Signal Processing 177 (2020) 107734

Contents lists available at ScienceDirect

Signal Processing

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

Bayesian fusion for infrared and visible images

Zixiang Zhao, Shuang Xu, Chunxia Zhang, Junmin Liu, Jiangshe Zhang*

School of Mathematics and Statistics, Xi'an Jiaotong University, China

ARTICLE INFO

Article history:
Received 28 January 2020
Revised 27 June 2020
Accepted 31 July 2020
Available online 11 August 2020

Keywords:
Image fusion
Hierarchical Bayesian model
Total-variation penalty
EM algorithm

ABSTRACT

Infrared and visible image fusion has been a hot issue in image fusion. In this task, a fused image containing both the gradient and detailed texture information of visible images as well as the thermal radiation and highlighting targets of infrared images is expected to be obtained. In this paper, a novel Bayesian fusion model is established for infrared and visible images. In our model, the image fusion task is cast into a regression problem. To measure the uncertainty in a better manner, we formulate the model in a hierarchical Bayesian manner. Aiming at making the fused image satisfy the human visual system, the model incorporates the total-variation (TV) penalty. Subsequently, the model is efficiently inferred by the expectation-maximization (EM) algorithm, where the fusion weight can be automatically inferred and adaptive to the source images. The performance of our algorithm is investigated and compared with several state-of-the-art approaches on TNO and NIR image fusion datasets. In comparison with other methods, the novel model can generate better fused images with highlighting targets and rich texture details, which may potentially improve the reliability of the target automatic detection and recognition system.

1. Introduction

Image fusion, an information-enhanced image processing technique to produce a robust or informative image [1], is a hot issue in computer vision today. There is a wide range of applications related to image fusion in pattern recognition [2], medical imaging [3], remote sensing [4–6] and modern military [7] whose research scenarios require to fuse two or more source images in same scene [8].

The fusion of visible and infrared images (IVIF) can greatly improve the perception ability of the human visual system in target detection and recognition [9]. As we know, a visible image has rich appearance information, and the features such as texture and detail information are often unclear in the corresponding infrared image. In contrast, an infrared image mainly reflects the heat radiation emitted by objects, which is less affected by illumination changes or artifacts and overcome the obstacles to target detection at night. However, the spatial resolution of infrared images is typically lower than that of visible images. Consequently, fusing thermal radiation and texture detail information into an image facilitates automatic detection and accurate positioning of targets [10].

Broadly speaking, the current algorithms for fusing visible and infrared images can be divided into four categories: multi-scale transformation, sparse representation, subspace and saliency methods [1]. The multi-scale transformation-based methods [11–14], in general, decompose source images into multiple levels and then fuse images from the same level of the decomposed layers with specific fusion strategies. Finally, the fused image is recovered by incorporating the fused layers. The second category is sparse representation-based methods [15–17], which assume that the natural image is a sparse linear combination of a set of atoms from an over-complete dictionary [18], and fused images can be recovered by the merging coefficients cooperating with a given dictionary. The third category is the subspace learning-based methods [19–21], which aims to project high-dimensional input images into low-dimensional subspaces to capture the intrinsic feature of the original image. The fourth category is saliency-based methods [22–24]. Based on the prior knowledge that humans usually pay more attention to the saliency objects rather than surrounding areas, they fuse images by maintaining the integrity of the salient target areas. Although the above-mentioned methods are relatively effective in accomplishing fusion tasks, they still have some shortcomings. Different decomposers in the transformation-based method have dissimilar effects, and they may not meet our demands in some circumstances. The sparse representation methods need to adjust the sparse intensity and control the sparseness, that is, each image needs to be fine-tuned and it is hence difficult to generalize to other scenes. In the subspace methods, information may be lost when they apply dimensionality reduction techniques to extract features. As for the saliency-based methods, manually designed saliency detectors are often suitable for specific parts of samples, and the generalization performance is relatively poor.

* Corresponding author.
E-mail address: jszhang@mail.xjtu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.sigpro.2020.107734
0165-1684/

The Bayesian method, by defining an appropriate prior to regularize the objective function, is widely used in signal processing tasks. For the multi-band image fusion (MBIF) issues, some researchers [25,26] construct Bayesian estimators by imposing an appropriate prior distribution to fuse a high spatial resolution panchromatic image and a low spatial resolution multispectral image. Similarly, Zhang et al. [27] put forward a robust noise-resistant MBIF fusion method by a statistical framework in the wavelet domain. Wei et al. [28] prove that the maximum likelihood estimator in an MBIF task can be solved by the explicit solution of the Sylvester equation without the iterative scheme. The model can be generalized to Bayesian fusion methods by taking the prior knowledge into account. Besides, in the area of signal processing, a variational Bayesian (VB) model is exploited in [29] to solve the issue of randomly missing output measurement in industrial processes. Likewise, Liu and Yang [30] propose a robust VB identification algorithm in industrial processes by assuming that noise is sampled from a mixture of Laplacians. To summarize, Bayesian methods integrate prior knowledge and the information provided by data (i.e., likelihood) via the Bayesian formula. More importantly, they output the posterior distributions, rather than only point estimations, of our interested parameters. Because the uncertainty can be measured and handled better, Bayesian methods greatly enhance the effectiveness of the inference process and have had a wide range of applications in various scenarios.

To the best of our knowledge, despite the success of Bayesian methods in multi-band fusion, the IVIF field still lacks the utilization of the Bayesian method. Therefore, in this paper, in order to combine the prior information embodied in source images, and to automatically infer the fusion weights of each pixel, we present a novel Bayesian model for fusing infrared and visible images. Our contributions can be summarized as follows.

(1) In our model, the image fusion task is cast into a regression problem. To measure the uncertainty, we formulate the problem in a hierarchical Bayesian manner. Besides, to make the fused image satisfy the human visual system, the model incorporates the TV penalty. Notably, the fusion weights of each pixel is adaptive to the input sources without manual adjustment in our model.
(2) This model is efficiently inferred by the EM algorithm. Cooperating with the half-quadratic splitting algorithm, the optimization problem can be decomposed into a least-squares issue, an ℓ_1-norm penalized regression issue and a deconvolution issue, with an explicit solution in each sub-problem.
(3) To the best of our knowledge, the current fusion methods [10,23,31,32] are only verified on the TNO dataset. In contrast, we test our algorithm on TNO and three scenes of NIR image fusion datasets including various foreground targets and illumination conditions. Compared with the state-of-the-art methods, our method can generate fused images with highlighting targets and rich texture details, which can improve the reliability of the target automatic detection and recognition system.

The rest paper is organized as follows. In Section 2, we introduced the Bayesian fusion method. In Section 3, some experiments are conducted to investigate and compare the proposed method with some state-of-the-art techniques. Finally, some conclusions are drawn in Section 4.

2. Bayesian fusion model

In this section, we present a novel Bayesian fusion model for infrared and visible images. Then, this model is efficiently inferred by the EM algorithm [33].

2.1. Model formulation

2.1.1. Optimization model

Given a pair of pre-registered infrared and visible images $U, V \in \mathbb{R}^{h\times w}$, image fusion technique aims at obtaining an informative image $I \in \mathbb{R}^{h\times w}$ from U and V.

It is well-known that visible images satisfy human visual perception, while they are significantly sensitive to disturbances, such as poor illumination, fog and so on. In contrast, infrared images are robust to these disturbances but may lose part of informative textures. In order to preserve the general profiles of two images, we minimize the difference between fused and source images, that is

$$\min_I f(U,I) + \phi g(V,I),$$

where f, g are loss functions and ϕ is a tuning parameter. Typically, we assume the difference is measured by ℓ_1-norm. Thus, the problem can be rewritten as

$$\min_I ||I-U||_1 + \phi||I-V||_1. \tag{1}$$

Let $X = I - V$ and $Y = U - V$, then we have

$$\min_X ||X-Y||_1 + \phi||X||_1. \tag{2}$$

2.1.2. Hierarchical Bayesian manner

Essentially, Eq. (2) corresponds to a linear regression model

$$Y = X + E,$$

where E denotes a Laplacian noise and X is governed by Laplacian distribution. By reformulating this problem in the Bayesian fashion, the conditional distribution of Y given X is

$$\begin{aligned} p(Y|X) &= \text{Laplace}(Y|X,\lambda_y) \\ &= \prod_{i,j} \frac{1}{2\lambda_y} \exp\left(-\frac{|y_{ij}-x_{ij}|}{\lambda_y}\right), \end{aligned}$$

and the prior distribution of X is

$$\begin{aligned} p(X) &= \text{Laplace}(X|0,\tau_x) \\ &= \prod_{i,j} \frac{1}{2\tau_x} \exp\left(-\frac{|x_{ij}|}{\tau_x}\right). \end{aligned}$$

where λ_y and τ_x are the scale parameters of Laplacian distributions $p(Y|X)$ and $p(X)$, respectively.

To avert from ℓ_1-norm optimization, we reformulate a Laplacian distribution as Gaussian scale mixtures with exponential distributed prior to the variance, that is,

$$\begin{aligned} &\text{Laplace}(\xi|\mu,\sqrt{\lambda/2}) \\ &= \frac{1}{2}\sqrt{\frac{2}{\lambda}} \exp\left(-\sqrt{\frac{2}{\lambda}}|\xi-\mu|\right) \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi a}} \exp\left(-\frac{(\xi-\mu)^2}{2a}\right) \frac{1}{\lambda} \exp\left(-\frac{a}{\lambda}\right) da \\ &= \int_0^\infty \mathcal{N}(\xi|\mu,a)\text{Exp}(a|\lambda) da \end{aligned} \tag{3}$$

where $\mathcal{N}(\xi|\mu,a)$ denotes a Gaussian distribution with mean μ and variance a, and $\text{Exp}(a|\lambda)$ denotes an exponential distribution with scale parameter λ. According to Eq. (3), the original model of $p(Y|X)$ and $p(X)$ can be rewritten in the hierarchical Bayesian manner, that is,

$$\begin{cases} y_{ij}|x_{ij}, a_{ij} \sim \mathcal{N}(y_{ij}|x_{ij}, a_{ij}) \\ a_{ij} \sim \text{Exp}(a_{ij}|\lambda) \\ x_{ij}|b_{ij} \sim \mathcal{N}(x_{ij}|0, b_{ij}) \\ b_{ij} \sim \text{Exp}(b_{ij}|\tau), \end{cases}$$

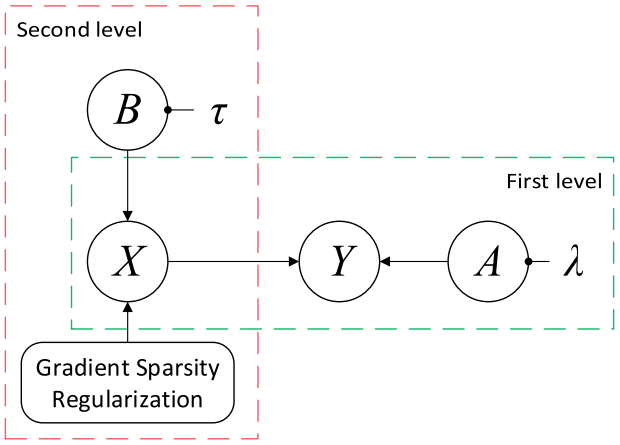

Fig. 1. Illustration of our Bayesian graph model.

for all $i = 1, \ldots, h$ and $j = 1, \ldots, w$, where h and w mean the height and the width of the input image. In what follows, we use matrices A and B to denote the collection of all latent variables a_{ij} and b_{ij}, respectively.

2.1.3. The TV penalty

Besides modeling the general profiles, the image textures should be taken into consideration so as to make fused image satisfy the human visual perception. As discussed above, there is plenty of high-frequency information in visible images, but the corresponding areas often cannot be observed in infrared images. In order to preserve the edge information of visible images, we regularize the fused image in gradient domain with a gradient sparsity regularizer expressed as

$$h(X) = \frac{1}{2}\lambda_g||\nabla I - \nabla V||_1 = \frac{1}{2}\lambda_g||\nabla X||_1, \tag{4}$$

where λ_g is a hyper-parameter controlling the strength of regularization, ∇ denotes the gradient operator. This regularizer makes the fused image have similar textures to the visible image. Besides, the gradient sparsity regularizer is calculated by

$$||\nabla X||_1 = ||\nabla_h X||_1 + ||\nabla_v X||_1, \tag{5}$$

where the ∇_h and ∇_v denote the gradient operator in horizontal and vertical directions, respectively. As a result, the gradient sparsity regularizer can be regarded as the TV penalty with respect to X.

By combining general profiles and gradients modeling, Fig. 1 displays the graphical expression of our hierarchical Bayesian model. Specifically, in the first level, A is a latent variable, while Y and X are observed and unknown variables, respectively. In the second level, X and B are observed and latent variables, respectively. And λ and τ are hyper-parameters. By ignoring the constant not depending on X, the log-likelihood of the model can be expressed as

$$\begin{aligned}\ell(X) &= \log p(Y, X) - h(X)\\ &= -\sum_{i,j}\left[\frac{(x_{ij} - y_{ij})^2}{2a_{ij}} + \frac{x_{ij}^2}{2b_{ij}}\right] - \frac{\lambda_g}{2}||\nabla X||_1.\end{aligned}$$

In next subsection, we will discuss how to infer this model.

2.2. Model inference

As is well-known, the EM algorithm is an effective tool to maximize the log-likelihood function of a problem which involves some latent variables. In detail, we firstly initialize the unknown variable X. Then, in E-step, it calculates the expectation of log-likelihood function with respect to $p(A, B|X^{(t)}, Y)$, which is often referred to as the so-called Q-function,

$$Q(X|X^{(t)}) = \mathbb{E}_{A,B|X^{(t)},Y}[\ell(X)].$$

In M-step, we find X to maximize the Q-function, i.e.,

$$X^{(t+1)} = \arg\max_X Q(X|X^{(t)}).$$

E-step: In order to obtain the Q-function in our model, $\mathbb{E}_{a_{ij}|x_{ij}^{(t)},y_{ij}}[1/a_{ij}]$ and $\mathbb{E}_{b_{ij}|x_{ij}^{(t)}}[1/b_{ij}]$ should be computed. For convenience, we had better compute the posterior distribution for $\tilde{a}_{ij} \equiv 1/a_{ij}$ and $\tilde{b}_{ij} \equiv 1/b_{ij}$. It has been assumed that the prior distribution of a_{ij} is $\mathrm{Exp}(a_{ij}|\lambda)$, so $\tilde{a}_{ij}$ is governed by an *inverse gamma distribution* with shape parameter of 1 and scale parameter of $1/\lambda$. And the probability density function of $\tilde{a}$ is given by

$$p(\tilde{a}) = \frac{1}{\lambda}\tilde{a}^{-2}\exp\left(-\frac{1}{\lambda\tilde{a}}\right).$$

According to the Bayesian formula, we have

$$\begin{aligned}&\log p(\tilde{a}_{ij}|y_{ij}, x_{ij})\\ &= \log p(y_{ij}|x_{ij}, a_{ij}) + \log p(\tilde{a}_{ij})\\ &= -\frac{\log a_{ij}}{2} - \frac{(y_{ij} - x_{ij})^2}{2a_{ij}} - 2\log\tilde{a}_{ij} - \frac{1}{\lambda\tilde{a}_{ij}} + \text{constant}\\ &= \frac{\log\tilde{a}_{ij}}{2} - \frac{\tilde{a}_{ij}(y_{ij} - x_{ij})^2}{2} - 2\log\tilde{a}_{ij} - \frac{1}{\lambda\tilde{a}_{ij}} + \text{constant}\\ &= -\frac{3}{2}\log\tilde{a}_{ij} - \frac{\tilde{a}_{ij}(y_{ij} - x_{ij})^2}{2} - \frac{1}{\lambda\tilde{a}_{ij}} + \text{constant}.\end{aligned}$$

Therefore, the posterior of $\tilde{a}_{ij}$ is an *inverse Gaussian distribution*, that is,

$$p(\tilde{a}_{ij}|y_{ij}, x_{ij}) = \mathcal{IN}(\tilde{a}_{ij}|\alpha_{ij}, \tilde{\lambda}),$$

where $\alpha_{ij} = \sqrt{2(y_{ij} - x_{ij})^2/\lambda}$ and $\tilde{\lambda} = 2/\lambda$. As for $\tilde{b}_{ij}$, we can compute its posterior in the same way,

$$\begin{aligned}&\log p(\tilde{b}_{ij}|x_{ij})\\ &= \log p(x_{ij}|b_{ij}) + \log p(\tilde{b}_{ij})\\ &= -\frac{\log b_{ij}}{2} - \frac{x_{ij}^2}{2b_{ij}} - 2\log\tilde{b}_{ij} - \frac{1}{\tau\tilde{b}_{ij}} + \text{constant}\\ &= -\frac{3}{2}\log\tilde{b}_{ij} - \frac{\tilde{b}_{ij}x_{ij}^2}{2} - \frac{1}{\tau\tilde{b}_{ij}} + \text{constant}.\end{aligned}$$

Similarly, the posterior of $\tilde{b}_{ij}$ is

$$p(\tilde{b}_{ij}|x_{ij}) = \mathcal{IN}(\tilde{b}_{ij}|\beta_{ij}, \tilde{\tau}),$$

where $\beta_{ij} = \sqrt{2x_{ij}^2/\tau}$ and $\tilde{\tau} = 2/\tau$. Note that the expectation of inverse Gaussian distribution is its location parameter. Thus, we have

$$\mathbb{E}_{a_{ij}|x_{ij}^{(t)},y_{ij}}\left[\frac{1}{a_{ij}}\right] = \alpha_{ij} = \sqrt{\frac{2(y_{ij} - x_{ij}^{(t)})^2}{\lambda}}, \tag{6}$$

$$\mathbb{E}_{b_{ij}|x_{ij}^{(t)}}\left[\frac{1}{b_{ij}}\right] = \sqrt{\frac{2[x_{ij}^{(t)}]^2}{\tau}}. \tag{7}$$

Thereafter, in E-step, the Q-function is given by

$$\begin{aligned}Q &= -\sum_{i,j}\left[\frac{1}{2}\alpha_{ij}(x_{ij} - y_{ij})^2 + \frac{1}{2}\beta_{ij}x_{ij}^2\right] - \frac{1}{2}\lambda_g||\nabla X||_1\\ &= -||W_1 \odot (X - Y)||_2^2 - ||W_2 \odot X||_2^2 - \lambda_g||\nabla X||_1.\end{aligned}$$

where the symbol $\odot$ means element-wise multiplication, and the (i, j)th entries of W_1 and W_2 are $\sqrt{\alpha_{ij}}$ and $\sqrt{\beta_{ij}}$, respectively. In particular, W_1 and W_2 are related to the pixel fusion weights, which can be adaptive to the input sources without manual-designed parameters.

M-step: Here, we need to minimize the negative Q-function with respect to X. The half-quadratic splitting algorithm is employed to deal with this problem, i.e.,

$$\min_{X,F,H} ||W_1 \odot (X-Y)||_2^2 + ||W_2 \odot X||_2^2 + \lambda_g ||F||_1,$$
$$\text{s.t.} \quad F = \nabla H, H = X. \tag{8}$$

It can be further cast into the following unconstraint optimization problem,

$$\min_{X,F,H} ||W_1 \odot (X-Y)||_2^2 + ||W_2 \odot X||_2^2 + \lambda_g ||F||_1 + \frac{\rho}{2}\left(||F-\nabla H||_2^2 + ||H-X||_2^2\right). \tag{9}$$

The unknown variables X, F, H can be solved iteratively in the coordinate descent fashion.

Update X: It is a least squares issue,

$$\min_X ||W_1 \odot (X-Y)||_2^2 + ||W_2 \odot X||_2^2 + \frac{\rho}{2}||H-X||_2^2.$$

The solution of X is

$$X = (2W_1^2 \odot Y + \rho H) \oslash (2W_1^2 + 2W_2^2 + \rho), \tag{10}$$

where the symbol $\oslash$ means the element-wise division.

Update F: It is an ℓ_1-norm penalized regression issue,

$$\min_F \lambda ||F||_1 + \frac{\rho}{2}||F-\nabla H||_2^2.$$

The solution is

$$F = S(\nabla H, \lambda/\rho), \tag{11}$$

where $S(x, \gamma) = \text{sign}(x)\max(|x| - \gamma, 0)$.

Update H: It is a deconvolution issue,

$$\min_H ||H-X||_2^2 + ||F-\nabla H||_2^2.$$

It can be efficiently solved by the fast Fourier transform (fft) and inverse fft (ifft) operators, and the solution is

$$H = \text{ifft}\left\{\frac{\text{fft}(X) + \overline{\text{fft}(k_h)} \odot \text{fft}(F)}{1 + \overline{\text{fft}(k_h)} \odot \text{fft}(k_h)}\right\}, \tag{12}$$

where $\bar{x}$ denotes the complex conjugation.

In order to make model more flexible, the hyper-parameters λ and τ are automatically updated. According to empirical Bayes, we have

$$\lambda = \frac{1}{hw}\sum_{i,j}\mathbb{E}[a_{ij}] = \frac{1}{hw}\sum_{i,j}\left(\frac{1}{\alpha_{ij}} + \frac{1}{\tilde{\lambda}}\right) \tag{13}$$

and

$$\tau = \frac{1}{hw}\sum_{i,j}\mathbb{E}[b_{ij}] = \frac{1}{hw}\sum_{i,j}\left(\frac{1}{\beta_{ij}} + \frac{1}{\tilde{\tau}}\right). \tag{14}$$

Algorithm 1 summarizes the workflow of our proposed model, where E-step and M-step alternate with each other until the maximum iteration number T^{out} is reached. Since there is no analytic solution in M-step, we maximize Q-function by updating (X, H, F) T^{in} times. The number of inner and outer loop iterations (i.e., T^{in} and T^{out}), the strength of gradient penalty (i.e., λ_g) and ℓ_2-norm penalty (i.e., ρ) in Eq. (9) can be determined with a validation dataset. More details are shown in Section 3.2.

Algorithm 1 Bayesian fusion

Input:
Infrared image U, Visible image V, Maximum iterations of outer and inner loops T^{out} and T^{in}.

Output:
Fused image I.

$Y = U - V$; Initialize $W_1, W_2 = 1, H, F = 0, \lambda, \tau = 1$;
for $t = 1, \ldots, T^{\text{out}}$ **do**
  % (M-step)
  for $j = 1, \ldots, T^{\text{in}}$ **do**
    Update X, F, H with Eqs. (10), (11) and (12), respectively.
  end for
  % (E-step)
  Evaluate expectations by Eqs. (6) and (7).
  Update hyper-parameters λ, τ by Eqs. (13) and (14).
end for
$I = X + V$.

Table 1
Dataset employed in this paper.

	Dataset(# images)	Illumination
Validation	FLIR(20)	Daylight&Nightlight
Test	TNO(20)	Nightlight
	NIR-Oldbuilding(51)	Daylight
	NIR-Street(50)	Daylight
	NIR-Urban(58)	Daylight

3. Experiments

We establish a battery of experiments to show the superiority of our proposed method in this section. All experiments are conducted with MATLAB R2019b on a computer with Intel Core i7-9750H CPU@2.60 GHz.

3.1. Experimental data and metrics

3.1.1. Datasets

In this experiment, we randomly choose 20 images in FLIR dataset [34] as a validation dataset and test the algorithms on 20 images of TNO image fusion dataset [35] and three scenes in NIR Scene dataset [36] ("Oldbuilding", "Street" and "Urban"). The number of image pairs and the illumination condition of each dataset are exhibited in Table 1. In TNO and FLIR datasets, the interesting objects are hard to be observed in visible images, as it was shot at night. In contrast, they are salient in infrared images, but without textures. Considering that the NIR image dataset was obtained in daylight, we study whether the fused image can have more detailed and highlight information. Since there is no reference or ground truth fusion image in each dataset, we will introduce some no-reference image quality assessment metrics to measure the performance of fusion algorithms in the next section.

3.1.2. Metrics

We employ Entropy (EN) [37], Mutual information (MI) [38], $Q^{AB/F}$ [38], Standard deviation (SD) [39] and Structure similarity index measure (SSIM) [40] metrics to evaluate the fusion performance of each algorithm. EN and SD measure how much information is contained in an image. $Q^{AB/F}$ reflects the edge information preserved in the fusion image. MI measures the agreement between source images and the fusion image, and SSIM reports the consistency in the light of structural similarities between fusion and source images. The larger the metric values are, the better a fused image is. Please refer to [1] to see more details on these metrics.

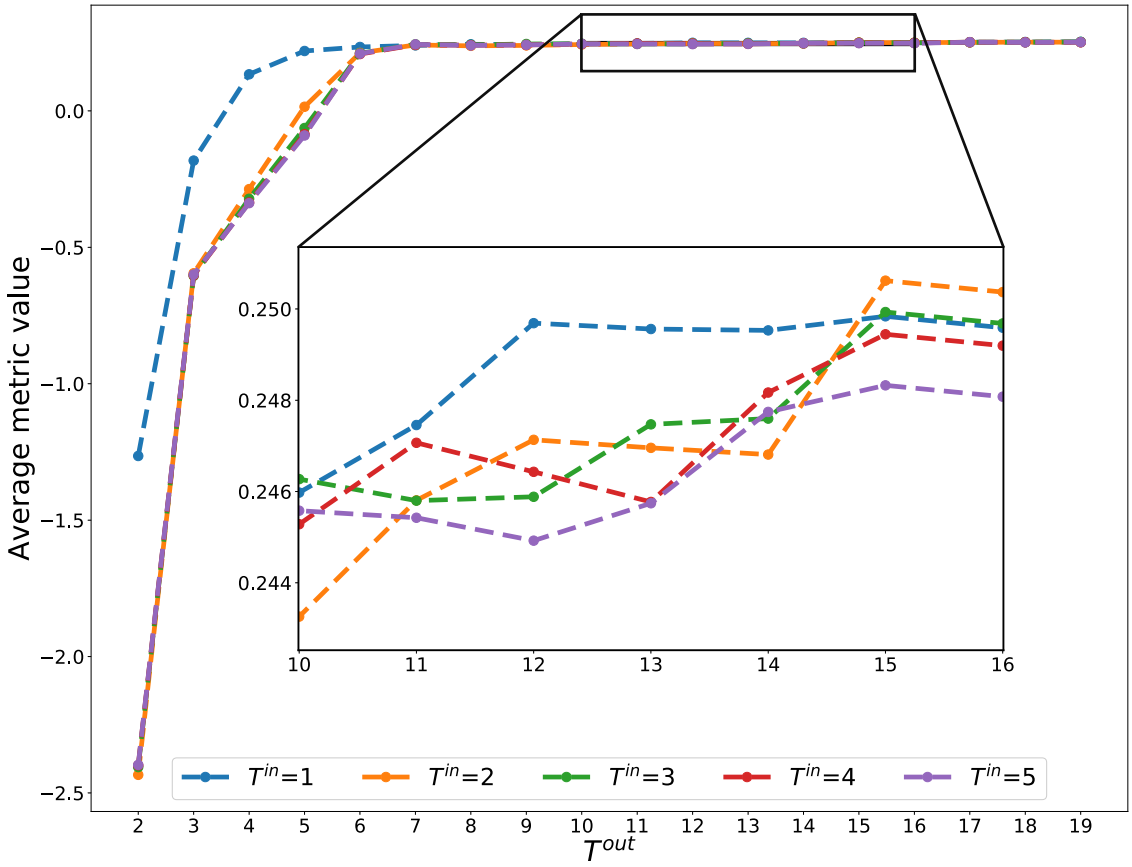

Fig. 2. The AMV of fusion results in validation set under the combination of $T^{\text{in}} = 1, \ldots, 5$ and $T^{\text{out}} = 1, \ldots, 20$. Each dashed line represents the change curve of AMV with respect to T^{out} for a certain T^{in}.

3.2. Hyperparameters and the implementation details

The hyperparameters can be determined by grid search. We utilize FLIR dataset as the validation set and find out the configuration which achieves a satisfactory fusion result on the validation set. Because it is difficult to make a consistent decision on all the five metrics, the average metric value (AMV), which is defined by the mean of five normalized metric values, is employed to quantitatively measure the fusion effect.

Take the determination of $\{T^{\text{in}}, T^{\text{out}}\}$ as an example. We show the AMV curves under the combination of $T^{\text{in}} = 1, \ldots, 5$ and $T^{\text{out}} = 1, \ldots, 20$ in Fig. 2. In Fig. 2, the different configurations of hyperparameters do not greatly affect the fusion performance, showing the robustness of our method in the selection of hyperparameters when $T^{\text{in}} \geq 2$ and $T^{\text{out}} \geq 7$. Then, we check the amplification on local details of curves and the configuration $\{T^{\text{in}} = 2, T^{\text{out}} = 15\}$ achieves the highest AMV value. Ultimately, to balance the computation time and fusion performance, T^{in} is set to 2 and T^{out} is set to 15, respectively. By the same token, we set $\lambda_g = 0.05$ and $\rho = 0.001$.

3.3. Ablation experiment

We conduct the following two ablation experiments to analyze the role of two critical ingredients in our Bayesian fusion model.

TV penalty (Exp. 1) Firstly, we verify the role of TV penalty. We set λ_g in Eqs. (4) and (8) to 0, i.e., eliminating the TV penalty in the optimization model. Then the fusion results on the validation set are measure by the above metrics.

Bayesian inference (Exp. 2) Another merit of our model lies in that Bayesian inference can adaptively calculate the fusion weights.

Table 2
Quantitative results of ablation experiments. The largest value is shown in bold.

	EN	MI	Qabf	SD	SSIM
BF	**6.584**	**2.807**	**0.365**	**25.441**	**0.882**
Exp. 1	6.566	2.748	0.338	25.300	0.872
Exp. 2	6.570	2.211	0.362	23.727	0.863

To illustrate the importance of Bayesian inference, we rewrite the optimization model of (2) as

$$\min_X ||X - Y||_1 + \phi_M ||X||_1, \tag{15}$$

where ϕ_M is set to 0.5 by perform a grid search on the validation set. It is worth noting that the role of ϕ_M is the same as that of W_1 as well as W_2 (in Eq. (8)). Both ϕ_M and W_1 (or W_2) are used to balance $||X - Y||_1$ and $||X||_1$. However, W_1 and W_2 are automatically determined by the EM algorithm. Here, we have to select a proper value for ϕ_M manually. The model (15) can be solved by the alternating direction method of multipliers (ADMM) algorithm [41].

The fusion results among the ablation experiments and our model on the validation set are exhibited in Table 2. Obviously, our model performs better than others in terms of all metrics. This shows the rationality of our model settings and also confirms the importance of TV penalty and Bayesian inference algorithm in our proposed fusion model.

3.4. Comparison with other models

This subsection aims to study the behaviors of our proposed model and other popular counterparts, including convolutional

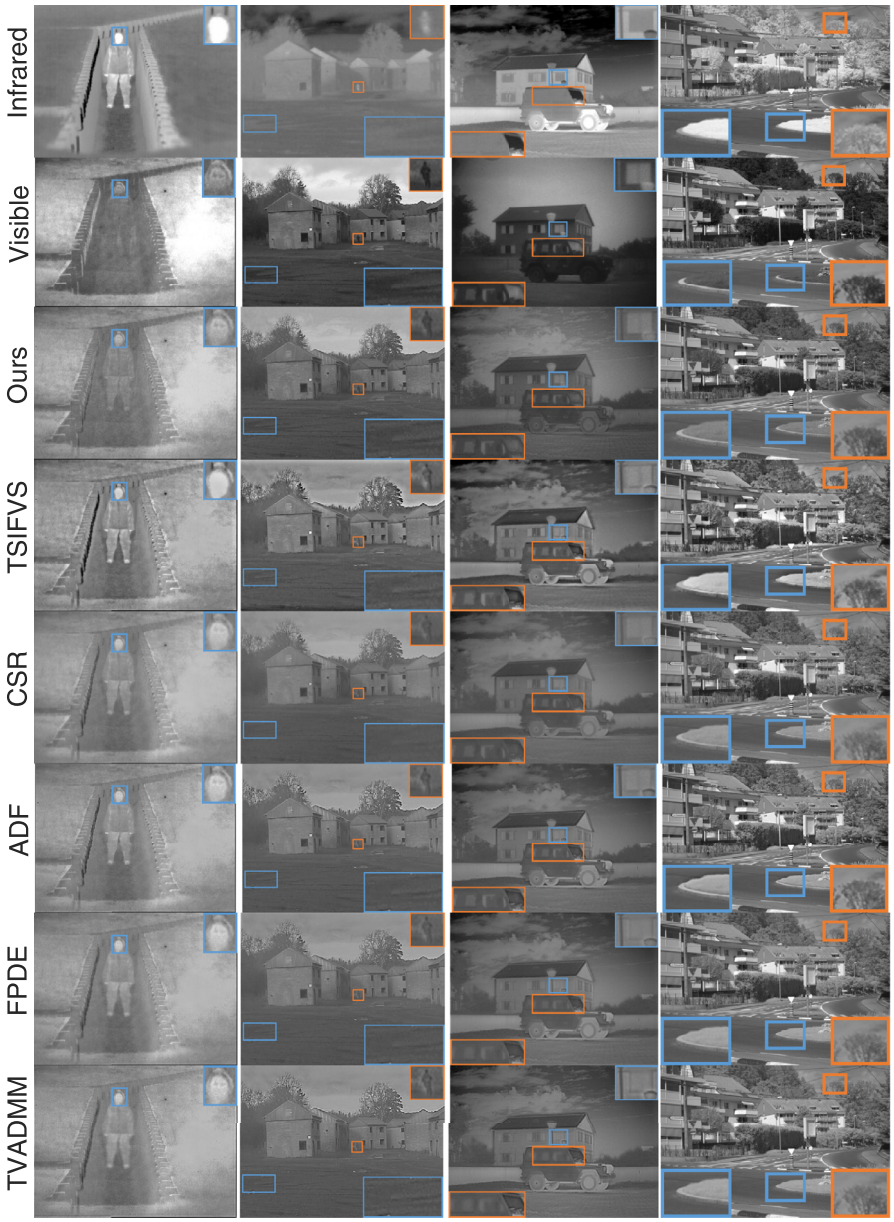

Fig. 3. Qualitative fusion results. From left to right: "Soldier_in_trench 1", "Image_04" and "Marne_04" in TNO dataset, "Image_34" in NIR-Street dataset. From top to bottom: infrared images, visible images, results of our method, TSIFVS, CSR, ADF, FPDE and TVADMM methods.

sparse representation (CSR) [42], anisotropic diffusion and karhunen-loeve transform (ADF) [43], fourth order partial differential equations (FPDE) [19], two-scale image fusion using saliency detection (TSIFVS) [22] and total variation and augmented lagrangian (TVADMM) [41].

3.4.1. Subjective visual evaluation

The qualitative fusion results of the counterparts and our model are displayed in Fig. 3, respectively. From left to right: "Soldier_in_trench_1", "Image_04", "Marne_04" in TNO dataset and "Image_34" in NIR-Street dataset. In the first column images, the TSIFVS and ADF methods have almost no face details. The TVADMM method has low target brightness, and the backgrounds of the CSR and FPDE methods (such as the trenches) are not clear enough. As for the fusion results listed in the second column, apparently, the house details of the CSR method is poor and the ground detail of the ADF method is not obvious enough. Meanwhile, the target objective of the TVADMM and TSIFVS methods have low brightness, and the background details(e.g. the trees) of the FPDE method are not clear enough. In the results of the third column images, the FPDE and ADF methods have lower brightness and fewer details, while the TVADMM and CSR methods have poorer window details, and the TSIFVS method has less obvious edge contours. In the results of the fourth column images, the edge contour of the TSIVIF and ADF methods does not fit the human visual system because of the clear boundary and artifacts. The CSR and FPDE methods are not salient enough in trees details and edges. Objects of the TVADMM method, especially the edge contour of the road (amplified in the blue box), has poor contrast and that of the FPDE method has visual blur with fewer details. On the contrary, our method can simultaneously focus on highlighting objects and the details of the backgrounds with high contrast. In short, compared with the other methods, our proposed Bayesian fusion model can generate better-fused images with highlighting targets and rich texture details.

Table 3
Quantitative results of different methods. The largest value is shown in bold, and the second largest value is underlined.

Dataset: TNO image fusion dataset						
Metrics	CSR	TSIFVS	TVADMM	ADF	FPDE	BF (Ours)
EN	6.435	**6.663**	6.407	6.387	6.438	6.633
MI	1.851	1.644	1.941	1.913	1.748	**2.828**
$Q^{AB/F}$	**0.555**	0.514	0.337	0.463	0.516	0.523
SD	23.066	27.425	22.518	22.097	22.742	**27.511**
SSIM	0.911	0.934	0.929	0.984	0.906	**1.008**
Dataset: NIR-Oldbuilding image fusion dataset						
Metrics	CSR	TSIFVS	TVADMM	ADF	FPDE	BF (Ours)
EN	7.107	**7.165**	7.114	7.084	7.077	7.127
MI	5.643	5.410	5.629	5.967	5.906	**6.225**
$Q^{AB/F}$	0.747	0.736	0.725	0.746	0.742	**0.749**
SD	42.971	**44.404**	42.867	42.210	42.498	43.899
SSIM	1.263	1.407	1.450	**1.492**	1.468	1.489
Dataset: NIR-Street image fusion dataset						
Metrics	CSR	TSIFVS	TVADMM	ADF	FPDE	BF (Ours)
EN	6.961	**7.052**	6.948	6.922	6.925	7.015
MI	4.525	4.262	4.580	4.837	4.827	**5.246**
$Q^{AB/F}$	0.685	0.653	0.629	0.656	0.665	**0.665**
SD	38.015	40.020	37.820	36.967	37.018	**40.419**
SSIM	1.321	1.418	1.471	**1.508**	1.500	1.502
Dataset: NIR-Urban image fusion dataset						
Metrics	CSR	TSIFVS	TVADMM	ADF	FPDE	BF (Ours)
EN	7.225	**7.265**	7.222	7.201	7.201	7.244
MI	5.957	5.903	6.026	6.410	6.410	**6.696**
$Q^{AB/F}$	0.813	0.813	0.811	0.829	0.829	**0.830**
SD	43.265	**44.269**	43.073	42.434	42.434	44.203
SSIM	1.431	1.575	1.602	**1.648**	1.645	1.646

Table 4
The computational times for each algorithm in four test datasets, in which the unit of time is seconds.

Methods	TNO	NIR-Oldbuilding	NIR-Street	NIR-Urban
TSIFVS	1.20 ± 0.01	1.82 ± 0.01	1.90 ± 0.01	2.12 ± 0.01
TVADMM	3.51 ± 0.01	4.74 ± 0.02	5.21 ± 0.02	5.62 ± 0.03
CSR	1291.53 ± 5.11	2113.66 ± 2.08	2339.80 ± 3.58	2600.70 ± 6.94
ADF	7.74 ± 0.02	11.03 ± 0.05	12.35 ± 0.17	13.23 ± 0.02
FPDE	21.91 ± 0.04	28.51 ± 0.04	33.36 ± 0.03	34.83 ± 0.02
BF (Ours)	10.92 ± 0.08	15.42 ± 0.04	17.31 ± 0.02	18.45 ± 0.05

3.4.2. Objective quantitative evaluation

We show a quantitative comparison of these fusion methods in Table 3. In TNO dataset, our method performs best in terms of the MI, SD, SSIM metrics, and is ranked second in EN, $Q^{AB/F}$. Meanwhile, in the three scenes of NIR Scene dataset, we get two first places in MI, $Q^{AB/F}$ and three second places in EN, SD, SSIM for "Oldbuilding" scenes, three first places in MI, $Q^{AB/F}$, SD and two second places in EN, SSIM for "Street" scenes, two first places in MI, $Q^{AB/F}$ and three second places in EN, SD, SSIM for "Urban" scenes. In summary, other methods may perform well under the measurement of a few metrics, but our model has good performance on almost all metrics, which demonstrates the excellent fusion effect of our method.

3.4.3. Computational time

To compare the time complexities of each fusion algorithm, Table 4 lists the average computational time of each algorithm over 10 independent runs. The results manifest that our method is faster than CSR and FPDE, but slightly slower than TSIFVS, TVADMM and ADF. In general, our method can quickly generate high-quality fusion images.

3.5. Result analysis

According to the loss function defined in Eq. (1), minimizing $||I-U||_1$ allows the fusion image to save more highlighting radiation information, and minimizing $||I-V||_1$ makes the details and textures be well preserved. Compared with other methods, our model can obtain adaptive fusion weights for each input image. Therefore, the qualitative and quantitative results reveal the outstanding performance of the proposed Bayesian fusion model on detail retention and image contrast. Additionally, the TV penalty lets the fused image better conform to human visual perception, preserves more texture details and prevents the artifacts and halos in the fused image.

In short, our model reasonably considers the characteristics of the fusion task and the Bayesian method effectively infers the fusion weights. This explains why our proposed method outperforms the other state-of-the-art fusion approaches.

4. Conclusions and future work

In this paper, with respect to the task of fusing infrared and visible images, we present a novel Bayesian fusion model to maintain thermal radiation and texture detail information from source images. In our model, the image fusion task is transformed into a regression problem, then a hierarchical Bayesian framework is established to convert the optimization problem into the inference of a probability model with latent variables, which can be solved by the EM algorithm with the half-quadratic splitting algorithm. In addition, the TV penalty is utilized to make the fused image satisfy human visual system. Notably, the parameters (i.e., fusion weights) are adaptive to the input images, which means that our model can better consider the characteristics of input images.

Our ablation experiments on the validation set demonstrate the effectiveness of the TV penalty and the Bayesian inference algorithm. On the test datasets, compared with the other methods in TNO and NIR datasets, our method can generate better fused images with highlighting thermal radiation targets and abundant texture details at a relatively fast speed. The superior fusion performance verifies the rationality and effectiveness of our model and algorithm, which may potentially facilitate automatic detection and precise positioning of targets.

In the future, we will focus on how to further improve the calculation speed of the algorithm, and apply our Bayesian model to other image fusion and signal processing fields, such as multi-exposure, multi-focus and multi-spectral fusion.

Declaration of Competing Interest

The authors declare that they have no competing interests.

CRediT authorship contribution statement

Zixiang Zhao: Software, Writing - original draft, Methodology, Investigation, Formal analysis. **Shuang Xu:** Methodology, Writing - original draft, Funding acquisition. **Chunxia Zhang:** Funding acquisition, Conceptualization. **Junmin Liu:** Funding acquisition, Conceptualization. **Jiangshe Zhang:** Funding acquisition, Conceptualization.

Acknowledgments

The research is supported by the National Key Research and Development Program of China under grant 2018AAA0102201, the National Natural Science Foundation of China under grants 61976174, 11671317 and 61877049, the Fundamental Research Funds for the Central Universities under grant number xzy022019059.

References

[1] J. Ma, Y. Ma, C. Li, Infrared and visible image fusion methods and applications: a survey, Inf. Fusion 45 (2019) 153–178.
[2] R. Singh, M. Vatsa, A. Noore, Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition, Pattern Recognit. 41 (3) (2008) 880–893.
[3] J.-j. Zong, T.-s. Qiu, Medical image fusion based on sparse representation of classified image patches, Biomed. Signal Process. Control 34 (2017) 195–205.
[4] G. Simone, A. Farina, F.C. Morabito, S.B. Serpico, L. Bruzzone, Image fusion techniques for remote sensing applications, Inf. Fusion 3 (1) (2002) 3–15.
[5] X. Li, Y. Yuan, Q. Wang, Hyperspectral and multispectral image fusion based on band simulation, IEEE Geosci. Remote Sens. Lett. 17 (3) (2019) 479–483.
[6] X. Li, Y. Yuan, Q. Wang, Hyperspectral and multispectral image fusion via nonlocal low-rank tensor approximation and sparse representation, IEEE Trans. Geoscie. Remote Sens. (2020) In press, doi:10.1109/TGRS.2020.2994968.
[7] C. Chen, Y. Li, W. Liu, J. Huang, Image fusion with local spectral consistency and dynamic gradient sparsity, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2760–2765.
[8] S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: a survey of the state of the art, Inf. Fusion 33 (2017) 100–112.
[9] 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.
[10] 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.
[11] Y. Liu, X. Chen, Z. Wang, Z.J. Wang, R.K. Ward, X. Wang, Deep learning for pixel-level image fusion: recent advances and future prospects, Inf. Fusion 42 (2018) 158–173.
[12] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transforms for image fusion, Inf. Fusion 12 (2) (2011) 74–84.
[13] G. Pajares, J.M. De La Cruz, A wavelet-based image fusion tutorial, Pattern Recognit. 37 (9) (2004) 1855–1872.
[14] Z. Zhang, R.S. Blum, et al., A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application, Proc. IEEE 87 (8) (1999) 1315–1326.
[15] B. Yang, S. Li, Visual attention guided image fusion with sparse representation, Opt.-Int. J. Light Electron. Opt. 125 (17) (2014) 4881–4888.
[16] J. Wang, J. Peng, X. Feng, G. He, J. Fan, Fusion method for infrared and visible images by using non-negative sparse representation, Infrared Phys. Technol. 67 (2014) 477–489.
[17] S. Li, H. Yin, L. Fang, Group-sparse representation with dictionary learning for medical image denoising and fusion, IEEE Trans. Biomed. Eng. 59 (12) (2012) 3450–3459.
[18] Y. Liu, Z. Wang, Multi-focus image fusion based on sparse representation with adaptive sparse domain selection, in: Proceedings of the 2013 Seventh International Conference on Image and Graphics, 2013, pp. 591–596.
[19] D.P. Bavirisetti, G. Xiao, G. Liu, Multi-sensor image fusion based on fourth order partial differential equations, in: Proceedings of the 2017 20th International Conference on Information Fusion (Fusion), IEEE, 2017, pp. 1–9.
[20] W. Kong, Y. Lei, H. Zhao, Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization, Infrared Phys. Technol. 67 (2014) 161–172.
[21] U. Patil, U. Mudengudi, Image fusion using hierarchical PCA., in: Proceedings of the 2011 International Conference on Image Information Processing, IEEE, 2011, pp. 1–6.
[22] D.P. Bavirisetti, R. Dhuli, Two-scale image fusion of visible and infrared images using saliency detection, Infrared Phys. Technol. 76 (2016) 52–64.
[23] X. Zhang, Y. Ma, F. Fan, Y. Zhang, J. Huang, Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition, J. Opt. Soc. Am. A 34 (8) (2017) 1400–1410.
[24] J. Zhao, Y. Chen, H. Feng, Z. Xu, Q. Li, Infrared image enhancement through saliency feature analysis based on multi-scale decomposition, Infrared Phys. Technol. 62 (2014) 86–93.
[25] Q. Wei, N. Dobigeon, J.-Y. Tourneret, Bayesian fusion of multi-band images, IEEE J. Sel. Top. Signal Process. 9 (6) (2015) 1117–1127.
[26] R.C. Hardie, M.T. Eismann, G.L. Wilson, Map estimation for hyperspectral image resolution enhancement using an auxiliary sensor, IEEE Trans. Image Process. 13 (9) (2004) 1174–1184.
[27] Y. Zhang, S. De Backer, P. Scheunders, Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images, IEEE Trans. Geosci. Remote Sens. 47 (11) (2009) 3834–3843.
[28] Q. Wei, N. Dobigeon, J.-Y. Tourneret, Fast fusion of multi-band images based on solving a sylvester equation, IEEE Trans. Image Process. 24 (11) (2015) 4109–4121.
[29] X. Yang, S. Yin, Variational bayesian inference for fir models with randomly missing measurements, IEEE Trans. Ind. Electron. 64 (5) (2016) 4217–4225.
[30] X. Liu, X. Yang, A variational bayesian approach for robust identification of linear parameter varying systems using mixture laplace distributions, Neurocomputing 395 (2020) 15–23.
[31] H. Li, X.-J. Wu, Densefuse: A fusion approach to infrared and visible images, IEEE Trans. Image Process. 28 (5) (2018) 2614–2623.
[32] H. Li, X.-J. Wu, J. Kittler, Infrared and visible image fusion using a deep learning framework, in: Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 2705–2710.
[33] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. Royal Stat. Soc.: Ser. B (Methodol.) 39 (1) (1977) 1–22.
[34] H. Xu, J. Ma, Z. Le, J. Jiang, X. Guo, Fusiondn: A unified densely connected network for image fusion., in: Proceedings of the AAAI, 2020, pp. 12484–12491.
[35] A. Toet, TNO Image Fusion Dataset, (2014). 10.6084/m9.figshare.1008029.v1
[36] M. Brown, S. Süsstrunk, Multispectral SIFT for scene category recognition, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR11), 2011, pp. 177–184. Colorado Springs
[37] J.W. Roberts, J.A. Van Aardt, F.B. Ahmed, Assessment of image fusion procedures using entropy, image quality, and multispectral classification, J. Appl. Remote Sens. 2 (1) (2008) 023522.
[38] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, Electron. Lett. 38 (7) (2002) 313–315.
[39] Y.-J. Rao, In-fibre bragg grating sensors, Meas. Sci. Technol. 8 (4) (1997) 355–375.
[40] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (3) (2002) 81–84.
[41] 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.
[42] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Image fusion with convolutional sparse representation, IEEE Signal Process. Lett. 23 (12) (2016) 1882–1886.
[43] D.P. Bavirisetti, R. Dhuli, Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform, IEEE Sensors J. 16 (1) (2016) 203–209.