

Infrared and visible images fusion by using sparse representation and guided filter

Qilei Li, Wei Wu, Lu Lu, Zuoyong Li, Awais Ahmad & Gwanggil Jeon

To cite this article: Qilei Li, Wei Wu, Lu Lu, Zuoyong Li, Awais Ahmad & Gwanggil Jeon (2019): Infrared and visible images fusion by using sparse representation and guided filter, Journal of Intelligent Transportation Systems, DOI: [10.1080/15472450.2019.1643725](https://doi.org/10.1080/15472450.2019.1643725)

To link to this article: <https://doi.org/10.1080/15472450.2019.1643725>



Published online: 01 Aug 2019.



Submit your article to this journal [↗](#)




Article views: 29



View Crossmark data [↗](#)



Infrared and visible images fusion by using sparse representation and guided filter

Qilei Li^{a,b} , Wei Wu^{a,b}, Lu Lu^a, Zuoyong Li^b, Awais Ahmad^c, and Gwanggil Jeon^{d,e}

^aCollege of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan, China; ^bFujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, China; ^cDepartment of Computer Science, Bahria University, Islamabad, Pakistan; ^dSchool of Electronic Engineering, Xidian University, Xi'an, China; ^eDepartment of Embedded Systems Engineering, Incheon National University, Incheon, Korea

ABSTRACT

Infrared and visible images play an important role in transportation systems since they can monitor traffic conditions around the clock. However, visible images are susceptible to the imaging environments, and infrared images are not rich enough in detail. The infrared and visible images fusion techniques can fuse these two different modal images into a single image with more useful information. In this paper, we propose an effective infrared and visible images fusion method for traffic systems. The weight maps are measured by utilizing the sparse coefficients. The next is to decompose the infrared and visible pair into high-frequency layers (HFLs) and low-frequency layers (LFLs). Since the two layers contain different structures and texture information, to extract the representative component, the guided filter is utilized to optimize weight maps in accordance with the different characteristic of the infrared and visible pairs. The final step is to reconstruct the two-scale layers according to the weight maps. Experimental results demonstrate our method outperforms other popular approaches in terms of subjective perception and objective metrics.

ARTICLE HISTORY

Received 5 December 2018
Revised 1 July 2019
Accepted 11 July 2019

KEYWORDS

Guided filter (GF); image fusion; sparse representation (SR); weight map

Introduction

Infrared and visible images play an important role in transportation systems (Li, Khoshelham, Sarvi, & Haghani, 2019). Infrared images are obtained according to the thermal radiation emitted from the objects, and they are less influenced by weather and light condition (Jeon, Lee, & Sohn, 2018). However, the background information in the infrared image is often missing. In contrast, visible images contain much more texture information, while it is very susceptible to imaging environments and lighting conditions. The infrared and visible images fusion technology (Dai, Zhao, & Li, 2018; Ma, Ma, & Li, 2019), as shown in Figure 1, can fuse infrared and visible images pairs into a single image. The fused image contains both texture information from the visible image and thermal radiation detail from the infrared image. Thus, it is more conducive to human observation and computer analysis.

Depending on the processing domain, the current major infrared and visible images fusion methods can be divided into two categories (Zhou & Omar, 2009): spatial domain methods and transform domain

methods. The spatial domain methods directly fuse the infrared and visible pair via the fusion rules. The classical fusion method is to average the infrared image and the visible image. Regrettably, the fused images obtained by this method are often unsatisfactory. To address this problem, Li, Kang, and Hu (2013) guided filter (GF) decomposed the infrared and visible images pairs into base layers and detail layers, then fuse them by using the GF. However, the activity level measurement in Li et al. (2013) is not accurate enough. Miao and Wang (2005) gauged the activity level in accordance with the image gradients. Zhou, Wang, Li, and Dong (2016) combined Gaussian filter with the bilateral filter to fuse the hybrid multi-scale decomposed images, while the decomposing process will consume much time. Recently, some deep learning-based image fusion methods were proposed. Liu, Chen, Peng, and Wang (2017) pioneered the use of convolutional neural networks for image fusion. For performance improvement, the Laplacian pyramid was used for multi-scale decomposition, and the image self-similarity was used to optimize the network

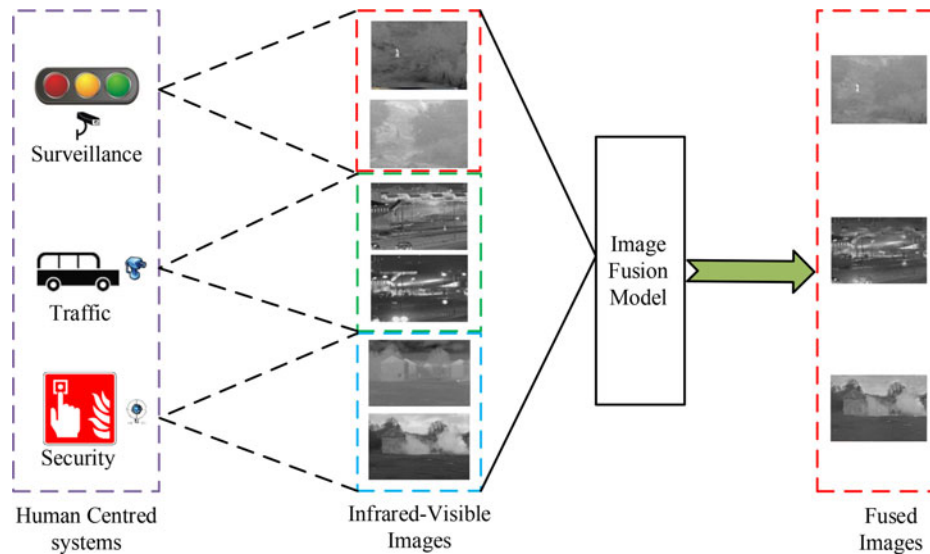


Figure 1. Fusion of infrared and visible images for urban traffic systems.

model (Liu, Chen, Cheng, Peng, & Wang, 2018). Following a different direction, Ma, Yu, Liang, Li, and Jiang (2018) proposed a “FusionGAN” model for infrared and visible images fusion based on generative adversarial network (GAN). Although these spatial domain methods can achieve good fusion results, they also cause many negative effects. They result in excessive smooth transitions at the edges, reduced contrast, and spectral distortion of the image.

For transform domain methods, multi-scale decomposition (MSD) is a powerful tool that used relatively broad. This study includes gradient pyramid (GP) (Petrovic & Xydeas, 2004), Laplacian pyramid (LP) (Burt & Adelson, 1983), discrete wavelet transform (DWT) (Li, Manjunath, & Mitra, 1995; Wang, Zhou, Yang, Liu, & Ren, 2019), dual-tree complex wavelet transform (DTCWT) (Lewis, O’Callaghan, Nikolov, Bull, & Canagarajah, 2007), and ratio of low-pass pyramid (RP) (Toet, 1989). Besides, several geometric analysis tools were widely deployed for image fusion. Nencini, Garzelli, Baronti, and Alparone (2007) employed the curvelet transform (CVT) to fuse the multi-modal images. Zhang and Guo (2009) decomposed the source images by using the nonsubsampling contourlet transform (NSCT). Sparse representation (SR) has been commonly employed in machine vision tasks, such as image super-resolution (Wu et al., 2016; Yang et al., 2017), face recognition (Wagner et al., 2012; Wright, Yang, Ganesh, Sastry, & Ma, 2009), and object detection (Agarwal, Awan, & Roth, 2004; Agarwal & Roth, 2002). For image fusion (Li, Yang, Wu, Liu, & Jeon, 2018, 2019; Li, Yin, & Fang, 2013; Liu, Liu, & Wang, 2015; Nejati, Samavi, & Shirani, 2015; Yang & Li, 2010), the general process is to

calculate the sparse coefficients of the source images, then to fuse the coefficient via a specific rule. The latter is to transform the fused coefficients to the result images. Liu and Wang (2015) trained various dictionaries, and selected the dictionary adaptively according to the gradient of the source image. Liu et al. tried to combine SR and NSCT together with the multi-scale decomposition in Liu et al. (2015) (NSCT-SR), which obtains satisfying fusion results. Liu et al. also introduced convolutional sparse representation (CSR) for image fusion in Liu, Chen, Ward, and Wang (2016). However, there are still some disadvantages remained. Since these methods fuse the coefficients in the transform domain, while a little change in the coefficients may lead to a great change in the spatial domain. As a result, such methods suffer from heavy undesirable artifacts.

To address these drawbacks, a novel infrared and visible images fusion approach is proposed in this paper. In the proposed method, we first utilize sparse coefficients to measure the activity level of the source infrared and visible images pairs. The activity level of the infrared and visible pair will be regarded as the weight map. The next is to decompose the infrared and visible pair into two-scale layers, namely high-frequency layers (HFLs) and low-frequency layers (LFLs). Because the two types of layers contain different structures and texture information, to extract the representative component, the GF (He, Sun, & Tang, 2010) was exploited to optimize weight maps in accordance with the different characteristic of the infrared and visible images pairs. The final step is to reconstruct the two-scale layers according to the weight maps. Compared with existing methods, three improvements can be listed as follows.

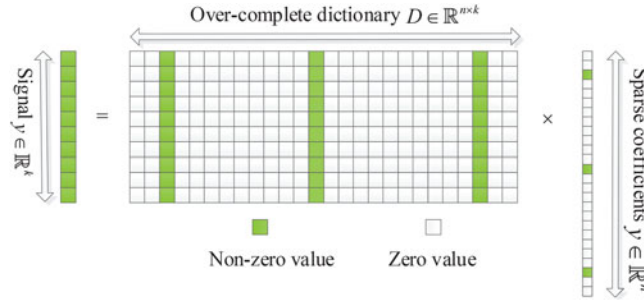


Figure 2. Illustration of sparse representative.

1. Traditional infrared and visible images fusion methods based on the MSD need to decompose the source images into more than two layers. This may cause a huge computational cost. Our method only needs to decompose the infrared and visible images pairs into LFL and HFL. Hence, our method enjoys lower computational complexity, owing to the fact that it does not require heavy decomposition scales.
2. The sparse coefficients are utilized to measure the activity level. This takes full advantage of the characteristics of the image. It also avoids the huge negative effects caused by the inaccurate measurement of activity level, such as ringing effect, blockiness, and edge blurring.
3. The GF is used to optimize the weight maps for LFL and HFL, respectively. By doing so, the representative information in the infrared and visible images can be well preserved. Hence, the visual effect of the fused image can be enhanced.

Related work

Sparse representation

SR has been used in many computer vision tasks, such as image super-resolution (Yang, Wu, Liu, Chen, & Zhou, 2018), image fusion (Li, Yang, Wu, Liu, & Jeon, 2018), and object detection (Zhang, Wang, Xie, & Li, 2018). It is based on the principle that a single can be interpreted by an over-complete dictionary and the corresponding sparse coefficients. Consider a set of signal $Y = \{y_1, \dots, y_M\} \in \mathbb{R}^{n \times M}$, each signal y_i can be expressed as

$$\min_{x_i} \|y_i - Dx_i\|_2^2 \quad \text{s.t.} \|x_i\|_0 \leq t, 1 \leq i \leq M, \quad (1)$$

where $y_i \in \mathbb{R}^n$, $D \in \mathbb{R}^{n \times k}$ is the dictionary with k atoms, $X = \{x_i\}_{i=1}^N$, $x_i \in \mathbb{R}^M$ is the sparse coefficient, and t is the number of nonzero elements. The illustration of SR is given in Figure 2.

Guided filter

GF (He et al., 2010) can extract the information from the guidance image and merge the information into the input image. GF functions differently depend on the input image and the guidance image. On one hand, it works as an edge-preserving filter when the guidance image is same as the input image. On the other hand, GF also merges the structure information from the guidance image into the input image when the guidance image and the input image are different. Figure 3 shows the working principle of the filtering process.

The output of GF is obtained by linear transform in the local window of the input image. We denote the guidance image, the input image, and the output image as G , I , and O , respectively. The output image O is obtained by the linear transform of G at pixel p , where p denotes the center of ω_k . This process formulated by

$$O_i = a_k G_i + b_k \quad \forall i \in \omega_k, \quad (2)$$

where ω_k is the local window with the size of $(2r + 1) \times (2r + 1)$, O_i is the output of ω_k , and a_k and b_k are the coefficients, which can be obtained by minimizing

$$E(a_k, b_k) = \sum_{i \in \omega_k} \left((a_k G_i - I_i)^2 + \lambda a_k^2 \right), \quad (3)$$

where $E(\cdot)$ denotes the mathematical expectation, λ is the manually set regularization factor, coefficients a_k and b_k can be obtained by the following linear regression:

$$a_k = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} G_i I_i - \mu_k \bar{P}_k}{\delta_k + \epsilon} \quad (4)$$

$$b_k = \bar{I}_k - a_k \mu_k,$$

where μ_k and δ_k are the mean and variance of G in ω_k , $|\omega|$ denotes the pixel amount in ω_k , and \bar{P}_k is mean of P_k . By solving the (2), O_i can be obtained

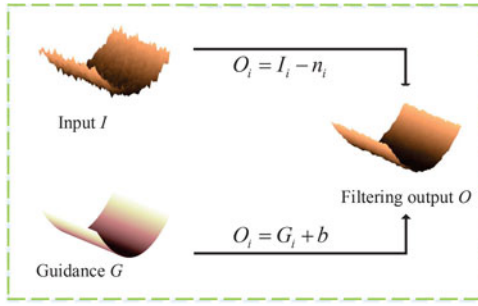


Figure 3. Illustration of GF.

$$O_i = \bar{a}_k G_i + \bar{b}_k. \quad (5)$$

Finally, O can be calculated by averaging the overlapping O_i . In this paper, we denote the procedure of GF as:

$$O = GF(I, G, r, \epsilon). \quad (6)$$

Proposed method

To obtain a fused image which preserves the representative information in the infrared and visible images, we measure the activity map to obtain an initial weight map. Then, we use GFs with different parameters to optimize the initial map to generate the optimized weight maps. The optimized weight maps, namely W^L and W^H , are elaborately designed for the high frequency layer and low frequency layer. The W^L can fuse the big scale structures and the W^H can fuse the texture detail. Finally, by combining the two components, the fused image can be obtained. Thus, the fused image preserves the representative information from the infrared and visible images.

The proposed method consists of four parts, as shown in Figure 4: (1) initial weight maps measurement by SR; (2) two-scales decomposition of infrared and visible; (3) weight maps optimization with the GF; and (4) two-scales reconstruction of infrared and visible.

Initial weight maps measurement by SR

The activity level is measured in accordance with the sparse coefficients, which stand for the characteristic of the source images. To calculate the over-complete dictionary, three infrared and visible images pairs are utilized. A large number of image patches with the size of $n \times n$ are sampled from the training images. By solving (1), the over-complete dictionary D can be obtained (See Figure 5). Then, the dictionary D is employed to calculate the SR coefficients of the input test infrared and visible images pairs. A sliding

window with the size of $n \times n$ is used to sample patches from the test images pixel by pixel. With dictionary D , these patches can be used to learn the coefficients via solving following formula:

$$\hat{X}_i = \min_{X_i} \|X_i\|_1 \quad \text{s.t.} \|Y_i - D\hat{X}_i\|_2 \leq \delta, \quad (7)$$

where $Y_i, (i = 1, 2)$ is the input image patches (i.e., infrared image or visible image), $\hat{X}_i, (i = 1, 2)$ is the calculated sparse coefficients, and δ is a pre-defined constant.

Then, we measure the activity level by calculating the L_0 -norm of the sparse coefficients $\hat{X}_i, (i = 1, 2)$ as

$$A_{ij} = \|\hat{X}_{i,j}\|_0 \quad i = (1, 2), j = (1, \dots, n), \quad (8)$$

where $\hat{X}_{i,j}$ denotes the coefficient of i_{th} source image with j_{th} patches, n is the total number of the patches, and A is the calculated activity map.

The element in A is obtained according to the corresponding patch from the source image. The next is to assign each element in $A_i, i = (1, 2)$ to the corresponding location in the source image to obtain the score map $S_i, i = (1, 2)$. The overlapping element in S will be averaged. This process can be expressed as

$$S_i = f(A_i), \quad i = (1, 2), \quad (9)$$

where $f(\cdot)$ is the element reshape function as discussed above.

The initial weight map is obtained as follows:

$$W_1(x, y) = \begin{cases} 1 & S_1(x, y) > S_2(x, y) \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

Since W_1 and W_2 are complementary, we have

$$W_2(x, y) = 1 - W_1(x, y). \quad (11)$$

Two-scale decomposition of infrared and visible images

In this subsection, we aim to decompose the infrared and visible source images into HFLs and LFLs. To achieve this goal, a simple yet effective method is to use an average filter. The LFLs can be obtained by

$$LFL_i = g(I_i) \quad i = 1, 2, \quad (12)$$

where $g(\cdot)$ denotes the average filtering operation, I_i denotes the source images (i.e., infrared image or visible image), and LFL_i is the corresponding LFL of I_i .

Once the LFLs are obtained, the corresponding HFLs can be easily calculated as follows:

$$HFL_i = I_i - LFL_i \quad i = 1, 2. \quad (13)$$

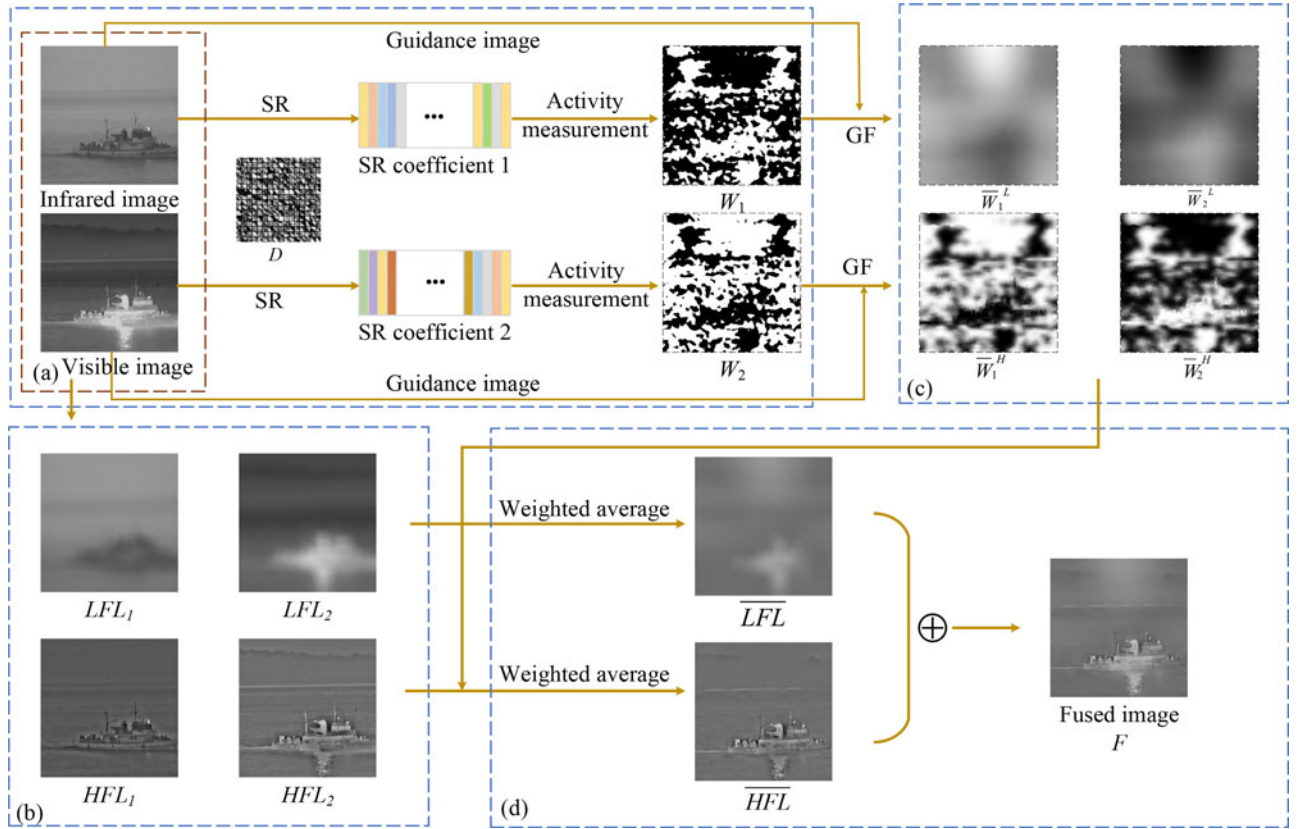


Figure 4. Schematic diagram of the proposed method. (a) Initial weight maps measurement by SR; (b) two-scales decomposition of infrared and visible; (c) weight maps optimization with the GF; and (d) two-scales reconstruction of infrared and visible.

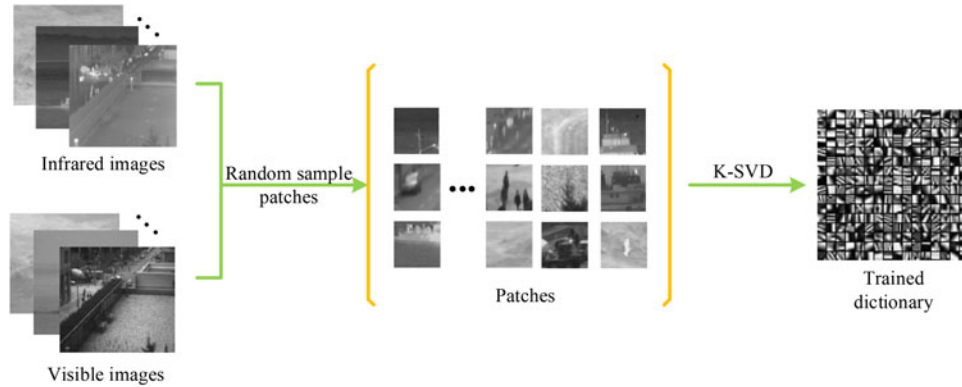


Figure 5. Training a dictionary.

By doing so, the infrared and visible images are decomposed into HFLs which contain big scale structures, and LFLs which contain small texture details.

Weight maps optimization with the GF

As is shown in Figure 4, the initial weight maps are noisy, which will produce undesirable artifacts. To solve this problem, GF is employed to achieve spatial consistency. We adjust the parameters of the GF to

suit the different conditions of HFLs and LFLs. This process can be summarized as:

$$\begin{aligned} \bar{W}_i^H &= GF(I_i, W_i, r_1, \epsilon_1) \\ \bar{W}_i^L &= GF(I_i, W_i, r_2, \epsilon_2), \end{aligned} \quad (14)$$

where $r_i, \epsilon_i, (i = 1, 2)$ are the parameter of GFs, \bar{W}_i^H , and \bar{W}_i^L are the optimized weight maps for HFLs and LFLs, respectively.

It can be seen in Figure 4, \bar{W}_i^L mainly aims to fuse the big scale structures, while \bar{W}_i^H focuses on texture detail.

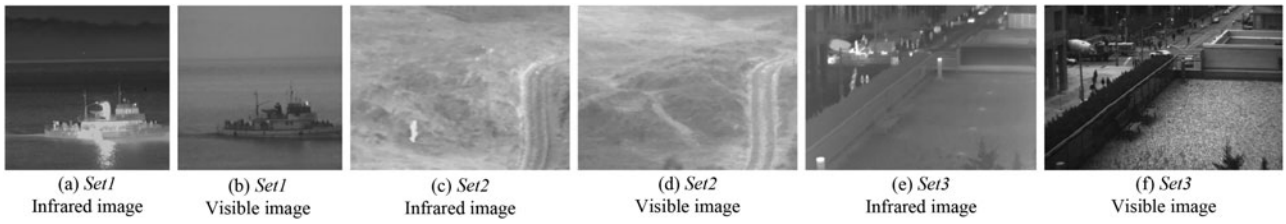


Figure 6. Three sets of infrared and visible images used to train the dictionary.

Two-scales reconstruction of infrared and visible

This subsection aims to reconstruct the two-scale images with the optimized weight maps. It can be divided into two steps. First, the LFLs and HFLs are fused, respectively via solving

$$\begin{aligned}\overline{LFL} &= \sum_{i=1}^2 \bar{W}_i^L \otimes LFL_i \\ \overline{HFL} &= \sum_{i=1}^2 \bar{W}_i^H \otimes LFH_i,\end{aligned}\quad (15)$$

where \otimes denotes the element-wise multiplication. \overline{LFL} and \overline{HFL} represent the fused low-frequency layer and high-frequency layer, respectively.

Second, the fusion result F can be obtained by combining the \overline{LFL} and \overline{HFL} , i.e.,

$$F = \overline{LFL} + \overline{HFL}.\quad (16)$$

Experimental results and analysis

Implementation details

To train the dictionary, we randomly sample 50,000 patches with the size of 8×8 from three sets of infrared and visible images (See Figure 6). The size of the dictionary is set to 64×256 , and the training iteration is fixed to 100. The nonzero limitation σ is set to 5. The average filter size used to separate LFLs from the source images are set to 31×31 conventionality. The parameters r_1 and ϵ_1 of the GF that used to optimize the W_i^H are set to 15 and 0.15, respectively. The parameters r_2 and ϵ_2 of the GF that used to optimize the W_i^L are set to 7 and $1e-6$, respectively. The parameter settings are summarized in Table 1.

The proposed method is compared with some popular image fusion methods on some classical infrared and visible images pairs. The compared methods include GFF (Li et al., 2013), LP (Burt & Adelson, 1983), RP (Toet, 1989), CVT (Nencini et al., 2007), NSCT (Zhang & Guo, 2009), SR (Yang & Li, 2010), NSCT-SR (Liu et al., 2015), ASR (Liu & Wang, 2015), and CSR (Liu et al., 2016). The parameters of these

Table 1. Parameters setting for the proposed method.

Training a dictionary			Refining weight maps			
Patch size	Number of patch	Dictionary size	r_1	ϵ_1	r_2	ϵ_2
8×8	50,000	64×256	15	0.15	7	$1e-6$

methods are set in accordance with the relevant publication.

We adopt four representative metrics to evaluate our method: mutual information (MI) (Qu, Zhang, & Yan, 2002), edge retention (Q^G) (Xydeas & Petrovic, 2000), feature mutual information (FMI) (Haghighat, Aghagolzadeh, & Seyedarabi, 2011), and the standard deviation (SD). In more detail, MI measures the information from the infrared and visible images pairs that the fused image contains; Q^G measures the gradient information of the fused image; FMI, a nonreference metric, aims to measure how much feature information the fused image contains; and SD measures the overall contrast of the fused image.

Subjective evaluation of the methods

Figures 7–10 exhibit the fusion results of four classical infrared and visible pairs. It can be seen the fusion results obtained via SR and NSCT-SR suffer from heavy artifacts. For instance, in Figures 7, 8, and 10, the fused images contain a large number of halos that are not part of the infrared and visible images pairs, which leads to degradation of image quality and seriously affects the visual effect. This is caused by that simply blending the coefficients, once improperly selected, can easily cause huge negative effects. As can be seen from these figures, the fusion results of our method better maintain the subjects while balanced the tone. That is because our method uses the sparse coefficient to mine the internal characteristics of the image. These MSD based methods, such as NSCT, CVT, LP, and RP, tend to be too close to a certain source map in the overall color. For example, the overall color of Figures 7 and 10 is close to the infrared images, which fail to take the advantages of the rich details in the visible images. Since MSD methods severely rely on the decomposition level, while

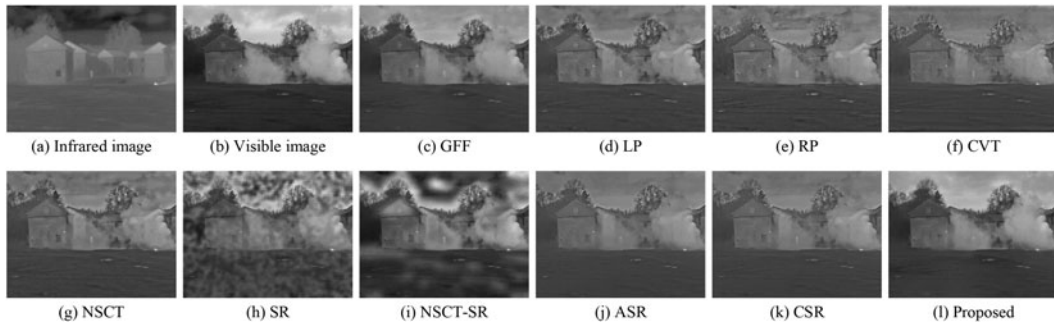


Figure 7. Fusion of “Smog” infrared and visible pair.

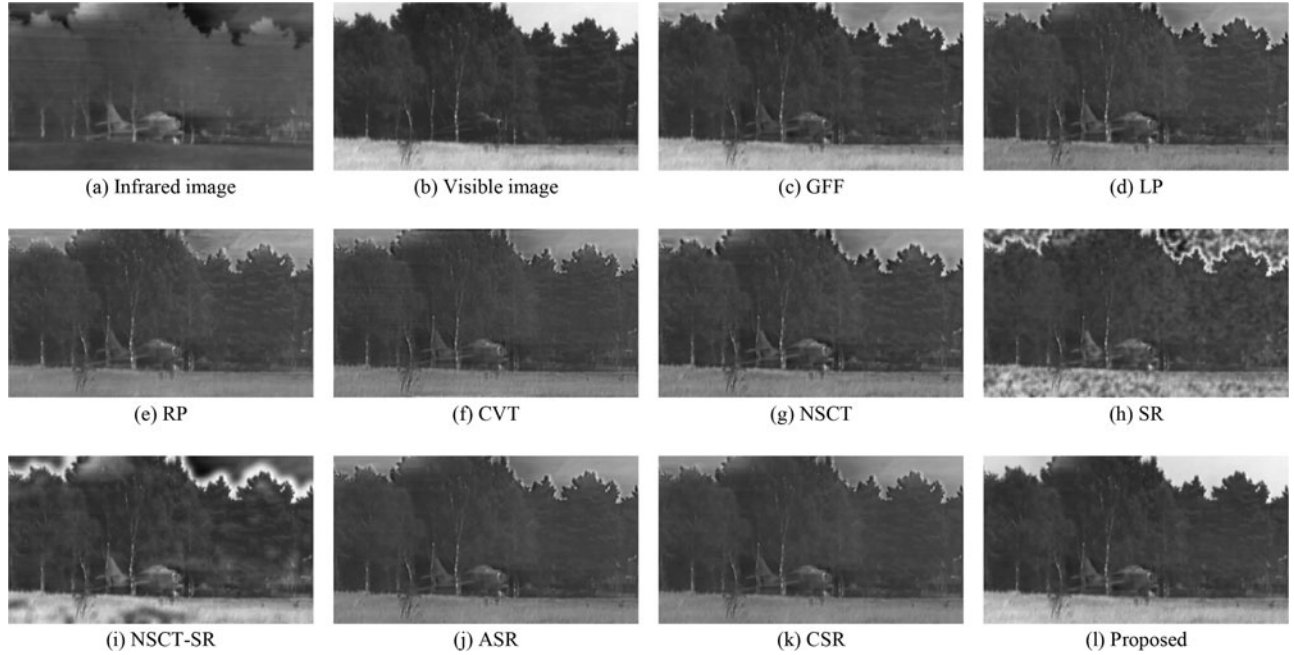


Figure 8. Fusion of “Airplane” infrared and visible pair.

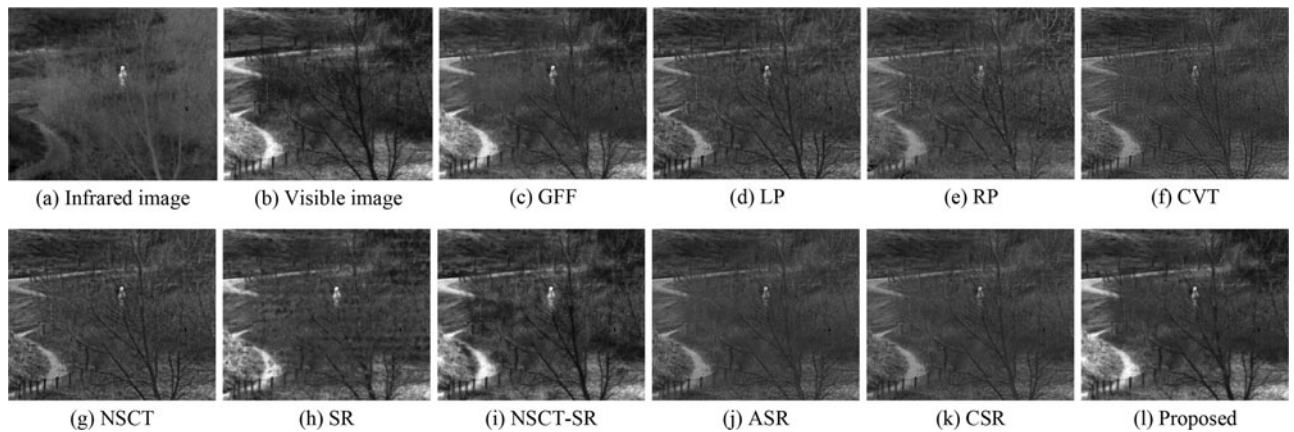


Figure 9. Fusion of “Jungle” infrared and visible pair.

different infrared and visible pairs require different decomposition levels. Our method effectively avoids this problem. It can be seen that our method is tone-balanced and is not significantly biased toward the

tone of a certain source image. In addition, the fusion results obtained by ASR and CSR are generally darker as compared with the result of the proposed method, which is not conducive to human visual observation.

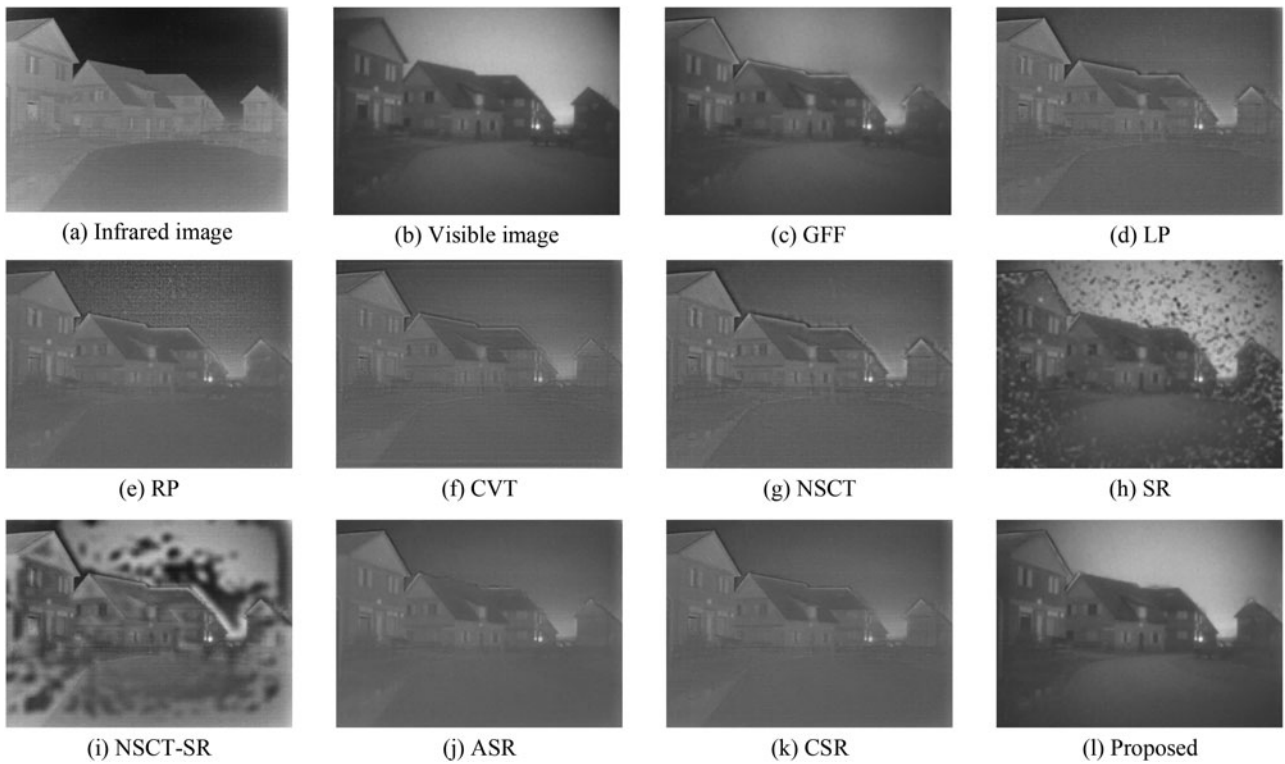


Figure 10. Fusion of “Town” infrared and visible pair.

Table 2. Objective evaluation of the proposed method.

Image	Metric	GFF	LP	RP	CVT	NSCT	SR	NSCT-SR	ASR	CSR	Proposed
Smog	MI	3.3223	2.5548	2.5233	2.4912	2.5044	2.5021	2.2735	2.8996	2.8888	3.5746
	Q^G	<u>0.5237</u>	0.5169	0.4715	0.4269	0.5087	0.3497	0.4723	0.4209	0.4604	0.5491
	FMI	<u>0.4455</u>	0.4104	0.4276	0.3533	0.3969	0.2719	0.3629	0.4056	0.3926	0.4592
	SD	40.9726	30.3130	30.6222	29.6668	29.8180	36.1872	<u>41.8924</u>	28.6578	28.5485	45.7470
Airplane	MI	2.5503	2.2380	2.2595	2.2532	2.2023	2.0277	<u>2.2201</u>	2.4782	2.5283	3.1435
	Q^G	<u>0.5905</u>	0.5482	0.4322	0.4020	0.5288	0.3762	0.5166	0.4903	0.4615	0.6078
	FMI	<u>0.5111</u>	0.4733	0.4260	0.4079	0.4642	0.3163	0.4432	0.4501	0.4202	0.5371
	SD	45.1806	32.5639	31.5790	31.1712	32.2404	38.5999	<u>45.8811</u>	31.3379	30.9745	57.2088
Jungle	MI	2.2234	1.0130	0.9983	0.9800	0.9876	1.9830	<u>2.1018</u>	1.0769	1.0805	3.0603
	Q^G	<u>0.5048</u>	0.4822	0.3725	0.3653	0.4664	0.3694	0.4655	0.4066	0.4453	0.5393
	FMI	<u>0.4982</u>	0.4419	0.4125	0.3866	0.4384	0.4039	0.4378	0.4475	0.4442	0.5171
	SD	33.6538	23.1978	22.3416	20.9122	23.0204	32.2006	37.6797	19.9913	20.1235	36.2073
Town	MI	3.4992	1.0970	0.9838	0.9968	1.0424	4.1491	1.5971	1.2775	1.2828	<u>4.2317</u>
	Q^G	0.3666	0.4348	0.2917	0.3497	0.4280	<u>0.2528</u>	0.3249	0.2856	0.3584	0.3711
	FMI	0.4919	0.4055	0.4205	0.3438	0.3892	0.3734	0.3412	0.4323	0.3868	0.5068
	SD	25.4264	12.7957	13.3216	12.5999	12.5765	32.5447	29.8037	11.1080	11.1877	31.4116

Note: The best performance is shown in bold and the second best performance is underlined.

Compared to GFF, our method can better express the subjects, such as the smog in Figure 7, the aircraft in Figure 8, and the pedestrians in Figure 9. This is because we use sparse coefficients to measure activity levels, which is more robust than GFF that simply measures the saliency maps by investigating spatial characteristics. Since our method can generate high-quality fusion results, the superiority is demonstrated subjectively.

Objective evaluation of the methods

To evaluate the effectiveness of the proposed method, we measure the performance of ten methods in four

metrics, namely MI, Q^G , FMI, and SD. The results are shown in Table 2. As can be seen that our method surpasses all the comparison methods on almost all metrics, especially in MI and FMI, which stand for the mutual information and the feature mutual information. It can be concluded that our method can effectively extract the representative informative from the source infrared and visible pair. Note that although in the “town” image pairs, our method is slightly lower in SD than SR method. However, as we mentioned previously, the SR method causes a lot of halos in the fused image. Similarly, in the “town”

image, our method is slightly lower in Q^G than LP method. However, the fusion result obtained by the LP method has a noticeable unbalanced tone, which is not conducive to human visual perception.

From the experiment results of the above examples, it is demonstrated that the proposed method can extract representative information for fusion. The fusion results are rich in detail and gradient information. Thus, the superiority is proved in both subjective aspect and objective aspect.

Conclusion

In this paper, an effective infrared and visible images fusion method is proposed. In the proposed method, the weight map is measured according to the L_0 -norm of the SR coefficients. Then, we use guided filter to optimize the weight map in accordance with the different characteristic of infrared and visible images. With the help of the guided filter, we solve two problems existed in the initial weight map, namely the binary problem and the inconsistency problem. Subsequently, the optimized weight map is employed to reconstruct the two-scale layers. Finally, the fused image can be obtained by combining the two-scale layers. By performing extensive experiments, the superiority of the proposed method as compared with other popular approaches is validated.

Funding

This work was partially supported by the National Natural Science Foundation of China (nos. 61711540303, 61701327, and 61601266), Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (no. MJUKF-IPIC201803), and China Postdoctoral Science Foundation Funded Project under Grant 2018M640916.

ORCID

Qilei Li  <http://orcid.org/0000-0002-9675-9016>

References

- Agarwal, S., Awan, A., & Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11), 1475–1490. doi:10.1109/TPAMI.2004.108
- Agarwal, S., & Roth, D. (2002). *Learning a sparse representation for object detection*. In European Conference on Computer Vision (pp. 113–127), Berlin, Heidelberg.
- Burt, P. J., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *Readings in Computer Vision*, 31(4), 532–540.
- Dai, X., Zhao, J., & Li, D. (2018). Effective detection by fusing visible and infrared images of targets for unmanned surface vehicles. *Automatika*, 59(3-4), 323–330. doi:10.1080/00051144.2018.1541150
- Haghighat, M. B. A., Aghagolzadeh, A., & Seyedarabi, H. (2011). A non-reference image fusion metric based on mutual information of image features. *Computers & Electrical Engineering*, 37(5), 744–756.
- He, K., Sun, J., & Tang, X. (2010). Guided image filtering. In *European conference on computer vision* (pp. 1–14). Berlin, Heidelberg: Springer.
- Jeon, H., Lee, J., & Sohn, K. (2018). Artificial intelligence for traffic signal control based solely on video images. *Journal of Intelligent Transportation Systems*, 22(5), 433–445. doi:10.1080/15472450.2017.1394192
- Lewis, J. J., O'Callaghan, R. J., Nikolov, S. G., Bull, D. R., & Canagarajah, N. (2007). Pixel- and region-based image fusion with complex wavelets. *Information Fusion*, 8(2), 119–130. doi:10.1016/j.inffus.2005.09.006
- Li, H., Manjunath, B. S., & Mitra, S. K. (1995). Multisensor image fusion using the wavelet transform. *Graphical models and image processing*, 57(3), 235–245.
- Li, Q., Yang, X., Wu, W., Liu, K., & Jeon, G. (2018). Multi-focus image fusion method for vision sensor systems via dictionary learning with guided filter. *Sensors*, 18(7), 2143.
- Li, Q., Yang, X., Wu, W., Liu, K., & Jeon, G. (2019). Pansharpening multispectral remote-sensing images with guided filter for monitoring impact of human behavior on environment. *Concurrency and Computation: Practice and Experience*, Published online, e5074.
- Li, S., Kang, X., & Hu, J. (2013). Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7), 2864–2875. doi:10.1109/TIP.2013.2244222
- Li, S., Yin, H., & Fang, L. (2013). Remote sensing image fusion via sparse representations over learned dictionaries. *IEEE Transactions on Geoscience and Remote Sensing*, 51(9), 4779–4789. doi:10.1109/TGRS.2012.2230332
- Li, Y., Khoshelham, K., Sarvi, M., & Haghani, M. (2019). Direct generation of level of service maps from images using convolutional and long short-term memory networks. *Journal of Intelligent Transportation Systems*, 23(3), 300–308. doi:10.1080/15472450.2018.1563865
- Liu, Y., Chen, X., Cheng, J., Peng, H., & Wang, Z. (2018). Infrared and visible image fusion with convolutional neural networks. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(03), 1850018. doi:10.1142/S0219691318500182
- Liu, Y., Chen, X., Peng, H., & Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36, 191–207. doi:10.1016/j.inffus.2016.12.001
- Liu, Y., Chen, X., Ward, R. K., & Wang, Z. J. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*, 23(12), 1882–1886. doi:10.1109/LSP.2016.2618776
- Liu, Y., Liu, S., & Wang, Z. (2015). A general framework for image fusion based on multi-scale transform and

- sparse representation. *Information Fusion*, 24, 147–164. doi:10.1016/j.inffus.2014.09.004
- Liu, Y., & Wang, Z. (2015). Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Processing*, 9(5), 347–357. doi:10.1049/iet-ipr.2014.0311
- Ma, J., Ma, Y., & Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, 45, 153–178. doi:10.1016/j.inffus.2018.02.004
- Ma, J., Yu, W., Liang, P., Li, C., & Jiang, J. (2018). FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion*, 48, 11–26. doi:10.1016/j.inffus.2018.09.004
- Miao, Q., & Wang, B. (2005). A novel adaptive multi-focus image fusion algorithm based on PCNN and sharpness. In *Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense IV*, Vol. 5778. International Society for Optics and Photonics.
- Nejati, M., Samavi, S., & Shirani, S. (2015). Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*, 25, 72–84. doi:10.1016/j.inffus.2014.10.004
- Nencini, F., Garzelli, A., Baronti, S., & Alparone, L. (2007). Remote sensing image fusion using the curvelet transform. *Information Fusion*, 8(2), 143–156. doi:10.1016/j.inffus.2006.02.001
- Petrovic, V. S., & Xydeas, C. S. (2004). Gradient-based multiresolution image fusion. *IEEE Transactions on Image Processing*, 13(2), 228–237.
- Qu, G., Zhang, D., & Yan, P. (2002). Information measure for performance of image fusion. *Electronics Letters*, 38(7), 313–315. doi:10.1049/el:20020212
- Toet, A. (1989). Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters*, 9(4), 245–253. doi:10.1016/0167-8655(89)90003-2
- Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., & Ma, Y. (2012). Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(2), 372–386.
- Wang, M., Zhou, S., Yang, Z., Liu, Z., & Ren, S. (2019). Image fusion based on wavelet transform and gray-level features. *Journal of Modern Optics*, 66(1), 77–86. doi:10.1080/09500340.2018.1512668
- Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 31(2), 210–227.
- Wu, W., Yang, X., Liu, K., Liu, Y., Yan, B., & Hua, H. (2016). A new framework for remote sensing image super-resolution: Sparse representation-based method by processing dictionaries with multi-type features. *Journal of Systems Architecture*, 64, 63–75. doi:10.1016/j.sysarc.2015.11.005
- Xydeas, C. S., & Petrovic, V. (2000). Objective image fusion performance measure. *Military Technical Courier*, 56(2), 181–193.
- Yang, B., & Li, S. (2010). Multifocus image fusion and restoration with sparse representation. *IEEE Transactions on Instrumentation & Measurement*, 59(4), 884–892.
- Yang, X., Wu, W., Liu, K., Chen, W., Zhang, P., & Zhou, Z. (2017). Multi-sensor image super-resolution with fuzzy cluster by using multi-scale and multi-view sparse coding for infrared image. *Multimedia Tools and Applications*, 76(23), 24871–24902.
- Yang, X., Wu, W., Liu, K., Chen, W., & Zhou, Z. (2018). Multiple dictionary pairs learning and sparse representation-based infrared image super-resolution with improved fuzzy clustering. *Soft Computing*, 22(5), 1385–1398. doi:10.1007/s00500-017-2812-3
- Zhang, Q., & Guo, B. L. (2009). Multifocus image fusion using the nonsubsampling contourlet transform. *Signal Processing*, 89(7), 1334–1346. doi:10.1016/j.sigpro.2009.01.012
- Zhang, Y., Wang, X., Xie, X., & Li, Y. (2018). Salient object detection via recursive sparse representation. *Remote Sensing*, 10(4), 652. doi:10.3390/rs10040652
- Zhou, Y., & Omar, M. (2009). Pixel-level fusion for infrared and visible acquisitions. *International Journal of Optomechatronics*, 3(1), 41–53. doi:10.1080/15599610902717835
- Zhou, Z., Wang, B., Li, S., & Dong, M. (2016). Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters. *Information Fusion*, 30, 15–26. doi:10.1016/j.inffus.2015.11.003