

Article

Multi-Spectral Fusion and Denoising of Color and Near-Infrared Images Using Multi-Scale Wavelet Analysis [†]

Haonan Su * , Cheolkon Jung and Long Yu

School of Electronic and Engineering, Xidian University, No. 2 South Taibai Road, Xi'an, Shaanxi 710071, China; zhengzk@xidian.edu.cn (C.J.); lyu@stu.xidian.edu.cn (L.Y.)

* Correspondence: hnsu@xidian.edu.cn

[†] This paper is an extended version of our published paper: Multi-Spectral Fusion and Denoising of RGB and NIR Images Using Multi-Scale Wavelet Analysis. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.

Abstract: We formulate multi-spectral fusion and denoising for the luminance channel as a maximum a posteriori estimation problem in the wavelet domain. To deal with the discrepancy between RGB and near infrared (NIR) data in fusion, we build a discrepancy model and introduce the wavelet scale map. The scale map adjusts the wavelet coefficients of NIR data to have the same distribution as the RGB data. We use the priors of the wavelet scale map and its gradient as the contrast preservation term and gradient denoising term, respectively. Specifically, we utilize the local contrast and visibility measurements in the contrast preservation term to transfer the selected NIR data to the fusion result. We also use the gradient of NIR wavelet coefficients as the weight for the gradient denoising term in the wavelet scale map. Based on the wavelet scale map, we perform fusion of the RGB and NIR wavelet coefficients in the base and detail layers. To remove noise, we model the prior of the fused wavelet coefficients using NIR-guided Laplacian distributions. In the chrominance channels, we remove noise guided by the fused luminance channel. Based on the luminance variation after fusion, we further enhance the color of the fused image. Our experimental results demonstrated that the proposed method successfully performed the fusion of RGB and NIR images with noise reduction, detail preservation, and color enhancement.

Keywords: image fusion; wavelet decomposition; color enhancement; near-infrared; denoising



Citation: Su, H.; Jung, C.; Yu, L. Multi-Spectral Fusion and Denoising of Color and Near-Infrared Images Using Multi-Scale Wavelet Analysis. *Sensors* **2021**, *21*, 3610. <https://doi.org/10.3390/s21113610>

Academic Editor: Kang Ryoung Park

Received: 16 April 2021

Accepted: 18 May 2021

Published: 22 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In low light conditions, the captured RGB images are degraded with serious noise. Although many denoising methods [1–3] have been proposed and have obtained good performance in noise reduction, the performance on low light images requires improvement due to complicated noise modeling after a series of operations in the camera processing pipeline. Recent advances in multi-spectral imaging provide techniques to capture near infrared (NIR) and RGB images simultaneously [4,5]. As NIR images provide fine details and clear structure in the challenging condition, this technique is applied to a lot of multi-spectral image restorations, such as image dehazing [6], contrast enhancement [7], and image denoising [8].

In low light conditions, NIR cameras and dark flash are used to capture NIR images [9]. The dark flash projects light containing NIR and visible bands, and the visible band is blocked. NIR cameras are sensitive to the NIR spectral band with the range of 700 to 1100 nm [10]. In low light conditions, NIR images have the advantages of texture rendering without noise corruption compared to noisy RGB images because strong near-IR flash improves the NIR light reflection of the captured scenes, and thus NIR images capture more visible structures without noise corruption. On the other hand, RGB images contain better color information over NIR images. Therefore, NIR and RGB images were both employed to generate fused images with good structure, little noise, and vivid colors.

1.1. Related Work

1.1.1. Gradient Processing

The basic idea of gradient-processing-based multi-spectral fusion is NIR images containing a clear structure without noise corruption compared to RGB images in low light conditions. Thus, researchers [8,11–14] propose multi-spectral image restoration guided by NIR gradients. However, RGB and NIR images have large discrepancies in gradients, which leads to blurred artifacts in fused images. Therefore, the correlation between the gradients of two images is applied to overcome the discrepancy problem. Zhuo et al. [11] developed dual weighted least square (WLS) smoothing, which employs two gradients of RGB and NIR images to remove the noise and preserve the main image structure in the luminance channel of RGB images.

Then, the structure from NIR images is transferred to denoised RGB images. However, the details of the results are also blurred with noise reduction. Li et al. [12] proposed blind image deblurring with the guidance of NIR image gradients. With a constraint on the difference between RGB and NIR image gradients, they generated clean results by conducting blur kernel estimation and image restoration alternatively. Shen et al. [8,13] introduced the gradient scale map (i.e., gradient ratio map) to deal with the discrepancy problem on two image gradients (e.g., gradient magnitude variation and gradient direction divergence between two images).

Using the scale map, they proposed two image restoration methods that denoise RGB images based on the common edges between two images. However, this leads to blurred artifacts when the edge magnitudes between two images were greatly different. Sugimura et al. [14] developed the simultaneous denoising and deblurring of low light images using several NIR images with a short exposure time.

They developed an energy function that used the gradient correlation between RGB and NIR images, between three RGB bands, and between temporal sequences to remove noise and improve the quality of the low light images. Yamashita et al. [4,5] developed a novel sensor that combined RGB and N sensors together, and they used this sensor to capture NIR and RGB raw data under low light conditions. Then, they employed traditional image demosaicing, motion deblurring, and denoising to extract RGB and NIR images, and enhanced the low light images with serious noise and blur.

1.1.2. Image Fusion

In the first group of methods, we assumed that the NIR images contained more useful information without noise corruption compared with RGB images. This always occurs in texture regions that are highly corrupted in low light conditions. However, NIR images also contain distorted regions with low contrast and reduced details due to a different spectral sensitivity of NIR and RGB sensors [8]. The multi-spectral fusion guided by NIR images distorts the original structure of RGB images in fusion results. Thus, the idea of the second group of methods borrows from image fusion, which combines the most useful information from both the RGB and NIR images.

Image denoising is applied to fusion results or before image fusion. Son et al. [15] proposed the multi-spectral fusion of RGB and NIR images based on layer decomposition. They divided RGB and NIR images into base and detail layers. In the NIR base layer, they generated new NIR data, which had the same appearance as the RGB data and preserved the local contrast of NIR data with contrast preservation regularization. Then, they generated three detail layers (from the noisy RGB image, NIR image, and its new NIR data) and fused them. Then, the residual-based sparsity priors were applied to denoise the fused detail layer. However, blur artifacts and color noise still remained in the fusion results.

Son et al. [7] proposed a NIR coloring method for RGB image restoration. The proposed method consists of three steps: a contrast-preserving conversion that is the same as the new NIR data generation in [15], detail fusion between RGB and NIR images, and color fusion from RGB colors. Shibata et al. [16] proposed multi-spectral image fusion

using local contrast and inconsistent measurements. They estimated fusion weights based on the high visibility from two images and the prevention of artifacts from inconsistency. Li et al. proposed a novel encoder-decoder architecture consisting of convolutional layers extracting RGB and NIR image features, a fusion layer with two image features, and a dense block [17]. They introduced a dense block to reuse features from the middle layer for the reconstruction layer.

Zhang et al. proposed a novel fusion method based on the Pulse Coupled Neural Network (PCNN) in a Non-Subsampled Shearlet Transform (NSST) [18]. The adaptive parameters of PCNN were set by the inherent characteristics (i.e., the gradient energy and the magnitude of NSST coefficients) of images. Zheng et al. proposed a novel image fusion for haze removal using a set of artificial multi-exposure images that were generated by gamma correction [19]. They selected the best visual quality regions for fusion based on the entropy of the image texture.

Jung et al. proposed an unsupervised deep image fusion network consisting of three steps: feature extraction from RGB and NIR images, feature fusion between two image features, and fusion image reconstruction [20]. The unsupervised loss function based on a structure tensor was utilized to constrain the contrast of the output images similar to that of the input images.

1.2. Contributions

In this paper, we propose a multi-spectral fusion and denoising (MFD) framework of RGB and NIR images using multi-scale wavelet analysis. We assumed that RGB and NIR images have the same spatial resolution with well calibration, which means that the paired images contain matched structures without geometry displacements. The RGB and NIR bands are perfectly geometrically overlapped. There are no shadows or object movement in them. This is because the RGB or NIR images were taken by switching two filters (i.e., the IR cut filter and visible cut filter in front of the camera sensor) [10].

Thus, object movement leads to unmatched objects between paired images under different capturing times. To deal with the discrepancy and noise problem in the fusion, we formulate two observation models: (1) the noise model: the RGB image are corrupted with the additive Gaussian noise; (2) the discrepancy model: the discrepancy is measured as the correlation between RGB and NIR images, and we call this a wavelet scale map. Based on two observed models, we formulate the MFD framework as a maximum a posteriori (MAP) estimation that conducts the wavelet scale map estimation and image fusion alternatively.

To estimate the wavelet scale map, we utilize the priors of the wavelet scale map and its gradients as the contrast preservation term and gradient denoising term, respectively. To estimate the fused wavelet coefficients, we model the fusion coefficients as the Laplacian distribution with adaptive scaling parameters for noise removal. We apply the MFD framework to the luminance channel. In the chrominance channels, we utilize the fused luminance as a guidance to remove the chroma noise and provide color enhancement based on the luminance variation after fusion. Experimental results demonstrate that the proposed method generated fusion images with reduced noise, preserved details, and saturated colors.

Figure 1 illustrates the entire diagram of the proposed method. In Figure 1, Y , C_b , and C_r are the luminance and chrominance channels after color space conversion; Y' is the fused luminance channel in the spatial domain after the inverse wavelet transform; g and u are the wavelet coefficients of NIR images and the luminance channel from RGB images in each subband; ω_3 is the visibility ratio of NIR data to RGB data in (14); ω_4 is composed of ω_{4x} and ω_{4y} in (15) and (16); α_0 is the local contrast of NIR wavelet coefficients; s is the wavelet scale map; v is the wavelet coefficients in MFD; and Y^d , C_b^d , and C_r^d are denoised luminance and chrominance channels.

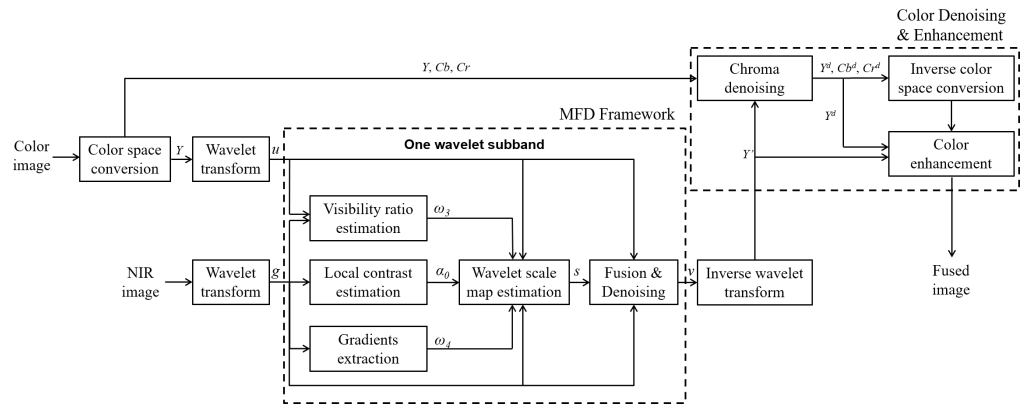


Figure 1. The entire framework of the proposed method (MFD: Multi-spectral fusion and denoising).

Compared to our previous work [21], we have four extensions: (1) We introduce two observation models of noise and discrepancy from RGB data, NIR data, and the wavelet scale map in problem formulation. These models provide the mathematical basis for Bayesian derivation. (2) We apply the MFD to only the luminance channel. We perform guided filtering to two chroma channels and color enhancement based on the luminance variation after fusion. This can save the computational cost in the other two channels and generate results with vivid colors. (3) We provide a simplified MFD framework for the low frequency band to generate the results with a more visible structure of the NIR component in the base layer. (4) We formulate the prior of the fused wavelet coefficients as the adaptive Laplacian distribution using the guidance of NIR data. This provides good performance in detail preservation.

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

- We propose the MFD framework for RGB and NIR image fusion in the wavelet domain to achieve both texture transfer and noise removal.
- We provide the discrepancy model based on the wavelet scale map (correlation between RGB and NIR data) to deal with the discrepancy between RGB and NIR images.
- We combine three probability terms of contrast preservation, gradient denoising, and fusion denoising into the MFD framework to resolve the discrepancy while reducing the noise in the fusion.
- We enhance the color based on the luminance variation after fusion. The enhanced colors are more vivid with less color distortion.

The remainder of this paper is as follows. We describe the details of the proposed method in Section 2, while we provide the experimental results and their corresponding analysis in Section 3. We draw conclusions of this paper in Section 4.

2. Proposed Method

2.1. Problem Formulation

In this work, we aim to generate a high quality image by fusion and denoising from noisy RGB and NIR images. We first normalize the color bands by the maximum pixel value 255 (8-bit image). We decompose the Y channel of noisy RGB and NIR images using wavelet analysis. We denote the RGB wavelet coefficients in one subband as the vector \mathbf{u} with N elements, the NIR wavelet coefficients in one subband as the diagonal matrix \mathbf{g} with $N \times N$ elements, where the diagonal elements are NIR wavelet coefficients, and the fusion image as the vector \mathbf{v} with N elements in the wavelet domain. The relationship between RGB, NIR and fusion images are formulated as follows:

$$\mathbf{u} = \mathbf{v} + \mathbf{N}_1 \quad (1)$$

$$\mathbf{v} = \mathbf{g}\boldsymbol{\alpha} + \mathbf{N}_2 \quad (2)$$

where the vector N_1 represents the random Gaussian noise with N elements, the vector α with N elements is defined as wavelet scale map to model the correlation between RGB and NIR images, and the vector N_2 means the random error with N elements. (1) represents the noise model. That means the observed noisy RGB wavelet component u is generated from the desired RGB wavelet component v with noise N_1 . (2) represents the discrepancy model. The relationship between RGB and NIR data are modeled as a linear function.

2.2. Multi-Spectral Fusion and Denoising Framework

We preform multi-spectral fusion and denoising (MFD) of NIR and RGB images based on multi-scale wavelet analysis and Bayesian theory [1]. Direct fusion of RGB and NIR images causes annoying artifacts, such as contrast attenuation due to the large discrepancy between two as shown in Figure 2c. Thus, we estimate a wavelet scale map to update the NIR wavelet coefficients and deal with the discrepancy between them. We formulate MFD as a maximum a posterior (MAP) estimation problem that finds clean RGB wavelet coefficients v and a wavelet scale map α given noisy RGB wavelet coefficients u and NIR wavelet coefficients g . For the MAP estimation, we maximize:

$$\begin{aligned} v, \alpha &= \max_{v, \alpha} p(v, \alpha | u, g) \\ &= \max_{v, \alpha} \{p(u, g | v, \alpha) \cdot p(v, \alpha) / p(u, g)\} \\ &\propto \min_{v, \alpha} \{-\log p(u, g | v, \alpha) - \log p(v, \alpha)\} \end{aligned} \quad (3)$$

where $p(v, \alpha | u, g)$ is the posterior and $p(u, g | v, \alpha)$ is the likelihood; $p(v, \alpha)$ is the prior with joint distribution of v and α . $p(u, g)$ has the same definition as $p(v, \alpha)$. $p(u, g)$ is a constant and thus is omitted in the optimization.

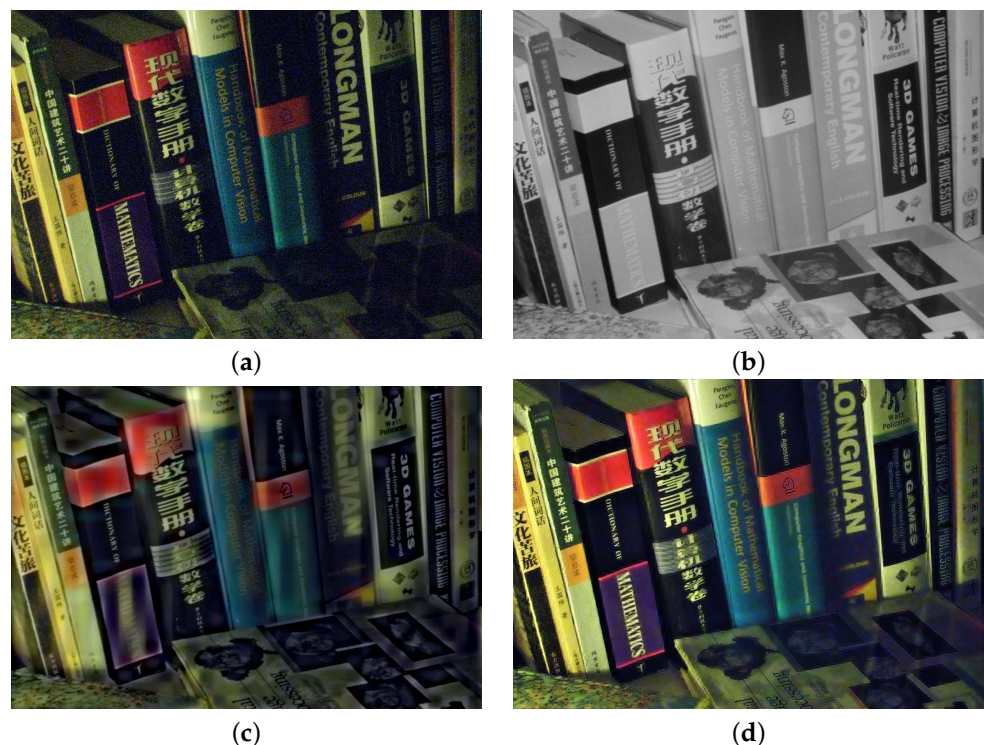


Figure 2. Fusion and denoising results in *Books* by Guided Wavelet Shrinkage (GWS) [1]. (a) RGB image. (b) NIR image. (c) GWS. (d) Proposed method.

First, we design the likelihood for the fusion of noisy RGB and NIR images based on the noise model and the discrepancy model in Section II.A. Given v and α , u and g are independent to each other. Thus, we split the likelihood into two parts as follows:

$$p(u, g|v, \alpha) = p(u|v, \alpha) \cdot p(g|v, \alpha) \quad (4)$$

$$\begin{aligned} p(u|v, \alpha) &\propto p(u|v) \\ &= N(u - v|0, \xi_1) \end{aligned} \quad (5)$$

$$p(g|v, \alpha) = N(v - g\alpha|0, \xi_2) \quad (6)$$

where $N(u - v|0, \xi_1)$ is a Gaussian distribution with zero mean and variance ξ_1 , $N(v - g\alpha|0, \xi_2)$ has the same definition as $N(u - v|0, \xi_1)$, and ξ_1 and ξ_2 represent random Gaussian noise and random error in (1) and (2). Based on the noise model in (1), u is independent of α , and thus α is removed in (5), and the $p(u|v)$ is defined as the Gaussian distribution in (5). Based on the discrepancy model in (2), we define $p(g|v, \alpha)$ as the Gaussian distribution in (6).

Second, we define the joint prior distribution of $p(v, \alpha)$. In (2), without considering g , v should be independent to α . The joint prior distribution of $p(v, \alpha)$ is separated into two parts as follows:

$$p(v, \alpha) = p(v) \cdot p(\alpha) \quad (7)$$

where $p(v)$ and $p(\alpha)$ are the priors of v and α . Then, to preserve the local contrast from the NIR data and remove noise in the scale map, we define the prior of the wavelet scale map α , which consists of two parts: (1) the prior of its magnitude $p_{lc}(\alpha)$; (2) the prior of its gradient $p_{gd}(\partial^* \alpha)$ ($\partial^* \alpha \in \{\frac{\partial \alpha}{\partial x}, \frac{\partial \alpha}{\partial y}\}$, $\frac{\partial \alpha}{\partial x}$ and $\frac{\partial \alpha}{\partial y}$ are the partial derivatives of α) as follows:

$$p(\alpha) = p_{lc}(\alpha) \cdot p_{gd}(\partial \alpha) \quad (8)$$

$$p_{lc}(\alpha) = N(\alpha|\alpha_0, \xi_3) \quad (9)$$

$$\begin{aligned} p_{gd}(\partial^* \alpha) &= N\left(\frac{\partial \alpha}{\partial x}|0, \xi_{4x}\left(\frac{\partial g}{\partial x}\right)\right) \cdot \\ &N\left(\frac{\partial \alpha}{\partial y}|0, \xi_{4y}\left(\frac{\partial g}{\partial y}\right)\right) \end{aligned} \quad (10)$$

where $p_{lc}(\alpha)$ is defined as the Gaussian distribution with the mean of α_0 and the variance of ξ_3 . This term is used as the local contrast preservation term that transfers the high contrast and visibility of NIR image to the fusion result. α_0 is the directive contrast from NIR components, which selects the high contrast component from the NIR wavelet coefficients. Moreover, we define $p_{gd}(\partial^* \alpha)$ as the gradient denoising term that utilizes the gradients of the NIR component as the guidance for adaptive noise removal in wavelet scale map estimation of the high-pass band. ξ_{4x} and ξ_{4y} are defined as the function of the gradient of the NIR components for denoising. $\frac{\partial g}{\partial x}$ and $\frac{\partial g}{\partial y}$ are the partial derivatives of g .

Finally, the prior of the fusion wavelet coefficient can be modeled as the heavy tailed distributions for noise removal, such as the Laplacian distribution and Generalized Gaussian distribution [22]. In this work, the prior $p(v)$ is defined as zero-mean Laplacian with the scaling parameter as follows:

$$p(v) = \frac{1}{2\xi_5} e^{-\frac{\|v\|}{\xi_5}} \quad (11)$$

where ξ_5 is the scaling parameter for the Laplacian distribution.

Based on (1)–(11), we perform MAP estimation by the minimization of energy function as follows:

$$E(\mathbf{v}, \boldsymbol{\alpha}) = \omega_1 \|\mathbf{v} - \mathbf{u}\|^2 + \omega_2 \|\mathbf{v} - \mathbf{g}\boldsymbol{\alpha}\|^2 + \omega_3 \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|^2 + \omega_{4x} \left\| \frac{\partial \boldsymbol{\alpha}}{\partial x} \right\|^2 + \omega_{4y} \left\| \frac{\partial \boldsymbol{\alpha}}{\partial y} \right\|^2 + \omega_5 \|\mathbf{v}\| \quad (12)$$

where the parameters ω_1 – ω_5 in (12) are the inverse of the variances ξ_1 – ξ_4 and the scaling parameter ξ_5 , i.e., $\omega_{1,2,5} = \xi_{1,2,5}^{-1}$, $\omega_{3,4x,4y} = -\xi_{3,4x,4y}^{-1}$.

2.3. Parameter Description

The parameters of each term in (12) are described as follows:

2.3.1. Parameters ω_1 and ω_2

The first term and the second term are the fusion weight of the NIR and RGB wavelet coefficients. We set $\omega_1 = \omega_2 = 0.5$ for fusion.

2.3.2. Parameters ω_3 and α_0

The third term represents the contrast preservation term in the wavelet scale map estimation. In the high-pass band, we obtain the directive contrast α_0 using Weber's law, i.e., the ratio of the Laplacian gradients in the high frequency subband to the local luminance in the low pass subband [23], as follows:

$$\alpha_0(x, y) = \begin{cases} \left(\frac{1}{g_l(x, y)}\right)^\gamma \cdot \frac{SML(x, y)}{g_l(x, y)} & \text{if } g_l(x, y) \neq 0 \\ SML(x, y) & \text{if } g_l(x, y) = 0 \end{cases} \quad (13)$$

where $SML(x, y)$ is the sum-modified-Laplacian gradient [23], $g_l(x, y)$ is the local luminance in the base subband and γ is the visual sensitivity to luminance that ranges from 0.6 to 0.7. We apply the visibility map to the parameter ω_3 for transferring the NIR component more than the RGB component as follows:

$$\omega_3 = \tau \cdot \phi(VI_{NIR}/VI_{RGB}|\sigma_1, \gamma_1) \quad (14)$$

where the visibility map VI is generated by [24], which evaluates the signal visibility of the human visual system (HVS) in the wavelet domain, $\phi(\cdot|\sigma_1, \gamma_1)$ uses the wavelet shrinkage function [1] with parameters σ_1 and γ_1 as the transfer function to compress the dynamic range into $[0, 1]$, τ is the constant value, which is set to 10^{-4} .

Severe noise degrades the main structure of the visibility map in high-pass bands, which significantly affects the performance of the contrast preservation term. We use the relative total variation [25] to reduce the noise of VI and produce the smoothing weight ω_3 for structure preservation. The ratio of VI_{NIR} to VI_{VIS} determines the visibility of NIR components over RGB components, i.e., a larger ratio, the more visible the contrast from the NIR components. Thus, a larger ratio, i.e., a larger ω_3 , provides more contrast transfer from the NIR to RGB components. Here, we set $\sigma_1 = 0.5$ and $\gamma_1 = 2$.

2.3.3. Parameters ω_{4x} and ω_{4y}

The fourth term considers the use of the gradients of NIR coefficients to guide denoising for the wavelet scale map. ω_{4x} and ω_{4y} are defined as follows:

$$\omega_{4x} = \lambda \cdot \left(\left\| \frac{\partial \mathbf{g}}{\partial x} \cdot \boldsymbol{\alpha} \right\|^\beta + \varepsilon \right)^{-1} \quad (15)$$

$$\omega_{4y} = \lambda \cdot \left(\left\| \frac{\partial \mathbf{g}}{\partial y} \cdot \boldsymbol{\alpha} \right\|^\beta + \varepsilon \right)^{-1} \quad (16)$$

where λ , β , and ε are the parameters of wavelet scale map denoising. We set $\lambda = 10^{-(M-1)} \cdot \omega_3/\tau$ and M are the decomposition level. λ is related to the maximum decomposition level M . ω_3 is used to adaptively control the weight (i.e., ω_{4x} and ω_{4y}) of gradient denoising. Large ω_3 provides coarse filtering (i.e., high ω_4 value) of the wavelet scale map, which selects whole regions of visible NIR pixels and then transfers them to RGB ones. Small ω_3 (i.e., low ω_4) achieves the careful filtering along the gradients of NIR wavelet coefficients, which makes them close to RGB data. Thus, ω_3 is large for visible NIR data and small for visible RGB data.

2.3.4. Parameters ω_5

The fifth term enforces the denoising of desirable RGB components v . We define the adaptive weights ω_5 by the adjusted NIR components as:

$$\omega_5 = \eta \cdot e^{-\frac{g \cdot \alpha}{avg(g \cdot \alpha)}} \quad (17)$$

where η is the parameter from 0.01 to 0.001 and $avg(\cdot)$ is the average function. ω_5 controls the denoising degree based on the adjusted NIR wavelet coefficient. The large magnitude of the adjusted NIR components means more visibility of the NIR data, which is transferred to the fusion result. Thus, it enforces weak denoising (i.e., small ω_5 value) for the fusion result because there is less noise in the NIR data. In contrast, a smaller adjusted NIR component means less fusion from the NIR data and, thus, stronger denoising (i.e., a larger ω_5 value) due to noisy RGB data.

2.4. Numerical Solution

We obtain the latent image by both the fusion and denoising of RGB and NIR images. MFD is iteratively performed by estimating α and v as follows:

Optimizing α : with the fixed v , α is calculated with the minimization of energy function $E(\alpha)$ as follows:

$$E(\alpha) = \omega_2 \cdot \|v - g\alpha\|^2 + \omega_3 \cdot \|\alpha - \alpha_0\|^2 + \omega_{4x} \cdot \left\| \frac{\partial \alpha}{\partial x} \right\|^2 + \omega_{4y} \cdot \left\| \frac{\partial \alpha}{\partial y} \right\|^2 \quad (18)$$

Based on the parameter design section, we rewrite (18) using the matrix notation as follows:

$$\alpha = \min_{\alpha} \{ \omega_2 (v - g\alpha)^T (v - g\alpha) + (\alpha - \alpha_0)^T \omega_3 (\alpha - \alpha_0) + \alpha^T D_x^T \omega_{4x} D_x \alpha + \alpha^T D_y^T \omega_{4y} D_y \alpha \} \quad (19)$$

where v , α , and α_0 are the vector forms of v , α , and α_0 , respectively; g , ω_3 , ω_{4x} , and ω_{4y} are the diagonal matrices of the original variables (g , ω_3 , ω_{4x} , and ω_{4y}); D_x and D_y are forward difference operators, respectively, and thus D_x^T and D_y^T are backward difference operators. We use a solver [26] to minimize (19). We obtain α by minimizing the energy function in (19); thus, the solution of the linear system as follows:

$$(\omega_2 g^T g + \omega_3 + D_x^T \omega_{4x} D_x + D_y^T \omega_{4y} D_y) \alpha = (\omega_2 g^T v + \omega_3^T \alpha_0) \quad (20)$$

$$\alpha = (\omega_2 g^T g + \omega_3 + D_x^T \omega_{4x} D_x + D_y^T \omega_{4y} D_y)^{-1} (\omega_2 g^T v + \omega_3^T \alpha_0) \quad (21)$$

Optimizing v : with the fixed α , v is calculated by minimizing $E(v)$ as follows:

$$E(v) = \omega_1 \|v - u\|^2 + \omega_2 \|v - g\alpha\|^2 + \omega_5 \|v\| \quad (22)$$

The solution of v in (22) is the soft-thresholding function as follows [1,27]:

$$\begin{aligned}\hat{v} &= \omega_1 \mathbf{u} + \omega_2 \mathbf{g} \boldsymbol{\alpha} \\ v &= \text{sign}(\hat{v}) \max(\|\hat{v}\| - \omega_5, 0)\end{aligned}\quad (23)$$

where $\max(\cdot)$ is the max function and $\text{sign}(\cdot)$ is the sign function, which is defined as $x = 1$ when $x > 0$, $x = -1$ when $x < 0$ and $x = 0$ when $x = 0$.

2.5. Application to Low Pass Fusion

In the low pass subband, we employ the same MFD framework to fuse the NIR and RGB components. However, the denoising term is removed because there is less noise in the low-pass band. (19) and (23) are simplified as follows:

$$\boldsymbol{\alpha} = (\omega_2 \mathbf{g}^T \mathbf{g} + \omega_3)^{-1} (\omega_2 \mathbf{g}^T v + \omega_3^T \boldsymbol{\alpha}_0) \quad (24)$$

$$v = \omega_1 \mathbf{u} + \omega_2 \mathbf{g} \boldsymbol{\alpha} \quad (25)$$

$$\alpha_0(i) = R_i g(i) / \text{avg}(R_i g) \quad (26)$$

where α_0 is the directive contrast of RGB components in the low-pass band, which is defined as the ratio of the center pixel intensity to average intensity in the window, $\text{avg}(\cdot)$ is average function in an window, R_i is a matrix that extracts the patch at the i th pixel location from an image, and the visibility degree VI of low-pass band is calculated similar to that in high-pass band. However, we use a luminance adaptation model [24] to calculate JND threshold without considering inter-band and intra-band masking.

2.6. Unified MFD Framework for RGB and NIR Image Fusion

As shown in Figure 1, we first apply the color space conversion, which converts RGB color space to the decorrelated color space (we use YCbCr color space) for the degraded RGB images. Then, we decompose luminance channels of RGB and NIR images by DT-CWT. Next, we use the MDF framework to fuse luminance channel of RGB images and NIR images. The fused luminance image is produced by inverse DT-DWT. Then, we employ a guided filter [28] to denoise the luminance and chrominance channels, and the denoised RGB image is obtained by inverse color space conversion.

Finally, we enhance colors based on the luminance variation after fusion. Algorithm 1 depicts the MDF framework. $\boldsymbol{\alpha}^n$, $\boldsymbol{\alpha}_0$, \mathbf{u} , \mathbf{g} and v^n mean the variables in high frequency band, where $\boldsymbol{\alpha}_1^n$, $\boldsymbol{\alpha}_{10}$, \mathbf{u}_1 , \mathbf{g}_1 and v_1^n mean the variables in the low-pass band. M is the maximum decomposition level chosen as a large value because DT-CWT with a large decomposition level (e.g., 3~5) extracts much noise in high pass subband, while N is maximum iteration number. Figure 2d shows that the proposed method successfully handles the discrepancy problem between RGB and NIR images while achieving the fusion result with noise reduction and good contrast based on the NIR image.

We provide the results of wavelet scale map estimation, and MFD framework for high pass and low-pass bands. Figure 3 shows the generation process for wavelet scale map in one wavelet subband. ω_3 represents the visibility of NIR wavelet coefficients compared to RGB components. α_0 represents the local contrast from NIR components. The wavelet scale map α without α_0 calculates the correlation (i.e., ratio) of wavelet coefficients between NIR and RGB components to deal with the discrepancy problem. The gradient denoising term guided by NIR wavelet coefficients reduces noise in the scale map and thus main structure in NIR image appears in the scale map. Then, α adjusts wavelet coefficients of NIR image to be compatible with RGB components.

Guided by the local contrast α_0 and visibility ω_3 , we select high contrast and visibility regions from NIR components by α (see the red blocks in Figure 3). Figure 4 shows the fusion and denoising results of NIR and RGB wavelet coefficients in high-pass band. We demonstrated that the fused results by the proposed method contained fine details from the NIR components (see the red blocks in Figure 4). Figure 5 shows the low-pass band

fusion results.

Algorithm 1 Multi-scale fusion and denoising of NIR and RGB images.

Input: Noisy gray image from RGB image, NIR image

Initialize: $\omega_1 = \omega_2 = 0.5, \sigma_1 = 0.5, \gamma_1 = 2, \eta = 0.005$

$\beta = 1.2, \varepsilon = 0.001, \alpha^0 = \mathbf{I}, v^0 = \mathbf{u}$

$M = 3 \sim 5, N = 20, \tau = 10^{-4}, \varepsilon_1 = \varepsilon_2 = 10^{-2}$.

1. Perform DT-CWT on noisy gray and NIR images.

2. **Detail layer:**

For $m=1:M$ (M : Maximum decomposition)

For $n=1:N$ (N : Maximum iteration number)

a. Calculate VI of v^n and g [24];

b. Calculate α_0 from g by (13);

c. Calculate $\omega_3 - \omega_5$ by (14)–(17);

d. Optimize α^{n+1} by (21);

e. Optimize v^{n+1} by (23);

if $\|\alpha^{n+1} - \alpha^n\|_2^2 / \|\alpha^n\|_2^2 < \varepsilon_1$ and

$\|v^{n+1} - v^n\|_2^2 / \|v^n\|_2^2 < \varepsilon_2$; **break**;

end For

end For

3. **Base layer:**

For $n=1:N$ (N : Maximum iteration number)

a. Calculate VI of v^n and g_l [24];

b. Calculate α_{l0} from g by (26);

c. Calculate ω_3 by (14);

d. Optimize α_l^{n+1} and v_l^{n+1} by (24)–(25);

if $\|\alpha_l^{n+1} - \alpha_l^n\|_2^2 / \|\alpha_l^n\|_2^2 < \varepsilon_1$ and

$\|v_l^{n+1} - v_l^n\|_2^2 / \|v_l^n\|_2^2 < \varepsilon_2$; **break**;

end For

4. Perform inverse DT-CWT.

Output: Fused gray image.

Compared to the fusion without α , the fusion results with α have a similar appearance to the original gray images, which indicates that the fusion results with α suffered less discrepancy artifacts (see Figure 5d,e). Figure 5c shows the local contrast map of the NIR data where bright pixels mean more details were transferred to the RGB components. Thus, the proposed base band fusion with α and α_0 solved the discrepancy problem and obtained visible details from NIR data as shown in the red blocks of Figure 5d,e.

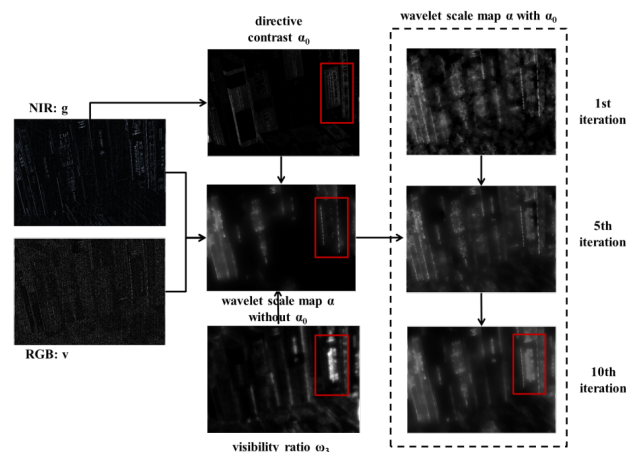


Figure 3. Wavelet scale map generation.

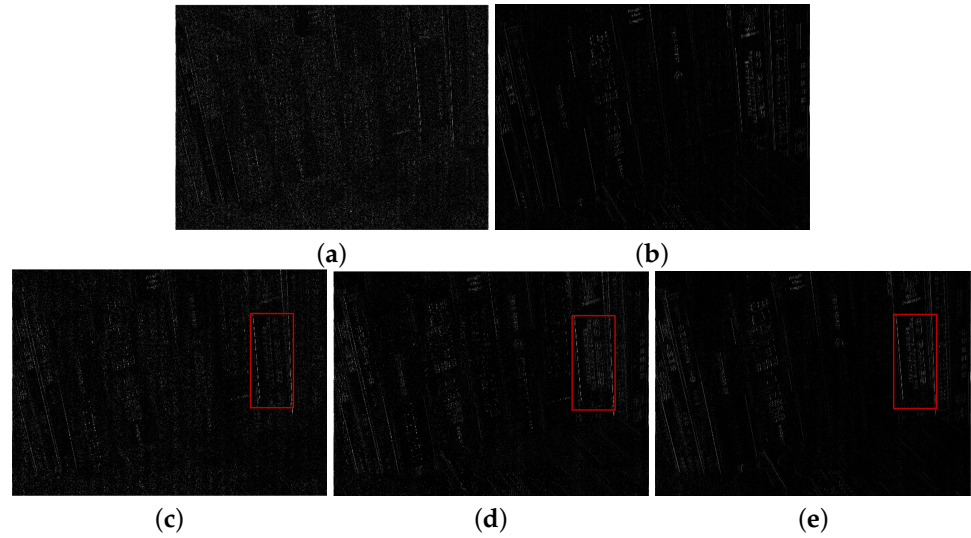


Figure 4. Fusion and denoising results of the high-pass bands in *Books*. (a) RGB image. (b) NIR image. (c) First iteration. (d) Third iteration. (e) Tenth iteration.

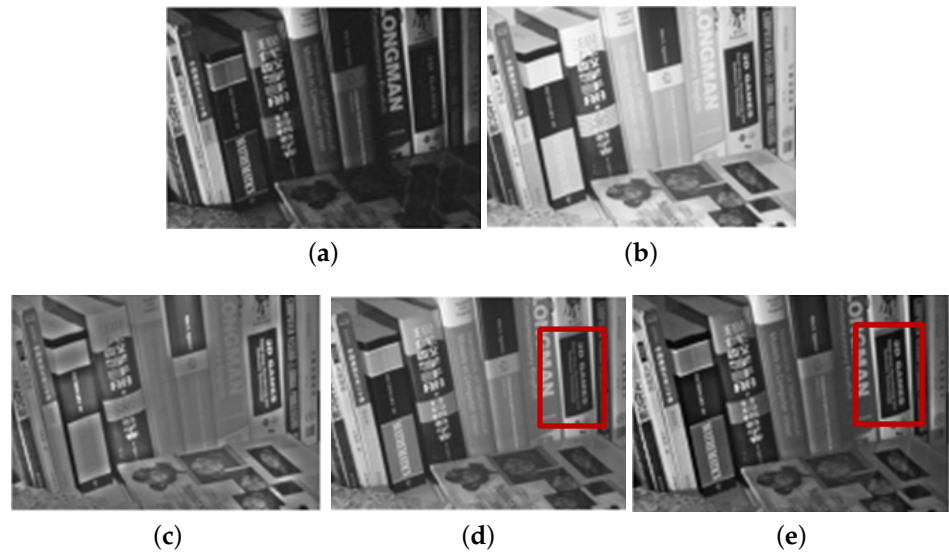


Figure 5. Fusion results of low-pass bands in *Books*. (a) RGB image. (b) NIR image. (c) α_0 . (d) Fusion without α . (e) Fusion with α .

2.7. Chroma Denoising and Color Enhancement

In Section IV.B, we obtained the fused luminance image by combining RGB and NIR images. Then, we employed the guided image filter [28] to denoise the chrominance channels guided by the fused luminance results [29]. The primary idea of [29] is the high correlation in the texture information over color channels. Based on the correlation, [29] transfers the texture information of the fused luma channel to the other two channels and removes the noise around the texture by the guided filter. Meanwhile, the guided filter is employed to remove noise in the noisy luma channel. Then, the denoised luma and chroma channels are combined and inverse color space conversion is used to obtain the denoised original images. Finally, we enhance the color of the fused image based on the variation of luminance after fusion as follows [30]:

$$M_e(x, y) = \left(\frac{Y'(x, y)}{Y^d(x, y)} \right)^\beta \cdot M_d(x, y) \quad (27)$$

where $M_e(x, y)$ and $M_d(x, y)$ are trichromatic (three channels, i.e., R,G,B channels) channel values of the output color and denoised images, respectively; $Y'(x, y)$ and $Y^d(x, y)$ are gray images from the fused result and the denoised original image, respectively; and β is the sensitive factor whose range is $[0.6, 1.0]$. In (27), we utilize the ratio of luminance variation to enhance the three channels independently.

As a result, the colors of the fusion images are enhanced more with the increase of β (see the trees in Figure 6). Figure 7 shows the fusion and color enhancement results of the proposed method. The proposed method produces fusion results with noise removal and detail transferring from the NIR images (see Figure 7a,b). Color enhancement provides saturated colors in images compared to the direct inverse color space conversion (see Figure 7c,d).

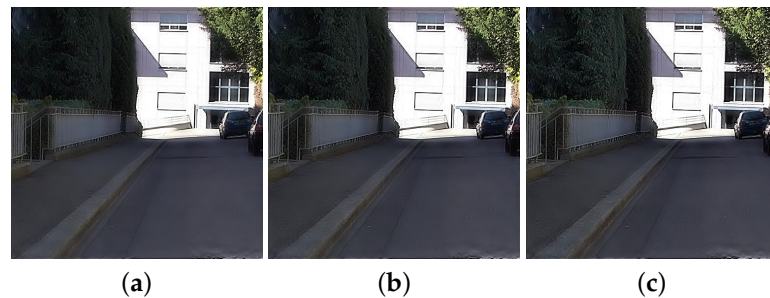


Figure 6. Color enhancement results with (a) $\beta = 0.6$, (b) $\beta = 0.8$, and (c) $\beta = 1.0$.

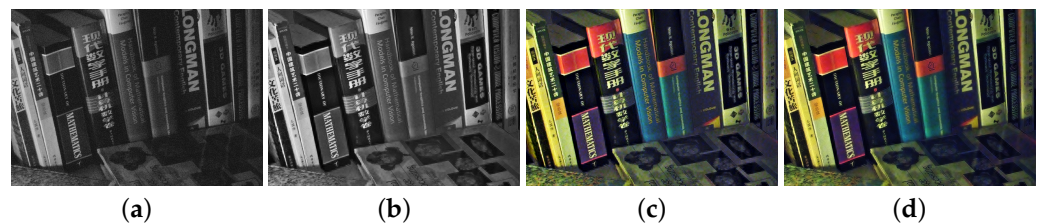


Figure 7. Fusion and color enhancement results in *Books*. (a) Original gray image. (b) Fusion results in the luma channel. (c) Inverse color space conversion (YCbCr \rightarrow RGB). (d) Color enhancement results.

3. Experimental Results

3.1. Multi-Spectral Fusion of NIR and RGB Images

In the experiments, we used fifteen pairs of RGB and NIR images obtained from [7,10,15] as shown in Figure 8. Figure 8a–d are from [15], Figure 8e–h are from [7], and Figure 8i–o are from [10]. We added Gaussian noise to the RGB images in Figure 8e–o. The RGB and NIR image pairs are well registered with the same spatial resolution from 436×512 to 1147×800 . For the tests, we used a PC with Intel (R) Core (TM) 547 i5 CPU (2.60 GHZ) and 4.00 GB RAM running a Windows environment and MATLAB 2012b. For deep learning methods, we used Nvidia GTX2080Ti with the Ubuntu 16.04 environment.

We compared the performance of the proposed method with those of the guided wavelet shrinkage (GWS) [1], dual WLS smoothing (DWLS) [11], scale map (SM) [8], DenseFuse [17], and Unsupervised Deep Image Fusion using structure tensor (UDIF) [20]. To conduct the experiments, we first removed noise using BM3D [31], and then we employed DenseFuse and UDIF for fusion. Figures 9–12 show the experimental results from different methods.

GWS [1] directly fuses RGB and NIR images without considering discrepancies, and thus there is much contrast attenuation in the results as shown in Figures 9e, 10e, 11e, and 12e. DWLS [11] uses the gradients of the NIR and Y channels of the initially denoised RGB images as the guidance for WLS smoothing and, then, transfers the details of NIR data to the smoothing results. However, the results seem blurred in the details because the initial denoising smoothes some details of the RGB data (see the red blocks in Figures 11c and 12c), and contains some noise (see the red blocks in Figures 9c and 10c).

SM [8] uses the scale map to correct the discrepancy of common edges between RGB and NIR images; however, some different details between two images seem blurred and are lost in the results (see the red blocks in Figures 10d and 11d). The zoomed red blocks in Figure 10d are provided in Figure 10i. DenseFuse [17] introduces the dense blocks in an encoder–decoder framework and preserves more extracted features of the middle layer. However, the generated results causes serious color distortion (see Figures 9–12f) and loss of details (see the red blocks of Figures 10 and 11f and the zoomed regions in Figures 10 and 11i).

UDIF [20] is an unsupervised deep learning fusion network with a structure tensor loss. They lead to the attenuated details (see the red blocks in Figures 10 and 11g and the zoomed regions in Figures 10 and 11i). Color distortion happens in the results (see Figure 12g). The proposed method adjusts the NIR coefficients to have the same distribution as RGB data and obtain a wavelet scale map for fusion. Thus, our method effectively handles the discrepancy between RGB and NIR images and produces both fusion and denoising without contrast attenuation compared to GWS (see Figures 9–12e,h).

Moreover, the contrast preservation term and visibility ratio detect high and visible contrasts from the NIR wavelet coefficients and then transfer them to RGB components. Thus, our results successfully preserve visible local contrast from the NIR data (see Figures 9–12h). The proposed method reduces more noise and artifacts compared with DWLS (see Figures 9 and 10c,h, and the zoomed region in Figure 10i) because of the Laplacian distribution modeling of the fusion wavelet coefficients. As shown in Figure 11, the proposed method preserves the local contrast of RGB images better than DWLS and SM (see the red blocks on the wall of Figure 11c,d,h).

This is because the low visibility ratio on the wall, i.e., the low visibility degree in the NIR data, reduces the effect of the contrast preservation term. As the gradient denoising term provides gradient enhancement for the scale map guided by the NIR wavelet coefficients, the proposed method is very effective in enhancing the details in the boxes compared to DenseFuse and UDIF (see the red blocks on the boxes in Figure 11f–h and its zoomed regions in Figure 11i).

As shown in Figure 12, the proposed method achieved better performance in detail preservation and color reproduction compared with the other methods. Finally, the proposed method generates the fusion results with vivid colors compared to DenseFuse and UDIF (see Figures 9, 11 and 12f–h) because we enhance the colors corresponding to the luminance variation.

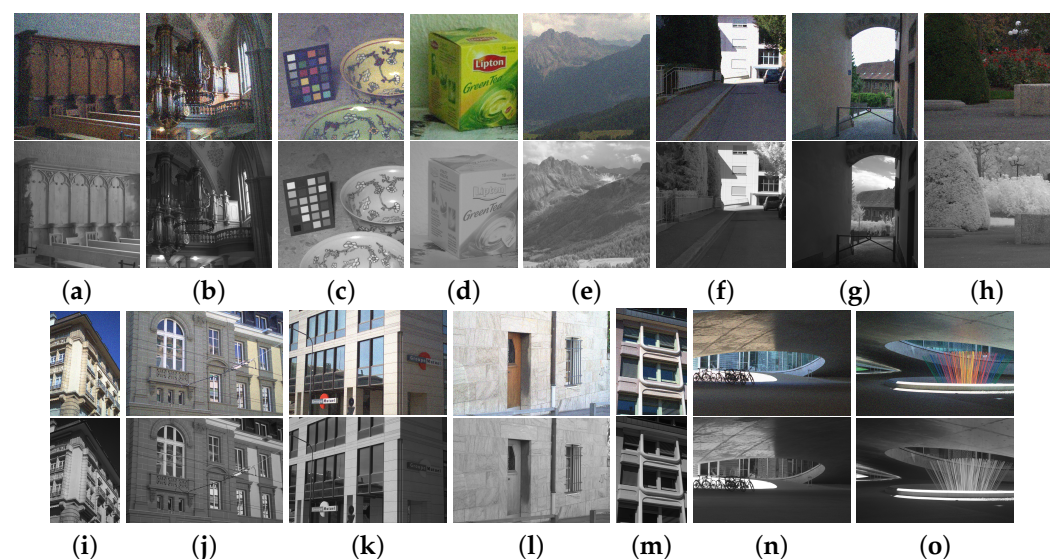


Figure 8. Test image pairs with noisy RGB images (top) and NIR images (bottom).

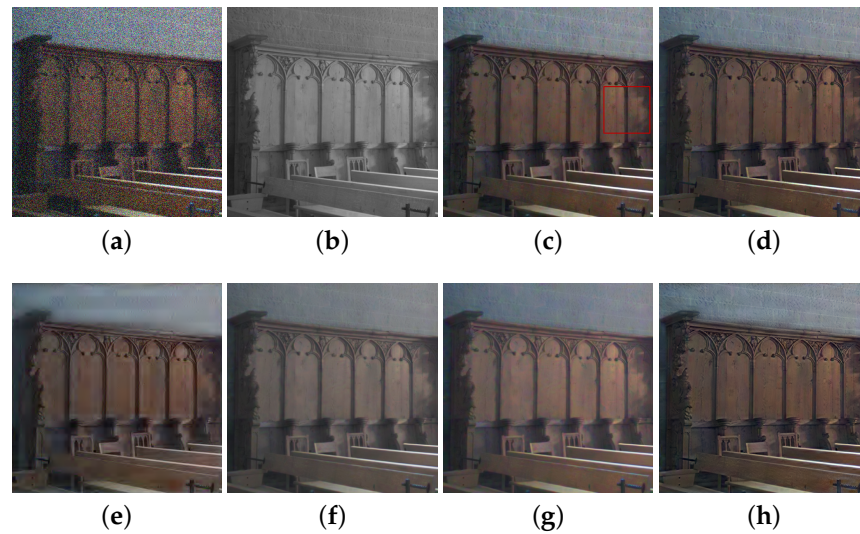


Figure 9. Experimental results for a test image. (a) Noisy RGB image. (b) NIR image. (c) DWLS [11]. (d) SM [8]. (e) GWS [1]. (f) DenseFuse [17]. (g) UDIF [20]. (h) Proposed method.

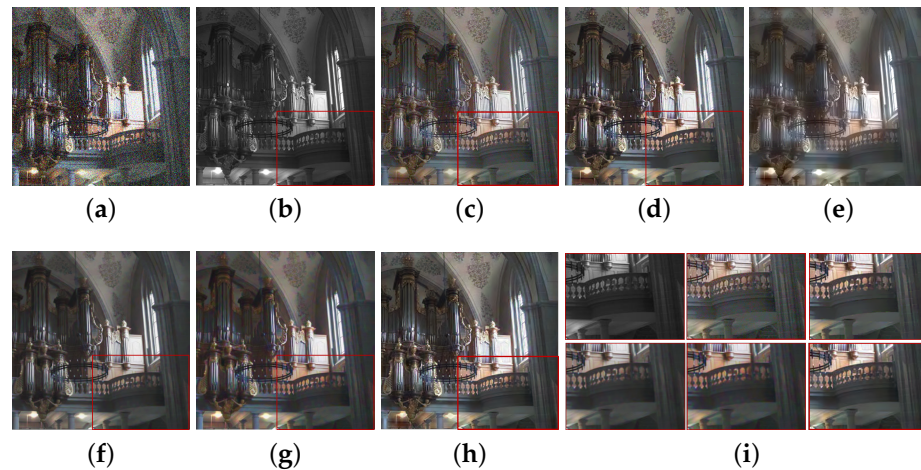


Figure 10. Experimental results for a test image. (a) Noisy RGB image. (b) NIR image. (c) DWLS [11]. (d) SM [8]. (e) GWS [1]. (f) DenseFuse [17]. (g) UDIF [20]. (h) Proposed method. (i) Zoomed regions of red blocks (From top left to bottom right: (b-d, f-h)).

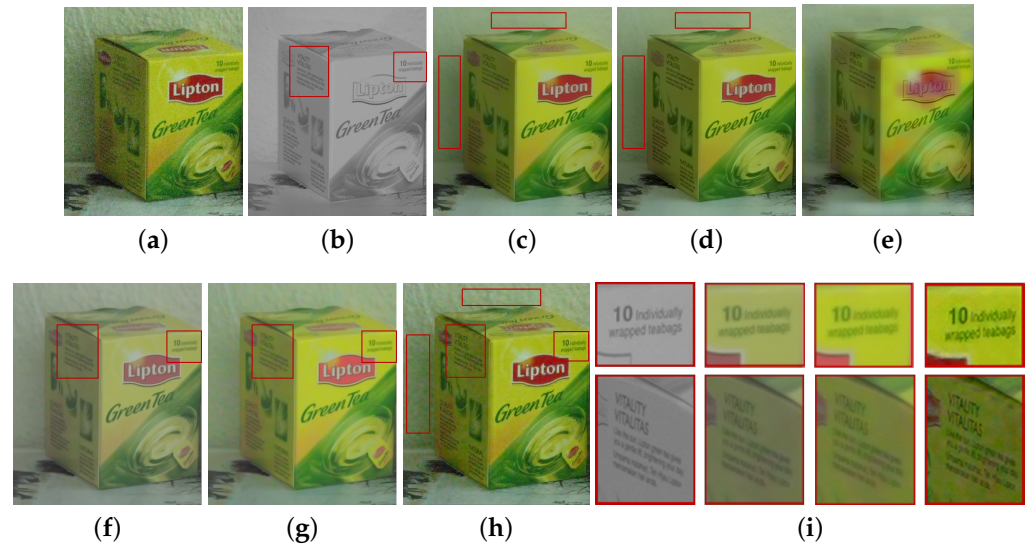


Figure 11. Experimental results for a test image. (a) Noisy RGB image. (b) NIR image. (c) DWLS [11]. (d) SM [8]. (e) GWS [1]. (f) DenseFuse [17]. (g) UDIF [20]. (h) Proposed method. (i) Zoomed regions of red blocks (From Left to Right: (b,f–h)).

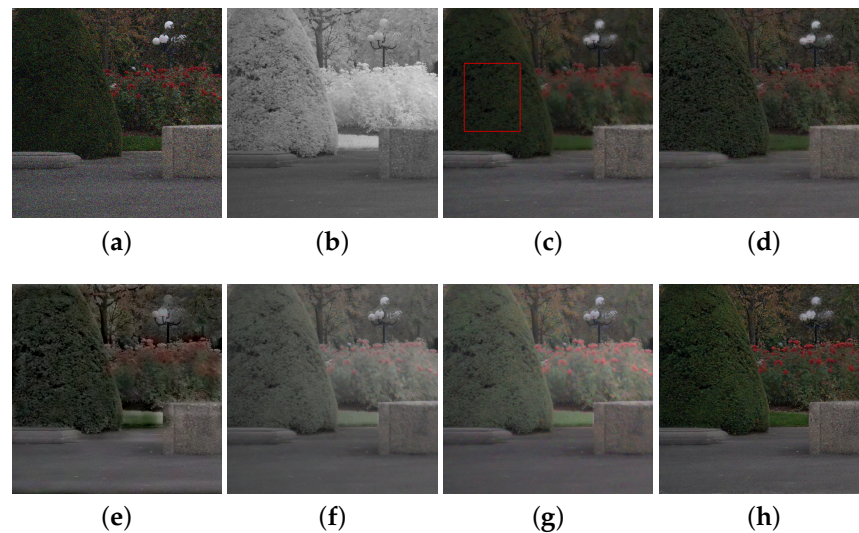


Figure 12. Experimental results for a test image. (a) Noisy RGB image. (b) NIR image. (c) DWLS [11]. (d) SM [8]. (e) GWS [1]. (f) DenseFuse [17]. (g) UDIF [20]. (h) Proposed method.

Moreover, we performed the quantitative measurements in terms of the discrete entropy (DE) [32], feature-based blind image quality evaluator (FBIQE) [33], and color image quality (CIQ) [34]. DE [32] is defined as

$$H(p) = - \sum_{i=0}^{L-1} p(i) \log_2(p(i)) \quad (28)$$

where $p(i)$ is the probability density function at the intensity level i and L is maximum pixel value ($L = 255$). DE estimates the detail amount of an image based on the histogram distribution. FBIQE measures the modified Bhattacharyya distance between the natural statistics of distorted images and the reference naturalness statistics in terms of the local structures, contrast, multiscale decomposition, and colors [33]. FBIQE is calculated as

$$q = \sqrt{(\mu - \mu')^T \left(\frac{\sigma + \sigma'}{2} (\mu - \mu') \right)} \quad (29)$$

where (μ, Σ) and (μ', Σ') represent the mean vector and variance matrix of natural statistics features from the test and reference images. The natural statistics features are modeled using multivariate Gaussian (MVG) model with (μ, Σ) as follows:

$$f(x) = \frac{1}{(2\pi)^{m/2} \|\sigma\|^{-1}} \exp\left(-\frac{1}{2}(x - \mu)^T \sigma^{-1} (x - \mu)\right) \quad (30)$$

where m is the dimension of the vector μ . The color image quality (CIQ) [34] metric assesses the image quality in terms of sharpness, colorfulness, and contrast characteristics. In the CIQ metric, the sharpness, colorfulness, and contrast measurement are calculated first and then combined with multiple linear regression (MLR) as follows:

$$CIQ = c_1 \times \text{colorfulness} + c_2 \times \text{sharpness} + c_3 \times \text{contrast} \quad (31)$$

where c_1 , c_2 , and c_3 are constants [34].

A larger DE indicates more details in the fused image, while a smaller FQBIE indicates good image quality with less noise, a clearer structure, and more natural colors in the fused images. Larger CIQ values indicate better image quality in the results with high contrast, sharp structure, and saturated colors. Table 1 shows the quantitative measurements on five methods.

The proposed method achieved the best performance in the average DE because the gradient denoising term performs gradient enhancement in a wavelet scale map guided by the gradients of the NIR wavelet coefficients. The detail-enhanced wavelet scale map provides detail enhancement of the fused RGB images. In addition, the proposed method achieved the best performance in FBQIE because the proposed method produced a fused RGB image with less noise and better details. Due to the good color enhancement and gradient enhancement in the proposed method, the fusion results achieved the best performance in the CIQ metric.

Table 1. Performance comparison between GWS [1], DWLS [11], SM [8], DenseFuse [17], UDIF [20], and the proposed method.

Metrics	DWLS	SM	GWS	Dense.	UDIF	Pro.
DE	7.082	7.046	7.030	6.841	6.865	7.128
FBQIE	29.639	27.269	27.909	29.600	30.469	26.836
CIQ	0.912	0.914	0.904	0.788	0.841	0.961

Bold numbers represent the best performance in each metric.

3.2. Parameter Analysis

We provide the effects of key parameters on the fusion results by the proposed method as follows. First, we perform the experiments to evaluate the effects of σ_1 and γ_1 (in the contrast preservation term) on the fusion performance as shown in Figure 13. We observed that the fusion results contained few NIR data with the increase of σ_1 and γ_1 . This is because increasing σ_1 and γ_1 converts most of the visibility ratio value (see (14)) into 0 as shown in Figure 13c. Thus, few NIR data are transferred to the fusion result. We assess the effects of λ and β (in the gradient denoising term) on the fusion result as shown in Figure 14.

From the figures, increasing β leads to the blur effects (see Figure 14a,b) because a large β reduces the gradient magnitude. By introducing the visibility ratio into λ , the fusion result contains sharp details (see Figure 14b,c) because the visibility ratio adaptively selects the NIR texture transfer and gradient denoising. A low visibility ratio enforces the gradient enhancement in gradient denoising terms. Finally, we provide experiments on the effect of η in the denoising term for the fusion result as shown in Figure 15. We concluded that the results with large η distorted the image details (see the red blocks) because a large η set the small wavelet coefficients of the fusion result to 0 in (23).

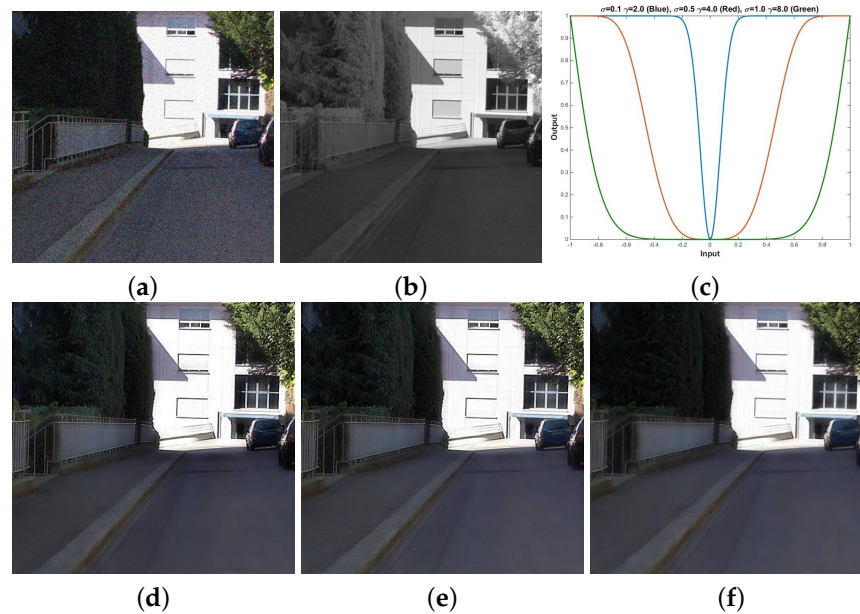


Figure 13. Fusion results from the MFD framework with different σ_1 and γ_1 pairs in the contrast preservation term. (a) Noisy RGB image. (b) NIR image. (c) Transfer function (blue, red, and green curves corresponding to (d–f)). (d) $\sigma_1 = 0.1$ and $\gamma_1 = 2.0$. (e) $\sigma_1 = 0.5$ and $\gamma_1 = 4.0$. (f) $\sigma_1 = 1.0$ and $\gamma_1 = 8.0$.

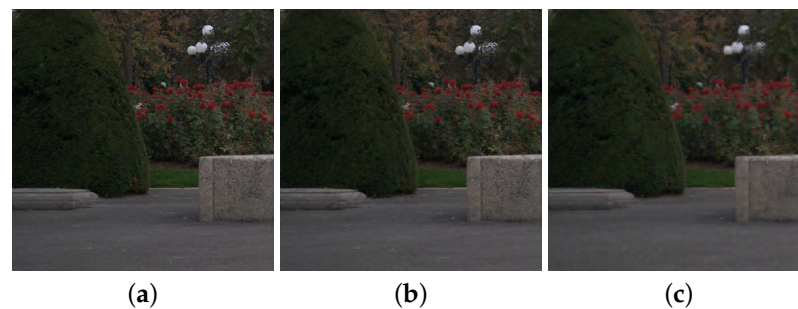


Figure 14. Fusion results from the MFD framework with different λ and β in the gradient denoising term for wavelet scale map estimation. (a) $\beta = 1.2$ and λ with visibility ratio; (b) $\beta = 2.0$ and λ with visibility ratio. (c) $\beta = 2.0$ and λ without visibility ratio. The visibility ratio means $V_{I_{NIR}}/V_{I_{RGB}}$ in (14).

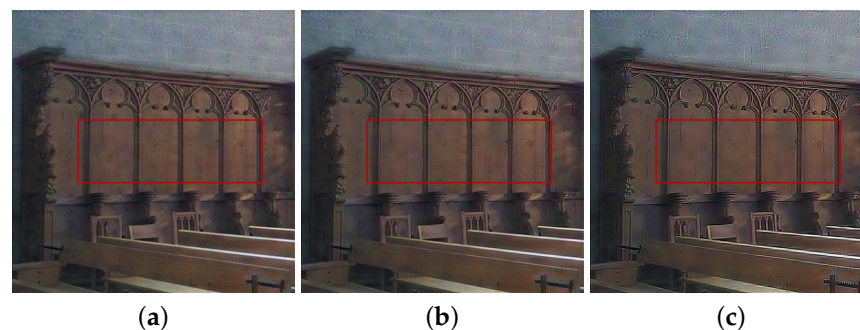


Figure 15. Fusion results from the MFD framework with different η in the denoising term for fused RGB image. (a) $\eta = 0.001$. (b) $\eta = 0.005$. (c) $\eta = 0.01$.

3.3. Comparison between Different Wavelet Transforms

We implement the proposed method on DWT and DT-CWT decompositions based on the visibility measurements provided by [24,35]. Figure 16 shows the fusion results

from the proposed method under DWT and DTCWT. From the figures, the results from DT-CWT contain less artifacts and sharper details compared with DWT (see the red boxes in Figure 16). Thus, we implement the proposed method under DT-CWT.

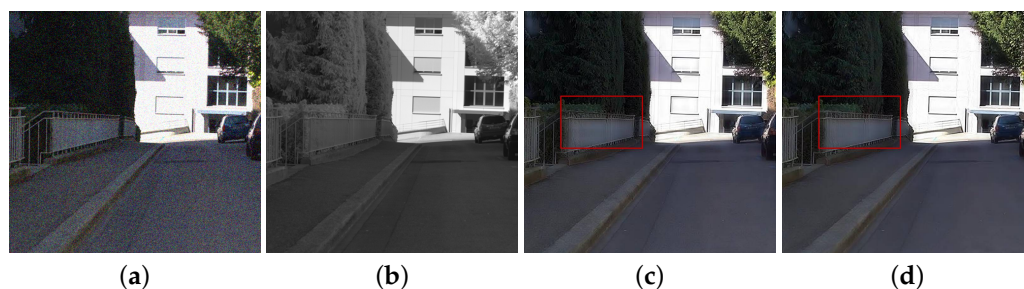


Figure 16. Fusion results from the MFD framework under DWT and DTCWT. (a) Noisy RGB image. (b) NIR image. (c) Fusion under DWT. (d) Fusion under DT-CWT.

3.4. Comparison with Different Color Spaces

We tested the proposed method on the YCbCr, HSV, and CIELAB color spaces. We used fifteen test images for the experiments in Figure 8. Figure 17 shows the visual comparison among three color spaces by the proposed method. The proposed method in YCbCr color space generated the results with the least color distortion (see the red blocks in Figure 17d,e and the greenish color artifacts in Figure 17c). Table 2 shows the performance comparison among the HSV, CIE LAB, and YCbCr color spaces with the proposed method. The proposed method achieved the best performance in the quantitative measurements. Thus, we adopt the YCbCr color space for the proposed method.

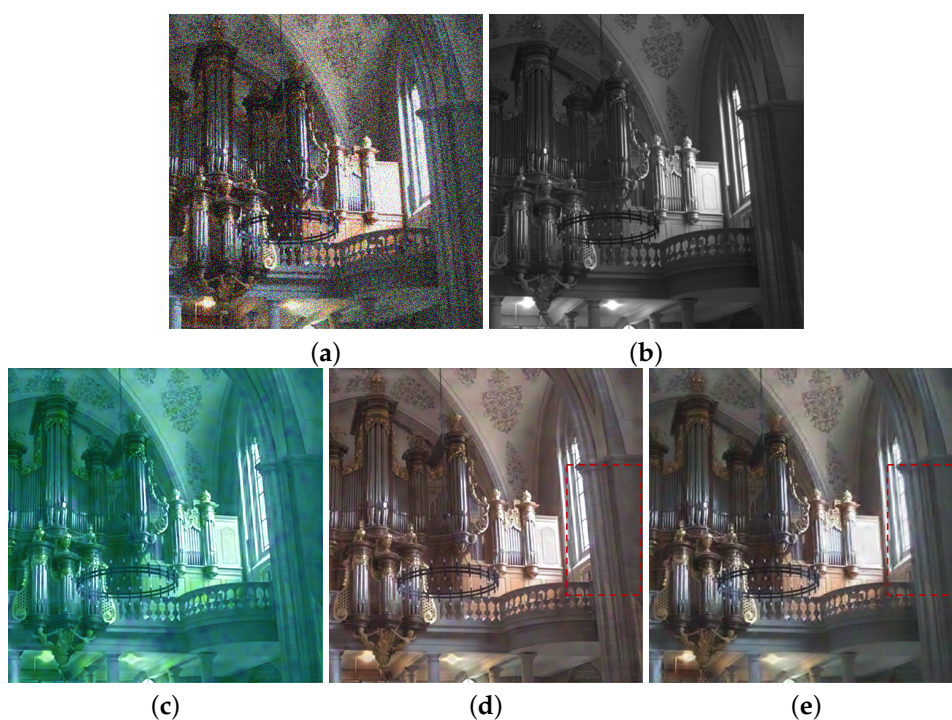


Figure 17. Fusion results from the proposed method on different color spaces. (a) Noisy RGB image; (b) NIR image; (c) Fusion result on HSV; (d) Fusion result on CIE LAB; (e) Fusion result on YCbCr.

Table 2. Performance comparison among the HSV, CIE LAB, and YCbCr color spaces with the proposed method.

Metrics	HSV	CIE LAB	YCbCr
DE	7.015	7.032	7.128
FBIQE	28.057	29.747	26.836
CIQ	0.764	0.896	0.961

Bold numbers represent the best performance in each metric.

3.5. Computational Complexity

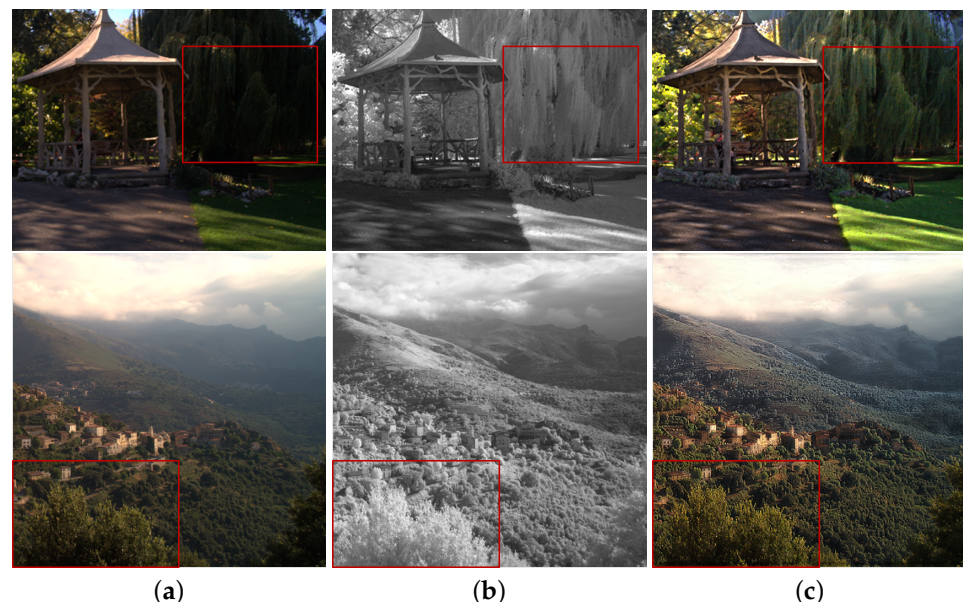
In this section, we provide the runtime of the proposed method and the compared methods on Table 3 (image size: 512×512). The average processing time of the proposed method was 14.94 s/image, which was the longest times among the methods. In the future, we will consider improving the speed of the proposed method by optimization in (18).

Table 3. Computation time of the six compared methods (image size: 512×512).

Methods	DWLS	SM	GWS	Dense.	UDIF	Pro.
Time(s/image)	3.42	7.99	0.60	0.40	0.29	14.94

3.6. Application to RGB-NIR Images under Normal Illumination

We applied the proposed method to the paired images captured under a normal illumination condition. As there is little noise in RGB images, we remove the chroma denoising in the proposed method and the denoising term in (23). (23) is rewritten to (25). However, the gradient denoising term is kept because it provides detail enhancement from the visible NIR components in the fusion results. As shown in Figure 18, the proposed method generates the enhanced results with a clearer structure and finer details (see the red blocks in Figure 18).

**Figure 18.** Experimental results for the paired images under normal illumination conditions. (a) RGB images; (b) NIR images; (c) Fusion results.

3.7. Fusion of RGB Luminance Channel and NIR Image in a Local Manner

We provide local scale map estimation and fusion on the RGB luminance channel and NIR image. Chroma denoising and color enhancement are then used to obtain fused color images. In the local scale map estimation and fusion stage, we first divide the RGB

luminance channel and NIR image into overlapping blocks. Then, we apply the proposed method to each block pair and estimate the local scale map. We obtain each fusion block y'_p corresponding to the local scale map. Finally, we combine all overlapping blocks with gaussian kernels as follows:

$$Y'_p = \frac{\sum_x \omega_p(x) \cdot y'_p(x)}{\sum_x \omega_p(x)} \quad (32)$$

where Y'_p is the global luminance at the pixel location p and $y'_p(x)$ is the local fused luminance in a $m \times m$ window at the center pixel p ; m is the window size; $\omega_p(x)$ is the weight function, which is inversely proportional to the l_2 norm distance between the center pixel p and the neighbor pixel x , which is formulated as follows:

$$\omega_p(x) = \exp\left(-\frac{(x-p)^2}{\gamma}\right) \quad (33)$$

where γ is the Gaussian parameter, which is set to 3. The relationship between the minimum size of window m and the maximum decomposition level M of wavelet is $m = 2^{(M+1)}$. This is because the image patches should be decomposed fully (i.e., the size of last decomposed image should be larger than or equal to 2×2). In this work, we set the maximum decomposition level to 3–5. Thus, the minimum window size corresponds to 16–64. Figure 19 shows the experimental results from the proposed method in a local manner. The proposed method worked on the test images in a local manner and achieved good performance.

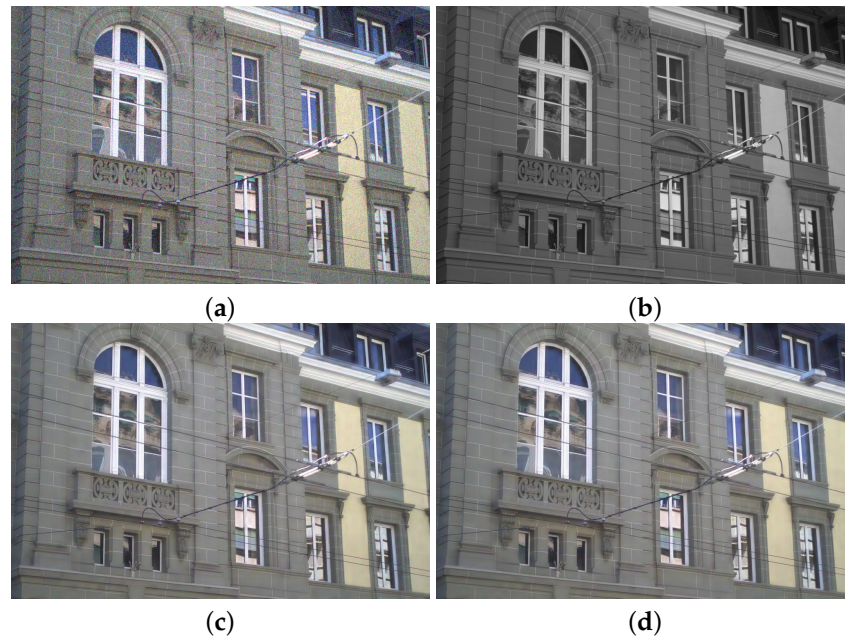


Figure 19. Experimental results from the proposed method in a local manner (the maximum decomposition level is 4). (a) RGB image; (b) NIR image; (c) Fusion result using a 64×64 local window; (d) Fusion result using a 128×128 local window.

3.8. Application to RGB-NIR Images with JPEG Compression

We applied the proposed method to the test images with JPEG compression. We performed JPEG compression on the test images using the *imwrite* function in MATLAB. In the *imwrite* function, we set the compression degree to 0.25, 0.5, and 0.75. In the previous manuscript, the lossless test images had the png format. Figure 20 shows the visual quality of the proposed method on the compressed test images. As shown in Figure 20, in a high compression degree (i.e., 0.25), the ringing artifacts appeared in the NIR images, and noisy blocking artifacts appeared in the RGB images (see Figure 20a,b).

The fusion results contained ringing artifacts (see Figure 20e,f). However, in a low

compression degree (i.e., 0.75), the proposed method achieved the same enhancement results on the compressed images as the lossless png format images (see Figure 20g,h). Table 4 shows the performance of the proposed methods on the compressed images with different compression degrees.

From the FBIQE and CIQ metrics, the results were corrupted under the high JPEG compression degree and had good quality under the low JPEG compression degree. This is because ringing artifacts decrease the performance of naturalness and contrast measurements in FBIQE and CIQ metrics. The DE values did not change significantly and were even better under the high compression degree because ringing artifacts appear in the fusion results (see Figure 20e,f), and the DE metric considers the artifacts as the amount of details.

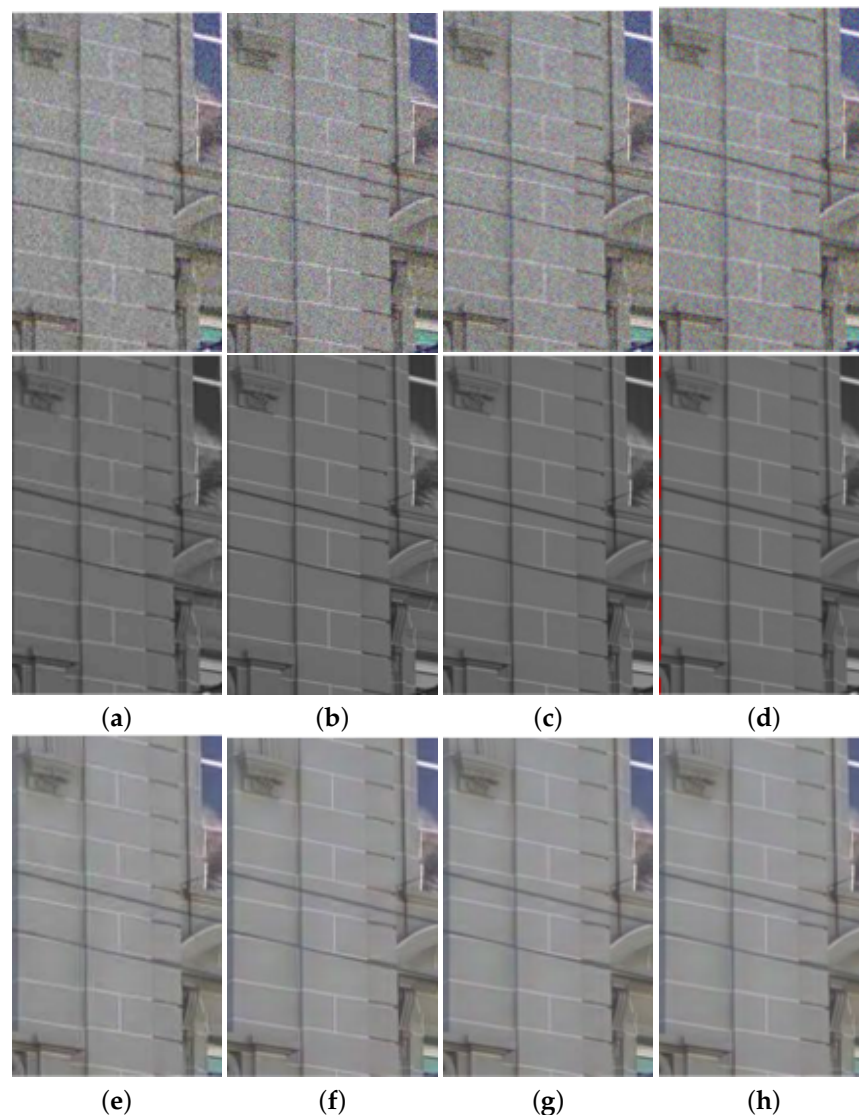


Figure 20. Fusion results from the proposed method on the compressed test images with different compression degrees: (a) 0.25; (b) 0.5; (c) 0.75; (d) png format; (e) Fusion results for (a); (f) Fusion results for (b); (g) Fusion results for (c); (h) Fusion results for (d). (From (a) to (d), **top**: RGB images; **bottom**: NIR images).

Table 4. Performance comparison of the proposed method on test images in terms of the DE, FBIQE, and CIQ metrics.

Metrics	0.25	0.5	0.75	png
DE	7.137	7.134	7.132	7.128
FBIQE	29.283	27.486	26.895	26.836
CIQ	0.803	0.826	0.934	0.961

Bold numbers represent the best performance in each metric.

3.9. Limitation and Future Work

There are two limitations of the proposed MFD framework. First, we did not consider the calibration problem for RGB and NIR image fusion. The proposed method generates fusion results with artificial edges when the data are not calibrated well. One solution is to combine the affine transform into the proposed MFD method to solve the registration problem [36]. Second, the proposed method does not work well on NIR regions with weak structure.

When NIR images contain weak structure and RGB images are seriously corrupted by noise, the weak gradients of the NIR regions smooth the fusion results, and the proposed method attenuates the details in the fusion regions (see the red blocks of grass in Figure 21). The adaptive selection of static denoising (NIR guidance denoising) and dynamic denoising (RGB self-denoising) is one solution to overcome this problem [37].

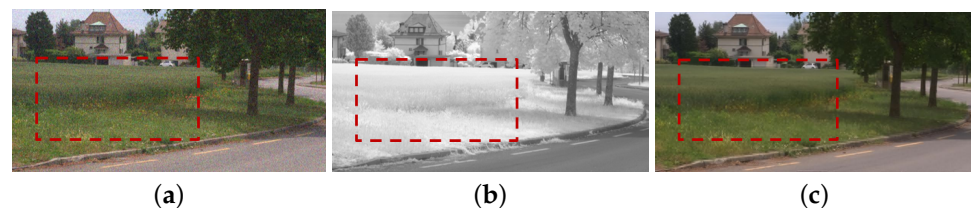


Figure 21. Experimental results on test paired images with weak structure in the NIR images. (a) RGB image; (b) NIR image; (c) Fusion result.

4. Conclusions

In this paper, we proposed MFD of RGB and NIR images based on multi-scale wavelet analysis. We conducted MFD by MAP estimation in the wavelet domain. In the luminance channel, we provided a discrepancy model to deal with the discrepancy between the RGB and NIR images. The discrepancy was obtained by the correlation between the two types of data. We used the priors of the wavelet scale map and its gradient as the contrast preservation term and gradient denoising term, respectively. Then, we adjusted the NIR image based on the scale map to fuse it with the RGB image.

The prior of the fusion wavelet coefficients was modeled as the Laplacian distribution with an adaptive scaling parameter based on the adjusted NIR image to reduce noise. In the chrominance channels, we used the guided filter to denoise the noise with the guidance of the fused luminance. Finally, we employed color enhancement based on the variation of the luminance after the fusion process. Our experimental results demonstrated that the proposed method achieved excellent fusion performance with clear structure and good details.

Author Contributions: Conceptualization, H.S. and C.J.; methodology, H.S. and C.J.; software, H.S.; validation, H.S. and L.Y.; formal analysis, H.S. and C.J.; investigation, H.S., C.J., and L.Y.; resources, H.S. and L.Y.; data curation, H.S. and L.Y.; writing—original draft preparation, H.S.; writing—review and editing, H.S. and C.J.; visualization, H.S. and L.Y.; supervision, C.J.; project administration, C.J.; funding acquisition, C.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (No. 61872280), the China Postdoctoral Science Foundation (No. 2019M663929XB) and the Fundamental Research Funds for the Central Universities (No. XJS200214).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found here: <https://sites.google.com/site/changhwan76son/home/project-icp2015/>, <https://sites.google.com/site/changhwan76son/home/multi-modal-image-pair-fusion/> as well as <https://www.epfl.ch/labs/ivrl/research/downloads/labouring-the-near-infrared/> (accessed on 22 May 2021).

Acknowledgments: The authors would like to thank the editors and the anonymous reviewers for their comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Deng, G. Guided Wavelet Shrinkage for Edge-Aware Smoothing. *IEEE Trans. Image Process.* **2017**, *26*, 900–914. [CrossRef] [PubMed]
2. Shen, Y.; Liu, Q.; Lou, S.; Hou, Y.L. Wavelet-Based Total Variation and Nonlocal Similarity Model for Image Denoising. *IEEE Signal Process. Lett.* **2017**, *24*, 877–881. [CrossRef]
3. Ding, Y.; Selesnick, I.W. Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization. *IEEE Signal Process. Lett.* **2015**, *22*, 1364–1368. [CrossRef]
4. Yamashita, H.; Sugimura, D.; Hamamoto, T. Enhancing low-light color images using an RGB-NIR single sensor. In Proceedings of the IEEE International Conference on Visual Communications and Image Processing (VCIP), Singapore, 13–16 December 2015; pp. 1–4.
5. Yamashita, H.; Sugimura, D.; Hamamoto, T. RGB-NIR imaging with exposure bracketing for joint denoising and deblurring of low-light color images. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 6055–6059.
6. Son, C.H.; Zhang, X.P. Near-Infrared Fusion via Color Regularization for Haze and Color Distortion Removals. *IEEE Trans. Circuits Syst. Video Technol.* **2017**, *28*, 3111–3126. [CrossRef]
7. Son, C.H.; Zhang, X.P. Near-Infrared Coloring via a Contrast-Preserving Mapping Model. *IEEE Trans. Image Process.* **2017**, *26*, 5381–5394. [CrossRef] [PubMed]
8. Shen, X.; Yan, Q.; Xu, L.; Ma, L.; Jia, J. Multispectral joint image restoration via optimizing a scale map. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 2518–2530. [CrossRef]
9. Krishnan, D.; Fergus, R. Dark Flash Photography. *ACM Trans. Graph.* **2009**, *28*, 96:1–96:11. [CrossRef]
10. Fredembach, C.; Süsstrunk, S. Colouring the near-infrared. In Proceedings of the Color and Imaging Conference. Society for Imaging Science and Technology, Portland, OR, USA, 10–14 November 2008; Volume 2008, pp. 176–182.
11. Zhuo, S.; Zhang, X.; Miao, X.; Sim, T. Enhancing low light images using near infrared flash images. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Hong Kong, China, 26–29 September 2010; pp. 2537–2540.
12. Li, W.; Zhang, J.; Dai, Q.H. Robust blind motion deblurring using near-infrared flash image. *J. Vis. Commun. Image Represent.* **2013**, *24*, 1394–1413. [CrossRef]
13. Yan, Q.; Shen, X.; Xu, L.; Zhuo, S.; Zhang, X.; Shen, L.; Jia, J. Cross-field joint image restoration via scale map. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, NSW, Australia, 1–8 December 2013; pp. 1537–1544.
14. Sugimura, D.; Mikami, T.; Yamashita, H.; Hamamoto, T. Enhancing color images of extremely low light scenes based on RGB/NIR images acquisition with different exposure times. *IEEE Trans. Image Process.* **2015**, *24*, 3586–3597. [CrossRef]
15. Son, C.H.; Zhang, X.P. Layer-based approach for image pair fusion. *IEEE Trans. Image Process.* **2016**, *25*, 2866–2881. [CrossRef]
16. Shibata, T.; Tanaka, M.; Okutomi, M. Versatile visible and near-infrared image fusion based on high visibility area selection. *J. Electron. Imaging* **2016**, *25*, 013016–013016. [CrossRef]
17. Li, H.; Wu, X. DenseFuse: A Fusion Approach to Infrared and Visible Images. *IEEE Trans. Image Process.* **2019**, *28*, 2614–2623. [CrossRef]
18. Zhang, L.; Zeng, G.; Wei, J.; Xuan, Z. Multi-Modality Image Fusion in Adaptive-Parameters SPCNN Based on Inherent Characteristics of Image. *IEEE Sens. J.* **2020**, *20*, 11820–11827. [CrossRef]
19. Zheng, M.; Qi, G.; Zhu, Z.; Li, Y.; Wei, H.; Liu, Y. Image Dehazing by an Artificial Image Fusion Method Based on Adaptive Structure Decomposition. *IEEE Sens. J.* **2020**, *20*, 8062–8072. [CrossRef]
20. Jung, H.; Kim, Y.; Jang, H.; Ha, N.; Sohn, K. Unsupervised Deep Image Fusion With Structure Tensor Representations. *IEEE Trans. Image Process.* **2020**, *29*, 3845–3858. [CrossRef]
21. Su, H.; Jung, C. Multi-spectral Fusion and Denoising of RGB and NIR Images Using Multi-scale Wavelet Analysis. In Proceedings of the 2018 24th IAPR International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
22. Bioucas-Dias, J.M. Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Trans. Image Process.* **2006**, *15*, 937–951. [CrossRef]

23. Bhatnagar, G.; Wu, Q.J.; Liu, Z. Directive contrast based multimodal medical image fusion in NSCT domain. *IEEE Trans. Multimed.* **2013**, *15*, 1014–1024. [[CrossRef](#)]
24. Hill, P.; Al-Mualla, M.E.; Bull, D. Perceptual Image Fusion Using Wavelets. *IEEE Trans. Image Process.* **2017**, *26*, 1076–1088. [[CrossRef](#)]
25. Xu, L.; Yan, Q.; Xia, Y.; Jia, J. Structure Extraction from Texture via Natural Variation Measure. *ACM Trans. Graph. (SIGGRAPH Asia)* **2012**, *31*. [[CrossRef](#)]
26. Farbman, Z.; Fattal, R.; Lischinski, D.; Szeliski, R. Edge-preserving decompositions for multi-scale tone and detail manipulation. *ACM Trans. Graph. (TOG)*. **2008**, *27*, 67. [[CrossRef](#)]
27. Wang, W.; Shui, P.; Feng, X. Variational Models for Fusion and Denoising of Multifocus Images. *IEEE Signal Process. Lett.* **2008**, *15*, 65–68. [[CrossRef](#)]
28. He, K.; Sun, J.; Tang, X. Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1397–1409. [[CrossRef](#)] [[PubMed](#)]
29. Dinh, K.Q.; Canh, T.N.; Jeon, B. Color image denoising via cross-channel texture transferring. *IEEE Signal Process. Lett.* **2016**, *23*, 1071–1075. [[CrossRef](#)]
30. Mantiuk, R.; Mantiuk, R.; Tomaszewska, A.; Heidrich, W. Color Correction for Tone Mapping. *Comput. Graph. Forum (Proc. Eurographics 2009)* **2009**, *28*, 193–202. [[CrossRef](#)]
31. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [[CrossRef](#)]
32. Shannon, C.E. A mathematical theory of communication. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2001**, *5*, 3–55. [[CrossRef](#)]
33. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [[CrossRef](#)]
34. Panetta, K.; Gao, C.; Aгаian, S. No reference color image contrast and quality measures. *IEEE Trans. Consum. Electron.* **2013**, *59*, 643–651. [[CrossRef](#)]
35. Zhen Liu.; Karam, L.J.; Watson, A.B. JPEG2000 encoding with perceptual distortion control. *IEEE Trans. Image Process.* **2006**, *15*, 1763–1778. [[CrossRef](#)]
36. Ma, J.; Chen, C.; Li, C.; Huang, J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [[CrossRef](#)]
37. Ham, B.; Cho, M.; Ponce, J. Robust Guided Image Filtering Using Nonconvex Potentials. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 192–207. [[CrossRef](#)] [[PubMed](#)]