

Article

Fusing Infrared and Visible Images of Different Resolutions via Total Variation Model

Qinglei Du ^{1,2}, Han Xu ¹, Yong Ma ¹, Jun Huang ¹ and Fan Fan ^{1,*}

¹ Electronic Information School, Wuhan University, Wuhan 430072, China; dql822@163.com (Q.D.); xu_han@whu.edu.cn (H.X.); mayong@whu.edu.cn (Y.M.); junhwong@whu.edu.cn (J.H.)

² Air Force Early Warning Academy, Wuhan 430019, China

* Correspondence: fanfan@whu.edu.cn; Tel.: +86-27-8720-9881

Received: 9 September 2018; Accepted: 5 November 2018; Published: 8 November 2018



Abstract: In infrared and visible image fusion, existing methods typically have a prerequisite that the source images share the same resolution. However, due to limitations of hardware devices and application environments, infrared images constantly suffer from markedly lower resolution compared with the corresponding visible images. In this case, current fusion methods inevitably cause texture information loss in visible images or blur thermal radiation information in infrared images. Moreover, the principle of existing fusion rules typically focuses on preserving texture details in source images, which may be inappropriate for fusing infrared thermal radiation information because it is characterized by pixel intensities, possibly neglecting the prominence of targets in fused images. Faced with such difficulties and challenges, we propose a novel method to fuse infrared and visible images of different resolutions and generate high-resolution resulting images to obtain clear and accurate fused images. Specifically, the fusion problem is formulated as a total variation (TV) minimization problem. The data fidelity term constrains the pixel intensity similarity of the downsampled fused image with respect to the infrared image, and the regularization term compels the gradient similarity of the fused image with respect to the visible image. The fast iterative shrinkage-thresholding algorithm (FISTA) framework is applied to improve the convergence rate. Our resulting fused images are similar to super-resolved infrared images, which are sharpened by the texture information from visible images. Advantages and innovations of our method are demonstrated by the qualitative and quantitative comparisons with six state-of-the-art methods on publicly available datasets.

Keywords: image fusion; different resolutions; total variation; infrared

1. Introduction

Compared with information provided by a single sensor image, multi-sensor images can supply additional information with feature complementarity for complex and complete scene representations and improved visual understanding. By integrating useful information from different sensor images, we can generate a single composite image, which is informative and suitable for follow-up processing and decision-making. In particular, infrared and visible image fusion is an indispensable branch that plays an important role in military and civilian applications [1] and has been used to enhance the performance in terms of human visual perception, object detection, and target recognition [2–8].

Visible sensors can capture reflected light to provide background details, such as vegetation, texture, area, and soil with high spatial resolution because they operate in the same waveband that can be recognized by human vision [9]. However, under unsatisfactory light conditions, scene information is lost. By contrast, infrared sensors map thermal radiation emitted from objects to gray images; therefore, these sensors exhibit unique advantages in overcoming adverse light conditions

and highlighting thermal targets, such as pedestrians and vehicles. Nevertheless, due to limitations of hardware and environments, infrared images are often accompanied by blurred details, serious noise, and considerably low resolution. These limitations encourage the fusion of infrared and visible images to produce a single image, which can simultaneously highlight thermal radiation for target detection and reserve texture information for characterization of appearance.

According to different representations and applications, the fusion process can be conducted at the following three levels [10]: pixel, feature, and symbol levels. Pixel-level fusion is on the lowest level and fuses different physical parameters. Feature-level fusion involves feature descriptor, probability variables, and object labels. Symbol-level fusion uses high abstraction of information, such as representative symbols or decisions, for fusion. The focus of this paper is on pixel level. From another perspective, existing infrared and visible image fusion methods are divided into eight categories according to their corresponding theories [1], including multi-scale transform-based methods [11,12], sparse representation-based methods [13,14], neural network-based methods [15,16], subspace-based methods [17,18], saliency-based methods [19,20], hybrid methods [21,22], deep learning-based methods [23,24], and other fusion methods [10,25].

Despite the notable progress of image fusion, fusing infrared and visible images of different resolutions remains a challenging task. Infrared images constantly suffer from considerably lower resolution compared with visible images. Under this condition, if existing methods are applied to fusing source images, then eliminating resolution differences, through downsampling visible images or upsampling infrared images as an example, is the first problem to solve. As shown in Figure 1, to intuitively illustrate the fusion process, we take a typical image pair and a typical fusion method, such as hybrid multi-scale decomposition (HMSD) [26], as an example. In procedure (a), the visible image is downsampled before using as a source image for fusion, and the fused image is upsampled considering consistency and comparability. Although the fused image can be forcibly converted into high resolution, the discarded high-quality texture information in the visible image caused by downsampling operation by image processing cannot be retrieved from the viewpoint of information theory; hence, the clarity of the fused image is not improved. In procedure (c), the original infrared image is upsampled to high resolution before fusion. As shown in the figure, upsampling results in blur and inaccuracy in the infrared image, which is also fused into the resulting image; hence, the fused image in procedure (c) is accompanied by serious noise. In addition, the principle of existing fusion rules typically focuses on preserving texture details in source images. This condition may be problematic in the fusion of infrared and visible images because infrared thermal radiation information is characterized by pixel intensities and usually contains few texture details that may fail to be fused in the resulting image. However, such information is often crucial to highlight the targets. This can be shown in procedures (a) and (c), in which the fused image loses the property of high contrast similar to that in the infrared image, submerging the target in the clutter background.

To address the above-mentioned challenges, in this paper, we propose a new algorithm termed as Different Resolution Total Variation (DRTV) for infrared and visible image fusion. In particular, we formulate the fusion problem as a convex optimization problem characterized by total variation model and seek the optimal fused image that minimizes an energy function. This formulation accepts different resolutions of infrared and visible images. The energy consists of a data fidelity term and a regularization term. The data fidelity term constrains that the downsampled fused image should present similar pixel intensities with the original infrared image. The regularization term ensures that the gradient distribution in the visible image can be transferred into the fused image. The two terms can guarantee that our fusion result contains the thermal radiation information from the infrared image and the texture details from the visible image. To optimize the solution procedure, we apply the fast iterative shrinkage-thresholding algorithm (FISTA) [27] framework to accelerate convergence, e.g., the convergence speed is increased to $O(1/k^2)$ with k being the iteration number. Owing to the preceding causes, the cost for acquiring high-quality infrared and visible fused images can be considerably reduced, thereby providing substantial benefit to military equipment.

and civilian industries relying on infrared and visible image fusion. For intuitive comparison, Figure 1b shows the fused result of DRTV on the same source image pair, where we zoom in a small region in the red box for clear comparison. Figure 1 shows that the fused image generated by our DRTV, i.e., Figure 1b, is clearer with less noise compared with Figure 1a,c due to the forced resolution conversion. In addition, our fusion result looks like a super-resolved infrared image, which preserves the property of high contrast in the infrared image, clearly highlighting the target.

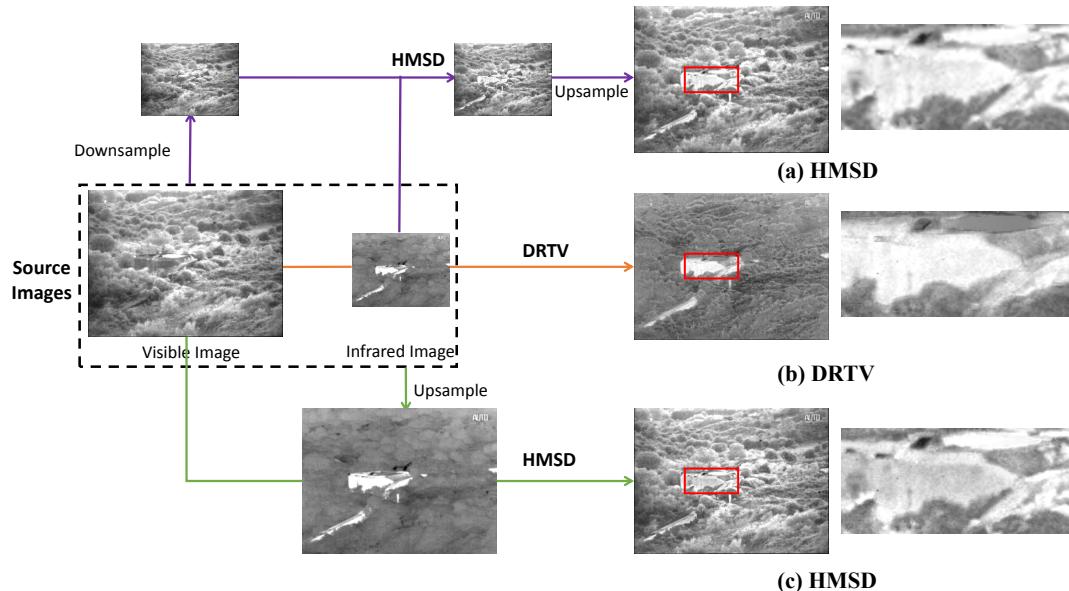


Figure 1. A typical example for fusing infrared and visible images of different resolutions through (a) downsampling the visible image; (b) our method; and (c) upsampling the infrared image.

The contributions of this paper include the following two aspects. On the one hand, the proposed DRTV makes a breakthrough in fusing visible and infrared images of different resolutions, which can considerably prevent loss of texture information caused by compression or blur and inaccuracy due to forced upsampling, without expenses on upgrading hardware devices. On the other hand, the fused images obtained by DRTV visually benefit from rich texture information; simultaneously, the thermal targets are easily detected. By changing the regularization parameter in our model, we can subjectively adjust the visual similarity between the fused image and visible/infrared image.

The remainder of this paper is organized as follows. Section 2 describes the problem statement and formulation for infrared and visible image fusion. Section 3 provides the optimization strategy of our algorithm. In Section 4, we compare DRTV with several state-of-the-art methods on publicly available datasets and analyze the results. Conclusions and future works are presented in Section 5.

2. Problem Statement and Formulation

We denote $\text{Vis} \in \mathbb{R}^{m \times n}$ the visible image of size $m \times n$ and $\text{IR} \in \mathbb{R}^{\frac{m}{c} \times \frac{n}{c}}$ the infrared image of size $\frac{m}{c} \times \frac{n}{c}$, where c is a constant denoting the ratio between the infrared image resolution and the resolution of the visible image. $\mathbf{X} \in \mathbb{R}^{m \times n}$ denotes the fused image. Note that the size of \mathbf{X} is $m \times n$ rather than $\frac{m}{c} \times \frac{n}{c}$.

Our goal for the given Vis and IR is to fuse them to generate a high-resolution fused image, which should virtually retain important information in source images, e.g., thermal radiation information in the infrared image, and texture features in the visible image.

2.1. Data Fidelity Term

Previous methods generally apply a restriction on source images, in which visible and infrared images should share the same resolution. When this restriction is not satisfied, downsampling the

visible image is one solution, but the information translated into the fused image also is similarly limited. Employing the upsampled infrared image as the prior knowledge to maintain thermal radiation information is another solution. However, thermal information after upsampling is constantly blurred and inaccurate. If we constrain fused information to follow the blurred source, then the fused result is even less accurate. Therefore, we assume in this study that the fused image after downsampling possesses similar pixel intensities to the original infrared image. Least squares fitting is used to model this relationship:

$$\varepsilon_1(\mathbf{X}) = \frac{1}{2} \|\psi\mathbf{X} - \mathbf{IR}\|_F^2, \quad (1)$$

where $\|\cdot\|_F$ denotes the Frobenius norm of matrix, which is defined as $\|\mathbf{X}\|_F = \sqrt{\sum_{i,j} X_{i,j}^2}$. ψ denotes the downsampling operator of the nearest interpolation. If the high-resolution image is defined as $\mathbf{S}_{M \times N}$ and the downsampled image is $\mathbf{O}_{m \times n} = \psi\mathbf{S}_{M \times N}$, then the operator ψ can be denoted as $\mathbf{O}_{i,j} = \mathbf{S}_{[\lceil i \times \frac{M}{m} \rceil], [\lceil j \times \frac{N}{n} \rceil]}$, where $[\cdot]$ represents rounding the data to the nearest integer. $\varepsilon_1(\mathbf{X})$ is a data fidelity term. According to the preceding constraint, the thermal target remains prominent in the fused image.

2.2. Regularization Term

The regularization term is essentially the prior knowledge used to avoid overfitting of least squares. One requirement is to allow thermal targets in the fused image to remain noticeable. Additional detailed appearance information in the visible image is expected to be preserved in the fused result. The direct approach is to constrain similar pixel intensity distribution in the fused and visible image. However, the same constraint is applied in $\varepsilon_1(\mathbf{X})$, and pixel intensity distributions of infrared and visible images considerably vary due to their different manifestations. Therefore, this approach is contradictory, and we attempt to search another feature to characterize appearance information in the visible image.

According to the human visual perception system, image edges and textures are abundant when the spatial frequency is large. Moreover, the spatial frequency is a metric based on gradient distribution. Understanding that the detailed appearance information is basically represented by gradient distribution of the image is easy. On the basis of this observation, we require that the fused image should present similar gradients with the visible image, which is represented by the ℓ_p norm ($p \geq 0$):

$$\varepsilon_2(\mathbf{X}) = \|\nabla\mathbf{X} - \nabla\mathbf{Vis}\|_p, \quad (2)$$

where ∇ is the gradient operator and $\varepsilon_2(\mathbf{X})$ should be as small as possible.

By using $\nabla_{i,j} = (\nabla_{i,j}^h, \nabla_{i,j}^v)$ to represent the image gradient at pixel (i, j) and $\mathbf{x}_{i,j}$ to denote the pixel of the i -th row and j -th column in the image, we obtain the following:

$$\nabla_{i,j}^h \mathbf{x} = \mathbf{x}_{i,j} - \mathbf{x}_{i,j+1}, \quad \nabla_{i,j}^v \mathbf{x} = \mathbf{x}_{i,j} - \mathbf{x}_{i+1,j}, \quad (3)$$

$$|\nabla_{i,j} \mathbf{x}| = \sqrt{(\nabla_{i,j}^h \mathbf{x})^2 + (\nabla_{i,j}^v \mathbf{x})^2}. \quad (4)$$

If i is the last row of the image, then we set $\nabla_{i,:}^v \mathbf{x} = 0$. Similarly, if j is the last column, then we set $\nabla_{:,j}^h \mathbf{x} = 0$.

We now consider the ℓ_p norm in $\varepsilon_2(\mathbf{X})$. Given that the image intensities are often piece-wise smooth across the image domain, their gradients tend to be sparse, and the non-zero elements usually correspond to the boundaries [28,29]. For constraining the fused image to exhibit similar pixel gradient distribution with the visible image, the differences of gradients between the images should be as sparse as possible; that is, the number of non-zero entries of gradient difference should be as small as possible. This condition is theoretically equivalent to minimizing the ℓ_0 norm of gradient difference,

i.e., $p = 0$. However, we must traverse all cases to obtain the ℓ_0 norm of a matrix; hence, obtaining the ℓ_0 norm is non-deterministic polynomial-time hard (NP-hard). Fortunately, for vectors, the ℓ_1 norm is the optimal convex approximation of the ℓ_0 norm and can be easily solved. The restricted isometry condition [30] also theoretically guarantees the exact recovery of sparse solution by ℓ_1 norm. For a vector $\mathbf{v} \in \mathbb{R}^{mn \times 1}$, the ℓ_1 norm is defined as $\|\mathbf{v}\|_1 = \sum_{i=1}^{mn} |\mathbf{v}_i|$. Specifically, if the energy function is the ℓ_1 norm of the gradient, it is a total variation model [10,31]. If $\varepsilon_2(\mathbf{X})$ is calculated by a norm of vectors, we can directly set the ℓ_p norm as ℓ_1 norm. However, in Equation (2), $\varepsilon_2(\mathbf{X})$ is the norm of the matrix. For the matrix $\nabla \mathbf{X} - \nabla \mathbf{Vis}$, if the optimal convex approximation of vectors above is to be applied, the norm of the matrix should be equal to the sum of absolute values of all elements. Note that $\|\nabla \mathbf{X} - \nabla \mathbf{Vis}\|_p$ is a norm on the gradient and the discrete isotropic TV-norm [32], i.e., $\|\cdot\|_{TV_I}$ is defined by the sum of gradients of all elements, which is defined as:

$$\|\mathbf{x}\|_{TV_I} = \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{(\mathbf{x}_{i,j} - \mathbf{x}_{i+1,j})^2 + (\mathbf{x}_{i,j} - \mathbf{x}_{i,j+1})^2} + \sum_{i=1}^{m-1} |\mathbf{x}_{i,n} - \mathbf{x}_{i+1,n}| + \sum_{j=1}^{n-1} |\mathbf{x}_{m,j} - \mathbf{x}_{m,j+1}|, \quad (5)$$

where \mathbf{x} is a matrix of size $m \times n$. It motivates us to convert the minimization of the ℓ_1 norm to that of the TV-norm.

In addition, another important advantage of the regularization based on $\|\cdot\|_{TV_I}$ is its insensitivity to outliers, which always correspond to sharp edges in image processing. Therefore, we use $\|\cdot\|_{TV_I}$ to define $\varepsilon_2(\mathbf{X})$ as follows:

$$\varepsilon_2(\mathbf{X}) = \|\mathbf{X} - \mathbf{Vis}\|_{TV_I}. \quad (6)$$

2.3. Formulation of the Fusion Problem

Taking comprehensive consideration of $\varepsilon_1(\mathbf{X})$ and $\varepsilon_2(\mathbf{X})$, positive regularization parameter λ is introduced herein to control the trade-off between the data fidelity term and the regularization term. Then, the fusion problem can be expressed as follows:

$$\mathbf{X} = \arg \min_{\mathbf{X}} \{\varepsilon(\mathbf{X}) = \varepsilon_1(\mathbf{X}) + \lambda \varepsilon_2(\mathbf{X})\}, \quad (7)$$

which is equivalent to:

$$\mathbf{X} = \arg \min_{\mathbf{X}} \left\{ \varepsilon(\mathbf{X}) = \frac{1}{2} \|\psi \mathbf{X} - \mathbf{IR}\|_F^2 + \lambda \|\mathbf{X} - \mathbf{Vis}\|_{TV_I} \right\}. \quad (8)$$

Thus far, the fusion problem is converted to solve the fused image \mathbf{X} in Equation (8), which minimizes the energy function $\varepsilon(\mathbf{X})$.

3. Optimization

The optimization problem (8) is clearly a convex optimization problem; therefore, a global optimal solution that minimizes $\varepsilon(\mathbf{X})$ exists. In our model, the data fidelity term is smooth, while the regularization term is non-smooth; thus, we can conclude that problem (8) is in the form of:

$$\mathbf{X} = \arg \min_{\mathbf{X}} \{F(\mathbf{X}) = f(\mathbf{X}) + g(\mathbf{X})\}, \quad (9)$$

where f is a smooth convex function, of which the gradient is Lipschitz continuous, and g is a continuous, but possibly non-smooth, convex function. Specifically,

$$f(\mathbf{X}) = \frac{1}{2} \|\psi \mathbf{X} - \mathbf{IR}\|_F^2, \quad g(\mathbf{X}) = \lambda \|\mathbf{X} - \mathbf{Vis}\|_{TV_I}. \quad (10)$$

The iterative shrinkage-thresholding algorithm (ISTA) [33] can be used to resolve the preceding problem. Although ISTA can solve general convex optimization problems, its convergence rate is proven to be only $O(1/k)$ with k being the iteration number. For optimizing the slow convergence of ISTA, Beck and Teboulle proposed FISTA [27]. Its convergence rate can be improved to $O(1/k^2)$. The iterative step in FISTA can be rewritten as follows:

$$\mathbf{X}^k = \arg \min_{\mathbf{X}} \left\{ \frac{L}{2} \left\| \mathbf{X} - \left(\mathbf{Y}^k - \frac{1}{L} \nabla f(\mathbf{Y}^k) \right) \right\|_F^2 + \lambda \|\mathbf{X} - \mathbf{Vis}\|_{TV_I} \right\}. \quad (11)$$

We denote ψ^T as the inverse operation of ψ and set $\mathbf{Y} = \mathbf{Y}^k - \psi^T(\psi\mathbf{X}^{k-1} - \mathbf{IR})/L$, where L is a Lipschitz constant of $\psi^T(\psi\mathbf{X}^{k-1} - \mathbf{IR})$, which can be set as 1 herein. By approximating $\nabla f(\mathbf{Y}^k)$ with $\nabla f(\mathbf{X}^{k-1})$ in the FISTA framework, the formula (11) then can be rewritten as:

$$\mathbf{X}^k = \arg \min_{\mathbf{X}} \left\{ \frac{L}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X} - \mathbf{Vis}\|_{TV_I} \right\}. \quad (12)$$

According to the standard FISTA framework, we can optimize Equation (8) by using Algorithm 1. From the iteration process, each iterate of FISTA depends on the previous two iterates and not only on the last iterate as in ISTA. Therefore, FISTA can achieve a fast convergence rate.

Algorithm 1: FISTA framework for DRTV.

Input: $L = 1, \lambda, t_1 = 1, \mathbf{X}^0 = \psi^T \mathbf{IR}, \mathbf{Y}^1 = \psi^T \mathbf{IR}$

Output: \mathbf{X}

- 1: **for** $k = 1$ **to** Maxiteration **do**;
 - 2: $\mathbf{Y} = \mathbf{Y}^k - \psi^T(\psi\mathbf{X}^{k-1} - \mathbf{IR})/L$;
 - 3: $\mathbf{X}^k = \arg \min_{\mathbf{X}} \left\{ \frac{L}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{X} - \mathbf{Vis}\|_{TV_I} \right\}; \quad \backslash\backslash$ solved by Algorithm 2;
 - 4: $t_{k+1} = \left[1 + \sqrt{1 + 4(t_k)^2} \right] / 2$;
 - 5: $\mathbf{Y}^{k+1} = \mathbf{X}^k + \frac{t_k - 1}{t_{k+1}} (\mathbf{X}^k - \mathbf{X}^{k-1})$;
 - 6: **end for**;
 - 7: \mathbf{X} is obtained as \mathbf{X}^k in the last iteration.
-

To solve the optimization problem in Line 3 in Algorithm 1, we set $\mathbf{Z} = \mathbf{X} - \mathbf{Vis}$. Then, we obtain the following minimization problem:

$$\mathbf{Z}^k = \arg \min_{\mathbf{Z}} \left\{ \frac{L}{2} \|\mathbf{Z} - (\mathbf{Y} - \mathbf{Vis})\|_F^2 + \lambda \|\mathbf{Z}\|_{TV_I} \right\}. \quad (13)$$

The minimization problem in Equation (13) is the problem of total variation (TV) minimization [31], and \mathbf{X}^k can be updated by $\mathbf{Z}^k + \mathbf{Vis}$. The first term reflects that \mathbf{Z} should be similar to $\mathbf{Y} - \mathbf{Vis}$ in terms of content. The second term, i.e., the TV term, is a constraint on the piece-wise smoothness of \mathbf{Z} . The smooth constraint has been proven to allow the existence of step edges; hence, the boundary is not blurred when minimizing the energy function [34].

We follow the standard procedure [32,35] for solving the TV minimization problem in Equation (13) and summarize the entire procedure in Algorithm 2. The convergence rate is also considerably superior to other methods based on gradient projection. Refer to [32] for the detailed derivation. For ease of understanding, we use the same symbol notations as those in [36]. The linear operator is defined as: $\mathcal{L}(\mathbf{R}, \mathbf{S})_{i,j} = \mathbf{R}_{i,j} - \mathbf{R}_{i-1,j} + \mathbf{S}_{i,j} - \mathbf{S}_{i,j-1}$. The corresponding inverse operator is presented as follows: $\mathcal{L}^T(\mathbf{X}) = (\mathbf{R}, \mathbf{S})$, where $\mathbf{R}_{i,j} = \mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}$, $\mathbf{S}_{i,j} = \mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}$. \mathbb{P} is a projection operator to ensure $\mathbf{R}_{i,j}^2 + \mathbf{S}_{i,j}^2 \leq 1$, $|\mathbf{R}_{i,n}| \leq 1$, and $|\mathbf{S}_{m,j}| \leq 1$.

Algorithm 2: TV minimization.

Input: $\lambda, \mathbf{Y}, \mathbf{Vis}, (\mathbf{U}^1, \mathbf{V}^1) = (\mathbf{R}^0, \mathbf{S}^0) = (\mathbf{0}, \mathbf{0}), t_1 = 1, \mathbf{B} = \mathbf{Y} - \mathbf{Vis}$
Output: \mathbf{X}

- 1: **for** $k = 1$ **to** Maxiteration **do**;
- 2: $(\mathbf{R}^k, \mathbf{S}^k) = \mathbb{P} \left[\left(\mathbf{U}^k, \mathbf{V}^k \right) + \frac{1}{8\lambda} \mathcal{L}^T \left(\mathbf{B} - \lambda \mathcal{L} \left(\mathbf{U}^k, \mathbf{V}^k \right) \right) \right];$
- 3: $t_{k+1} = \left[1 + \sqrt{1 + 4(t_k)^2} \right] / 2;$
- 4: $(\mathbf{U}^{k+1}, \mathbf{V}^{k+1}) = (\mathbf{R}^k, \mathbf{S}^k) + \frac{t_k - 1}{t_{k+1}} (\mathbf{R}^k - \mathbf{R}^{k-1}, \mathbf{S}^k - \mathbf{S}^{k-1});$
- 5: **end for**;
- 6: $\mathbf{Z} = \mathbf{B} - \lambda \mathcal{L} \left(\mathbf{R}^k, \mathbf{S}^k \right);$
- 7: $\mathbf{X} = \mathbf{Z} + \mathbf{Vis}.$

4. Experimental Results

In this section, we conduct experiments on publicly available datasets in comparison with several state-of-the-art methods to verify the effectiveness of our DRTV. We first introduce the testing data and compare methods and evaluation metrics. Then, we report the experimental results, followed by parameter analysis and computational cost comparison. All experiments are performed on a laptop with 1.9 GHz Intel Core i3 4030 CPU, 8 GB RAM, and Matlab code.

4.1. Dataset

The publicly available image dataset *TNO Human Factors*, which contains multispectral night images of different military scenarios, is used for performance evaluation. We select 20 infrared/visible images with different types of scenes for testing: *airplane_in_trees*, *helicopter*, *nato_camp_sequence*, *Farm*, *house_with_3_men*, *Kaptein_01*, *Kaptein_1123*, *Marne_02*, *Marne_03*, *Marne_11*, *Movie_01*, *Duine_7408*, *heather*, *house*, *Kaptein_1654*, *man_in_doorway*, *Movie_12*, *Movie_18*, *Movie_24*, and *Reek*.

Notably, visible and infrared images in the dataset are of the same resolution. As in real-world applications, the infrared images often present considerably low resolution. Here, we downsample the infrared images in our testing data for the experimental conditions to approximate the real-world situation. Therefore, all infrared source images present a low resolution (e.g., ratio c described in Section 2 is fixed to 2). We do not need image registration [37–42] before fusion because the source image pairs are all aligned.

We compare our DRTV with six state-of-the-art fusion methods, including pyramid transform-based methods such as Laplacian pyramid (LP) [43] and ratio of low-pass pyramid (RP) [44], multi-scale transform-based method such as HMSD [26], edge-preserving decomposition-based method such as cross-bilateral filter (CBF) [45], subspace-based method such as directional discrete cosine transform and principal component analysis (DDCTPCA) [46], and total variation-based method such as gradient transfer fusion (GTF) [10]. Parameters in these methods are set by default as suggested in the original papers.

4.2. Evaluation Metrics

In numerous cases, the visual differences between fused results are insufficiently significant; therefore, fusion methods should be evaluated comprehensively from different perspectives. At present, evaluations are mainly divided into two directions: subjective evaluation and objective evaluation.

Subjective evaluation relies on human eyes and concentrates on the perception of details, contrast, sharpness, and target highlighting. While the quality of image is also affected by other factors, our eyes are insensitive to such factors. Numerous types of objective evaluation, which are mainly based on information theory, structural similarity, image gradients, and statistics, are available. Objective evaluation is consistent with human visual perception system and subject to experimental conditions.

The first two metrics, i.e. standard deviation (SD) [47] and entropy (EN) [48], are only dependent on the fused image. The three remaining metrics, i.e. mutual information (MI) [49], structural similarity index measure ($SSIM$) [50] and Peak signal-to-noise ratio ($PSNR$) [51], are determined by the fused and source images. In this paper, if the resolution of the fused image \mathbf{X} is higher than that of the source image \mathbf{A} (either the infrared image \mathbf{IR} or the visible image \mathbf{Vis}), then we must downsample the fused image first and then calculate the metrics.

(1) SD

SD is a statistical concept reflecting distribution and contrast. SD is mathematically defined as follows:

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (\mathbf{X}(i,j) - \mu)^2},$$

where μ is the mean value of \mathbf{X} . Areas with high contrast are more likely to attract attention. Therefore, a large SD means that the fused image is improved.

(2) EN

The EN of the fused image can measure the amount of information contained in it and is mathematically defined as:

$$EN = - \sum_{l=0}^{L-1} p_l \log_2 p_l,$$

where L is the number of gray levels and p_l is the normalized histogram of the corresponding gray level. When the EN is large, additional information is contained in the fused image, and the performance of the fusion method is improved. However, EN is conventionally used as an auxiliary metric because it may be influenced by noise.

(3) MI

MI is used to measure the amount of information transmitted from the source image to the fused image. MI is defined as follows:

$$MI = MI_{\mathbf{IR},\mathbf{X}} + MI_{\mathbf{Vis},\mathbf{X}},$$

where $MI_{\mathbf{IR},\mathbf{X}}$ and $MI_{\mathbf{Vis},\mathbf{X}}$ respectively represent the amount of information transmitted from the infrared image and the visible image to the fused image. MI between two random variables can be calculated by the Kullback–Leibler metric:

$$MI_{\mathbf{A},\mathbf{X}} = \sum_{a,x} p_{\mathbf{A},\mathbf{X}}(a,x) \log_2 \frac{p_{\mathbf{A},\mathbf{X}}(a,x)}{p_{\mathbf{A}}(a)p_{\mathbf{X}}(x)},$$

where $p_{\mathbf{A}}(a)$ and $p_{\mathbf{X}}(x)$ respectively represent the edge histograms of \mathbf{A} and \mathbf{X} . $p_{\mathbf{A},\mathbf{X}}(a,x)$ is the joint histogram of \mathbf{A} and \mathbf{X} . A large MI metric means that considerable information is transferred from source images to the fused image, thereby indicating good fusion performance.

(4) $SSIM$

To model image loss and distribution, $SSIM$ was proposed by Wang et al. [52] as a universal image quality index based on structural similarity. The formula of the index is as follows:

$$SSIM_{\mathbf{A},\mathbf{X}} = \sum_{a,x} b(a,x) \cdot c(a,x) \cdot s(a,x).$$

In the formula above, to measure the structural similarity between source image \mathbf{A} and fused image \mathbf{X} , $SSIM_{\mathbf{A},\mathbf{X}}$ is calculated and it mainly consists of three components: similarities of light, contrast and structure information, denoted as $b(a,x)$, $c(a,x)$ and $s(a,x)$, respectively. The image

patches of source image \mathbf{A} and fused image \mathbf{X} in a sliding window are respectively denoted by a and x . The specific definitions of the three components are as follows:

$$b(a, x) = \frac{2\mu_a\mu_x + C_1}{\mu_a^2 + \mu_x^2 + C_1},$$

$$c(a, x) = \frac{2\sigma_a\sigma_x + C_2}{\sigma_a^2 + \sigma_x^2 + C_2},$$

$$s(a, x) = \frac{\sigma_{ax} + C_3}{\sigma_a\sigma_x + C_3},$$

where μ_a and μ_x indicate the mean values of image patches a and x , respectively, reflecting brightness information; σ_a and σ_x represent SD , reflecting contrast information; σ_{ax} denotes the covariance of source and fused images, reflecting the similarity of structure information; C_1 , C_2 , and C_3 are positive parameters used to avoid unstable phenomenon. Specifically, when $C_1 = C_2 = C_3 = 0$, the index is equivalent to the universal image quality index [53]. Thus, for all source and fused images, the index measure $SSIM$ is defined as follows:

$$SSIM = \omega_{Vis}SSIM_{Vis,X} + \omega_{IR}SSIM_{IR,X},$$

where $SSIM_{Vis,X}$ and $SSIM_{IR,X}$ indicate structural similarities between visible/infrared and fused images, respectively.

(5) PSNR

$PSNR$ is used to reflect the distortion by the fusion operation and is defined by the ratio of peak value power and noise power in the generated fused image:

$$PSNR = 10\log_{10} \frac{r^2}{MSE},$$

where r denotes the peak value of the fused image and MSE denotes the mean square error between the fused image and source images. The larger the $PSNR$, the less distortion the fusion method produces.

4.3. Results and Analysis

To maintain a considerable amount of information in source images, we select the strategy of upsampling infrared images rather than downsampling visible images for the comparison methods before fusion, that is, e.g., the procedure (c) in Figure 1. Our DRTV involves only one parameter λ , which we set to 50 as its default value.

We first conducted several qualitative comparisons of different methods on four typical image pairs, including *airplane_in_trees*, *Kaptein_01*, *nato_camp_sequence*, and *sandpath*, as shown in Figure 2. Compared with other six fusion methods, DRTV considerably improves the resolution of the fused images and does not suffer from strong noise caused by upsampling infrared images before fusion. Although the resolution of infrared images is only half of that of visible images, DRTV allows preservation of additional detailed texture information, which benefits from the downsampling operator in $\varepsilon_1(\mathbf{X})$ and the $\|\cdot\|_{TV_I}$ norm of gradient differences in $\varepsilon_2(\mathbf{X})$, in visible images. Furthermore, the fused results generated by our DRTV are similar to high-resolution infrared images with detailed scene representations, which are beneficial for detecting and recognizing potential targets.



Figure 2. Fusion results of different fusion methods on four image pairs. The four groups of results from left to right: *airplane_in_trees*, *Kaptein_01*, *nato_camp_sequence* and *sandpath*, where the corresponding subfigures are: (a) infrared image; (b) visible image; (c) our DRTV; (d) LP [43]; (e) RP [44]; (f) HMSD [26]; (g) CBF [45]; (h) DDCTPCA [46]; (i) GTF [10].

Next, we provide quantitative comparisons on the total dataset with the following five metrics described in the last section: *SSIM*, *MI*, *SD*, *EN* and *PSNR*. The results are shown in Figure 3. For the metrics *SSIM*, *MI* and *PSNR*, our DRTV can generate the largest average values and exhibits the best performance on 17, 15 and 9 image pairs, respectively. For the metric *SD*, DRTV can produce comparable results. Meanwhile, our method for the metric *EN* does not achieve the best results, which can be explained as follows. Entropy is defined as a measure of uncertainty. When the probability of each gray value is evenly distributed, the uncertainty and the entropy are both large. For our DRTV, we constrain the fused image and the infrared image to present similar intensity distribution to facilitate discrimination of the thermal target by human eyes. Therefore, our fusion result maintained the intensity distribution of the original infrared image, where the gray value is unevenly distributed as that in the visible image, leading to a small *EN* value.

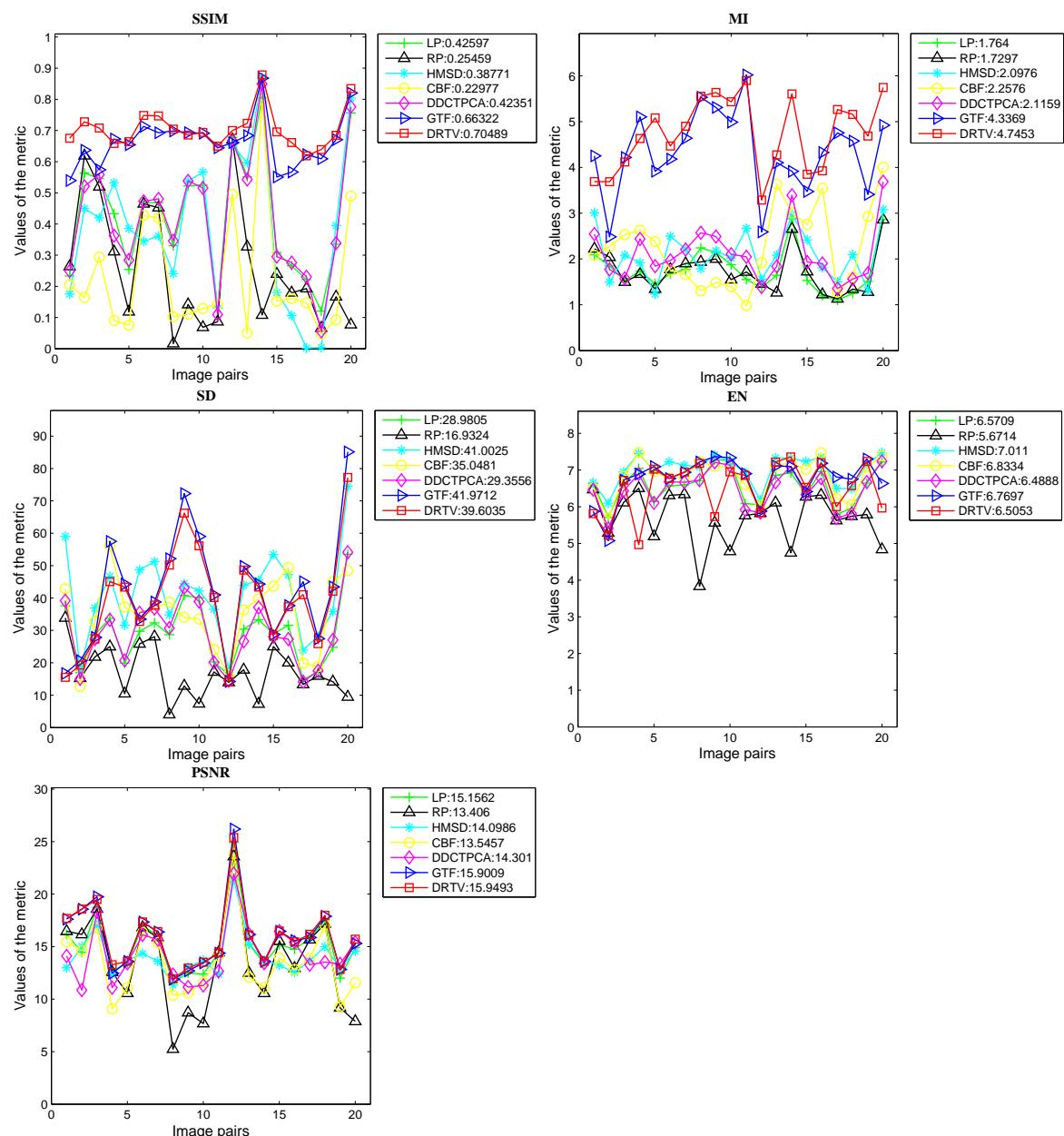


Figure 3. Quantitative comparisons of six fusion methods on twenty image pairs.

4.4. Parameter Analysis

Only one parameter is presented in our proposed method, i.e., the regularized parameter λ , which controls the trade-off between $\varepsilon_1(\mathbf{X})$ and $\varepsilon_2(\mathbf{X})$. When λ is small, the fused image preserves additional thermal radiation information in the infrared image, allowing the visual effect to be highly similar to the infrared image. Conversely, when λ is sufficiently large, the texture information in the visible image remains in the fused image to a large extent; hence, the fused image is similar to the visible image. When $\lambda \rightarrow +\infty$, the fused image is the original visible image. To analyze the influence of parameter λ on fusion performance in detail, we test different values of λ on two image pairs, including *nato_camp_sequence* and *lake*. We gradually increase the value of λ and observe the changes of the fused images. Figures 4 and 5 show the changes of the fused images of the two image pairs when λ is fixed to 1, 10, 50, 100, 200, and 500.

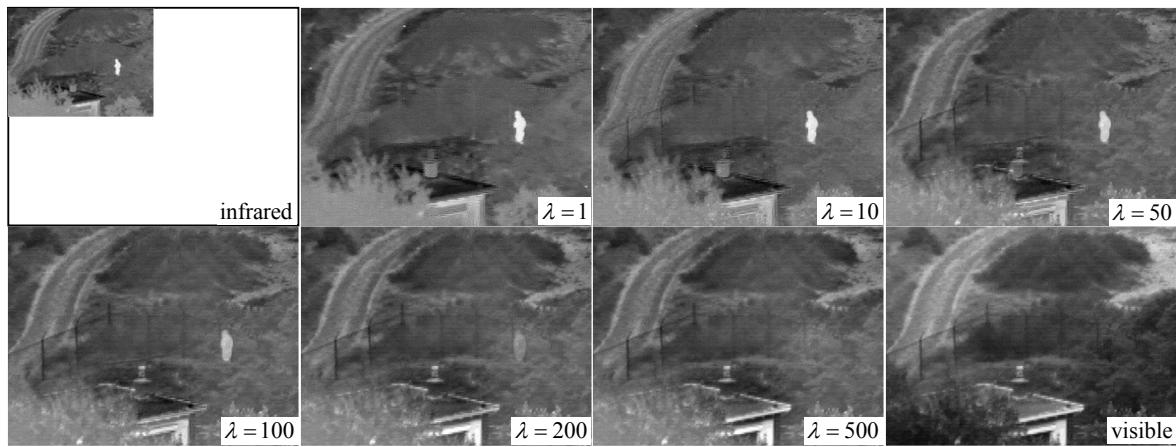


Figure 4. Fusion results of our DRTV on *nato_camp_sequence* when parameter λ increases.

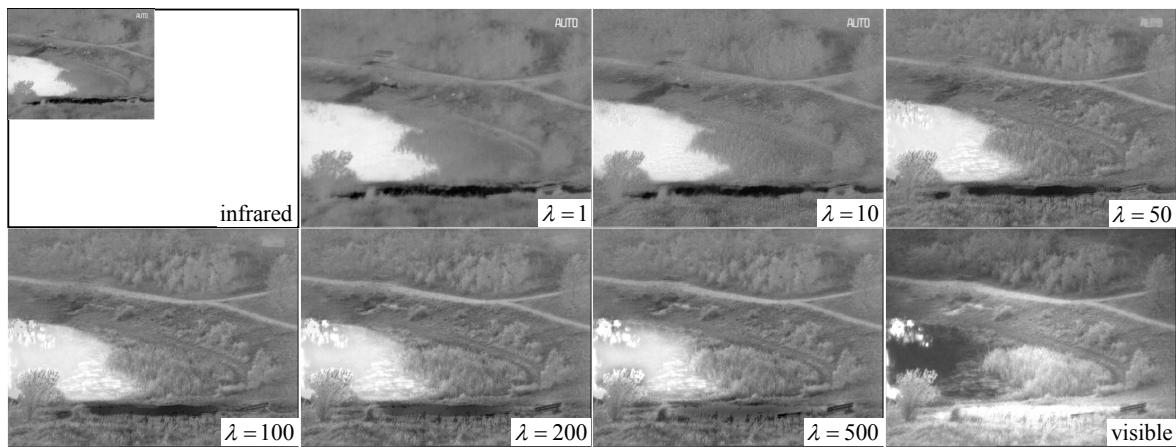


Figure 5. Fusion results of our DRTV on *lake* when parameter λ increases.

As λ changes from 1 to 500, the fence and hill around the house in Figure 4 gradually appear. In addition, the outline of the chimney in the bottom left corner becomes gradually clear. For the results in Figure 5, the most evident change is that the texture of vegetation on the river is clearly delineated, and the chair is gradually separated from the road. Thus, we can conclude that texture features in the fused image gradually become clear as the value of λ increases.

By contrast, we can observe that, in Figure 5, when λ is smaller than 200, the pedestrian in *nato_camp_sequence* is legible, and the gray distribution of the house in the bottom left corner is similar to that in the infrared image. Meanwhile, when λ is set as 500, the pedestrian's thermal radiation information is completely lost in the fused image and the house is similar to that in the visible image.

In Figure 5, the word “AUTO” in the upper right corner gradually disappears in the fused image as λ changes from 1 to 500. Therefore, we conclude that, as λ increases, thermal radiation information becomes gradually lost in the fused image. This finding is also the flexibility and advantage offered by our method. By adjusting the value of parameter λ , users can conveniently control the similarity of the fused image to the infrared image or the visible image.

4.5. Computational Cost Comparison

Given that the FISTA framework is introduced in our model to accelerate the convergence rate, to evaluate it, we compare the model solved by FISTA framework and that solved by the previous variational model, ISTA. The convergence rate is observed through the value of $f(\mathbf{X}^k)$. The objective function $f(\mathbf{X})$ is denoted by $\varepsilon(\mathbf{X})$ in Equation (8) and \mathbf{X}^k denotes the fused image obtained by the k -th iteration. Figure 6 demonstrates the convergence rate comparison corresponding to two image pairs in Figure 2, e.g., airplane and Kaptein_01, as an example. Inheriting the advantage of the FISTA framework, our method typically converges only after 20 to 50 iterations while the method based on ISTA often converges in 70 to 90 iterations. Additional iterations will increase the computational cost without substantial performance gain. To stop the iteration automatically by the value of $f(\mathbf{X}^k)$, we apply the residual ratio of $f(\mathbf{X}^k)$ to stop the iteration process. The residual ratio (RR) is defined as:

$$RR = \frac{|f(\mathbf{X}^k) - f(\mathbf{X}^{k-1})|}{f(\mathbf{X}^{k-1})}. \quad (14)$$

During the iteration process, the iteration is stopped when the value of RR is less than a predefined threshold and in our method, and the threshold is set as 0.3% empirically.

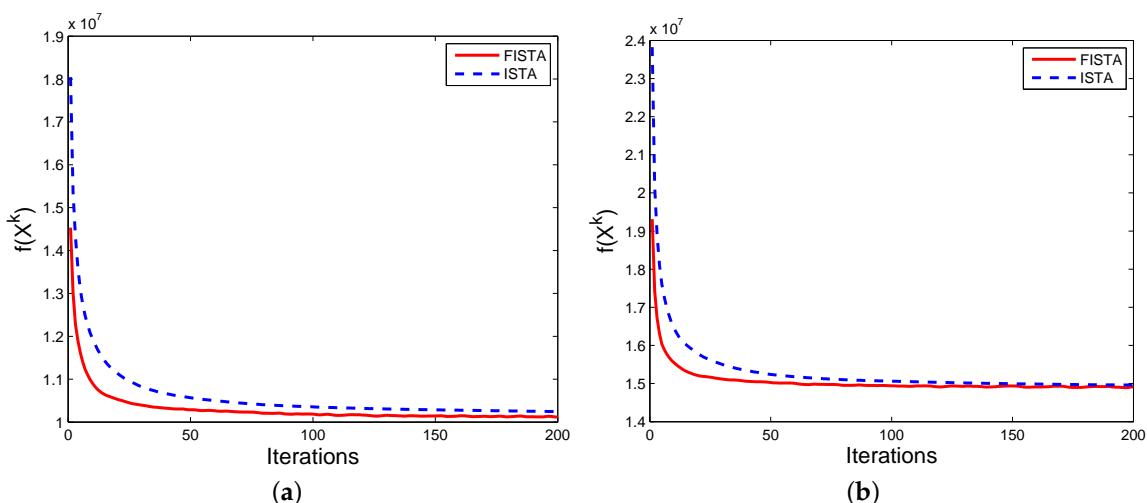


Figure 6. Convergence rate comparison between FISTA framework and previous variational model. ((a) result on *airplane_in_trees*; (b) result on *Kaptein_01*).

To evaluate the efficiency of our proposed method, we compare our method with six other methods in terms of runtime. The computational costs of seven methods on 20 image pairs are shown in Figure 7, and the average runtime of seven algorithms are reported in the legend. In DDCTPCA, the source images are divided into many non-overlapping square blocks and there are eight directional modes to be performed on each block. All the coefficients in eight modes are used for fusion. Thus, the process of obtaining and fusing the coefficients for all the blocks takes up most of the runtime. Therefore, the runtime has a lot to do with the number of blocks, i.e., the size of source images. This is the reason why the runtime of DDCTPCA is oscillating, while, in our method, we formulate the

problem as a convex optimization problem characterized by total variation model and sizes of source images have less effects on computational cost. This finding demonstrates that our method can achieve comparable efficiency.

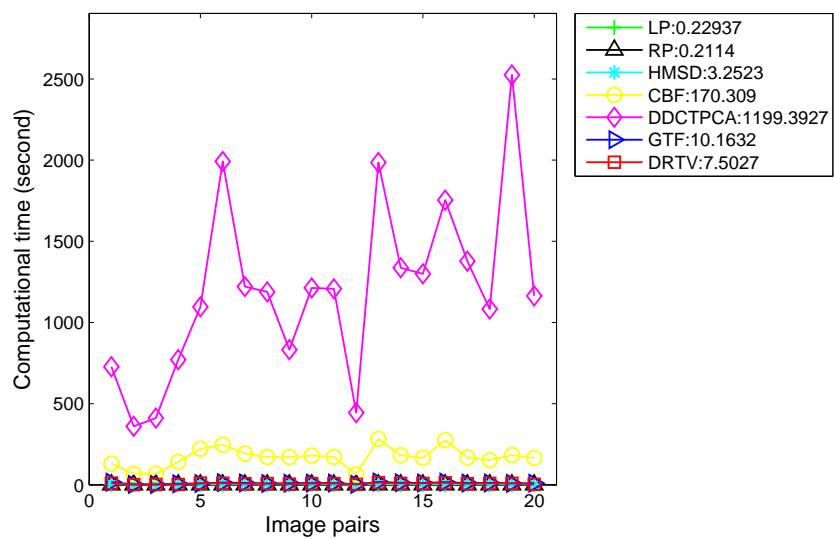


Figure 7. Runtime comparison of different methods on the whole test dataset.

5. Conclusions and Future Work

In this paper, we proposed a method for fusing infrared and visible images of different resolutions called DRTV. The fusion process is similar to conducting infrared image super-resolution, which simultaneously integrates the texture detail information in the visible image. The quantitative comparisons on several metrics with six state-of-the-art fusion methods demonstrate that our method can not only keep more detailed information in the source images, but also considerably reduce noise in the results caused by upsampling infrared images before fusion which typically occurs in existing methods. The fused image can also retain the thermal radiation information to a large extent, thereby benefiting fusion-based target detection and recognition systems.

In our objective function, we have used the first-order TV to preserve the texture detail information in the visible image. The first-order TV performs well in preserving edges of object in the piecewise constant image compared with other total variational models. However, it will produce staircase effects. For the purpose of eliminating the staircase effect in the first-order TV model, high-order models have been proposed by modifying the TV regularizer in the first order variational model to other regularizers and have shown promising performance, e.g., the TGV regularizer in [54], the bounded Hessian regularizer together with an edge diffusivity function in [55], Laplacian regularizer and regularizers combining the aforementioned regularizers with the TV regularizer [56]. Thus, it deserves attention to modify our model with high-order variational models to remove staircase effects in the follow-up work.

Author Contributions: Q.D., H.X. and F.F. have made substantial contributions to the algorithm design. H.X. has made substantial contributions to the algorithm implementation and manuscript preparation. Q.D. has made substantial contributions to the experimental data collection and the result analysis. Q.D., Y.M. and J.H. revised the final manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant No. 61605146.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [[CrossRef](#)]

2. Huang, Y.; Bi, D.; Wu, D. Infrared and Visible Image Fusion Based on Different Constraints in the Non-Subsampled Shearlet Transform Domain. *Sensors* **2018**, *18*, 1169. [[CrossRef](#)] [[PubMed](#)]
3. Kong, S.G.; Heo, J.; Boughorbel, F.; Zheng, Y.; Abidi, B.R.; Koschan, A.; Yi, M.; Abidi, M.A. Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. *Int. J. Comput. Vis.* **2007**, *71*, 215–233. [[CrossRef](#)]
4. Gao, Y.; Ma, J.; Yuille, A.L. Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples. *IEEE Trans. Image Process.* **2017**, *26*, 2545–2560. [[CrossRef](#)] [[PubMed](#)]
5. Yang, M.; Su, T.; Lin, H. Fusion of Infrared Thermal Image and Visible Image for 3D Thermal Model Reconstruction Using Smartphone Sensors. *Sensors* **2018**, *18*, 2003. [[CrossRef](#)] [[PubMed](#)]
6. Singh, R.; Vatsa, M.; Noore, A. Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognit.* **2008**, *41*, 880–893. [[CrossRef](#)]
7. Ma, J.; Jiang, J.; Liu, C.; Li, Y. Feature guided Gaussian mixture model with semi-supervised EM and local geometric constraint for retinal image registration. *Inf. Sci.* **2017**, *417*, 128–142. [[CrossRef](#)]
8. Liu, T.; Liu, H.; Chen, Z.; Lesgold, A.M. Fast blind instrument function estimation method for industrial infrared spectrometers. *IEEE Trans. Ind. Inform.* **2018**. [[CrossRef](#)]
9. Bavirisetti, D.P.; Dhuli, R. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Phys. Technol.* **2016**, *76*, 52–64. [[CrossRef](#)]
10. Ma, J.; Chen, C.; Li, C.; Huang, J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf. Fusion* **2016**, *31*, 100–109. [[CrossRef](#)]
11. Pajares, G.; De La Cruz, J.M. A wavelet-based image fusion tutorial. *Pattern Recognit.* **2004**, *37*, 1855–1872. [[CrossRef](#)]
12. Dogra, A.; Goyal, B.; Agrawal, S. From Multi-Scale Decomposition to Non-Multi-Scale Decomposition Methods: A Comprehensive Survey of Image Fusion Techniques and Its Applications. *IEEE Access* **2017**, *5*, 16040–16067. [[CrossRef](#)]
13. Li, S.; Yin, H.; Fang, L. Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 3450–3459. [[CrossRef](#)] [[PubMed](#)]
14. Zhang, Q.; Liu, Y.; Blum, R.S.; Han, J.; Tao, D. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Inf. Fusion* **2018**, *40*, 57–75. [[CrossRef](#)]
15. Kong, W.; Zhang, L.; Lei, Y. Novel fusion method for visible light and infrared images based on NSST-SF-PCNN. *Infrared Phys. Technol.* **2014**, *65*, 103–112. [[CrossRef](#)]
16. Liu, Y.; Chen, X.; Cheng, J.; Peng, H.; Wang, Z. Infrared and visible image fusion with convolutional neural networks. *Int. J. Wavelets Multiresolut. Inf. Process.* **2018**, *16*, 1850018. [[CrossRef](#)]
17. Bavirisetti, D.P.; Xiao, G.; Liu, G. Multi-sensor image fusion based on fourth order partial differential equations. In Proceedings of the International Conference on Information Fusion, Xi'an, China, 10–13 July 2017; pp. 1–9.
18. Patil, U.; Mudengudi, U. Image fusion using hierarchical PCA. In Proceedings of the International Conference on Image Information Processing, Shimla, India, 3–5 November 2011; pp. 1–6.
19. Zhang, X.; Ma, Y.; Fan, F.; Zhang, Y.; Huang, J. Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition. *J. Opt. Soc. Am. A* **2017**, *34*, 1400–1410. [[CrossRef](#)] [[PubMed](#)]
20. Shibata, T.; Tanaka, M.; Okutomi, M. Visible and near-infrared image fusion based on visually salient area selection. In Proceedings of the Digital Photography XI. International Society for Optics and Photonics, San Francisco, CA, USA, 27 February 2015; Volume 9404, p. 94040G.
21. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **2015**, *24*, 147–164. [[CrossRef](#)]
22. Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys. Technol.* **2017**, *82*, 8–17. [[CrossRef](#)]
23. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [[CrossRef](#)]
24. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* **2018**, *42*, 158–173. [[CrossRef](#)]
25. Ma, Y.; Chen, J.; Chen, C.; Fan, F.; Ma, J. Infrared and visible image fusion using total variation model. *Neurocomputing* **2016**, *202*, 12–19. [[CrossRef](#)]

26. Zhou, Z.; Wang, B.; Li, S.; Dong, M. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters. *Inf. Fusion* **2016**, *30*, 15–26. [CrossRef]
27. Beck, A.; Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [CrossRef]
28. Chen, C.; Li, Y.; Liu, W.; Huang, J. Image fusion with local spectral consistency and dynamic gradient sparsity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2760–2765.
29. Mei, X.; Ma, Y.; Li, C.; Fan, F.; Huang, J.; Ma, J. Robust GBM hyperspectral image unmixing with superpixel segmentation based low rank and sparse representation. *Neurocomputing* **2018**, *275*, 2783–2797. [CrossRef]
30. Candes, E.J.; Romberg, J.K.; Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **2006**, *59*, 1207–1223. [CrossRef]
31. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [CrossRef]
32. Beck, A.; Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **2009**, *18*, 2419–2434. [CrossRef] [PubMed]
33. Daubechies, I.; Defrise, M.; Mol, C.D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.* **2003**, *57*, 1413–1457. [CrossRef]
34. Wang, C. Research on Image Processing Technology Based on Variational Problems and Partial Differential Equations. Ph.D. Thesis, Science and Technology University of China, Hefei, China, 2007.
35. Chen, C.; Li, Y.; Liu, W.; Huang, J. SIRF: Simultaneous Satellite Image Registration and Fusion in a Unified Framework. *IEEE Trans. Image Process.* **2015**, *24*, 4213. [CrossRef] [PubMed]
36. Chen, C.; Li, Y.; Huang, J. Calibrationless parallel MRI with joint total variation regularization. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; pp. 106–114.
37. Ma, J.; Zhao, J.; Tian, J.; Bai, X.; Tu, Z. Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.* **2013**, *46*, 3519–3532. [CrossRef]
38. Ma, J.; Zhao, J.; Jiang, J.; Zhou, H.; Guo, X. Locality preserving matching. *Int. J. Comput. Vis.* **2018**. [CrossRef]
39. Yan, J.; Cho, M.; Zha, H.; Yang, X.; Chu S.M. Multi-graph matching via affinity optimization with graduated consistency regularization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 1228–1242. [CrossRef] [PubMed]
40. Yan, J.; Li, C.; Li, Y.; Cao, G. Adaptive discrete hypergraph matching. *IEEE Trans. Cybern.* **2018**, *48*, 765–779. [CrossRef] [PubMed]
41. Ma, J.; Wu, J.; Zhao, J.; Jiang, J.; Zhou, H.; Sheng, Q.Z. Nonrigid Point Set Registration with Robust Transformation Learning under Manifold Regularization. *IEEE Trans. Neural Netw. Learn. Syst.* to be published. [CrossRef] [PubMed]
42. Ma, J.; Jiang, J.; Zhou, H.; Zhao, J.; Guo, X. Guided locality preserving feature matching for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4435–4447. [CrossRef]
43. Burt, P.J.; Adelson, E.H. The Laplacian Pyramid as a Compact Image Code. *Read. Comput. Vis.* **1987**, *31*, 671–679.
44. Toet, A. Image fusion by a ratio of low-pass pyramid. *Pattern Recognit. Lett.* **1989**, *9*, 245–253. [CrossRef]
45. Kumar, B.K.S. Image fusion based on pixel significance using cross bilateral filter. *Signal Image Video Process.* **2015**, *9*, 1193–1204. [CrossRef]
46. Naidu, V.P.S. Hybrid DDCT-PCA based multi sensor image fusion. *J. Opt.* **2014**, *43*, 48–61. [CrossRef]
47. Hu, H.M.; Wu, J.; Li, B.; Guo, Q.; Zheng, J. An Adaptive Fusion Algorithm for Visible and Infrared Videos Based on Entropy and the Cumulative Distribution of Gray Levels. *IEEE Trans. Multimed.* **2017**, *19*, 2706–2719. [CrossRef]
48. Zhou, X.; Wang, W. *Infrared and Visible Image Fusion Based on Tetrolet Transform*; Springer: Berlin/Heidelberg, Germany, 2016.
49. Zhou, Y.; Mayyas, A.; Omar, M.A. Principal Component Analysis-Based Image Fusion Routine with Application to Automotive Stamping Split Detection. *Res. Nondestruct. Eval.* **2011**, *22*, 76–91. [CrossRef]
50. Zhao, C.; Shao, G.; Ma, L.; Zhang, X. Image fusion algorithm based on redundant-lifting NSWMDA and adaptive PCNN. *Optik-Int. J. Light Electron Opt.* **2014**, *125*, 6247–6255. [CrossRef]
51. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition (ICPR), Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.

52. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
53. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [[CrossRef](#)]
54. Duan, J.; Lu, W.; Tench, C.; Gottlob, I.; Proudlock, F.; Samani, N.N.; Bai, L. Denoising optical coherence tomography using second order total generalized variation decomposition. *Biomed. Signal Process. Control* **2016**, *24*, 120–127. [[CrossRef](#)]
55. Duan, J.; Qiu, Z.; Lu, W.; Wang, G.; Pan, Z.; Bai, L. An edge-weighted second order variational model for image decomposition. *Digit. Signal Process.* **2016**, *49*, 162–181. [[CrossRef](#)]
56. Lu, W.; Duan, J.; Qiu, Z.; Pan, Z.; Liu, R.W.; Bai, L. Implementation of high-order variational models made easy for image processing. *Math. Methods Appl. Sci.* **2016**, *39*, 4208–4233. [[CrossRef](#)]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).