Signal Processing 172 (2020) 107546

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

# Infrared and visible image fusion and denoising via $\ell_2 - \ell_p$ norm minimization

Li Liu [a,*], Luping Xu [a], Houzhang Fang [a]

*Xidian University, Xi'an, 710071, China*

ARTICLE INFO

*Article history:*
Received 13 October 2019
Revised 13 February 2020
Accepted 18 February 2020
Available online 20 February 2020

*Keywords:*
Image Fusion and Denoising
$\ell_p$ Sparsity
Inverse Problem
Half-quadratic Splitting

ABSTRACT

Most traditional infrared and visible image fusion methods often ignore noise in acquisition or transmission and their performance inevitably decreases in practical applications. To address this problem, a new and effective variational model is proposed for simultaneous image fusion and denoising. In an $\ell_2 - \ell_p$ norm minimization setting with $p = 0$ and $p = 1$ respectively, the hybrid $\ell_2$ norm fidelity term is built to preserve image intensity and details from both infrared and visible images. And the nonconvex $\ell_0$ norm and convex $\ell_1$ norm sparsity constraints are applied to reduce noise while preserving important image fine features. Furthermore, a computationally efficient numerical algorithm based on half-quadratic splitting iteration is used to solve the complex optimization problem. Experimental results demonstrate that the proposed method can achieve a superior performance compared with existing fusion methods in both subjective and objective assessments.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Image fusion aims at extracting and synthesizing complementary information from different kinds of sensors in order to produce a more accurate, more reliable composite image which can facilitate subsequent image processing. Medical image fusion helps physicians extract the features that may not be normally visible in images by different modalities, and also reduces the storage cost by reducing storage to a single fused image instead of multiple-source images [1]. Satellite image fusion combines remotely sensed images with different spectral and spatial resolutions to identify the textures and shapes of the various objects and map updating [2]. Due to the limited depth-of-focus of optical lenses, multifocus image fusion can create a single image where all objects are in focus [3]. Therefore, image fusion plays an increasingly important role in image analysis and computer vision, and investigations on effective image fusion algorithms are very necessary and highly meaningful.

Images from different sensors, such as infrared, visible, computed tomography, magnetic resonance imaging, multispectral images and panchromatic images, are commonly used in image fusion. Among them, infrared and visible image fusion technique is more widely applied in video surveillance, agricultural automation, biometric identification and electronics testing. Infrared images are obtained by sensing radiation in the infrared spectrum. The radiation sensed is either emitted by an object with a temperature greater than absolute zero, or is reflected by it [4]. However, limited by its imaging principle and working conditions, infrared images are often noisy, low contrast, blurred, and contain poor visual details, and this often leads to incorrect interpretation in practical applications [5]. On the contrary, visible light sensors work at the same spectrum of wavelength as the human eye, and can obtain more detail information from reflected light from the observed object at high temporal and spatial resolution. However, they have poor anti-interference and imaging capabilities in bad weather and at night. Thus, combining visible and infrared images can yield information-rich and accurate representations of scenes. The resulting fused image should contain rich details of the background while highlighting the targets. In image fusion literature, compared with the traditional fusion methods focusing on sharp source images, less attention has been paid to develop effective fusion algorithms for noisy ones. Those traditional approaches are often based on the assumption that source images are noise-free, assumption that, more often than not, does not hold true as a precondition for image fusion application. Often, in practice, the input images are degraded by noise from various sources at different stages of processing from acquisition to transmission or storage. Especially the quality of infrared images is seriously affected by different types of noise, mainly electronic noise and nonuniformity noise [6]. Therefore, investigations on the fusion of infrared and visible images are challenging and promising in practical applications.

As far as we know, only a little work has been done for noisy image fusion. A straightforward approach is to average multi-

* Corresponding author.
*E-mail address:* dr.liliu@hotmail.com (L. Liu).

https://doi.org/10.1016/j.sigpro.2020.107546
0165-1684/© 2020 Elsevier B.V. All rights reserved.

sensor images. Although this simple operation can reduce noise to some degree, it can not effectively preserve edge structure information of source images and will lead to several undesired side effects. Nowadays, in order to address the noise problem in image fusion, some robust wavelet-based algorithms are proposed [7]. Especially in [8], a noisy image fusion method based on non-Gaussian statistical modelling of wavelet coefficients was proposed. Generalised Gaussian and alpha-stable distributions of wavelet coefficients were used to model image wavelet coefficients. For joint image fusion and denoising, a total variation (TV) based approach in conjunction with principal component analysis was used to estimate the fused image [9]. TV models do not require knowledge of the probability of pixel intensities and can preserve the edge details of the fused image while suppressing noise. However, the resulting images may be affected by the staircase effect in TV model. In the variational framework, a weighted energy function with TV (WTV) regularization term were built for fusion and denoising of noisy source images [10]. The weighted fidelity term served for image fusion, and the TV regularization term served for image denoising. Only considering the fusion of image intensities in the weighted fidelity term, details of fused image may be lost in the process of image fusion. An adaptive sparse representation model was proposed for image fusion and denoising [11]. In their fusion model, the high-quality image patches were classified into several categories in advance. Then, based on the gradient information from the source images, they learned a set of compacted subdictionaries from numerous high-quality image patches and designed an algorithm to select adaptively one from the subdictionaries. This method is a patch-based dictionary learning one, and in order to learn the dictionaries it requires some extra computations comparing with simple gradient-based variational methods such as WTV.

Recently, various image fusion methods in gradient domain have been investigated and obtained satisfactory fusion performance [12–15]. The intrinsic principle of those methods is that human visual system is sensitive to the changes of local details in an image, which are actually image gradients. Importantly, image fusion in gradient domain can obtain the fine characteristics from source images. Therefore, the resulting images seem to be more authentic. A variational model was proposed to reconstruct a fused image from gradient domain. The weighted structure tensor was used to fuse detailed information of multispectral images [12]. It assumes that the great contrast corresponds with visual saliency. Thus, the maximum contrast from multispectral images is selected as that of the resulting fused image. Although the algorithm can effectively preserve edge details of the source images, the fusion performance may be limited for noisy images due to the fact that the maximum contrast may be related to noise. Other novel TV-based fusion methods have obtained a great success for image fusion. The main ideas in [16,17] were to retain intensity information from infrared images and preserve gradient information from visible images. In [18], the fusion performance was improved by not only keeping intensity information from both infrared and visible images, but also preserving gradient information from visible images. All aforementioned fusion methods have a limited performance for noisy image fusion.

In this paper, a variational fusion method is proposed in order to fuse noisy images. The key idea of such method is not only to keep both intensity and gradient information from infrared and visible images, but also to remove noise in fused images. In Fig. 1 (a) and (b), both image details in infrared and visible images, marked in green rectangles, should be fused into the resulting image, and simultaneously both intensity information from them, marked in blue rectangles, also should be preserved in the fused image. The resulting images of the proposed method are shown in Fig. 1 (c) and (d). The homogeneous regions are marked in red rectangles, and the standard deviations (SD) of the homogeneous regions in visible and infrared images are 7.20 and 7.50. After processed by the proposed fusion method, the SDs of the same homo-

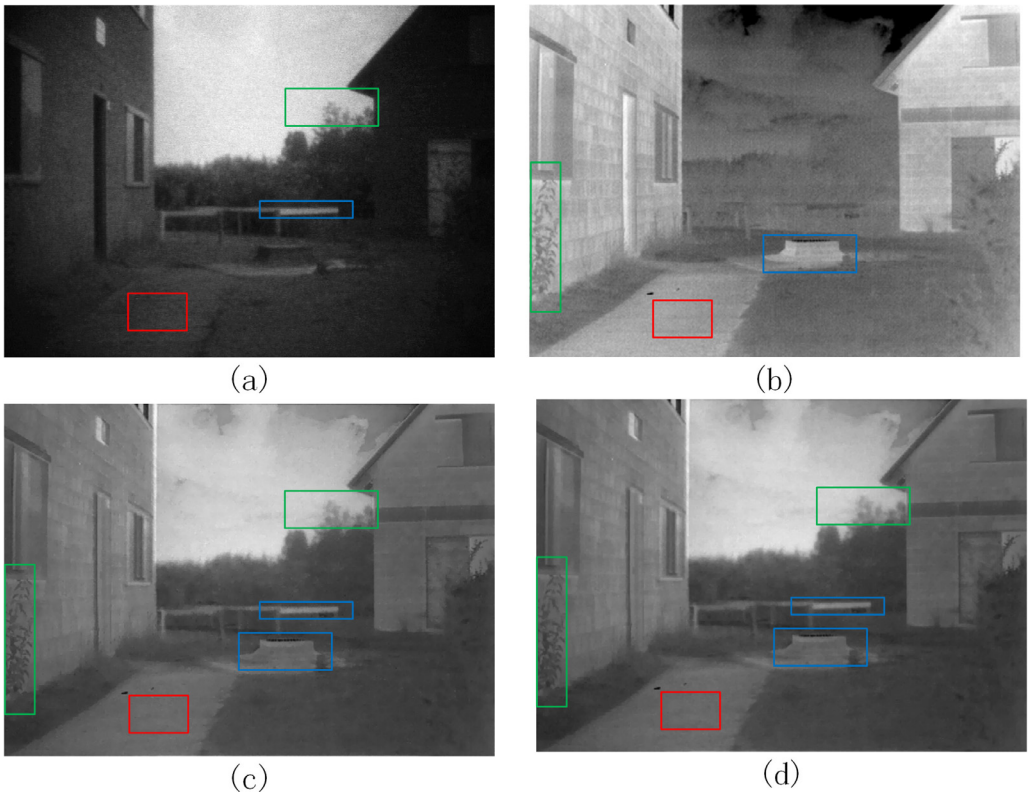



**Fig. 1.** (a) Visible image. (b) Infrared image. (c) Fused image of $\ell_2 - \ell_0$ minimization. (d) Fused image of $\ell_2 - \ell_1$ minimization. Green rectangle highlights image detail structures of infrared, visible and fused images, and blue and red rectangles highlight image intensity information and image homogeneous region, respectively.

geneous in Fig. 1 (c) and (d) are 4.18 and 4.26. The proposed fusion model is formulated as an $\ell_2 - \ell_p$ minimization problem setting with $p = 0$ and $p = 1$ respectively. The novel hybrid fidelity term simultaneously preserves intensity and texture information from infrared and visible images while the nonconvex $\ell_0$ norm and convex $\ell_1$ norm regularization term plays a role of denoising the fused image. To tackle the $\ell_2 - \ell_p$ norm minimization problem, a computationally efficient numerical algorithm based on half-quadratic splitting is introduced due to its fast convergence. The resulting optimization problem is split into two sub-problems that can be easily solved. Experimental results of public data sets are presented to illustrate the effectiveness of the proposed fusion method.

The main ideas and contributions of the proposed approach are summarized as follows:

(i) A new infrared and visible image fusion algorithm based on $\ell_2 - \ell_p$ minimization model is proposed. The hybrid fidelity $\ell_2$ norm term in intensity and gradient domain serves for image fusion and can achieve a good balance between pixel intensities from infrared and visible images while preserving detailed information. And the $\ell_p$-norm regularization term plays a significant role of image denoising.
(ii) The computationally efficient numerical algorithm based on half-quadratic regularization is introduced to optimize the $\ell_p$-norm minimization problem setting with $p = 0$ and $p = 1$ respectively. The resulting optimization problem is split into two sub-problems that can be easily solved.
(iii) Both qualitative and quantitative comparisons with state-of-the-art approaches on public data sets are provided.

The remainder of this paper is organized as follow: In Section 2, we discuss the fusion model based on the $\ell_2 - \ell_p$ norm minimization and solve the optimization problem using a computational efficient numerical algorithm based on half-quadratic regularization. In Section 3, we present and analyze the experimental results. Finally, we summarize this paper in Section 4.

## 2. Proposed image fusion and denoising method

In this section, the key components of the proposed fusion model are analyzed thoroughly. Then, the half-quadratic splitting method is introduced to optimize the complex minimization problem.

### 2.1. Image fusion and denoising based on $\ell_2 - \ell_p$ norm minimization

For the majority of infrared and visible CCD sensors, there are two common kinds of noise, viz., the dark current noise and photo generated sampling noise. In general, the mixed noise can be modeled as additive white Gaussian noise [19]. Thus, the degradation model can be expressed as

$$s = f + n \tag{1}$$

where $s$ is degraded source image, $f$ denotes sharp image, and $n$ is system noise.

What we concern here is to integrate noisy infrared and visible images into a more reliable and friendly visual image by highlighting targets and preserving edges details of source images. Infrared radiosity depends on the image object's temperature, the object's thermal emissivity as well as the reflected thermal radiance from other objects or thermal sources [4]. The difference in wavelength between the visible and infrared spectra causes dramatic changes in transmission and reflection. The appearance of targets in an fused image includes two primary parts, intensity information and detail information. To highlight the targets, both salient characteristics of infrared and visible images in intensity domain should be preserved. Thus, the data fidelity in intensity domain can be expressed as

$$R_1(f) = \eta_0 \|r - f\|_2^2 + \eta_1 \|v - f\|_2^2 \tag{2}$$

where $\eta_0$ and $\eta_1$ are positive weights, and $r$ and $v$ denote the infrared and visible images, respectively.

In image fusion, saliency is usually defined as edge detail information. The fusion process is aimed at enhancing and synthesizing this type of information. On one hand, details in visible image captured in day time are more abundant and more reliable than ones in the infrared image. On the other hand, the salient characteristics of the infrared image in weak light condition are more distinct. Edge details should be attached to optimize the variational fusion model to improve the human vision perception. Therefore, another data fidelity in gradient domain is formulated as

$$R_2(f) = \eta_2 \|\mathbf{D}r - \mathbf{D}f\|_2^2 + \eta_3 \|\mathbf{D}v - \mathbf{D}f\|_2^2 \tag{3}$$

where $\eta_2$ and $\eta_3$ are positive weights, and $\mathbf{D}$ denotes the gradient operator.

Both data fidelity terms provide the similarity between the desired fused image and source images. In order to obtain a sharp fused image, the regularization term should be imposed on the latent image, aiming at penalizing the undesirable components on images. Thus, the regularization term in the proposed fusion model is expressed as

$$R_3(f) = \lambda \|\mathbf{D}f\|_p \tag{4}$$

$\lambda$ is the regularization parameter, which controls the tradeoff between the data fidelity and regularization term. The key is to construct appropriate regularization terms to constrain the solution so as to remove random noise.

In our work, simultaneous image fusion and denoising task is to estimate a sharp fused image from source images in the presence of random noise. Based on the above theory, the fusion model can be formulated as the combination of data fidelity term and regularization term as follows:

$$\begin{aligned} \mathbf{min}\ & \eta_0 \|r - f\|_2^2 + \eta_1 \|v - f\|_2^2 + \eta_2 \|\mathbf{D}r - \mathbf{D}f\|_2^2 \\ & + \eta_3 \|\mathbf{D}v - \mathbf{D}f\|_2^2 + \lambda \|\mathbf{D}f\|_p \end{aligned} \tag{5}$$

When $p$ equals to 1, it is the well-known anisotropic TV regularization. It has been widely applied in various applications, due to its desirable properties, such as convexity and the significant ability to preserve edge structures. Although TV tries to model statistics of sharp images, a more powerful and discriminative model is needed to separate sharp image from noisy source ones. Since $\ell_0$ gradient minimization can control the global number of non-zero gradients between neighbours in an image, recently it is widely applied in image denoising, artifact removal, edge detection and detail enhancement [20]. Xu demonstrated that $\ell_0$ norm of image gradient is better than total variation and weighted $\ell_2$ norm of gradient [21]. Nguyen presented a descent approach based on region fusion to make $\ell_0$ gradient minimization converge faster [22]. Liu used $\ell_0$ norm prior to tackle the optics temperature-dependent nonuniformity issue in thermal images [23].Due to the powerful capability of $\ell_0$ gradient minimization in detail preservation and noise removal, an $\ell_0$ norm regularization term is chosen in the proposed fusion model.

### 2.2. Optimization with half-quadratic splitting

Our algorithm is derived from the well-known variable-splitting and penalty techniques in optimization [23–25]; Specifically, at each pixel an auxiliary variable $w$ is introduced to transfer $\mathbf{D}f$ out of the nondifferentiable term $\|\cdot\|_p$, and the difference between $w$

and $\mathbf{D}f$ is penalized, yielding the following approximation model to (5)

$$\min \eta_0 \parallel r-f\parallel_2^2+\eta_1 \parallel v-f\parallel_2^2+\eta_2 \parallel \mathbf{D}r-\mathbf{D}f\parallel_2^2 +\eta_3 \parallel \mathbf{D}v-\mathbf{D}f\parallel_2^2+\lambda \parallel w\parallel_p+\beta \parallel w-\mathbf{D}f\parallel_2^2 \quad (6)$$

with a sufficiently large penalty parameter $\beta$. It is well known that the solution of (6) converges to that of (5) as $\beta \to \infty$. We obtain the solution of (6) by alternatingly solving the following two subproblems.

**Subproblem 1: computing *f*** The $f$ estimation subproblem corresponds to minimizing

$$\min \eta_0 \parallel r-f\parallel_2^2+\eta_1 \parallel v-f\parallel_2^2+\eta_2 \parallel \mathbf{D}r-\mathbf{D}f\parallel_2^2 +\eta_3 \parallel \mathbf{D}v-\mathbf{D}f\parallel_2^2+\beta \parallel w-\mathbf{D}f\parallel_2^2 \quad (7)$$

by omitting the terms not involving $w$ in (6). The function is quadratic and thus has a global minimum. Alternatively, we diagonalize derivative operators after Fast Fourier Transform (FFT) for speedup. This yields solution

$$f=\mathcal{F}^{-1}\left(\frac{(\eta_0\mathcal{F}(r)+\eta_1\mathcal{F}(v)+\eta_2\mathcal{F}(\mathbf{D}^T\mathbf{D})\mathcal{F}(r)+\eta_3\mathcal{F}(\mathbf{D}^T\mathbf{D})\mathcal{F}(v)+\beta\mathcal{F}(\mathbf{D}^Tw)}{\eta_0+\eta_1+(\eta_2+\eta_3+\beta)\mathcal{F}(\mathbf{D}^T\mathbf{D})}\right) \quad (8)$$

where $\mathcal{F}$ denotes the two-dimensional (2D) Fourier transform, and $\mathcal{F}^{-1}$ denotes the 2D inverse Fourier transform. The plus, multiplication and division are all component-wise operators. Compared to minimizing (7) directly in image space, which involves very-large-matrix inversion, computation in the Fourier domain is much faster due to the simple component-wise division.

**Subproblem 2: computing $w$ When $p=1$, the minimization function for $w$ is**

$$\min \lambda \parallel w\parallel_1+\beta \parallel w-\mathbf{D}f\parallel_2^2 \quad (9)$$

The unique solution of (9) is given by the following soft shrinkage formula:

$$w=\max\left(|\mathbf{D}f|-\frac{\lambda}{2\beta},0\right)sgn(\mathbf{D}f) \quad (10)$$

When $p=0$, the minimization function for $w$ is

$$\min \lambda \parallel w\parallel_0+\beta \parallel w-\mathbf{D}f\parallel_2^2 \quad (11)$$

Note that (11) is a pixel-wise minimization problem, thus, $w$ is obtained based on [21],

$$w=\begin{cases}\mathbf{D}f, & |\mathbf{D}f|^2>\frac{\lambda}{\beta}\\ 0, & otherwise\end{cases} \quad (12)$$

Our alternating minimization algorithm is sketched in Algorithm 1. Parameter $\beta$ is automatically adjusted in iterations starting from a small value $\beta_0$, and is multiplied by $\kappa$ each time. This scheme is effective to speed up convergence.

**Algorithm 1** Image fusion and denoising via $\ell_2-\ell_p$ norm minimization.

**Input:** infrared image $r$, visible image $v$, parameters $\eta_0$, $\eta_1$, $\eta_2$, $\eta_3$, $\beta_0$, $\beta_{max}$ and rate $\kappa$
**Initialization:** $\beta \leftarrow \beta_0$
**repeat**
with $w$, solve for $f$ in Eq.(8) by applying FFT;
with $f$, if $p=1$, solve for $w$ in Eq.(11) by applying the soft shrinkage operator;
if $p=0$, solve for $w$ with Eq.(12);
$\beta \leftarrow \beta^*\kappa$;
**until** $\beta \geq \beta_{max}$
**Output:** fused image $f$.

## 3. Experiment results and analysis

In this section, we test the performance of the proposed infrared and visible image fusion approach and compared it with several state-of-the-art fusion methods, viz., gradient transfer and TV minimization (GTF) [16], weighted least square optimization (WLS) [26], deep convolutional neural network (DCNN) [27], hybrid multi-scale decomposition transform fusion with guided filter (HMSD-GD) [28], adaptive sparse representation (ASR) [11], and weighted variational method based on TV regularization (WTV) [10]. To evaluate the performance of the proposed method thoroughly, experimental images from publicly available TNO image fusion dataset[1] and Bristol Eden project multisensor data set[2] are utilized. The above-mentioned algorithms for comparison are implemented based on publicly available codes and the corresponding parameters are set as their defaults. All the comparative experiments are executed on a desk computer with 3.2GHz Intel Core i7-8700 CPU, 16GB memory and Matlab code.

### 3.1. Subjective evaluation

The fused images of ASR, DCNN, and the proposed method are shown in Figs. 2–4. We magnify the details of each fused image so that it is much easier for observers to compare the performance of the different fusion methods. In Fig. 2 (a), the cloudy sky of $\ell_2-\ell_0$, $\ell_2-\ell_1$, is much brighter than that of the other two fusion methods, which means the proposed fusion method can effectively balance both infrared and visible intensity information. After zooming in on the details of fused images, the rich edge structures of the proposed method show that $\ell_2-\ell_0$ and $\ell_2-\ell_1$ can successfully transfer both the important detail features from source images into the fused images. More importantly, the proposed fusion method can effectively remove random noise mainly from the visible image when noise level of visible image is higher than one of infrared image. Although ASR has the capability of removing simulated noise in source images, it can not effectively tackle noise issue in real-world source images. In Fig. 2 (c), the noise level in the infrared image is greater than that in the visible image. Both sky areas of $\ell_2-\ell_0$ and $\ell_2-\ell_1$ are brighter than ones of the other two methods. Unfortunately, the fused image of DCNN simultaneously has excessively bright and dark structures, which means it does not balance intensity information from source images well. It can be observed further that $\ell_2-\ell_1$ has the best performance of noise removal, while successfully preserving both image detail information from infrared and visible images.

The same conclusions are drawn from the magnified image patches in Figs. 3–4. It is observed from the shapes of the clouds in Fig. 3 (b) and (d) that the proposed fusion method can not only effectively preserve both intensity and detail information from infrared and visible images, but also significantly reduce noise. In Fig. 3 (b) and (d), the performance of ASR and DCNN is affected by noise, and thus they have a limited capability of transferring detail information from source images into the fused one. It is demonstrated that from the homogeneous areas in Fig. 3 (b) and

[1] https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029.
[2] http://www.cis.rit.edu/pelz/scanpaths/data/bristol-eden.htm.

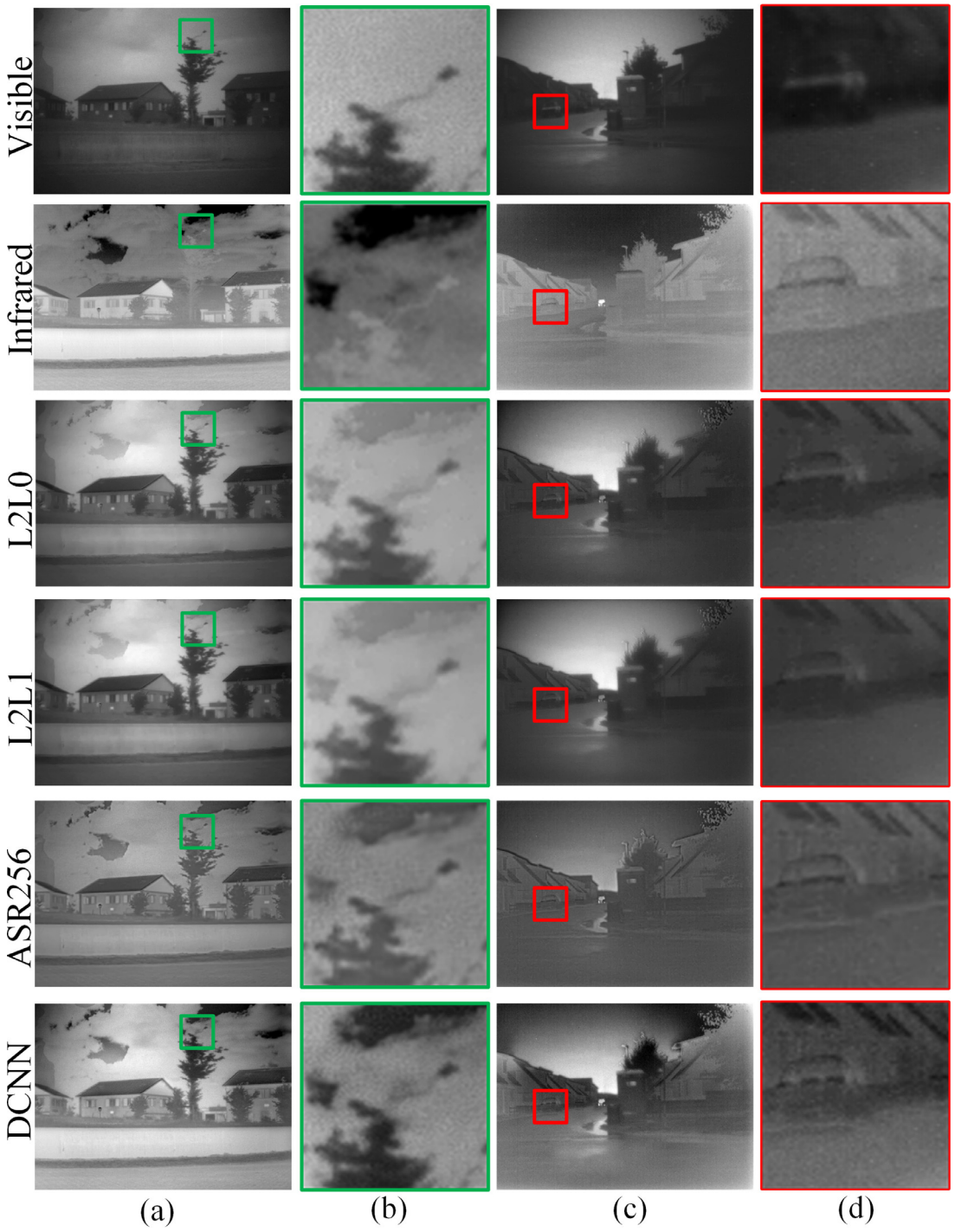



**Fig. 2.** The resulting fused images of different algorithms. (a) and (c) are the source images and fused images. (b) and (d) are the magnified image patches cropped from (a) and (c).

(d) that the proposed fusion method has excellent ability of removing noise. Another finding in Fig. 4 (c) is that both ASR and DCNN induce unfriendly visual artifacts in the edges of the building, which means they are not capable of balancing the intensity information from source images, and weak characteristics of visible image are effectively preserved in Fig. 4 (d) as the proposed method is used. Therefore, the qualitative comparisons from the detail structures demonstrate that our fusion method outperforms the state-of-the-art fusion ones.

To verify the effectiveness of the proposed method further, 41 image frames from Bristol Eden project multisensor data set and their corresponding fused images are used. Two fused image frames among them are shown in Fig. 5. It is observed from the magnified detail patches that the infrared image contains more conspicuous random noise than the visible image. The resulting fused images from $\ell_2 - \ell_0$ and $\ell_2 - \ell_1$ show that they have excellent capability of preserving edge structures and reducing noises. The WTV fusion method effectively removes noise, but it obviously oversmooths edge structures. The ASR128, ASR256, GTF, WLS, HMSD-GF and DCNN methods can transfer salient features into the resulting fused images, but they all have limited capability of tackling noise issue.

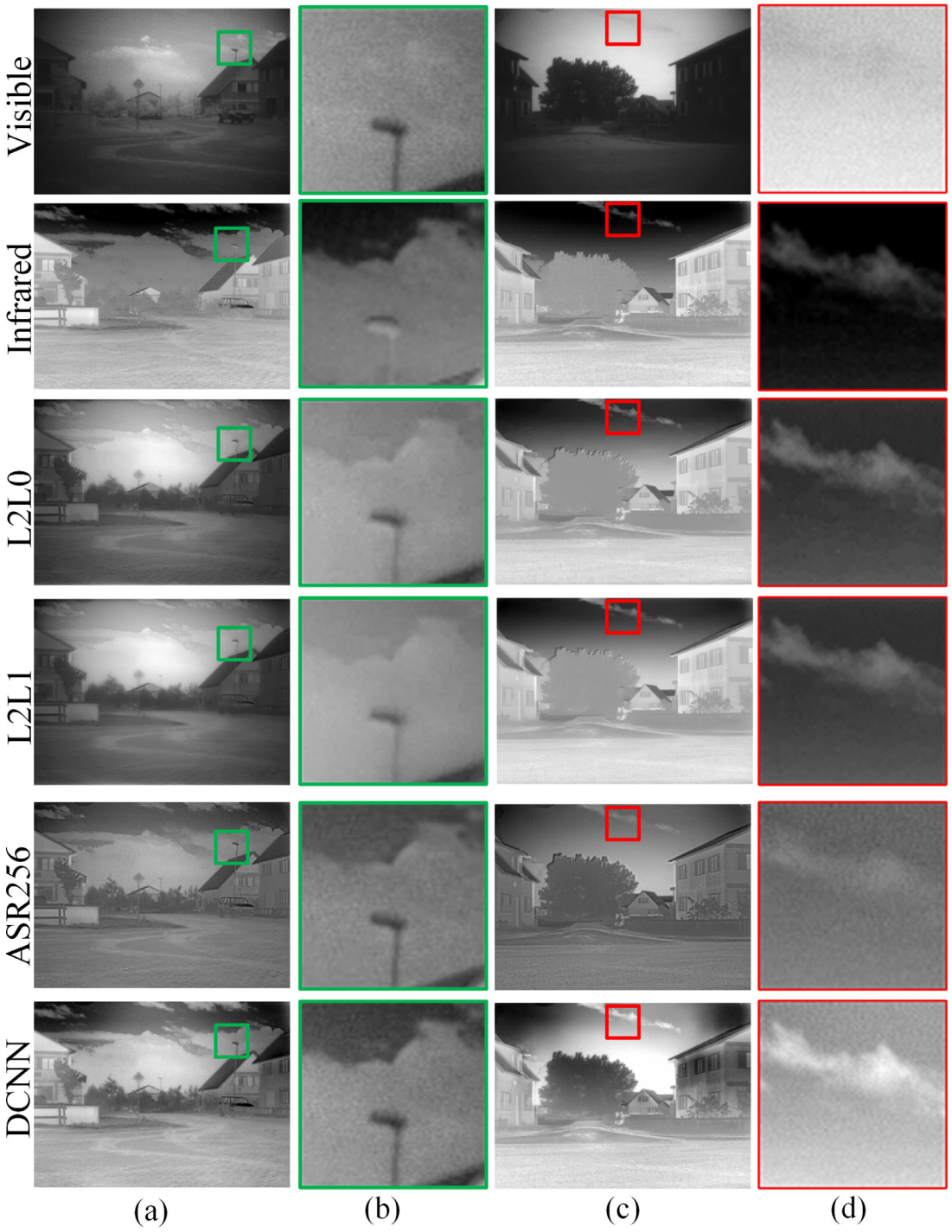



**Fig. 3.** The resulting fused images of different algorithms. (a) and (c) are the source images and fused images. (b) and (d) are the magnified image patches cropped from (a) and (c).

In summary, the above comparison algorithms can better deal with fusion problem and have achieved better performance in some cases. However, for degraded image fusion, the proposed method has better visual performance than other methods. The most advantage of the proposed method is that both intensity and edge information are preserved without introducing any additional artifacts, and at the same time system noises from infrared and visible images are significantly removed. The sharpened fused images without noise will be beneficial for automatic target detection and recognition. This is the significant difference between our method and the other existing fusion methods.

### *3.2. Objective evaluation*

Only with visual comparison, can the performance of all fusion methods not be evaluated thoroughly. Thus, it is necessary to evaluate fusion results in some objective quality indexes. This paper selects the following five objective quality indicators as the evaluation criteria:

1. **Mutual information (MI)** [16]: A quality index that measures the amount of information that is transferred from source images to the fused image. A large MI metric means that consider-

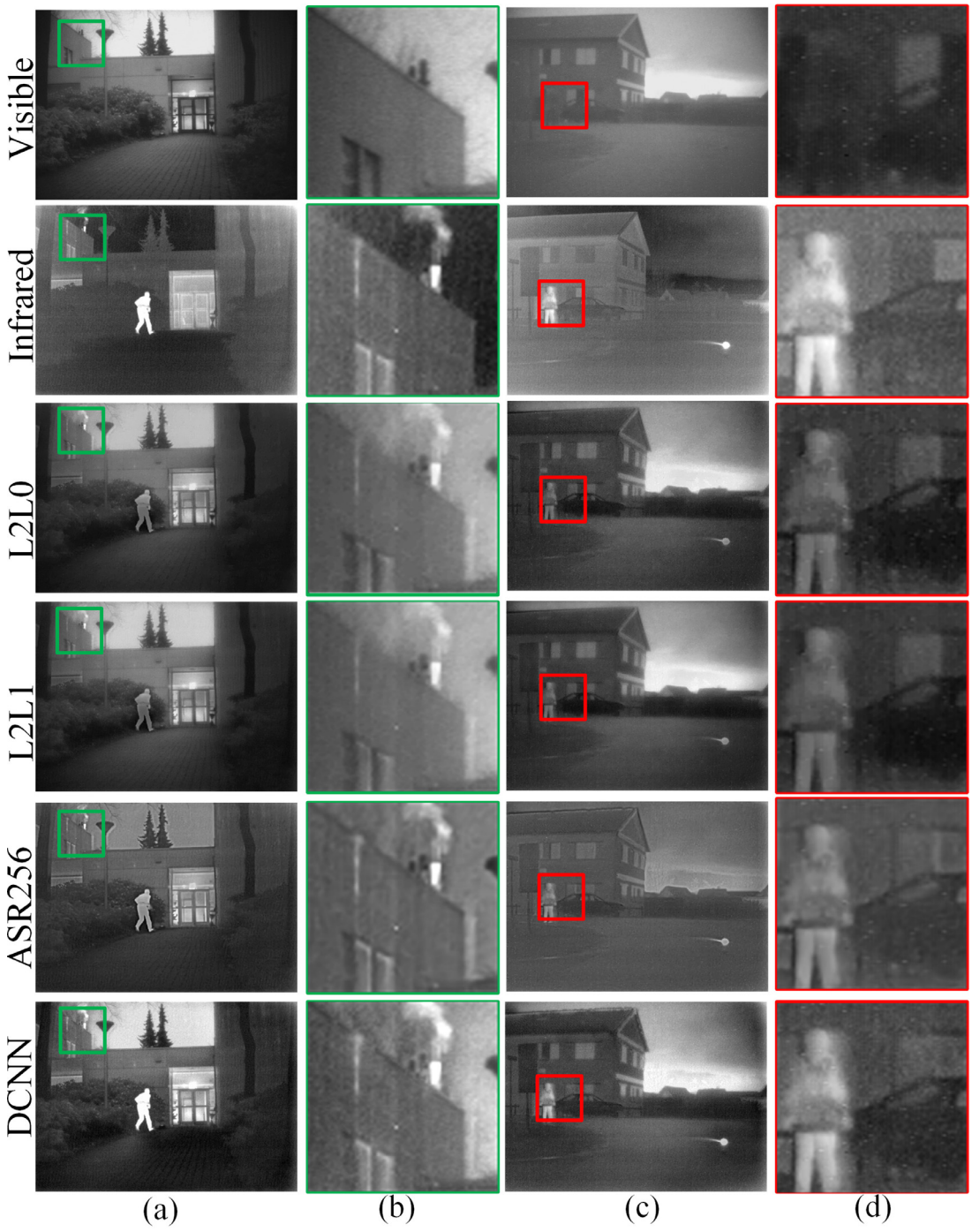



**Fig. 4.** The resulting fused images of different algorithms. (a) and (c) are the source images and fused images. (b) and (d) are the magnified image patches cropped from (a) and (c).

able information is transferred from source images to the fused one, which indicates a good fusion performance.

2. **Normalized mutual information (NMI)** [29]: A modified MI index to deal with instability of raw MI. The larger NMI is, the better a fusion performance is.
3. **Entropy (EN)** [30]: EN measures the amount of information contained in a fused image on the basis of information theory. The larger the EN, the more information is contained in the fused image and the better the performance of the fusion method.
4. **Structural similarity index (SSIM)** [31]: SSIM mainly consists of three components: loss of correlation, luminance and contrast distortion. SSIM can measure the structural similarity between source image and fused image. A large SSIM means a better fusion performance.
5. **Phase congruency (PC)** [29]: Phase congruency measurement provides an absolute measurement of image features with a value between zero and one. This metric counts how the features from the source images are fused into the resulting fused image.

For all the above five indicators, the larger the value, the more the fused image contains the source image information. However, since the source images inevitably include noise, these indicators

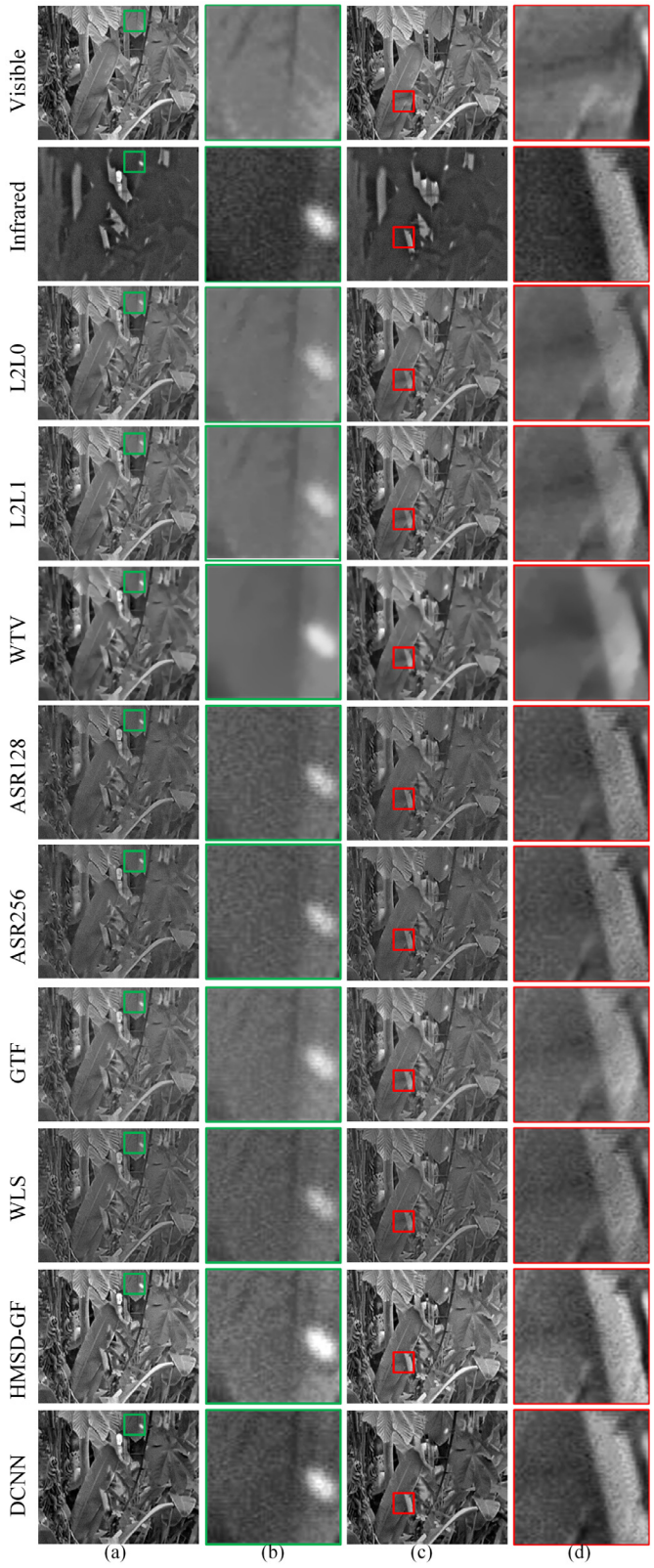



**Fig. 5.** The resulting fused images of different algorithms. From top to bottom: visible image, infrared image, results of $\ell_2 - \ell_0$, $\ell_2 - \ell_1$, WTV, ASR128, ASR256, GTF, WLS, HMSD-GD and DCNN, respectively. Clearly, only our method has the capability of simultaneous image fusion and denoising.

will be artificially high and become unreal. Therefore, the objective indicators need to be evaluated together with subjective perception. The detailed quantitative evaluation of Figs. 2–5 is shown in Table 1 and Fig. 6. The values in boldface means the highest results.

In general, it is clearly observed that MIs, NMIs, PCs are the highest in all experiments. Since the proposed fusion method transfers not only the intensity information from the source images, but also the detail information of them, our fused images have the highest degree of sharpness. In addition, $\ell_0$ and $\ell_1$ regularization terms make edge information more comprehensive because noises from source images are effectively removed. Therefore, it is demonstrated from the aspect of fused edge information and noise removal that the proposed fusion method greatly improves the overall performance. Unfortunately, the EN and SSIM values of the proposed method do not reach the highest in each group. This is because reducing noise in fused images will make the values of EN and SSIM become smaller.

### 3.3. Parameter selection

In our algorithm, several parameters must be properly selected in order to obtain a good fusion result, including $\eta_0$, $\eta_1$, $\eta_2$, $\eta_3$, $\lambda$, $\beta$ and $\kappa$. $\eta_0$ and $\eta_1$ play important roles in balancing intensity information from infrared and visible image, and both are set empirically with the range [0, 1]. $\eta_2$ and $\eta_3$ have similar function to balance fine detail structures from both source images, and are set with the range [0, 20]. The regularization parameter $\lambda$ can play a significant role of smoothing image noise, which is set with the range [0, 1]. The penalty parameter $\beta$ has a fixed initial value 0.1, and $\kappa$ that is set to $\sqrt{2}$ is a good balance between efficiency and performance. For example, $\eta_0$, $\eta_1$, $\eta_2$ and $\eta_3$ in the fusion experiment of Bristol Eden project multisensor data set are 0.5, 1, 5 and 5, respectively, and the values of $\lambda$ in $\ell_2 - \ell_0$ and $\ell_2 - \ell_1$ are 0.0025 and 0.15, respectively.

Finally, we compare the average running time of all fusion methods for source images with different sizes. In our method, half-quadratic splitting is introduced to solve the resulting minimization problem in order to effectively reduce the large computation load. The complex optimization model is decomposed into two independent subproblems which can be calculated or updated in parallel. Moreover, to estimate the final fused image, the efficient FFT is used to solve the corresponding quadratic minimization problem in (8). Consequently, the proposed fusion method is relatively fast. The average running time is show in Table 2.

## 4. Conclusion

In this paper, we have proposed a new fusion model for simultaneous image fusion and denoising. Different from previous fusion models, both intensity and gradient information from source images have been taken into full consideration. Meanwhile, the nonconvex $\ell_0$ norm and convex $\ell_1$ norm sparsity constraints have been designed for denoising. The half-quadratic splitting method has been applied to solve the complex optimization model due to its fast convergence and superior numerical stability. Several public data sets were tested in our experiments. Both the qualitative and quantitative assessments confirm that the proposed model can output better fusion results than other state-of-the-art techniques of WTV, ASR, GTF, WLS, HMSD-GF and DCNN. Since the fusion performance of the proposed method is closely dependent on proper parameter selection, strong emphasis should be put on the automatic parameter selection that is more useful for practical applications in the future research.

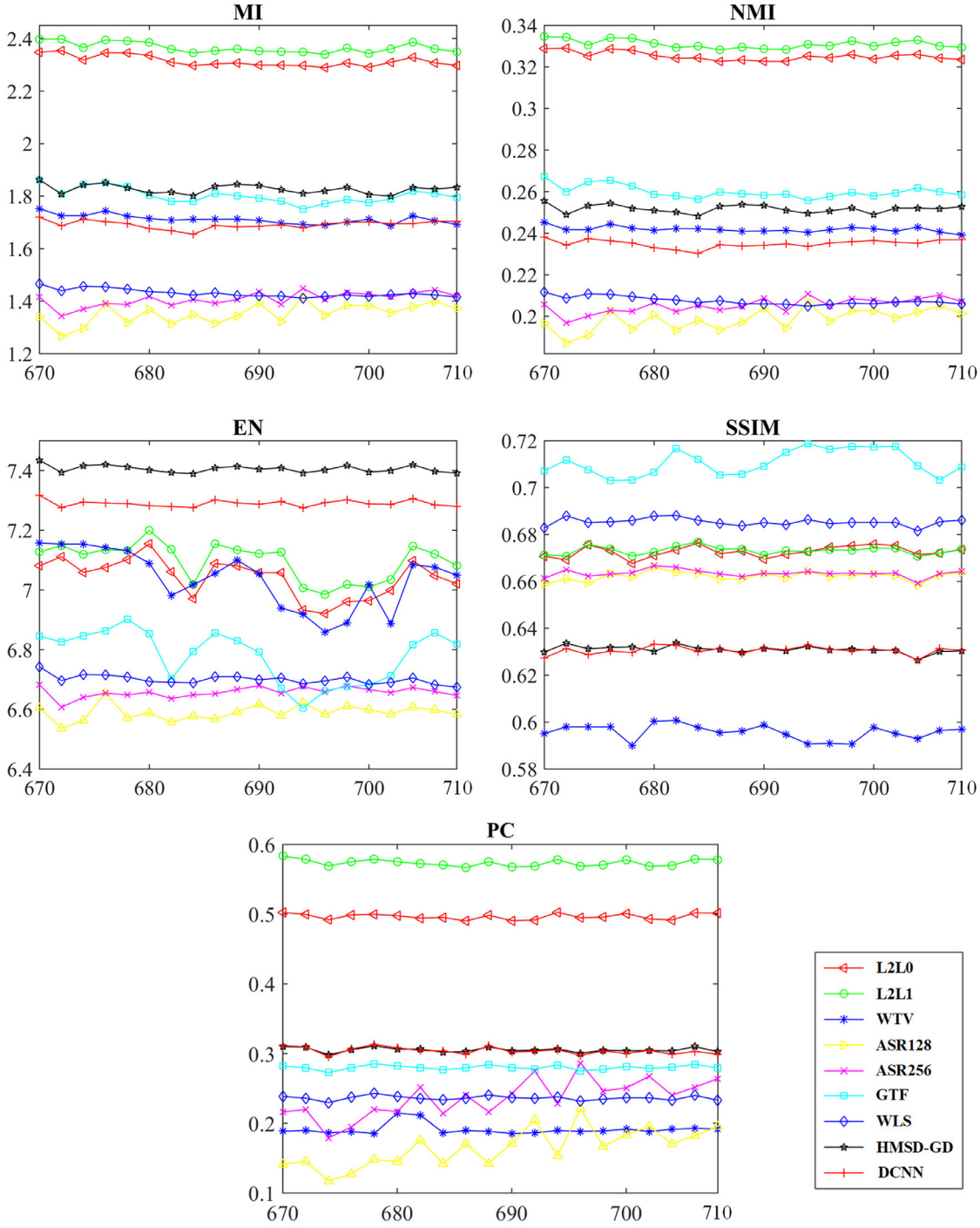



**Fig. 6.** Quantitative comparisons of the five metrics, i.e., MI, NMI, EN, SSIM and PC on the Bristol Eden project multisensor data set (670-710 image frames). The seven state-of-the-art methods such as WTV, ASR128, ASR256, GTF, WLS, HMSD-GD and DCNN, are used for comparison. For all five metrics, larger values mean better performance.

**Table 1**
MI, NMI, EN, SSIM and PC of the original images and the images fused by different fusion algorithms.

| Image | Metric | $\ell_2-\ell_0$ | $\ell_2-\ell_1$ | WTV | ASR128 | ASR256 | GTF | WLS | HMSD-GD | DCNN |
|---|---|---|---|---|---|---|---|---|---|---|
| Fig. 2(a) | MI | **2.91** | 2.86 | 2.30 | 2.17 | 2.18 | 2.70 | 2.08 | 2.12 | 1.91 |
| | NMI | **0.40** | **0.40** | 0.31 | 0.31 | 0.31 | 0.37 | 0.30 | 0.29 | 0.26 |
| | EN | 7.20 | 7.24 | 7.61 | 6.83 | 6.83 | 7.41 | 6.84 | 7.47 | **7.63** |
| | SSIM | 0.71 | **0.72** | 0.67 | 0.71 | 0.71 | **0.72** | **0.72** | 0.68 | 0.68 |
| | PC | 0.40 | **0.48** | 0.18 | 0.33 | 0.34 | 0.34 | 0.33 | 0.34 | 0.38 |
| Fig. 2(c) | MI | 3.70 | **3.73** | 2.30 | 1.76 | 1.77 | 2.39 | 1.66 | 2.34 | 1.64 |
| | NMI | 0.46 | **0.55** | 0.34 | 0.28 | 0.28 | 0.36 | 0.26 | 0.33 | 0.23 |
| | EN | 6.74 | 6.79 | 6.64 | 5.81 | 5.81 | 6.51 | 5.82 | 7.26 | **7.32** |
| | SSIM | **0.67** | **0.67** | 0.63 | **0.67** | **0.67** | **0.67** | **0.67** | 0.65 | 0.64 |
| | PC | 0.31 | **0.32** | 0.08 | 0.23 | 0.24 | 0.26 | 0.25 | 0.27 | 0.28 |
| Fig. 3(a) | MI | 2.77 | **2.94** | 2.25 | 1.98 | 1.98 | 1.69 | 1.36 | 1.57 | 1.41 |
| | NMI | 0.38 | **0.40** | 0.31 | 0.29 | 0.29 | 0.23 | 0.20 | 0.22 | 0.19 |
| | EN | 7.40 | 7.42 | **7.44** | 6.69 | 6.69 | 7.18 | 6.77 | 7.29 | 7.41 |
| | SSIM | 0.66 | 0.67 | 0.61 | **0.71** | 0.70 | 0.56 | 0.53 | 0.50 | 0.52 |
| | PC | 0.23 | **0.24** | 0.12 | 0.18 | 0.18 | 0.15 | 0.14 | 0.15 | 0.14 |
| Fig. 3(c) | MI | 3.33 | **3.41** | 2.57 | 2.33 | 2.33 | 2.85 | 2.49 | 1.85 | 2.52 |
| | NMI | 0.46 | **0.47** | 0.37 | 0.34 | 0.34 | 0.41 | 0.36 | 0.25 | 0.34 |
| | EN | 7.21 | 7.26 | 6.95 | 6.56 | 6.56 | 6.88 | 6.61 | **7.65** | 7.62 |
| | SSIM | 0.57 | 0.58 | 0.53 | 0.57 | 0.57 | **0.59** | 0.56 | 0.56 | 0.56 |
| | PC | 0.54 | **0.58** | 0.11 | 0.49 | 0.50 | 0.42 | 0.40 | 0.48 | 0.53 |
| Fig. 4(a) | MI | 2.98 | **3.01** | 2.34 | 2.03 | 2.33 | 2.42 | 1.92 | 2.00 | 2.02 |
| | NMI | **0.42** | **0.42** | 0.33 | 0.30 | 0.34 | 0.35 | 0.28 | 0.28 | 0.29 |
| | EN | 7.00 | 7.21 | 7.06 | 6.73 | 6.73 | 6.91 | 6.77 | **7.26** | 7.19 |
| | SSIM | **0.72** | **0.72** | 0.60 | 0.70 | 0.70 | **0.72** | 0.71 | 0.67 | 0.67 |
| | PC | 0.30 | **0.31** | 0.12 | 0.27 | 0.28 | 0.27 | 0.26 | 0.29 | 0.29 |
| Fig. 4(c) | MI | 3.43 | **3.49** | 2.19 | 1.81 | 1.81 | 2.56 | 1.66 | 1.87 | 2.48 |
| | NMI | 0.49 | **0.50** | 0.31 | 0.28 | 0.28 | 0.38 | 0.26 | 0.27 | 0.36 |
| | EN | 7.12 | 7.13 | 7.18 | 6.04 | 6.05 | 6.89 | 6.12 | 7.01 | **7.18** |
| | SSIM | 0.73 | **0.74** | 0.63 | 0.72 | 0.72 | 0.73 | **0.74** | 0.69 | 0.69 |
| | PC | **0.31** | **0.31** | 0.08 | 0.28 | 0.28 | 0.24 | 0.22 | 0.24 | 0.25 |

**Table 2**
Average running time (s) of different fusion algorithms.

| Image size | $\ell_2-\ell_0$ | $\ell_2-\ell_1$ | WTV | ASR128 | ASR256 | GTF | WLS | HMSD-GD | DCNN |
|---|---|---|---|---|---|---|---|---|---|
| 450 × 620 | 0.77 | 0.77 | 3.55 | 126.92 | 130.30 | 1.32 | 1.91 | 1.16 | 46.77 |
| 576 × 768 | 1.21 | 1.19 | 6.02 | 277.15 | 311.14 | 2.29 | 3.18 | 1.99 | 71.33 |

## Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

**Li Liu:** Conceptualization, Methodology, Software. **Luping Xu:** Supervision. **Houzhang Fang:** Resources.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant no. 61801355 and the Aeronautical Science Foundation under Grant no. ASFC-20170181002.

## References

[1] Z. Xu, Medical image fusion using multi-level local extrema, Inform. Fusion 19 (2014) 38–48.
[2] J. Choi, K. Yu, Y. Kim, A new adaptive component-substitution-based satellite image fusion by using partial replacement, IEEE Trans. Geosci. Remote Sens. 49 (1) (2011) 295–309.
[3] M.S. Farid, A. Mahmood, S.A. Al-Maadeed, Multi-focus image fusion using content adaptive blurring, Inform. Fusion 45 (2019) 96–112.
[4] N.J.W. Morris, S. Avidan, W. Matusik, H. Pfister, Statistics of infrared images, Proc. IEEE Conf. Comput. Vis. Pattern Recognit (2007) 1–7.
[5] C. Gao, L. Wang, Y. Xiao, Q. Zhao, D. Meng, Infrared small-dim target detection based on markov random field guided noise modeling, Pattern Recognit. 76 (2018) 463–475.
[6] D.S. Kim, Noise power spectrum measurements in digital imaging with gain nonuniformity correction, IEEE Trans. Image Process. 25 (8) (2016) 3712–3722.
[7] P. Scheunders, S.D. Backer, Wavelet denoising of multicomponent images using gaussian scale mixture models and a noise-free image as priors, IEEE Trans. Image Process. 16 (7) (2007) 1865–1872.
[8] A. Loza, D. Bull, N. Canagarajah, A. Achim, Non-gaussian model-based fusion of noisy images in the wavelet domain, Comput. Vision Image Understand. 114 (1) (2010) 54–65.
[9] M. Kumar, S. Dass, A total variation-based algorithm for pixel-level image fusion, IEEE Trans. Image Process 18 (9) (2009) 2137–2143.
[10] W.W. Wang, P.L. Shui, X.C. Feng, Variational models for fusion and denoising of multifocus images, IEEE Signal Process. Lett. 15 (2008) 65–68.
[11] Y. Liu, Z. Wang, Simultaneous image fusion and denoising with adaptive sparse representation, IET Image Process. 9 (5) (2015) 347–357.
[12] G. Piella, Image fusion for enhanced visualization: a variational approach, Int. J. Comput. Vis. 83 (2009) 1–11.
[13] D.A. Socolinsky, L.B. Wolff, Multispectral image visualization through first-order fusion, IEEE Trans. Image Process. 11 (8) (2002) 923–931.
[14] Z. Zhou, S. Li, B. Wang, Multi-scale weighted gradient-based fusion for multi–focus images, Inform. Fusion 20 (2014) 60–72.
[15] F. Li, T. Zeng, Variational image fusion with first and second-order gradient information, J. Comput. Math. 34 (2) (2016) 200–222.
[16] Y. Ma, J. Chen, C. Chen, F. Fan, J. Ma, Infrared and visible image fusion using total variation model, Neurocomputing 202 (2016) 12–19.
[17] J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, Inform. Fusion 31 (2016) 100–109.
[18] H. Guo, Y. Ma, X. Mei, J. Ma, Infrared and visible image fusion based on total variation and augmented lagrangian, J. Opt. Soc. Am. A 34 (11) (2017) 1961–1968.
[19] V.S. Petrovic, C.S. Xydeas, Sensor noise effects on signal-level image fusion performance, Inform. Fusion 4 (3) (2003) 167–183.
[20] J. Pan, Z. Hu, Z. Su, M. Yang, L0-regularized intensity and gradient prior for deblurring text images and beyond, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2) (2017) 342–355.
[21] L. Xu, C. Lu, Y. Xu, J. Jia, Image smoothing via l0 gradient minimization, ACM T. Graphic 30 (6) (2011) 174.
[22] R.M.H. Nguyen, M.S. Brown, Fast and effective l0 gradient minimization by region fusion, Proc. IEEE int'l. Conf. Comput. Vis. (2015) 208–216.

[23] L. Liu, T. Zhang, Optics temperature-dependent nonuniformity correction via l0-regularized prior for airborne infrared imaging systems, IEEE Photon. J. 8 (5) (2016) 3900810.
[24] Y. Wang, J. Yang, W. Yin, Y. Zhang, A new alternating minimization algorithm for total variation image reconstruction, SIAM J. Imag. Sci. 1 (3) (2008) 248–272.
[25] W. Zuo, D. Meng, L. Zhang, X. Feng, D. Zhang, A generalized iterated shrinkage algorithm for non-convex sparse coding, Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (2013) 217–224.
[26] J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, Infrared Phys. Technol. 82 (2017) 8–17.
[27] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inform. Fusion 36 (2017) 191–207.
[28] Z. Zhou, M. Dong, X. Xie, Z. Gao, Fusion of infrared and visible images for night-vision context enhancement, Appl. Opt. 55 (23) (2016) 6480–6490.
[29] Z. Liu, E. Blasch, Z. Xue, J. Zhao, R. Laganiere, W. Wu, Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: A comparative study, IEEE Trans. Pattern Anal. Mach. Intell. 34 (1) (2012) 94–109.
[30] X. Bai, Y. Zhang, F. Zhou, B. Xue, Quadtree-based multi-focus image fusion using a weighted focus-measure, Inform. Fusion 22 (2015) 105–118.
[31] Y. Ma, J. Ma, C. Li, Infrared and visible image fusion methods and applications: a survey, Inform. Fusion 45 (2019) 153–178.