IEEE Open Journal of Signal Processing

Received 18 June 2021; revised 4 October 2021; accepted 14 October 2021. Date of publication 10 October 2021; date of current version 10 November 2021. The review of this article was arranged by Associate Editor H. Vicky Zhao.

Digital Object Identifier 10.1109/OJSP.2021.3122074

Multispectral Fusion of RGB and NIR Images Using Weighted Least Squares and Convolution Neural Networks

CHEOLKON JUNG (Member, IEEE), QIHUI HAN, KAILONG ZHOU, AND YUANQUAN XU

School of Electronic Engineering, Xidian University, Xian 710071, China

CORRESPONDING AUTHOR: CHEOLKON JUNG (e-mail: ckjung@skku.edu).

This work was supported by the National Natural Science Foundation of China under Grant 61872280.

ABSTRACT In low light condition, color (RGB) images captured by visible sensors suffer from severe noise causing loss of colors and textures. However, near infrared (NIR) images captured by NIR sensors are robust to noise even in low light condition without color. Since RGB and NIR images are complementary in low light condition, the multispectral fusion of RGB and NIR images provides a viable solution to low light imaging. In this paper, we propose multispectral fusion of RGB and NIR images using weighted least squares (WLS) and convolution neural networks (CNNs). We combine traditional WLS filtering for layer decomposition and denoising with latest deep learning for image enhancement and texture transfer into the multispectral fusion to take both advantages. We build two networks based on CNN: image enhancement network (IEN) for image enhancement and texture transfer network (TTN) for NIR texture transfer. First, we perform RGB image denoising based on WLS filtering and generate the base layer. We use both RGB and NIR images for WLS filtering as weights to filter out noise in low light RGB images. Second, we conduct IEN to enhance contrast of the base layer. Third, we perform TTN to deliver NIR details completely and naturally to the fusion result. The combination of WLS, TTN and IEN leads to noise reduction, contrast enhancement, and detail preservation in fusion. Experimental results show that the proposed method achieves good performance in both noise reduction and detail transfer as well as outperforms state-of-the-art methods in terms of visual quality and quantitative measurements.

INDEX TERMS Image fusion, convolution neural networks, multispectral, near-infrared, sensor fusion, weighted least squares.

I. INTRODUCTION

The quality of RGB images depends on the shooting environment. Under the proper shooting environment, the captured RGB image is of excellent quality, which is suitable for human visual perception. However, when the shooting environment of RGB images is poor like low light condition, the quality of RGB images is degraded by noise or other artifacts. To improve the imaging quality of RGB images in low light condition, many studies have been done. One solution is to enhance the color and luminance of low light images through data-driven approach [5], [24] and illumination adjustment approach [1], [9], [41]. As image acquisition of different types of sensors becomes more and more convenient, improving the quality of RGB images through multi-sensor image fusion has received increasing attention. Fusion of RGB and NIR images is able to improve imaging quality in low light condition [38], [40], [47]. In low light condition, RGB images are degraded by much noise causing loss of detail and color by increasing camera ISO. Fortunately, NIR images captured at the same scene have high resolution and clear texture, which are robust to noise. However, compared with RGB images, the most significant disadvantage of NIR images is no color information. Thus, as RGB and NIR images are more and more easily accessible, their fusion becomes a possible solution to the low light imaging. In the fusion of RGB and NIR images, the most prominent problem is the inconsistency of the

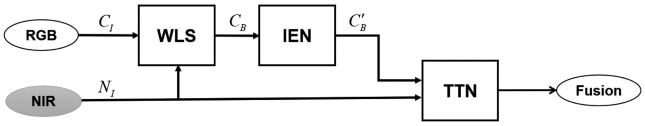


FIGURE 1. Entire framework of the proposed fusion method for RGB and NIR images. WLS: Weighted least squares. IEN: Image enhancement network. TTN: Texture transfer network.

luminance channels of RGB images and the contrast of NIR images. Directly using NIR images as luminance channel results in color distortion and structural loss.

In this paper, we propose multispectral fusion of RGB and NIR images using weighted least squares (WLS) and convolution neural networks (CNNs). As shown in Fig. 1, we divide the fusion task of RGB and NIR images into three modules: RGB image denoising, image enhancement and NIR detail transfer. For RGB image denoising, we adopt WLS to remove the noise in RGB images through joint guidance and generate the base layer. For image enhancement, we build an image enhancement network (IEN) based on CNN to enhance contrast of the base layer. For NIR detail transfer, we design a texture transfer network (TTN) based on CNN to transfer NIR details to the fusion result. Based on the three modules, the proposed method combines multispectral advantages of RGB and NIR images, and generates a fusion image with noise reduction, contrast enhancement, and detail preservation. We perform IEN first and then TTN to achieve natural looking fusion results. Fig. 2 shows the fusion results of RGB and NIR images by different methods. We capture the pair of RGB and NIR images at night 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, the proposed method successfully removes noise and produces details in fusion, especially in the background regions (see the trees and road). However, Yan *et al.*'s method [47] produces an oversmoothed fusion image in the background regions, thus not natural-looking. DIF-Net [13] and DenseFuse [17] change the color tone of the person's clothe affected by the NIR image.

Compared with existing methods, the main contributions of this paper are as follows:

- We combine traditional WLS filtering with latest deep learning to take both advantages in fusion. We use WLS filtering for layer decomposition and denoising, while we utilize deep learning for image enhancement and texture transfer.
- We build IEN for image enhancement and TTN for NIR texture transfer based on CNNs. IEN enhances the base layer obtained by WLS, thus it does not amplify noise nor change color tone after contrast enhancement. TTN effectively transfers NIR details to the fusion thanks to the feature extraction and fusion by CNN.
- The combination of WLS, TTN, and IEN leads to noise reduction, contrast enhancement, and detail preservation in fusion. The proposed method is beyond simple combination of existing techniques, and provides a viable solution to low light imaging using multiple sensors.

II. RELATED WORK

Up to the present, many outstanding studies have been done for the fusion of different types of images. Similar to RGB/NIR fusion, there is flash/no-flash image fusion among them. In low light condition, no-flash images contains a lot of noise and detail loss, while flash images include details with little noise. Thus, flash images provide details for no-flash image similar to NIR images. However, flash images cause some unnecessary shadows and contain specular highlights. Eisemann *et al.* [4] proposed a flash/no-flash fusion method that decomposed flash and no-flash images into large scale image structures and small scale details. Petschnigg *et al.* [34] proposed a flash/no-flash fusion method based on bilateral filtering. As a general fusion method, He *et al.* [10] proposed guided image filtering (GF) based on a local linear model. Another type of image fusion is fusion of infrared and RGB images. Infrared images distinguish the target from their backgrounds based on radiation difference, and perform well for imaging in bad weather and night conditions. NIR and infrared images are very similar since they have strong anti-interference performance against difficult surroundings. However, infrared images typically have low resolution and poor textures. The previous work of RGB/infrared image fusion are mainly divided into seven categories according to their adopted theories, i.e. multi-scale transform [20], [33], [53], sparse representation [21], [42], neural network [16], [44], subspace methods [2], [15], and saliency-based methods [26], [52], [54], hybrid models [23], [30], and other methods [25], [55]. Compared with flash images, NIR images provide better details with high resolution in night vision. Thus, the fusion of RGB and NIR images is able to produce high quality fusion results especially in low light condition. However, luminance channel of RGB images has contrast and structural differences from NIR images. The direct use of NIR images as the luminance channel will cause color distortion and structural loss. To solve the contrast difference, Son *et al.* [38] proposed low light color image denoising based on contrast conversion between NIR images and luminance channels. Son *et al.* [40] 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.* [37] proposed a fusion method based on high visibility area selection. Son *et al.* [39] proposed a layer-based approach for image pair fusion. Zhuo *et al.* [58] constructed a framework for the fusion of RGB and NIR. Yan *et al.* [47] explicitly modelled derivative-level confidence and proposed cross field joint image fusion by optimizing a scale map.

So far, deep learning-based RGB and NIR image fusion methods are relatively few. As a new research direction, deep learning-based RGB/NIR image fusion is of significant importance. Traditional fusion methods often fail to produce satisfactory natural-looking fusion results due to the fact that

FIGURE 2. RGB-NIR image fusion results. Left to right: Input RGB image, input NIR image, Yan *et al.*'s method [47], DIF-Net [13], DenseFuse [17], and the proposed method (WLS+TTN). We capture them at night by JAI AD-130 GE camera which is able to simultaneously capture RGB and NIR images through the same optical path with two CCDs.

the extraction and fusion of details are not accurate and natural enough. Considering the powerful feature extraction and representation capabilities of convolutional neural networks (CNN), deep learning-based RGB/NIR fusion is an effective way of yielding more natural and higher quality fusion images. However, the biggest obstacle to deep learning-based RGB/NIR image fusion is the acquisition of training data. Unlike the restoration of degraded images such as image super-resolution and image enhancement, the goal of image fusion is not to recover the lost information but to fuse different types of image information to obtain a natural-looking fusion image. Therefore, compared to the image restoration task, the ground truth of the image fusion task is almost impossible to obtain. For RGB/NIR image pairs in low light condition, one way to obtain the ground truth is to capture daytime RGB images in the same scene. Such approach is costly and difficult to implement in practice. Jung *et al.* [12] proposed multispectral fusion of RGB and NIR images using two stage convolutional neural networks (CNNs), called FusionNet. They synthesized noisy RGB images for training data by adding noise in clean RGB images, and use the clean RGB images as ground truth. Jung *et al.* [13] proposed an unsupervised deep learning framework for image fusion with structure tensor representations, called DIF-Net. They provided an unsupervised loss function using the structure tensor representation of the multi-channel image contrasts. The loss function was minimized by a stochastic deep learning solver, thus directly producing a fusion image without iterations. In the deep learning-based image fusion, the most similar task is the fusion of visible and infrared (IR) images. Ren *et al.* [36] proposed an IR and visible image fusion method based on convolutional neural networks (CNN). Li *et al.* [19] proposed a deep learning framework for the fusion of infrared and visible image. They used VGG19 to extract features of IR and visible images. Li *et al.* [17] further proposed an encoder-decoder network structure for the fusion of IR and visible images, called DenseFuse. Li *et al.* [18] then proposed a visible and IR fusion method based on ResNet and zero-phase component analysis. Liu *et al.* [22] proposed a fusion method based on CNN and saliency detection. As a very similar study, Prabhakar *et al.* [35] proposed a deep unsupervised approach for exposure fusion with extreme exposure image pairs, called DeepFuse. Ma *et al.* [28] adopted a generative adversarial network (GAN) for the fusion of visible and IR images, called FusionGAN. They further achieved fusion methods of visible and IR images such as DcGAN, GANMcC, STDFusionNet, U2Fusion, and SDNet [26], [27], [29], [45], [48], [49]. In our previous work [56], [57], we built a fusion framework of RGB and NIR images based on WLS. This framework is fully based on WLS that is one of traditional filtering methods, and provides alternating guidance for fusion based on it. In this work, we combine traditional filtering (WLS) and latest deep learning (CNN) into the fusion of RGB and NIR images to take advantage of both approaches. We use WLS to achieve good denoising performance in low light condition, while we adopt CNN to transfer NIR details and enhance RGB images, thus resulting in natural-looking fusion images.

III. PROPOSED METHOD

A. RGB IMAGE DENOISING BY WLS

To remove noise in the RGB image, we adopt WLS to denoise through the joint guidance of luminance channel of RGB image and NIR image. WLS is a globally optimized image filter with data and smoothing terms [31], which can smooth images effectively. Given an input image f and a guidance image g, an output image u is obtained by minimizing the following WLS energy function as follows:

$$\varepsilon(u) = \sum_{p} \left(u_p - f_p\right)^2 + \lambda \sum_{p} \sum_{q \in N(p)} \omega_{p,q}(g)\left(u_p - u_q\right)^2 \quad (1)$$

where $N(p)$ represents a set of four adjacent pixels of P; λ controls the balance between data and smoothing terms, and

increasing λ results in smoothing output; and $\omega_{p,q}(g)$ is the weight calculated from the guidance image f and measure the similarity between pixels p and q. $\omega_{p,q}(g)$ is defined as follows:

$$\omega_{p,q}(g) = \exp(-\left\|g_p - g_q\right\| / \sigma) \tag{2}$$

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

$$\varepsilon(\mathbf{u}) = (\mathbf{u} - \mathbf{v})^{\mathrm{T}}(\mathbf{u} - \mathbf{v}) + \lambda \mathbf{u}^{\mathrm{T}} \mathbf{A}_g \mathbf{u} \tag{3}$$

where $\mathbf{u}$ and $\mathbf{v}$ 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 $\mathbf{A}_g$ is $S \times S$ Laplacian matrix as follows [8]:

$$\mathbf{A}_g(m,n) = \begin{cases} \sum\limits_{l \in N(m)} \omega_{m,l}(g) & n = m \\ -\omega_{m,n}(g) & n \in N(m) \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

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

$$(\mathbf{I} + \lambda \mathbf{A}_g)\mathbf{u} = \mathbf{v} \tag{5}$$

However, solving it by matrix inversion is of high computational complexity. By fast global smoothing of WLS, the time complexity reaches $O(N)$. 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:

$$\sum_x ((u_x^h - f_x^h)^2 + \lambda_t \sum_{i \in N_h(x)} \omega_{x,i}(g^h)(u_x^h - u_i^h)^2) \tag{6}$$

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

$$(\mathbf{I}_h + \lambda_t \mathbf{A}_h)\,\mathbf{u}_h = \mathbf{f}_h \tag{7}$$

where $\mathbf{I}_h$ is an identity matrix with a size of $W \times W$; $\mathbf{u}_h$ and $\mathbf{f}_h$ represent the vector notations of u_h and f_h, respectively; $\mathbf{A}_h$ is a three-point Laplacian matrix with a size of $W \times W$. The linear system in Eq. (7) is written as follows:

$$\begin{bmatrix} b_0 & c_0 & 0 & \cdots & 0 \\ \ddots & \ddots & \ddots & 0 & 0 \\ 0 & a_x & b_x & c_x & 0 \\ 0 & 0 & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & a_{W-1} & b_{W-1} \end{bmatrix} \begin{bmatrix} u_0^h \\ \vdots \\ u_x^h \\ \vdots \\ u_{W-1}^h \end{bmatrix} = \begin{bmatrix} f_0^h \\ \vdots \\ f_x^h \\ \vdots \\ f_{W-1}^h \end{bmatrix} \tag{8}$$

where u_x^h and f_x^h are the x_{th} elements of $\mathbf{u}_h$ and $\mathbf{f}_h$, respectively; a_x, b_x, and c_x represent three nonzero elements in the x_{th} row of $(\mathbf{I}_h + \lambda_t \mathbf{A}_h)$. In boundary condition, $a_0 = 0$ and $c_{W-1} = 0$. a_x, b_x and c_x are written as:

$$a_x = \lambda_t \mathbf{A}_h\,(x, x-1) = -\lambda_t \omega_{x,x-1}$$

$$b_x = 1 + \lambda_t \mathbf{A}_h(x,x) = 1 + \lambda_t \left(\omega_{x,x-1} + \omega_{x,x+1}\right)$$

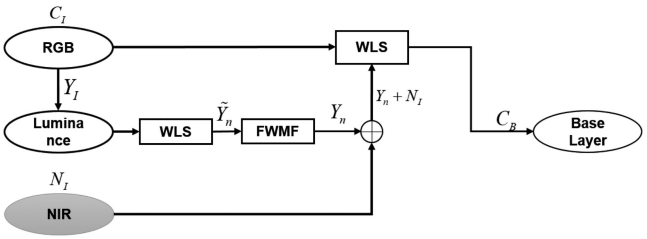


FIGURE 3. RGB image denoising module based on WLS. FWMF: Fast weighted median filtering. ⊕: Pixel-wise addition operator.

$$c_x = \lambda_t \mathbf{A}_h\,(x, x+1) = -\lambda_t \omega_{x,x+1} \tag{9}$$

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

$$\tilde{c}_x = c_x / (b_x - \tilde{c}_{x-1} a_x)$$

$$\tilde{f}_x^h = \left(f_x^h - \tilde{f}_{x-1}^h a_x\right) / (b_x - \tilde{c}_{x-1} a_x), (x = 1, \ldots, W-1) \tag{10}$$

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

$$u_x^h = \tilde{f}_x^h - \tilde{c}_x u_{x+1}^h, (x = W-2, \ldots, 0) \tag{11}$$

with $u_{W-1}^h = \tilde{f}_{W-1}^h$. 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 [7], we perform 2D smoothing by applying sequential 1D global smoothing to a multiple number of iterations [31]. In this scheme, λ_t in each iteration is computed as follows:

$$\lambda_t = \frac{3}{2} \frac{4^{T-t}}{4^T - 1} \lambda \tag{12}$$

where T represents the total number of iterations along each dimension. In each iteration, we perform 1D solver with parameter λ_t along x dimension and y dimension of 2D images continuously.

Based on WLS, a framework for denoising RGB images was designed. Fig. 3 shows the denoising module for RGB image. Firstly, we transfer the RGB image to YUV space and take out the luminance channel Y_I. Then, we utilize WLS and FWMF [51] to remove noise in luminance channel Y_I to get Y_n. Here, Y_n becomes very blurry although the noise is removed. However, using the joint guidance of Y_n and NIR image N_I to denoise the input RGB image C_B can maintain the edge information of the image while denoising. The base layer C_B is obtained by minimizing the following energy function:

$$\varepsilon\,(\mathbf{C}_B) = (\mathbf{C}_B - \mathbf{C}_I)^{\mathbf{T}}\,(\mathbf{C}_B - \mathbf{C}_I) + \lambda \mathbf{C}_B{}^T \mathbf{A}_{G_n} \mathbf{C}_B \tag{13}$$

where $\mathbf{C}_I$ represents column vectors containing values of C_I; $\mathbf{A}_{G_n}$ denotes Laplacian matrix defined by $Y_n + N_I$; and range parameter is σ.

FIGURE 4. RGB denoising comparison between several strategies. Top-left: Noisy RGB image. Top-right: Guided by C_I. Bottom-left: Guided by N_I. Bottom-right: Guided by $Y_n + N_I$.

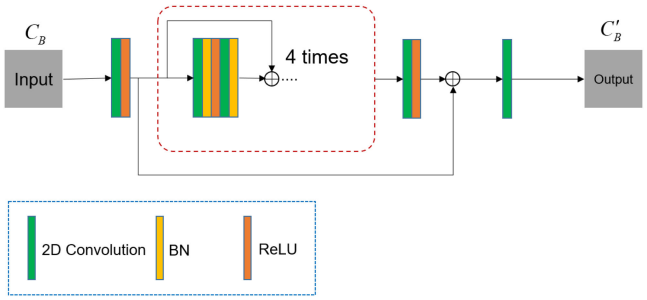


FIGURE 5. Network structure of the image enhancement network (IEN). IEN is based on CNN, which consists of four residual blocks.

FIGURE 6. Samples of training data for IEN. In this dataset, iPhone images are of low quality used as input, and Canon images are of high quality used as ground truth.

In low light condition, the RGB image contains severe noise causing destruction of image structure due to low SNR. Since some details of the RGB image are lost by WLS filtering, we adopt the joint guidance of luminance channel of RGB image and NIR image in WLS filtering to retain details in the RGB image while removing noise. By adopting this joint

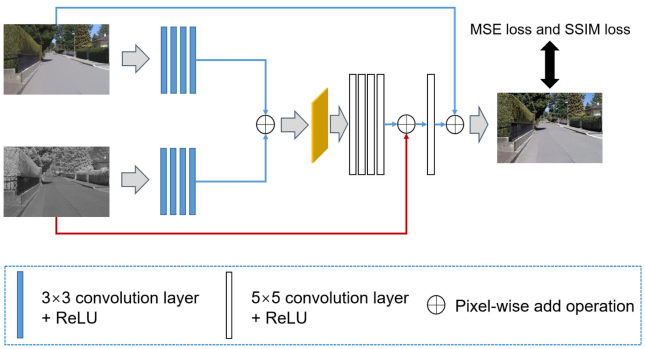


FIGURE 7. Network structure of the texture transfer network (TTN).

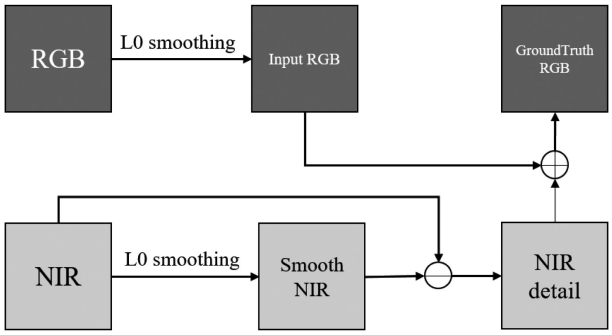


FIGURE 8. Training data generation for TTN. We use the L0 smoothing to synthesize texture lost images for training.

guidance strategy, the base layer obtained by the denoising module achieves a good denoising effect and maintains edge information. To verify the effectiveness of joint guidance, we perform a comparison of several denoising methods. As shown in Fig. 4, the guidance of C_I is not able to produce satisfactory denoising performance. When using NIR image only as guidance, the contrast difference between RGB and NIR images would cause serious blurs in edges (see the red boxes). Guided by $Y_n + N_I$, the denoising module successfully removes noise while preserving the structure of RGB image.

B. IMAGE ENHANCEMENT BY IEN

Since the contrast of RGB images taken in low light condition is low with a dark tone, we build an image enhancement network (IEN) for contrast enhancement. The network structure of IEN is shown in Fig. 5. IEN is based on CNN, which consists of four residual blocks. Each residual block has two convolutional layers, two BN layers, and one ReLU layer. This network is trained from the pairs of low quality images by smartphone camera and high quality images by digital single-lens reflex (DSLR) camera. We use MSE loss, SSIM loss and perceptual loss to form the total loss function. For IEN, the contrast change degrades color distortion in the fusion image, and thus we constrain MSE and SSIM by adding a new term of perceptual loss. The perceptual loss is calculated by the pre-trained VGG19. Since our task pays more attention to the color information of the image, we adopt

FIGURE 9. Samples of training data. Left: Input RGB image. Middle: Input NIR image. Right: Ground truth.

several shallower layers in VGG19 to calculate the perceptual loss. The total loss function for IEN, L_{total1}, is as follows:

$$\mathcal{L}_{total1} = \lambda_1 \mathcal{L}_{MSE} + \lambda_2 \mathcal{L}_{SSIM} + \lambda_3 \mathcal{L}_{perceptual} \tag{14}$$

where λ_1, λ_2 and λ_3 represent the weight of each loss.

We adopt the DPED dataset [11] for network training as shown in Fig. 6. In the dataset, iPhone images are of low quality and used as input, and Canon images are of high quality and used as ground truth.

C. NIR DETAIL TRANSFER BY TTN

Since NIR image has good detail information, it is required to extract the details of NIR image and transfer them to the fusion image. To achieve this, we build a texture transfer network (TTN) that transfers the details of the NIR image to the fusion image. The network structure of TTN is shown in Fig. 7. As shown in the figure, TTN consists of two pathways (RGB and NIR) and their fusion module. The four layers in two pathways have a series of filters 32, 64, 64 and 1 with kernel size 3×3. The four layers in the fusion module has a series of filters 32, 32, 64 and 1 with kernel size 5×5, while the last layer in the fusion module is filter 1 with kernel size 5×5. First, the features of the RGB (IEN result) and NIR images are extracted through two pathways separately. We perform ResNet model as the network backbone and extract features from RGB and NIR images to estimate residual. Then, the information extracted by the two pathways is added at the pixel level in the fusion module. Finally, the information is fused through a convolution channel in the fusion module. The input RGB image is connected with the end of the network to get the fusion image. The loss functions used in TTN are MSE loss and SSIM loss [43], which are defined as follows:

$$\mathcal{L}_{SSIM} = \frac{1}{N} \sum_{i=1}^{N} (1 - SSIM_i) \tag{15}$$

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{x=1}^{X} \sum_{y=1}^{Y} \left\| I_g(x, y) - I_o(x, y) \right\|_2^2 \tag{16}$$

where N is the number of images, x and y represent pixel coordinates, I_o is the output image of TTN, and I_g is ground truth. The total loss function for TTN, $\mathcal{L}_{total2}$, is defined as follows:

$$\mathcal{L}_{total2} = \lambda_4 \mathcal{L}_{MSE}(I_g, I_o) + \lambda_5 \mathcal{L}_{SSIM}(N_I, I_o) + \lambda_6 \mathcal{L}_{SSIM}(I_g, I_o) \tag{17}$$

where λ_4, λ_5 and λ_6 represent the weight of each loss.

Since there is no suitable dataset for the fusion task of RGB and NIR images, we synthesize training data based on the dataset [3]. Fig. 8 illustrates the training data generation for TTN. We adopt the L_0 smoothing in the training data generation [46] to synthesize texture lost images by smoothing. Samples of the synthesized training data are shown in Fig. 9. The input RGB image is synthesized from the ground truth by the L_0 smoothing.

IV. EXPERIMENTAL RESULTS

To verify the effectiveness of the proposed method, we conduct visual and quantitative comparison of the experimental results on synthetic image pairs (see Figs. 12–13) and real image pairs (see Figs. 14–16). The synthetic image pairs are indoor scenes, which are obtained by adding Gaussian noise and salt-and-pepper noise into the clean RGB images in a publicly available dataset [6]. The real image pairs are outdoor scenes, which are captured by JAI AD-080GE camera at night. This camera is able to capture RGB and NIR images simultaneously using the same optical path with two CCDs. We compare the proposed method with some state-of-the-art ones: 1) Traditional methods: Yan *et al.* [47]; 2) Deep learning methods: DIF-Net [13], DenseFuse [17], GANMcC [29], and STDFusionNet [26].

A. PARAMETER SETTING

For training and testing, we use a PC with Intel Core i7-7700 CPU and GTX1080 GPU running Ubuntu 16.04, Pytorch and C++. In RGB denoising module, excessive smoothing may result in loss of details. To balance denoising and detail preservation, we set the parameters of WLS $\lambda = 10^2$ and $\sigma = 0.1$. For FWMF, we set the window radius to 5 and sigma to 150 to control the weight between two pixels. For experiments, we train the proposed IEN and TTN using Adam optimizer [14]. The number of epochs is 50, and the batch size is 64. During the training process, we set learning rate to 0.0001, and batch size to 64 for hyperparameters.

FIGURE 10. Stepwise results by the proposed method. Left to right: Input RGB image, input NIR image, WLS, IEN, and TTN.

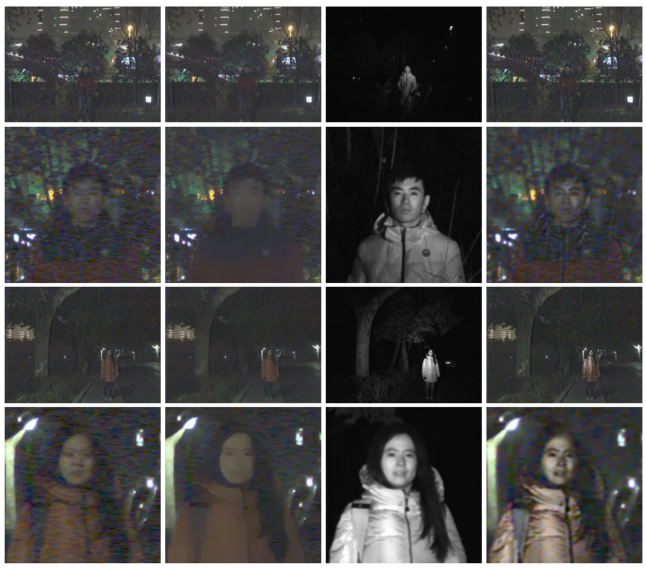


FIGURE 11. Fusion results on real images by the proposed method. Left to right: Input RGB images, WLS results, input NIR images, fusion results by the proposed method (WLS+IEN+TTN). The second and fourth rows are the zoomed images of the first and third rows, respectively.

B. PERFORMANCE EVALUATION AND ABLATION STUDY

Fig. 10 shows stepwise results (WLS, IEN, and TTN) by the proposed method. In low light condition, the RGB images contain much noise even destroying image textures. WLS successfully removes noise from the input RGB images by the joint guidance of RGB luminance channel and NIR image, and generates the base layer. IEN enhances contrast of the base layer, while TTN extracts the details of NIR image and fuses them with the base layer. We provide more fusion results on real images by the proposed method in Fig. 11. As shown in the figures, the proposed method generates natural-looking fusion results with noise reduction and detail preservation. Since IEN enhances the base layers obtained by WLS, the fusion results are enhanced and thus can not be directly compared with the other fusion methods. Thus, we provide the fusion results by WLS+TTN in Figs. 12–16 to see the detail transfer effect by TTN. Then, we further provide the fusion results by WLS+IEN+TTN in Figs. 17 and 18 to see the image enhancement effect by IEN. As shown in Figs. 12–16, compared with the fusion results by other methods, our fusion results by WLS+TTN achieve outstanding denoising performance while retaining good details in both foreground objects and background. From the viewpoint of color, the proposed method minimizes color distortion while transferring NIR details to fusion. That is, our method preserves the original color tone of the input RGB image after fusion. Moreover, our fusion results in Figs. 13–16 do not produce edge blurring artifacts in areas where the NIR and RGB luminance channels are inconsistent (see the red boxes in Fig. 13 and the background in Figs. 14–16). The proposed method produces

FIGURE 12. Visual comparison in *bowls*. Top: Input RGB image, input NIR image, STDFusionNet [26], and GANMcC [29]. Bottom: Yan *et al.* [47], DIF-Net [13], DenseFuse [17], the proposed method (WLS+TTN). Since IEN enhances the base layer obtained by WLS and leads to enhancement of the fusion result, we separately provide the fusion result of the proposed method (WLS+IEN+TTN) in Fig. 17.

FIGURE 13. Visual comparison in *teapot*. Top: Input RGB image, input NIR image, STDFusionNet [26], and GANMcC [29]. Bottom: Yan *et al.* [47], DIF-Net [13], DenseFuse [17], the proposed method (WLS+TTN). Since IEN enhances the base layer obtained by WLS and leads to enhancement of the fusion result, we separately provide the fusion result of the proposed method (WLS+IEN+TTN) in Fig. 17.

FIGURE 14. Visual comparison in a real image pair. Top: Input RGB image, input NIR image, STDFusionNet [26], and GANMcC [29]. Bottom: Yan *et al.* [47], DIF-Net [13], DenseFuse [17], the proposed method (WLS+TTN). Since IEN enhances the base layer obtained by WLS and leads to enhancement of the fusion result, we separately provide the fusion result of the proposed method (WLS+IEN+TTN) in Fig. 18.

FIGURE 15. Visual comparison in a real image pair. Top: Input RGB image, input NIR image, STDFusionNet [26], and GANMcC [29]. Bottom: Yan *et al.* [47], DIF-Net [13], DenseFuse [17], the proposed method (WLS+TTN). Since IEN enhances the base layer obtained by WLS and leads to enhancement of the fusion result, we separately provide the fusion result of the proposed method (WLS+IEN+TTN) in Fig. 18.

FIGURE 16. Visual comparison in a real image pair. Top: Input RGB image, input NIR image, STDFusionNet [26], and GANMcC [29]. Bottom: Yan *et al.* [47], DIF-Net [13], DenseFuse [17], the proposed method (WLS+TTN). Since IEN enhances the base layer obtained by WLS and leads to enhancement of the fusion result, we separately provide the fusion result of the proposed method (WLS+IEN+TTN) in Fig. 18.

natural-looking fusion results with little noise and fine details. This is because the proposed method takes advantage of both filtering (WLS) and deep learning (CNN) into fusion. Figs. 14–16 show the fusion results in low light condition, which indicates that the proposed fusion method provides a solution to low light imaging. We compare the fusion results without and with IEN in Figs. 17 and 18, i.e. WLS+TTN and WLS+IEN+TTN. It can be observed that WLS+IEN+TTN recovers more details in the fusion results than WLS+TTN while retaining the color tone of the original images. IEN only enhances the base layer obtained by WLS, thus it does not amplify noise nor change color tone after contrast enhancement. The ablation experiments indicate that IEN enhances the contrast and color of the fusion images without tone change.

C. QUANTITATIVE MEASUREMENTS

For quantitative measurements, we choose blind image quality assessment (BIQA) [50] and natural image quality evaluator (NIQE) [32] as evaluation metrics, which are no-reference metric for image quality assessment. NIQE measures naturalness without subjective tests based on a multivariate Gaussian model. The reason for choosing BIQA and NIQE as the evaluation metric is that neither RGB nor NIR images can be used as reference images for the fusion task. Tables 1

TABLE 1. BIQA Comparison Between Different Methods

Method	*teapot*	*bowls*	*doll*	*books*	Fig. 2	Fig. 14	Fig. 15	Fig. 16	Average
Yan et al. [47]	31.72	20.94	25.75	28.63	37.49	39.47	27.82	24.24	29.51
DIF-Net [13]	34.40	30.23	29.79	30.27	31.49	33.27	44.18	33.64	33.41
DenseFuse [17]	31.90	27.32	30.24	27.52	32.78	28.44	42.79	32.30	31.66
GANMcC [29]	46.52	20.23	28.27	34.66	45.33	40.89	32.17	28.64	34.59
STDFusionNet [26]	35.74	22.82	**21.40**	26.78	38.94	33.08	29.08	21.31	28.64
Proposed (WLS+TTN)	**27.44**	**19.82**	22.50	25.34	36.66	29.74	**24.94**	22.93	26.17
Proposed (WLS+IEN+TTN)	28.82	19.92	23.68	**22.19**	**21.45**	**26.13**	29.72	**19.63**	**23.94**

Bold symbol represents the best performance, while underline symbol represents the second performance.

TABLE 2. NIQE Comparison Between Different Methods

Method	*teapot*	*bowls*	*doll*	*books*	Fig. 2	Fig. 14	Fig. 15	Fig. 16	Average
Yan et al. [47]	2.757	2.946	**2.633**	**3.001**	5.610	5.009	5.408	4.763	4.016
DIF-Net [13]	**2.642**	3.086	4.260	4.452	4.042	4.373	4.355	4.684	3.987
DenseFuse [17]	3.692	3.584	4.651	4.487	5.763	6.023	5.755	6.043	5.000
GANMcC [29]	4.430	3.526	4.913	4.178	4.367	5.404	4.833	5.139	4.599
STDFusionNet [26]	4.514	5.828	7.678	6.018	6.184	6.276	6.012	5.960	6.059
Proposed (WLS+TTN)	3.058	2.944	3.896	3.194	**3.652**	**3.657**	**3.845**	**3.741**	**3.498**
Proposed (WLS+IEN+TTN)	2.847	**2.732**	4.110	3.133	3.734	3.999	4.376	4.204	3.642

Bold symbol represents the best performance, while underline symbol represents the second performance.

FIGURE 17. Fusion results without and with IEN. Left: WLS+TTN. Right: WLS+IEN+TTN. Top to bottom: *teapot*, *bowls*, *doll*, *books*. IEN enhances the contrast and color of the fusion images without tone change.

and 2 show BIQA and NIQE scores of different methods, respectively. Smaller scores represent better performance. Bold and underlined numbers indicate the best and second performance, respectively. In most scenes, the proposed method

FIGURE 18. Fusion results without and with IEN. Left: WLS+TTN. Right: WLS+IEN+TTN. Top to bottom: Fig. 2, Fig. 14, Fig. 15, Fig. 16. IEN enhances the contrast and color of the fusion images without tone change.

(WLS+TTN) achieves the minimum BIQA and NIQE scores, and outperforms the others in average performance.

V. CONCLUSION

In this paper, we have proposed multispectral fusion of RGB and NIR images using WLS and CNN. We have combined traditional filtering (WLS) and latest deep learning (CNN) into the fusion of RGB and NIR images to take advantage of both approaches. The denoising module based on WLS achieves good denoising performance by the joint guidance strategy. IEN effectively enhances the contrast of the base layer, thus improving the color information of the image. TTN transfers the details of NIR image to the fusion image with the help of feature extraction and fusion by CNN. Experimental results demonstrate the superiority of the proposed method for RGB/NIR image fusion in noise removal and detail transfer over the state-of-the-art methods.

REFERENCES

[1] P. P. Banik, R. Saha, and K.-D. Kim, "Contrast enhancement of low-light image using histogram equalization and illumination adjustment," in *Proc. Int. Conf. Electron., Inf., Commun.*, 2018, pp. 1–4.

[2] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proc. Int. Conf. Inf. Fusion*, 2017, pp. 1–9.

[3] M. Brown and S. Süsstrunk, "Multi-spectral sift for scene category recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 177–184.

[4] E. Eisemann and F. Durand, "Flash photography enhancement via intrinsic relighting," *ACM Trans. Graph.*, vol. 23, pp. 673–678, 2004.

[5] K. Fotiadou, G. Tsagkatakis, and P. Tsakalides, "Low light image enhancement via sparse representations," in *Proc. Int. Conf. Image Anal. Recognit*, 2014, pp. 84–93.

[6] C. Fredembach and S. Süsstrunk, "Colouring the near-infrared," in *Proc. Color Imag. Conf.*, 2008, pp. 176–182.

[7] E. S. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–12, 2011.

[8] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.

[9] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[10] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 1–14.

[11] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis.*, 2017, pp. 3277–3285.

[12] C. Jung, K. Zhou, and J. Feng, "FusionNet: Multispectral fusion of RGB and NIR images using two stage convolutional neural networks," *IEEE Access*, vol. 8, pp. 23912–23919, 2020.

[13] H. Jung, Y. Kim, H. Jang, N. Ha, and K. Sohn, "Unsupervised deep image fusion with structure tensor representations," *IEEE Trans. Image Process.*, vol. 29, pp. 3845–3858, Jan. 2020.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.

[15] 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," *Infrared Phys. Technol.*, vol. 67, pp. 161–172, 2014.

[16] W. Kong, L. Zhang, and Y. Lei, "Novel fusion method for visible light and infrared images based on NSST-SF-PCNN," *Infrared Phys. Technol.*, vol. 65, pp. 103–112, 2014.

[17] H. Li and X.-J. Wu, "Densefuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, May 2019.

[18] H. Li, X.-J. Wu, and T. S. Durrani, "Infrared and visible image fusion with resnet and zero-phase component analysis," *Infrared Phys. Technol.*, vol. 102, 2019, Art. no. 103039.

[19] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proc. Int. Conf. Pattern Recognit.*, 2018, pp. 2705–2710.

[20] 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, 2011.

[21] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 12, pp. 3450–3459, Dec. 2012.

[22] D. Liu, D. Zhou, R. Nie, and R. Hou, "Infrared and visible image fusion based on convolutional neural network model and saliency detection via hybrid l0-l1 layer decomposition," *J. Electron. Imag.*, vol. 27 no. 6, 2018, Art. no. 063036.

[23] 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, 2015.

[24] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, 2017.

[25] 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, 2016.

[26] J. Ma, L. Tang, M. Xu, H. Zhang, and G. Xiao, "STDFusionnet: An infrared and visible image fusion network based on salient target detection," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, Apr. 2021.

[27] J. Ma, H. Xu, J. Jiang, X. Mei, and X. P. Zhang, "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4980–4995, Mar. 2020.

[28] J. Ma, W. Yu, P. Liang, C. Li, and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11–26, 2019.

[29] J. Ma, H. Zhang, Z. Shao, P. Liang, and H. Xu, "GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–14, Dec. 2020.

[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," *Infrared Phys. Technol.*, vol. 82, pp. 8–17, 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] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[33] G. Pajares and J. M. De LaCruz, "A wavelet-based image fusion tutorial," *Pattern Recognit.*, vol. 37, no. 9, pp. 1855–1872, 2004.

[34] 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, pp. 664–672, 2004.

[35] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4724–4732.

[36] X. Ren, F. Meng, T. Hu, Z. Liu, and C. Wang, "Infrared-visible image fusion based on convolutional neural networks (CNN)," in *Proc. Int. Conf. Intell. Sci. Big Data Eng.*, 2018, pp. 301–307.

[37] T. Shibata, M. Tanaka, and M. Okutomi, "Versatile visible and near-infrared image fusion based on high visibility area selection," *J. Electron. Imag.*, vol. 25 no. 1, 2016, Art. no. 013016.

[38] C.-H. Son, "Near-infrared fusion via a series of transfers for noise removal," *Signal Process.*, vol. 143, pp. 20–27, 2018.

[39] C.-H. Son and X.-P. Zhang, "Layer-based approach for image pair fusion," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2866–2881, Jun. 2016.

[40] C.-H. Son, X.-P. Zhang, and K.-W. Lee, "Near-infrared coloring via a contrast-preserving mapping model," in *Proc. GlobalSIP*, 2015, pp. 677–681.

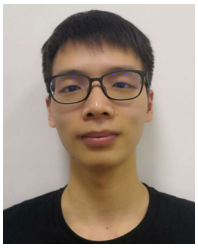
[41] Q. Song and H. Liu, "Enhancing low-light color image via l_0 regularization and reweighted group sparsity," *IEEE Access*, vol. 9, pp. 101614–101626, 2021.

[42] J. Wang, J. Peng, X. Feng, G. He, and J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infrared Phys. Technol.*, vol. 67, pp. 477–489, 2014.
[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[44] T. Xiang, L. Yan, and R. Gao, "A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in nsct domain," *Infrared Phys. Technol.*, vol. 69, pp. 53–61, 2015.
[45] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2020.3012548.
[46] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l 0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 1–12, 2011.
[47] Q. Yan *et al.*, "Cross-field joint image restoration via scale map," in *Proc. IEEE Conf. Comput. Vis.*, 2013, pp. 1537–1544.
[48] H. Zhang and J. Ma, "SDNet: A versatile squeeze-and-decomposition network for real-time image fusion," *Int. J. Comput. Vis.*, vol. 129, pp. 2761–2785, 2021.
[49] H. Zhang, H. Xu, X. Tian, J. Jiang, and J. Ma, "Image fusion meets deep learning: A survey and perspective," *Inf. Fusion*, vol. 76, pp. 323–336, Dec. 2021.
[50] L. Zhang, L. Zhang, and A. C. Bovik, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2579–2591, Aug. 2015.
[51] Q. Zhang, L. Xu, and J. Jia, "100 times faster weighted median filter (WMF)," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2830–2837.
[52] X. Zhang, Y. Ma, F. Fan, Y. Zhang, and J. Huang, "Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition," *JOSA A*, vol. 34, no. 8, pp. 1400–1410, 2017.
[53] 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*, vol. 87, no. 8, pp. 1315–1326, Aug. 1999.
[54] J. Zhao, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Infrared image enhancement through saliency feature analysis based on multi-scale decomposition," *Infrared Phys. Technol.*, vol. 62, pp. 86–93, 2014.
[55] J. Zhao, G. Cui, X. Gong, Y. Zang, S. Tao, and D. Wang, "Fusion of visible and infrared images using global entropy and gradient constrained regularization," *Infrared Phys. Technol.*, vol. 81, pp. 201–209, 2017.
[56] K. Zhou and C. Jung, "Multispectral fusion of RGB and NIR images using weighted least squares and alternating guidance," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 1489–1493.
[57] K. Zhou, C. Jung, and S. Yu, "Scale-aware multispectral fusion of RGB and NIR images based on alternating guidance," *IEEE Access*, vol. 8, pp. 173197–173207, 2020.
[58] S. Zhuo, X. Zhang, X. Miao, and T. Sim, "Enhancing low light images using near infrared flash images," in *Proc. IEEE Conf. Image Process.*, 2010, pp. 2537–2540.

CHEOLKON JUNG (Member, IEEE) is a Born Again Christian. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from Sungkyunkwan University, Republic of Korea, in 1995, 1997, and 2002, respectively. From 2002 to 2007, he was a Research Staff Member with Samsung Advanced Institute of Technology, Samsung Electronics, Republic of Korea. From 2007 to 2009, he was also a Research Professor with the School of Information and Communication Engineering, Sungkyunkwan University. Since 2009, he has been with the School of Electronic Engineering, Xidian University, China, where he is currently a Full Professor and the Director of the Xidian Media Laboratory. His main research interests include image and video processing, computer vision, pattern recognition, machine learning, computational photography, video coding, virtual reality, information fusion, multimedia content analysis and management, and 3DTV.

QIHUI HAN received the B.S. degree in automation engineering from Henan Polytechnic University, China, in 2013. He is currently working toward the Ph.D. degree in electronic engineering with Xidian University, China. His main research interests include image processing, computational photography, virtual reality, and deep learning.

KAILONG ZHOU received the B.S. degree in electronic engineering from Xidian University, China, in 2017. He is currently working toward the M.S. degree with Xidian University. His research interests include image processing and deep learning.



YUANQUAN XU received the B.S. degree in physical engineering from Zhengzhou University, China, in 2018. He is currently working toward the M.S. degree with Xidian University, China. His research interests include computer vision, 3D reconstruction, and machine learning.