



## Region-aware RGB and near-infrared image fusion

Jiacheng Ying<sup>a</sup>, Can Tong<sup>a</sup>, Zehua Sheng<sup>a</sup>, Bowen Yao<sup>a</sup>, Si-Yuan Cao<sup>b,a</sup>, Heng Yu<sup>d</sup>,  
Hui-Liang Shen<sup>a,b,c,\*</sup>

<sup>a</sup> College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

<sup>b</sup> Ningbo Innovation Center, Zhejiang University, Ningbo 315100, China

<sup>c</sup> Key Laboratory of Collaborative Sensing and Autonomous Unmanned Systems of Zhejiang Province, Hangzhou 310015, China

<sup>d</sup> School of Computer Science, University of Nottingham Ningbo China, Ningbo 315199, China

### ARTICLE INFO

#### Article history:

Received 15 November 2022

Revised 24 April 2023

Accepted 26 May 2023

Available online 26 May 2023

#### Keywords:

Image fusion

RGB and near-infrared

overexposed sky recovery

vegetation enhancement

gradient-domain optimization

### ABSTRACT

This paper proposes a region-aware fusion method, called RaIF, for RGB and near-infrared (NIR) outdoor scenery image fusion. The method is motivated by the observation that current fusion approaches produce gray appearance in overexposed sky regions and distortion in vegetation regions. RaIF generates the region probability maps by exploiting their specific characteristics in the visible and NIR spectra. It recovers the overexposed sky regions by employing the intrinsic channel correlation between RGB and NIR images, and enhances the vegetation regions in an adjustable manner. RaIF formulates image fusion problem as a gradient-domain optimization problem with luminance and chromaticity regularizations. Experimental results validate the superiority of RaIF that produces fused images with improved appearance in the sky and vegetation regions, and achieves the state-of-the-art performance quantitatively and qualitatively. Furthermore, RaIF can act as a refinement module that improves the fusion results of current deep learning based approaches. It is also capable of recovering specular highlight regions other than sky overexposure.

© 2023 Elsevier Ltd. All rights reserved.

### 1. Introduction

Image fusion integrates information of different images with complementary characteristics and thus has a wide range of applications in visual enhancement [1,2], remote sensing [3], image segmentation [4], image classification [5,6], etc. This work focuses on image fusion of outdoor scenery images acquired in the visible and near-infrared (NIR) spectra for the purpose of image enhancement.

RGB and NIR images have distinct characteristics because of the radiation variation of scenes in the visible and NIR spectra, respectively. Hence, image enhancement can be achieved by properly fusing RGB and NIR images. Most of the current image fusion approaches aim to improve the visibility of image textures using either traditional or deep learning techniques [7,8]. Those approaches make full use of the structural information of the source images and generate detail-enhanced fusion results. However, some practical issues have not been well addressed up to

date. For example, color loss caused by sky overexposure is a common phenomenon in RGB images. The current fusion approaches only fuse the luminance of the source images and adopt the original color of RGB images. As a result, the fused images may lack color and appear fake in the overexposed regions. Furthermore, the vegetation regions seem “glowing” in NIR images, while they appear darker in RGB images because vegetation reflects more NIR radiations. The current approaches tend to assign larger fusion weights to the NIR image, and consequently the fused image may exhibit brightness and color distortions in the vegetation regions.

This work proposes a region-aware RGB and NIR image fusion method, called RaIF, to tackle the mentioned issues. The method first detects the sky and vegetation regions and generate their corresponding probability maps by exploiting the regions' specific characteristics. To deal with the sky overexposure, RaIF predicts the color and texture of the overexposed sky using a gain-gamma model based on the channel correlation between the RGB and NIR images. To improve the visual fidelity of the vegetation appearance, RaIF enhances the brightness of the vegetation regions in an adjustable manner according to the probability map. Finally, we solve the RaIF problem in the gradient domain with necessary luminance and chromaticity constraints. An example of RGB and NIR image fusion is illustrated in Fig. 1. It is observed that, compared with DenseFuse [8], the proposed RaIF method can re-

\* Corresponding author.

E-mail addresses: [yingjiacheng@zju.edu.cn](mailto:yingjiacheng@zju.edu.cn) (J. Ying), [21931021@zju.edu.cn](mailto:21931021@zju.edu.cn) (C. Tong), [shengzehua@zju.edu.cn](mailto:shengzehua@zju.edu.cn) (Z. Sheng), [bwyao36@zju.edu.cn](mailto:bwyao36@zju.edu.cn) (B. Yao), [karlcao@hotmail.com](mailto:karlcao@hotmail.com) (S.-Y. Cao), [heng.yu@nottingham.edu.cn](mailto:heng.yu@nottingham.edu.cn) (H. Yu), [shenhl@zju.edu.cn](mailto:shenhl@zju.edu.cn) (H.-L. Shen).

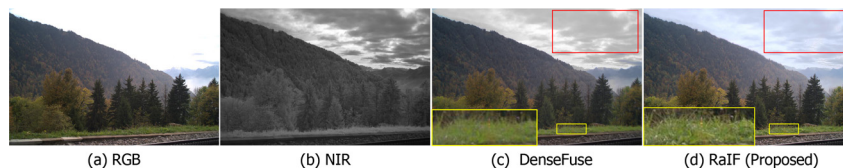


Fig. 1. An example of RGB and NIR image fusion. The red and yellow boxes highlight the fusion differences of DenseFuse [8] and the proposed RaIF method in the overexposed sky regions and vegetation regions, respectively.

cover the color and texture of the sky regions and improve the visibility of the vegetation regions with less distortion. Furthermore, RaIF can act as a module to refine the fusion results of the current deep learning-based image fusion approaches, and can also recover specular highlight regions. In summary, the main contributions are as follows:

- This work proposes a region-aware RGB and NIR image fusion method, called RaIF, to enhance the outdoor scenery images, especially in the sky and vegetation regions.
- The method generates the region probability maps by exploring the specific characteristics in the RGB and NIR images, with which the color and texture of the sky regions can be reliably generated and the vegetation regions can be precisely enhanced in the fused image.
- The method formulates the region-aware image fusion as a gradient-domain optimization problem with additional luminance and chromaticity regularizations under the guidance of the region probability maps.

## 2. Related work

We briefly review the RGB and NIR image fusion approaches, which are approximately divided into four categories, *i.e.*, multi-scale fusion, saliency-based fusion, gradient fusion, and deep learning fusion.

Multi-scale fusion decomposes the RGB and NIR images into multi-scale components and fuses the two source images in multiple scales. Two-scale decomposition is a common way to extract the detail and base layers of the source images for detail-preserving fusion [9]. Multi-scale decomposition leverages the multi-frequency components of the source images to improve the naturalness of the fused image by using wavelet transform [10] or Laplacian pyramid [7].

Saliency-based fusion enhances image contrast and texture using RGB and NIR saliency maps. FIE [11] computes saliency fusion maps by extracting the local contrast of the NIR and RGB images to produce detail-enhanced RGB images. WLS [12] designs visual saliency maps using the sum of local absolute differences and employs the saliency maps for base layer fusion.

Gradient fusion methods fuse two source images in the gradient domain and reconstruct the fused image using gradient information. GTF [13] converts the gradient of the visible image to the infrared image using total variation minimization. SpE [14] fuses the gradient of the RGB image and a higher-channel image based on the texture tensor, and recovers the final color image using a lookup-table-based reconstruction algorithm [15].

Deep learning-based methods achieve image fusion by constructing multi-input neural networks and training them in a self-supervised or unsupervised manner. DenseFuse [8] employs a dense block in the feature encoder to exploit shallow and deep features of the input images for image fusion. SEDRFuse [16] employs skip connections from the encoders to the decoder to reuse some missing details in the fusion network. CSF [17] pre-trains a pixel-wise classifier to explain the importance of the extracted feature maps and fuses the features using importance maps. STDFusionNet

[18] introduces saliency masks in the loss function to encourage the two encoders to extract background details in the visible image and salient targets in the infrared image, respectively.

Overall, most of the current approaches focus on enhancing the details and contrast, but do not pay much attention to the limitations in fusing the sky and vegetation regions. To the best of our knowledge, the proposed RaIF method is the first one to fuse RGB and NIR images based on the spectral characteristics of different regions, and produces fused images with good appearance.

## 3. Proposed approach

Fig. 2 illustrates the framework of the proposed RaIF method. It consists of five main modules.

- 1) In the region map generation module, feature maps are extracted from the RGB and NIR images to generate the sky and vegetation probability maps.
- 2) In the overexposed sky recovery module, a gain-gamma model is employed to establish the relationship between the NIR and RGB sky pixels.
- 3) In the gradient fusion module, the RGB and NIR gradients are fused using structure tensor.
- 4) In the image reconstruction module, the fused image is reconstructed by solving an optimization problem with gradient constraint, luminance regularization, and chromaticity regularization.
- 5) Finally, the luminance range is compressed and the vegetation map is used to enhance the vegetation regions in an adjustable manner.

More details will be elaborated in the following.

### 3.1. Region map generation

#### 3.1.1. Sky Map Generation

The sky map generation process is based on the following three priors: (a) The sky region is always far away from the shooting position, (b) the sky region is usually smooth, and (c) the sky region is usually at the top of the image. Accordingly, we generate three maps: distance map  $\mathbf{M}_{dis}$ , smoothness map  $\mathbf{M}_{smo}$ , and height map  $\mathbf{M}_{height}$ .

**Distance Map.** According to the haze degradation model, the transmission value denotes the portion of the light that reaches the camera without scattering [19]. It is negatively related to the distance from the shoot scene to the camera. Hence, the distance map  $\mathbf{M}_{dis}$  is computed as

$$\mathbf{M}_{dis} = \mathbf{1} - f_{norm}(\mathbf{M}_{tr}), \quad (1)$$

where  $f_{norm}(\cdot)$  is the normalization operation that linearly scales the range of the map to [0,1],  $\mathbf{M}_{tr}$  is the transmission map computed according to [20].

**Smoothness Map.** The smoothness map reflects the smoothness of each pixel. The degree of pixel smoothness is measured using local entropy

$$E(x, y) = - \sum_{(x', y') \in \mathcal{N}} p(f(x', y')) \log p(f(x', y')), \quad (2)$$

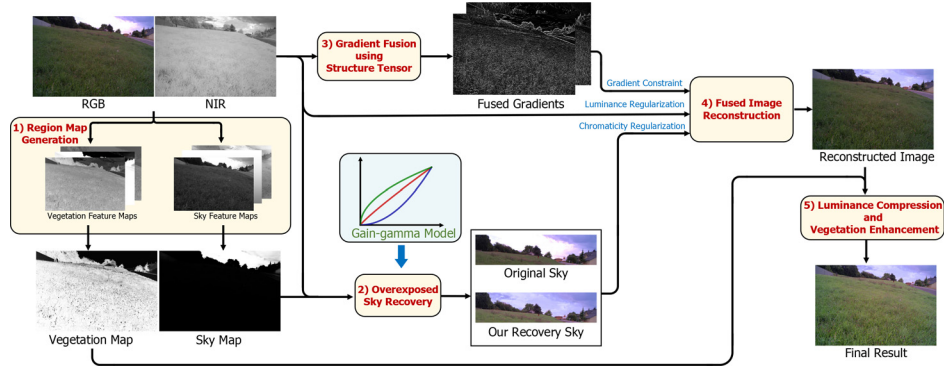


Fig. 2. The framework of the proposed RaIF method.

where  $p(f(x', y'))$  denotes the occurrence probability of the  $f$ -th gray level at pixel position  $(x', y')$ , and  $\mathcal{N}$  denotes the local neighborhood around pixel  $(x, y)$ . We compute the local entropy maps  $\mathbf{E}_{\text{gray}}$  and  $\mathbf{E}_{\text{nir}}$  from the grayscale image  $\mathbf{G}$  (average of the red, green, and blue channel images) and the NIR image  $\mathbf{N}$ . We further refine the maps  $\mathbf{E}_{\text{gray}}$  and  $\mathbf{E}_{\text{nir}}$  using guided filtering  $GF_{r,\epsilon}(\cdot, \cdot)$  with  $\mathbf{G}$  and  $\mathbf{N}$ , respectively, as

$$\mathbf{E}'_{\text{gray}} = f_{\text{norm}}(GF_{r,\epsilon}(\mathbf{E}_{\text{gray}}, \mathbf{G})), \quad (3)$$

$$\mathbf{E}'_{\text{nir}} = f_{\text{norm}}(GF_{r,\epsilon}(\mathbf{E}_{\text{nir}}, \mathbf{N})), \quad (4)$$

where  $r$  denotes the filter radius, and  $\epsilon$  denotes the regularization parameter. Here,  $r = 32$  and  $\epsilon = 1 \times 10^{-6}$ . Then, the smoothness map  $\mathbf{M}_{\text{smo}}$  is computed as

$$\mathbf{M}_{\text{smo}} = f_{\text{norm}}((1 - \mathbf{E}'_{\text{gray}}) \odot (1 - \mathbf{E}'_{\text{nir}})), \quad (5)$$

where  $\odot$  denotes element-wise multiplication.

**Height Map.** Inspired by [21], the height map  $\mathbf{M}_{\text{height}}$  at position  $(x, y)$  is computed as

$$M_{\text{height}}(x, y) = \exp\left(-\left(\frac{y}{1.2H}\right)^2\right), \quad (6)$$

where  $H$  denotes the height of the image.

Finally, the sky probability map is computed using the above three maps as

$$\mathbf{M}_{\text{sky}} = h_{\text{sky}}(\mathbf{M}_{\text{tr}} \odot \mathbf{M}_{\text{smo}} \odot \mathbf{M}_{\text{height}}), \quad (7)$$

where  $h_{\text{sky}}(\cdot)$  denotes a probability mapping function aiming at expanding the probability difference between the sky and non-sky regions. Its form and parameters will be discussed later in this section.

### 3.1.2. Vegetation Map Generation

The vegetation probability map is generated based on the following three priors: (a) The sky region should not be the vegetation region, (b) normalized difference vegetation index (NDVI) [22] is usually high in the vegetation region, and (c) the intensity of the green channel image is usually higher than those of the red and blue channel images in the vegetation region. Accordingly, we generate three maps, *i.e.*, non-sky map  $(1 - \mathbf{M}_{\text{sky}})$ , NDVI map  $\mathbf{M}_{\text{NDVI}}$ , and ratio map  $\mathbf{M}_{\text{ratio}}$ .

**NDVI Map.** The NDVI map  $\mathbf{M}_{\text{NDVI}}$  that reflects vegetation coverage [22] is computed as

$$\mathbf{M}_{\text{NDVI}} = f_{\text{norm}}((\mathbf{N} - \mathbf{R}_r) \oslash (\mathbf{N} + \mathbf{R}_r)), \quad (8)$$

where  $\oslash$  denotes element-wise division, and  $\mathbf{R}_r$  the red channel image.

**Ratio Map.** The ratio map  $\mathbf{M}_{\text{ratio}}$  is computed as

$$\mathbf{M}_{\text{ratio}} = f_{\text{norm}}(\mathbf{R}_g \oslash (\mathbf{R}_r + \mathbf{R}_g + \mathbf{R}_b)), \quad (9)$$

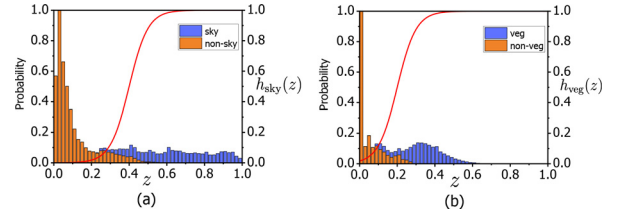


Fig. 3. The stacked histograms of (a) sky maps and (b) vegetation maps in ten scenes before mapping. The red curves in (a) and (b) are the probability mapping functions  $h_{\text{sky}}(\cdot)$  and  $h_{\text{veg}}(\cdot)$ , respectively.

where  $\mathbf{R}_g$  and  $\mathbf{R}_b$  denote the green and blue channel images respectively.

Finally, the vegetation probability map is computed using the three maps as

$$\mathbf{M}_{\text{veg}} = h_{\text{veg}}((1 - \mathbf{M}_{\text{sky}}) \odot \mathbf{M}_{\text{NDVI}} \odot \mathbf{M}_{\text{ratio}}), \quad (10)$$

where  $h_{\text{veg}}(\cdot)$  denotes a probability mapping function to expand the probability difference between the vegetation and non-vegetation regions.

### 3.1.3. Probability Mapping Function

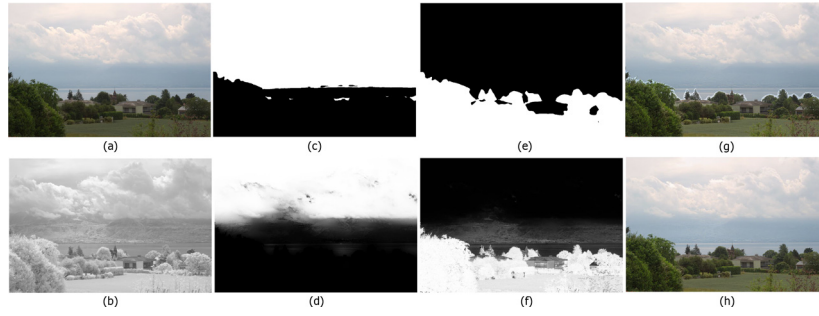
The probability mapping functions  $h_{\text{sky}}(\cdot)$  in (7) and  $h_{\text{veg}}(\cdot)$  in (10) are in the form of sigmoid function,

$$h(z) = 1 - \frac{1}{1 + \exp(-\alpha(\beta - z))}, \quad (11)$$

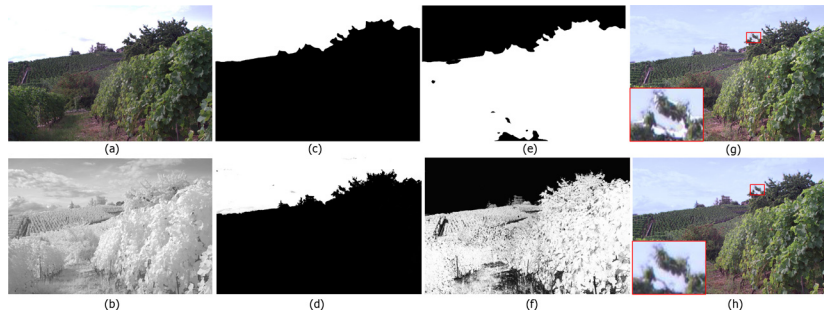
where the parameter  $\alpha$  adjusts the degree of probability,  $\beta$  determines the adjustment threshold. The probability increases when  $z$  is above the threshold  $\beta$  and decreases when below the threshold. Fig. 3 shows the stacked histograms of the sky and vegetation maps in 10 scenes before mapping, respectively. The probability boundary between sky and non-sky pixels is around 0.4 and that between vegetation and non-vegetation pixels is around 0.2. Therefore, we set  $\beta = 0.4$  in  $h_{\text{sky}}(\cdot)$  and  $\beta = 0.2$  in  $h_{\text{veg}}(\cdot)$ . Besides, we empirically set  $\alpha = 20$  in both  $h_{\text{sky}}(\cdot)$  and  $h_{\text{veg}}(\cdot)$ . The red curves in Fig. 3(a) and Fig. 3(b) are the final mapping functions  $h_{\text{sky}}(\cdot)$  and  $h_{\text{veg}}(\cdot)$ , respectively.

### 3.1.4. Region Maps Comparison

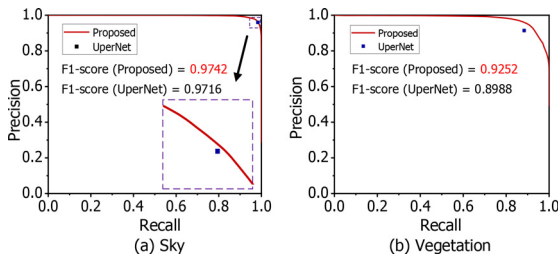
To illustrate the advantages of the proposed region map generation algorithm, we compare it with the semantic segmentation approach UperNet [23]. Fig. 4 and Fig. 5 illustrate the resultant region maps and fused images of the *Lakeside* and *Plants* scenes, respectively. The proposed algorithm produces soft probability maps that are more reliable and flexible for image enhancement. Brightness artifacts are obvious at the sky-vegetation boundary when using the region maps produced by UperNet (see Fig. 4(g) and Fig. 5(g)).



**Fig. 4.** The sky and vegetation region maps produced by UperNet [23] and the proposed algorithm of the *Lakeside* scene, as well as their corresponding fusion results. (a) RGB image. (b) NIR image. (c) Sky map produced by UperNet. (d) Sky map produced by the proposed algorithm. (e) Vegetation map produced by UperNet. (f) Vegetation map produced by the proposed algorithm. (g) Fusion result using the region maps produced by UperNet. (h) Fusion result using the region maps produced by the proposed algorithm.



**Fig. 5.** The sky and vegetation region maps produced by UperNet [23] and the proposed algorithm of the *Plants* scene, as well as the corresponding fusion results using the different maps. Red boxes highlights the details in the fusion results. (a) RGB image. (b) NIR image. (c) Sky map produced by UperNet. (d) Sky map produced by the proposed algorithm. (e) Vegetation map produced by UperNet. (f) Vegetation map produced by the proposed algorithm. (g) Fusion result using the region maps produced by UperNet. (h) Fusion result using the region maps produced by the proposed algorithm.



**Fig. 6.** The precision-recall curves of the proposed probability map generation algorithm and the precision-recall points of UperNet [23], as well as their corresponding F1-scores. (a) Sky region. (b) Vegetation region.

In contrast, the fusion results are fine when using the region maps produced by the proposed algorithm. Furthermore, we manually annotate the sky and vegetation regions of 30 images as ground truths and compare the proposed algorithm with UperNet. Fig. 6 shows the precision-recall curve of the proposed algorithm and the precision-recall point of UperNet, as well as their F1-scores [24]. It is observed that the proposed algorithm obtains higher F1-scores than UperNet in both the sky and vegetation regions, indicating its higher detection performance.

### 3.2. Overexposed sky recovery

In this module, we first divide the sky regions into overexposed and non-overexposed parts by judging whether the maximum channel value of each pixel is equal to 1. We then attempt to recover the overexposed sky regions by mapping the sky pixel intensities from the NIR spectrum to the RGB spectrum with the assistance of the non-overexposed sky regions. According to the imaging model [25], for an object under the same lighting con-

dition, its NIR irradiances should be proportional to the red, green, and blue irradiances. This is validated in Fig. 7(c), which clearly shows a linear relationship between the NIR and red intensities in a captured image pair. However, as shown in Fig. 7(d), the linear relationship will break when the nonlinear camera response function [26] in the image signal processing (ISP) pipeline is enabled during image acquisition. By using inverse response functions, the linear relationship can be established as

$$f_{\text{rgb}}^{-1}(\mathbf{R}_{\text{sky},i}) = k_i \cdot f_{\text{nir}}^{-1}(\mathbf{N}_{\text{sky}}), \quad i \in \{r, g, b\}, \quad (12)$$

where  $k_i$  denotes the irradiance ratio,  $f_{\text{rgb}}^{-1}$  and  $f_{\text{nir}}^{-1}$  denote the inverse response functions applied to the RGB and NIR images, respectively. Previous work have collected a diverse database of real-world camera response functions (DoRF) and introduced empirical model to obtain accurate camera response functions [26]. Since the sky pixel range is usually limited, this work employs a gain-gamma model [27] because it needs fewer parameters and has better adaptability:

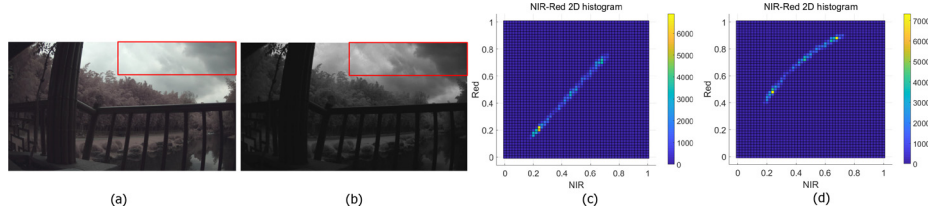
$$(\mathbf{R}_{\text{sky},i})^{\frac{1}{\gamma_1}} = k_i \cdot (\mathbf{N}_{\text{sky}})^{\frac{1}{\gamma_2}}, \quad i \in \{r, g, b\}, \quad (13)$$

where  $\gamma_1$  and  $\gamma_2$  are gamma parameters. Equation (13) can be simplified to

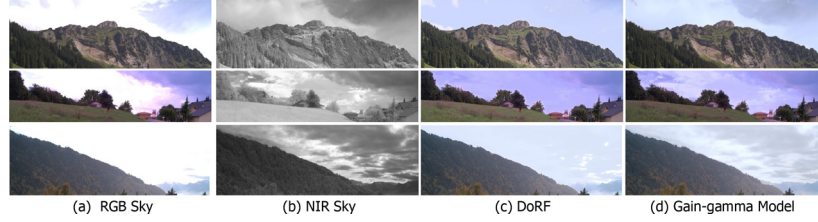
$$\mathbf{R}_{\text{sky},i} = a_i \cdot \mathbf{N}_{\text{sky}}^{\gamma}, \quad i \in \{r, g, b\}, \quad (14)$$

with gain  $a_i = k_i^{\gamma_1}$  and gamma  $\gamma = \frac{\gamma_1}{\gamma_2}$ . To compute the parameters, we sample  $\gamma$  in the range [0.5,2.2] with an interval 0.1, and solve  $a_i$  using weighted least-squares. The weight of each pixel is its probability belonging to the sky region. The parameters with the lowest fitting error are regarded as the final solution.

Fig. 8 illustrates the sky recovery results of three example images using DoRF [26] and the gain-gamma model. It is observed that the recovered sky using DoRF has less textures because the



**Fig. 7.** The histogram of NIR-red pixels of the sky regions in an example image pair. The red boxes indicate the pixels used for histogram plotting. (a) Sample RGB image. (b) Sample NIR image. (c) NIR-red histogram of sampling pixels in directly acquired image pair. (d) NIR-red histogram of sampling pixels when applying tone mapping to the image pair.



**Fig. 8.** Sky recovery results using DoRF [26] and the gain-gamma model of three example sky image pairs. (a) RGB sky. (b) NIR sky. (c) Recovered sky using DoRF. (d) Recovered sky using the gain-gamma model.

functions from DoRF strongly compress the luminance range of the sky regions and thus reduce textures. In comparison, the recovered sky using the gain-gamma model has vivid colors and textures.

It is noted that the above computation assumes the sky regions are partially overexposed. However, in practical photography, the sky regions can be fully overexposed. To discriminate these two cases, this work computes the ratio of the weighted sum of the non-overexposed sky pixels and that of all the sky pixels. If the ratio is below a threshold (e.g., 0.05), the sky regions are judged as fully overexposed. In this case, this work uses the default parameters computed from the NIR-RGB image dataset.

With the estimated parameter  $\hat{a}_i$  and  $\hat{\gamma}$ , the image intensity of the overexposed sky can be recovered as

$$\tilde{\mathbf{R}}_{\text{sky},i} = \hat{a}_i \cdot \mathbf{N}_{\text{sky}}^{\hat{\gamma}}. \quad (15)$$

### 3.3. Gradient fusion

Inspired by the spectral edge (SpE) algorithm [14], this work fuses the NIR and RGB images in the gradient domain and reconstructs the image using the fused gradients. We keep the fused structure tensor equal to that of the RGB-NIR four-channel image, and the fused gradients approximately close to the gradients of the input RGB image. It is noted that SpE [14] cannot obtain the correct structure tensor in the overexposed sky regions with zero RGB gradients. This work can ensure correct gradient fusion by using the RGB gradients recovered according to (15).

### 3.4. Fused image reconstruction

This work solves the gradient domain reconstruction problem by optimization to obtain a desired image fusion result.

We denote by  $\mathbf{P}_{i,x}$  and  $\mathbf{P}_{i,y}$  the  $i$ -th channel of the fused gradient field in the horizontal and vertical directions, respectively, and denote by  $\mathbf{X}_i$  the  $i$ -th channel of the color image to be reconstructed. Then  $\mathbf{X}$  is computed from the gradient domain as

$$\mathbf{X} = \arg \min_{\mathbf{X}} \Theta(\mathbf{X}) + \lambda_1 \Psi(\mathbf{X}) + \lambda_2 \Phi(\mathbf{X}), \quad (16)$$

where  $\Theta(\mathbf{X})$  denotes the gradient data term,  $\Psi(\mathbf{X})$  and  $\Phi(\mathbf{X})$  the luminance and chromaticity regularization terms, respectively. The two balance parameters  $\lambda_1$  and  $\lambda_2$  are both set to be 1.0 in this

work. Compared to SpE [14] that only uses multichannel gradients in image reconstruction, RaIF employs the two additional regularization terms,  $\Psi(\mathbf{X})$  and  $\Phi(\mathbf{X})$ . This treatment encourages the fused image to obtain close luminance and chromaticity to the RGB image.

The gradient data term  $\Theta(\mathbf{X})$  penalizes the difference between the gradients of the reconstructed image and the fused gradients, formulated as

$$\Theta(\mathbf{X}) = \sum_{i \in \{r,g,b\}} \sum_{d \in \{x,y\}} \|\nabla_d \mathbf{X}_i - \mathbf{P}_{i,d}\|_F^2, \quad (17)$$

where  $\|\cdot\|_F$  denotes Frobenius-norm,  $\nabla_x$  and  $\nabla_y$  the gradient operators in the horizontal and vertical directions, respectively.

The regularization term  $\Psi(\mathbf{X})$  imposes two luminance constraints: 1) The luminance of the overexposed sky regions should be close to that of the recovered sky regions, and 2) the luminance of the other regions should be close to that of the original RGB image. Accordingly, this term is defined as

$$\Psi(\mathbf{X}) = \|(\mathbf{L} - \mathbf{L}_{\text{sky}}) \odot \mathbf{M}_1\|_F^2 + \|(\mathbf{L} - \mathbf{L}_{\text{org}}) \odot \mathbf{M}_2\|_F^2, \quad (18)$$

where  $\mathbf{L}$  is the luminance of the image  $\mathbf{X}$ , formulated as

$$\mathbf{L} = (\mathbf{X}_r + \mathbf{X}_g + \mathbf{X}_b)/3. \quad (19)$$

$\mathbf{L}_{\text{sky}}$  and  $\mathbf{L}_{\text{org}}$  are the luminance of the image  $\tilde{\mathbf{R}}_{\text{sky}}$  and  $\mathbf{R}$ , computed using the similar form as  $\mathbf{L}$ .  $\mathbf{M}_1$  denotes the overexposed sky weight mask, computed as

$$\mathbf{M}_1 = \mathbf{M}_{\text{sky}} \odot \mathbf{M}_{\text{overexposed}}, \quad (20)$$

and  $\mathbf{M}_2$  is the complementary weight mask of  $\mathbf{M}_1$  and is computed as

$$\mathbf{M}_2 = \mathbf{1} - \mathbf{M}_{\text{sky}} \odot \mathbf{M}_{\text{overexposed}}. \quad (21)$$

Similarly, the regularization term  $\Phi(\mathbf{X})$  imposes two chromaticity constraints, i.e., 1) the chromaticity of the overexposed sky regions should be close to that of the recovered sky regions, and 2) the chromaticity of the other regions should be close to that of the original RGB image. This term is defined as

$$\Phi(\mathbf{X}) = \sum_{i \in \{r,g\}} \|(\mathbf{C}_i - \mathbf{C}_{\text{sky},i}) \odot \mathbf{M}_1\|_F^2 + \|(\mathbf{C}_i - \mathbf{C}_{\text{ori},i}) \odot \mathbf{M}_2\|_F^2, \quad (22)$$

where  $\mathbf{C}_i$  is the chromaticity channels of the image  $\mathbf{X}$ , formulated as

$$\mathbf{C}_i = \mathbf{X}_i \odot (\mathbf{X}_r + \mathbf{X}_g + \mathbf{X}_b). \quad (23)$$

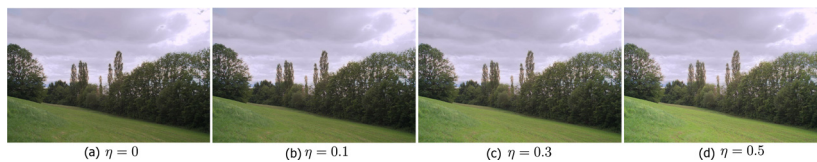


Fig. 9. Fused images with various parameters  $\eta$ . (a)  $\eta = 0$ . (b)  $\eta = 0.1$ . (c)  $\eta = 0.3$ . (d)  $\eta = 0.5$ .

$\mathbf{C}_{\text{org},i}$  and  $\mathbf{C}_{\text{sky},i}$  are the chromaticity channels of the image  $\tilde{\mathbf{R}}_{\text{sky}}$  and  $\mathbf{R}$ , computed using similar form as  $\mathbf{C}_i$ .

We transform  $\{\mathbf{X}_r, \mathbf{X}_g, \mathbf{X}_b\}$  to  $\{\mathbf{L}, \mathbf{C}_r, \mathbf{C}_g\}$  using (19) and (23), and apply alternative optimization [28] to solve (16). Our experiments indicate that 2 iterations suffice to obtain a stable solution  $\{\mathbf{L}, \mathbf{C}_r, \mathbf{C}_g\}$ .

### 3.5. Luminance compression and vegetation enhancement

It is noted that the solved luminance of the overexposed sky regions usually exceeds 1. Linear normalization is a straightforward way to compress the luminance to the range [0,1], but it can make dark regions further duller. To deal with this issue, this work employs a nonlinear function that keeps the luminance unchanged in the range [0,0.5] but compresses the luminance beyond 0.5, formulated as

$$f(z) = \begin{cases} z, & 0 \leq z < 0.5 \\ a \ln(z + b) + c, & 0.5 \leq z \leq z_m \end{cases}, \quad (24)$$

where  $z_m$  denotes the maximal luminance. The function should be continuous at  $z = 0.5$  and properly compress the luminance in the range  $[0.5, z_m]$ . Accordingly,  $f(z)$  should satisfy three conditions, i.e.,  $f(0.5) = 0.5$ ,  $f'(0.5) = 1$ , and  $f(z_m) = 1$ . These conditions uniquely determine the three parameters  $a$ ,  $b$ , and  $c$  in (24).

We then enhance the compressed luminance channel using the vegetation probability map to improve the visibility of the vegetation regions,

$$\mathbf{L}_{\text{enh}} = \mathbf{L} + \eta \cdot \mathbf{L} \odot \mathbf{M}_{\text{veg}}, \quad (25)$$

where  $\eta$  is an adjustable parameter to control the luminance of the vegetation regions. Fig. 9 presents the fused images when  $\eta = 0, 0.1, 0.3$  and  $0.5$ . It is observed that with different  $\eta$ , the vegetation is enhanced with varying degrees. Without loss of generality, we set  $\eta = 0.3$  in the experiments (Section 4).

Finally, the fused image  $\mathbf{X}_{\text{enh}}$  is reconstructed using the enhanced luminance channel  $\mathbf{L}_{\text{enh}}$  and chromaticity channels  $\mathbf{C}_r$  and  $\mathbf{C}_g$ .

### 3.6. Algorithm summary

For clarity, Algorithm 1 lists the entire algorithm of the proposed region-aware image fusion (RaIF) method.

## 4. Experiments

In this section, we first introduce the datasets used in the experiments. Then, we introduce 14 evaluating metrics, five of which evaluate the sky recovery results and the others evaluate the image fusion results. Finally, we present and analyze the experimental results.

We perform four kinds of experiments. First, we compare the proposed sky recovery algorithm with four grayscale image colorization approaches including TCG [29], CUO [30], DEBC [31], and PCTDSC [32], and six state-of-the-art image fusion approaches including GTF [13], SpE [14], DenseFuse [8], CSF [17], SEDRFuse [16], and STDFusionNet [18]. Similar to RaIF, GTF and SpE fuse RGB and NIR images in the gradient domain. DenseFuse, CSF, SEDRFuse, and

---

### Algorithm 1: Region-aware Image Fusion (RaIF)

---

**Input:** NIR image  $\mathbf{N}$ , RGB image  $\mathbf{R}$ .

**Output:** Fused image  $\mathbf{X}_{\text{enh}}$ .

1. Compute the sky map using (7);
  2. Compute the vegetation map using (10);
  3. Compute the parameters  $\hat{a}_i, i \in \{r, g, b\}$  and  $\hat{\gamma}$  of the gain gamma model;
  4. Predict the overexposed sky regions using (15);
  5. Gradient fusion according to Section 3.3;
  6. Compute the luminance and chromaticity channels from the gradient domain using (16);
  7. Compress the luminance range using (24);
  8. Enhance the vegetation regions using (25).
  9. Reconstruct the fused image  $\mathbf{X}_{\text{enh}}$  from luminance  $\mathbf{L}_{\text{enh}}$  and chromaticity  $\mathbf{C}_r$  and  $\mathbf{C}_g$ .
- 

STDFusionNet employ deep learning for image fusion. Second, we compare the proposed image fusion method with the above six image fusion approaches. Then, we employ RaIF as a module to refine the deep learning based image fusion approaches. Finally, we test the RaIF method on images with specular highlight regions other than sky overexposure.

### 4.1. Datasets

We build a dataset with 20 RGB-NIR image pairs for the sky recovery experiment. The images are acquired using a prototype camera equipped with an OmniVision ov4686-H67A sensor<sup>1</sup>. We properly adjust the exposure time so that the RGB images are all well exposed and can serve as ground truths. We then intentionally overexposed the sky regions by multiplying the image intensity with a scale larger than 1.

We collect 52 public RGB-NIR image pairs from the EPFL dataset [33] for the image fusion experiment. Each image contains both sky and vegetation regions. Among these images, 24 ones are divided into the SKY-OVEREXPOSED group and the rest are divided into the SKY-WELLEXPOSED group, judged by the criterion if the sky regions are overexposed.

### 4.2. Evaluation metrics

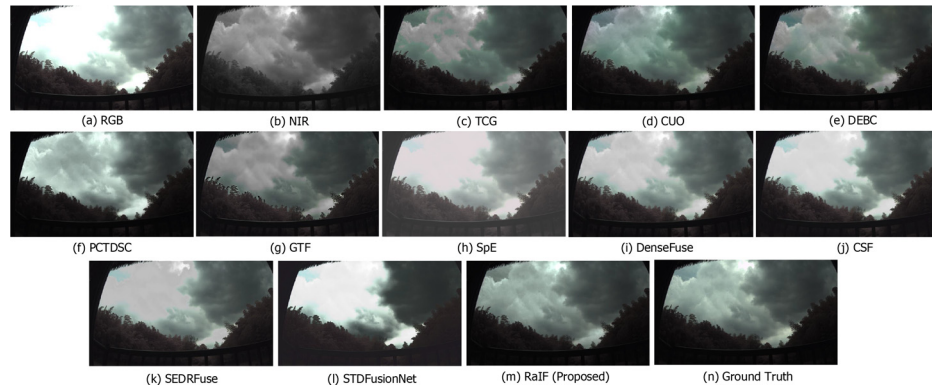
Five quality metrics are used for sky recovery evaluation, including  $\Delta E$  [34], spectral angle mapper (SAM) [35], root mean square error (RMSE) [35], peak signal to noise ratio (PSNR) [36], and structural similarity (SSIM) [36].  $\Delta E$  measures the accuracy of color restoration. SAM and RMSE measure the spectral quality and the spatial quality, respectively. PSNR measures the global quality of the recovered image, while SSIM measures the structure consistency of two images.

Nine objective metrics are used for image fusion evaluation, including entropy (EN) [36], average gradient (AG) [36], spatial frequency (SF) [36], color naturalness index (CNI) [37], colorfulness

<sup>1</sup> <https://www.ovt.com/products/ov04686-h67a/>

**Table 1**  
The average metric values of the sky recovery results produced by RaIF, four grayscale image colorization approaches, and six image fusion approaches. The best ones are in bold.

	Method/Metric	$\Delta E \downarrow$	SAM $\downarrow$	RMSE $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$
Colorization approaches	TCG [29]	20.94	4.44	50.66	14.50	0.61
	CUO [30]	17.17	3.93	42.76	16.21	0.75
	DEBC [31]	17.95	2.65	43.66	15.96	0.74
	PCTDSC [32]	27.23	1.55	76.39	10.72	0.73
	GTF [13]	17.58	1.57	41.35	16.20	0.71
Fusion approaches	SpE [14]	36.05	2.74	102.98	8.41	0.50
	DenseFuse [8]	12.57	1.54	29.84	19.06	0.77
	CSF [17]	15.87	1.54	41.76	15.69	0.74
	SEDRFuse [16]	23.22	1.54	62.49	12.67	0.64
	STDFusionNet [18]	24.11	1.55	67.65	11.38	0.67
	RaIF (Ours)	<b>4.36</b>	<b>1.18</b>	<b>7.99</b>	<b>27.30</b>	<b>0.89</b>



**Fig. 10.** The sky recovery results of the *handrail* scene produced by the proposed RaIF, four grayscale image colorization approaches, and six image fusion approaches. (a) RGB image. (b) NIR image. (c) TCG [29]. (d) CUO [30]. (e) DEBC [31]. (f) PCTDSC [32]. (g) GTF [13]. (h) SpE [14]. (i) DenseFuse [8] (j) CSF [17]. (k) SEDRFuse [16]. (l) STDFusionNet [18] (m) RaIF (Proposed) (n) Ground truth.

metric (CM) [38], mutual information (MI) [36], feature mutual information (FMI) [39], visual information fidelity (VIF)[40], and root mean square error (RMSE). EN measures the amount of information contained in a fused image. AG quantifies the gradient information of a fused image. SF measures the horizontal and vertical gradient distribution of a fused image. CNI evaluates the color statistical naturalness of a fused image. CM measures the degree of colorfulness of a fused image. MI and FMI measure the amount of image information and feature information transferred from source images to the fused image, respectively. VIF measures the visual information fidelity of the fused image. RMSE measures the distortion from the input RGB image to the fused image, and RMSE is only computed on the non-overexposed regions to eliminate the influence of the overexposed areas. It is worth noting that each objective metric only reflects one aspect of the fused images and does not ensure the superiority of image fusion approaches. For example, exaggerated contrast of fused images and artifacts can improve the AG and SF values but do not provide a good appearance. Therefore, the quantitative results are analyzed together with the fused image appearance produced by image fusion approaches.

### 4.3. Evaluation of sky recovery

We evaluate the proposed sky recovery algorithm on the self-collected 20 RGB-NIR image pairs, compared with four grayscale image colorization approaches including TCG [29], CUO [30], DEBC [31], and PCTDSC [32], and six image fusion approaches including GTF [13], SpE [14], DenseFuse [8], CSF [17], SEDRFuse [16], and STDFusionNet [18]. It is worth noting that, for the sake of quantitative evaluation, we adopt linear normalization instead of the non-linear mapping function (24) in luminance compression, because in this dataset the sky regions are intentionally made overexposed

by manual scaling. For a fair comparison, the metrics are computed only in the sky regions.

Table 1 lists the average  $\Delta E$ , SAM, RMSE, PSNR, and SSIM values of the sky recovery results. It is observed that the proposed RaIF method obtains the lowest  $\Delta E$ , SAM and RMSE values and highest PSNR and SSIM values among these approaches. This demonstrates that RaIF outperforms the competitors in terms of color accuracy, spatial accuracy, global quality, and structure fidelity. Fig. 10 illustrates the sky recovery results of the *handrail* scene. It is observed that TCG [29], CUO [30], and DEBC [31] paint false colors to the sky regions. PCTDSC [32] produces a quite bright sky that still contains overexposed luminance. GTF [13] and DenseFuse [8] generate gray sky. SpE [14], CSF [17], SEDRFuse [16], and STDFusionNet [18] can hardly recover the texture of the overexposed sky. In comparison, RaIF produces colored sky that is very close to the ground truth.

### 4.4. Evaluation of image fusion

We evaluate the image fusion approaches on the 52 public RGB-NIR image pairs, comparing RaIF with the state-of-the-art approaches including GTF [13], SpE [14], DenseFuse [8], CSF [17], SEDRFuse [16], and STDFusionNet [18].

1) *Qualitative Comparison:* Fig. 11 shows the fused images of 5 scenes in the SKY-OVEREXPOSED group produced by different image fusion approaches. The original RGB images all suffer from sky overexposure while the NIR images still contain sky information. It is observed that all the competitors except SpE [14] produce fake or gray skies in the *Plantation*, *Trees1*, and *Grass* scenes, while SpE fails to fuse the textures of the overexposed regions from the NIR image. It is also observed that GTF [13] generates blurred and unclear fusion results, especially in the *Lake* and *Trees2* scenes. Furthermore, artifacts arise (see red boxes) in the fused images pro-



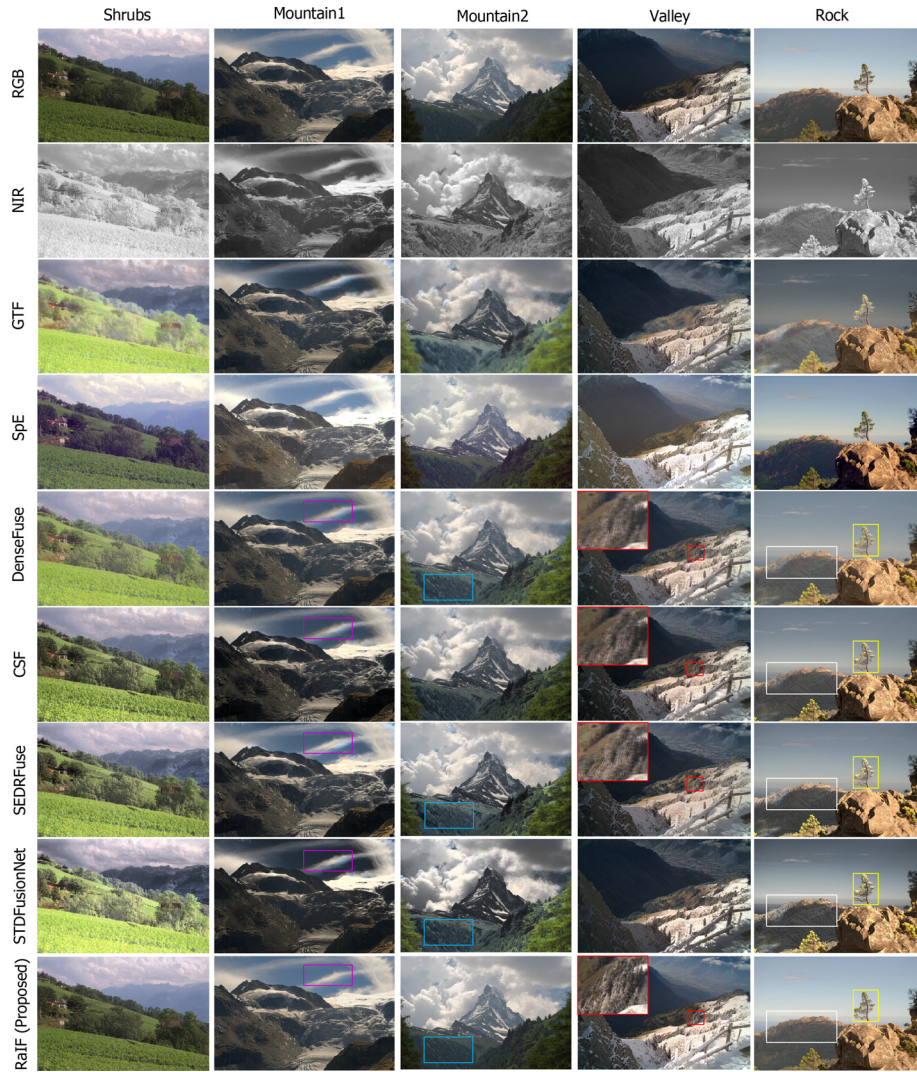
**Fig. 11.** Image fusion results of 5 scenes in the SKY-OVEREXPOSED group, produced by GTF [13], SpE [14], DenseFuse [8], CSF [17], SEDRFuse [16], STDFusionNet [18], and RaIF. Red boxes highlight the image fusion quality of GTF, DenseFuse, CSF, SEDRFuse, STDFusionNet, and RaIF.

duced by GTF [13], DenseFuse [8], CSF [17], SEDRFuse [16], and STDFusionNet [18] in the *Lake* and *Trees2* scenes. The reason is that these fusion approaches are sensitive to minor misalignment when fusing textures in the source images. In comparison, RaIF generates colorful sky that is harmonious with the scene appearances. Moreover, RaIF also improves the visibility of the vegetation regions, which is clearly observed in the *Grass* and *Lake* scenes. In addition, in the *Lake* scene, RaIF can improve the texture details of the mountain without introducing artifacts.

Fig. 12 illustrates the image fusion results of 5 scenes in the SKY-WELL EXPOSED group produced by different approaches. In this group, RGB images are not overexposed but the details are degraded by fog and haze. It is observed that GTF [13] generates blurred details in the *Mountain2* and *Rock* scenes, and also produces too bright vegetation in the *Shrubs* scene. SpE [14] generates overexposed regions that lose textures in the *Mountain1* and *Valley* scenes. DenseFuse [8], CSF [17], SEDRFuse [16], and STDFusionNet [18] significantly improve the contrast of the fuse images but still have limitations. For example, STDFusionNet produces too bright vegetation in the *Shrubs* scene and too dark mountain in the *Mountain1* scene. Color distortion appears in the purple boxes in the *Mountain1* scene. Moreover, in the *Mountain2* scene, the green trees turn to be cyan in the fusion results produced by DenseFuse,

SEDRFuse, and STDFusionNet, as highlighted by blue boxes. Artifacts also arise in the fused images produced by DenseFuse, CSF, and SEDRFuse in the *Valley* scene (see red boxes). In the *Rock* scene, the vegetation turns to be gray (see white boxes), and fake white edges arise around the tree (see yellow boxes) in the fusion results of DenseFuse, CSF, SEDRFuse, and STDFusionNet. In comparison, RaIF improves the visibility of textures, as can be observed in the mountain areas in the *Shrubs*, *Valley*, and *Rock* scenes. Moreover, the resulting images produced by RaIF are free of color distortion, fake edges, and artifacts.

2) *Quantitative Comparison:* Table 2 lists the average metric values of the fused images produced by different approaches of the SKY-OVEREXPOSED group at the levels of the whole image and different regions (i.e., sky and vegetation regions). The metrics CM, CNI, MI, and FMI are suitable to evaluate the performance of the sky region recovery. The metrics AG, SF, VIF, and RMSE are effective to evaluate the visibility, as well as luminance and color distortions, in the vegetation regions. It is observed that SpE obtains the highest average MI and FMI values at the level of the whole image. This is because SpE [14] aims to produce fused images with the distributions similar to those of the original RGB images. Therefore SpE obtains very low MI and FMI values between the fused and NIR images, and quite high MI and FMI values between the fused



**Fig. 12.** Image fusion results of 5 scenes in the SKY-WELLEXPOSED group, produced by GTF [13], SpE [14], DenseFuse [8], CSF [17], SEDRFuse [16], STDFusionNet [18], and RaIF of the SKY-WELLEXPOSED Group. Purple, blue, red, white and yellow boxes highlight the image fusion quality of DenseFuse, CSF, SEDRFuse, STDFusionNet, and RaIF, in terms of color distortion, color cast, artifacts, color fading, and fake edges.

**Table 2**

Average metric values of fused images in the SKY-OVEREXPOSED group using different fusion approaches. The best ones are in bold and the second best ones are underlined.

Whole Image								
Method	EN↑	AG↑	SF↑	CNI↑	MI↑	FMI↑	VIF↑	RMSE↓
GTF [13]	7.215	5.184	39.356	<u>0.596</u>	3.593	0.409	0.434	61.905
SpE [14]	6.742	3.981	31.826	0.519	<b>5.913</b>	<b>0.509</b>	<u>0.521</u>	<u>26.083</u>
DenseFuse [8]	7.358	4.314	31.553	0.575	4.849	0.432	0.448	37.013
CSF [17]	<b>7.440</b>	4.803	35.918	0.574	4.556	0.359	0.452	29.468
SEDRFuse [16]	7.185	4.683	35.379	0.569	4.234	0.392	0.435	33.941
STDFusionNet [18]	7.175	<u>5.207</u>	<u>39.891</u>	<b>0.625</b>	4.768	0.411	0.453	49.033
RaIF (Proposed)	<u>7.438</u>	<b>5.369</b>	<b>40.567</b>	0.582	<u>5.674</u>	<u>0.476</u>	<b>0.528</b>	<b>15.214</b>
Sky Region				Vegetation Region				
Method	CM↑	CNI↑	MI↑	FMI↑	AG↑	SF↑	VIF↑	RMSE↓
GTF [13]	8.772	0.471	4.313	0.740	7.473	29.151	0.499	58.421
SpE [14]	<u>13.096</u>	0.599	4.257	<b>0.787</b>	5.469	22.261	<u>0.597</u>	23.445
DenseFuse [8]	12.811	0.573	<b>5.454</b>	0.765	6.289	23.478	0.564	38.892
CSF [17]	12.476	0.611	4.667	0.735	7.069	27.124	0.514	<u>23.183</u>
SEDRFuse [16]	12.551	0.597	3.444	0.765	7.134	26.899	0.513	30.320
STDFusionNet [18]	13.069	<b>0.685</b>	4.752	0.750	<b>8.108</b>	<u>30.239</u>	0.520	46.864
RaIF (Proposed)	<b>15.190</b>	<u>0.678</u>	<u>5.245</u>	<u>0.784</u>	<u>7.967</u>	<b>30.473</b>	<b>0.639</b>	<b>16.764</b>

**Table 3**

Average metric values of fused images in the SKY-WELLEXPOSED group using different fusion approaches. The best ones are in bold, and the second best ones are underlined.

Method	EN↑	AG↑	SF↑	CNI↑	MI↑	FMI↑	VIF↑	RMSE↓
GTF [13]	7.319	5.014	35.599	0.576	3.768	0.463	0.449	53.285
SpE [14]	7.419	5.233	38.085	0.578	<b>5.866</b>	<b>0.522</b>	<b>0.581</b>	<u>24.444</u>
DenseFuse [8]	7.368	4.525	31.129	0.567	4.231	0.453	0.462	33.128
CSF [17]	<b>7.547</b>	5.232	36.464	0.561	4.081	0.383	0.468	29.109
SEDRFuse [16]	<u>7.545</u>	5.176	36.169	0.577	4.391	0.435	0.465	33.195
STDFusionNet [18]	7.506	<b>6.315</b>	<b>44.518</b>	<b>0.595</b>	4.532	0.457	0.512	50.298
RaIF (Proposed)	7.535	<u>5.508</u>	<u>38.750</u>	<u>0.590</u>	<u>5.429</u>	<u>0.489</u>	<u>0.524</u>	<b>12.746</b>

and RGB images. However, as mentioned above, SpE has poor performance in fusing NIR textures, especially in the overexposed regions (see Fig. 11). CSF [17] obtains the highest EN value at the level of the whole image, but it cannot recover the overexposed sky and avoid artifacts. STDFusionNet [18] obtains the highest CNI value at the level of the whole image and the sky region, indicating that the fused images have the highest degree of naturalness. However, STDFusionNet produces too large global luminance contrast in the *Trees1*, *Grass*, and *Lake* scenes in Fig. 11. The proposed RaIF method has the highest VIF value and the lowest RMSE value at the level of the whole image and the vegetation region, verifying that RaIF can produce fused images with good visual fidelity. Furthermore, RaIF always has the best or the second best AG and SF values at the level of the whole image and the vegetation region, indicating that RaIF performs well in texture enhancement. RaIF has the highest CM value at the level of the sky region since it can generate sky regions with rich color information. Moreover, RaIF has the second best CNI values, which indicates that the recovered sky has a relatively natural color distribution.

Table 3 lists the average metric values of the fused images produced by different approaches of the SKY-WELLEXPOSED group for the whole image. It is observed that the proposed RaIF method obtains the lowest average RMSE value. This indicates that RaIF produces fused images with the least distortion from the RGB images. Besides, RaIF obtains the second-best average AG, SF, MI, and

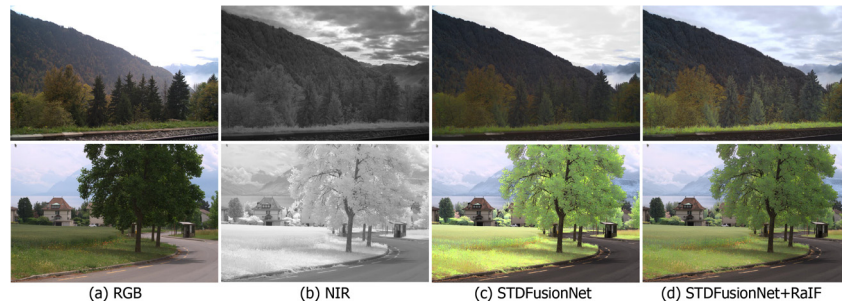
FMI values, which validates that it keeps a good balance in detail enhancement and information preservation. Moreover, RaIF also ranks second in CNI and VIF metrics, meaning that RaIF has relatively good color naturalness and visual fidelity. The above quantitative results are expected since RaIF aims to tackle the sky overexposure and weak vegetation visibility issues but not texture enhancement.

#### 4.5. Image fusion by using RaIF as a refinement module

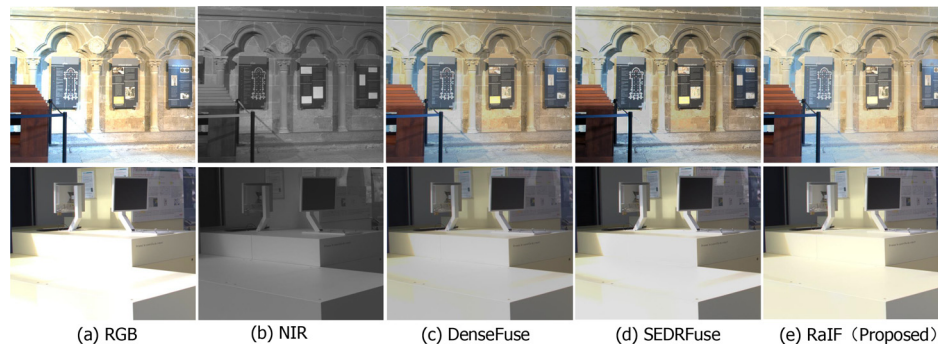
As discussed above, some state-of-the-art approaches have advantages in detail preserving and texture enhancement, but cannot recover the overexposed sky and obtain vegetation with appropriate luminance. In the following, we show that RaIF can act as a refinement module to overcome the limitation of these approaches but still keep their advantages.

For this purpose, we compute the guided gradients  $\mathbf{P}$  in (17), guided luminance  $\mathbf{L}_{org}$  in (18), and guided chromaticity  $\mathbf{C}_{org}$  in (22) using the fused images produced by the state-of-the-art approaches. In addition, we also adjust the vegetation luminance of the fused images to obtain a relatively real appearance.

Fig. 13 shows the image fusion results of two scenes produced by the state-of-the-art STDFusionNet [18] with and without RaIF refinement. It is observed that the original STDFusionNet produces gray sky and too bright vegetation, respectively, in the two scenes.



**Fig. 13.** Image fusion refinement of the STDFusionNet [18] result using RaIF module of two scenes.



**Fig. 14.** Image fusion results of 2 scenes under the circumstance of specular highlights, produced by DenseFuse [8], SEDRFuse [16], and the proposed RaIF method.

In contrast, STDFusionNet with RaIF refinement successfully recovers the sky color and produces a more realistic appearance in the vegetation regions with appropriate luminance. Overall, the RaIF module can obviously improve the visual appearance of the state-of-the-art approach in the sky and vegetation regions.

#### 4.6. Recovery of overexposed highlight regions

It is noted that specular highlights can also result in overexposure in RGB images. In the following, we demonstrate the capability of RaIF to deal with this circumstance. Fig. 14 shows the image fusion results of two overexposed scenes produced by DenseFuse [8], SEDRFuse [16], and the proposed RaIF method. It is observed that DenseFuse and SEDRFuse result in color fading in the overexposed wall and desk regions. In comparison, RaIF successfully recovers the luminance and color information in these overexposed regions, thanks to the generalization ability of the sky region map generation algorithm on other overexposed regions.

## 5. Conclusions

This work proposes a region-aware RGB and near-infrared image fusion method, called RaIF, based on the spectral characteristics of different regions. Different from the existing image fusion methods, this work focuses on tackling the appearance issues of RGB images in the sky and vegetation regions. The proposed RaIF method recovers the textures and colors of the overexposed sky by exploiting the relationship of NIR and RGB pixels, and improves the visibility of the vegetation regions in an adjustable manner using the generated probability map. Experimental results validate that RaIF performs well on both sky overexposed and well-exposed images. Furthermore, RaIF can be used as a module to improve the fusion results produced by other state-of-the-art approaches. It can also recover specular highlight regions other than sky overexposure. The practical applications of RaIF include photo beautification and RGB-NIR joint high dynamic range imaging for outdoor scenery.

A limitation of the proposed method is that it currently only deals with the sky and vegetation regions. In the future we will extend our region-aware fusion method to a wider range of scenarios by focusing on other interested regions.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The authors do not have permission to share data.

## Acknowledgment

This work was supported in part by the "Pioneer" and "Leading Goose" R & D Program of Zhejiang under grant 2023C03136 and in part by the Ten Thousand Talents Program of Zhejiang Province under grant 2020R52003.

## References

- [1] Q. Zhang, T. Shi, F. Wang, R.S. Blum, J. Han, Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency, *Pattern Recognit.* 83 (2018) 299–313.
- [2] Q. Zhang, G. Li, Y. Cao, J. Han, Multi-focus image fusion based on non-negative sparse representation and patch-level consistency rectification, *Pattern Recognit.* 104 (2020) 107325.
- [3] S. Li, R. Dian, L. Fang, J.M. Bioucas-Dias, Fusing hyperspectral and multispectral images via coupled sparse tensor factorization, *IEEE Trans. Image Process.* 27 (8) (2018) 4118–4130.
- [4] J. Zhang, C. Li, S. Kosov, M. Grzegorzec, K. Shirahama, T. Jiang, C. Sun, Z. Li, H. Li, Lcu-net: a novel low-cost U-net for environmental microorganism image segmentation, *Pattern Recognit.* 115 (2021) 107885.
- [5] H. Chen, C. Li, G. Wang, X. Li, M.M. Rahaman, H. Sun, W. Hu, Y. Li, W. Liu, C. Sun, et al., Gashis-transformer: a multi-scale visual transformer approach for gastric histopathological image detection, *Pattern Recognit.* 130 (2022) 108827.
- [6] X. Zhang, Z. Sheng, H.-L. Shen, Focusnet: classifying better by focusing on confusing classes, *Pattern Recognit.* 129 (2022) 108709.
- [7] A.V. Vanmali, V.M. Gadre, Visible and NIR image fusion using weight-map-guided Laplacian-Gaussian pyramid for improving scene visibility, *Sādhanā* 42 (7) (2017) 1063–1082.
- [8] H. Li, X.-J. Wu, DenseFuse: a fusion approach to infrared and visible images, *IEEE Trans. Image Process.* 28 (5) (2018) 2614–2623.
- [9] A. Fang, X. Zhao, J. Yang, Y. Zhang, X. Zheng, Non-linear and selective fusion of cross-modal images, *Pattern Recognit.* 119 (2021) 108042.
- [10] H. Su, C. Jung, Multi-spectral fusion and denoising of RGB and NIR images using multi-scale wavelet analysis, in: *International Conference on Pattern Recognition*, 2018.
- [11] M. Awad, A. Elliethy, H.A. Aly, Adaptive near-infrared and visible fusion for fast image enhancement, *IEEE Trans. Comput. Imag. PP* (99) (2019), 1–1.
- [12] J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, *Infrared Phys. Technol.* 82 (2017) 8–17.
- [13] J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, *Inf. Fusion* 31 (2016) 100–109.
- [14] D. Connah, M.S. Drew, G.D. Finlayson, Spectral edge image fusion: theory and applications, in: *European Conference on Computer Vision*, Springer, 2014, pp. 65–80.
- [15] G.D. Finlayson, D. Connah, M.S. Drew, Lookup-table-based gradient field reconstruction, *IEEE Trans. Image Process.* 20 (10) (2011) 2827–2836.
- [16] L. Jian, X. Yang, Z. Liu, G. Jeon, M. Gao, D. Chisholm, SEDRFuse: a symmetric encoder-decoder with residual block network for infrared and visible image fusion, *IEEE Trans. Instrument. Measur.* 70 (2020) 1–15.
- [17] H. Xu, H. Zhang, J. Ma, Classification saliency-based rule for visible and infrared image fusion, *IEEE Trans. Comput. Imag.* 7 (2021) 824–836.
- [18] J. Ma, L. Tang, M. Xu, H. Zhang, G. Xiao, STDFusionNet: an infrared and visible image fusion network based on salient target detection, *IEEE Trans. Instrument. Measur.* 70 (2021) 1–13.
- [19] C. Feng, S. Zhuo, X. Zhang, L. Shen, S. Süssstrunk, Near-infrared guided color image dehazing, in: *2013 IEEE international conference on image processing, IEEE, 2013*, pp. 2363–2367.
- [20] S.E. Kim, T.H. Park, L.K. Eom, Fast single image dehazing using saturation based transmission map estimation, *IEEE Trans. Image Process.* 29 (2019) 1985–1998.
- [21] B. Zafarifar, et al., Blue sky detection for picture quality enhancement, in: *International Conference on Advanced Concepts for Intelligent Vision Systems*, Springer, 2006, pp. 522–532.
- [22] F.J. Kriegler, W.A. Malila, R.F. Nalepka, W. Richardson, Preprocessing transformations and their effects on multispectral recognition, *Remote Sens. Environ.* 6 (1969).
- [23] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, A. Torralba, Semantic understanding of scenes through the ADE20K dataset, *Int. J. Comput. Vis.* 127 (3) (2019) 302–321.
- [24] H. Huang, H. Xu, X. Wang, W. Silamu, Maximum f1-score discriminative training criterion for automatic mispronunciation detection, *IEEE/ACM Trans. Audio Speech Lang. Process.* 23 (4) (2015) 787–797.
- [25] H. Haneishi, T. Hasegawa, A. Hosoi, Y. Yokoyama, N. Tsumura, Y. Miyake, System design for accurately estimating the spectral reflectance of art paintings, *Appl. Opt.* 39 (35) (2000) 6621–6632.
- [26] M.D. Grossberg, S.K. Nayar, Modeling the space of camera response functions, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (10) (2004) 1272–1282.
- [27] A.A. Bell, J. Brauers, J.N. Kaftan, D. Meyer-Ebrecht, A. Bocking, T. Aach, High dynamic range microscopy for cytopathological cancer diagnosis, *IEEE J. Sel. Top. Signal Process.* 3 (1) (2009) 170–184.
- [28] K. Zhang, M. Wang, S. Yang, L. Jiao, Convolution structure sparse coding for fusion of panchromatic and multispectral images, *IEEE Trans. Geosci. Remote Sens.* 57 (2) (2018) 1117–1130.
- [29] T. Welsh, M. Ashikhmin, K. Mueller, Transferring color to greyscale images, in: *Conference on Computer Graphics and Interactive Techniques*, 2002, pp. 277–280.
- [30] A. Levin, D. Lischinski, Y. Weiss, Colorization using optimization, in: *ACM SIGGRAPH 2004 Papers*, 2004, pp. 689–694.
- [31] M. He, D. Chen, J. Liao, P.V. Sander, L. Yuan, Deep exemplar-based colorization, *ACM Trans. Graph.* 37 (4) (2018) 1–16.
- [32] M. He, J. Liao, D. Chen, L. Yuan, P.V. Sander, Progressive color transfer with dense semantic correspondences, *ACM Trans. Graph.* 38 (2) (2019) 1–18.
- [33] M. Brown, S. Süssstrunk, Multi-spectral SIFT for scene category recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 177–184.
- [34] D.H. Brainard, et al., Color appearance and color difference specification, *Sci. color* 2 (191–216) (2003) 5.
- [35] L. Loncan, L.B. De Almeida, J.M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G.A. Licciardi, M. Simoes, et al., Hyperspectral pansharpening: a review, *IEEE Geosci. Remote Sens. Mag.* 3 (3) (2015) 27–46.

- [36] J. Ma, Y. Ma, C. Li, Infrared and visible image fusion methods and applications: a survey, *Inf. Fusion* 45 (2019) 153–178.
- [37] K.-Q. Huang, Q. Wang, Z.-Y. Wu, Natural color image enhancement and evaluation algorithm based on human visual system, *Comput. Vis. Image Understand.* 103 (1) (2006) 52–63.
- [38] D. Hasler, S.E. Suesstrunk, Measuring colorfulness in natural images, in: *Human vision and electronic imaging VIII*, volume 5007, SPIE, 2003, pp. 87–95.
- [39] M.B.A. Haghighat, A. Aghagolzadeh, H. Seyedarabi, A non-reference image fusion metric based on mutual information of image features, *Comput. Electric. Eng.* 37 (5) (2011) 744–756.
- [40] H.R. Sheikh, A.C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2) (2006) 430–444.

**Jiacheng Ying** received the B.Eng. degree in information engineering from Zhejiang University, Hangzhou, China, in 2020, where he is pursuing the Ph.D. degree with the College of Information Science and Electronic Engineering. His research interests are multi-modal image fusion and image processing.

**Can Tong** received the B.Eng. and Master degree in information engineering from Zhejiang University, Hangzhou, China, in 2019 and 2022, respectively. He is currently working as an ISP engineer in NIO, Shanghai, China. His research interests are image processing and image fusion.

**Zehua Sheng** received the B.E. degree in 2017 from Zhejiang University, Hangzhou, China, where he is currently working toward the Ph.D. degree with the College of Information Science and Electronic Engineering, Zhejiang University. His research interests include image denoising and multimodal image processing.

**Bowen Yao** received the B.Eng. degree in information engineering from Zhejiang University, Hangzhou, China, in 2020, where he is pursuing the Master degree with the College of Information Science and Electronic Engineering. His research interests are color constancy and image alignment.

**Si-Yuan Cao** received his B.Eng. degree in electronic information engineering from Tianjin University in 2016, and Ph.D. degree in electronic science and technology from Zhejiang University in 2022. He is currently a lecturer in Ningbo Innovation Center, Zhejiang University, China. His research interests are multispectral/multimodal image registration, homography estimation, place recognition and image processing.

**Heng Yu** obtained his Ph.D. degree from the National University of Singapore (NUS) in 2012. Since his graduation, he has been working as a research scientist and a research fellow at the University of Erlangen-Nuremberg, Germany and NUS, respectively. Prior to his appointment as an Assistant Professor at UNNC, he was an Assistant Professor at the United Arab Emirates University in UAE, and briefly as a Xinghai Associate Professor at Dalian Maritime University. From 2014 to 2016, he served as a R&D director in a big data enterprise in Shenzhen, China. His current research interest lies in intelligent hardware/infrastructure, including but not limited to implementing AI algorithms/applications on dedicated hardware platforms, empowering existing computing systems with intelligence, etc.

**Hui-Liang Shen** received the B.Eng. and Ph.D. degrees in electronic engineering from Zhejiang University, China, in 1996 and 2002, respectively. He is currently a Full Professor with the College of Information Science and Electronic Engineering, Zhejiang University. His research interests include multispectral imaging, image processing, computer vision, and machine learning.