Scale-Aware Multispectral Fusion of RGB and NIR Images Based on Alternating Guidance

KAILONG ZHOU, (Student Member, IEEE), CHEOLKON JUNG®, (Member, IEEE), AND SHENGTAO YU
School of Electronic Engineering, Xidian University, Xi’an 710071, China
Corresponding author: Cheolkon Jung (zhengzk@xidian.edu.cn)

This work was supported in part by the Huawei Innovation Research Program under Grant HF2018095025, in part by the National Natural Science Foundation of China under Grant 61872280, and in part by the International Science and Technology Cooperation Program of China under Grant 2014DFG12780.

ABSTRACT In low light condition, color (RGB) images captured by increasing the camera ISO contain much noise and detail loss. However, near infrared (NIR) images are robust to noise and have clear textures without color. In this paper, we propose scale-aware multispectral fusion of RGB and NIR images based on alternating guidance. Low light RGB images provide large-scale image structure and color information, while NIR images have fine details lost in RGB images. Since they are complementary, we adopt alternating guidance for the fusion of them using weighted least squares (WLS). First, we perform the first guidance to denoise the RGB image and obtain base layer. Then, we conduct the second guidance for scale-aware detail transfer of the NIR image and yield detail layer. Finally, we combine the base and detail layers to generate a fusion image. We maximize the multispectral advantage of RGB and NIR images for fusion based on alternating guidance. Experimental results show that the proposed method achieves good performance in noise reduction, detail transfer and color reproduction, and is superior to the state-of-the-art ones in terms of quantitative measurement and computational efficiency.

INDEX TERMS Image fusion, alternating guidance, detail transfer, weighted least squares, scale-aware.

I. INTRODUCTION

With advances in the sensor technology, the image types are highly diversified. In addition to the widely used visible RGB cameras, there exist depth cameras to record depth information, infrared (IR) and NIR cameras for invisible wavelength band imaging, and X-ray cameras for medical imaging. Due to the increasing demands for computer vision applications, the requirements for imaging quality are becoming higher and higher. To maximize the advantages of various sensors, image fusion uses multiple sensors to improve imaging quality and accuracy of vision applications [20], [28]. Recent image fusion methods include fusion of flash/no-flash images [19], infrared/RGB images [17], near infrared (NIR)/RGB images [13], [21], [34]. In this paper, we focus on the multispectral fusion of RGB and NIR images in low light condition.

A. RELATED WORK

In low light condition, it is hard to capture high quality RGB images without flash due to low signal-to-noise ratio (SNR). A commonly used method is to increase the camera ISO setting in low light condition. However, the quality of RGB images is seriously degraded by noise at a higher ISO value, and thus the edges and details are severely destructed. IR band imaging acquires images stably in adverse environments such as low light. It is widely used in object recognition [25], object detection [5], video surveillance [9], and remote sensing [24]. Although IR images have the advantage of resisting unfavorable environments, they generally have low resolution with poor textures. These defects limit IR images in certain applications requiring high-quality images, such as night vision systems. However, in the NIR band (750-1400 nm) close to the visible band [2], the NIR images have high resolution, clean textures, and robustness to noise. Therefore, compared with IR images, NIR images are more suitable for night vision systems that require higher image quality such as video surveillance [22]. Although NIR images do not have color information, it is easy to simultaneously acquire RGB
and NIR images which are complementary through hybrid camera systems. Therefore, acquiring a fusion image with color information and clean details through the fusion of RGB and NIR images provides a solution to high-quality imaging in low light condition.

Prior to introducing RGB and NIR image fusion, we first review RGB and IR image fusion. The research on RGB and IR image fusion has a long history, which can be divided into seven categories: multiscale transform [10], [18], [35], sparse representation [11], [28], neural networks [8], [29], subspace methods [1], [7], and saliency-based methods [33], [36], hybrid models [12], [15], and other methods [14], [37]. NIR images provide higher resolution and better details than IR ones in low light condition. Therefore, the fusion of RGB and NIR images is more suitable for producing high-quality fusion results in low light condition. However, due to the difference in contrast and structure between the luminance channel of the RGB image and the NIR image, the NIR image cannot be directly replaced with the luminance channel of the RGB image. The direct replacement of the luminance channel by the NIR image leads to the color distortion and structural destruction. To solve the contrast difference, Son [26] proposed a method for low-light color image denoising based on contrast conversion between NIR images and luminance channels. Son et al. [27] further proposed an NIR coloring method using a contrast-preserving mapping model. To successfully preserve structural information of RGB and NIR images, Shibata et al. [23] proposed a fusion method based on high visibility area selection. Yan et al. [30] explicitly modelled derivative-level confidence and proposed cross field joint image fusion by optimizing a scale map.

B. CONTRIBUTIONS

In this paper, we propose scale-aware fusion of RGB and NIR images in low light condition based on alternating guidance. Noisy RGB images contain color information and big-scale image structures, while NIR images include small-scale fine textures. We adopt alternating guidance for fusion of RGB and NIR images based on WLS to make full use of their own advantage. First, we perform the first guidance for denoising on noisy RGB image and obtain base layer. For the first guidance, the joint guidance of NIR image and denoised luminance of RGB image was employed to remove noise while retaining edges. Since the NIR imaging highly depends on the NIR light strength, NIR images have no or very small values beyond the range that the NIR light reaches. This usually happens in the outdoor environment, especially for the background at night time, which makes the noisy RGB information also useful for fusion (see Fig. 1). Thus, we perform sigmoid-based NIR weighting for base layer generation (BLG) to achieve different smoothing degrees according to the NIR intensity. Then, we conduct the second guidance for scale-aware detail transfer on NIR image and obtain detail layer. Compared with direct smoothing on NIR image, the second guidance is able to generate more complete small-scale textures lost in noisy RGB image. Finally, we combine base and detail layers to produce a fusion result. As shown in Fig. 1, the proposed method successfully transfers details...
from the NIR image to the fusion result with noise removal and color reproduction. We capture them in low light condition by JAI AD-130 GE camera. This camera is able to simultaneously capture RGB and NIR images through the same optical path with two CCDs. As shown in the figure, RGB images captured in the low illumination condition have color with severe noise and detail loss. However, NIR images captured in the same condition have fine details with little noise. Thus, they are complementary, and the fusion of them is able to take both advantages. The preliminary result of this paper was presented in [38]. In this paper, we extend our previous work in the following three points. First, we perform sigmoid-based NIR weighting to selectively take both NIR textures (foreground) and RGB information (background) in fusion. Second, we remove NIR highlights in fusion by the joint guidance of NIR image and Y channel (RGB image). Third, we capture real image pairs by JAI AD-130 GE camera, and verify the effectiveness of the proposed method on them.

Compared with existing methods, main contributions of the proposed method are as follows:

- We propose scale-aware fusion of RGB and NIR images based on alternating guidance to make full use of the multispectral advantage. Low light RGB images contain big-scale image structures with much noise, while NIR images include small-scale fine textures without color. Since they are complementary, we adopt alternating guidance to achieve scale aware fusion of the paired images.
- We adopt WLS to alternately use RGB and NIR images as guidance for noise removal, detail transfer and NIR highlight removal. WLS is an edge-aware smoothing filter based on global optimization, which is used in the first guidance (RGB image denoising) to get base layer and in the second guidance (NIR texture transfer) to obtain detail layer. We combine the base and detail layers to generate a fusion image.
- We achieve high computational efficiency with $O(N)$ time complexity by WLS-based fast global smoothing.

II. PROPOSED METHOD

Fig. 2 shows the whole framework of the proposed scale-aware RGB/NIR image fusion based on alternating guidance. We adopt weighted least squares (WLS) for fast global smoothing that is a global edge preservation filter. We perform scale-aware alternating guidance for fusion of RGB and NIR images based on WLS. We use the combination of NIR image and denoised RGB luminance channel as joint guidance to get base layer in the first guidance. The first guidance effectively removes noise while protecting structure of RGB image. We utilize an over-smoothed RGB image as guidance to obtain detail layer in the second guidance. Since this over-smoothed RGB image only provides big-scale structure of the RGB image, the details of different scales and contrasts lost in the RGB image are completely taken from the NIR image by guiding the NIR image smoothing. The second guidance successfully takes clear texture from NIR image. Finally, we combine base and detail layers to generate fusion image.

A. SIGMOID-BASED NIR WEIGHTING

To effectively aggregate the multispectral information, we perform independent processing of the foreground and background segmented by the NIR intensity. Since the NIR imaging highly depends on the NIR intensity, the NIR image contains rich information within the range that the NIR light can reach. However, the NIR intensity has little or no information beyond the range. Thus, we mainly use the NIR information within the range, while we mainly utilize the RGB information above the range. To achieve different smoothing

1 https://www.jai.com/products/ad-130-ge
degrees to the NIR intensity, we perform sigmoid-based NIR weighting as follows:

\[ M(i) = \frac{1}{1 + e^{-\varepsilon(i-\tau)}} \]  

(1)

where \( \varepsilon \) represents the NIR intensity, \( \varepsilon \) is a parameter that adjusts the steepness of the curve, and \( \tau \) is a parameter that controls the threshold.

### B. WEIGHTED LEAST SQUARES

WLS is an edge-aware smoothing filter based on global optimization formulation [16], which consists of a data term and a prior term. The prior term is represented for smoothing by a weighted \( L_2 \) norm. Given an input image \( f \) and a guidance image \( g \), an output image \( u \) is obtained by minimizing the following WLS energy function:

\[ \varepsilon(u) = \sum_p (u_p - f_p)^2 + \lambda \sum_{q \in N(p)} \omega_{p,q}(g)(u_p - u_q)^2 \]  

(2)

where \( N(p) \) represents a set of four adjacent pixels of \( p \); \( \lambda \) controls the balance between data and smoothing terms, and increasing \( \lambda \) results in smoothing output; and \( \omega_{p,q}(g) \) is the weight calculated from the guidance image \( g \) and a measure the similarity between pixels \( p \) and \( q \). \( \omega_{p,q}(g) \) is defined as follows:

\[ \omega_{p,q}(g) = \exp(-\|g_p - g_q\|/\sigma) \]  

(3)

where \( \sigma \) is a range parameter. The energy function in Eq. (1) is transformed into a vector form as follows:

\[ \varepsilon(u) = (u - v)^T(u - v) + \lambda u^T A_g u \]  

(4)

where \( u \) and \( v \) denote \( S \times 1 \) column vectors containing values of \( u \) and \( v \), respectively, and \( S \) is the total number of pixels; \( T \) denotes the transposition; and \( A_g \) is \( S \times S \) Laplacian matrix as follows [4]:

\[ A_g(m, n) = \begin{cases} \sum_{l \in N(m)} \omega_{m,l}(g) & n = m \\ -\omega_{m,n}(g) & n \in N(m) \\ 0 & \text{otherwise} \end{cases} \]  

(5)

Based on a large sparse matrix, this energy function can be solved through a linear system as follows:

\[ (I + \lambda A_g) u = v \]  

(6)

However, solving it by matrix inversion is of high computational complexity. By fast global smoothing of WLS, the time complexity can reach \( O(N) \). First, we consider one-dimensional (1D) case assuming that WLS energy function works on a 1D horizontal input signal \( f^h \) and a 1D guiding signal \( g^h \) along \( x \) dimension (\( x = 0, \ldots, W - 1 \)). The energy function of the 1D signal is as follows:

\[ \sum_{x} (u^h_x - f^h_x)^2 + \lambda_t \sum_{i \in N_h(x)} \omega_{x,i}(g^h)(u^h_i - u^h_x)^2 \]  

(7)

where \( N_h(x) \) represents two neighbors of \( x \). This energy function is minimized by the following linear equation:

\[ (I_h + \lambda_t A_h) u^h = f^h \]  

(8)

where \( I_h \) is an identity matrix with a size of \( W \times W \); \( u^h \) and \( f^h \) represent the vector notations of \( u^h \) and \( f^h \), respectively; \( A_h \) is a three-point Laplacian matrix with a size of \( W \times W \). The linear system in Eq. (7) is written as follows:

\[
\begin{bmatrix}
b_0 & c_0 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \cdots & \vdots \\
0 & a_x & b_x & c_x & 0 \\
0 & 0 & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & a_{W-1} & b_{W-1} \\
\end{bmatrix}
\begin{bmatrix}
u^h_0 \\ \vdots \\ u^h_k \\ \vdots \\ u^h_{W-1} \end{bmatrix}
= \begin{bmatrix}
f^h_0 \\ \vdots \\ f^h_k \\ \vdots \\ f^h_{W-1} \end{bmatrix}
\]  

(9)

where \( u^h_k \) and \( f^h_k \) are the \( x_{ih} \) elements of \( u^h \) and \( f^h \), respectively; \( a_x, b_x, \) and \( c_x \) represent three nonzero elements in the \( x_{ih} \) row of \( (I_h + \lambda_t A_h) \). In boundary condition, \( a_0 = 0 \) and \( c_{W-1} = 0 \), \( a_x, b_x \) and \( c_x \) are written as:

\[ a_x = \lambda_t A_h(x, x-1) = -\lambda_t \omega_{x-1, x} \]
\[ b_x = 1 + \lambda_t A_h(x, x) = 1 + \lambda_t (\omega_{x-1, x} + \omega_{x, x+1}) \]
\[ c_x = \lambda_t A_h(x, x+1) = -\lambda_t \omega_{x, x+1} \]  

(10)

Matrix \((I_h + \lambda_t A_h)\) is a tridiagonal matrix whose nonzero elements exist only in the left and right diagonals. By Gaussian elimination, it reaches \( O(N) \) complexity. In Gaussian elimination, intermediate \( \tilde{c}_x \) and \( \tilde{f}_x^h \) are computed as follows:

\[ \tilde{c}_x = c_x / (b_x - \tilde{c}_{x-1} a_x) \]
\[ \tilde{f}_x^h = (f_x^h - \tilde{f}_{x-1}^h a_x) / (b_x - \tilde{c}_{x-1} a_x), \quad (x = 1, \ldots, W - 1) \]  

(11)

where \( \tilde{c}_0 = c_0 / b_0 \) and \( \tilde{f}_0^h = f_0^h / b_0 \). Then, output \( u^h_k \) is obtained by:

\[ u^h_k = \tilde{f}_k^h - \tilde{c}_k u^h_{k+1}, \quad (x = W - 2, \ldots, 0) \]  

(12)

with \( u^h_{W-1} = \tilde{f}_{W-1}^h \). To process a two-dimensional (2D) image signal by using 1D solver, we perform 1D global smoothing operations along each dimension of 2D signal. To prevent the streaking artifact which commonly appears in separable algorithms [3], we perform 2D smoothing by applying sequential 1D global smoothing to a multiple number of iterations [16]. In this scheme, \( \lambda_t \) in each iteration is computed as follows:

\[ \lambda_t = \left( \frac{3}{4} \right)^{\frac{2t}{T}} \frac{T}{2} \]  

(13)

where \( T \) represents the total number of iterations along each dimension. In each iteration, we perform 1D solver with parameter \( \lambda_t \) along \( x \) dimension and \( y \) dimension of 2D images continuously.
C. FIRST GUIDANCE FOR NOISY RGB DENOISING

We perform the first guidance for RGB denoising and obtain the base layer. For the first guidance, we utilize both NIR image and denoised luminance channel of RGB image. NIR images contain image structure with clean textures and little noise. Thus, using NIR image as guidance achieves good denoising performance on noisy RGB image. However, due to the contrast difference between NIR image and RGB luminance channel in Fig. 3, direct adoption of NIR image as guidance would ruin some edges of RGB image when the contrast of NIR image is not matched with that of RGB luminance channel. Moreover, using RGB image itself as the guidance (self-guidance) can not use NIR clean textures for denoising. Thus, we adopt both NIR image and denoised luminance channel for guidance to remove noise while keeping image structure [32].

To remove salt-and-pepper noise in $Y_I$, we use fast weight median filter (FWMF) and get $Y_n$. Then, we perform element-wise addition of $Y_n$ and NIR image $N_l$ to obtain the guidance for denoising noisy RGB image $C_I$. The base layer $C_B$ is obtained by minimizing the following energy function:

$$
\varepsilon(C_B) = (C_B - C_I)^T(C_B - C_I) + \lambda_2 C_B^T A_G C_B
$$

(14)

where $C_I$ represents column vectors containing values of $C_I$; $A_G$ denotes Laplacian matrix defined by $Y_n + N_l$; and range parameter $\sigma$ is $\sigma_2$. Fig. 4 shows RGB denoising results by the first guidance. Guided by $C_I$, i.e. self-guidance, the first guidance can not achieve satisfactory denoising performance (see Fig. 4(b)). Guided by $N_l$, i.e. NIR image, the first guidance makes the picture much blurry (see Fig. 4(c)). When only NIR image is used as guidance, the contrast difference between RGB and NIR images such as red boxes in Fig. 3 would cause serious blurs in edges. Guided by $Y_n + N_l$, the first guidance successfully removes noise while preserving the structure of RGB image (see Fig. 4(d)).

D. SECOND GUIDANCE FOR SCALE-AWARE DETAIL TRANSFER

In low light condition, RGB images often lose details due to the severe noise. Thus, we use the input NIR image to recover the lost details in the noisy RGB image. To transfer multiscale details of the NIR image, we use the RGB image with the big-scale structure to guide the NIR detail transfer, called scale-aware detail transfer. We obtain the second guidance image by over-smoothing the base layer of the input RGB image to contain the big-scale structure. Therefore, based on the second guidance, the details of multiple scales and contrasts in the NIR image are successfully transferred. As shown in Fig. 5(a), the black texts on the tea box in noisy RGB image are mostly destructed and only their color is identified. In contrast, they are very clear in NIR image. However, the lost details in noisy RGB image have various scales and textures as shown in Fig. 5(a) (see the red box in NIR image). In practice, details in NIR image have various scales and textures from small to large, and thus it is inappropriate to apply the same smoothing to NIR image as the second guidance for detail transfer. Directly applying smoothing to NIR image, i.e. self-guidance, causes loss of some textures and color distortion. If we utilize small scale smoothing on NIR image as the second guidance, large scale textures would not be transferred (see Fig. 5(b)). On the contrary, if we use large scale smoothing on NIR image as the second guidance, most details are smoothed so that some unwanted details are transferred (see Fig. 5(c)). The unwanted details would cause serious color distortion in fusion results compared with the input color image. Thus, we use over-smoothed RGB image, $C_S$, as the second guidance for detail transfer. We obtain $C_S$ from base layer $C_B$, i.e. the output of the first guidance. Guided by $C_S$, we successfully generate detail layer $N_D$ from input NIR image $N_l$. Since $C_S$ maintains large scale structure of RGB image, the second guidance acquires details with

![Figure 3. Contrast difference between RGB and NIR images. Left: RGB image. Right: NIR image.](image)

![Figure 4. RGB denoising comparison by the first guidance. Top-left: Noisy RGB image. Top-right: Guided by $C_I$ with small denoising coefficient. Bottom-left: Guided by $N_l$ with big denoising coefficient. Bottom-right: Guided by $Y_n + N_l$.](image)
FIGURE 5. (a) Input NIR (top) and RGB (down) images. (b) Detail transfer guided by small scale NIR smoothing. (c) Detail transfer guided by large scale NIR smoothing. (d) Proposed method guided by $C_S$.

FIGURE 6. NIR highlights in the human eyes.

Various scales from NIR image that are lost in noisy RGB image (see Fig. 5(d)). For the second guidance, we first obtain $C_S$ from $C_B$ using WLS as follows:

$$\varepsilon(C_S) = (C_S - C_B)^T (C_S - C_B) + \lambda_3 C_S^T A C_B C_S$$

(15)

where $C_B$ represents vector norm of $C_B$; $A_{C_B}$ denotes Laplacian matrix defined by $C_B$; and the range parameter $\sigma$ is $\sigma_3$. By minimizing the energy function Eq. (15), we obtain $C_S$.

Then, we obtain the smoothed NIR image $N_S$ under the guidance of $C_S$ by minimizing the following energy function:

$$\varepsilon(N_S) = (N_S - N_I)^T (N_S - N_I) + \lambda_4 N_S^T A C_S N_S$$

(16)

where $N_I$ represents vector norm of $N_I$; $A_{C_S}$ denotes Laplacian matrix defined by $C_S$, and the range parameter $\sigma$ is $\sigma_4$. We acquire $N_D$ through a pixel-wise subtraction between $N_I$ and $N_S$.

E. REMOVAL OF NIR HIGHLIGHTS

In NIR images, highlights often appear in the human eyes. They usually happen at night because the pupils dilate to allow in more light. Much of the NIR light passes into the eye through the pupils, and then it is reflected through them. The
K. Zhou et al.: Scale-Aware Multispectral Fusion of RGB and NIR Images

NIR camera records this reflected light. This is very similar to the red-eye effect in flash photography. Fig. 6 shows the NIR highlights that appear in the human eyes. To suppress them, we perform the second guidance using the image smoothed by joint NIR image and Y channel of CS, instead of the smoothed base layer CS.

For a comparison, we provide the fusion results by two smoothed images. As shown in Fig. 7, the second guidance by the joint guidance of NIR and Y channel of CS successfully removes the NIR highlights of the human eyes in fusion.

F. LAYER FUSION
We reconstruct fusion image by combining base layer CB and detail layer ND. First, we convert CB to YUV color space as follows:

\[
\begin{align*}
Y_B &= 0.299R + 0.587G + 0.114B \\
U_B &= -0.147R - 0.289G + 0.436B \\
V_B &= 0.615R - 0.55G - 0.100B
\end{align*}
\] (17)

where R, G, B represent red, green and blue channels of CB, respectively. We use YUV color space to transfer the NIR details to the fusion without color shift.

Then, we combine YB and ND through pixel-wise add operation to generate the fused luminance channel Y as follows:

\[
Y = Y_B + N_D
\] (18)

Finally, we convert Y, UB, VB to the RGB color space to generate fusion RGB image CO as follows:

\[
\begin{align*}
R_O &= Y + 1.14VB \\
G_O &= Y - 0.39UB - 0.58VB \\
B_O &= Y + 2.03UB
\end{align*}
\] (19)

III. EXPERIMENTAL RESULTS
For experiments, we use synthetic image pairs in Figs. 8-10 and real image pairs in Figs. 11-13. The synthetic image pairs are indoor scenes, while the real image pairs are outdoor scenes. We synthesize low light RGB images in the synthetic image pairs (Figs. 8-10) by adding Gaussian noise and salt-and-pepper noise into the clean RGB images in a publicly available dataset [2]. Moreover, we capture real image pairs in the real image pairs (Figs. 11-13) using JAI AD-080GE camera at night time. This camera can capture RGB and NIR images simultaneously using the same optical path with two CCDs. In the real image pairs, the RGB images are heavily corrupted by noise with severe loss of details. Thus, they are much more degraded and challenging than the synthetic ones for the fusion. We perform our experiments on a PC with Intel i5-6500 CPU (3.2GHz) and 16GB RAM using Matlab 2015b and C++.

A. PARAMETER SETTING
In the proposed method, we use WLS four times for alternating guidance, i.e. images denoising and detail transfer. Moreover, we utilize fast weight median filtering (FWMF) [32] to remove salt-and-pepper noise. We set the number of WLS iterations to T = 3 for the balance of efficiency and quality. For each WLS, we set \(\lambda_1 = 15^2\), \(\lambda_2 = 30^2\), \(\lambda_3 = 40^2\) and \(\lambda_4 = 10^2\), \(\sigma_1 = 0.15\), \(\sigma_2 = 0.1\), \(\sigma_3 = 0.03\), \(\sigma_4 = 0.01\). For FWMF, we set the window radius to 5 and sigma to 150 to control the weight between two pixels.

B. PERFORMANCE EVALUATION
To verify effectiveness and efficiency of the proposed method, we compare the proposed method with...
state-of-the-art fusion methods: He et al.’s method [6], Petschnigg et al.’s method [19], Zhuo et al.’s method [39] and Yan et al.’s method [30]. As shown in Figs. 8-13, the RGB images captured in low light condition contain severe noise, while their NIR images captured in the same condition have clear texture and less noise. As shown in the figures, detail loss of low light RGB images is very serious due to much noise. As shown in Fig. 10, the proposed method achieves good denoising performance, while successfully retaining textures in the fusion result. Moreover, there is little color shift in our fusion result. In the indoor scenes (Figs. 8-10), the proposed method produces very similar fusion results to Yan et al.’s work, and achieves slightly better denoising performance than it. In Fig. 9, Petschnigg et al.’s and Zhuo et al.’s methods tend to produce over-enhancement amplifying noise. In the outdoor scenes (Figs. 11-13), the proposed method apparently outperforms the others in both noise removal and detail transfer. That is, the proposed method achieves outstanding fusion performance especially for the outdoor scenes. At the same time, our method successfully...
preserves color information of RGB images by preventing color shift in fusion.

For quantitative measurements, we evaluate performance on the fusion results in terms of blind image quality assessment (BIQA) [31]. As an evaluation metric, we adopt a no-reference metric for quality assessment, BIQA, because neither RGB nor NIR images are used as reference images. Table 1 shows BIQA scores for different methods on the test image pairs. Smaller scores represent better performance. Bold and underlined numbers indicate the best and second performance, respectively. The proposed method achieves the minimum BIQA scores and outperforms the others in average performance. Furthermore, the proposed method achieves higher computational efficiency than the others. The time complexity of our method reaches $O(N)$ by fast global smoothing of WLS [16]. We estimate their runtime on our
testing dataset (total 29 pairs of RGB and NIR images) with resolution of 1920 × 1080. Table 2 lists the average runtime (unit: sec/pair). Among them, Yan et al.’s work [30] achieves comparable performance in fusion to ours, however, the proposed method is 30+ times faster. This is because we use fast global smoothing for fusion of RGB and NIR images based on WLS.

**IV. CONCLUSION**

We have proposed scale-aware multispectral fusion of RGB and NIR images based on alternating guidance. Low light RGB images contain colors with coarse image structure, while NIR images include clean textures without color. We have adopted scale-aware multispectral fusion to contain multiscale structures of RGB and NIR images in the fusion results. Moreover, we have used alternating guidance based on WLS to maximize the multispectral advantage. We have utilized joint guidance of NIR image and denoised RGB luminance in the first guidance to remove noise while keeping edges. We have used an over-smoothed RGB image as guidance in the second guidance to perform scale-aware detail transfer of NIR image to fusion. Experimental results show that the proposed fusion method achieves good performance in noise removal and detail transfer with high computational efficiency.

**ACKNOWLEDGMENT**

The authors would like to thank Dr. C. Su and Dr. H. Liu from Huawei Technologies for their helpful advice and discussion.

**REFERENCES**

[1] D. P. Bavirisetti, G. Xiao, and G. Liu, “Multi-sensor image fusion based on fourth order partial differential equations,” in *Proc. 20th Int. Conf. Inf. Fusion (Fusion)*, Jul. 2017, pp. 1–9.

[2] C. Fredembach and S. Süssstrunk, “Colouring the near-infrared,” in *Proc. Color Imag. Conf.* Springfield, VA, USA: Society Imaging Science Technology, 2008, pp. 176–182.

[3] E. S. L. Gastal and M. M. Oliveira, “Domain transform for edge-aware image and video processing,” *ACM Trans. Graph.*, vol. 30, no. 4, p. 69, 2011.

[4] L. Grady, “Random walks for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[5] J. Han and B. Bhanu, “Fusion of color and infrared video for moving human detection,” Pattern Recognit., vol. 40, no. 6, pp. 1771–1784, Jun. 2007.

[6] K. He, J. Sun, and X. Tang, “Guided image filtering,” in Proc. Eur. Conf. Comput. Vis., Berlin, Germany: Springer, 2010, pp. 1–14.

[7] W. Kong, Y. Lei, and H. Zhao, “Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization,” Inf. Phys. Technol., vol. 67, pp. 161–172, Nov. 2014.

[8] W. Kong, L. Zhang, and Y. Lei, “Novel fusion method for visible light and infrared images based on NSST–SF–PCNN,” Inf. Phys. Technol., vol. 65, pp. 103–112, Jul. 2014.

[9] P. Kumar, A. Mittal, and P. Kumar, “Fusion of thermal infrared and visible spectrum video for robust surveillance,” in Computer Vision, Graphics and Image Processing. Berlin, Germany: Springer, 2006, pp. 528–539.

[10] S. Li, B. Yang, and J. Hu, “Performance comparison of different multi-resolution transforms for image fusion,” Inf. Fusion, vol. 12, no. 2, pp. 74–84, Apr. 2011.

[11] S. Li, H. Yin, and L. Fang, “Group-sparse representation with dictionary learning for medical image denoising and fusion,” IEEE Trans. Med. Imaging, vol. 39, no. 12, pp. 3450–3459, Dec. 2012.

[12] Y. Liu, S. Liu, and Z. Wang, “A general framework for image fusion based on multi-scale transform and sparse representation,” Inf. Fusion, vol. 24, pp. 147–164, Jul. 2015.

[13] Y. M. Lu, C. Fredembach, M. Vetterli, and S. Susstrunk, “Fast global image smoothing based on weighted least square,” IEEE Trans. Image Process., vol. 23, no. 12, pp. 5638–5653, Dec. 2014.

[14] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, “Statistics of image patches,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–7.

[15] G. Pujares and J. M. de la Cruz, “A wavelet-based image fusion tutorial,” Pattern Recognit., vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[16] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, “Digital photography with flash and no-flash image pairs,” ACM Trans. Graph., vol. 23, no. 3, pp. 664–672, Aug. 2004.

[17] G. Pueella, “A general framework for multisresolution image fusion: From pixels to regions,” Inf. Fusion, vol. 4, no. 4, pp. 259–280, Dec. 2003.

[18] Z. Sadeghipoor, Y. M. Lu, and S. Susstrunk, “Correlation-based joint acquisition and demosaicing of visible and near-infrared images,” in Proc. 16th IEEE Int. Conf. Image Process. (ICIP), Nov. 2009, pp. 3797–3800.

[19] J. Ma, C. Chen, C. Li, and J. Huang, “Infrared and visible image fusion via gradient transfer and total variation minimization,” Inf. Fusion, vol. 31, pp. 100–109, Sep. 2016.

[20] J. Ma, Z. Zhou, B. Wang, and H. Zong, “Infrared and visible image fusion based on visual saliency map and weighted least square optimization,” Inf. Phys. Technol., vol. 82, pp. 8–17, May 2017.

[21] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, “Fast global image smoothing based on weighted least squares,” IEEE Trans. Image Process., vol. 23, no. 12, pp. 5638–5653, Dec. 2014.

[22] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, “Statistics of image patches,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–7.

[23] G. Pujares and J. M. de la Cruz, “A wavelet-based image fusion tutorial,” Pattern Recognit., vol. 37, no. 9, pp. 1855–1872, Sep. 2004.

[24] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama, “Digital photography with flash and no-flash image pairs,” ACM Trans. Graph., vol. 23, no. 3, pp. 664–672, Aug. 2004.

[25] G. Pueella, “A general framework for multisresolution image fusion: From pixels to regions,” Inf. Fusion, vol. 4, no. 4, pp. 259–280, Dec. 2003.

[26] Z. Sadeghipoor, Y. M. Lu, and S. Susstrunk, “Correlation-based joint acquisition and demosaicing of visible and near-infrared images,” in Proc. 16th IEEE Int. Conf. Image Process. (ICIP), Nov. 2009, pp. 3797–3800.

[27] J. Ma, C. Chen, C. Li, and J. Huang, “Infrared and visible image fusion via gradient transfer and total variation minimization,” Inf. Fusion, vol. 31, pp. 100–109, Sep. 2016.

[28] J. Ma, Z. Zhou, B. Wang, and H. Zong, “Infrared and visible image fusion based on visual saliency map and weighted least square optimization,” Inf. Phys. Technol., vol. 82, pp. 8–17, May 2017.

[29] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, “Fast global image smoothing based on weighted least squares,” IEEE Trans. Image Process., vol. 23, no. 12, pp. 5638–5653, Dec. 2014.

[30] J. Ma, Z. Zhou, B. Wang, and H. Zong, “Infrared and visible image fusion based on visual saliency map and weighted least square optimization,” Inf. Phys. Technol., vol. 82, pp. 8–17, May 2017.

[31] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do, “Fast global image smoothing based on weighted least squares,” IEEE Trans. Image Process., vol. 23, no. 12, pp. 5638–5653, Dec. 2014.

[32] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister, “Statistics of image patches,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2007, pp. 1–7.