Neurocomputing 202 (2016) 12–19

Contents lists available at ScienceDirect

Neurocomputing

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

Infrared and visible image fusion using total variation model

Yong Ma [a], Jun Chen [b], Chen Chen [c], Fan Fan [a,*], Jiayi Ma [a]

[a] Electronic Information School, Wuhan University, Wuhan 430072, China
[b] School of Automation, China University of Geosciences, Wuhan 430074, China
[c] Department of Electrical and Computer Engineering, UIUC, Urbana, IL 61801, USA

ARTICLE INFO

Article history:
Received 25 November 2015
Received in revised form
2 March 2016
Accepted 8 March 2016
Available online 28 March 2016

Keywords:
Image fusion
Infrared
Total variation

ABSTRACT

Image fusion is a process of combining complementary information from multiple images of the same scene into an image, so that the resultant image contains a more accurate description of the scene than any of the individual source images. In this paper, we propose a novel fusion strategy for infrared (IR) and visible images based on total variation (TV) minimization. By constraining the fused image to have similar pixel intensities with the IR image and similar gradients with the visible image, it tends to simultaneously keep the thermal radiation and appearance information in the source images. We evaluate our method on a publicly available database with comparisons to other seven fusion methods. Our results have a major difference that the fused images look like sharpened IR images with detailed appearance information. The quantitative results demonstrate that our method also can achieve comparable metric values with other state-of-the-art methods.

1. Introduction

Multi-sensor data often provides complementary information about the region surveyed, and image fusion which aims to create new images from such data offering more complex and detailed scene representation has then emerged as a promising research strategy for scene analysis in the areas of remote sensing [1,2], pattern recognition [3,4], medical imaging [5,6] and modern military [7,8]. In this paper, we concentrate on the fusion of multi-sensor data such as thermal infrared (IR) and visible images, which can lead to better performance for human visual perception, object detection, as well as target recognition [9,10].

The goal of image fusion is to identify the most important information in the source images and to transfer, without distortion or loss, this information into a fused image. For the problem of IR and visible image fusion, visible sensors capture reflected lights with abundant appearance information, and it is better for establishing a discriminative model. In contrast, IR sensors capture principally thermal radiations emitted by objects, which are not affected by illumination variation or disguise and hence, it can overcome some of the obstacles to discover the target and work day and night. However, IR image has lower spatial resolution than visible image, where appearance features such as textures in a visible image often get lost in the corresponding IR image since textures seldom influence the heat emitted by an object. Therefore, it is beneficial for automatic target detection and unambiguous localization to fuse the thermal radiation and texture information into a single image.

The process of image fusion can be performed at different levels depending on the information representation and applications. A common categorization is to distinguish between pixel, feature and symbol levels [11]. Fusion at pixel-level represents fusion at the lowest level referring to combining the raw source images into a single image [12]. Fusion at higher level such as feature-level or symbol level combines information in the form of feature descriptors and probabilistic variables [13]. However, pixel-level fusion is still a popular strategy for most image fusion applications, as it has the main advantage that the original measured quantities are directly involved in the fusion process. Besides, pixel-level fusion algorithms are computationally efficient and easy to implement [14]. In this paper, we only focus on the pixel-level image fusion problem.

A prerequisite for pixel-level fusion is that multi-sensor images have to be correctly registered on a pixel-by-pixel basis. Image registration techniques have been discussed extensively in the literature [15–19]. Throughout this paper, it will be assumed that all source images have been registered. To address the pixel-level fusion problem, many methods have been proposed in the past decades [20–31]. The simplest strategy is to take the average of the

* Corresponding author.
E-mail addresses: mayong@whu.edu.cn (Y. Ma), junchen@cug.edu.cn (J. Chen), cchen156@illinois.edu (C. Chen), fanfan@whu.edu.cn (F. Fan), jyma2010@gmail.com (J. Ma).

http://dx.doi.org/10.1016/j.neucom.2016.03.009
0925-2312/

source images pixel by pixel. However, such direct method will lead to several undesired side effects including reduced contrast. In order to solve this problem, multi-scale transform (MST) based methods have been proposed which are able to provide much better performance, since they are consistent with human visual system and real-world objects usually consist of structures at different scales [11,32]. Examples of these methods include Laplacian pyramid [20], discrete wavelet transform [33], and nonsubsampled contourlet transform [34]. The MST-based methods have achieved great success in many situations; however, they use the same representations for different source images and try to preserve the same salient features such as edges and lines in the source images. For the problem of IR and visible image fusion, the thermal radiation information in an IR image is characterized by the pixel intensities, and the target typically has larger intensities compared to the background and hence can be easily detected; while the texture information in a visible image is mainly characterized by the gradients, and the gradients with large magnitude (e.g. edges) provide detail information for the scene. Therefore, it is not appropriate to use the same representations for these two types of images during the fusion process. Instead, to preserve the important information as more as possible, the fused image is desirable to keep the main intensity distribution in the IR image and the gradient variation in the visible image. To this end, in this paper we proposed a novel algorithm based on total variation (TV) minimization for IR and visible image fusion.

More precisely, we formulate the fusion as a TV minimization problem, where the data fidelity term constrains that the fused image should have the similar pixel intensities with the given IR image, and the regularization term ensures that the gradient distribution in the visible image can be transferred into the fused image. The ℓ^1 norm is employed to encourage the sparseness of the gradients [35], and the optimization problem can then be solved via existing TV minimization techniques [36].

The contribution of this paper is two-fold. On the one hand, we propose a new IR and visible image fusion algorithm based on total variation minimization. It tends to simultaneously preserve the thermal radiation information as well as the detailed appearance information in the source images, and to the best of our knowledge, such fusion strategy has not yet been studied. On the other hand, we provide both qualitative and quantitative comparisons with several state-of-the-art approaches on a publicly available dataset. Compared to previous methods, our method can generate fusion results looking like sharpened IR images with detailed scene representation and hence, it is able to improve the reliability of automatic target detection and recognition systems.

The rest of the paper is organized as follows. Section 2 presents the formulation of the proposed fusion algorithm for IR and visible image fusion. In Section 3, we demonstrate our method for fusion on a publicly available dataset with comparisons to several state-of-the-art approaches, followed by some concluding remarks in Section 4.

2. Method

In this section, we present the layout of our IR and visible image fusion method. To this end, we first introduce the TV minimization problem, and then present our fusion method based on TV minimization.

2.1. Total variation minimization

The TV model was proposed by Beck et al. [37] as a regularizing criterion to solve the image denoising problem due to its property of effectively preserving edge information, which has evolved from an image denoising method [38] into a more general technique for inverse problems [39], including deblurring [40], blind deconvolution [41], inpainting [42], super-resolution [43], texture analysis [44] and smoothing [45]. For an image of size $m \times n$, we denote by $\mathbf{u} \in \mathbb{R}^{mn\times 1}$ the column-vector form of its pixel intensities, which has gray-scale values ranging from 0 to 255. The TV model of $\mathbf{u}$ is defined as follows:

$$J(\mathbf{u}) = \sum_{i=1}^{mn} |\nabla_i \mathbf{u}| = \sum_{i=1}^{mn} \sqrt{(\nabla_i^h \mathbf{u})^2 + (\nabla_i^v \mathbf{u})^2}, \tag{1}$$

where $|u| := \sqrt{u_1^2 + u_2^2}$ for every $u = (u_1, u_2) \in \mathbb{R}^2$, $\nabla_i = (\nabla_i^h, \nabla_i^v)$ denotes the image gradient ∇at pixel i with ∇^{h} and ∇^{v} being linear operators corresponding to the horizontal and vertical first-order differences, respectively. More specifically, $\nabla_i^h \mathbf{u} = \mathbf{u}_i - \mathbf{u}_{r(i)}$ and $\nabla_i^v \mathbf{u} = \mathbf{u}_i - \mathbf{u}_{b(i)}$, where $r(i)$ and $b(i)$ represent the nearest neighbor to the right and below the pixel i. Besides, if pixel i is located in the last row or column, $r(i)$ and $b(i)$ are both set to be i.

With the TV model in Eq. (1), many image processing tasks can be formulated as the following inverse problem with regularization constraint:

$$\mathbf{u}^* := \arg\min_{\mathbf{u}} \frac{1}{2} \|\mathbf{u} - \mathbf{f}\|^2 + \lambda \mathbf{J}(\mathbf{u}), \tag{2}$$

where the first term $\|\mathbf{u} - \mathbf{f}\|^2$ is the data fidelity item, which stands for the fidelity between the observed image $\mathbf{f}$ and the original unknown image $\mathbf{u}$. The total variation $J(\mathbf{u})$ in the second term plays a role of regularization. λ is the regularization parameter controlling the tradeoff between the data fidelity and regularization item. Eq. (2) is the classic TV minimization problem which can be efficiently solved by using existing algorithms [36]. And it has been investigated for solving many image processing tasks such as denoising, deblurring, and reconstruction.

2.2. The proposed fusion method

Given a pair of aligned IR and visible images, our goal is to generate a fused image that simultaneously preserves the thermal radiation information and the detailed appearance information in the two images, respectively. Here the IR, visible and fused images are all supposed to be gray scale images of size $m \times n$, and their column-vector forms are respectively denoted by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{x} \in \mathbb{R}^{mn\times 1}$ with gray-scale values ranging from 0 to 255. Typically, the thermal radiation is characterized by the pixel intensities, and then the fused image is expected to have the similar pixel intensities with the IR image, for example, the following empirical error should be as small as possible

$$\mathcal{E}_1(\mathbf{x}) = \tfrac{1}{2} \|\mathbf{x} - \mathbf{u}\|_2^2. \tag{3}$$

To fuse the detailed appearance information, a straightforward scenario is to require the fused image also to have the similar pixel intensities with the visible image. However, the intensity of a pixel in the same physical location may be significantly different for IR and visible images, as they are manifestations of two different phenomena and hence, it is not appropriate to generate $\mathbf{x}$ by simultaneously minimizing $\|\mathbf{x} - \mathbf{u}\|_2^2$ and $\|\mathbf{x} - \mathbf{v}\|_2^2$. Note that the detailed appearance information about the scene is essentially characterized by the gradients in the image. Therefore, we propose to constrain the fused image to have similar pixel gradients rather than similar pixel intensities with the visible image. As visible images are often piece-wise smooth, their gradients tend to be sparse and gradients with large magnitude correspond to the edges. It is widely known that the ℓ^1 norm encourages sparsity and ℓ^2 norm does not, thus we consider minimizing the gradient differences with ℓ^1 norm to encourage sparseness of the gradients:

$$\mathcal{E}_2(\mathbf{x}) = \|\nabla \mathbf{x} - \nabla \mathbf{v}\|_1. \tag{4}$$

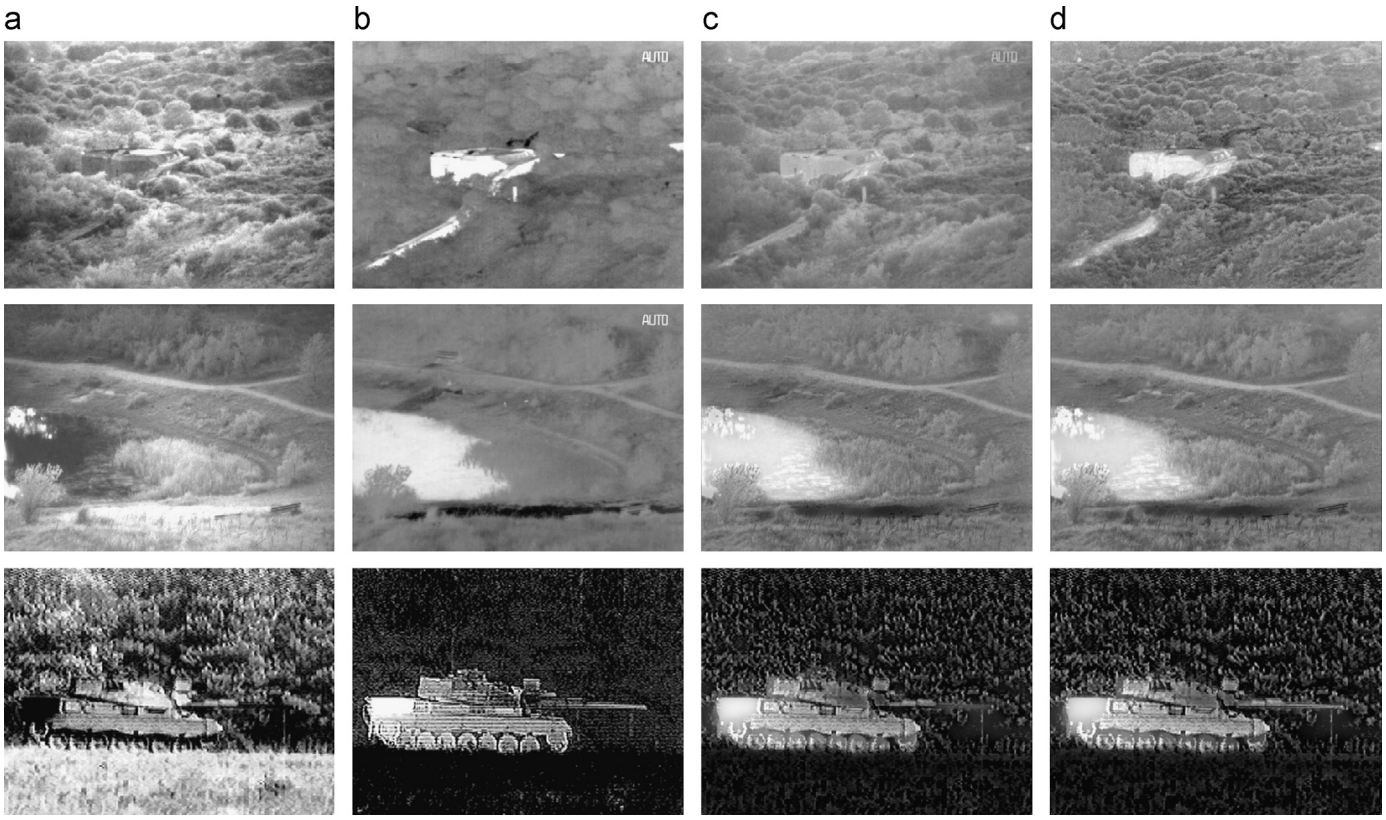


Fig. 1. Quantitative illustration of our fusion strategy. (a) visible images; (b) IR images; (c) results of a popular method Wavelet [23]; (d) results of the proposed method.

Combining Eqs. (3) and (4), we formulate the fusion problem as minimizing the following objective function:

$$\mathcal{E}(\mathbf{x}) = \mathcal{E}_1(\mathbf{x}) + \lambda \mathcal{E}_2(\mathbf{x}) = \tfrac{1}{2}\|\mathbf{x}-\mathbf{u}\|_2^2 + \lambda \|\nabla\mathbf{x}-\nabla\mathbf{v}\|_1, \tag{5}$$

where the first term constrains the fused image $\mathbf{x}$ to have the similar pixel intensities with the IR image $\mathbf{u}$, the second term requires that the fused image $\mathbf{x}$ and the visible image $\mathbf{v}$ have the similar gradients, more specifically, the similar edges in corresponding positions, and λ is a positive parameter controlling the trade-off between the two terms. The objective function (5) to some extent aims to transfer the gradients/edges in the visible image onto the corresponding positions in the IR image. Thus the fused image should still look like an IR image, but with more appearance details, i.e., an IR image with more complex and detailed scene representation. It plays a role of IR image sharpness or enhancement, which is the major difference between our method and other typical fusion methods [46].

Clearly, the objective function (5) is convex and thus has a global optimal solution. The first term is smooth while the second term is non-smooth. Let $\mathbf{y} = \mathbf{x} - \mathbf{v}$, Eq. (5) can be rewritten as:

$$\mathbf{y}^* = \arg\min_{\mathbf{y}} \left\{ \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{mn}} \frac{\mathbf{1}}{\mathbf{2}} \|\mathbf{y}_i - (\mathbf{u}_i - \mathbf{v}_i)\|_2^2 + \lambda \mathbf{J}(\mathbf{y}) \right\}. \tag{6}$$

The new objective function (6) is a standard TV minimization problem, as shown in Eq. (2). With a regularization parameter λ, it merely requires to compute $\mathbf{y}$ by optimizing the problem (6) via an existing TV minimization technique, and the global optimal solution of the fused image $\mathbf{x}^*$ is then determined by: $\mathbf{x}^* = \mathbf{y}^* + \mathbf{v}$. After we obtain the estimate of $\mathbf{x}^*$, we normalize it so that it has gray-scale value ranging from 0 to 255 by using the following rule:

$$\mathbf{x} = \begin{cases} 0 & \text{if } \mathbf{x}^* < 0 \\ 255 & \text{if } \mathbf{x}^* > 255 \\ \mathbf{x}^* & \text{otherwise.} \end{cases} \tag{7}$$

We summarize the proposed fusion method in Algorithm 1.

Algorithm 1. The proposed fusion algorithm.

Input: IR image $\mathbf{u}$, visible image $\mathbf{v}$, parameter λ
Output: Fused image $\mathbf{x}$
1 Compute $\mathbf{y}^*$ by optimizing the problem (6) using the algorithm introduced in [36];
2 Calculate the optimal estimation of $\mathbf{x}^*$ as $\mathbf{x}^* = \mathbf{y}^* + \mathbf{v}$;
3 The final fused image $\mathbf{x}$ is obtained by normalizing $\mathbf{x}^*$ using Eq. (7).

3. Experimental results

In this section, we test the performance of our method on a publicly available dataset, and compare it with seven state-of-the-art fusion methods such as Laplacian pyramid (LP) [20], ratio of low-pass pyramid (RP) [22], Wavelet [23], dual-tree complex wavelet transform (DTCWT) [24], curvelet transform (CVT) [25], multi-resolution singular value decomposition (MSVD) [26], and Laplacian pyramid with sparse representation (LP-SR) [29]. All the seven algorithms are implemented based on publicly available codes, where the parameters are set according to the original paper, and we try our best to tune some details. The experiments are performed on a laptop with 2.5 GHz Intel Core CPU, 8 GB memory and Matlab Code.

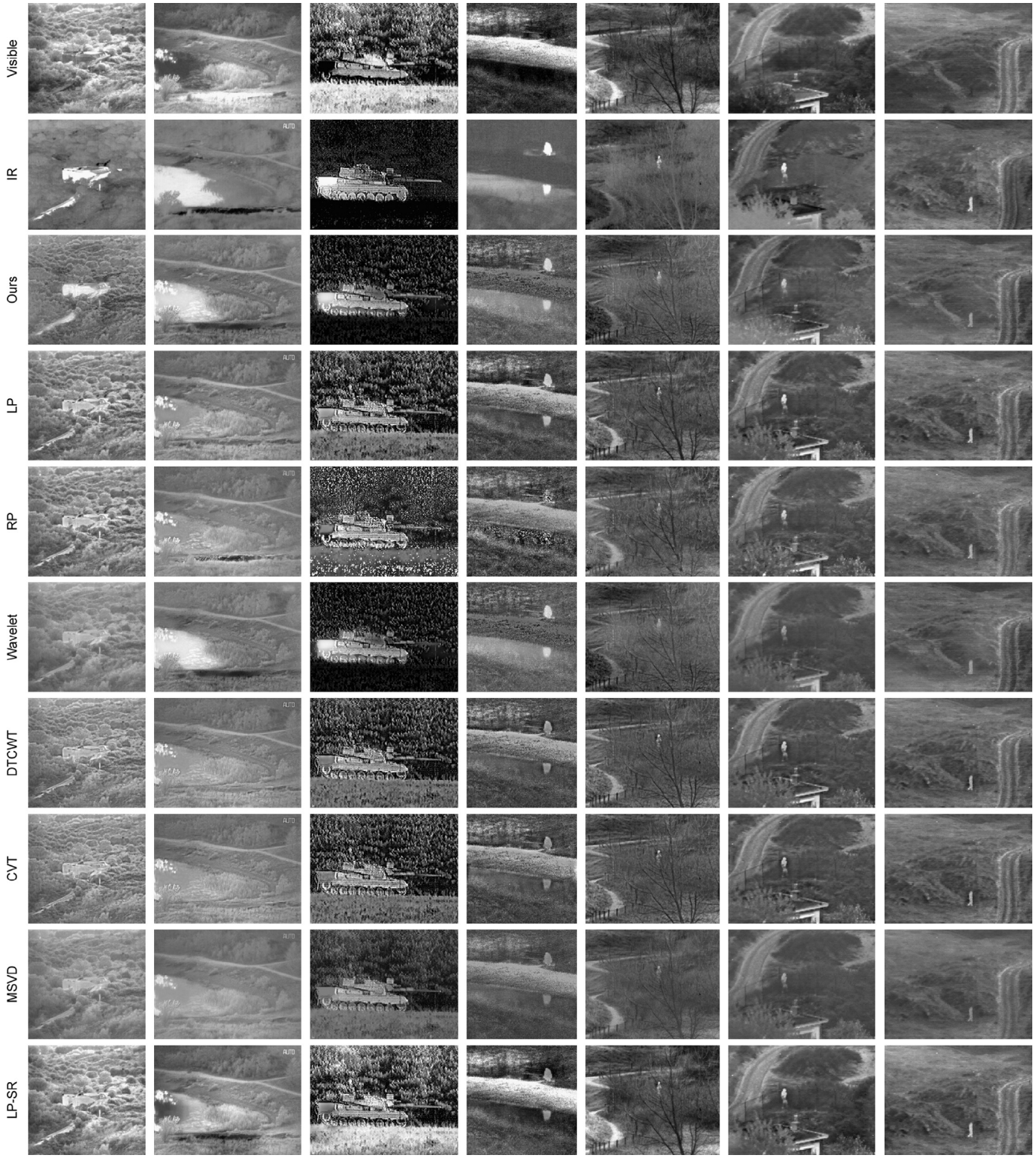


Fig. 2. Qualitative fusion results on the seven image pairs of *Bunker*, *Lake*, *Tank*, *Bench*, *Sandpath*, *Nato_camp* and *Duine* (from left to right). From top to bottom: visible images, IR images, our fusion results, results of LP [20], RP [22], Wavelet [23], DTCWT [24], CVT [25], MSVD [26], and LP-SR [29], respectively. Clearly, only our method can preserve both the thermal radiation and the texture information in the source images.

3.1. Datasets and settings

We test our method on the surveillance images from TNO Human Factors, which contains multispectral (intensified visual, near-infrared, and longwave infrared or thermal) nighttime imagery of different military relevant scenarios, registered with different multiband camera systems such as Athena, DHV and FEL. The dataset is available at http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029. We choose seven typical pairs and an image sequence for both qualitative illustration and quantitative comparison. For Athena camera system, the visual and infrared images were acquired by the Charge Coupled Device

Table 1
Quantitative comparisons of MI on images in Fig. 2. Bold indicates the best result.

Alg.	*Bunk.*	*Lake*	*Tank*	*Bench*	*Sand.*	*Nato.*	*Duine*
LP [20]	1.52	1.54	1.99	1.94	1.07	1.52	1.44
RP [22]	1.44	1.62	1.20	2.04	1.04	1.42	1.36
Wavelet [23]	1.57	2.03	**3.53**	2.69	1.13	1.61	1.59
DTCWT [24]	1.33	1.61	1.88	1.83	1.01	1.49	1.41
CVT [25]	1.29	1.54	1.80	1.78	0.98	1.42	1.33
MSVD [26]	1.43	1.86	2.28	1.99	1.03	1.52	1.43
LP-SR [29]	**3.42**	**2.29**	3.52	3.49	**2.89**	1.74	1.55
Ours	3.14	1.35	2.28	**4.45**	2.34	**4.58**	**5.40**

Table 2
Quantitative comparisons of Q^G on images in Fig. 2. Bold indicates the best result.

Alg.	*Bunk.*	*Lake*	*Tank*	*Bench*	*Sand.*	*Nato.*	*Duine*
LP [20]	0.66	0.62	0.51	0.70	0.49	0.48	0.48
RP [22]	0.53	0.50	0.19	0.58	0.38	0.41	0.41
Wavelet [23]	0.26	0.28	0.18	0.32	0.25	0.28	0.30
DTCWT [24]	0.59	0.57	0.49	0.67	0.45	0.41	0.42
CVT [25]	0.57	0.55	0.40	0.65	0.40	0.37	0.37
MSVD [26]	0.35	0.38	0.15	0.44	0.31	0.28	0.28
LP-SR [29]	0.67	0.62	**0.51**	0.72	0.50	0.47	0.48
Ours	**0.70**	**0.66**	0.46	**0.74**	**0.60**	**0.55**	**0.51**

Table 3
Quantitative comparisons of FMI on images in Fig. 2. Bold indicates the best result.

Alg.	*Bunk.*	*Lake*	*Tank*	*Bench*	*Sand.*	*Nato.*	*Duine*
LP [20]	0.54	0.51	0.46	0.52	0.45	0.42	0.45
RP [22]	0.46	0.46	0.39	0.51	0.41	0.40	0.42
Wavelet [23]	0.46	0.45	0.41	0.46	0.41	0.41	0.42
DTCWT [24]	0.50	0.45	0.43	0.50	0.43	0.38	0.42
CVT [25]	0.49	0.47	0.39	0.49	0.40	0.36	0.39
MSVD [26]	0.37	0.37	0.23	0.39	0.35	0.31	0.31
LP-SR [29]	0.55	0.52	**0.47**	0.54	0.46	0.42	0.45
Ours	**0.60**	**0.57**	0.46	**0.57**	**0.58**	**0.58**	**0.58**

Table 4
Quantitative comparisons of EN on images in Fig. 2. Bold indicates the best result.

Alg.	*Bunk.*	*Lake*	*Tank*	*Bench*	*Sand.*	*Nato.*	*Duine*
LP [20]	7.14	6.76	7.45	7.11	6.59	6.69	6.57
RP [22]	6.98	6.78	7.57	7.31	6.55	6.59	6.48
Wavelet [23]	6.70	6.53	7.18	6.53	6.08	6.25	6.29
DTCWT [24]	6.92	6.66	7.40	6.94	6.41	6.51	6.47
CVT [25]	6.92	6.67	7.41	6.97	6.45	6.56	6.50
MSVD [26]	6.74	6.57	7.24	6.69	6.16	6.29	6.32
LP-SR [29]	7.43	7.20	**7.92**	**7.45**	**7.11**	6.91	**6.73**
Ours	**7.44**	**7.34**	6.73	6.83	7.05	**6.93**	6.29

(CCD) sensor and forward-looking infrared (FLIR) sensor, respectively. The visual image of the DHV camera system was acquired by the visual and near-infrared intensified low-light CCD sensor, and the IR image of the DHV camera system was acquired by the thermal middle wavelength infrared sensor. For FEL camera system, the visible-light camera was a Siemens K235 CCD video camera, with a 756×581 CCD chip, and the infrared (IR) camera was an Amber Radiance 1 (Goleta, CA, USA) Focal Plane Array (FPA) camera, with an array of 256×256 pixels, operating in the 3–5 μm (mid-range) band. Moreover, the image pairs have all been aligned in advances. For handling unaligned images, an off-the-shelf multimodal image registration algorithm could be used to align the image pairs as preprocessing [3].

Usually, the performance of image fusion results can be both evaluated in subjective and objective ways. Since there is little difference among fusion results under most circumstances, it is difficult to correctly evaluate the fusion results in a subjective way. In recent years, many fusion metrics have been proposed, which are generally based on the measurement of the transfer of a feature (e.g., edges, amount of information) from the source images into the fused image [47]. However, none of them is definitely better than others. Therefore, it is necessary to apply multiple metrics to make a comprehensive evaluation of different fusion methods. In this paper, we quantitatively evaluate the performances of different fusion methods using four metrics, i.e., mutual information (MI) [48], gradient based fusion metric (Q^G) [49], feature mutual information (FMI) [50], and entropy (EN) [29]. In detail, MI is a basic concept of information theory measuring the amount of information that one variable contains about another, Q^G is a popular fusion metric which computes the amount of gradient information injected into the fused image from the source images, FMI calculates the mutual information of the image features which is in consistency with the subjective quality measures, and EN is used to measure the amount of information in the fused image. For all the four metrics, larger values indicate better performance.

Parameter Initialization: There is only one parameter in our proposed algorithm, i.e., the regularization parameter λ. It controls the trade-off of the information preservation between the two source images. Generally, small value of λ indicates keeping more thermal radiation information, while large value of λ corresponds to transferring more texture information. The fusion result will be the original IR image when $\lambda = 0$ and the original visible image when $\lambda = +\infty$. Throughout our experiments, we fix λ to 0.4 as an empirical value, which can achieve good visual effects in most cases.

3.2. Qualitative comparisons

We first give an intuitive impression on the performance of the proposed method. To this end, we perform experiments on three typical image pairs, i.e., *Bunker*, *Lake* and *Tank*, and use a popular fusion method Wavelet [23] for comparison. The results are illustrated in Fig. 1. The left two columns are the visible and IR images to be fused, where the visible images contain detailed backgrounds and the IR images highlight the targets, i.e. the building, lake and tank. The third column is the fusion results of Wavelet. We see that the fusion results of Wavelet cannot well preserve the details in the source images and fail to highlight the target as the backgrounds have the similar pixel intensities with the targets, such as the building in the first row and the tank in the third row. This demonstrates the importance of keeping the thermal radiation distribution in the IR image, and the advantage will be magnified when the scene contains false targets which often occurs in military applications. The final column in Fig. 1 is the fusion results of our proposed algorithm, which aims to fuse both the thermal radiation information in IR images and the texture information in visible images. Our results tend to preserve the thermal radiation distribution in the IR images and hence, the targets can be easily detected. Meanwhile, the details of the backgrounds in the visible images are also kept in our results.

The qualitative comparison on all the seven image pairs with seven state-of-the-art methods are presented in Fig. 2. We see that only our method tends to simultaneously preserve the thermal radiation and the texture information in the source images. The fused images look like sharpened IR images with clear highlighted targets, and hence it will be beneficial for automatical target detection. This is the major difference between our method and the other existing fusion methods.

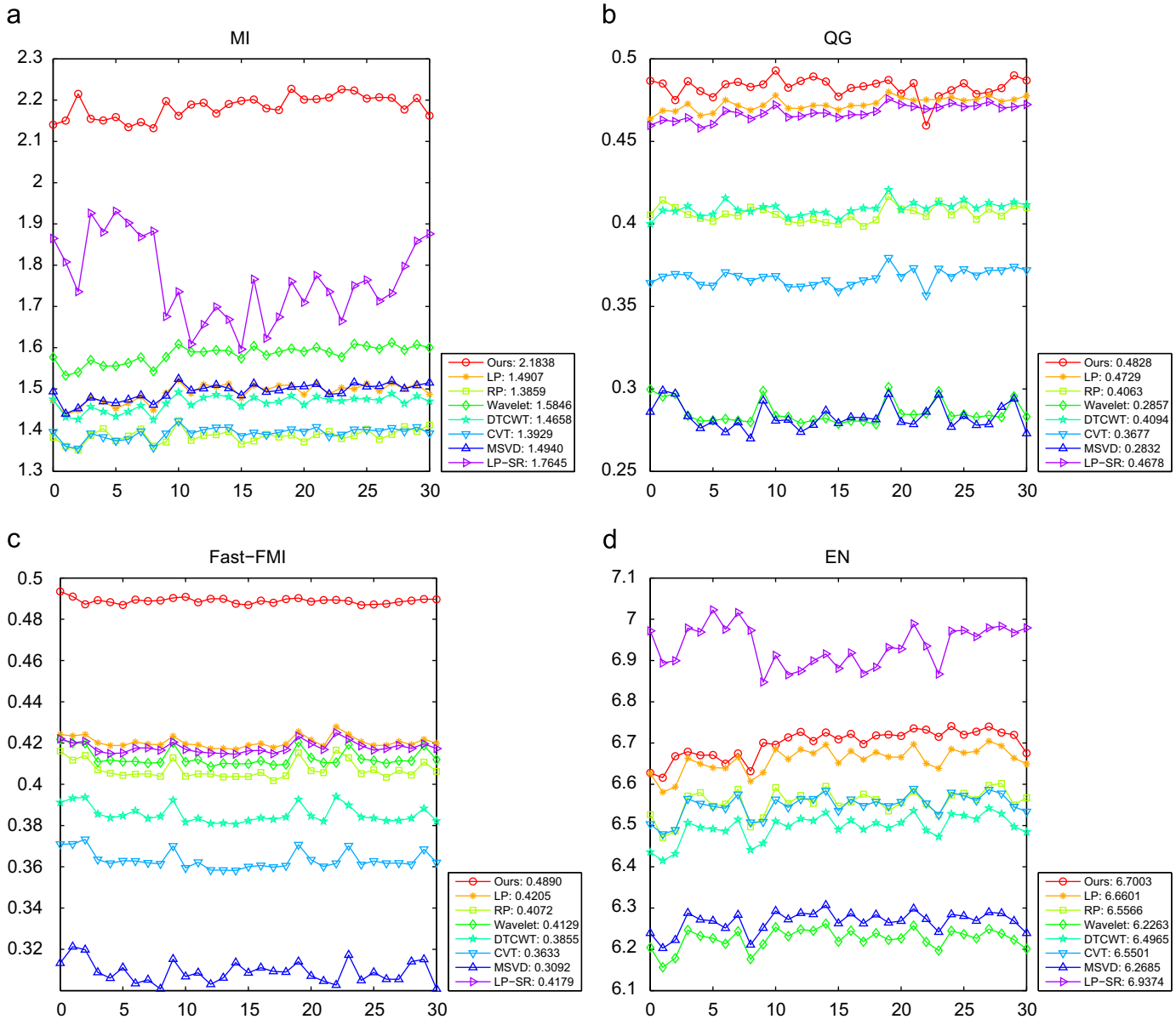


Fig. 3. Quantitative comparisons of the four metrics, i.e., (a) MI, (b) Q^G, (c) FMI and (d) EN on the *Nato_camp* sequence. The seven state-of-the-art methods such as LP [20], RP [22], Wavelet [23], DTCWT [24], CVT [25], MSVD [26], and LP-SR [29] are used for comparison. For all the four metrics, larger values indicate better performance. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

3.3. Quantitative comparisons

We next provide quantitative comparisons of the eight methods on the seven testing image pairs. The results are reported in Tables 1–4, where the best values are highlighted with red color. We see that our proposed method can produce the best metric values in most cases, followed by the LP-SR [29] method. More specifically, LP-SR has comparable good MIs and ENs to our method, and our method has better Q^Gs and FMIs than the other compared fusion methods. It is interesting that even the traditional LP [20] method can produce comparable results with several state-of-the-art methods. This means that it may not be difficult to preserve details in the source images, which further highlights the significance of preserving the thermal radial information in IR images.

We further provide quantitative comparisons of the eight methods on an image sequence (i.e., *Nato_camp*) which contains 32 image pairs, and an example pair can be seen from the sixth column in Fig. 2. The quantitative results are reported in Fig. 3, where our proposed method marked with red circles consistently has the best MIs and FMIs, the second best ENs, and has the best Q^G on most image pairs. This demonstrates that our method tends to not only keep the important thermal radial information, but also preserve a large amount of details in the source images. The average run time of our method on the sequence is about 0.15 s per pair, where the images are all of size 270×360.

4. Conclusion

Within this paper, we propose a novel IR and visible fusion method based on TV minimization. It tends to simultaneously keep the thermal radiation information in the IR image and preserve appearance information in the visible image. The fusion results look like high-resolution IR images with clear highlighted targets and hence, it will be beneficial for fusion-based target detection and recognition systems. The quantitative comparisons on several

metrics with other seven state-of-the-art fusion methods demonstrate that our method can not only identify the most important information, but also can keep the largest or approximately the largest amount of information in the source images.

The model proposed in this paper deals with IR and visible image fusion problem, however, it is general and can be also applied to other image processing problems such as super-resolution [51–54], whose basic is the fusion of a sequence of low-resolution noisy blurred images to produce a higher-resolution image or sequence.

Acknowledgments

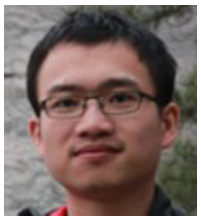
This work was supported by the National Natural Science Foundation of China under Grant nos. 61503288 and 41501505, the China Postdoctoral Science Foundation under Grant no. 2015M570665, the China Scholarship Council Foundation under Grant no. 201506415020, and the Open foundation of China University of Geoscience under Grant no. AU2015CJ05.

References

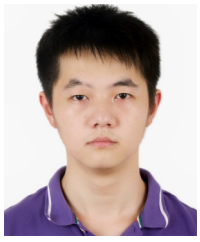
[1] C. Chen, Y. Li, W. Liu, J. Huang, Image fusion with local spectral consistency and dynamic gradient sparsity, in: IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2760–2765.
[2] L. Dong, Q. Yang, H. Wu, H. Xiao, M. Xu, High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform, Neurocomputing 159 (2015) 268–274.
[3] J. Ma, J. Zhao, Y. Ma, J. Tian, Non-rigid visible and infrared face registration via regularized gaussian fields criterion, Pattern Recognit. 48 (3) (2015) 772–784.
[4] N. Wang, Y. Ma, K. Zhan, Spiking cortical model for multifocus image fusion, Neurocomputing 130 (2014) 44–51.
[5] H. Li, B. Manjunath, S.K. Mitra, Multisensor image fusion using the wavelet transform, Graph. Models Image Process. 57 (3) (1995) 235–245.
[6] G. Bhatnagar, Q.J. Wu, Z. Liu, A new contrast based multimodal medical image fusion framework, Neurocomputing 157 (2015) 143–152.
[7] F. Meng, B. Guo, M. Song, X. Zhang, Image fusion with saliency map and interest points, Neurocomputing 177 (2016) 1–8.
[8] Y. Li, C. Tao, Y. Tan, K. Shang, J. Tian, Unsupervised multilayer feature learning for satellite image scene classification, IEEE Geosci. Remote Sens. Lett. 13 (2) (2016) 157–161.
[9] A. Toet, J.K. IJspeert, A.M. Waxman, M. Aguilar, Fusion of visible and thermal imagery improves situational awareness, Displays 18 (2) (1997) 85–95.
[10] S.G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B.R. Abidi, A. Koschan, M. Yi, M. A. Abidi, Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition, Int. J. Comput. Vis. 71 (2) (2007) 215–233.
[11] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, Inf. Fusion 4 (4) (2003) 259–280.
[12] J. Saeedi, K. Faez, Infrared and visible image fusion using fuzzy logic and population-based optimization, Appl. Soft Comput. 12 (3) (2012) 1041–1054.
[13] V.S. Petrovic, C.S. Xydeas, Gradient-based multiresolution image fusion, IEEE Trans. Image Process. 13 (2) (2004) 228–237.
[14] M. Kumar, S. Dass, A total variation-based algorithm for pixel-level image fusion, IEEE Trans. Image Process. 18 (9) (2009) 2137–2143.
[15] J. Ma, J. Zhao, A.L. Yuille, Non-rigid point set registration by preserving global and local structures, IEEE Trans. Image Process. 25 (1) (2016) 53–64.
[16] J. Ma, J. Zhao, J. Tian, A.L. Yuille, Z. Tu, Robust point matching via vector field consensus, IEEE Trans. Image Process. 23 (4) (2014) 1706–1721.
[17] X. Bai, L.J. Latecki, Path similarity skeleton graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 30 (7) (2008) 1282–1292.
[18] X. Bai, B. Wang, C. Yao, W. Liu, Z. Tu, Co-transduction for shape retrieval, IEEE Trans. Image Process. 21 (5) (2012) 2747–2757.
[19] J. Ma, W. Qiu, J. Zhao, Y. Ma, A.L. Yuille, Z. Tu, Robust L_2E estimation of transformation for non-rigid registration, IEEE Trans. Signal Process. 63 (5) (2015) 1115–1129.
[20] P.J. Burt, E.H. Adelson, The laplacian pyramid as a compact image code, IEEE Trans. Commun. 31 (4) (1983) 532–540.
[21] A. Toet, L.J. Van Ruyven, J.M. Valeton, Merging thermal and visual images by a contrast pyramid, Opt. Eng. 28 (7) (1989) 287789.
[22] A. Toet, Image fusion by a ratio of low-pass pyramid, Pattern Recognit. Lett. 9 (4) (1989) 245–253.
[23] P. Zeeuw, Wavelets and Image Fusion, CWI, Amsterdam, March, 1998.
[24] J.J. Lewis, R.J. O'Callaghan, S.G. Nikolov, D.R. Bull, N. Canagarajah, Pixel-and region-based image fusion with complex wavelets, Inf. Fusion 8 (2) (2007) 119–130.
[25] F. Nencini, A. Garzelli, S. Baronti, L. Alparone, Remote sensing image fusion using the curvelet transform, Inf. Fusion 8 (2) (2007) 143–156.
[26] V. Naidu, Image fusion technique using multi-resolution singular value decomposition, Def. Sci. J. 61 (5) (2011) 479–484.
[27] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013) 2864–2875.
[28] S. Li, X. Kang, J. Hu, B. Yang, Image matting for fusion of multi-focus images in dynamic scenes, Inf. Fusion 14 (2) (2013) 147–162.
[29] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, Inf. Fusion 24 (2015) 147–164.
[30] C. Chen, Y. Li, W. Liu, J. Huang, Sirf: simultaneous satellite image registration and fusion in a unified framework, IEEE Trans. Image Process. 24 (11) (2015) 4213–4224.
[31] J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, Inf. Fusion 31 (2016) 100–109.
[32] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, Inf. Fusion 13 (1) (2012) 10–19.
[33] G. Pajares, J.M. De La Cruz, A wavelet-based image fusion tutorial, Pattern Recognit. 37 (9) (2004) 1855–1872.
[34] A.L. Da Cunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.
[35] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[36] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging Vis. 20 (1–2) (2004) 89–97.
[37] A. Beck, M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, IEEE Trans. Image Process. 18 (11) (2009) 2419–2434.
[38] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D: Nonlinear Phenom. 60 (1) (1992) 259–268.
[39] T. Chan, S. Esedoglu, F. Park, A. Yip, Recent developments in total variation image restoration, Math. Models Comput. Vis. 17 (2005) 2.
[40] C.R. Vogel, M.E. Oman, Iterative methods for total variation denoising, SIAM J. Sci. Comput. 17 (1) (1996) 227–238.
[41] T.F. Chan, C.-K. Wong, Total variation blind deconvolution, IEEE Trans. Image Process. 7 (3) (1998) 370–375.
[42] J. Shen, T.F. Chan, Mathematical models for local nontexture inpaintings, SIAM J. Appl. Math. 62 (3) (2002) 1019–1043.
[43] Q. Yuan, L. Zhang, H. Shen, Regional spatially adaptive total variation super-resolution with spatial information filtering and clustering, IEEE Trans. Image Process. 22 (6) (2013) 2327–2342.
[44] G. Gilboa, A total variation spectral framework for scale and texture analysis, SIAM J. Imaging Sci. 7 (4) (2014) 1937–1961.
[45] Q. Liu, B. Xiong, D. Yang, M. Zhang, A generalized relative total variation method for image smoothing, Multimed. Tools Appl. (2015) 1–22.
[46] B. Khaleghi, A. Khamis, F.O. Karray, S.N. Razavi, Multisensor data fusion: a review of the state-of-the-art, Inf. Fusion 14 (1) (2013) 28–44.
[47] N. Cvejic, T. Seppanen, S.J. Godsill, A nonreference image fusion metric based on the regional importance measure, IEEE J. Sel. Top. Signal Process. 3 (2) (2009) 212–221.
[48] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315.
[49] C. Xydeas, V. Petrović, Objective image fusion performance measure, Electron. Lett. 36 (4) (2000) 308–309.
[50] M.B.A. Haghighat, A. Aghagolzadeh, H. Seyedarabi, A non-reference image fusion metric based on mutual information of image features, Comput. Electr. Eng. 37 (5) (2011) 744–756.
[51] J. Jiang, X. Ma, Z. Cai, R. Hu, Sparse support regression for image super-resolution, IEEE Photonics J. 7 (5) (2015) 1–11.
[52] J. Jiang, R. Hu, Z. Han, T. Lu, Efficient single image super-resolution via graph-constrained least squares regression, Multimed. Tools Appl. 72 (3) (2014) 2573–2596.
[53] Z.-Y. Wang, Z. Han, R.-M. Hu, J.-J. Jiang, Noise robust face hallucination employing Gaussian–Laplacian mixture model, Neurocomputing 133 (2014) 153–160.
[54] J. Jiang, R. Hu, Z. Wang, Z. Han, J. Ma, Facial image hallucination through coupled-layer neighbor embedding, IEEE Trans. Circuits Syst. Video Technol. http://dx.doi.org/10.1109/TCSVT.2015.2433538, in press.

Yong Ma graduated from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997. He received the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003. His general field of research is in signal and systems. His current research projects include remote sensing of the Lidar and infrared, as well as Infrared image processing, pattern recognition, interface circuits to sensors and actuators. Between 2004 and 2006, he was a Lecturer at the University of the West of England, Bristol, U.K. Between 2006 and 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, Wuhan, where he was a Professor of electronics. He is now a Professor with the Electronic Information School, Wuhan University.

Jun Chen received her B.S. and M.E. degrees from the School of Mechanical and Electronic Information in China University of Geosciences, Wuhan, China, in 2002 and 2004, respectively. She received her Ph.D. degree from the Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2014. She has been an assistant professor in China University of Geosciences, Wuhan, China, since 2008. She is now a visiting scholar with the Department of Computer Science, Northwest University. Her research interests include computer vision and machine learning.



Chen Chen received the B.E. degree and M.S. degree both from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2011, respectively. He received a master degree in the Department of Computer Science and Engineering, University of Texas at Arlington in 2015. He is now a Ph.D. student in Department of Electrical and Computer Engineering in University of Illinois at Urbana–Champaign. His major research interests include image processing and computer vision.



Fan Fan received the B.S. degree of Communication Engineering, and the Ph.D. degree of Electronic Circuit and System, both from the Huazhong University of Science and Technology, Wuhan, China, in 2009, and 2015, respectively. He is currently a post doctor in the School of Electronic Information, Wuhan University, China. His current research interests include infrared thermal imaging, machine learning and computer vision.

Jiayi Ma received the B.S. degree of Information and Computing Science, and the Ph.D. degree of Control Science and Engineering, both from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was with the Department of Statistics, University of California at Los Angeles. He is currently an Associate Professor with the Electronic Information School, Wuhan University. His current research interests include in the areas of computer vision, machine learning, and pattern recognition.