Target-aware Fusion of Infrared and Visible Images

Yingjie Zhou, Kun Gao, Zeyang Dou, Zizheng Hua, Hong Wang

Abstract—Image fusion technology involves the use of complementary information from multiple sensors to generate a composite image that can highlight details in the region of interest. Some recent methods have tackled the problem of characterization of different source image features that are lacking in conventional methods. However, during fusion processing, these methods may lose some information of interest like smoke. This study proposes a method called Target-aware Decomposition and Parallel Gradient Fusion (TAD-PGF) that fuses infrared and visible images to maintain the high brightness characteristics of infrared targets while transferring the appearance of both source images to the fused image, where “appearance” means details pertaining to the environment and “infrared target” often means hot objects like human body. The target layer is extracted from the infrared image and used as a guide to extract appearance-related information from the visible image. Given that the background of the infrared image contains useful background information, a parallel gradient fusion scheme is proposed to fuse the relevant features with appearance-related information in the visible image. The final blended image is obtained by adding a target layer and a fused appearance layer directly to the fused image. Numerous experiments using publicly available databases were conducted to provide qualitative and quantitative comparisons between state-of-the-art methods and the proposed TAD-PGF. The results reveal that the TAD-PGF can attain good visual effect in various scenarios and maintain useful information from source images to enhance the details of interest.

Keywords — infrared image, visible image, image fusion, $L_s$ filter, WLS filter, parallel gradient fusion

I. INTRODUCTION

Due to limitations of time, space, and the spectrum of light visible to the human eye, different sensors have been designed to obtain information in various scenarios that is otherwise not available to us [1]. However, in case of visual information, it is challenging to process massive amounts of that at the same time. The purpose of image fusion is to combine redundant and complementary information concerning the same scene extracted from multiple images to create a new, single image that offers more a complex and detailed representation of the scene than any of the original images [2]. Image fusion has lately emerged as a field of research for scene analysis in pattern recognition [5], remote sensing [4], medical imaging [6], and modern military applications. In particular, multi-sensor data, such as thermal infrared images generated by detecting differences in thermal radiation, are not easily influenced by such complex factors as wind and sand. This characteristic has significantly expanded the ability to reveal concealed targets. However, limited by their imaging principle and working conditions, infrared images are often noisy, low contrast, blurred, and contain poor visual details, and this often leads to incorrect interpretation in applications. By comparison, visible light sensors work at the same spectrum of wavelength as the human eye, and can obtain information from reflected light at high temporal and spatial resolution. However, they have poor anti-interference and imaging capabilities in bad weather and at night. Combining visible and infrared images can yield information-rich and accurate representations of scenes. The resulting fusion image should contain rich details of the background while highlighting the targets.

In light of the requirements of applications, image fusion can be performed at three levels: pixel, feature, and decision. Pixel-level image fusion is the lowest level of fusion but the basis of the other levels. It involves directly processing initial information without feature extraction and classification, with the aim of synthesizing a better visual appearance of the image than before [7]. By contrast, higher-level fusion combines information that is more abstract, such as feature descriptors and probabilistic variables [8]. Feature-level fusion typically uses feature vectors, such as edges, angles, and textures, to conduct comprehensive analysis and processing, whereas decision-level fusion carries out logical or statistical reasoning using multiple images. Compared with the other two levels, pixel-level image fusion can make fully use of the input data. In this paper, we mainly focus on pixel-level fusion.

In pixel-level image fusion, extracting salient information from images and integrating it together without artifacts are the primary problems. However, it is impossible to design a universal method suitable for all image fusion tasks due to the diversity of images. Over the past few decades, many pixel-level image fusion methods have been proposed for different situations. The easiest method involves combining pixel values by weight. However, such a direct method may reduce the contrast of the image and cause obvious stitch marks when the grayscale between the images is relatively large. Sparse representation-based methods use sparse linear combinations of dictionary items to describe images. An over-complete dictionary is built by simulating the sparse coding mechanism of the human visual system [7], [18]. Subjecting to the limited
amount of the dictionary, even though it yields good performance for low-frequency image fusion, the lack of details needs to be remedied. Methods from different domains, such as the intensity–hue–saturation (IHS) transform [19], [20], principal component analysis (PCA) [21], and the Gram–Schmidt (GS) transform [22]-based methods, are typically used on the infrared image to replace the intensity or first principal component of the multi-spectral image, which may generate significant color distortions in the fusion image.

Multi-scale decomposition (MSD) methods are the most commonly used pixel-level fusion method, and include the Laplacian pyramid [9], wavelet decomposition (DWT) [10], curvelet transform [27], and contourlet transform [11], [12]. These methods have been successful in medical imaging [12], surveillance [13], [14], and remote sensing applications [15], [16]. Nevertheless, the relevant algorithms ignore many characteristics of the source images and preserve the same salient features in different images, such as edges [3], because of which the fusion image often features the blurring effect or loses a considerable amount of visual information [17].

In recent years, some novel methods have been proposed. Bavirisetti and Dhuli introduced an image fusion method based on saliency detection and two-scale image decomposition [23] that can highlight saliency information to some extent, but fails to maintain the difference in brightness between the infrared target and the background. Xiaqing Luo et al. introduced an method using nonsubsampled contourlet transform(NSCT) and stacked sparse autoencoders (SSAE) to integrate the infrared object into the fused image[39]. This successful preserves the complete target, but fails to keep the clear edges. Zhou Z et al. introduced a method through a hybrid multi-scale decomposition with Gaussian and bilateral filters aims to separate the details and edges of the picture and merge them[40]. Unlike our method, the purpose is to separate the complete infrared goal.Ma et al. introduced an optimization method called gradient transfer TV minimization (GTF) [3] that transfers the visible gradient to an infrared image, and this successfully highlights the infrared target while preserving some details in visible image. However, owing to constraints of magnitude on the visible gradient, it distorts the original brightness values, which can hide small targets. Some important information in visible images, such as smoke, is also missed as a consequence.

In this paper, we propose an algorithm that fuses infrared and visible images called Target-aware Decomposition and Parallel Gradient Fusion (TAD-PGF). The proposed method retains the target information in infrared images while transferring appearance-related information from visible images. Compared with other methods, our method will not ignore the variance of feature information. Our algorithm has two stages, the first stage is target-aware decomposition and the second stage is the parallel gradient fusion. The global framework of our proposed algorithm is shown in Fig. 1. First, we use the base layer (target layer) of the infrared image decomposed by the $L_0$ filter as a guide to separate detail related to appearance in the visible image using the weighted least squares(WLS) filter. Thus, we get the detail layers and base layers of both source images. Second, to preserve appearance-related information in both source images, we use the proposed the PGF method for generating the fusion detail layers from two detail layers. The objective function of this method consists of four terms—two intensity control parts and two gradient control parts (regularization term)—so that the fusion detail layer have similar pixel intensities and gradient direction distribution to those of each of the input images. Finally, the fusion detail layer is added to the infrared base layer to form the output image.

The main contributions of this study are as follows:

1. We propose an algorithm to fuse visible and infrared images, called Target-aware decomposition and Parallel Gradient Fusion. The target-aware decomposition stage uses the target layer of the infrared image to guide the decomposition of the visible image, which is a novel design that can accurately separate details in the background from the target region. The proposed PGF method takes advantage of the features of intensity and gradient direction to generate a fused detail image from appearance-related detail layers without artifacts.

2. We derive split Bregman-based alternating minimization to iteratively minimize the proposed PGF objective function.

3. Our methods achieves the best visual effects and keeps the largest quantitative results in most cases.
The remainder of this paper is organized as follows: The proposed method, the motivation for it, and the method of calculation based on split Bregman are introduced in Section II. The experimental setup and an analysis of the results are provided in Section III, and the conclusions of this study are stated in Section IV.

II. THE PROPOSED METHOD

In this section, we comprehensively describe the proposed fusion algorithm.

A. Target-aware Decomposition

Given a pair of infrared and visible images, in light of their characteristics, extracting saliency information and processing them separately are the most intuitive and effective means of obtaining the best results. Therefore, separating the infrared target from the background and extracting the details of appearance in the visible image form the first problem.

In general, the infrared radiation of targets often embodies thermal anomalies with a sharp outline and a lack of environment details in infrared images in contrast to visible images. Based on these features, we choose the L0 filter to extract it while removing details of the background. The L0 filter is effective in maintaining the main edges of the image while eliminating details with small changes in the amplitude of the gradient amplitude, which is well suited to this task [35].

The objective function of the L0 filter is as follows:

$$\min_{u} \left\{ \sum_{p} \left( u_p - g_p \right)^2 + \lambda_{L0} \cdot C(u) \right\}$$  \hspace{1cm} (1)

where $g$ is the input image, $u$ is the computed result, $C$ is a counter of the magnitude of the gradient, $\lambda_{L0}$ is a hyperparameter and $p$ are pixels of the input image. (1) can be solved by the half-quadratic division algorithm [35]. By setting different values of $\lambda_{L0}$, we can obtain the results of different effects of smoothing. We can then simply decompose the original image at different scales according to the practical situation.

The desired appearance layer of visible image is one that contains abundant background detail except in the part that influences the outline of the target. Using the target layer of the infrared image as a guide, undesirable details can be easily removed. We choose the WLS filter to carry out this task. Other filters such as the guided filter are also suitable.

The WLS filter is an edge-preserving multi-scale image decompositions method based on the weighted least squares’ optimization framework. The main idea of the WLS is to smoothen the input image under the guidance of its log-luminance channel to control the gradient along each directions of the new image [36]. By replacing the log-luminance channel by the infrared target layer, this method can be expressed as:

$$\min_{u} \left\{ \sum_{p} \left( u_p - g_p \right)^2 + \lambda_{WLS} \left( a_{s,p}(g) \left( \frac{\alpha}{\lambda} \right) + a_{s,p}(g) \left( \frac{\alpha}{\lambda} \right) \right) \right\}$$  \hspace{1cm} (2)

where $a_{s,p}(g) = \left( \frac{\alpha}{\lambda} \right)$ and $a_{s,p}(g) = \left( \frac{\alpha}{\lambda} \right)$ and where $\alpha$ is the calculated result, $g$ is the input image, and $a_{s,p}$ and $\alpha_{s,p}$ are the smoothness weights controlled by the infrared target layer $g$. $\lambda_{WLS}$ is a balance parameter between the terms. Increasing the value of $\lambda_{WLS}$ can cause image $u$ to gradually become smooth. The exponent $\alpha$ determines the sensitivity to the gradients of $g$ while $\epsilon$ is a small constant.

B. Detail layer fusion method: Parallel gradient fusion

Once we complete the decomposition, the infrared target and the details of the background are completely separated. As shown in Fig 1, the detail layers of different source image contains different details of the same scene. To fuse all appearance information, we propose a parallel gradient fusion method to integrate these detail layers into a fused detail layer.

1) Fusion objective function

To clarify, the infrared, visible, and fused detail layers are all supposed to be grayscale of size $m \times n$, denoted separately by $u_g, v_d, f_d \in \mathbb{R}^{m \times n}$. Equation (3) can be seen as the difference between the fusion image and the source images, which should be as small as possible:

$$\zeta(f_d) = \frac{1}{2} \| f_d - u_g \|_2^2 + \frac{1}{2} \| f_d - v_d \|_2^2$$  \hspace{1cm} (3)

Though (3) forced the pixel intensity of the fused details similar with those of source images, one should also consider the gradient constraints as well. Image gradient has two attributes: gradient magnitude and gradient direction. Because we don’t want to alter the intensity of the interest target, we only
need to make the gradients of the fused details parallel to those in source images. And that is why call it parallel gradient fusion. This is in contrast to previous work, which only considers the gradient magnitude as a constraint. The gradient magnitude constraint spoils the pixel intensity distribution, which leads to a loss of a small amount of information or distorts the original brightness values of infrared target. The gradient direction control part can be express as:

\[
\zeta(f_d) = \langle \nabla f_d, \nabla u_d \rangle - \langle f_d, \nabla u_d \rangle + \langle \nabla f_d, \nabla v_d \rangle - \langle f_d, \nabla v_d \rangle
\]

(4)

where \( \nabla \) is the gradient operator defined in detail later. \( \| \nabla f_d \|, \| \nabla u_d \|, \) and \( \| \nabla v_d \| \) denote the gradient magnitudes of the fusion, infrared, and visible detail layers, respectively. \( \langle \nabla f_d, \nabla u_d \rangle \) and \( \langle \nabla f_d, \nabla v_d \rangle \) are the inner products of the gradients of the fusion image and the corresponding source image, respectively.

The final objective function can be written as:

\[
\zeta(f_d) = \zeta_0(f_d) + \zeta_2(f_d)
\]

\[
= \frac{1}{2} \lambda \| f_d - u_d \|^2 + \frac{1}{2} \| f_d - v_d \|^2 + \alpha_1 \left( \langle \nabla f_d, \nabla u_d \rangle - \langle f_d, \nabla u_d \rangle \right) + \alpha_2 \left( \langle \nabla f_d, \nabla v_d \rangle - \langle f_d, \nabla v_d \rangle \right)
\]

(5)

where the first and second terms constrain the fusion image \( f_d \) to have similar pixel intensity to those of the source images, and \( \lambda \) is a hyper-parameter that controls the impact ratio. The third and fourth terms represent the direction of the control of the gradient, and are called regularization terms. \( \alpha_1 \) and \( \alpha_2 \) are hyper-parameters used to control the degree of similarity along the direction of the gradient.

Using this novel algorithm, we can combine details of the appearance from different detectors in a harmonious way. Both the intensity distribution and the gradient of the images are considered. In the next section, we provide a solution to the PGF.

2) Numerical implementation employing split Bregman iteration and alternating minimization algorithm

According to the split Bregman framework \([37]\) and the alternating minimization algorithm \([38]\), we derive an alternating split Bregman algorithm to solve (5). Specifically, the objective function can be expressed as an optimization problem in discrete version:

\[
f_d = \arg \min_{f_d} \sum_{i=1}^{m} \left[ \frac{1}{2} \lambda (f_d - u_d)^2 + \frac{1}{2} (f_d - v_d)^2 + \alpha_1 \langle \nabla f_d, \nabla u_d \rangle - \langle f_d, \nabla u_d \rangle + \alpha_2 \langle \nabla f_d, \nabla v_d \rangle - \langle f_d, \nabla v_d \rangle \right]
\]

(6)

where

\[
\| \nabla f_d \| = \sqrt{\| \nabla u_d \|^2 + \| \nabla v_d \|^2}
\]

\[
\| \nabla u_d \| = \sqrt{\| \nabla u_d \|^2 + \| \nabla v_d \|^2}
\]

\[
\| \nabla v_d \| = \sqrt{\| \nabla u_d \|^2 + \| \nabla v_d \|^2}
\]

The gradient magnitude \( |x| = \sqrt{x_1^2 + x_2^2} \) for every \( x = (x_1, x_2) \in \mathbb{R}^2 \), \( \nabla_i = (\nabla_i^h, \nabla_i^v) \) denotes the image gradient \( \nabla \).

TABLE I

<table>
<thead>
<tr>
<th>ALGORITHM PROCESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Algorithm:</strong> Target-aware Decomposition and Parallel Gradient Fusion (TAD-PGF)</td>
</tr>
<tr>
<td><strong>Input:</strong> infrared image ( u ), visible image ( v )</td>
</tr>
<tr>
<td><strong>Output:</strong> fusion image ( F )</td>
</tr>
<tr>
<td>1 Decomposition ( u ) using ( L_0 )-filter to get base layer ( u_b ) and detail layer ( u_d )</td>
</tr>
<tr>
<td>2 Decomposition ( v ) using WLS filter under the guidance of ( u_b ) to get base layer ( v_b ) and detail layer ( v_d )</td>
</tr>
<tr>
<td>3 Repeat</td>
</tr>
<tr>
<td>1) Fix ( b ) and ( t ); compute ( \hat{f}_d ) using Equation ( 12 )</td>
</tr>
<tr>
<td>2) Update ( \hat{f}_d ); compute ( b ) using the Equation ( 15 )</td>
</tr>
<tr>
<td>3) Update ( b ); compute ( t ) using Equation ( 16 )</td>
</tr>
<tr>
<td>4) Update ( t )</td>
</tr>
<tr>
<td>Until convergence, acquire the fusion detail layer ( f_d )</td>
</tr>
<tr>
<td>The final fusion image ( F ) is calculated as ( F = f_d + u_b )</td>
</tr>
</tbody>
</table>

at pixel \( i \), \( \nabla_i^h \) and \( \nabla_i^v \) are the horizontal and vertical gradients, respectively.

To derive the alternating split Bregman algorithm, we introduce dual variable \( b \) to replace \( \nabla f_d \). Then, using the augmented Lagrangian method, we have the discrete constrained optimization problem:

\[
f_d = \arg \min_{f_d} \sum_{i=1}^{m} \left[ \frac{1}{2} \lambda (f_d - u_d)^2 + \frac{1}{2} (f_d - v_d)^2 + \alpha_1 \langle \nabla f_d, \nabla u_d \rangle - \langle f_d, \nabla u_d \rangle + \alpha_2 \langle \nabla f_d, \nabla v_d \rangle - \langle f_d, \nabla v_d \rangle \right]
\]

(7)

s.t. \( b = \nabla f_d \)

The proposed iteration scheme is:

\[
f_d^{k+1} = \arg \min_{f_d} \left\{ \sum_{i=1}^{m} \left[ y(f_d) + \frac{1}{2} \tau (\theta_i - \nabla f_d - t_i) \right] \right\}
\]

(8)

\[
t^{k+1} = b^k + \nabla f_d^{k+1} - t^{k+1}
\]

(9)

where

\[
y(f_d) = \frac{1}{2} \lambda (f_d - u_d)^2 + \frac{1}{2} (f_d - v_d)^2 + \alpha_1 \langle \nabla f_d, \nabla u_d \rangle - \langle f_d, \nabla u_d \rangle + \alpha_2 \langle \nabla f_d, \nabla v_d \rangle - \langle f_d, \nabla v_d \rangle
\]

(8) can be decoupled into several sub-problems. In each, the rest of the variables are fixed when solving for any one. The sub-problems are looped until convergence.

1. Calculate sub-problem \( f_d \) with fixed \( b \) and \( t \):

\[
f_d^{k+1} = \arg \min_{f_d} \left\{ \sum_{i=1}^{m} \left[ \frac{1}{2} \lambda (f_d - u_d)^2 + \frac{1}{2} (f_d - v_d)^2 \right] \right\}
\]

(10)

The optimal \( f_d^{k+1} \) satisfies

\[
(1 + \lambda) f_d^{k+1} - u_d - v_d - \nabla f_d^{k+1} - t^k = 0
\]

(11)

We use the Fourier transform and the inverse Fourier
transform to get the result:

\[ f_d = \text{FFT}^{-1} \left( \frac{\text{FFT}(\lambda u_d + v_r) + \tau \text{FFT}(\nabla) \text{FFT}(b^i - t^i)}{1 + \lambda + \tau \text{FFT}(\Delta)} \right) \]  

(12)

2. Calculate sub-problem \( b \) with fixed \( f_d^{k+1} \) and \( t \) :

\[ b^{k+1} = \arg \min_b \left\{ \sum_{i=1}^{n} \left[ \frac{a_i (\| \nabla u_d \| - \langle b, \nabla u_d \rangle \} \right] + \frac{a_i (\| \nabla v_d \| - \langle b, \nabla v_d \rangle \right]}{b^k + \frac{a_i}{\tau} \nabla u_d + \frac{a_i}{\tau} \nabla v_d, \alpha_i u_d] + \alpha_i v_d \right\} \]  

(13)

where \( b = (b_h, b_v) \). Thus, the optimal condition for (13) becomes

\[ \frac{b (\| \nabla u_d \| + a_i \| \nabla v_d \|)}{b^k + \frac{a_i}{\tau} \nabla u_d + \frac{a_i}{\tau} \nabla v_d, \alpha_i u_d] + \alpha_i v_d \right\} \]  

(14)

The above Equation (14) has an explicit solution as follows:

\[ b = \text{shrink}(\nabla f_d^{k+1} + t^k + \frac{a_i}{\tau} \nabla u_d + \frac{a_i}{\tau} \nabla v_d, \alpha_i u_d] + \alpha_i v_d \right) \]  

(15)

where \( \text{shrink}(x, r) = \max\{\|x\| - r, 0\} x \) \[|x| \]

3. Update \( t \):

\[ t^{k+1} = t^k + \nabla f_d^{k+1} - b^{k+1} \]  

(16)

C. Final result

At the end of the algorithm, once the fusion detail layer \( f_d \) and the infrared target layer of the infrared image \( u_b \) have been prepared, we add them up to obtain the final result. Note that the add operation doesn’t decrease the spatial resolution because of the texture removal nature of L0 smooth filter.

\[ F = f_d + u_b \]  

(17)

The flow of the complete algorithm is shown in Table I.

An example is shown in Fig. 2. Figs. 2(a) and 2(b) are the source images taken at visible and infrared wavelengths, respectively. The visible image contains the vast majority of environmental details, such as the details of bushes, while the infrared image shows information pertaining to the infrared target with highlights. Fig. 2(c) is the result of GTF and has better visual effects. However, compared with the result of our proposed method in Fig. 2(d), it reduces the brightness of certain parts of the target and blurs the edges of the bunker. Our method successfully highlights the target tank and contains rich detail.

III. EXPERIMENTAL RESULTS

In this section, we experimentally validated the performance of the proposed TAD-PGF and compare with nine state-of-the-art fusion methods. The methods used to compare with ours are the Laplacian pyramid (LP) [9], ratio of low-pass pyramid (RP) [24], wavelet transform [25], dual-tree complex wavelet transform (DT-CWT) [26], curvelet transform (CVT) [27], multi-resolution singular value decomposition (MSVD) [28], guided filtering-based fusion (GFF) [29], Laplacian pyramid with sparse representation (LP-SR) [30], and gradient transfer fusion (GTF) [3]. We evaluate all the fusion images qualitatively and quantitatively. All nine algorithms were executive based on publicly available codes and the parameters were set accordingly to achieve consistent results.

According to the proposed TAD-PGF, the Lo filter and the WLS filter have one hyper-parameter each, and the PGF has four hyper-parameters that need to be set.

Some empirical values are recommended in Table II that are suitable for most occasions.

A. Result on TNO Human Factor

To test the proposed method in different scenarios, we chose aligned images containing typical military targets and other natural targets from the TNO Human Factor database. The background was also complex, and contained a series of scenes from forest to plain. Six typical image pairs called Tank, Bunker, Sandpath, Nato-camp, Soldier_behind_smoke, and Bench were prepared for qualitative illustration and quantitative evaluation, whereas the Nato_camp_sequence was used for objective quantitative evaluation to validate the universality.

We first provided a series of intuitive results as shown in Fig. 3. The columns from (a) to (f) are the results of fusion of the six image pairs of Tank, Bunker, Sandpath, Nato-camp, Soldier_behind_smoke, and Bench. The rows from top to bottom are visible images, infrared images, and the results of fusion of LP [9], RP [24], wavelet [25], DT-CWT [26], CVT [27], MSVD [28], GFF [29], LP-SR [30], GTF[3] and our method, respectively.

As is clear, all 10 algorithms delivered good performance. They all successfully combined the information in the source images and retained appearance-related details to some extent. On this standard, they were all reasonable. However, in a complex environment, the nine state-of-the-art methods could not always maintain the highlighted targets and background details at the same time, as shown in Fig. 3. On the contrary, our TAD-PGF method overcame these problems. It not only highlighted important targets, but also preserved the details in both the infrared and the visible images, which rendered its results more accurate. It did not miss small targets and details or mixed some misinformation in with the data. All these advantages render our method suitable for target recognition and location, and valuable for quickly representing the overall situation.

Subjective evaluation was performed by visual interpretation, which was simple and intuitive. However, it may lead to different evaluations according to different interpreters due to incompleteness and subjectivity. It is neither accurate nor appropriate to judge the results of fusion using only subjective criteria. Thus, some objective parameters should be used to assess the fusion image.
The commonly used evaluation indicators evaluate the spatial resolution, detail, content, and classification capability of the fusion image, including the root-mean-square error (RMSE) [34], entropy (EN) [30], and mutual information (MI) [31]. We used the following to quantitatively evaluate the performance of the fusion methods.

- Entropy (EN): EN reflects the amount of information contained in the image [30]. It defined as follows:

\[
\text{Entropy (EN)} = -\sum p(x) \log p(x)
\]

where \( p(x) \) is the probability of occurrence of the pixel intensity \( x \).

**TABLE II**

<table>
<thead>
<tr>
<th>ALGORITHM PARAMETER INITIALIZATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization of proposed algorithm</td>
</tr>
<tr>
<td>( L_0 ) filter: ( \lambda_{L_0} = 0.01 )</td>
</tr>
<tr>
<td>( WLS ) filter: ( \lambda_{WLS} = 1 )</td>
</tr>
<tr>
<td>( PGT ) method: ( \lambda = 0.5 \ a_1 = 1 \ a_2 = 2 \ \tau = 10 )</td>
</tr>
<tr>
<td>Iterations : 30</td>
</tr>
</tbody>
</table>

Fig. 3. From (a) to (f) are the results of fusion of the six image pairs of Tank, Bunker, Sandpath, Nato-camp, Soldier\_behind\_smoke, and Bench. From top to bottom: infrared images, visible images, results of LP [9], RP [24], wavelet [25], DTCWT [26], CVT [27], MSVD [28], GFF [29], LP-SR [30], GTF [31] and the proposed method.
where $\mu_x$, $\mu_y$ are the mean of $x$ and $y$, respectively, $\sigma^2_x$, $\sigma^2_y$ are the variances of $x$ and $y$, respectively, and $\sigma_{xy}$ is the covariance of both $x$ and $y$. Two variables $c_1$ and $c_2$ are used to stabilize the division with a weak denominator. Having calculated the SSIM between the fusion image and the source images, the result is given in the form of a weighted sum.

The results of the quantitative comparison of the above six images are shown in Fig. 4. According to these data, no matter the circumstance, our method obtained the best visual results and maintained useful information from the source images. Especially in the case of rich background details or smoke obscuration targets, our method performs very well.

To verify the universality of this method, we used the Nato-camp sequence that contained 32 sets of images as a test sequence. The quantitative results of the Nato-camp sequence are given in Fig. 5, where our method TAD-PGF is represented by the black box. It yielded the best performance in terms of EN, 1/K-L divergence, and SSIM. This shows that our method is universal and effective. Moreover, it is interesting that the traditional wavelet method performed as well as our method and was nearly superior to all other enumeration methods in terms of SSIM.

B. Parameter analysis

As mentioned above, the $L_0$ and WLS filters were used to decompose the original images. Both have hyper-parameters to control the scale of the detail. Setting different hyper-parameters can help extract corresponding scales of detail, as shown in Fig. 6. Its flexibility enhances the applicability of our method to various scenarios.
We changed the hyper-parameters of the $L_0$ filter to determine the impact on the results of infrared decomposition. As shown in Fig. 6, the scale of decomposition increased with $\lambda_{L0}$, and the extracted details changed accordingly. Note that the results of decomposition perfectly separated the highlighted infrared targets from the environment details when $\lambda_{L0} = 0.01$. With a smaller value of $\lambda_{L0}$, some details mix in with the base layer. As $\lambda_{L0}$ increased, it caused the infrared target to lose its structural feature. Throughout the above analysis and our experiments, we used 0.01 as the empirically chosen value of $\lambda_{L0}$, as it yielded good visual effects in most cases.

We used the target layer as a guide to test the performance of the WLS filter on the decomposition of the visible image with different parameters. As show in Fig. 7, the greater the value of $\lambda_{WLS}$, the fewer the details remaining in the base layer. It is worth noting that the effect of the target’s brightness is eliminated when extracting its detail. It is an advantage to preserve the difference in brightness between the infrared target and the background.

Similarly, according to the objective function of the fusion of the detail layers, we can deduce that fixing parameters $\lambda$, $\alpha_1$, $\alpha_2$, and $\tau$ of the PGF yields a unique result. We need only
change the value of these hyper-parameters to obtain the desired characteristics. This is a significant advantage in terms of meeting the requirements of different applications. \( \tau \) is a regular parameter in the PGF algorithm that also controls the smoothness of the fused image, and can be used for de-noising. We demonstrate the process on a series of images in Figs. 8, 9, and 10.

Due to constraints related to the hardware and the environment, there may be different levels of noise in the source images. As shown in Fig. 8, the original visible image contains a considerable amount of noise. In the fusion process, this noise is gradually smoothed out as parameter \( \tau \) increases. In practice, we can adjust this value flexibly. Two other advantages of our method are its robustness and flexibility in the face of the background. As shown in Figs. 9 and 10, changing the ratio of the parameters of the infrared image to those of the visible image (intensity or gradient) can make fusion image shows a similar detail appearance to the source image of a great proportion. Nevertheless, because of the particularity of the algorithm design, the background does not make a significant difference on intensity to the results. This ensures the recognition of the target. Throughout the above analyses and experiments, we used \( \lambda = 0.5 , \ \alpha_1 = 1 , \ \alpha_2 = 2 , \ \text{and} \ \tau = 10 \) as empirical values, which can achieve a good separation effect in most cases.

IV. CONCLUSION

In this study, we proposed a method for infrared and visible image fusion called Target-aware Decomposition and Parallel Gradient Fusion (TAD-PGF) based on the \( L_0 \) filter, the WLS filter, and parallel gradient fusion. It can maintain information concerning the target of interest in infrared images while transferring appearance-related information contained in visible images. Moreover, using a tunable hyper-parameter, the proposed method can selectively highlight the information of interest according to the practical application considered. All these features are beneficial for fusion-based target detection and recognition systems. Our method yielded the best visual effects and preserved useful information from the source images to enhance the details of interest.

REFERENCES

transform for infrared and visible light image," *Infrared Physics & Technology*, vol. 61, no. 5, pp. 94-100, 2013.


Yingjie Zhou received the B.E. degree in Optoelectronic Information Engineering from Beijing Institute of Technology, China, in 2016. She is pursuing the M.S. degree in Optical Engineering with the Beijing Institute of Technology. Her research interests include image processing and deep learning.
Kun Gao received his B.A. in electrical engineering and Ph.D. in instrument science and engineering from Zhejiang University, China, in 1995 and 2002, respectively. From 2002 to 2004, he was post-doctoral fellow at Tsinghua University in China. Since 2005, he has been at the Beijing Institute of Technology in China, working on infrared technology and real-time image processing. He is a member of the Optical Society of China.

Zeyang Dou received his B.E. in mathematics from Baoding University in China in 2012, and his M.S. in computational mathematics from the Communication University of China in 2016. He is currently pursuing his Ph.D. at the Beijing Institute of Technology. His research interests include image processing, machine learning, and deep learning.

Zizheng Hua received his B.E. in electrical and information engineering from SouthWest JiaoTong University in 2015. He is currently working toward his M.S. in the School of Optoelectronics at the Beijing Institute of Technology in optical engineering. His current research focuses on deep learning and parallel processing of images based on multi-core parallel computing.

Hong Wang received the B.E. degree in electronic science and technology from Dezhou University, China, in 2014, and the M.S. degree in electronic circuit and system from Tianjin Normal University, China, in 2018. She is currently pursuing the Ph.D. degree at the Beijing Institute of Technology. Her research interests include image processing, pattern recognition, machine learning, and deep learning.