Visible and Near Infrared Image Fusion Based on Texture Information
Yongpeng Wu\textsuperscript{a}, Hairong Zhang\textsuperscript{b}, Fan Qin\textsuperscript{c}, Yang Liu\textsuperscript{c}

\textsuperscript{a}Southern Marine Science and Engineering Guangdong Provincial Laboratory, Zhanjiang, 524000, China; \textsuperscript{b}College of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, 524000, China; \textsuperscript{c}College of Instrumentation & Electrical Engineering, Jilin University, Changchun, 130021, China

\textsuperscript{*}Corresponding author: oceanliew@live.com

ABSTRACT

In this paper, we proposed a novel visible and near-infrared image fusion (VIS-NIR fusion) method based on texture filter, aiming at the problems of artifact, color distortion and noise in traditional VIS-NIR fusion methods. Firstly, the structure information of the visible (VIS) image and the near infrared (NIR) image after texture removal is obtained by relative total variation (RTV) calculation as the base layer of the fused image; secondly, the image noise weight is calculated by establishing a Bayesian classification model, and the noise information in the VIS information is adaptively filtered by joint bilateral filtering; finally, the fused image is acquired by color space conversion. The experimental results show that the proposed algorithm can preserve the spectral characteristics and the unique information of VIS and NIR images without artifacts and color distortion, and has good robustness as well as preserving the unique texture.

Keywords: signal processing, image fusion, near infrared, noise reduction, texture

1. INTRODUCTION

At present, image processing through a single light source is continuously developing to multi-spectral image processing technology. To ensure the imaging requirements under different lighting conditions, image fusion can take advantage of the characteristics of each light source to produce images that are more conducive to human visual perception \cite{1} and has powerful advantages in target detection, object recognition. VIS cameras can capture clear and colorful visible light information, but the intensity of the light can affect the clarity of the captured target and is more susceptible to interference. However, NIR cameras are less affected by external conditions and capture more information about the structure and texture of the target. Therefore, the fusion of the images captured by the two light sources can solve the visual interference caused by different lighting conditions and produce images with high robustness and better quality.

Currently, most image fusion algorithms use a multi-scale transform to extract and analyze feature information in the target image and then calculate the weight coefficients between the source images as specific fusion strategies. Multi-scale transformation is usually divided into two categories: pyramid transformation in the spatial domain and frequency domain. The pyramid transformation method has many advantages, such as highlighting the details of pictures \cite{2}-\cite{3}, and keeping specific scale information. Methods in the frequency domain include wavelet transform \cite{4}, dual-tree discrete wavelet transform (DTCWT) \cite{5}, nonsampled contourlet transform (NSCT) \cite{6}, etc. After completing the multi-scale transformation, it is necessary to reconstruct the information of each scale with specific fusion strategies. The maximum absolute rule is adopted to fuse low-frequency information \cite{7}, and the weighted average underlying information fusion technology based on visual saliency map \cite{3}.

The existing NIR and RGB image fusion methods are eliminated the impact of different lighting conditions on the image. Li et al. \cite{8} proposed a method to preserve the spectral characteristics with the reflection and diffusion transport models. The proposed model can compensate for the visible loss spectral details caused by light scattering. This algorithm can improve the quality of image visibility and avoid image distortion. However, it causes partial loss of NIR texture information. Shibata et al. \cite{9} proposed a fusion method based on Markov Random Field (MRF) energy learning to reduce artefacts caused by geometry and illumination inconsistency, but color distortion occurred in the fusion image. Under low light conditions, there is noise in the images taken by the visible camera cause of insufficient light, but the near-infrared camera is not easily affected. Mei et al. \cite{10} established a convolutional neural network including three subnetworks:
DenoisingNet, EnhancingNet and FusionNet to eliminate the noise of noisy pictures under low illumination, but a huge challenge is to obtain a suited and adaptive data set.

Although the current algorithms have made progress, these algorithms do not maintain the spectral characteristics of the visible image, the fused image appears color distortion, and some algorithms will appear artefacts at the edge of the image. In the case of insufficient illumination, it may cause noise in the visible image, so it is essential to remove the influence of noise on the fusion image as well as improve the overall robustness and visual effect of the image.

To solve these problems, this paper proposes a novel and effective fusion framework based on image texture information extraction, automatically fusing NIR and RGB images with Gaussian noise and noiseless environments, and retaining information specific to each light source. Section II introduces a texture filter based on relative total variance and Joint bilateral filter. In Section III, the workflow of the algorithms in this paper is presented. In Section IV, the performance of the algorithms in this paper is evaluated. Finally, Section V concludes the paper.

2. EDGE PRESERVE FILTER

2.1 Texture filter

Texture filter is a filter that retains the main structure of the image edge by smoothing texture, which makes subsequent image processing easier. The texture filter based on relative total variation (RTV) believes that the main edge of local window has a more similar direction gradient than the complex texture part \(^{[11]}\). A new windowing variant form is adopted to effectively extract the main structure of the image. To highlight the contrast between texture and structure in the prominent area, a more powerful regularization term of structure and texture decomposition is formed.

2.2 Joint bilateral filter

Joint bilateral filtering (JBLF) is used as a nonlinear filter \(^{[12]}\), which achieves edge-preserving and noise-reducing smoothing effects. It adds the guide image on top of the bilateral filter, and the whole smoothing process is made more robust and stable by the intervention of the guide image to get better results. It is defined as follows:

\[
S_p = \frac{1}{k} \sum_{q \in N(p)} f(p, q) g(G_p, G_q) I_q
\]

where \(S_p, G_p, I_p\) represent the smoothed, guided and input images under \(p\) pixels, respectively. \(p = p(p_x, p_y)\), \(q = (q_x, q_y)\) are the spatial coordinates of the two pixels, and \(f = (\cdot)\) and \(g = (\cdot)\) are the spatial and color distances between the two pixels.

3. THE PROPOSED ALGORITHM

The VIS-NIR fusion method proposed in this paper mainly includes the following four steps: (1) separation of structure and texture information, (2) calculation of noise weight, (3) noise removal, and (4) image fusion. The Fig.1 shows the workflow.

Figure 1: The workflow of our algorithm.
3.1 Separation of structure and texture information

One image can be divided into two parts, the primary structure and the texture, as shown in equation (2):

\[ I_C = T_C + S_C \]  

where \( C \in (vis, nir) \) represents the visible and near infrared channels, \( I_C \) represents the original image, \( S_C \) is the structure image, \( T_C \) expresses the final coarse texture image, whose details are smoothed out by RTV texture filtering. The significant structure \( S_{vis} \) of visible images can be extracted by texture structure separation. \( T_{vis} \) is a coarse texture image that needs to be denoised and \( V_{nir} \) is used to calculate the luminance channel of the fused image.

3.2 Calculation of noise weight

In low light condition, there will be a lot of noise in the VIS image collected by the camera. The coarse texture image \( T_{vis} \) needs to be reprocessed to separate the fine texture from the noise to extract the effective information.

Zhang et al. \[13\] mentioned in decomposition and differentiation of noise and texture that different components of images may have diverse local variances. The formula for local variance is as follows:

\[ P_N = \left( \frac{1}{N} \right) \int \left( f(x) - m_N \right)^2 f(x) \, dx = \frac{1}{N} \int f(x) \, dx \]  

(3)

Where \( f(x) \) represents the local area of the image, \( N \) is the scale information of this part, and \( m_N f \) represents the mean value of this area. For different parts of the image, it has the following relationship: \( P_{smooth} > P_{texture} > P_{edge} > P_{noise} \).

In other words, the noise part has a weaker variance compared with other regions of the image. We consider the property of noise as a prerequisite for separating fine texture information from noise. However, plentiful texture and noise are mixed in the RGB image, and the local variance image obtained in this way cannot be straightly used as the evaluation standard for judging the robust texture and noise, so the feature factors need to be extracted through reprocessing.

We use histogram to calculate the distribution of the local variance of each picture, divide the local variance of each image into several parts, solve the maximum value of the first-order gradient and obtain the coordinate information where the value is located as the two feature elements of the image. The two feature elements are put into the naive Bayes model for classification, and the prediction label of each image is obtained, which is marked as a noisy image or noiseless image.

Firstly, we select a 3×3 window to traverse the whole image and calculate the local variance of the brightness channel of the visible image through Formula (3) to obtain the local variance image \( L \). Then we normalize it. Finally, we take the histogram \( h \) of \( L \). The expression for acquiring the maximum value of the \( h \) first-order gradient and its index is as follows:

\[ f_1 = \max \nabla h \]  

(4)

\[ f_2 = \arg \max \nabla(h) \]  

(5)

where we set a training instance \( l = (x, c), x = (f_1, f_2) \) is the characteristic attribute, and \( c \) is the category attribute divided into noisy and noise-free.

3.3 Noise removal

After the Bayesian model trained in Section 3.2, the following formula is used to calculate the probability \( P_{NOI} \) and \( P_{DNOI} \) of the input image in the noisy and noiseless data.

\[ P_T = \frac{1}{2\pi\delta_1\delta_2\sqrt{1-\rho^2}} e^{-\left( \frac{1}{2(1-\rho^2)} \left[ \frac{(x_1-\mu_1)^2}{\delta_1^2} + \frac{(x_2-\mu_2)^2}{\delta_2^2} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\delta_1\delta_2} \right] \right)} \]  

(6)
where $T \in (NOI, DNOI)$ represents two types of noise and noiseless, $x_1$ and $x_2$ respectively represent the normalized $(f_1, f_2)$ of the input image obtained by Section 3.2, $\delta_1$ and $\delta_2$ respectively represent the variances of the two feature vectors under T type, $\mu_1$ and $\mu_2$ represent their mean values, and $\rho$ is the correlation coefficient of the two variables. By default, $\rho$ is 0.5, $\delta_1$ and $\delta_2$, $\mu_1$ and $\mu_2$ are obtained from the training model.

### 3.4 Image fusion

NIR image is a single-channel image and should not be considered as additional color channels, but rather as channels carrying luminance and spatial information $^{[14]}$. Therefore, we need to convert the visible image from the RGB channel to one of the luminance containing channels such as HSV, YCrCb, or YUV channel and use its luminance channel as part of the fusion with the NIR, so that the color of the fused image can be closer to the color of the source image. We have chosen the V channel in HSV to fuse the NIR information to effectively preserve the NIR information and to prevent color distortion. We use $S_{vis}$ obtained from Section 3.1 to get its tones, contrast, and brightness images $H_{rgb}, S_{rgb}, V_{rgb}$. The fusion brightness image is obtained by the following formula.

$$F_v = V_{rgb} + JBLF(T_{vis}, s) + T_{nir} \tag{7}$$

where $T_{nir}$ is the near-infrared image texture obtained by formula (5) and $JBFL(T_{vis}, s)$ is the result of Section 3.3 combined with bilateral filtering. Finally, we convert the fused brightness image $F_v$, $H_{rgb}$ and $S_{rgb}$ into RGB image as the final fusion image.

### 4. EXPERIMENTAL RESULTS

In the experiments, we use two evaluation criteria structural similarity index measure (SSIM) $^{[15]}$, visual information fidelity for fusion (VIFF) $^{[16]}$ to verify the merits of the proposed fusion algorithm by dividing the experimental part into two categories, noiseless and noisy environments.

To evaluate the performance of the proposed algorithm, several contrast experiments were carried out with the state-of-the-art VIS and NIR image fusion methods, including Image Restoration Via Scale Map (VSM) $^{[17]}$, Spectrum Characteristics Preservation (SCP) $^{[8]}$, adaptive and fast image enhancement (LC) $^{[18]}$, Guided Filter for Fusion (GFF) $^{[19]}$, and fusion using Laplacian-Gaussian Pyramid Decomposition (LGPD) $^{[2]}$. To demonstrate that the algorithm can achieve good results in both noisy and noise-free environments, we test it in the noise-free environment Fig.2-Fig.4 and the noisy environment Fig. 5-Fig.7, where the noisy images are obtained by adding Gaussian noise to the noisy images.

#### 4.1 Structural similarity index measure (SSIM)

Wang et al. $^{[15]}$ introduced a complementary framework for quality assessment based on structural similarity. Structural similarity index measure (SSIM) defines structure information from the perspective of image composition, which is different from brightness, contrast, and reflects the attributes of object structure in the scene.
Figure 2: Visual experiment in image Tree (a) VIS image (b) NIR image (c) LGPD [2] (d) GFF [19] (e) LC [18] (f) SCP [8] (g) VSM [17] (h) OUR.

Figure 3: Visual experiment in image Water (a) VIS image (b) NIR image (c) LGPD [2] (d) GFF [19] (e) LC [18] (f) SCP [8] (g) VSM [17] (h) OUR.
Figure 4: Visual experiment in image Mountain (a) VIS image (b) NIR image (c) LGPD\textsuperscript{2} (d) GFF\textsuperscript{19} (e) LC\textsuperscript{18} (f) SCP\textsuperscript{8} (g) VSM\textsuperscript{17} (h) OUR.

Figure 5: Visual experiment in image Forest (a) VIS image (b) NIR image (c) LGPD\textsuperscript{2} (d) GFF\textsuperscript{19} (e) LC\textsuperscript{18} (f) SCP\textsuperscript{8} (g) VSM\textsuperscript{17} (h) OUR.
4.2 Visual information fidelity for fusion (VIFF)

The visual information fidelity for fusion (VIFF) is an index to measure the quality of fused image based on the fidelity of visual information [16], and the multi-resolution image fusion metric using Visual Information Fidelity (VIF) is used to quantitatively evaluate the effect of image fusion methods. This process includes the following four steps. (1) The original image and the output image are filtered respectively and divided into multiple blocks. (2) The distortion of the VIS information of each block is evaluated. (3) The VIFF of each sub-band is accumulated. (4) Superimpose sub-band information to obtain VIFF of the whole image.
In Table 1, our algorithm maintains a high value for VIFF in the noiseless case, because our algorithm is more capable of separating texture details, making the fused image more realistic and maintaining the natural characteristics of the visible light image capably. LC method has better figures than the other methods. LC adds the structural information of the NIR subject to the VIS image, adding richer detail while keeping the source image unchanged, so it has higher values than the other methods, but artefacts can appear at the edges of the image. SCP algorithm is a method of transferring spectral variations by modelling the transmission. This algorithm allows better preservation of the spectral characteristics of VIS images and reduces artefacts, so the VIFF is relatively high. Although GFF has a high SSIM, color distortion is already present in the fused image and it is not conducive to human eye observation. In general, the quality of an image needs to be evaluated from both a subjective and objective point of view.

As for noisy environment in the Table 2, our algorithm maintains better results. Other methods are not effective in removing noise in the presence of noise and cannot remove the noise pollution caused by visible light except the VSM method. The VSM algorithm has good denoising capabilities as it is guiding the VIS image through the NIR, however, it does not take into account that the VIS image may contain unique information, so it fails to retain the rich textural properties of the noisy case, leaving the fusion information incomplete. Compared to the VSM method, our algorithm achieves better results by separating texture from noise in VIS images and retaining their textural properties. Since our algorithm is adaptive, it could be applied in noisy and noise-free environments, and there is no color distortion, so the fused image is more realistic and more natural.

### Table 1. Quantitative comparison of noiseless data

| Image   | Method | LGPD | GFF | LC  | SCP | VSM  | OUR  |
|---------|--------|------|-----|-----|-----|------|------|
| Tree    | SSIM   | 0.5449 | 0.5733 | 0.5762 | 0.5433 | 0.4747 | 0.5851 |
|         | VIFF   | 0.5239 | 0.5736 | 0.6008 | 0.6032 | 0.5036 | 0.6755 |
| Water   | SSIM   | 0.6876 | **0.7317** | 0.7239 | 0.6477 | 0.6916 | 0.7201 |
|         | VIFF   | 0.7586 | 0.8390 | 0.8275 | **0.9185** | 0.7361 | 0.9006 |
| Mountain| SSIM   | 0.6811 | **0.6952** | 0.6778 | 0.6444 | 0.6504 | 0.6734 |
|         | VIFF   | 0.8478 | 0.9194 | 0.8587 | 0.9547 | 0.7905 | **0.9859** |

### Table 2. Quantitative comparison of noise data

| Image | Method | LGPD | GFF | LC  | SCP | VSM  | OUR  |
|-------|--------|------|-----|-----|-----|------|------|
| Forest| SSIM   | 0.1721 | 0.2064 | 0.2226 | 0.1747 | 0.5247 | **0.5643** |
|       | VIFF   | 0.5060 | 0.3153 | 0.3616 | 0.3824 | 0.5484 | **0.7119** |
| Urban | SSIM   | 0.1803 | 0.1863 | 0.1896 | 0.1692 | **0.6653** | 0.6391 |
|       | VIFF   | 0.6736 | 0.5950 | 0.6029 | 0.7031 | 0.7329 | **0.7386** |
| Street| SSIM   | 0.1073 | 0.1041 | 0.1059 | 0.0965 | **0.4582** | 0.4535 |
|       | VIFF   | 0.5067 | 0.6006 | 0.6061 | 0.6755 | 0.6193 | **0.7334** |

### 5. CONCLUSION

In this paper, we propose a visible and near infrared fusion algorithm. In contrast to the multi-scale decomposition type of image fusion, we classify the image into structure and texture by absorbing the texture information of the two source images and no longer subdividing the detail layers. The images are subsequently categorized into noisy and noiseless terms with a Bayesian model and guided by their due environment using a joint bilateral filter for adaptive fusion.

Our algorithm has the following virtues:

1. Compared to previous algorithms, we are not guiding the VIS image directly by NIR image to fusion and denoise. Instead, we maintain the information unique to each of the VIS and NIR.

2. We can retain the spectral characteristics of the color image without causing color distortion and creating artefacts at the edges of the image. In addition, the method in this paper shows better performance than other methods in the results of vision and quantization experiments.
STATEMENT

The current manuscript (Visible and Near Infrared Image Fusion Based on Texture Information) is an article that we have independently completed, and we have not yet found any significant similarities with other people's articles. It should be noted that we have previously submitted an earlier version of this article to the preprint website arxiv. However, arxiv is a preprint website designed to prevent plagiarism during the review process, and articles on it cannot be considered published. Submitting manuscripts to arxiv is not a practice of multiple submissions. The arxiv website is https://arxiv.org/abs/2207.10953#:~:text=In%20this%20paper%2C%20a%20novel%20visible%20and%20near-infrared,traditional%20visible%20and%20near%20infrared%20image%20fusion%20methods.

Meanwhile, we assure that any potential influence from websites containing information about the arXiv preprint of this article such as https://www.researchgate.net/publication/362229890_Visible_and_Near_Infrared_Image_Fusion_Based_on_Texture_Information et al. will be eliminated.

REFERENCES

[1] C. Pohl, J. L. Van Genderen, Review article multisensor image fusion in remote sensing: Concepts, methods and applications, International Journal of Remote Sensing 19 (5), 823–854(1998). https://doi.org/10.1080/014311698215748
[2] Vannali, A.V., Gadre, V.M. Visible and NIR image fusion using weight-map-guided Laplacian–Gaussian pyramid for improving scene visibility. Sādhanā 42, 1063–1082 (2017). https://doi.org/10.1007/s12046-017-0673-1
[3] J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, Infrared Physics Technology 82, 8–17(2017). https://doi.org/10.1016/j.infrared.2017.02.005.
[4] S. G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis Ma chine Intelligence 11 (4) (1989). https://doi.org/10.1109/34.192463.
[5] Y. Zuo, J. Liu, G. Bai, W. Xuan, M. Sun, Airborne infrared and visible image fusion combined with region segmentation, Sensors (Basel, Switzerland) 17 (5) (2017). https://doi.org/10.3390/s17051127
[6] J. Adu, J. Gan, Y. Wang, J. Huang, Image fusion based on nonsampled contourlet transform for infrared and visible light image, Infrared Physics Technology 61 (Complete) 94–100 (2013). https://doi.org/10.1016/j.infrared.2013.07.010.
[7] J. Hu, S. Li, The multiscale directional bilateral filter and its application to multisensor image fusion, Information Fusion 13 (3), 196–206 (2012). https://doi.org/10.1016/j.inffus.2011.01.002.
[8] Z. Li, H.-M. Hu, W. Zhang, S. Pu, B. Li, Spectrum characteristics preserved visible and near-infrared image fusion algorithm, IEEE Transactions on Multimedia 23, 306–319 (2021). https://doi.org/10.1109/TMM.2020.2978640.
[9] T. Shibata, M. Tanaka, M. Okutomi, Versatile visible and near-infrared image fusion based on high visibility area selection, Journal of Electronic Imaging 25 (1), 013016 (2016). https://doi.org/10.1117/1.JEI.25.1.013016
[10] L. Mei, C. Jung, Deep fusion of rgb and nir paired images using convolutional neural networks, in: 2020 25th International Conference on Pattern Recognition (ICPR), pp. 6802–6803 (2021). https://doi.org/10.1109/ICPR48806.2021.9412411
[11] L. Xu, Q. Yan, Y. Xia, J. Jia, Structure extraction from texture via relative total variation, ACM Transactions on Graphics 31 (6), 1 (2012). https://doi.org/10.1145/2366145.2366158
[12] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, K. Toyama, Digital photography with flash and no-flash image pairs, Acm Transactions on Graphics 23 (3), 664–672 (2004). https://doi.org/10.1145/1015706.1015777
[13] Z. Tao, Q. Fan, Wavelet characterization of dyadic bmo norm and its application in image decomposition for distinguishing between texture and noise, International Journal of Wavelets Multiresolution and Information Processing 9 (3), 445–457 (2011). https://doi.org/10.1142/S0219691311004183
[14] C. Fredembach, S. Süsstrunk, Colouring the near infrared, in: Color and Imaging Conference, 2008.
[15] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861

[16] Y. Han, Y. Cai, Y. Cao, X. Xu, A new image fusion performance metric based on visual information fidelity, Information Fusion 14 (2), 127–135 (2013). https://doi.org/10.1016/j.inffus.2011.08.002

[17] Q. Yan, X. Shen, L. Xu, S. Zhuo, X. Zhang, L. Shen, J. Jia, Cross-field joint image restoration via scale map, in: 2013 IEEE International Conference on Computer Vision, pp. 1537–1544, (2013). https://doi.org/10.1109/ICCV.2013.194

[18] M. Awad, A. Elliethy, H. A. Aly, Adaptive near-infrared and visible fusion for fast image enhancement, IEEE Transactions on Computational Imaging 6, 408–418 (2020). https://doi.org/10.1109/TCI.2019.2956873

[19] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Transactions on Image Processing 22 (7), 2864–2875 (2013). https://doi.org/10.1109/TIP.2013.2244222