Image Fusion Method Based on Near-Infrared and Visible Light Difference Features (Postprint)

Authors: Zhao Guohui, Zhang He

Date: 2022-06-06T00:00:00+00:00

Abstract

Images captured in low-light environments typically suffer from degradation issues such as darkness, noise, and color cast. To address this, we propose a visible and near-infrared fusion method based on differential features. First, by observing the imaging inconsistency between near-infrared and visible images, differential features are constructed through their structural differences. Second, these differential features are mapped into fusion weights for weighted fusion. Finally, a color cast correction method is employed to restore brightness in regions prone to color cast after fusion, rendering the colors in these regions authentic and natural. The proposed method can effectively utilize near-infrared information to output high-quality, clear images. Experimental results demonstrate that our method outperforms existing algorithms in both subjective perception and objective evaluation.

Full Text

Preamble

Near-Infrared and Visible Image Fusion Based on Differential Features

Zhao Guohui, Zhang He†
(Hangzhou Hikvision Digital Technology Co., Ltd., R&D Center, Hangzhou 310051, China)

Abstract: Images captured in dim light often suffer from severe degradations such as low visibility, intensive noise, and color distortion. To address these issues, this paper proposes a near-infrared and visible light fusion method based on differential features. First, we observe the imaging inconsistency between near-infrared and visible light images and construct differential features from their structural differences. Second, these differential features are mapped into fusion weights for weighted fusion. Finally, a color cast correction method is applied to restore brightness in regions prone to color deviation after fusion, ensuring that the fused image exhibits authentic and natural colors. The proposed method can effectively utilize near-infrared information to produce high-quality, clear images. Experimental results demonstrate that our approach outperforms existing algorithms in both subjective perception and objective evaluation.

Keywords: image fusion; differential feature; near-infrared; color correction

0 Introduction

Low-light conditions at night result in insufficient illumination, causing captured images to suffer from various degrees of degradation including dark brightness, low contrast, high noise, and color imbalance. These issues affect both human visual perception and automatic extraction of useful information. Research on restoring low-light nighttime images can not only improve image aesthetics but also finds important practical applications in intelligent surveillance, autonomous driving, and terrain reconnaissance. A straightforward approach to this problem is to increase visible light supplementation, but this inevitably introduces light pollution. In video surveillance applications, for instance, strong light can cause temporary blindness in drivers, leading to road safety hazards. Other image adjustment methods such as enlarging the aperture (which increases light intake but reduces depth of field) or slowing the shutter speed (which improves signal-to-noise ratio but causes motion blur and ghosting) present similar trade-offs.

Although numerous excellent results have emerged in the low-light image enhancement field, methods relying solely on a single input image struggle to distinguish between random noise and genuine image information under conditions of information scarcity. Denoising [1, 2] and enhancement [3-6] approaches often result in loss of detailed textures and are inadequate at adaptively compensating for contrast loss caused by insufficient illumination, yielding suboptimal improvements. Consequently, low-light enhancement algorithms using paired visible and near-infrared images [7-16] have attracted increasing attention. These methods employ two image sensors with specialized optical components: one captures near-infrared grayscale images containing brightness information and object contours with near-infrared supplementation, while the other captures visible color images containing scene color information. Image fusion algorithms then combine information from both sensors to obtain images possessing both visible color characteristics and near-infrared brightness. Since human eyes have weak perception of near-infrared wavelengths, near-infrared supplementation does not interfere with normal human activities or cause glare distraction to drivers. Therefore, near-infrared and visible light fusion technology holds significant application value in nighttime surveillance scenarios.

Unlike traditional long-wave infrared thermal imaging [17], near-infrared imaging utilizes reflected light from objects, capturing geometric shapes and aiming to produce clean, transparent, and clear color images when fused with visible light. However, maximizing the use of near-infrared brightness information while ensuring authentic and trustworthy colors suitable for human vision or subsequent computer analysis remains a challenging research problem. Due to significant differences in imaging appearance between near-infrared and visible light, directly replacing the brightness channel of visible images with near-infrared grayscale images causes severe color deviation. Krishnan et al. [7] proposed using near-infrared gradient structural information to guide color image denoising. Shen et al. [9] constructed a scale map to explicitly represent the credibility of near-infrared gradients, recovering visible images by solving for the optimal scale map. Su et al. [11] combined multi-scale wavelet analysis to decompose noisy visible and near-infrared images, fusing them by balancing visibility and contrast to reconstruct clear images. Considering that near-infrared provides ideal detail perception, Jee et al. [10] employed joint bilateral filtering with near-infrared images as gradient structural guidance to assist visible light denoising, subsequently superimposing near-infrared texture details. However, due to imaging differences between visible and near-infrared, this approach erroneously blurs some inherent edge gradients in visible images.

Inspired by deep learning theory, several sample-based image fusion methods [13–16] have emerged. Li et al. [13] proposed a deep learning framework comprising an encoder, fusion layer, and decoder model for visible and infrared image fusion, where the end-to-end network eliminates the need for manual prior information extraction. Tang et al. [14] adopted a U-Net architecture with larger receptive fields to more effectively encode multi-scale contextual information, ensuring favorable visual results for visible and near-infrared fusion. Mei et al. [16] designed independent sub-networks for denoising, enhancement, and fusion, with input images sequentially processed through three sub-networks to obtain enhanced results. Such sub-networks have relatively simple tasks and converge easily during training. While these deep learning-based methods have achieved remarkable success, they suffer from requiring large training datasets and substantial computational resources. Fusion quality depends on numerous factors including training label materials, model capacity, loss function selection, and hyperparameter choices.

Currently, low-light visible and near-infrared fusion technology remains primarily in the academic research stage with limited industrial application, mainly due to three reasons: First, structural differences in how objects reflect visible versus near-infrared light often cause color casts due to brightness imbalance in restored images. Second, most methods utilize near-infrared structural information to assist low-light image denoising and detail enhancement but make limited use of near-infrared brightness information, yielding unsatisfactory restoration results. Third, existing methods have high computational complexity and demanding computational requirements, making them difficult to implement for real-time HD video processing and embedded systems.

Observing the inconsistent brightness performance of near-infrared images for objects of the same color, this paper proposes a near-infrared and visible image fusion method based on differential features, referred to as the differential fusion method. The overall framework is shown in Figure 1 [Figure 1: see original paper]. First, the differential feature is obtained by differencing visible and near-infrared brightness, which physically reflects the regions and degrees of spectral reflection differences and provides effective constraints for color cast-prone areas during fusion. Second, based on the correlation between differential features and fusion, a feature mapping function transforms differential features into corresponding fusion weight terms. Third, visible and near-infrared images undergo brightness-domain weighted fusion to obtain the fused image brightness. Given that texture details are weakened in dark environments, near-infrared texture migration is also integrated. Finally, brightness and color information are combined to reconstruct a clear color image. Additionally, this paper proposes a differential feature-based color cast correction method that maximizes infrared information utilization while ensuring visually bright, transparent, high signal-to-noise ratio results with accurate and authentic color reproduction. Experimental results demonstrate that the proposed method is simple yet effectively overcomes color cast issues during image fusion, significantly improves visual quality, and operates at high speed suitable for real-time monitoring tasks.

---

1 Methodology

1.1 Data Acquisition

To obtain authentic low-light visible and near-infrared image pairs, this paper employs the HIK iDS-2CD9736-AMS dual-sensor 9-megapixel camera (resolution 4096$×$2160). This camera uses a beam splitter prism to direct visible and near-infrared light to corresponding sensors. The visible and near-infrared images have only minor rigid spatial transformations, with a calibration matrix obtained using a checkerboard pattern and automatically applied to all image pairs.

The test camera is mounted on a road pole at a height of 6 meters to capture vehicles at a distance of 24 meters. It is paired with an environmentally friendly strobe light (LED strobe and near-infrared flash), where near-infrared images use near-infrared flash supplementation at wavelengths of 730–1000 nm, and visible images use LED strobe supplementation with relatively weak brightness for environmental color acquisition only. Since the core target of road monitoring is moving vehicles, the camera shutter time is set to 2 ms to avoid motion blur, and gain is conventionally set to 50. Due to the short shutter time and weak supplementation, insufficient sensor photosensitivity results in very low signal-to-noise ratio in visible images with substantial noise.

To simplify the problem, input images are pre-processed through Image Signal Processing (ISP) denoising, transforming a low-light noisy image into a low-light blurred image. This allows subsequent processing to focus on image fusion without considering noise issues. The input visible image format is YCbCr, where Y represents the visible brightness channel and CbCr represent the chrominance channels. The input near-infrared image is single-channel, denoted as N for the near-infrared brightness channel. The near-infrared and visible brightness domain image fusion formula is defined as:

$$Y_{\text{fusion}}(x) = a(x) \cdot N(x) + (1 - a(x)) \cdot Y(x)$$

where x is a 2D vector representing pixel coordinates, $Y(x)$ is the visible image brightness radiance map, $N(x)$ is the near-infrared image brightness radiance map, and $Y_{\text{fusion}}(x)$ is the desired multispectral fused brightness map. The parameter a represents the fusion weight of the near-infrared image, describing the proportion of near-infrared brightness in the output image, with values between 0 and 1. Since fusion weights vary with pixel position, the overall map is called a fusion weight map. This equation shows that if the fusion weight information a is known for a scene, the desired result can be calculated accordingly. As the fusion weight map a is unknown, the essence of multispectral fusion based on this physical model can be viewed as estimating parameter a to reconstruct Y_{fusion} from Y and N.

1.2 Differential Feature

Near-infrared light is spectrally very close to human-visible light. Under normal circumstances, the brightness distributions of near-infrared and visible images are considered similar. However, for certain objects, particularly colored ones, there are significant differences in reflective imaging, even completely opposite behaviors. Statistical analysis of numerous naturally captured visible-near-infrared image pairs reveals three main differences: (a) near-infrared images exhibit inconsistent brightness for the same color; (b) visible texture information may disappear in near-infrared images; and (c) brightness reversal can occur between visible and near-infrared images. For example, Figure 2 [Figure 2: see original paper] shows a set of nighttime traffic surveillance images: vehicles of the same red color appear with varying brightness in the corresponding near-infrared images—some bright, some dark. The stripes on the upper-left clothing are invisible in its near-infrared counterpart, while the black seatbelt in the upper-left becomes gray-white in the lower-left near-infrared image. Failure to distinguish these differences during fusion inevitably leads to color cast distortion and texture detail loss.

Assuming that local regions of natural images have identical reflection characteristics for visible and near-infrared light, their local brightness should also be similar. Applying low-pass edge-preserving filtering (such as median filtering, bilateral filtering [18], or guided filtering [19]) to both brightness images smooths local regions and eliminates singularities, yielding brightness feature maps Y_{base} and N_{base}. Taking bilateral filtering of the visible image brightness map as an example:

$$Y_{\text{base}}(x) = \frac{1}{k(x)} \sum_{y \in \Omega} Y(y) \cdot \exp\left(-\frac{\|x-y\|^2}{2\sigma_s^2}\right) \cdot \exp\left(-\frac{\|Y(x)-Y(y)\|^2}{2\sigma_r^2}\right)$$

where the normalization factor is:

$$k(x) = \sum_{y \in \Omega} \exp\left(-\frac{\|x-y\|^2}{2\sigma_s^2}\right) \cdot \exp\left(-\frac{\|Y(x)-Y(y)\|^2}{2\sigma_r^2}\right)$$

Y_{base} is a locally smoothed brightness feature image. The left regularization term defines spatial domain weights based on pixel distances, while the right regularization term defines range weights based on intensity differences. Both are Gaussian distance functions defined over a local region Ω. Empirically, setting σ_s to a 20-pixel neighborhood window and σ_r to 1% of the image' s color range yields good results. The differential image is obtained by subtracting the visible brightness feature from the near-infrared brightness feature:

$$\text{diff}(x) = N_{\text{base}}(x) - Y_{\text{base}}(x)$$

The differential image physically reflects differences in how local regions of natural images reflect visible versus near-infrared light. As shown in Figure 3 [Figure 3: see original paper], brighter regions in the differential map (absolute values displayed for visualization) indicate larger discrepancies between near-infrared and visible reflection characteristics. For instance, the black short-sleeved shirt worn by a driver appears exceptionally bright in the near-infrared image. Directly using near-infrared brightness from these regions can easily cause grayish color casts. Darker regions in the differential map indicate smaller reflection differences, such as the dashboard and black seatbelt, where near-infrared imaging closely matches visible imaging. In these regions, using a higher proportion of near-infrared brightness makes the restored image brighter, more transparent, and higher in signal-to-noise ratio without introducing color cast issues.

1.3 Feature Mapping

The differential feature naively captures imaging discrepancies between different spectra. Leveraging this valuable prior information enables maximizing near-infrared information utilization during fusion while avoiding color bias from incorporating problematic infrared information, achieving desirable image quality. Based on observed correlations between differential features and fusion color casts, we derive the following inference regarding near-infrared fusion weights:

$$\begin{cases} \text{diff}(x) \to 0 \Rightarrow a(x) \to 1 \\ \text{diff}(x) \to \infty \Rightarrow a(x) \to 0 \end{cases}$$

When the differential value approaches zero, visible and near-infrared reflection characteristics are similar, and the near-infrared fusion weight approaches the maximum value of 1. When the differential value approaches infinity, reflection characteristics are opposite, and the weight approaches the minimum value of 0. Note that differential values are positive in regions where near-infrared brightness exceeds visible brightness. Since near-infrared images are captured using flash, this constitutes the vast majority of the image. However, negative differential values also exist in shadow-occluded near-infrared regions. Shadows do not affect image realism but rather enhance the sense of transparency. Based on these observations, we establish a piecewise linear feature mapping function as illustrated in Figure 4 [Figure 4: see original paper].

Figure 4 shows that when differential feature values are negative or small positive values, the near-infrared fusion weight is set to the maximum a_{max} to maximize near-infrared information utilization. When differential features are positive and exceed threshold t_0, the fusion weight gradually decreases to a_{min} to reduce color cast effects from strong near-infrared reflection. Through this mapping, differential feature images are transformed into near-infrared fusion weight maps. In Figure 4, thresholds t_0 and t_1 are empirically set to 20 and 120, respectively. Fusion weight limits a_{max} and a_{min}, which control the degree of near-infrared fusion, are typically set to 0.8 and 0.1. Since the input differential feature data range is deterministic and enumerable (for 8-bit depth images, differential values fall between -255 and 255), this step is often pre-computed as a lookup table to accelerate processing.

1.4 Image Fusion

After obtaining the fusion weight map a through feature mapping, the fused image brightness radiance map can be directly computed using Equation (1). However, visible images captured in low-light conditions suffer from blurred texture lines and degraded detail gradients due to insufficient photosensitivity. Therefore, surface texture details from the near-infrared image must be extracted and migrated into the fused image to enrich details and improve visual hierarchy.

The final fused brightness is defined as:

$$Y_{\text{fusion}}(x) = Y_{\text{base}}(x) + D(x)$$

where Y_{base} is the weighted fusion result of visible and near-infrared brightness domains, and $D(x)$ is the migrated texture detail layer:

$$D(x) = \lambda_{\text{VIS}} \cdot (Y(x) - Y_{\text{base}}(x)) + \lambda_{\text{NIR}} \cdot a(x) \cdot (N(x) - N_{\text{base}}(x))$$

Parameters λ_{VIS} and λ_{NIR} control the texture superposition strength from visible and near-infrared images, respectively. Larger values increase sharpness but may cause over-sharpening and pixel overflow. These are generally set to 1.0.

For near-infrared texture, we also multiply by the fusion weight to avoid artifacts and maladaptation when gradient reversal occurs between near-infrared and visible images.

For color acquisition in fused images, the low-light visible UV channels are multiplied by corresponding brightness gain values:

$$U_{\text{fusion}}(x) = U(x) \cdot \text{Gain}(x), \quad V_{\text{fusion}}(x) = V(x) \cdot \text{Gain}(x)$$

where $\text{Gain}(x) = Y_{\text{fusion}}(x)/Y(x)$ is the brightness gain value. Finally, the fused YUV color space is converted to RGB to obtain the color fusion result. As shown in Figure 5 [Figure 5: see original paper], the fused image becomes clean and transparent with clearer facial details, while maintaining color consistency with the visible image for authentic and trustworthy colors.

1.5 Color Cast Correction

Low-light visible images have lower signal-to-noise ratio compared to near-infrared flash images. Although enhancement methods like LIME [3] can brighten visible images, they cannot increase information content and introduce negative effects such as color noise and halation, as shown in Figure 6(d). Therefore, incorporating more near-infrared brightness improves signal-to-noise ratio. We experimented with weight design based on maximum brightness principles for visible-near-infrared pairs. While this significantly improves brightness and clarity, it suffers from severe color distortion that contradicts human perception—for example, dark red car bodies appear pinkish and dark green grass appears cyan, as shown in Figure 6(e). Variational methods primarily utilize near-infrared gradient information while neglecting near-infrared brightness, resulting in dark images lacking the high-contrast characteristics of near-infrared. Reference [14]' s deep learning approach avoids color cast but incorrectly enhances headlight halation and illuminated road areas, likely due to the difficulty of obtaining real low-light visible-near-infrared fusion datasets; they used long-exposure images as training labels, causing the network to function as visible denoising and enhancement without effectively leveraging near-infrared information for improved visibility. In contrast, our differential fusion method produces clear vehicle textures with vivid colors, prominent road markings, and overall visual effects closer to near-infrared flash images.

While the differential fusion method avoids color distortion, it cannot guarantee high signal-to-noise ratio in fused images due to limited near-infrared utilization. Increasing infrared usage further introduces perceivable color casts, particularly for red colors to which human eyes are sensitive. As shown in Figure 6(c), dark red car bodies appear slightly pinkish. Additionally, nighttime images often suffer from dull colors or casts due to insufficient supplementation or white balance issues, requiring further correction.

First, the differentially fused image is converted to RGB color space for color feature calculation. Common colors such as red, green, and blue often experience color casts due to excessive near-infrared brightness. Therefore, we design the following correction formula based on brightness changes and pixel color intensity characteristics:

$$I_c^{\text{corrected}}(x) = I_c(x) \cdot \min(1 - k(x), I_c(x), s(x)), \quad c \in \{r, g, b\}$$

where $k(x) = \max(Y_{\text{fusion}}(x) - Y(x), 0)$ represents the brightness increase after fusion, and $s(x) = \max_c I_c(x) - \min_c I_c(x)$ represents the color intensity value. The corrected fusion result is shown in Figure 6(f), demonstrating effective infrared information utilization with bright, transparent, high signal-to-noise ratio images featuring authentic colors. The overall style approaches near-infrared flash brightness while preserving visible light' s true color characteristics.

---

2 Experimental Results

To validate the proposed method' s effectiveness, this section compares our approach with state-of-the-art methods using both real traffic scene images and publicly available near-infrared/visible datasets. Comparison methods include the representative and influential BM3D denoising algorithm [1], variational method [9], and deep learning fusion method [14].

2.1 Visual Subjective Analysis

Figure 7 [Figure 7: see original paper] presents fusion results for low-light traffic scene images captured with near-infrared flash. The challenges in traffic scene imaging include fast-moving vehicles (requiring short 2 ms shutter times, resulting in dark, noisy images) and complex nighttime lighting (streetlights, vehicle headlights causing large halos). Near-infrared images benefit from flash supplementation with balanced brightness, high clarity, and minimal noise/halo. Visually, BM3D denoising maintains the original style with low noise but insufficient clarity and contrast. Reference [9] produces blurry results in regions with abnormal infrared reflection and yields negative visible edge values in headlight-illuminated road areas. Reference [14] effectively avoids color cast but underutilizes near-infrared information, failing to suppress headlight halation effectively. Our method avoids color distortion and halo artifacts while fully extracting near-infrared information, achieving brightness and clarity close to flash images with the best e and $\bar{r}$ scores.

Figure 8 [Figure 8: see original paper] shows low-light images captured in a darkroom with near-infrared flash. Although a static scene, the extremely dark environment produces significant noise that overwhelms information. BM3D suppresses noise well but with far lower clarity than near-infrared fusion results. Reference [9] yields poor contrast with unnatural visual coordination. Reference

[14] cannot effectively utilize near-infrared information, leaving some regions unclear (e.g., blurred text on packaging boxes). Our differential fusion method ideally restores object colors and clarity, with packaging text clearly visible.

Figure 9 [Figure 9: see original paper] compares ceramic artifact photos captured in a darkroom. The original images suffer from significant color noise and 朦胧 appearance. BM3D eliminates noise but produces dull, blurry images. Reference [9] restores clarity but cannot effectively remove noise (e.g., color noise remains on terracotta warriors). Reference [14] produces clear images but with insufficient contrast and transparency. Our method combines visible color advantages with near-infrared brightness while overcoming their respective limitations, achieving superior fusion performance in noise reduction, color accuracy, visible information preservation, contrast, and near-infrared information content.

2.2 Objective Quantitative Comparison

While image fusion lacks definitive objective metrics and relies primarily on subjective evaluation, low-light image fusion can be viewed as an image enhancement process. Since low-light imaging reduces edge features and detail information, restoration quality can be measured by edge strength and detail contrast. A widely used metric in image enhancement is the visible edge gradient method proposed by Hautiere et al. [20], which uses the ratio of newly visible edges (e), average gradient ratio ($\bar{r}$), and saturated pixel ratio (σ) to describe enhancement effects:

$$e = \frac{n_r - n_o}{n_o}, \quad \bar{r} = \exp\left(\frac{1}{n_r}\sum_{p_i \in R} \log r_i\right), \quad \sigma = \frac{n_s}{M \times N}$$

where n_o and n_r represent the number of visible edges in visible and fused images (reflecting local contrast), $\bar{g}_o$ and $\bar{g}_r$ denote average gradients of original and restored images, and σ represents the proportion of newly saturated pixels from over-enhancement. Visible images are denoised before evaluation to mitigate noise effects. Generally, larger e and $\bar{r}$ values indicate more recovered texture details and clearer images, while higher σ suggests more detail loss from overflow.

Table 1 demonstrates that our method achieves clear advantages in enhancing real traffic and darkroom scenes, with higher visible edge counts and average gradient ratios. Reference [9]' s algorithm fails when near-infrared and visible reflections diverge, producing smoothed, detail-lacking results sometimes inferior to BM3D. Reference [14] avoids color cast but underutilizes near-infrared information (e.g., ineffective headlight halo suppression in Figure 7). Our method successfully avoids color distortion and halo artifacts while fully extracting near-infrared information, achieving brightness and clarity close to flash images with the best e and $\bar{r}$ scores. Our higher σ values stem from incorporating more near-infrared shadows rather than over-enhancement.

2.3 Running Time

Our method offers significant performance advantages. Reference [9]' s variational approach requires multiple iterations for optimal fusion. Reference [14]' s deep learning method employs a U-Net with dozens of convolution operations, making it computationally expensive. For an $M \times N$ image, our algorithm has time complexity $O(MN)$ independent of other parameters, making it substantially faster than other methods with linear scalability for larger inputs. Implemented on a HiSilicon embedded platform using SIMD instructions for 32-pixel parallel processing, our method processes 4096$×$2160 HD images in 31 ms, meeting real-time video processing requirements.

---

3 Conclusion

This paper presents a low-light image enhancement method based on differential feature fusion of near-infrared and visible images. By observing imaging differences between near-infrared and visible light, we propose constructing fusion weights from differential features to maximize near-infrared brightness utilization, enhancing image signal-to-noise ratio and producing results approaching near-infrared flash brightness style while preserving authentic visible colors. The method is simple and easily optimized for embedded platforms. Compared with recent research, our approach demonstrates clear advantages in both effectiveness and real-time performance suitable for industrial applications.

The differential feature-based fusion method offers a universal principle for multi-source heterogeneous image restoration. Future work will explore its application in other image restoration domains such as near-infrared/visible image dehazing and flash/non-flash image joint denoising.

References

[1] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. IEEE Trans on Image Processing, 2007, 16 (8): 2080-2095.

[2] Zhang Kai, Zuo Wangmeng, Chen Yunjin, et al. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising [J]. IEEE Trans on Image Processing, 2017, 26 (7): 3142-3155.

[3] Guo Xiaojie, Li Yu, Ling Haibin. LIME: low-light image enhancement via illumination map estimation [J]. IEEE Trans on Image Processing, 2017, 26 (2): 982-993.

[4] Chen Chen, Chen Qifeng, Xu Jia, et al. Learning to see in the dark [C]// Proc of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE Press, 2018: 3291-3300.

[5] Wang Ping, Sun Zhenming. Multi-level decomposition Retinex low-light image enhancement algorithm [J]. Application Research of Computers, 2020, 37 (04): 1204-1209.

[6] Liu Shouxin, Long Wei, Li Yanyan, et al. Low-light image enhancement based on HSV color space [J]. Computer Engineering and Design, 2021, 42 (09): 2552-2560.

[7] Krishnan D, Fergus R. Dark flash photography [J]. ACM Trans on Graphics, 2009, 28 (3): 96.

[8] Zhuo Shaojie, Zhang Xiaopeng, Miao Xiaoping, et al. Enhancing low light images using near infrared flash images [C]// Proc of IEEE International Conference on Image Processing. Piscataway, NJ: IEEE Press, 2010: 2537-2540.

[9] Shen Xiaoyong, Yan Qiong, Xu Li, et al. Multispectral joint image restoration via optimizing a scale map [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2015, 37 (12): 2518-2530.

[10] Jee S, Kang M G. Sensitivity improvement of extremely low light scenes with RGB-NIR multispectral filter array sensor [J]. Sensors, 2019, 19 (5): 1065.

[11] Su Haonan, Jung C, Yu Long. Multi-spectral fusion and denoising of color and near-infrared images using multi-scale wavelet analysis [J]. Sensors, 2021, 21 (11): 3610.

[12] Li Zhuo, Hu Haimiao, Zhang Wei, et al. Spectrum characteristics preserved visible and near-infrared image fusion algorithm [J]. IEEE Trans on Multimedia, 2021, 23: 306-319.

[13] Li Hui, Wu Xiaojun. Densefuse: a fusion approach to infrared and visible images [J]. IEEE Trans on Image Processing, 2019, 28 (5): 2614-2623.

[14] Tang Chaoying, Pu Shiliang, Ye Pengzhao, et al. Fusion of low-illuminance visible and near-Infrared images based on convolutional neural networks [J]. Acta Optica Sinica, 2020, 40 (33): 31-39.

[15] Jung C, Zhou Kailong, Feng Jiawei. Fusionnet: multispectral fusion of RGB and NIR images using two stage convolutional neural networks [J]. IEEE Access, 2020, 8: 23912-23919.

[16] Mei Lin, Jung C. Deep fusion of RGB and NIR paired images using convolutional neural networks [C]// Proc of IEEE International Conference on Pattern Recognition. Piscataway, NJ: IEEE Press, 2021: 10245-10252.

[17] Li Guofa, Lin Yongjie, Qu Xingda. An infrared and visible image fusion method based on multi-scale transformation and norm optimization [J]. Information Fusion, 2021. 71: 109-129.

[18] Tomasi C, Manduchi R. Bilateral filtering for gray and color images [C]// Proc of IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE Press, 1998: 839-846.

[19] He Kaiming, Sun Jian, Tang Xiaoou. Guided image filtering [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35 (6): 1397-1409.

[20] Hautiere N, Tarel J P, Aubert D, et al. Blind contrast enhancement assessment by gradient ratioing at visible edges [J]. Image Analysis and Stereology, 2011, 27 (2): 87-95.

Note: Figure translations are in progress. See original paper for figures.