Infrared and Visible Image Fusion Based on Spatial Convolution Sparse representation

Luling Shao¹, Jin Wu*¹ and Minghui Wu²

¹ School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan, Hubei, 430081, China
² Electronic Information School, Wuhan University, Wuhan, Hubei, 430072, China
*Corresponding author’s e-mail: wujin@wust.edu.cn

Abstract. In the traditional sparse representation based infrared and visible image fusion method, the detail information is not able to be effectively extracted, resulting in the decrease of infrared target intensity and the blurry of visible background information. In order to solve the above issue, a new image fusion method based on spatial convolution sparse representation is proposed. Firstly, a spatial convolution sparse representation is used to perform two-scale decomposition of infrared and visible images by introducing a gradient regularization, and the detail and intensity information are extracted effectively from the source images. Then, the weighted average rule and the maximum selection rule are used to fuse the base layer and detail layer images, respectively. Finally, the fusion image is constructed. Experimental results illustrate that the proposed method is superior to the traditional fusion method based on sparse representation.

1. Introduction

Infrared imaging can highlight objects by infrared radiation, even in poor visual environments such as low light, smoke, and camouflage. Visible imaging can capture the reflected light of the object, which contains rich details and texture information. The fusion of infrared and visible images can complement their advantages and disadvantages, which not only can retain the highlighted signal of infrared images, but also retain the details of visible images. It has been used in the military[1], detection[2], remote sensing[3], and other fields.

In the last decades, many methods have been designed to fuse the infrared and visible images[4-7], and sparse representation (SR) is one of them. The SR uses as few atoms as possible to represent the source image in a given overcomplete dictionary, which can effectively obtain the information contained in the image and improve the quality of the fused image.

Yang et al.[8] first applied the SR to the fusion of infrared and visible images. Then, researchers have proposed many improved fusion methods based on the SR. Wang et al.[9] proposed a fusion method based on a non-negative SR model, which effectively analyses the image features through the activity level of the representation coefficients. Yin et al.[10] introduced a joint sparse representation (JSR) to extract different components of the source image to achieve fusion. Besides, Liu et al.[11] found that an over-complete dictionary can increase calculation complexity and noise sensitivity. Thus, they proposed a clustering-based dictionary construction method to reduce the redundancy of the dictionary and improve the robustness of noise image fusion.
In the above method, the construction of the dictionary and sparse representation is mainly performed on independent image patches or sets of image patches, which may cause the spatial consistency of the same pixel in different image patches to be ignored. Further, some details may be smooth or lost in the fused image[12]. Therefore, Liu et al.[12] proposed a fusion method based on convolutional sparse representation (CSR), and it used a global sparse representation to enhance spatial consistency.

However, there is a drawback in the fusion method based on the CSR. Usually, the source image is divided by a two-scale decomposition pre-processing into the base layer and the detail layer. Then the CSR is used to represent the detail layer. Figure 1 shows the framework of the traditional fusion method based on the CSR. It can be seen that the detail layer acquisition and the CSR are processed separately. In this case, the original image information cannot be effectively expressed under the CSR. As a result, in the final fusion results, the intensity of the infrared target may be reduced or the visible image texture may be blurred.

![Figure 1. The framework of the traditional fusion method based on the CSR](image)

In order to solve the above problems, this paper proposes an infrared and visible fusion method based on spatial convolutional sparse representation. First, in order to effectively extract the details of the source image, this paper proposes a new image decomposition model based on the spatial convolutional sparse representation. This model achieves the separation of the base layer and the detail layer by introducing a gradient regularization. Then, the base layer and the detail layer are fused separately, according to the weighted average rule and the maximum value selection rule. Finally, under the subjective and objective evaluations, the proposed method is compared with traditional image fusion methods based on the sparse representation.

2. Proposed method
The framework of the proposed fusion method in this paper is shown in figure 2. First, the spatial convolutional sparse representation (SCSR) is used to decompose the infrared and visible images into the base layer and detail layer. Then, according to the weighted average rule and the maximum selection value, the base layer and the detail layer are fused separately. Finally, the fused image is constructed.
2.1. Spatial convolutional sparse representation for image decomposition

In [13], Papyan et al. proposed a spatial convolutional sparse representation model, the specific form as follow:

\[ I = D\Omega \]  \hspace{2cm} (1)

Where \( D \in \mathbb{R}^{N \times N_m} \) is a banded convolution dictionary, \( \Omega \in \mathbb{R}^{N_m} \) is a global sparse representation coefficient. Further, they formulated Equation (1) to be:

\[ I = \sum_{i=1}^{N} U^T_i D_i \varphi_i \]  \hspace{2cm} (2)

Where \( U^T_i \in \mathbb{R}^{N \times m} \) is the operator that puts \( D_i \varphi_i \) in the \( i \)-th position and pads the rest of the entries with zeros. Let \( s_i = D_i \varphi_i \), then Equation (2) can be changed to:

\[ I = \sum_{i=1}^{N} U^T_i s_i \]  \hspace{2cm} (3)

According to this theory, we proposed a two-scale decomposition model as follow:

\[
\begin{align*}
\min_{D_b, D_d, \varphi_b} & \quad \frac{1}{2} \left\| I - D_b \varphi_b - I^d \right\|_F^2 + \lambda \left\| \Omega_d \right\|_1 + \xi \left\| \varphi^d \right\|_1 \\
\text{s.t.} & \quad \varphi_b = D_b \varphi_b,
\end{align*}
\]  \hspace{2cm} (4)

Where image \( I \) is decomposed into base layer \( I^b \) and detail layer \( I^d \) (\( D_d \Omega_d \)), \( D_d \) is a convolution dictionary, and \( \Omega_d \) is its corresponding sparse vector. The gradient constraint mentioned in [12] is used as the regularization term for the base layer extraction. The expression of \( \left\| \varphi^d \right\|_1 \) is:

\[
\text{arg min}_{\varphi^d} \left\| I - I^b \right\|_F^2 + \eta \left( \left\| g_x * I^b \right\|_1 + \left\| g_y * I^b \right\|_1 \right)
\]  \hspace{2cm} (5)

Where \( g_x = [-1,1] \) and \( g_y = [-1,1]^T \) are the horizontal and vertical gradient operators, respectively; \( \eta \) is the regularization parameter. Equation (4) can be equivalent to:

\[
\begin{align*}
\min_{D_b, D_d, \varphi_b, \varphi^d} & \quad \frac{1}{2} \left\| I - \sum_{i=1}^{N} U^T_i s_i^b - I^d \right\|_F^2 + \lambda \sum_{i=1}^{N} \left\| \varphi^d_j \right\|_1 + \xi \left\| \varphi^d \right\|_1 \\
\text{s.t.} & \quad s_i^b = D_b \varphi_b, \quad I^b = I^d
\end{align*}
\]  \hspace{2cm} (6)
Where we let $I^b = L_d^b$ to facilitate the minimization over the $GR$ norm. Its corresponding ADMM formulation is given by:

$$
\min_{d,\phi_d} \frac{1}{2} \left\| I - \sum_{i=1}^{N} U_i^T s_d^i - I^b \right\|^2_F + \sum_{i=1}^{N} \left( \frac{\lambda}{2} \left\| s_d^i - D_i \phi_d^i + u_d^i \right\|_1^2 + \frac{\eta}{2} \left\| t^b - L_d^b + V_d^b \right\|_1^2 + \xi \left\| t^b \right\|_{GR} \right)
$$

(7)

Where $u_d^i$ and $V_d^b$ are the dual variable of $s_d^i$ and $t^b$, respectively. For Equation (7), we can modify algorithm 1 given in [13] to obtain $s_d^i = D_i \phi_d^i$ and $t^b$.

2.2. Spatial convolutional sparse representation for image decomposition

In order to achieve the fusion of infrared and visible images, the proposed decomposition method in 2.1 is first used to decompose the infrared image $I_{IR}$ and visible image $I_{VS}$ at two scales. After the decomposition, we can obtain: the detail layer $I_{IR}^d = \sum_{i=1}^{N} U_i^T D_i \phi_{Id}$ and the base layer $I_{IR}^b$ of the infrared image, and the detail layer $I_{VS}^d = \sum_{i=1}^{N} U_i^T D_i \phi_{VId}$ and the base layer $I_{VS}^b$ of the visible image.

For detail layer fusion, we adopt the maximum value selection rule. The expression is:

$$
\phi_{Id}^f = \begin{cases} 
\phi_{Id}^d & \text{if } \left\| \phi_{Id}^d \right\|_1 > \left\| \phi_{VId}^d \right\|_1 \\
\phi_{VId}^d & \text{else}
\end{cases}
$$

(8)

$$
I_{Id}^f = \sum_{i=1}^{N} U_i^T D_i \phi_{Id}^f
$$

(9)

Where $I_{Id}^f$ represents the fused detail layer, $\phi_{Id}^f$ represents the fusion coefficient of the corresponding detail layer, $\| \cdot \|_1$ represents the $\ell_1$ norm.

For base layer fusion, we use the weighted average rule. The expression is:

$$
I_{Ib}^f = w_1 * I_{IR}^b + w_2 * I_{VS}^b
$$

(10)

Where $w_1$ and $w_2$ are weights, and set as $w_1 = w_2 = 0.5$.

2.3. Fusion image construction

According to the detail layer fusion and base layer fusion, we construct the fused image $I_f$:

$$
I_f = I_{Id}^f + I_{Ib}^f
$$

(11)

3. Experiments

In our experiments, five pairs of visible and infrared images from the TNO Image Fusion Dataset are used as test images, which are shown in figure 3. In addition, we compare our method with current representative fusion algorithms, including SR-based fusion method[8], JSR-based fusion method[10], ASR-based fusion method[11], CSR-based fusion method[12].
3.1. Subjective evaluation

Figure 4 shows the fusion result of the first pair of source images. It can be seen that in the SR, JSR, and ASR, the texture of the ground is almost lost. Although the CSR retains part of the texture, the intensity of the infrared object is weak. Our method has a good performance on the preservation of the intensity and detail.

Figure 5 gives the fusion result of the second pair of source images. It can be found that our method clearly presents the building and the human in front of it. In the SR, JSR, and ASR, the building is integrated with the environment, which can hardly be identified. In addition, although the CSR contains visible details, the building is not as salient as ours.

Figure 6 shows the fusion result of the third pair of source images. It can be seen that our method can highlight the sentinel and retain details such as fence and bushes. In the SR, JSR, and ASR, the details of the fence and bushes are blurry. Besides, the intensity of the sentry is not salient in the CSR.
The comparison results demonstrate that our method can preserve the infrared target and visible details effectively.

3.2. Objective evaluation
In order to objectively and comprehensively evaluate the performance of the fusion method, four objective metrics, namely, the mutual information (MI)[14], the weighted fusion quality index (Qw)[15], the visual information fidelity fusion metric (VIFF)[16] and the standard deviation (STD), are adopted to evaluate fusion performance. In this paper, the objective evaluation index is calculated using the code provided in [17].

The objective evaluation indexes of the five pairs of the source images are shown in Table 1. In our method, the MI is slightly lower than the CSR, and the other indicators are optimal.

Table 1. The objective evaluation indexes of the five pairs of the source images.

| Indexes | Method | I      | II     | III    | IV     | V      |
|---------|--------|--------|--------|--------|--------|--------|
| MI      | SR     | 2.1857 | 1.8305 | 1.5683 | 2.3323 | 2.0650 |
|         | JSR    | 2.1325 | 1.6874 | 1.5085 | 2.2563 | 2.0706 |
|         | ASR    | 2.2139 | 1.8343 | 1.5873 | 2.3527 | 2.0469 |
|         | CSR    | 2.3204 | 1.9031 | 1.6147 | 2.3987 | 2.0843 |
|         | Ours   | 2.3119 | 1.9640 | 1.6393 | 2.4436 | 2.0956 |
| Qw      | SR     | 0.7891 | 0.8179 | 0.7366 | 0.8237 | 0.8106 |
|         | JSR    | 0.7742 | 0.7820 | 0.7429 | 0.8123 | 0.7748 |
|         | ASR    | 0.8001 | 0.8210 | 0.7557 | 0.8427 | 0.8140 |
|         | CSR    | 0.8195 | 0.8258 | 0.8083 | 0.8634 | 0.8149 |
|         | Ours   | 0.8284 | 0.8361 | 0.8156 | 0.8747 | 0.8289 |
| VIFF    | SR     | 0.2727 | 0.1892 | 0.2435 | 0.1549 | 0.3077 |
|         | JSR    | 0.2548 | 0.1614 | 0.2400 | 0.1500 | 0.2510 |
|         | ASR    | 0.2830 | 0.1952 | 0.2632 | 0.1720 | 0.3094 |
|         | CSR    | 0.3007 | 0.2111 | 0.3426 | 0.2016 | 0.3210 |
|         | Ours   | 0.4042 | 0.3554 | 0.5102 | 0.2987 | 0.4514 |
| STD     | SR     | 32.8481| 31.4279| 23.3819| 27.2381| 30.2247|
|         | JSR    | 31.5955| 27.8666| 22.2013| 25.7130| 27.5492|
|         | ASR    | 33.1058| 31.3788| 23.6471| 27.6668| 30.4366|
|         | CSR    | 33.3670| 30.7343| 24.7646| 27.7245| 30.7639|
|         | Ours   | 36.8500| 35.6237| 28.7490| 30.0651| 36.7528|

The MI indicates that our method retains the source image information effectively. For example, the visible details and infrared intensity on the fusion results in figure 6 are well preserved. The Qw reflects that our fusion image is more similar to the source image and the details are clear. For example, the ground texture details in figure 4 are clear. The VIFF shows that our fusion image is suitable for human visual perception. The STD indicates that our fusion images have good contrast.

4. Conclusion
This paper focuses on the fusion of infrared and visible images and proposes a fusion method based on spatial convolutional sparse representation. This method can perform global representation on local image patches with a gradient constraint, so it can effectively extract the details and intensity information of the source image. In this case, the quality of the fused image is improved. Experimental results illustrate that the proposed method is superior to the current image fusion method based on sparse representation.
References

[1] Muller, A., Narayanan, S. (2009) Cognitively-engineered multisensor image fusion for military applications. Information Fusion, 10(2): 137-149.

[2] Bulanon, D., Burks, T., Alchanatis, V. (2009) Image fusion of visible and thermal images for fruit detection. Biosystems Engineering, 103(1): 12-22.

[3] Li, H., Ding, W., Cao, X., et al. (2017) Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing. Remote Sensing, 9: 441.

[4] Adu, J., Wang, M., Wu, Z., et al. (2012) Infrared image and visible light image fusion based on nonsubsampled contourlet transform and the gradient of uniformity. International Journal of Advancements in Computing Technology, 4(5): 114-121.

[5] Li, S., Kang, X., Hu, J. (2013) Image fusion with guided filtering. IEEE Transactions on Image processing, 22(7): 2864-2875.

[6] Ma, J., Chen, C., Li, C., et al. (2016) Infrared and visible image fusion via gradient transfer and total variation minimization. Information Fusion, 31: 100-109.

[7] Liu, Y., Liu, S., Wang, Z. (2015) A general framework for image fusion based on multi-scale transform and sparse representation. Information Fusion, 24: 147-164.

[8] Yang, B., Li, S. (2012) Pixel-level image fusion with simultaneous orthogonal matching pursuit. Information Fusion, 13: 10-19.

[9] Wang, J., Peng, J., Feng, X., et al. (2014) Fusion method for infrared and visible images by using non-negative sparse representation. Infrared Physics & Technology, 67: 477-489.

[10] Yin, H., Li, S. (2011) Multimodal image fusion with joint sparsity model. Optical Engineering, 50(6): 7007(1-10).

[11] Liu, Y., Wang, Z. (2014) Simultaneous image fusion and denoising with adaptive sparse representation. IET Image Processing, 9(5): 347-357.

[12] Liu, Y., Chen, X., Ward, R., et al. (2016) Image fusion with convolutional sparse representation. IEEE signal processing letters, 23(12): 1882-1886.

[13] Panyan, V., Romano, Y., Sulam, J., et al. (2017) Convolutional dictionary learning via local processing. In: Proceedings of the IEEE International Conference on Computer Vision. Venice. pp.5306-5314.

[14] Qu, G., Zhang, D., Yan, P. (2002) Information measure for performance of image fusion. Electronics Letters, 38: 313-315.

[15] Piella, G., Heijmans, H. (2003) A new quality metric for image fusion. In: Proceedings 2003 International Conference on Image Processing. Barcelona. pp.173-176.

[16] Sheikh, H., Bovik, A. (2006) Image information and visual quality. IEEE Transactions on Image Processing, 15(2): 430-444.

[17] Liu, Z., Blasch, E., Xue, Z., et al. (2012) Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study. IEEE Transactions on Pattern Analysis Machine Intelligence, 34(1): 94-109.