Multi-Modal Image Fusion via Sparse Representation and Multi-Scale Anisotropic Guided Measure

SHUAI ZHANG, FUYU HUANG, HUI ZHONG, BINGQI LIU, YICHAO CHEN, AND ZIANG WANG

1Department of Electronic and Optical Engineering, Army Engineering University, Shijiazhuang 050003, China
2PLA ARMY 32137, Zhangjiakou 075600, China

Corresponding author: Fuyu Huang (hfyoptics@163.com)

This work was supported by the Natural Science Foundation of China under Grant 61801507.

ABSTRACT The multi-modal image fusion plays an important role in various fields. In this paper, a novel multi-modal image fusion method based on robust principal component analysis (RPCA) is proposed, which consists of low-rank components fusion and sparse components fusion. In the low-rank components fusion part, a universal low-rank dictionary is constructed for sparse representation (SR) and the low-rank fusion is converted to sparse coefficients fusion by adopting the batch-OMP. In the sparse components fusion part, the anisotropic weight map is constructed to express salient structures of the images. Moreover, a multi-scale anisotropic guided measure is proposed to guide the fusion process, which can extract and preserve the scale-aware salient details of sparse components. Finally, the multi-modal fusion can be achieved by combining two fusion parts together. The experimental results validate that the proposed method outperforms nine state-of-the-art methods in multi-modal fusion both at gray-gray and gray-color scales, in terms of qualitative and quantitative evaluations.

INDEX TERMS Multi-modal image fusion, robust principal component analysis, sparse representation, multi-scale anisotropic guided measure.

I. INTRODUCTION

With the development of the modern technology, the requirement for the completeness of the information acquisition is increasing, so the multi-sensors play an important role in many fields. To better obtain the composite image for further visual and processing tasks, the image fusion has become a research hotspot and been widely employed in computer vision, military surveillance, medical imaging, remote sensing, and so on [1]–[5].

Current image fusion methods are mainly divided into spatial domain based and transform domain based according to their processing domain [6], [7]. In spatial domain, the fused image can be constructed through the combination of the input images at pixel-level or block-level. These methods mainly select salient pixels or regions with higher clarity to fuse the multi-modal images [8]. The direct fusion in pixel-level will lead to decreasing the edge contrast, and the region fusion effect relies too much on segmentation of different details. Meanwhile, such methods are often subject to noise interference and blocking artifacts though their computation complexity is simple [9]. Zhang proposed to fuse infrared and visible images through infrared feature extraction and visual information preservation, and it is especially designed for low-light circumstance [10]. In this fusion type, the image details may be preserved well, however, the continuity is hardly guaranteed.

The transform based methods always adopt the way of multi-scale analysis. Generally, these methods decompose the image into different subbands to extract edge details or salience structures at different scales. Then, the image fusion is converted to these subbands fusion by using some rules. Finally, the fused image can be obtained through inverse multi-scale transform. The major advantage of such methods is that more details of multi-modal images are well preserved. The traditional methods include curvelet transform (CVT) [11], [12], discrete wavelet transform (DWT) [13], [14], shearlet transform [15]–[17],
dual-tree complex wavelet transform (DTCWT) [18], [19] and nonsubsampled contourlet transform (NSCT) [20]–[22]. Different transformation methods have their own unique advantages on features preservation, and many researches are performed to find optimal combination ways. Li and Yang [23] proposed a multifocus image fusion method by combining CVT and wavelet transform. Chen et al. [24] proposed an image fusion method based on neighborhood characteristic and regionalization in NSCT domain. However, the fused effect heavily depends on the fusion rule and many rules don’t take the global features consistency into account. So finding an optimal transform basis is not so easy to realize. Zhao designed a visual weight map to extract saliency in decomposed scales [25]. Feng proposed a total variation to retain the texture details and maintain the edges in infrared and visible image fusion [26]. Zhu proposed a new activity measure consisting of phase congruency, sharpness change and local sharpness change measure to realize the multiple highpass subbands fusion with high informative [27]. Zhu also proposed a total-variation based method to accomplish fusion on cartoon and texture components, respectively [8]. These studies above better extract the details and structures on transformation space rather on original image through designing corresponding extraction rules.

Additionally, the sparse representation (SR) becomes popular in fusion due to exploiting the self-similarity of the natural images [28], [29] and revealing the intrinsic nature of human visual system (HVS) characteristics. The main idea of the SR is that a given signal can be represented by a linear combination of a few atoms in an overcomplete dictionary [30]. Yang firstly introduces the SR into image fusion in [31]. Wei adopts the SR in the hyper-spectral and multispectral image fusion [32]. Liu proposed the image fusion method with convolutional sparse representation [33], and so on. Meanwhile, the Group SR, Robust SR and Non-negative SR are proposed in recent years. However, the fusion effect of the SR-based methods relies on the dictionary to a great extent. The diversity of the high-frequency signals restricts the dictionary establishment. Moreover, the SR based on single image component lacks of flexibility in fusion. Liu proposed a convolutional sparsity based morphological component analysis and achieved multi-component and global SRs of source images [34]. Thus, the SR based fusion of multi-component can decrease the complexity of the dictionary and improve fusion adaptive.

Motivated by these methods above, we study the fusion scheme under the robust principal component analysis (RPCA) [35] to puruse more complete background and abundant structure information in this paper. On one hand, instead of adopting SR on original image, the SR based fusion is performed on low-rank components of RPCA result to preserve background more flexibly and robust. In this way, the low-rank component overcomplete dictionary can be constructed easily with higher representative and university, and the problem of low-rank fusion is converted to find the best fusion way of the sparse coefficients encoded by batch-OMP (orthogonal matching pursuit). On the other hand, a multi-scale anisotropic guided measure is proposed to guide the sparse components fusion in the paper. Specifically, the anisotropic weight map is constructed by combining the Laplace filtering, dilation operation and heat diffusion. Then, multi-scale guided map is constructed by using the designed multi-scale guided filter and taking the original images as the guidance image. Finally, the sparse components fusion is completed with the guidance of multi-scale anisotropic guided map. Thus, the structure information can be preserved according to its contrasts with a certain direction at multi-scale sparse components, which is not fully considered in many existing coefficients fusion ways.

Both multi-scale structures information and global spatial consistency are preserved in the proposed measure.

Specifically, the contributions of the proposed fusion scheme can be summarized as:

- An overcomplete dictionary of low-rank components is constructed and it gives a rise to the robust and universal representation of the low-rank components. The batch-OMP is utilized to enhance the efficiency of sparse coding.
- The anisotropic weight map is constructed, which well respects the anisotropic directional structures and salient objects of the images. Moreover, a multi-scale guided filter is designed to extract the scale-aware structures and improve the spatial consistency.
- A multi-scale anisotropic guided measure is proposed to guide the sparse components fusion, which integrates the original and complementary information into final fusion result.

The rest of the paper is organized as follows. Section II presents the proposed methodology in detail. The experiment results and analysis are presented in Section III. Finally, the conclusions are given in Section VI.

II. METHODOLOGY

A. FUSION FRAMEWORK

The key issues in the multi-modal image fusion are how to extract the effective information in different bands and how to combine them together. Taking the color image and the gray image as example, the flow of the proposed fusion method is depicted in Fig. 1. The color map is converted into YUV space, in which Y component represents the signal intensity and it is used for fusion with gray map. In detail, they are firstly decomposed into low-rank components and sparse components, which represent the background and structure details, respectively. The low-rank components are encoded by batch-OMP under the overcomplete dictionary, and their corresponding coefficients are fused to reflect the low-rank components fusion result. Meanwhile, the sparse components are fused using the proposed multi-scale anisotropic guided measure. Finally, the low-rank fusion components, sparse components and chrominance components (U and V) are combined to obtain final fusion image.
B. RPCA DECOMPOSITION MODEL

The core idea of RPCA [36] is that a data matrix can be represented as the superposition of a low-rank component and a sparse component under the low-rank and sparse optimization criteria. Assuming an input matrix, it always contains the background, structure information and noise, so it can be decomposed as:

\[ D = L + S, \quad \text{rank}(L) \ll \min(M, N) \]  

(1)

where \( L \) is a principal matrix known to be low-rank, and \( S \) is sparse for it represents some salient detail and noise. Although the Eq. (1) is NP-hard, the Wright et al. have proposed a best solution by using tractable convex optimization:

\[
\begin{align*}
\min_{L, S} & \quad \|L\|_* + \lambda \|S\|_1 \\
\text{s.t.} & \quad L + S = D
\end{align*}
\]

(2)

where \( \|\cdot\|_* \) represents the nuclear norm of a matrix, \( \|\cdot\|_1 \) is the \( l_1 \)-norm, and \( \lambda \) is a weighting parameter that is usually set to be \( 1/\sqrt{\max(M,N)} \).

As shown in Fig. 2, the objects with high temperature difference can be modeled as the sparse component in the infrared image. A visible image contains abundant scene background information and texture details [36]. To separate the sparse details from the background more accurately, we adopt the IALM (inexact alternating direction methods) instead of ALM (augmented Lagrange multipliers) to solve the convex problem in Eq. (2) [37], in which the exact solution is not need. In RPCA decomposition model, the higher \( \lambda \) is, the higher weight of low-rank components obtains, i.e., the more map information is judged as background and the fusion map lacks of distinct details. As shown in Fig. 3, the IR image are decomposed into low-rank components (first row)
and sparse components (second row) with different $\lambda$. It can be seen that the people, chimney, road and other salient objects are better extracted in sparse components when $\lambda$ is set to 0.02. Therefore, a appropriate $\lambda$ is so important to preserve enough background and structure details. In additional, the $\lambda = \frac{1}{\sqrt{\max(m,n)}}$ is suggested in [38], but it doesn’t work well in our many experiments. We find a more optimal parameter $\lambda$ definition that can realize a well compromise fusion effect of two different components. The $\lambda$ of the original image with lower average value is defined as $\lambda = \frac{10^{-3}}{\min(m,n)}$, and the other is defined as $\lambda = \frac{10^{-3}}{\max(m,n)}$.

C. LOW-RANK COMPONENT FUSION

The basic assumption of the SR is that a signal can be approximately represented by a linear combination of a few atoms from an overcomplete dictionary [23]. Most of existing methods directly obtain natural image patches and train them, but the dictionary sometimes performs unsatisfied because the types and quantity of the training sample are far from enough. Learning-based approach can overcome the problem in dictionary construction, however, it need constant attention of experimenter. The main reason is that the natural image patches are so diverse. The goal of the SR is to find the sparsest coefficients that contain the fewest nonzero atoms among all feasible solution. We find the similarity of the low-rank component patches is higher than natural image patches. So we construct the overcomplete dictionary by training the low-rank patches to improve its robust and flexibility. Supposing there are $K$ training patches $y$ of size $\sqrt{n} \times \sqrt{n}$ and they are arranged to column vectors. Note that their mean values should be subtracted to zero before construction. The K-SVD [39] is selected for dictionary learning for its simplicity and efficiency. The dictionary $D$ satisfies the follows condition:

$$\min_D \sum_{i=1}^{K} \|x_i\|_0 \text{s.t.} \|y_i - D \alpha_i\|_2 < \varepsilon$$

The pre-fused images $I_A$ and $I_B$ are divided into $\sqrt{n} \times \sqrt{n}$ patches and they are ordered as $\{v_i^A, v_i^B\}$ which are expressed using $D$:

$$v_i^A = D \alpha_i^A, v_i^B = D \alpha_i^B$$

where $\alpha_i$ is the sparse coefficient. In this process, OMP is a traditional and effective way for sparse coding. Thus, the original image patches are represented as sparse coefficients in a common dictionary $D$.

The sliding window technique is utilized to divide the training images and pre-fused images into patches of size $\sqrt{n} \times \sqrt{n}$. Referring a typical example [22], we verify the fusion results of $\sqrt{n} = 2, 4, 8, 12, 16$ in dictionary training respectively, and make a compromise between computation and redundancy when $\sqrt{n}$ is defined as 8. After obtaining the dictionary $D$, the batch-OMP is adopted for sparse coding because its complexity is simpler and it doesn’t depend on the dictionary implementation. Due that the low-rank component reflects the basic information of the background, we adopt the maximum principle for sparse coefficients fusion:

$$\alpha_i^F = \max(\alpha_i^A, \alpha_i^B)$$

Then the fused result of $L_F$ is calculated by:

$$L_i^F = D \alpha_i^F$$

D. MULTI-SCALE ANISOTROPIC GUIDED MEASURE

The guided filter is based on a local model, making it qualified for other applications such as image matting, up-sampling and colorization [40], [41]. The guided filter is used to guide the fusion measure of the sparse component because its good edge preserving smoothing ability. In proposed fusion framework, the details guided measure plays an important role in constructing a weight map and guiding the sparse components fusion. To improve the performance
of details guided measure, a multi-scale anisotropic guided measure is introduced in this section.

The guided filtering principle can be described as:

\[ Z_i = a_k Y_i + b_k \forall i \in r_k \]  

(7)

where \( Z_i \) and \( Y_i \) is the ith pixel value of the output image \( Z \) and guidance image \( Y \), \( r_k \) is its local window. The linear coefficients \( a_k \) and \( b_k \) can be estimated by:

\[ a_k = \frac{1}{|r|^2} \sum_{Y_i} X_i \delta_k - \mu_k \Delta_k, \quad b_k = \tilde{X}_k - a_k \mu_k \]  

(8)

where \( X \) is the input image, \( \delta_k \) and \( \mu_k \) are the variance and the mean of \( Y \) in \( r_k \), \( \eta \) is the regularization parameter, \(|r|\) is the number of pixels in \( r_k \), \( \tilde{X}_k \) is the mean of \( X \) in \( r_k \). In Eq. (5), the value of \( Z \) will change when the window size \( r_0 \) of \( r_k \) is different. So, all the values of coefficients \( a_k \) and \( b_k \) are first averaged if there are several different \( r_0 \). Thus, the output image is estimated as:

\[ Z_i = \tilde{a}_i Y_i + \tilde{b}_i \]  

(9)

The weight map is calculated to guide the fusion according to actual pre-fused image feature. However, the low frequency information always causes much interference. Moreover, the details extraction is not so satisfied because many filtering methods are isotropic so that the magnitude and direction are not fully considered in filtering process. To overcome the above two deficits, we propose to construct an anisotropic map. Firstly, the high frequency details \( H_n \) are obtained by applying the Laplace filtering to original images:

\[ H_n = I_n \ast L \]  

(10)

where \( L \) is a 3 x 3 filter referring from [41]. Then, we adopt dilation operation on \( H_n \) to avoid the inner loss of salient object. Therefore, the weight map can better correspond to the decomposition images of the sparse component:

\[ D_n = H_n \ast Se \]  

(11)

where \( Se \) is a dilation template. Next, the anisotropic heat diffusion is utilized to reveal the intrinsic structure and discover the local anatomical structure. The heat diffusion process over the manifold M is governed by the heat equation:

\[ \Delta_M u(x, t) = -\frac{\partial u(x, t)}{\partial t} \]  

(12)

where \( \Delta_M \) is the Laplace-Beltrami operator of M. The heat kernel is one of the generic solutions and defined as:

\[ k(x, y, t) = \sum_{i=0}^{\infty} e^{-\lambda_i t} \phi_i(x) \phi_i(y) \]  

(13)

where \( \lambda_i \) and \( \phi_i \) are the eigenvalue of the Laplacian matrix and its corresponding eigenvector. Thus, the heat diffusion based image \( \Phi_n \) can be obtained that preserve the anisotropic feature.

\[ \Phi_n = \int_M k(x, y, t) I(x, y) \, dx \, dy \]  

(14)

The higher of the value of heat diffusion result is, the more weight of corresponding pre-fused image should take in fusion. Finally, the initial anisotropic weight map \( W_{initial}^t \) for rth pre-fused image are obtained by:

\[ W_{initial}^t(x, y) = \begin{cases} 1 & \text{if } \Phi_t(x, y) = \max(\Phi_t(x, y))_{r=1}^n \\ 0 & \text{else} \end{cases} \]  

(15)

The weight map at first scale is calculated using \( W_{initial} \) map guided by original image, and the next weight map is calculated using the previous result.

\[ W_{MGF}^j = \begin{cases} G(I, W_{initial}) & j = 1 \\ G(I, W_{MGF}^{j-1}) & j > 1 \end{cases} \]  

(16)

where \( I \) is the original image used as guided image, \( W_{initial} \) is the initial anisotropic weight map, \( W_{MGF}^j \) is the guided filtering weight map at jth scale, \( G(\cdot) \) represents the guided filter function.

To explore the relationship between the window sizes with the multi-scale guided filtering results, we take the initial anisotropic weight map and original IR image as the input and guidance map, respectively. In first group of experiment, the scale is set 3, and the window sizes are set 2, 4 and 8. As shown in Fig. 4(c)-(e), the weight maps can extract more distinct details, but brings more smooth background information at the same time. The second group experiment is performed under the window size \( r = 2, 4 \) and 8. The results indicate that the region containing the high-frequency targets are well segmented, and the representation ability of edge details is improved, as shown in Fig. 4(f)-(h).

After obtaining the weight maps, they are treated as multi-scale anisotropic guided measure for sparse component fusion. The multi-modal images are merged through normalized summation at different scale, respectively.

\[ S_F^j = \sum_{k=1}^{n} \text{norm}(W_{MGF}^j(k)) \cdot S_k \]  

(17)

where \( S_F^j \) is the fusion result of the sparse component at jth scale, \( W_{MGF}^j(k) \) is the weight map of image \( k \) at jth scale, \( S_k \) is the sparse component of image \( k \).

Finally, the final fusion result of the sparse component is obtained by select the maximum value among fusion results at l scales.

\[ S_F = \max(S_F^{j=1, 2, \ldots, l}) \]  

(18)

Similarly, the proposed method can be extended to fuse the color and gray image. The direct fusion in RGB of the color image may cause the color distortion. Considering the human perception, the YUV space can be used to encode a color image into one luminance component (Y) and two chrominance components (U and V). Therefore, the YUV color space is adopted to accomplish the gray and color images fusion. The conversion from RGB to YUV is defined.
as follows [42]:
\[
\begin{bmatrix}
Y \\
U \\
V
\end{bmatrix} =
\begin{bmatrix}
0.2568 & 0.5041 & 0.0979 \\
-0.1482 & -0.2910 & 0.4392 \\
0.4392 & -0.3678 & -0.0714
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix} +
\begin{bmatrix}
16 \\
128 \\
128
\end{bmatrix}
\] (19)

Thus, the Y component is treated as the gray component of the color image and used to fusion with the gray image. Finally, the fused color image can be obtained through the inverse YUV transformation.

In color images fusion, such as PET and SPECT, the color images are first converted into YUV space. Thus, the fusion can be treated as luminance components (Y) fusion and chrominance components (U and V) fusion, i.e., the former can be accomplished according to gray-gray fusion method, and the latter can select the maximum value of the chrominance components as the fused results to preserve more color information. Finally, the color-color fused image can be obtained through inverse YUV transformation.

III. EXPERIMENT RESULTS AND ANALYSIS
To demonstrate the multi-modal fusion effectiveness of the proposed method, we validate the effect of the proposed method by conducting lots of experiments. Here we select six typical pairs of multi-modal images including three color-gray pairs to analysis in detail and other ten pairs for fusion results present. The proposed method is compared with nine image fusion methods, i.e., the morphological difference pyramid (MDP) based method [43], the gradient pyramid (GP) based method [44], the SR based method [22], the DTCWT based method [19], the NSCT based method [20], the NSCT pulse coupled neural network based method [45], zhang’s method [10], ASR method [46] and GFCE method [47]. The parameter settings of these methods are as follows. Four decomposition levels, the ‘averaging’ scheme for the low-pass sub-band and the absolute maximum choosing scheme for the high-pass sub-band are selected for the MDP based method, GP based method and DTCWT based method. Specifically, ‘legall’ and ‘qshift_06’ are selected as the DTCWT decomposition filters. For the SR based method, the number of the overlapped pixels is set 6, and the max-L1 is selected for coefficients fusion rule. Three decomposition levels, ‘pyrexc’ for the directional decomposition and ‘vk’ for the pyramidal decomposition are adopted in NSCT based method. Five decomposition levels, ‘9-7’ for the directional decomposition, ‘pkva’ for the pyramidal decomposition and the maximum choosing scheme for high-pass and low-pass are
adopted in NSCT-PCNN based method. The parameters of zhang’s method, ASR method and GFCE method are set up as the default in respective literatures. In our method, the \( \lambda \) is defined according to Section B. The level of multi-scale anisotropic guided measure is set three and its filter window sizes are set 2, 4 and 8, respectively. Note that there will be artifacts if the window size is too small or too large.

The Laplace filter is set as \[
\begin{bmatrix}
0 & 1 & 0 \\
1 & -4 & 1 \\
0 & 1 & 0
\end{bmatrix},
\]
which refers from [41]. The dilation template is set as \[
\begin{bmatrix}
0 & 1 & 0 \\
1 & 1 & 1 \\
0 & 1 & 0
\end{bmatrix}.
\]

All the experiments are implemented using MATLAB 2014b on a notebook PC with a 2.2 GHz Intel Core CPU and 4 GB memory.

A. QUALITATIVE EVALUATION

In image fusion, the qualitative evaluation means visually comparing the fusion images by human vision system (HVS) according to their prior knowledge. In this subsection, twenty people are selected as observers and two representation regions are enlarged as close-ups in each fused image for better comparison.

Fig. 5 shows the fused images obtained with different methods for the “UN camp” images. All the methods can achieve the fusion effect to a certain extent. The MDP based method can only extract details in some regions. For example, there are distortions at the house roof and its edge. The contrast of the human with the background is low to distinguish the human in GP and NSCT based methods. The consistency of whole fused image is poor in the result of the SR based method. There are some aircrafts around the human in the result of the DTCWT based method. There are too many low-frequency signals to highlight the details in ASR method result. In whole view, the proposed method performs better than others comparison methods except GFCE method on energy preservation and detail extraction.

Fig. 6 shows the fused images obtained with different methods for the “Road2” images. The pedestrian disappears in the light in the results of MDP based methods. The image energies are low in the results of the GP based method, NSCT based method and ASR method, in which the contrast of the roof is not obvious enough. The objects are interfered by the lights of all the comparison methods in fused images, such as the car, the roof lamps, the pedestrian, and so on. On the contrary, the proposed method achieves the better fusion result due to more prominent pedestrian and richer background information.

Fig. 7 shows the fused results of the medical images. Artificially distorted traces appear in brain boundary of the fused images of the MDP based method. In NSCT based result, the high-frequency signals are so weak to distinguish them. Similarly, the bright parts in brain of source 2 are more or less lost in all the comparison methods. In contrast, the brain structures and the global spatial consistency are so clearly in the result of the proposed method.

Next, three image pair examples of gray-scale and color-scale are presented in Fig. 8-10. In Fig. 8, the ‘OETEC’ image
pairs are used to verify the fusion effect of all methods. All comparison methods fail to preserve the color fidelity of the sky. The humans in NSCT and NSCT-PCNN based results are so weak to identify. The sky is not so continuous in the GFCE fusion result.

In Fig. 9, the background details of far scene are preserved well in the results of the DTCWT, NSCT,
TABLE 1. Evaluation indices for the Fig. 5.

| Metrics      | $Q_M$ | $Q_G$ | $Q_E$ | $Q_P$ | $Q_T$ |
|--------------|-------|-------|-------|-------|-------|
| MDP          | 0.2327| 0.3942| 0.1732| 0.4346| 0.6774|
| GP           | 0.2882| 0.4098| 0.3099| 0.4988| 0.7622|
| SR           | 0.2720| 0.3561| 0.1417| 0.4141| 0.6777|
| DTCW         | 0.2297| 0.4166| 0.2586| 0.4987| 0.7744|
| NSCT         | 0.2339| 0.3613| 0.2488| 0.2121| 0.7105|
| NSCT-PCNN    | 0.2849| 0.4167| 0.1542| 0.4936| 0.7982|
| Zhang’s method | 0.3816| 0.4408| 0.1175| 0.3481| 0.8315|
| GFCE         | 0.3033| 0.5263| 0.1857| 0.5403| 0.8616|
| ASR          | 0.2904| 0.3474| 0.1530| 0.5363| 0.5181|
| Proposed Method | 0.3157(2) | 0.4443(2) | 0.2759(2) | 0.5558 | 0.8909(3) |

TABLE 2. Evaluation indices for the Fig. 6.

| Metrics | $Q_M$ | $Q_G$ | $Q_E$ | $Q_P$ | $Q_T$ |
|---------|-------|-------|-------|-------|-------|
| MDP     | 0.3545| 0.5328| 0.4023| 0.5969| 0.7511|
| GP      | 0.3837| 0.5322| 0.5613| 0.6091| 0.8225|
| SR      | 0.5818| 0.5253| 0.3960| 0.6600| 0.7835|
| DTCW    | 0.3221| 0.5228| 0.4981| 0.6679| 0.8163|
| NSCT    | 0.4013| 0.4424| 0.3742| 0.4964| 0.7592|
| NSCT-PCNN | 0.3719| 0.5850| 0.3425| 0.6656| 0.8910|
| Zhang’s method | 0.7066| 0.6005| 0.1992| 0.5094| 0.7562|
| GFCE    | 0.3604| 0.5483| 0.3358| 0.6340| 0.8585|
| ASR     | 0.3698| 0.5611| 0.3858| 0.6299| 0.8505|
| Proposed Method | 0.6632(2) | 0.6556 | 0.6067 | 0.6587(4) | 0.9216 |

TABLE 3. Evaluation indices for the Fig. 7.

| Metrics | $Q_M$ | $Q_G$ | $Q_E$ | $Q_P$ | $Q_T$ |
|---------|-------|-------|-------|-------|-------|
| MDP     | 0.5797| 0.4018| 0.2254| 0.3732| 0.6350|
| GP      | 0.5900| 0.3692| 0.4040| 0.4905| 0.6341|
| SR      | 0.7801| 0.4011| 0.2038| 0.5428| 0.6912|
| DTCW    | 0.5484| 0.3645| 0.3040| 0.5157| 0.5937|
| NSCT    | 0.6822| 0.3403| 0.3203| 0.4163| 0.6250|
| NSCT-PCNN | 0.6655| 0.5510| 0.2663| 0.5407| 0.7372|
| Zhang’s method | 0.4316| 0.5766| 0.3361| 0.5656| 0.8814|
| GFCE    | 0.5516| 0.4461| 0.2170| 0.4956| 0.6959|
| ASR     | 0.6638| 0.6352| 0.2688| 0.5381| 0.8329|
| Proposed Method | 0.8669| 0.4532(4) | 0.5289 | 0.5701 | 0.7057(4) |

TABLE 4. Evaluation indices for the Fig. 8.

| Metrics | $Q_M$ | $Q_G$ | $Q_E$ | $Q_P$ | $Q_T$ |
|---------|-------|-------|-------|-------|-------|
| MDP     | 0.7663| 0.7369| 0.2117| 0.4432| 0.6068|
| GP      | 0.6345| 0.2538| 0.3338| 0.6453| 0.1828|
| SR      | 0.8472| 0.7792| 0.3062| 0.5619| 0.7866|
| DTCW    | 0.5632| 0.2446| 0.3179| 0.5452| 0.5578|
| NSCT    | 0.7301| 0.6875| 0.3086| 0.4728| 0.6730|
| NSCT-PCNN | 0.6198| 0.5636| 0.3411| 0.6229| 0.7038|
| Zhang’s method | 0.6806| 0.4147| 0.2772| 0.6367| 0.7674|
| GFCE    | 0.6384| 0.5754| 0.4022| 0.6484| 0.7827|
| ASR     | 0.5079| 0.7473| 0.5772| 0.6245| 0.7876|
| Proposed Method | 0.7873(2) | 0.7468(2) | 0.4155 | 0.6782 | 0.8152 |

NSCT-PCNN based methods, Zhang’s method, GFCE based methods and proposed method. However, the MDP and GP based methods can’t tackle the color information well. The yellow smoke is not discontinuous in NSCT-PCNN based fusion image. Moreover, there is too much noise in the GFCE fusion result, which interferes with the grass and roof edges. The cloud contrast in the sky is so low in the Zhang’s fusion result. The proposed method makes cloud contrast high, preserves more details and performs continuous.

In Fig. 10, a set of MR and PET image fusion results is taken as the example. The color distortions appear in the fused images of comparison method excepting NSCT-PCNN based method, Zhang’s method and GFCE method. The color distortion in SR based result is so high and it is obviously not suitable to observe. The high-frequency signals of source 1 are interfered by color in Zhang’s fusion result. In overview, the NSCT-PCNN based method, GFCE method and proposed method obtain relatively satisfied effects.

B. QUANTITATIVE EVALUATION

In order to assess the fusion more objectively, five fusion metrics, i.e., information theory based metric ($Q_M$), gradient-based metric ($Q_G$), phase congruency based metric ($Q_P$), structural similarity based metric ($Q_E$), Yang’s metric ($Q_Y$) are adopted. The default parameters given in the related publications are adopted for these quality indexes [48], [49]. The quantitative performances of different methods are shown in Table 1-6, in which the bold values indicate the greatest value in each row and mean the best performance among the all methods. We can see from six tables that the proposed method nearly obtains the highest scores in most of the metric values.

Specially, the whole image of the GP based method is smooth, especially at corners and edges, so its phase congruency based metrics is higher than ours in Table 1. The GFCE method obtains better effect on energy and high-frequency details preservation than our method. In Fig. 5 experiment, the light on the square of our fusion image is too bright to...
express the structure details, so the $Q_E$ metric of proposed method only ranks fourth. The light of the cars and motors are preserved more in Zhang’s fusion result, so its $Q_{MI}$ obtains higher value in Table 2. However, too much light of the cars and motors is not beneficial to identify these targets. In Table 3, the proposed method tends to preserve the bright parts from source 2, which covers up some low gray details. Therefore, many gradients in the fusion image are lost, that leads to structure weakness, so the $Q_G$ and $Q_Y$ only rank fourth. In Table 4, the OETEC fusion result of SR nearly fails for too many discontinuous blocks exist in the image. However, these bright blocks increase the amount of information and gradients, which is described by $Q_{MI}$ and $Q_G$ metrics, respectively. In Table 5, the reason that GP get higher $Q_P$ is similar to Table 1. In Table 6, the $Q_E$ of proposed method is nearly equal with highest value, which indicates the proposed method also obtains well effect.

In six groups of fusion experiments under five assessment metrics, there are 30 metrics in total, and the number of times of proposed method ranking first to fourth are 17, 8, 1 and 4 times, respectively. According to [41], the five quality metrics should be considered together to evaluate the real performance. In our view, the fusion quality should be evaluated by combining qualitative and quantitative evaluations, rather than based on single one. In summary, it is demonstrated that the proposed method outperforms nine comparison fusion methods and achieves the state-of-the-art fusion effect.
obtain final color fused image. Thus, the details and the color are preserved well in fused result, as shown in Fig. 8-10.

The average time cost of all the methods are shown in Table 7, the resolution of six image pairs are $320 \times 240$, $256 \times 256$, $256 \times 256$, $340 \times 255$, $616 \times 454$ and $256 \times 256$, respectively. It can be seen that the method only ranks sixth and its computation cost is much higher than that of the first five methods. The main reason of high computation cost is the sparse representation occupies 96.4% of the running time in proposed method. However, the proposed method is more efficiency than other SR based methods, which indicates the proposed low-rank components SR fusion is effective.

Furthermore, Fig. 12 and Fig. 13 give the multi-modal image fusion examples at gray-gray and gray-color scales, respectively. According to the fusion results, the fusion results preserve the details and color information well and improve the visual observation quality. Therefore, the proposed method can well accommodate multi-modal image fusion.

IV. CONCLUSION

In this paper, a novel multi-modal image fusion method is proposed by decomposing the fusion into low-rank components fusion and sparse components fusion. The low-rank components have the similar feature, so a universal low-rank dictionary is constructed and the batch-OMP is utilized to improve effect and efficiency of coefficients. To preserve more energy and details of the sparse components, a multi-scale anisotropic measure is proposed to construct multi-scale weight maps as the guidance for the sparse components fusion. Moreover, the multi-scale weight maps improve the spatial consistency and suppress the artifacts and faults. The qualitative and quantitative evaluations verify the better performance of the proposed method on multi-modal images fusion. In the future, we will investigate how to expand the application area and select the more appropriate parameters.

REFERENCES

[1] S. T. Li, X. D. Kang, L. Y. Fang, J. W. Hu, and H. T. Yin, “Pixel-level image fusion: A survey of the state of the art,” Information Fusion, vol. 33, pp. 100–112, Jan. 2017.

[2] Z. Sheng, S. Wen-Zhong, J. Liu, G.-X. Zhu, and J.-W. Tian, “Multisource image fusion method using support value transform,” IEEE Trans. Image Process., vol. 16, no. 7, pp. 1831–1839, Jul. 2007.

[3] V. Bhateja, H. Patel, A. Krishn, A. Sahu, and A. Lay-Ekuakille, “Multi-modal medical image sensor fusion framework using cascade of wavelet and contourlet transform domains,” IEEE Sensors J., vol. 15, no. 12, pp. 6783–6790, Dec. 2015.

[4] Y. A. V. Phamila and R. Amutha, “Discrete cosine transform based fusion of multi-focus images for visual sensor networks,” Signal Process., vol. 95, pp. 161–170, Feb. 2014.

[5] N. Jasim Hussein, F. Hu, and F. He, “Multisensor of thermal and visual images to detect concealed weapon using harmony search image fusion approach,” Pattern Recognit. Lett., vol. 94, pp. 219–227, Jul. 2017.

[6] Z. M. Liu and C. J. Liu, “Fusion of color, local spatial and global frequency information for face recognition,” Pattern Recognit., vol. 43, pp. 2882–2890, Aug. 2010.

[7] L. Cao, L. Jin, H. Tao, G. Li, Z. Zhaang, and Y. Zang, “Multi-focus image fusion based on spatial frequency in discrete cosine transform domain,” IEEE Signal Process. Lett., vol. 22, no. 2, pp. 220–224, Feb. 2015.

[8] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, “A novel multi-modality image fusion method based on image decomposition and sparse representation,” Inf. Sci., vol. 432, pp. 516–529, Mar. 2018.

[9] T. Wan, C. Zhi, and Z. Qin, “Multifocus image fusion based on robust principal component analysis,” Pattern Recognit. Lett., vol. 34, no. 9, pp. 1001–1008, Jul. 2013.

[10] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, “Infrared and visual image fusion through infrared feature extraction and visual information preservation,” Infr. Phys. Technol., vol. 83, pp. 227–237, Jun. 2017.

[11] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, “Remote sensing image fusion using the curvelet transform,” Inf. Fusion, vol. 8, no. 2, pp. 143–156, 2007.

[12] Y. Yang, S. Tong, S. Y. Huang, P. Lin, and Y. M. Fang, “A hybrid method for multi-focus image fusion based on fast discrete curvelet transform,” IEEE Access, vol. 5, pp. 14986–14913, 2017.

[13] R. Soundrapandiyam, M. Karappillai, S. Kumar, S. Kumar Tyagi, F. Wu, and K.-H. Jung, “An efficient DWT and intuitionistic fuzzy based multi-modality medical image fusion,” Int. J. Imag. Syst. Technol., vol. 27, no. 2, pp. 118–132, Jun. 2017.

[14] Z. He, W. Lu, W. Sun, and J. Huang, “Digital image splicing detection based on Markov features in DCT and DWT domain,” Pattern Recognit., vol. 45, no. 12, pp. 4292–4299, Dec. 2012.

[15] X. Liu, W. Mei, and H. Xu, “Structure tensor and nonsampled shearlet transform based algorithm for CT and MRI image fusion,” Neurocomputing, vol. 235, pp. 131–139, Apr. 2017.

[16] M. Yin, P. Duan, W. Liu, and X. Liang, “A novel infrared and visible image fusion algorithm based on shift-invariant dual-tree complex shearlet transform and sparse representation,” Neurocomputing, vol. 226, pp. 182–191, Feb. 2017.

[17] S. Dahle, G. Kutyniok, P. Maass, C. Sagiv, H.-G. Stark, and G. Teschke, “The uncertainty principle associated with the continuous Shearlet transform,” Int. J. Wavelets, Multiresolution Inf. Process., vol. 6, no. 2, pp. 157–181, 2008.

[18] R. Vijayaraj and S. Musttan, “Discrete wavelet transform based principal component averaging fusion for medical images,” AEU-Int. J. Electron. Commun., vol. 69, no. 6, pp. 896–902, Jun. 2015.

[19] I. W. Selesnick, R. G. Baraniuk, and N. C. Kingsbury, “The dual-tree complex wavelet transform,” IEEE Signal Process. Mag., vol. 22, no. 6, pp. 123–151, 2005.

[20] L. Tang, F. Zhao, and Z. Zhao, “The nonsampled contourlet transform for image fusion,” in Proc. Int. Conf. Wavelet Anal. Pattern Recognit., Beijing, China, Nov. 2007, pp. 305–310.

[21] Y. Yang, Y. Que, S. Huang, and P. Lin, “Multimodal sensor medical image fusion based on type-2 fuzzy logic in NSCT domain,” IEEE Sensors J., vol. 16, no. 10, pp. 3735–3745, May 2016.

[22] Y. Liu, S. Liu, and Z. Wang, “A general framework for image fusion based on multi-scale transform and sparse representation,” Inf. Fusion, vol. 24, pp. 147–164, Jul. 2015.

[23] S. Li and B. Yang, “Multifocus image fusion by combining curvelet and wavelet transform,” Pattern Recognit. Lett., vol. 29, no. 9, pp. 1295–1301, Jul. 2008.

[24] Y. Chen, J. Xiong, H.-L. Liu, and Q. Fan, “Fusion method of infrared and visible images based on neighborhood characteristic and regionalization in NSCT domain,” Optik, vol. 125, no. 17, pp. 4980–4984, Sep. 2014.

[25] J. Zhao, H. Feng, Z. Xu, Q. Li, and T. Liu, “Detail enhanced multi-focus image fusion using visual weight map extraction based on multi scale edge preserving decomposition,” Opt. Commun., vol. 287, pp. 45–52, Jan. 2013.

[26] X. Feng, K. Hu, and X. Lou, “Infrared and visible image fusion based on the total variational model and adaptive wolf pack algorithm,” IEEE Access, vol. 8, pp. 2348–2361, 2020.

[27] Z. Zhu, M. Zheng, G. Qi, D. Wang, and Y. Xiang, “A phase congruency and local Laplacian Energy Based Multi-Modality Medical Image Fusion Method in NSCT Domain,” IEEE Access, vol. 7, pp. 20811–20824, 2019.

[28] E. Shechtman and M. Irani, “Matching local self-similarities across images and videos,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Minneapolis, MN, USA, Jun. 2007, pp. 1–8.

[29] J. Mairal, M. Elad, and G. Sapiro, “Sparse representation for color image restoration,” IEEE Trans. Image Process., vol. 17, no. 1, pp. 53–69, Jan. 2008.

[30] H. Yin, S. Li, and L. Fang, “Simultaneous image fusion and super-resolution using sparse representation,” Inf. Fusion, vol. 14, no. 3, pp. 229–240, Jul. 2013.
[31] B. Yang and S. Li, “Multifocus image fusion and restoration with sparse representation,” IEEE Trans. Instrum. Meas., vol. 59, no. 4, pp. 884–892, Apr. 2010.

[32] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tornéret, “Hyperspectral and multispectral image fusion based on a sparse representation,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 7, pp. 3658–3668, Jan. 2015.

[33] Y. Liu, X. Chen, R. K. Ward, and Z. Jane Wang, “Image fusion with convolutional sparse representation,” IEEE Signal Process. Lett., vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[34] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, “Medical image fusion via convolutional sparsity based morphological component analysis,” IEEE Signal Process. Lett., vol. 26, no. 3, pp. 485–489, Mar. 2019.

[35] M. Partridge and M. Jabri, “Robust principal component analysis,” in Proc. IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process., 2000, pp. 289–298.

[36] J. Li, M. Song, and Y. Peng, “Infrared and visible image fusion based on robust principal component analysis and compressed sensing,” Inf. Phys. Technol., vol. 89, pp. 129–139, Mar. 2018.

[37] J. R. Shi, X. Y. Zhen, Z. T. Wei, and W. Yang, “Survey on algorithms of low-rank matrix recovery,” Appl. Res. Comput., vol. 30, no. 6, pp. 1601–1605, 2013.

[38] E. J. Candès, X. D. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, pp. 1–37, 2011.

[39] M. Haron, M. Elad, and A. Bruckstein, “K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” IEEE Trans. Signal Process., vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[40] K. He, J. Sun, and X. Tang, “Guided image filtering,” in Proc. Eur. Conf. Comput. Vis., Heraklion, Greece, 2010, pp. 1–14.

[41] S. T. Li, X. D. Kang, and J. W. Hu, “Image Fusion with Guided Filtering,” IEEE Trans. Image Process., vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[42] RGB to YUV Format Conversion and Inverse Conversion Method and Circuit for Depth Packing and Depacking, Accessed: Jan. 20, 2020. [Online]. Available: https://patents.glgoo.top/patent/US10242646B2/en

[43] N. Uniyal and S. K. Verma, “Image fusion using morphological pyramid consistency method.” Int. J. Control Autom., vol. 95, no. 25, pp. 34–38, Jun. 2014.

[44] V. S. Petrovic and C. S. Xydeas, “Gradient-based multiresolution image fusion,” IEEE Trans. Image Process., vol. 13, no. 2, pp. 228–237, Feb. 2004.

[45] S. Ding, X. Zhao, H. Xu, Q. Zhu, and Y. Xue, “NSCT-PCNN image fusion based on image gradient motivation,” IET Comput. Vis., vol. 12, no. 4, pp. 377–383, Jun. 2018.

[46] Y. Liu and Z. Wang, “Simultaneous image fusion and denoising with adaptive sparse representation,” IET Image Process., vol. 9, no. 5, pp. 347–357, May 2015.

[47] Z. Zhou, M. Dong, X. Xie, and Z. Gao, “Fusion of infrared and visible images for night-vision context enhancement,” Appl. Opt., vol. 55, no. 23, p. 6480, Aug. 2016.

[48] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, and W. Wu, “Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 1, pp. 94–109, Jan. 2012.

[49] Y. Chen and R. S. Blum, “A new automated quality assessment algorithm for image fusion,” Image Vis. Comput., vol. 27, no. 10, pp. 1421–1432, Sep. 2009.

FUYU HUANG was born in Hebei, China, in 1985. He received the B.S., M.S., and Ph.D. degrees from the Department of Optical and Electronic Engineering, Mechanical Engineering College, Shijiazhuang, China, in 2008, 2010, and 2014, respectively. His research interests include optical signal processing, infrared detection, and optical design.

HUI ZHONG was born in Sichuan, China, in 1991. She received the B.S. degree from the Beijing University of Chemical Technology, in 2014, and the M.S. degree from the Mechanical Engineering College, in 2016. Her research interests include material designing and armor maintenance.

YICHAO CHEN received the B.S. degree in material physics from Anhui University, in 2013, and the M.S. degree in optical engineering from the Mechanical Engineering College, in 2016. He is currently pursuing the Ph.D. degree in mechanical engineering with Army Engineering University, Shijiazhuang, China. His main research interests include machine vision and infrared technology.

SHUAI ZHANG was born in Shanxi, China, in 1992. He received the B.S. and M.S. degrees in optics from the Mechanical Engineering College, in 2016. He is currently pursuing the Ph.D. degree in optics with Army Engineering University, Shijiazhuang, China. His research interests include the non-linear imaging, camera calibration, and image fusion.

ZIANG WANG was born in Heilongjiang, China, in 1994. He received the B.S. degree in optics from the Academy of Ordnance Engineering, in 2017. He is currently pursuing the M.S. degree in optics with Army Engineering University, Shijiazhuang. His research interests include computer vision and camera calibration.

VOLUME 8, 2020