. B. K. S. Kumar, "Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform," Sig. Image Video Process. **7**, 1125–1143 (2012).
31. Y. Han, Y. Cai, and X. Xu, "A new image fusion metric based on visual information fidelity," Inf. Fusion **14**, 127–135 (2013).
32. N. Mitanoudis and T. Stathaki, "Optimal contrast correction for ICA-based fusion of multimodal images," IEEE Sens. J. **8**, 2016–2026 (2008).