

Reflectance estimation for infrared and visible image fusion

Yan Gu¹, Feng Yang¹, Weijun Zhao¹, Yiliang Guo¹, Chaobo Min^{2*}

¹North Night Vision Technology Corp., Ltd., Nanjing, 211102, China
[E-mail: jsguyan@163.com]

²The College of Internet of Things Engineering, HoHai University, Changzhou, 213000, China
[E-mail: chaobomin@outlook.com]

*Corresponding author: Chaobo Min

*Received November 15, 2020; revised March 29, 2021; accepted July 3, 2021;
published August 31, 2021*

Abstract

The desirable result of infrared (IR) and visible (VIS) image fusion should have textural details from VIS images and salient targets from IR images. However, detail information in the dark regions of VIS image has low contrast and blurry edges, resulting in performance degradation in image fusion. To resolve the troubles of fuzzy details in dark regions of VIS image fusion, we have proposed a method of reflectance estimation for IR and VIS image fusion. In order to maintain and enhance details in these dark regions, dark region approximation (DRA) is proposed to optimize the Retinex model. With the improved Retinex model based on DRA, quasi-Newton method is adopted to estimate the reflectance of a VIS image. The final fusion outcome is obtained by fusing the DRA-based reflectance of VIS image with IR image. Our method could simultaneously retain the low visibility details in VIS images and the high contrast targets in IR images. Experiment statistic shows that compared to some advanced approaches, the proposed method has superiority on detail preservation and visual quality.

Keywords: Image fusion, reflectance estimation, Retinex, infrared, detail preservation

1. Introduction

Infrared (IR) images reflecting the thermal radiation emitted by objects are influenced little by illumination variations and disguises [1]. Thus, IR imaging has been applied widely, such as medical imaging [2], night vision [3] and remote sensing [4], etc. However, it typically suffers from the loss of image details and textures. In contrast, visible (VIS) images could easily capture the information of the reflectance in target scenes. Hence, VIS images have rich details and conform well to the human visual system. Nevertheless, VIS images are sensitive to scene variations such as illumination, smoke, rain and disguises so that the targets in VIS images maybe not salient. Image fusion is put forward to produce a single fused image with both abundant details of VIS images and salient target areas of IR images, harnessing adequately of complementary information from IR and VIS images. Therefore, our major concern of this work is IR and VIS image fusion.

In recent years, numerous VIS and IR image fusion approaches have been proposed as the demand is burgeoning. Generally, fusion algorithms could be classified into six categories [1]: sparse representation, subspace, multi-scale transform, neural network, saliency and hybrid models. Studies of sparse representation [5,6] think that an image is a linearly combination of sparse basis in an over-complete dictionary. As a result, a source image can be resolved into two kinds of layers: detail and base, and following that, the different layers of two to-be-fused images are fused by different fusion regulations. Subspace-based methods [7,8] can extract the intrinsic structures of the source image by projecting high-dimensional image into low-dimensional subspace, in order to remove redundant information from the fused image. Multi-scale transformation [9] are actively employed for image fusion, making an assumption that a source image could be resolved into a variety of layers. The different layers can be fused by certain particular fusion tactic in order to obtain the final fusion. Some transform models, such as pyramid [10], wavelet [11], curvelet [12], Gaussian filters [13] and non-subsampled contourlet transform (NSCT) [14], are often applied for multi-scale decomposition. With the springing up of deep learning techniques, neural network, imitating the perceptual behaviour system of the human brain, performs well on image fusion because of strong adaptability, fault tolerance and noise immunity [15,16]. Saliency-based methods [17,18] could enhance the visual effect of the fused image by using human visual attention. The sixth category is hybrid methods [19,20] combining the superior properties of various fusion algorithms. Take an instance, Yin et al. [21] proposed a hybrid strategy for IR and VIS image fusion based on sparse representation and shift-invariant dual-tree complex shearlet transformation. Recently, with the development of deep learning, some relative fusion approaches have emerged. Convolutional neural network [22,23] and generative adversarial network [24] have been applied successfully.

Essentially, the abovementioned approaches pay special attention to the extraction of detail information from IR and VIS images to ensure rich textures of the fused outcome. However, as for IR images, people are more interested in salient targets than details. Therefore, the high-contrast property in IR images is more valuable to image fusion. If only textures are considered for image fusion, the benefit of the thermal radiation in IR images will disappear. Several methods are proposed for this issue, such as gradient transfer fusion (GTF) [25], fusion generative adversarial network [26], image fusion via detail preserving adversarial learning [1]. The main idea of these methods is that the detail information from VIS images is fused with the intensity information of the thermal radiation in IR images. Thus, the performance of detail preservation directly determines fusion performance. However, the visual quality of VIS

images is greatly influenced by illumination conditions. In the bright regions of VIS images, the textures of objects are clear so that they are easily captured and remained in the fused images. Nevertheless, the detail information of dark regions suffer from low contrast, blurry edges and high noise. These raise a tough challenge for detail preservation and will blur the targets in fusion results.

Retinex decomposition [27], regarding the captured picture as the outcome of illumination and reflectance layers, is often used for reflectance estimation to recover details from the captured images. The Retinex model has been applied well to image enhancement, dehazing, fusion and denoising [28]. However, since the details of dark regions in the source images are easily lost in reflectance recovery using the Retinex model [29], the performance of detail preservation on image fusion is limited.

In this work, IR and VIS image fusion via evaluating reflectance with dark region approximation (DRA) is introduced to overcome the abovementioned challenges. At first, the DRA is proposed for the Retinex model to preserve more details of those dark regions in VIS images when processing reflectance estimation. After that, derived from Retinex model with DRA, reflectance estimation could be formulated as an unconstrained optimization problem through Gaussian field criterion and then solved by the quasi-Newton method. Finally, DRA-based reflectance estimated from VIS image is fused with IR image by a strategy called Gamma correction. Qualitative and quantitative comparisons of the method in this paper and 3 advanced strategies are conducted on a public dataset. Unlike the counterparts, our fusion work can illuminate the dark regions of VIS images and produce the fused images with highlighted details that have even low visibility in the source images.

The contributions of this paper lie in two aspects:

- 1) Reflectance estimation on the DRA-version is developed to exploit detail information from the bright and dark regions of VIS images via two different imaging models respectively. Consequently, our fused images have abundant details and high visibility.

- 2) Because reflectance estimation based on the DRA is formulated as an unconstrained optimization problem that can be solved by gradient-based numerical optimization technology, our method has a straightforward structure and is easy to be implemented in real applications.

The rest of this paper is unfolded as following. Section 2 presents some basic theories and implementation details of our method. Experiment results and discussion are shown in Section 3. Finally, the concluding remarks are given in Section 4.

2. Methodology

In this part, we describe the basic theory of the DRA and present the implementation details of the proposed method. Fig. 1 shows the framework of our approach.

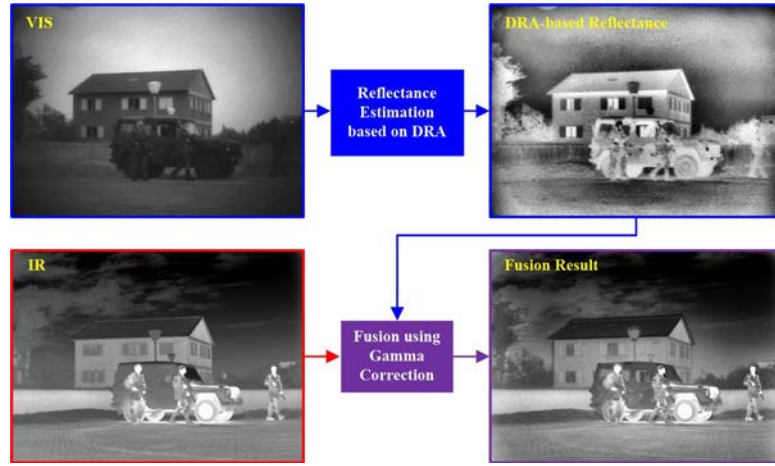


Fig. 1. Framework of the proposed method. Firstly, utilizing DRA, the reflectance is estimated from the VIS image. The IR image is then fused with the DRA-based reflectance layer of the VIS image by a fusion strategy based on Gamma correction.

2.1 Dark Region Approximation (DRA)

In the Retinex model, one VIS image denoted by \mathbf{L} is the element-wise multiplication of two layers, namely an illumination layer \mathbf{T} and a reflectance layer \mathbf{R}

$$\mathbf{L} = \mathbf{R} \circ \mathbf{T}, \quad (1)$$

Eq. (1) could be expressed in logarithmic domain, given as

$$\hat{\mathbf{L}} = \hat{\mathbf{R}} + \hat{\mathbf{T}}, \quad (2)$$

in which

$$\hat{\mathbf{L}} = \log(\mathbf{L}),$$

$$\hat{\mathbf{R}} = \log(\mathbf{R}), \text{ thought as a map describing detailed structures of the object.}$$

$$\hat{\mathbf{T}} = \log(\mathbf{T}).$$

In some previous-proposed Retinex image enhancement algorithm, like single-scale Retinex (SSR) [30] and multi-scale Retinex (MSR) [31], the reflectance $\hat{\mathbf{R}}$ is $\hat{\mathbf{L}}$ minus its Gaussian-filtered value $\hat{\mathbf{T}}$ (that is, $\hat{\mathbf{R}} = \hat{\mathbf{L}} - \hat{\mathbf{T}}$). Whereas, the decomposition of VIS image into bright and dark regions gives rise to practical problems about Retinex model so that above expression $\hat{\mathbf{R}} = \hat{\mathbf{L}} - \hat{\mathbf{T}}$ changes to

$$\hat{\mathbf{R}} = \hat{\mathbf{R}}_L + \hat{\mathbf{R}}_H = (\hat{\mathbf{L}}_L - \hat{\mathbf{T}}_L) + (\hat{\mathbf{L}}_H - \hat{\mathbf{T}}_H), \quad (3)$$

$\hat{\mathbf{R}}$, $\hat{\mathbf{L}}$ and $\hat{\mathbf{T}}$ with certain subscript means these are relative region of $\hat{\mathbf{R}}$, $\hat{\mathbf{L}}$ and $\hat{\mathbf{T}}$ respectively. Subscript L means the dark region while H indicates the bright region. Thanks to good illumination, the discrepancy between $\hat{\mathbf{L}}_H$ and $\hat{\mathbf{T}}_H$ are apparent. That is to say, the structural details $\hat{\mathbf{R}}_H$ could be remained well in $\hat{\mathbf{R}}$. However, due to the weak intensity of $\hat{\mathbf{L}}_L$, the intensity of $\hat{\mathbf{T}}_L$ merely stays at a low level. As a result, there is no significant discrepancy between $\hat{\mathbf{L}}_L$ and $\hat{\mathbf{T}}_L$ in comparison with $\hat{\mathbf{R}}_H$, which demonstrates that the detail loss of dark regions (, losing $\hat{\mathbf{R}}_L$,) is more likely to happen in $\hat{\mathbf{R}}$. **Fig. 2** is a descriptive graph of Retinex model based reflectance estimation. It could be seen that details in the blue box (a bright region) is more distinguishable and apparent than those in the red box (a dark region). Correspondingly, the direct reflectance estimation in logarithmic space loses the details we need [26].

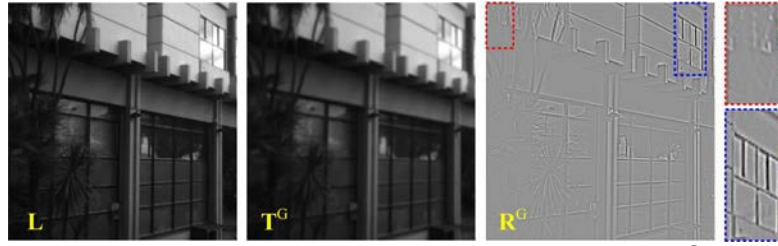


Fig. 2. Descriptive graph of Retinex model-based reflectance estimation. T^G , a Gaussian-filtered form of L , is the illumination layer. An estimated reflectance R^G is $\hat{L} - \log(T^G)$. The red and blue box is dark and bright area separately.

Observing the dark regions, then we put forward DRA: assuming there is few illuminations in the dark area i.e. $\hat{T}_L = 0$. Then Eq. (3) is modified as Retinex model in DRA version as (4):

$$\hat{R} = \hat{R}_L + \hat{R}_H \approx \hat{L}_L + (\hat{L}_H - \hat{T}_H). \quad (4)$$

With DRA's participation, the intensity distribution in dark areas could be thought as the reflectance in VIS images. Examining Eq. (3) and (4), it looks obvious that $\hat{L}_L > \hat{L}_L - \hat{T}_L$, indicating the details of dark regions in \hat{R} has been enhanced. The DRA-based Retinex model reconciles the details of bright and dark regions in \hat{R} .

But how to distinguish \hat{L}_L and \hat{L}_H ? This hanging matter prevents model (4) from being utilized directly on reflectance recovery. In the following section, we will solve this issue.

2.2 DRA-based reflectance estimation

2.2.1 Objective function

It is common sense that illuminating an object partly arouses reflectance so that we defined the reflectance \hat{R} as a map function ψ relying on the illumination layer \hat{T} and then (2) is rewritten as

$$\hat{L} = \psi(\hat{T}) + \hat{T}. \quad (5)$$

With Eq. (5), adopting Gaussian field criterion to construct an objective function for solving the map function ψ . The objective function based on Gaussian-field is given by

$$\min_{\psi} E(\psi) = \min_{\psi} - \sum_{i=1}^M \exp \left\{ -\frac{[\hat{L}_i - (\psi(\hat{T}_i) + \hat{T}_i)]^2}{2\sigma^2} \right\}, \quad (6)$$

σ is a range parameter, M is pixels' quantity in \hat{L} . \hat{L}_i and \hat{T}_i is the intensity of the i -th pixel in \hat{L} and \hat{T} , separately. The objective function emphasizes the closeness between $\hat{L} - \hat{T}$ and map function $\psi(\hat{T})$. With the merits of continuous differentiability and convenient computation, Gaussian field criterion is a good distance measurement. Otherwise, this criterion in (6) lays a solid grounding on the distinguishment of the two regions in single-channel images, about which would be talked later.

2.2.2 Reflectance model

In our research, reflectance is seen as one kind of transformation of illumination. It is supposed that a regular pattern of the transformation may exist in VIS image. Therefore, the previous mentioned map function ψ (, that is, the reflectance model) could be given as (7)

$$\psi_N(\hat{T}_i) = \sum_{n=1}^N \left(\sum_{k=0}^n \alpha_{nk} x_i^k y_i^{n-k} + \beta_n \hat{T}_i^n \right), \quad (7)$$

α_{nk} and β_n : reflectance parameters.

N : order of the reflectance model.

$[x_i, y_i]^T$: coordinate vector of the i^{th} pixel.

The first term of (7) represents the spacial distribution of the reflectance. Another term expresses the regular pattern of the transformation from illumination to reflectance.

Fundamentally, the reflectance model is a mixture of diverse polynomials depending on intensity values and coordinate vectors, whose highly nonlinearity helps to represent the complex pattern in reflectance.

Write (7) in the matrix form:

$$\psi_N(\hat{\mathbf{T}}_i) = \mathbf{C}^N \mathbf{P}_i^N = [\mathbf{c}_c^N \mid \mathbf{c}_d^N] [\mathbf{p}_{ci}^N \mid \mathbf{p}_{di}^N]^T, \quad (8)$$

where $\mathbf{c}_c^N = [\alpha_{10}, \alpha_{11}, \alpha_{20}, \alpha_{21}, \alpha_{22}, \dots, \alpha_{NN}]$ is the $1 \times N_p$ ($N_p = N(N+3)/2$) dimensional vector including every α_{nk} of the first term in Eq. (7). $\mathbf{c}_d^N = [\beta_1, \beta_2, \dots, \beta_N]$ is the $1 \times N$ dimensional vector including every β_n of the second term in (7). Thus \mathbf{C}^N is the $1 \times (N_p + N)$ dimensional reflectance parameter vector. \mathbf{p}_{ci}^N is the $1 \times N_p$ dimension vector including all $x_i^k y_i^{n-k}$ and \mathbf{p}_{di}^N is the $1 \times N$ -dimension vector including every $\hat{\mathbf{T}}_i^n$. Therefore, \mathbf{P}_i^N is the $(N_p + N) \times 1$ dimensional polynomial vector of the i -th pixel in $\hat{\mathbf{T}}$. $[\cdot]^T$ is matrix transposition.

Substituting (8) into (6), the optimization function would be

$$\min_{\mathbf{C}^N} E(\mathbf{C}^N) = \min_{\mathbf{C}^N} - \sum_{i=1}^M \exp \left\{ - \frac{(\hat{\mathbf{L}}_i - \mathbf{C}^N \mathbf{P}_i^N - \hat{\mathbf{T}}_i)^2}{2\sigma^2} \right\}. \quad (9)$$

2.2.3 Optimization

It could be observed that Eq. (9) is continuously differentiable to \mathbf{C}^N , the reflectance parameter. Owing to the proposed model in the form of polynomial, it is not hard to derive (9) correspondingly as below:

$$\frac{\partial E(\mathbf{C}^N)}{\partial \mathbf{C}^N} = \frac{1}{\sigma^2} \sum_{i=1}^M \mathbf{P}_i^N (\mathbf{C}^N \mathbf{P}_i^N + \hat{\mathbf{T}}_i - \hat{\mathbf{L}}_i) \exp \left\{ - \frac{(\hat{\mathbf{L}}_i - \mathbf{C}^N \mathbf{P}_i^N - \hat{\mathbf{T}}_i)^2}{2\sigma^2} \right\}. \quad (10)$$

With the derivation (10), it could be accepted that solving \mathbf{C}^N , the optimal parameter, with gradient-based numerical optimization approaches, for instance, quasi-Newton method [32]. Prior to that, the acquisition of the illumination layer \mathbf{T} is still a hard nut to crack in estimating reflectance. It is an ill-posed problem to estimate the illumination and reflectance layers from the captured image in the meantime [33], which objective function (9) and derivation (10) both fail to solve, making the illumination layer determination must be earlier than reflectance estimation.

2.2.4 Illumination layer calculation

According to objective function (9), it is obvious that we deem reflectance estimation to a fitting problem. The captured image $\hat{\mathbf{L}}$ and the illumination $\hat{\mathbf{T}}$ are known data. The reflectance model (7) is a fitting function while the objective function is a fitting criterion aiming at making the reflectance model approximate to $\hat{\mathbf{L}} - \hat{\mathbf{T}}$. Since this model is made up of polynomials, it could illustrate the regular pattern of the reflectance, $\hat{\mathbf{R}} = \hat{\mathbf{L}} - \hat{\mathbf{T}}$. In some previous-proposed Retinex image enhancement algorithm like SSR and MSR, the illumination layer is often gotten from the Gaussian-filtered captured image. If the objective function is optimized by Gaussian-filtered illumination layer acquired by $\hat{\mathbf{T}}^G = \log(\mathbf{T}^G)$, the reflectance estimation might have similar defect to direct-estimated reflectance as in Section 2.1.

The penalty curve plotted in **Fig. 3** shows the Gaussian criterion is pretty tolerant to large x (e.g. $1 - \exp \{-x^2/(2\sigma^2)\}$), indicating the property of poor response to large x , with which coarse blur is introduced to obtain the illumination and to estimate reflectance by DRA-based Retinex model (4) then.

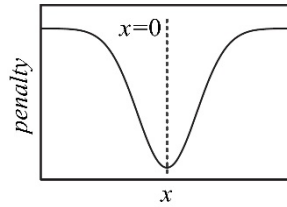


Fig. 3. Penalty curve of Gaussian criterion

To be frankly, the process of coarse blur is pretty straightforward. Make the pixel value zero if the parity of its x and y coordinates numbers are the same, and then the Gaussian-based illumination layer $\hat{\mathbf{T}}^G$ is further degraded (as shown in **Fig. 4**). $\hat{\mathbf{T}}_i^C$ is the pixel value in the coarse-blur-based illumination layer $\hat{\mathbf{T}}^C$. When optimizing the objective function (9) with $\hat{\mathbf{T}}_i^C \neq 0$ pixels, we fit the reflectance model to the differences between $\hat{\mathbf{L}}$ and $\hat{\mathbf{T}}^G$. On condition that $\hat{\mathbf{T}}_i^C$ is non-zero, the reflectance model is more willing to show reflectance's regular pattern of bright area with Retinex model.

At the same time, if $\hat{\mathbf{T}}_i^C = 0$, the term $[\hat{\mathbf{L}}_i - (\psi(\hat{\mathbf{T}}_i^C) + \hat{\mathbf{T}}_i^C)]^2$ in the objective function would be $[\hat{\mathbf{L}}_i - \psi(\hat{\mathbf{T}}_i^C)]^2$. The objective function responses weakly to large numbers of $[\hat{\mathbf{L}}_i - \psi(\hat{\mathbf{T}}_i^C)]^2$ in **Fig. 3**. Thus, we could fit the reflectance to the weak intensities of the captured picture by arranging a suitable value of the range parameter σ . This is to say, when $\hat{\mathbf{T}}_i^C = 0$, the reflectance model is prone to show the intensity distribution's regular pattern in the dark regions.

With Eq. (9), we could initially set a soft threshold to make decomposition of a single-channel picture into bright and dark areas. Followed by modelling the reflectance layer of these areas by $\hat{\mathbf{R}}_H = \hat{\mathbf{L}}_H - \hat{\mathbf{T}}_H$ and $\hat{\mathbf{R}}_L = \hat{\mathbf{L}}_L$ respectively. Ending with reconciling the reflectance's regular patterns in bright and dark areas with the reflectance model. Therefore, the objective function attaching the coarse-blurred illumination layer estimates the optimal reflectance model and it is this model that could express the regular patterns with $\hat{\mathbf{R}} = \hat{\mathbf{L}}_L + (\hat{\mathbf{L}}_H - \hat{\mathbf{T}}_H)$, DRA-based Retinex model.

After determining the optimal reflectance model, DRA-based reflectance layer \mathbf{R}^C is

$$\mathbf{R}^C = \exp\{\psi_N^C(\hat{\mathbf{L}})\}, \quad (11)$$

where ψ_N^C is the N-dimensional optimal reflectance model.

Making a comparison between the DRA-estimated reflectance estimation in **Fig. 4** and the direct-estimated reflectance in **Fig. 2**, it is salient that DRA-based reflectance makes details in bright regions more clear and illuminates those in dark regions as well.

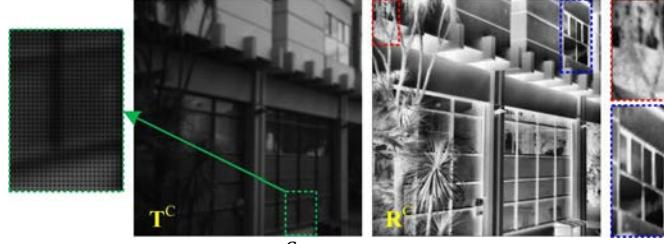


Fig. 4. DRA-based reflectance estimating. T^C is the coarse-blurred illumination layer from T^G . The red rectangle means a dark region and the blue rectangle means a bright region.

Fig. 5 presents several DRA-based results of reflectance estimation of VIS images. We could see that the DRA-based reflectance layers have abundant details and good contrast both in the dark and the bright regions. Some details hard to observe in the source images are more salient in the DRA-based reflectance layers.

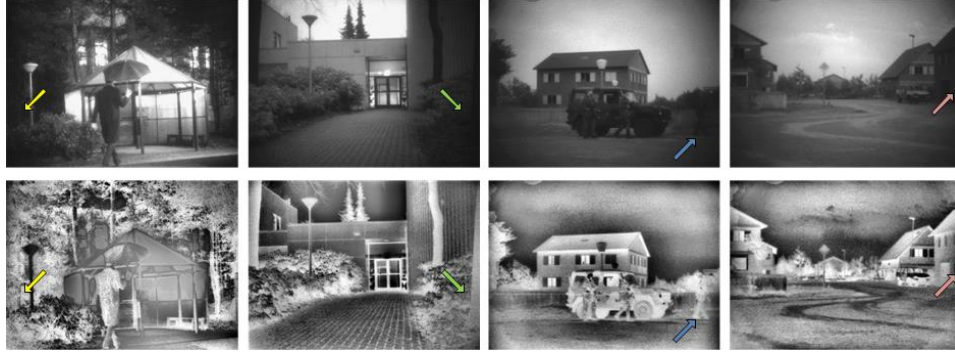


Fig. 5. DRA-based results of reflectance estimation from VIS images. From top to bottom: VIS images and the corresponding DRA-based reflectance layers. The regions pointed by the arrows are the dark regions.

2.3 Fusion based on Gamma correction

According to the DRA-based reflectance, a strategy based on Gamma correction is developed to IR and VIS image fusion, which could be formulated as

$$\mathbf{F} = \mathbf{L}_{ir} \diamond (\mathbf{R}_{vis}^C + \mathbf{t}), \quad (12)$$

where the operator \diamond denotes element-wise power. \mathbf{F} denotes the final fusion result. \mathbf{L}_{ir} denotes IR image and \mathbf{R}_{vis}^C denotes the DRA-based reflectance of VIS image. The intensity ranges of IR and VIS images have to be scaled into $[0,1]$. \mathbf{t} is a threshold matrix which has the same size with \mathbf{R}_{vis}^C and the every element is defined as $t_i = 1 - \text{otsu}(\mathbf{R}_{vis}^C)$, where $\text{otsu}(\mathbf{R}_{vis}^C)$ represents the threshold calculated from \mathbf{R}_{vis}^C by OTSU algorithm [34].

By using the Eq. (12), the details estimated from VIS images by the DRA-based Retinex model are fused with IR images, because the Gamma value of each pixel in IR images is evaluated according to \mathbf{R}_{vis}^C . Moreover, according to the threshold matrix \mathbf{t} , the Gamma values of the pixels with low intensity in VIS images are greater than 1, while those of the pixels with high intensity are less than 1. As a result, the contrast of the final fusion can be enhanced further. The process of our method is outlined in [Algorithm 1](#).

Algorithm 1. Image fusion using DRA reflectance estimation

Input: VIS image \mathbf{L}_{vis} , IR image \mathbf{L}_{ir} , the order N of the reflectance model, range parameter σ

1. $\hat{\mathbf{L}}_{vis} = \log(\mathbf{L}_{vis})$;
2. Deduce coarse-blurred illumination $\hat{\mathbf{T}}_{vis}^c$ from \mathbf{L}_{vis} ;
3. Estimate ψ_N^c , the optimal reflectance model, with quasi-Newton method via Eq. (9) & (10);
4. Obtain DRA-based reflectance \mathbf{R}_{vis}^c by Eq. (11) with ψ_N^c ;
5. Calculate the fused image \mathbf{F} via Eq. (12).

Output: \mathbf{F} , final fusion result

3. Experiment

3.1 Parameter setting

Algorithm 1 tells us we need 2 input parameters: the range parameter σ and the order N , determining the width of the response range of Gaussian fields criterion in the objective function (9) and the nonlinearity of the reflectance model, respectively. Here, after plenty of trials, we got the optimal setting is $\sigma = 0.6$ and $N = 4$. In addition, because we map the grey range of the input image into $[0,1]$, the optimal σ and N would be similar on different samples.

3.2 Dataset

The dataset tested in this work contains 30 pairs of IR and VIS images from TNO dataset¹ containing aligned pairs of IR and VIS image. To explore the fusion performance of the proposed way, we compare it with 3 fusion methods, namely, two-scale image fusion based on visual saliency (TSIFVS) [35], GTF [25] and image fusion based on gradient transfer optimization model (GTOM) [36]. All of these methods are programmed in Matlab and parameters set goes along with original papers. Each experiment is managed on a computer with the configuration of Windows 10 OS (64 bits) with Intel Core i7-9700K CPU @3.60GHz and 16GB RAM. The average runtime of our method is 2.42 seconds for a pair of IR and VIS images with 620×450 pixels.

3.3 Qualitative comparison

We pick up 30 pairs of IR and VIS images from TNO dataset as the testing set of qualitative comparison. Five typical pairs are selected for qualitative illustration from the testing set as **Fig. 6**.

The original IR and VIS images are located at the first two columns of **Fig. 6**. The fusion results of our method are shown in the 3rd row, while the rest of **Fig. 6** correspond to the results of the other approaches. It is apparent that all methods could to some extent fuse the information from the original IR and VIS images. However, from the qualitative comparison we can see that compared to the competitors, our method have an advantage on detail preservation. For instance, in our fusion images of **Fig. 6(a)** and **(b)**, the brightness of the dark regions in the VIS images has substantial growth. Hence, the hard-visible details in the original VIS images, such as shrubs and trees, are preserved and enhanced in our fusion results. From **Fig. 6(c)** we can see that the texture of the jeep has high saliency in the image fusion of our method. As for **Fig. 6(d)**, the targets, such as vehicle, signpost and street lamp, have good visibility in our fusion result. On the contrary, the detail of VIS images is not well preserved by TSIFVS, GTF and GTOM. Undoubtedly, the superiority of our approach on preserving

¹ http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.

details thanks to the DRA-based reflectance estimated from VIS images. Furthermore, exploiting the details information of the dark areas of VIS images via DRA-based reflectance estimation is very helpful to improve the visibility of fused images.

In the boxes of Fig. 6(e), the details in the IR image are more salient than those in the VIS image. From the fused images of the competitors, we can see that although the content of IR and VIS images are fused, details from the IR image become rather blurry. This is because the fused image suffers from the low saliency of the VIS image. It is obvious that our method could handle this problem. This demonstrates that the DRA-based reflectance estimated from VIS images could retain the salient information of the source images. It is also proved that the fusion model (12) could simultaneously preserve the textural detail information of IR and VIS images.

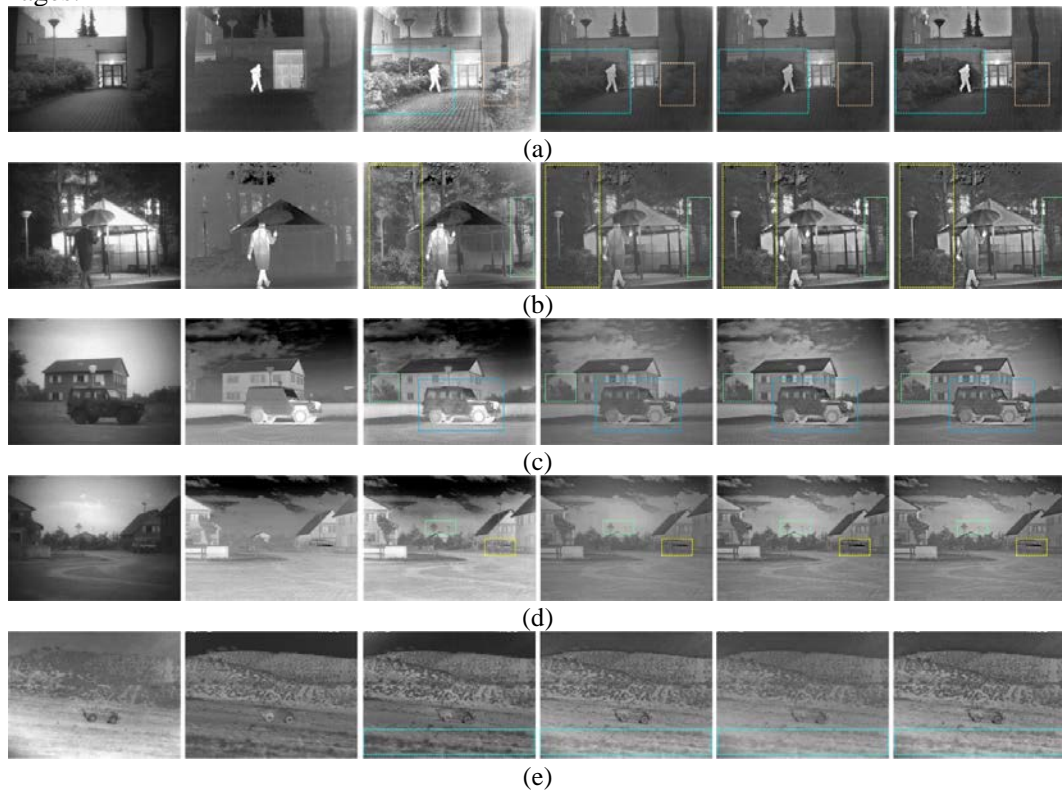


Fig. 6. Qualitative fusion results of 5 typical pairs of IR and VIS image from TNO dataset. The blue&yellow boxes emphasize the area of emerging marked differences between our result and the others'. From left: VIS images, IR images, our proposed method, TSIFVS, GTF and GTOM.

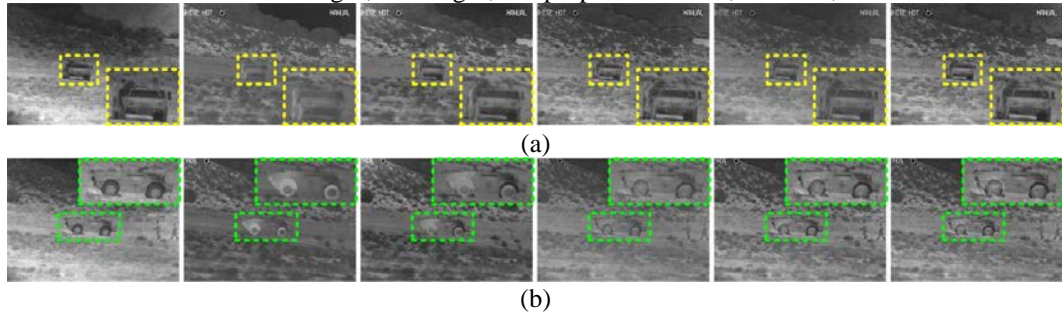


Fig. 7. Fusion results on 2 pairs of IR and VIS image pairs with high noise. We have zoomed some regions in the box in and put them at the corner of each subplot. From left: VIS images, IR images, our proposed method, TSIFVS, GTF and GTOM.

Fig. 7 shows the fusion outcome of 2 pairs of IR and VIS image pairs with high noise. We can see that noise affects us more slightly than other competitors. The reason is that DRA-estimated reflectance is able to limit noise via the regularity of the reflectance model (7).

3.4 Quantitative comparison

In this work, six metrics are introduced to quantitative evaluation. They are:

Entropy (EN) [37]: information amount in an image.

Correlation coefficient (CC) [38]: the degree of linear correlation between the source image and the fused image.

Standard deviation (SD) [39]: the contrast distribution of an image.

Spatial frequency (SF) [40]: the richness of the textural details of an image.

Visual information fidelity (VIF) [41]: the information fidelity of the fused image.

Structural similarity index measure (SSIM) [42]: the structural similarity between the source image and the fused image.

In terms of these 6 metrics, the larger the value, the better the performance would be.

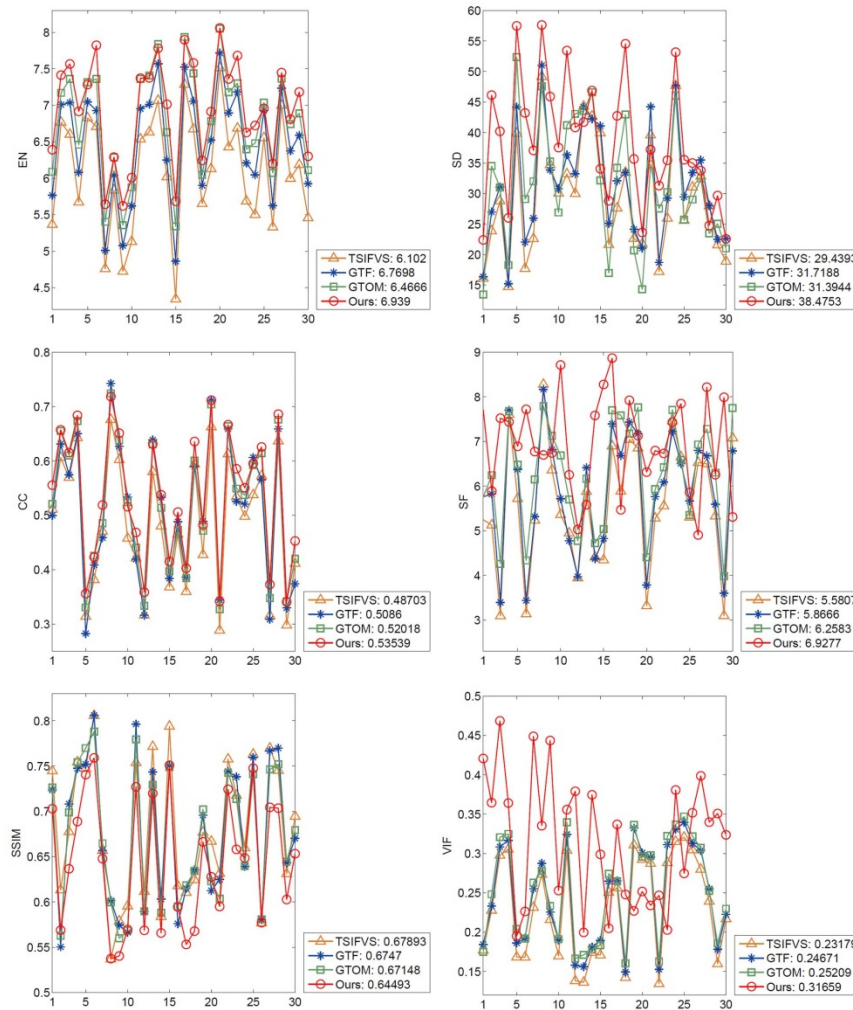


Fig. 8. Quantitative comparison of 6 evaluation metrics on TNO dataset. Three advanced methods, TSIFVS, GTF and GTOM, are used for comparison. The numbers in the legends are the average values of evaluation metrics.

The results of the six metrics are reported in **Fig. 8**. It is clear that our experiment has the best SD, EN, CC, VIF and SF for most pairs of image. Moreover, the mean values of these 5 metrics of our method respectively are the highest among all these approaches. The highest SD indicates that the contrast of our fused images is the best among the approaches. The highest EN demonstrates that the information of our fusion results are more abundant than those of the competitors. The greatest CC suggests that our fused images have strong correlation with the original images. The highest VIF demonstrates that our fused images have good visual quality and conforms well to the human visual system. The best SF means that there are rich details and textures in our fused images. However, SSIM of our method is relatively low because our method have good performance on detail preservation. Some barely visible details in the source images could be clearly maintained in our fused images. Moreover, our method could simultaneously exploit the textural detail information from IR and VIS images. Thus, some areas in our fused images might look neither like VIS images nor like IR images, resulting in low structural similarity. A representative example can be seen in **Fig. 6(a)**, in which the details of the shrubs have low intensity in the VIS image and are not visible in the IR image; however, in order to preserve the details in dark regions, the brightness of the shrubs in our fused image is increased greatly. Hence, the area of the shrubs in our fused image looks neither like VIS images nor like IR images. Similar situation also occurs in the trees of **Fig. 6(b)** and the jeep of **Fig. 6(c)**.

4. Conclusion

In this paper, a reflectance estimation way with dark region approximation (DRA) has been proposed to IR and VIS image fusion. The performance of Retinex model has been better taking advantages of DRA, especially that of detail preservation of dark regions of VIS images. DRA-based reflectance of VIS image with the Gaussian field criterion is evaluated by the quasi-Newton method. Then, it is fused with IR image by a strategy based on Gamma correction to generate the final fusion result. Qualitative and quantitative comparisons unveil the superiority of our method over other advanced approaches on preserving details and visual quality.

Admittedly, the proposed method has done well in our experiments evaluated from many respects, however, the high computation complexity of it is still a problem, increasing the difficulty in practical applications. Accordingly, we will focus on simplifying the structure of our method and improving the time cost of reflectance estimation in future.

Acknowledgements

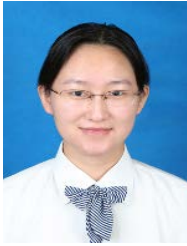
The authors gratefully acknowledge the financial supports from the National Natural Science Foundation of China (No. 61901157).

Reference

- [1] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, J. Wu and J. Jiang, "Infrared and visible image fusion via detail preserving adversarial learning," *Inf. Fusion*, vol. 54, pp. 85-98, Jul. 2019.
[Article \(CrossRef Link\)](#).
- [2] F. An, X. Zhou and D. Lin, "Multiscale self-coordination of bidimensional empirical mode decomposition in image fusion," *KSII Trans. Int. Inf. Syst.*, vol. 9, no. 4, pp. 1441-1456, Apr. 2015.
[Article \(CrossRef Link\)](#).

- [3] W. Jeong, B. Han, H. Yang and Y. Moon, "Real-time visible-infrared image fusion using multi-guided filter," *KSII Trans. Int. Inf. Syst.*, vol. 13, no. 6, pp. 3092-3107, Jun. 2019. [Article \(CrossRef Link\)](#).
- [4] C. Deng, Z. Wang, X. Li, H. Li and C. C. Cavalcante, "An improved remote sensing image fusion algorithm based on IHS transformation," *KSII Trans. Int. Inf. Syst.*, vol. 11, no. 3, pp. 1633-1649, Mar. 2017. [Article \(CrossRef Link\)](#).
- [5] Y. Liu, X. Chen, R. K. Ward and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882-1886, Dec. 2016. [Article \(CrossRef Link\)](#).
- [6] J. Wang, J. Peng, X. Feng, G. He and J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infrared Phys. Technol.*, vol. 67, pp. 477-489, Nov. 2014. [Article \(CrossRef Link\)](#).
- [7] D. P. Bavirisetti, G. Xiao and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proc. of Int. Conf. on Inf. Fusion*, pp. 701-709, July 10-13, 2017.
- [8] W. Kong, Y. Lei and H. Zhao, "Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization," *Infrared Phys. Technol.*, vol. 67, pp. 161-172, Nov. 2014. [Article \(CrossRef Link\)](#).
- [9] S. Li, X. Kang, J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864-2875, Jul. 2013. [Article \(CrossRef Link\)](#).
- [10] A. Toet, "Image fusion by a ratio of low-pass pyramid," *Pattern Recognit. Lett.*, vol. 9, no. 4, pp. 245-253, May. 1989. [Article \(CrossRef Link\)](#).
- [11] P. Hill, M. E. Al-Mualla and D. Bull, "Perceptual image fusion using wavelets," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1076-1088, Mar. 2017. [Article \(CrossRef Link\)](#).
- [12] M. Choi, R. Kim, M. Nam and H. Kim, "Fusion of multispectral and panchromatic satellite images using the curvelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 136-140, Apr. 2005. [Article \(CrossRef Link\)](#).
- [13] Z. Zhou, B. Wang, S. Li and M. Dong, "Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters," *Inf. Fusion*, vol. 30, pp. 15-26, Jul. 2016. [Article \(CrossRef Link\)](#).
- [14] F. Meng, M. Song, B. Guo, R. Shi and D. Shan, "Image fusion based on object region detection and non-subsampled contourlet transform," *Comput. Electr. Eng.*, vol. 62, pp. 375-383, Aug. 2017. [Article \(CrossRef Link\)](#).
- [15] W. Kong, L. Zhang and Y. Lei, "Novel fusion method for visible light and infrared images based on NSST-SF-PCNN," *Infrared Phys. Technol.*, vol. 65, pp. 103-112, Jul. 2014. [Article \(CrossRef Link\)](#).
- [16] T. Xiang, L. Yan and R. Gao, "A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain," *Infrared Phys. Technol.*, vol. 69, pp. 53-61, Mar. 2015. [Article \(CrossRef Link\)](#).
- [17] X. Zhang, Y. Ma, F. Fan, Y. Zhang and J. Huang, "Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition," *JOSA A.*, vol. 34, no. 8, pp. 1400-1410, Aug. 2017. [Article \(CrossRef Link\)](#).
- [18] J. Zhao, Y. Chen, H. Feng, Z. Xu and Q. Li, "Infrared image enhancement through saliency feature analysis based on multi-scale decomposition," *Infrared Phys. Technol.*, vol. 62, pp. 86-93, Jan. 2014. [Article \(CrossRef Link\)](#).
- [19] 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. [Article \(CrossRef Link\)](#).
- [20] W. Zhao, H. Lu and D. Wang, "Multisensor image fusion and enhancement in spectral total variation domain," *IEEE Trans. Multimedia*, vol. 20, no. 4, pp. 866-879, Apr. 2018. [Article \(CrossRef Link\)](#).
- [21] 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. [Article \(CrossRef Link\)](#).

- [22] Y. Liu, X. Chen, H. Peng, Z. Wang, "Multi-focus image fusion with a deep convolutional neural network," *Inf. Fusion*, vol. 36, pp. 191-207, Jul. 2017. [Article \(CrossRef Link\)](#).
- [23] H. Li, X.-J. Wu, J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proc. of International Conference on Pattern Recognition*, pp. 2705-2710, Nov. 2018. [Article \(CrossRef Link\)](#).
- [24] X. Mao, Q. Li, H. Xie, R.Y. Lau, Z. Wang, S.P. Smolley, "Least squares generative adversarial networks," in *Proc. of IEEE International Conference on Computer Vision*, pp. 2813-2821, Dec. 2017. [Article \(CrossRef Link\)](#).
- [25] J. Ma, C. Chen, C. Li and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100-109, Sep. 2016. [Article \(CrossRef Link\)](#).
- [26] J. Ma, W. Yu, P. Liang, C. Li and J. Jiang, "FusionGAN: a generative adversarial network for infrared and visible image fusion," *Inf. Fusion*, vol. 48, pp. 11-26, Aug. 2019. [Article \(CrossRef Link\)](#).
- [27] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108-128, Dec. 1977. [Article \(CrossRef Link\)](#).
- [28] J. Qu, Y. Li, Q. Du and H. Xia, "Hyperspectral and panchromatic image fusion via adaptive tensor and multi-scale retinex algorithm," *IEEE Access*, vol. 8, pp. 30522-30532, Feb. 2020. [Article \(CrossRef Link\)](#).
- [29] X. Fu, D. Zeng, Y. Huang, X. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. of IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2782-2790, Jun. 2016. [Article \(CrossRef Link\)](#).
- [30] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround Retinex," *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451-462, Mar. 1997. [Article \(CrossRef Link\)](#).
- [31] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965-976, Jul. 1997. [Article \(CrossRef Link\)](#).
- [32] M. H. Loke and R. D. Barker, "Rapid least-squares inversion of apparent resistivity pseudosections by a quasi-Newton method," *Geophysical Prospecting*, vol. 44, no. 1, pp. 131-152, Jan. 1996. [Article \(CrossRef Link\)](#).
- [33] X. Ren, W. Yang, W. Cheng and J. Liu, "LR3M: robust low-light enhancement via low-rank regularized Retinex model," *IEEE Trans. Image Process.*, vol. 29, pp. 5862-5876, Apr. 2020. [Article \(CrossRef Link\)](#).
- [34] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cy.*, vol. 9, no. 1, pp. 62-66, Jan. 1979. [Article \(CrossRef Link\)](#).
- [35] D. P. Bavirisetti and R. Dhuli, "Two-scale image fusion of visible and infrared images using saliency detection," *Infrared Phys. Technol.*, vol. 76, pp. 52-64, May. 2016. [Article \(CrossRef Link\)](#).
- [36] R. Yu, W. Chen and D. Zhou, "Infrared and visible image fusion based on gradient transfer optimization model," *IEEE Access*, vol. 8, pp. 50091-50106, Mar. 2020. [Article \(CrossRef Link\)](#).
- [37] J. W. Roberts, J. Van Aardt and F. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, pp. 023522, Jan. 2008. [Article \(CrossRef Link\)](#).
- [38] M. Deshmukh and U. Bhosale, "Image fusion and image quality assessment of fused images," *Int. J. Image Process.*, vol. 4, no. 5, pp. 484-508, Dec. 2010. [Article \(CrossRef Link\)](#).
- [39] Y. Rao, "In-fibre bragg grating sensors," *Meas. Sci. Technol.*, vol. 8, no. 4, pp. 355, 1997. [Article \(CrossRef Link\)](#).
- [40] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959-2965, Dec. 1995. [Article \(CrossRef Link\)](#).
- [41] Y. Han, Y. Cai, Y. Cao and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. fusion*, vol. 14, no. 2, pp. 127-135, Apr. 2013. [Article \(CrossRef Link\)](#).
- [42] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no 3, pp. 81-84, Mar. 2002. [Article \(CrossRef Link\)](#).



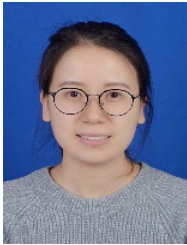
Yan Gu received the M.S. degrees in optical engineering from Nanjing University of Science & Technology, China, in 2009. She is currently a senior engineer with North Night Vision Technology Co., Ltd, Nanjing, China. Her research interest includes infrared imaging and image fusion. (e-mail: jsguyan@163.com)



Feng Yang received the M.S. degrees in optical engineering from Nanjing University of Science & Technology, China, in 2014. He is now a vice-senior engineer with North Night Vision Technology Co., Ltd, Nanjing, China. His research interest includes image processing and optoelectronic imaging system. (e-mail: yangfeng_nnvt@163.com)



Weijun Zhao received the M.S. degree from the School of Physics and Technology, Nanjing Normal University, Nanjing, China, in 2015. He is now an engineer with the North Night Vision Technology Co., Ltd. His current research interests include electro-optic detection and infrared imaging. (e-mail: 15195867099@163.com)



Yiliang Guo received the M.S. degree from the School of Electronics and Optical Technology, Nanjing University of Technology, Nanjing, China, in 2013. She is now an engineer with the North Night Vision Technology Co., Ltd. Her current research interests include low-light imaging and ultraviolet detection. (e-mail: gyl@nvt.com.cn)



Chaobo Min received the Ph.D. degree from the School of Electronics and Optical Technology, Nanjing University of Technology, Nanjing, China, in 2014. He is now a lecturer with the College of Internet of Things Engineering, HoHai University. His current research interests include image processing and computer vision. (e-mail: chaobomin@outlook.com)