

An RGB-NIR Image Fusion Method for Improving Feature Matching

Hanhoon Park*

Department of Electronic Engineering, Pukyong National University, Busan, Korea

Received 12 January 2020; received in revised form 05 March 2020; accepted 14 May 2020

DOI: <https://doi.org/10.46604/ijeti.2020.5177>

Abstract

The quality of RGB images can be degraded by poor weather or lighting conditions. Thus, to make computer vision techniques work correctly, images need to be enhanced first. This paper proposes an RGB image enhancement method for improving feature matching which is a core step in most computer vision techniques. The proposed method decomposes near-infrared (NIR) image into fine detail, medium detail, and base images by using weighted least squares filters (WLSF) and boosts the medium detail image. Then, the fine and boosted medium detail images are combined, and the combined NIR detail image replaces the luminance detail image of an RGB image. Experiments demonstrate that the proposed method can effectively enhance RGB image; hence more stable image features are extracted. In addition, the method can minimize the loss of the useful visual (or optical) information of the original RGB image that can be used for other vision tasks.

Keywords: image fusion, feature matching, near-infrared (NIR), detail boosting

1. Introduction

With the increasing availability and deployment of imaging sensors operating in multiple spectral bands, the methods of fusing images observed in various spectrum bands, including both visible (RGB) and near-infrared (NIR) bands, have been extensively explored and widely used in remote sensing and surveillance applications [1-7]. As each spectral band provides a different type of information, the spectral bands could be used selectively and enhanced by different bands.

In particular, as the NIR band is especially useful under poor weather or lighting conditions, RGB images can be enhanced in their clearness, contrast, and details by fusing with NIR images. During the last decade, various RGB-NIR fusion methods have been proposed for different applications. For example, Schaul et al. [8] proposed a dehazing method that decomposed RGB and NIR images into multi-resolution representations by using weighted least squares filters (WLSF) and fused the detail images of RGB and NIR at each level to maximize the contrast of hazy color images. Jang and Park [9] proposed a similar dehazing method where a Gaussian low-pass filter was used for image decomposition. Fredembach et al. [10] proposed a skin smoothing method that separated RGB and NIR images into base and detail layers by using a fast bilateral filter (fast-BF) of [11] and replaced the detail layer of RGB images with that of NIR images. Salamati and Süssstrunk [12] proposed a material-based image segmentation method that formed an intrinsic image which is independent of illumination from R, G, and B to NIR ratio images and took the intersection between the segmentations obtained from the intrinsic image and those obtained from NIR images that are independent of color. Honda et al. [13] proposed a denoising method that denoised the difference between NIR and RGB images and added back the denoised difference image to the original RGB image. Su and Jung [14] proposed another denoising method that generated the wavelet scale map and used gradients of NIR

* Corresponding author. E-mail address: hanhoon.park@pknu.ac.kr

Tel.: +82-51-6296225; Fax: +82-51-6296210

wavelet coefficients as the weight for WLSF-based smoothing in the scale map. Li et al. [15] proposed a RGB-NIR fusion method for improving the classification accuracy in citrus fruit identification. The method fused the contrast from RGB images and the homogeneity from NIR images using an entropy filter and a discrete wavelet transformation. Jiang et al. [16] proposed a multi-spectral image classification method using a double-channel convolutional neural network (CNN) model. The features of the RGB and NIR bands were calculated in two different CNNs, and subsequently, the feature fusion was performed in the fully connected layer. Liu et al. [17] also used CNNs to extract multi-layered features from the RGB and NIR images. Kurihara et al. [18] proposed a photoplethysmography-based heart rate estimation method which is robust to challenging illumination conditions by computing the confidence scores of RGB and NIR images. They used the fact that NIR images can be captured and independent of changes in the background illuminations. By measuring correlations between signals extracted from background and face regions, they adaptively utilized RGB and NIR images for heart rate estimation. Li and Wu [19-20] proposed an RGB-NIR fusion method by using deep learning. They used a network comprising encoder, fusion layer, and decoder. NIR and RGB images were first fed into the encoder. Features maps were obtained by convolutional blocks, then fused in the fusion layer by using two different strategies (addition and L1-norm). Finally, the fused image was reconstructed by decoder network which has the fused feature map as input. Sharma et al. [21] proposed a RGB-NIR fusion method for robust face or palmprint recognition. The method separated NIR images into base and detail layers by using fast-BF and WLSF respectively, and the NIR detail images were obtained by taking an average/maximum of the detail layers from both filters. Then, in order to reconstruct the final enhanced RGB-NIR images, the NIR detail images were fused with the RGB base images separated by WLSF. Their method produced more high-quality image features and thus could improve the face or palmprint recognition accuracy.

This paper also proposes the RGB-NIR fusion method. Similar to the conventional RGB-NIR fusion method (i.e., [21]), the purpose of the proposed method is to enhance RGB images so that more stable image features can be extracted from them. However, unlike the conventional RGB-NIR fusion method [21], the proposed method tries to fuse the RGB and NIR images more carefully. NIR images usually have much luminance details, and thus fusing RGB images with NIR images can enhance the details of RGB images, which makes more image features extracted. Nevertheless, the detail enhancement may result in undesired artifacts or noise, which makes the features match more difficult. In addition, RGB images lose their information contents (e.g., texture or color details) which are useful for other vision tasks. Therefore, the method tries to reduce the difference between the enhanced and the original RGB images as much as possible. This is because we do not aim at fusing RGB and NIR images on the fly for a given application, but making enhanced RGB images used later for a variety of computer vision applications. The method was inspired by the WLSF-based detail enhancement method in [22] which enabled to boost the image details at different scales. The method attempts to boost the medium details of NIR images and to replace the RGB luminance detail images with the boosted NIR detail images to reproduce the enhanced RGB images.

The remaining of this paper is organized as follows. In Section 2, WLSF is briefly introduced and the previous method that used the WLSF to enhance the image detail is explained. Then, the proposed method is explained in detail, and its performance is evaluated and compared to other existing methods in Sections 3 and 4 respectively. Finally, conclusion and future works are presented in Section 5.

2. Detail Boosting Using WLSF

2.1. WLSF

WLSF is a non-linear and edge-preserving smoothing filter enabling to construct multi-scale decompositions that capture details at a variety of scales as explained in [22]. It smooths a gray-scale image g with a few changes in pixel values as possible. Formally, a filtered image u can be obtained by minimizing the energy function as follows.

$$u = WLSF(g, \lambda) = \min_u \sum_p [(u_p - g_p)^2 + \lambda(a_p(\partial u / \partial x)_p^2 + b_p(\partial u / \partial y)_p^2)] \quad (1)$$

where the subscript p denotes the spatial location of a pixel and λ is a constant that controls the level of smoothness. The smoothness weights a and b depend on g and are defined as:

$$a_p = \left(\left| \frac{\partial \ln(g)}{\partial x} \right|_p^\alpha + \varepsilon \right)^{-1} \quad (2)$$

$$b_p = \left(\left| \frac{\partial \ln(g)}{\partial y} \right|_p^\alpha + \varepsilon \right)^{-1} \quad (3)$$

where $\ln()$ is the natural logarithm function, the exponent α determines the sensitivity to the gradients of g , and ε is a small constant that prevents division by zero. By rewriting Eq. (1) using matrix notation, u is the unique solution of a linear system.

2.2. Detail boosting

By smoothing an image using WLSF with progressively increasing λ s, the detail images at multi-scales can be obtained by subtracting smoothed images with the next larger λ s from the original or smoothed images. Letting u_1, u_2, \dots, u_k progressively denote coarser versions of a gray-scale input image g , and the k -level detail images are defined as:

$$d^i = u^{i-1} - u^i \quad (4)$$

where

$$i = 1, 2, \dots, k \text{ and } u^0 = g \quad (5)$$

Then, the detail images can be boosted while avoiding artifacts and using a boosting function, which is defined as:

$$S(\delta_b, d_p) = 1 / (1 + e^{-\delta_b d_p}) \quad (6)$$

where d is a detail image and δ_b is a boosting factor that controls the level of boosting (refer to [22]). The sigmoid function in Eq. (6) avoids the hard clipping that would otherwise occur when the detail layers are significantly boosted. Finally, the boosted detail images are added up to reproduce an enhanced image.

3. Proposed Method

The proposed method enhances an RGB image by fusing it with an NIR image. To enhance an RGB image so that more stable image features can be extracted from it and minimize its change after enhancement, more cares, e.g., how to extract the good-quality detail images from the NIR image and how to combine the NIR detail images with the RGB image, should be taken on. In this paper, we separate out two levels of detail images from the NIR image, properly boost one of them, and reproduce the NIR detail image by weighted-summing the two detail images again. Then, we substitute the RGB luminance detail image with the reproduced NIR detail image.

Fig. 1 illustrates the entire procedure of the proposed RGB-NIR image fusion method. The method first transforms RGB image into YCbCr image. Then, it smooths the luminance (Y) image by using WLSF. The smoothed image serves as base image, and a detail image is obtained by subtracting the base image from the original Y image. NIR image is also smoothed by using WLSF but with two different λ s. The smoothed image with the larger λ serves as base image (I_{NIR}^B). Thus, a medium detail image is obtained by subtracting the base image from the smoothed image with the smaller λ , and a fine detail image is obtained by subtracting the smoothed image with the smaller λ from the original NIR image as follows (refer to Fig. 2).

$$I_{NIR}^B = WLSF(I_{NIR}, \lambda_2) \quad (7)$$

$$I_{NIR}^{MD} = WLSF(I_{NIR}, \lambda_1) - I_{NIR}^B \quad (8)$$

$$I_{NIR}^{FD} = I_{NIR} - WLSF(I_{NIR}, \lambda_1) \quad (9)$$

$$\lambda_1 < \lambda_2 \quad (10)$$

The medium detail image is boosted by using Eq. (6) and scaled by 0.5. Next, the NIR fine and boosted medium detail images are combined with the Y base image, which results in a new Y image. Finally, the new Y image and the Cb and Cr images are transformed to an enhanced RGB image. In Eq. (6), by adjusting δ_b , the proposed method controls the boosting strength and thus the difference between the enhanced and the original RGB images.

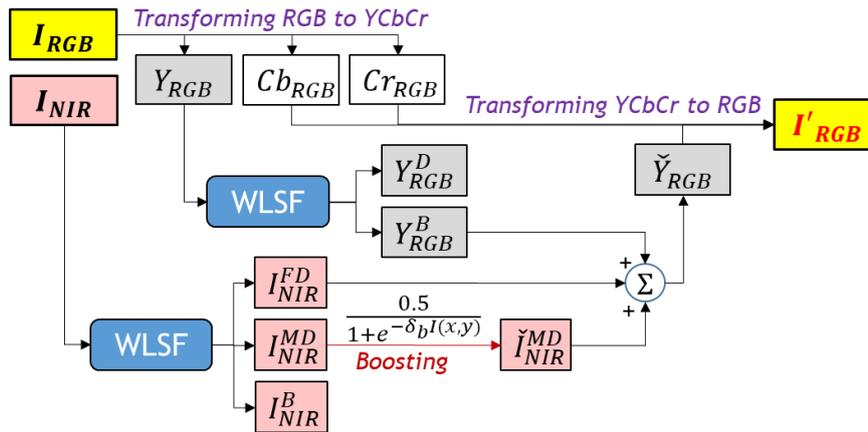


Fig. 1 Process flow of the proposed method

In Fig. 2, an NIR image was decomposed into base, medium detail, and fine detail images. The values in detail images were normalized to $[0, 255]$ for clearly showing the contents. Notice that the fine detail image contains the information related to the scene texture or detailed shape, while the medium detail image contains the information mainly related to the scene structure or shape. The fine detail (also containing noise) can be significantly changed by image transformation. Therefore, boosting the fine detail image may cause decrease in the similarity between an RGB image and its transformations, which indicates that less features are matched between the RGB image and its transformations. This is why only the medium detail image is boosted.



Fig. 2 Image decomposition by WLSF

4. Experimental Results and Discussion

We evaluated the proposed method on images from an RGB-NIR Scene Dataset used in [23] (refer to the examples in Fig. 3). The dataset has 9 categories (*country, field, forest, indoor, mountain, oldbuilding, street, urban, and water*) and 954 (477 RGB and 477 NIR) images. The size of most images is close to 1024×700 . We conducted the same experiments as done in [21]. To quantify the feature quality, the number of SIFT features [24] that were extracted in common from both the images, and its transformations was counted; their matching scores, the squared Euclidean distance between the matched feature

descriptors, were computed. Then, the performance of the proposed method was compared with other fusion methods, i.e., [8, 10, 21] which were named Fre_BF, Sch_WLS, BF_WLS_Viv, and BF_WLS_Avg, respectively. Fre_BF used fast-BF [10]; Sch_WLS used WLSF [8]; BF_WLS_V and BF_WLS_Avg used both fast-BF and WLSF [21]. We applied synthetic transformations, such as rotation (45°, 90°, and 180°) and scaling (0.5, 0.75), to each image of the dataset. The feature matching which used vl_sift and $vl_ubcmatch$ with default parameters in the Vlfeat library (refer to [25]) was carried out between the original and transformed image pairs.



Fig. 3 A part of the images used in the experiments

The number of matched features has changed from hundreds to a few thousands depending on the image contents and transformations. Then, we computed the changes (in %) in the number of matched features and the matching scores, between before and after enhancing the RGB images using each fusion method. We also applied random sample consensus (RANSAC) with homography [26-27] to remove the outliers (as done in [21]) and computed the changes (in %) in the numbers of inliers and the matching scores achieved by using each fusion method. Finally, in order to evaluate whether each fusion method can maintain the information content of the original RGB images, we computed the peak signal-to-noise ratio (PSNR) in dB between the enhanced and the original RGB images. The parameters used in each fusion method are shown in Table 1, which are the same as those used in [21].

Table 1 Parameters of each fusion method used in the experiments

| Fusion method | Parameters |
|-------------------|--------------------------------------------------------------------------------------------|
| Fre_BF ([10]) | All parameters (for space and range Gaussian functions) are adapted for each image |
| Sch_WLS ([8]) | $\lambda = 0.1$, $c = 2$, $\alpha = 1.2$, and $n^\dagger = 6$ |
| BF_WLS_Viv ([21]) | $\lambda = 0.125$, $\alpha = 1.2$, and $n = 1$ |
| BF_WLS_Avg ([21]) | $\lambda = 0.125$, $\alpha = 1.2$, and $n = 1$ |
| Proposed | $\lambda_1 = 0.05$, $\lambda_2 = 0.5$, $\lambda_y = 0.05$, $\alpha = 1.2$, and $n = 1$ |

$^\dagger n$ is the total number of levels of decomposition

Fig. 4 shows the intermediate images produced in each processing step of the proposed method. The detail images were normalized to [0, 255] for clearly showing the contents. The enhanced luminance image \check{Y}_{RGB} has better texture than the original luminance Y_{RGB} .

Fig. 5 shows some of fusion results. Tables 2-4 show the feature matching and PSNR results of each fusion method. In [21], BF_WLS_Avg performed the best among the existing fusion methods. It was the same in our experiments, although the results of each method were slightly different, Fre_BF increased the number of feature matches by 4% but had the worst performance in terms of matching scores. Its matching scores even got worse in most images, which indicates that the feature matching got worse qualitatively. Sch_WLS had poor performance in increasing the number of feature matches and matching scores. Furthermore, its PSNR was too low, which indicates that Sch_WLS severely loses the information content of the original RGB images (refer to Fig. 5).

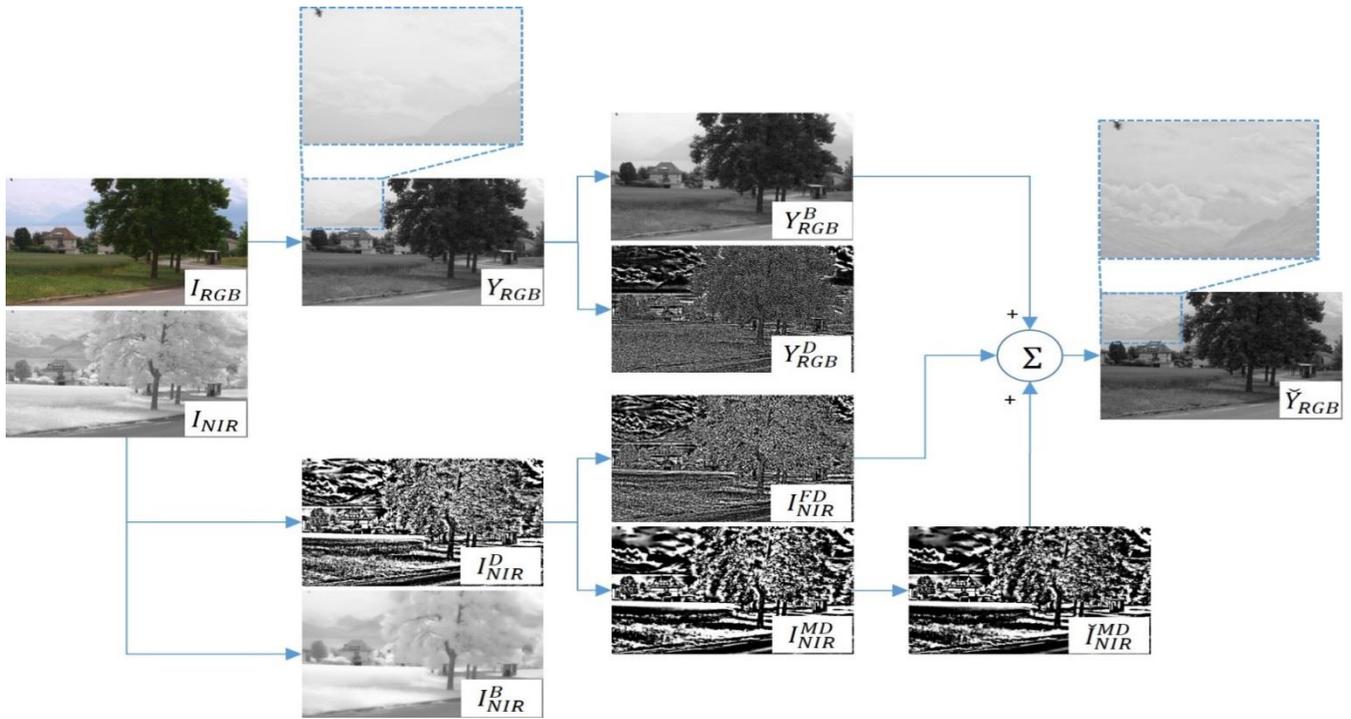


Fig. 4 Intermediate images produced in each processing step of the proposed method



Fig. 5 Visual comparison of RGB images enhanced by different RGB-NIR fusion methods

The performance of the proposed method depends on the boosting factor δ_b . As shown in Fig. 6, larger boosting factors increased the number of feature matches at an expense of a loss in the PSNR. In addition, the increase in the number of feature matches is not correlated with the increase in matching scores. On the contrary, smaller boosting factors retained high PSNR, but the number of feature matches are not increased as much as BF_WLS_Avg in most images. Therefore, we set the boosting factor so that the proposed method worked well in terms of both feature matching and PSNR. The optimal boosting factor is not largely different across different images. For the dataset used in our study, the proposed method worked the best in terms of both feature matching and PSNR when the boosting factor was set to 4.

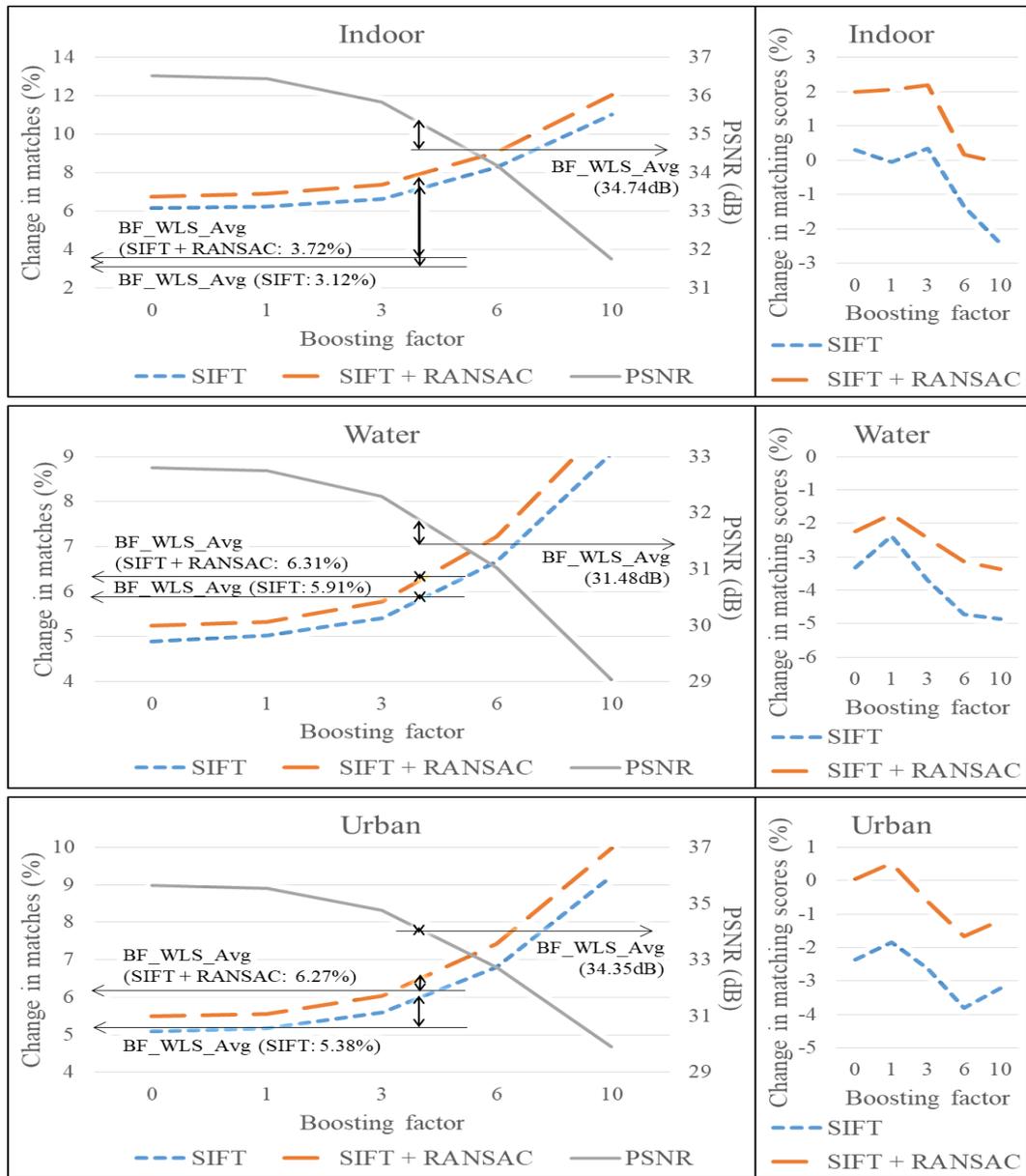


Fig. 6 Performance of the proposed method adjustable by the boosting factor

In Fig. 5 and Tables 2-4, the results of the proposed method are those obtained when the boosting factor is set to 4. The proposed method can extract better NIR details, which is exactly useful for feature matching, by boosting and combining details extracted at different scales. As a result, the proposed method outperforms BF_WLS_Avg in terms of both feature matching and PSNR. Specifically, the method results in increased feature matches over RGB images by 7.17% (with SIFT) and 7.64% (with SIFT + RANSAC) on average while retaining the PSNRs to more than 33dB. The increase in feature matches was approximately 0.3% larger than BF_WLS_Avg, and the PSNR was approximately 0.2dB better than BF_WLS_Avg. The improvement by the proposed method might not be that much. However, the proposed method worked more consistently across images than BF_WLS_Avg (refer to the standard deviation values in Table 2).

Table 2 Changes in the numbers of matched features

| Fusion method | Using SIFT (%) | Using SIFT + RANSAC (%) |
|-----------------------------|-----------------|-------------------------|
| Fre_BF ([10]) | 3.83 ± 2.39 | 4.02 ± 2.39 |
| Sch_WLS ([8]) | 1.43 ± 0.86 | 1.88 ± 0.96 |
| BF_WLS_Viv ([21]) | 1.56 ± 3.04 | 1.70 ± 3.08 |
| BF_WLS_Avg ([21]) | 6.84 ± 4.22 | 7.31 ± 4.13 |
| Proposed ($\delta_b = 4$) | 7.17 ± 2.22 | 7.64 ± 2.25 |

In Table 3, the decrease in matching scores ensures that the feature matching was not just improved numerically but also qualitatively. In terms of matching score, the method was also comparable to BF_WLS_Avg, but with better consistency. In addition, depending on applications, we can adjust the results of the method by changing the boosting factor.

Table 3 Changes in the matching scores

| Fusion method | Using SIFT (%) | Using SIFT + RANSAC (%) |
|-----------------------------|------------------|-------------------------|
| Fre_BF ([10]) | -0.88 ± 2.30 | -0.36 ± 1.34 |
| Sch_WLS ([8]) | -1.09 ± 1.72 | -0.49 ± 0.53 |
| BF_WLS_Viv ([21]) | -1.24 ± 0.63 | -0.18 ± 1.34 |
| BF_WLS_Avg ([21]) | -3.26 ± 0.86 | -2.83 ± 1.60 |
| Proposed ($\delta_b = 4$) | -3.25 ± 0.68 | -2.06 ± 0.50 |

Table 4 PSNR and processing time resulted from each fusion method

| Fusion method | PSNR (dB) | Time (s) |
|-----------------------------|------------------|----------|
| Fre_BF ([10]) | 32.77 ± 1.58 | 0.10 |
| Sch_WLS ([8]) | 19.89 ± 4.27 | 19.13 |
| BF_WLS_Viv ([21]) | 33.52 ± 2.58 | 3.32 |
| BF_WLS_Avg ([21]) | 32.86 ± 1.62 | 3.33 |
| Proposed ($\delta_b = 4$) | 33.01 ± 1.74 | 4.67 |

BF_WLS_Viv, where RGB and NIR images were decomposed into base and detail images by using WLSF, and the RGB detail image replaced by the NIR detail image (refer to [21]) had high PSNRs but showed the worst performance in increasing the feature matches. It indicates that a simple use of WLSF cannot improve the performance for feature matching. BF_WLS_Viv is worse than Fre_BF that uses fast-BF instead of WLSF at the same framework. Although the proposed method also uses WLSF, its performance is much better than BF_WLS_Viv. This is because the proposed method more precisely decomposes the NIR details and boosts the details in an effective manner.

However, the thing we need to know is that NIR images do not always have better details than RGB images as shown in Fig. 7. For example, in far water regions, NIR images lost the underwater details. Therefore, instead of simply substituting the RGB luminance detail image with the enhanced NIR detail image, which of the NIR image and the RGB images contains better local details and should be determined first. Depending on the results, the image fusion should be done adaptively and locally. This would be a good direction for future improvement of the proposed method.



Fig. 7 Details in NIR and RGB images (The yellow circles represent the regions that have fine details)

All the fusion methods were implemented by using Matlab R2017b. Their processing times on a PC equipped with i7-7700 3.6GHz CPU and 32GB RAM are shown in Table 4. WLSF was relatively time-consuming than BF, thus the processing time was in proportion to the number of running WLSF. Therefore, in terms of processing time, Fre_BF was the best. Since the proposed method needs to run WLSF once more compared to BF_WLS_Viv and BF_WLS_Avg, it was a little slower than BF_WLS_Viv or BF_WLS_Avg.

5. Conclusion

This paper proposed an RGB-NIR image fusion method that decomposed RGB and NIR image by using WLSF into detail and base images respectively and substituted the RGB detail image with the NIR detail image. The NIR detail image was obtained in two different levels: fine detail and medium detail. Then, the medium detail image was boosted in an analytical manner. The proposed method outperformed the existing fusion methods in improving the performance for feature matching and maintaining the information content of the original RGB image which can be useful for other vision tasks. Specifically, the increase in feature matches was approximately 0.3% larger, and the PSNR was approximately 0.2dB better than the state of the art. In addition, the proposed method worked much more consistently.

In our experiments, we set the boosting factor that controls the strength of detail boosting. However, the optimal boosting factor can be different across images or applications. Therefore, we are currently studying an automatic way to compute the optimal boosting factor.

In this paper, we assumed that RGB and NIR images were geometrically well aligned. Nevertheless, some images in the dataset used in our experiments were slightly misaligned. Therefore, the method for geometric registration between RGB and NIR images, such as [28-31], is required to improve the fusion results. This remains as future work.

Acknowledgment

This research was supported by the Ministry of Trade, Industry and Energy for its financial support of the project titled “The establishment of advanced marine industry open laboratory and development of realistic convergence content”.

Conflicts of Interest

The authors declare no conflict of interest.

References

- [1] C. Pohl and J. L. V. Genderen, Remote Sensing Image Fusion: A Practical Guide, CRC Press, 2016.
- [2] C. Pohl and J. L. V. Genderen, “Multisensor image fusion in remote sensing: concepts, methods and applications,” International Journal of Remote Sensing, vol. 19, pp. 823-854, March 1998.
- [3] A. E. Hayes, R. Montagna, and G. D. Finlayson “New applications of Spectral Edge image fusion,” Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XXII, vol. 9840, p. 984009, 2016.
- [4] M. Ehlers, Multi-image fusion in remote sensing: spatial enhancement vs. spectral characteristics preservation, Lecture Notes in Computer Science, 2008.
- [5] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, J. Hai, and K. He, “A survey of infrared and visual image fusion methods,” Infrared Physics & Technology, vol. 85, pp. 478-501, July 2017.
- [6] G. He, J. Ji, D. Dong, J. Wang, and J. Fan, “Infrared and visible image fusion method by using hybrid representation learning,” IEEE Geoscience and Remote Sensing Letters, vol. 16, no. 11, pp. 1796-1800, April 2019.
- [7] J. Wang, B. Huang, H. K. Zhang, and P. Ma, “Sentinel-2A image fusion using a machine learning approach,” IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 12, pp. 9589-9601, December 2019.
- [8] L. Schaul, C. Fredembach, and S. Süsstrunk, “Color image dehazing using the near-infrared,” 2009 16th IEEE International Conference on Image Processing, IEEE Press, November 2009, pp. 1629-1623.

- [9] D. W. Jang and R. H. Park, "Colour image dehazing using near-infrared fusion," *IET Image Processing*, vol. 11, no. 8, pp. 587-594, April 2017.
- [10] C. Fredembach, N. Barbuscia, and S. Süsstrunk, "Combining visible and near-infrared images for realistic skin smoothing," *Color and Imaging Conference, 17th Color and Imaging Conference Final Program and Proceedings*, pp. 242-247, January 2009.
- [11] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," *Proc. of the 29th annual conference on Computer graphics and interactive techniques*, July 2002, pp. 257-266.
- [12] N. Salamati and S. Süsstrunk, "Material-based object segmentation using near-infrared information," *Color and Imaging Conference, 18th Color and Imaging Conference Final Program and Proceedings*, vol. 2010, no. 1, pp. 196-201, January 2010.
- [13] H. Honda, R. Timofte, and L. V. Gool, "Make my day - high-fidelity color denoising with near-infrared," *Proc. of the IEEE conference on computer vision and pattern recognition workshops*, pp. 82-90, 2015.
- [14] H. Su and C. Jung, "Multi-spectral fusion and denoising of RGB and NIR images using multi-scale wavelet analysis," *2018 24th International Conference on Pattern Recognition, IEEE Press*, August 2018, pp. 1779-1784.
- [15] P. Li, S.-H. Lee, H. Y. Hsu, and J.-S. Park, "Nonlinear fusion of multispectral citrus fruit image data with information contents," *Sensors*, vol. 17, no. 1, pp. 142, January 2017.
- [16] J. Jiang, X. Feng, F. Liu, Y. Xu, and H. Huang, "Multi-spectral RGB-NIR image classification using double-channel CNN," *IEEE Access*, vol. 7, pp. 20607-20613, January 2019.
- [17] Y. Liu, L. Dong, Y. Ji, and W. Xu, "Infrared and visible image fusion through details preservation," *Sensors*, vol. 19, 4556, November 2019.
- [18] K. Kurihara, D. Sugimura, and T. Hamamoto, "Adaptive fusion of RGB/NIR signals based on face/background cross-spectral analysis for heart rate estimation," *IEEE International Conference on Image Processing*, September 2019, pp. 4534-4538.
- [19] H. Li and X. J. Wu, "DenseFuse: a fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614-2623, December 2018.
- [20] H. Li, X. J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," *Proc. of IEEE, International Conference on Pattern Recognition*, August 2018, pp. 2705-2710.
- [21] V. Sharma, J. Y. Hardeberg, and S. George, "RGB-NIR image enhancement by fusing bilateral and weighted least squares filters," *Journal of Imaging Science & Technology*, vol. 61, no. 4, pp. 1-9, August 2017.
- [22] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Transactions on Graphics*, vol. 27, pp. 2205-2221, August 2008.
- [23] M. Brown and S. Süsstrunk, "Multispectral SIFT for scene category recognition," *Proc. of CVPR, USA*, 2011.
- [24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, November 2004.
- [25] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," *Proceedings of the 18th ACM international conference on Multimedia*, October 2010, pp. 1469-1472.
- [26] R. Hartely and A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge University Press, January 2003.
- [27] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381-395, January 1981.
- [28] S. J. Chen, H. L. Shen, C. Li, and J. H. Xin, "Normalized total gradient: a new measure for multispectral image registration," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1297-1310, November 2017.
- [29] N. Ofir, S. Silberstein, H. Levi, D. Rozenbaum, Y. Keller, and S. D. Bar, "Deep multi-spectral registration using invariant descriptor learning," *2018 25th IEEE International Conference on Image Processing*, October 2018, pp. 1238-1242.
- [30] X. Zhao, J. Zhang, C. Yang, H. Song, Y. Shi, X. Zhou, D. Zhang, and G. Zhang, "Registration for optical multimodal remote sensing images based on FAST detection, window selection, and histogram specification," *Remote Sensing*, vol. 10, no. 5, p. 663, April 2018.
- [31] R. Zhu, D. Yu, S. Ji, and M. Lu, "Matching RGB and infrared remote sensing images with densely-connected convolutional neural networks," *Remote Sensing*, vol. 11, no. 23, p. 2836, November 2019.

