



## Infrared and visible image fusion methods and applications: A survey

Jiayi Ma, Yong Ma\*, Chang Li

*Electronic Information School, Wuhan University, Wuhan 430072, China*



### ARTICLE INFO

**Keywords:**

Image fusion  
Infrared image  
Visible image  
Image registration  
Evaluation metric

### ABSTRACT

Infrared images can distinguish targets from their backgrounds based on the radiation difference, which works well in all-weather and all-day/night conditions. By contrast, visible images can provide texture details with high spatial resolution and definition in a manner consistent with the human visual system. Therefore, it is desirable to fuse these two types of images, which can combine the advantages of thermal radiation information in infrared images and detailed texture information in visible images. In this work, we comprehensively survey the existing methods and applications for the fusion of infrared and visible images. First, infrared and visible image fusion methods are reviewed in detail. Meanwhile, image registration, as a prerequisite of image fusion, is briefly introduced. Second, we provide an overview of the main applications of infrared and visible image fusion. Third, the evaluation metrics of fusion performance are discussed and summarized. Fourth, we select eighteen representative methods and nine assessment metrics to conduct qualitative and quantitative experiments, which can provide an objective performance reference for different fusion methods and thus support relative engineering with credible and solid evidence. Finally, we conclude with the current status of infrared and visible image fusion and deliver insightful discussions and prospects for future work. This survey can serve as a reference for researchers in infrared and visible image fusion and related fields.

### 1. Introduction

Image fusion is an enhancement technique that aims to combine images obtained by different kinds of sensors to generate a robust or informative image that can facilitate subsequent processing or help in decision making. The keys to an excellent fusion method are effective image information extraction and appropriate fusion principles, which allow useful information to be extracted from source images and integrated in the fused image without introducing any artifact in the process [1,2]. In the age of information explosion, sensor techniques are being developed rapidly. The appearance of complicated applications requires comprehensive information about a certain scenario for enhanced understanding of various conditions. Sensors of the same type acquire information from only one aspect and are thus unable to provide all required information. Therefore, fusion techniques play an increasingly important role in modern applications and computer vision.

Images of different types, such as visible, infrared, computed tomography (CT), and magnetic resonance imaging (MRI), are good source images for fusion. Among the combinations of these types, infrared and visible image fusion is superior in many aspects. First, their signals come from different modalities, thereby providing scene information from different aspects; i.e., visible images capture reflected light, whereas infrared images capture thermal radiation. Therefore,

this combination is more informative than that of single-modality signals. Second, infrared and visible images present characteristics that are inherent in nearly all objects. Furthermore, these images can be acquired by relatively simple equipment, as opposed to such imaging techniques as CT and MRI, which have rigorous requirements. Finally, infrared and visible images share complementary properties, thus producing robust and informative fused images. Visible images typically have high spatial resolution and considerable detail and chiaroscuro; thus, they are suitable for human visual perception. However, these images can be easily influenced by severe conditions, such as poor illumination, fog, and other effects of bad weather. Meanwhile, infrared images, which depict the thermal radiation of objects, are resistant to these disturbances but typically have low resolution and poor texture. Fusion techniques for visible and infrared images can thus be used in more fields than other fusion schemes due to the ubiquitous and complementary characteristics of the utilized images. Object recognition [3], detection [4], image enhancement [5], surveillance [6], and remote sensing [7] are typical applications of infrared and visible image fusion.

Fusion algorithms can be developed with different schemes and generally divided into seven categories according to their adopted theories, i.e., multi-scale transform- [8–10], sparse representation- [2,11], neural network- [12,13], subspace- [14,15], and saliency-

\* Corresponding author.

E-mail addresses: [jyma2010@gmail.com](mailto:jyma2010@gmail.com) (J. Ma), [mayong@whu.edu.cn](mailto:mayong@whu.edu.cn) (Y. Ma), [lichang1214@gmail.com](mailto:lichang1214@gmail.com) (C. Li).

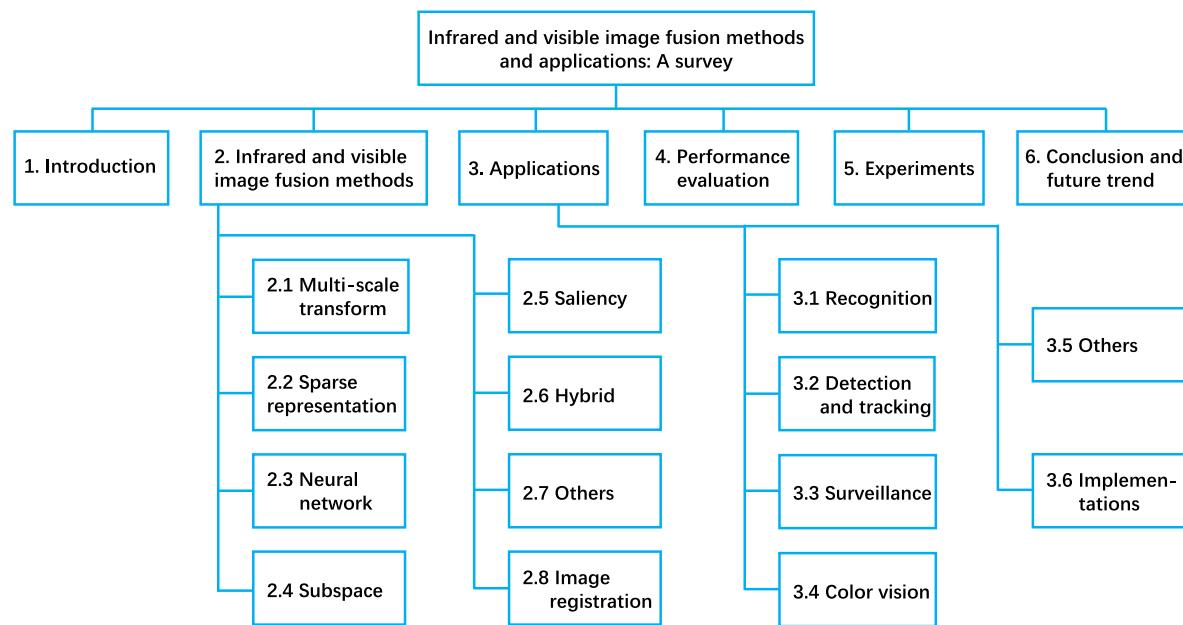


Fig. 1. Structure of this survey.

based [16,17] methods, hybrid models [18,19], and other methods [20,21]. Multi-scale transform-based methods are the most active field in image fusion and assume images to be represented by various layers in different grains. These methods decompose source images into several levels, fuse corresponding layers with particular rules, and reconstruct the target images accordingly. Popular transforms used for decomposition and reconstruction include wavelet [22], pyramid [23], curvelet [24], and their revised versions. Sparse representation-based methods work on the basis of the possible representation of images with linear combinations of sparse bases in over-complete dictionaries, which are key to their good performance. Neural network-based methods imitate the perception behavior of the human brain to deal with neural information; this procedure has the advantages of good adaptability, fault tolerance, and anti-noise capacity. Subspace-based methods use similar ideas but implement them in spaces with complete bases instead of over-complete ones. Principal component analysis (PCA), non-negative matrix factorization (NMF), and independent component analysis (ICA) are the main methods in this category. Saliency-based methods are based on the fact that human visual attention is often captured by objects or pixels that are more significant than their neighbors. In addition, these methods can retain the integrity of salient object regions and improve the visual quality of the fused images. The abovementioned infrared and visible image fusion methods have their advantages and disadvantages. Meanwhile, hybrid models combine the advantages of these schemes to improve image fusion performance. The other techniques are based on total variation [20], fuzzy theory [25], and entropy [21]. These methods can lead to a new vision for infrared and visible image fusion, which can inspire the design of novel models.

The research on fusion algorithms in theory and application has been rapidly developing in recent years. However, certain problems require attention. An appropriate information extraction method is key to ensuring good fusion performance. Several current approaches, especially multi-scale transform-based methods, depend on predefined transforms and corresponding levels for decomposition and reconstruction. However, no criteria are used to evaluate these transforms and levels. Consequently, transforms are selected blindly, thereby easily degrading performance [26]. Moreover, current fusion rules are mostly excessively simple that certain artifacts, such as halo, are introduced in the results. Methods in other categories also suffer from several problems. For example, constructing suitable over-complete dictionaries

that have good representation of target data using sparse representation-based methods is still difficult. Additionally, suitable sparse coding techniques are also important for fusion performance. Designing appropriate neural networks and adjusting the corresponding parameters remain challenging. Furthermore, applying state-of-the-art deep learning techniques to infrared and visible image fusion is yet to be addressed. Finding a subspace with powerful expressive ability in subspace-based methods is difficult. Meanwhile, designing niche targeting and anti-noise saliency detection methods in accordance with the human visual system is also challenging, and using the saliency of source images is still an open problem. In addition, combining the advantages of different methods to create enhanced hybrid image fusion models is a prominent issue in infrared and visible image fusion. Designing new and good infrared and visible image fusion methods with novel ideas is also difficult.

Infrared and visible image fusion is useful, but its current methods still struggle with several problems. Thus, this survey aims to specifically and comprehensively review the existing fusion algorithms for infrared and visible images and hopes to shed light on the improvement of such approaches. This work has certain advantages in comparison with previous reviews. The current study finds that infrared and visible image fusion is considerably superior to other combinations of source images. However, current reviews are mainly about general fusion approaches and pay only a small amount of attention to methods for infrared and visible image fusion, thus lacking pertinence. This survey is comprehensive and specific, reviewing and concluding nearly all main approaches for infrared and visible image fusion. Furthermore, this work includes theories and applications in infrared and visible image fusion and discusses corresponding tendency in future study. Finally, extensive experiments are conducted in this survey to provide an objective comparison of the typical approaches and support relative engineering with credible and solid evidence.

The structure of this survey is schematically illustrated in Fig. 1. Section 2 discusses image fusion and registration methods in different categories. Section 3 introduces the main applications of infrared and visible image fusion. Section 4 provides an overview of the metrics used to measure fusion quality. In Section 5, we extensively experiment to compare the typical algorithms in each category and provide an objective performance reference. Finally, Section 6 presents the conclusion and discussion on future work.

## 2. Infrared and visible image fusion methods

In this section, we comprehensively review infrared and visible image fusion methods. These methods are divided into seven categories according to their corresponding theories, and we review each of them in detail. Successful image fusion has a prerequisite; *i.e.*, different source images should be strictly aligned in advance. Therefore, we also provide a brief review on registration techniques, especially in the context of infrared and visible image registration.

### 2.1. Multi-scale transform-based methods

Over the past few decades, multi-scale transforms have been effective in such fields as infrared and visible image fusion [27,28]. Multi-scale transforms can decompose original images into components of different scales, where each component represents the sub-image at each scale and real-world objects typically comprise components at different scales [1]. Several studies have demonstrated that multi-scale transforms are consistent with human visual characteristics, and this property can enable fused images to have good visual effect [18,29,30].

In general, infrared and visible image fusion schemes based on multi-scale transforms comprise three steps [1], as shown in Fig. 2. First, each source image is decomposed into a series of multi-scale representations. Then, the multi-scale representations of the source image are fused according to a given fusion rule. Finally, the fused image is acquired using corresponding inverse multi-scale transforms on the fused representations. The key in multi-scale transform-based fusion

schemes lies in the selection of the transforms and fusion rules. Next, we review the techniques in this category on the basis of these two aspects.

#### 2.1.1. Multi-scale transform

##### (1) Pyramid transform

The concept of pyramid transform was proposed in the 1980s and aims to decompose original images into sub-images with different scales of spatial frequency band, which have a pyramid data structure [31]. Since then, various types of pyramid transforms have been proposed for infrared and visible image fusion, Laplacian [32–34], steerable [35–37], and contrast [38–41] pyramids.

The Laplacian pyramid transform requires iteratively performing the following steps: low-pass filtering, sampling, interpolating, and differencing [31,32]. For example, Bulanon et al. adopted the Laplacian pyramid transform and fuzzy logic for fusing infrared and visible images of fruits [32], and their method could achieve better fruit detection performance than using only the infrared image. The steerable pyramid transform decomposes the source image into a collection of sub-images in different scales and orientations. This transform has the advantages of self-inverting, aliasing-free, translation and rotation-invariant [35,42] properties. For example, Liu et al. proposed an image fusion method that was based on the steerable pyramid and expectation maximization algorithm, and their proposed method could outperform the traditional steerable pyramid fusion method [36]. The contrast pyramid transform stems from the Laplacian pyramid transform and calculates the ratio of two adjacent low-pass filtering images of the Gaussian pyramid; this model could consider local contrast [38]. Jin

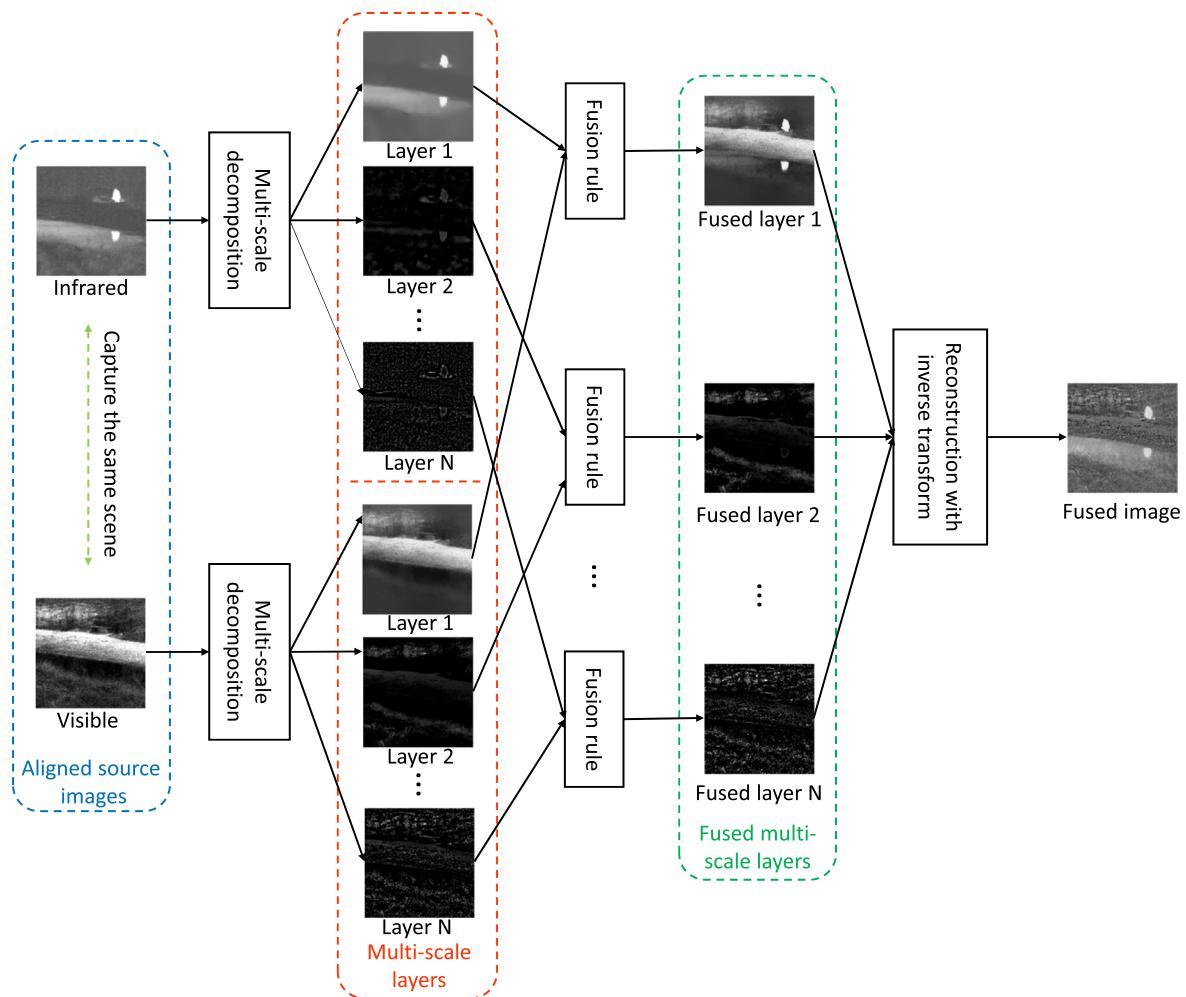


Fig. 2. Multi-scale transform based infrared and visible image fusion scheme.

et al. proposed an infrared and visible image fusion method that was based on the contrast pyramid and multi-objective evolutionary algorithm, where the latter was used to optimize fusion coefficients.

#### (2) Wavelet transform

The concept of wavelet transform was proposed by Grossman and Morlet in the 1980s. Then, Mallet established multi-resolution theory for wavelet transform; this theory was inspired by the tower algorithm for signal decomposition and reconstruction [43]. The wavelet transform has since been widely applied in image fusion and related fields. Unlike the multi-scale representation coefficients of the pyramid transform, those of the wavelet transform are uncorrelated. Typical wavelet transform-based infrared and visible image fusion methods include discrete wavelet transform [44–48], dual-tree discrete wavelet transform [49,50], lifting wavelet transform [51], lifting stationary wavelet transform [52], redundant-lifting non-separable wavelet multidirectional analysis [53], spectral graph wavelet transform [54], quaternion wavelet transform [55], motion-compensated wavelet transform [56] and multi-wavelet [57].

The discrete wavelet transform can decompose the source image into a series of high- and low-frequency sub-images, which can be implemented by a bank of filters. However, the discrete wavelet transform has the problems of oscillation, shift variance, aliasing, and lack of directionality [58]. The dual-tree discrete wavelet transform is based on the separable filter bank, which has the advantages of computational efficiency, near shift-invariance, and directional selectivity over the discrete wavelet transform [49,58]. The lifting wavelet transform has also been successfully applied to infrared and visible image fusion [51] and can be seen as a completely spatial domain transform method; it has the advantages of adaptive design, irregular sampling, and integral transform over the traditional wavelet transform. Yan et al. proposed a new infrared and visible image fusion that was based on the spectral graph wavelet transform [54], and their method represented the irregular regions of images. Moreover, Chai et al. applied an infrared and visible image fusion model that was based on the quaternion wavelet transform, which could be considered a generalization of the traditional wavelet transform on the basis of quaternion algebra. Xu et al. proposed an infrared and visible video fusion method formed by combining the wavelet transform with motion compensation [56].

#### (3) Nonsubsampled contourlet transform

The wavelet transform is a fast and efficient method of representing 1D piecewise smooth signals and has been successfully applied in many signal processing and communication domains. Extending the 1D signal wavelet transform to a 2D image wavelet transform can isolate discontinuities at edge points. However, the 2D wavelet transform cannot capture the abundant directional information of images. To overcome this problem, Minh N. Do and Martin Vetterli proposed an efficient multi-direction multi-resolution image representation method named contourlet transform [59]. This model is based on the Laplacian pyramid [32] and directional filter bank [60], which can capture the geometry of image edges well [61–63]. However, the contourlet transform suffers from the problems of shift variance due to down-sampling and up-sampling and redundancy caused by the pyramidal filter bank structure.

The nonsubsampled contourlet transform (NSCT) [64] was proposed as a flexible and fully shift-invariant model with fast implementation to overcome the abovementioned problems of the contourlet transform. The NSCT is based on the nonsubsampled pyramid and nonsubsampled directional filter banks. The nonsubsampled pyramid filter bank decomposes each source image into a set of high- and low-frequency sub-images to attain multi-resolution decomposition, whereas the nonsubsampled directional filter bank decomposes these high-frequency sub-images to attain multi-direction decomposition [64]. The NSCT is shift-invariant because it eliminates the down-sampling and up-sampling operations of the contourlet transform, thereby ensuring good frequency selectivity and regularity. Hence, the nonsubsampled directional filter bank can reduce the pseudo-Gibbs filter bank structure.

phenomenon and the effect of misregistration [64,65]. The NSCT has been widely applied in infrared and visible image fusion [12,65–77] because of these advantages. For example, Meng et al. proposed an infrared and visible image fusion that was based on object region detection and the NSCT, and their model could clarify objects and preserve details and visual artifacts in the fused image [65]. Adu et al. proposed a new fusion method that was based on the NSCT and gradient characteristics, and their method could not only extract the target from the infrared image but also preserve details in the visible image [67].

#### (4) Edge-preserving filter

The edge-preserving filter emerged as an effective tool for computational photography and other image processing applications in the 2000s. This filter decomposes the source image into a smooth base layer and one or more detail layers [78]. The edge-preserving filter can preserve the spatial consistency of structures and reduce halo artifacts around edges. The base layer is typically obtained by using an edge-preserving filter on the image, thereby potentially capturing big changes in intensity. The detail layers comprise a series of difference images that can preserve details at various progressively fine scales [78]. The edge-preserving filter has been widely used in various fields due to these advantages. In recent years, this filter has also been applied in infrared and visible image fusion, such as mean filter [79,80], nonlocal means [81], anisotropic diffusion [82], bilateral filter [83,84], weighted least squares filter [85,86], multi-scale edge-preserving filter [87], local edge-preserving filter [16],  $L_1$  fidelity with  $L_0$  gradient [88], guided filter [89,90], fourth order partial differential equation [14] and hybrid filter [19,91].

The mean filter is a classic and simple denoising method that aims to remove noise from each pixel using the mean value of its spatial neighboring pixels. The nonlocal means and the mean filter are similar, except that the former adopts the neighboring pixels according to the similarity of the intensity gray level. Anisotropic diffusion is based on partial differential equations, which can not only smooth an image in homogeneous regions but also preserve edges, shapes, and positions [92]. The bilateral filter emerged as a fast alternative to the time-consuming anisotropic diffusion model and combines low-pass filtering with an edge-stopping function [93]. Mathematically, the bilateral filter removes noise from each pixel using the Gaussian weighted average of the neighboring pixels according to spatial and spectral distances. Hence, this filter is local and nonlinear and can thus simultaneously smooth images and preserve details. The weighted least squares filter aims to balance the smoothing and approximation of original images, which can simultaneously reduce ringing and deblur the images [78,85]. The multi-scale edge-preserving filter is based on the Laplacian pyramid and weighted least squares filter and can take advantage of multi-scale decomposition and edge-preserving filters [87].

The local edge-preserving filter can simultaneously preserve global salient edges and local shapes, thereby enhancing local details [16,94]. Shen et al. proposed a new image decomposition method that was based on the  $L_1$  fidelity with  $L_0$  gradient [88,95], which is used to preserve details of base and detail layers. The guided filter is general, translation-variant [96], and based on the local linear model; it is a fast linear time filter method that has the advantages of edge-preserving smoothing and structural transfer. Hence, the guided filter has been applied to numerous fields, including infrared and visible image fusion [89,90]. The fourth-order partial differential equation-based method is an edge-preserving filter that decomposes the source image into base and detail sub-images [14]. Certain hybrid filters for image decomposition consist of two or more filters. For example, Zhou et al. proposed a new method that was based on the Gaussian and bilateral filters; their model could decompose the image into multi-scale edge and texture details [91]. Ma et al. proposed a new method that was based on the Gaussian and rolling guidance filters, and their approach reduced halos and preserved information as specific scales [19].

#### (5) Other multi-scale transform methods

Other types of multi-scale transform methods have also been applied in infrared and visible image fusion. These models include curvelet transform [97–101], framelet transform [13,102,103], shearlet transform [15,104–109], tetrolet transform [110,111], top-hat transform [112–115], discrete cosine transform [116], directionlets transform [117], empirical mode decomposition [118–121], Internal Generative Mechanism [122] and multi-resolution singular value decomposition (SVD) [123].

The concept of the curvelet transform was proposed by Candes and Donoho [124] and aims to overcome the inherent limitations of the wavelet transform, which is ill-suited to representing the edges and geometric structures of images. The curvelet transform is a multi-resolution, multi-direction pyramid that can preserve geometric regularity along edges [125]. However, image fusion methods based on the wavelet and curvelet transforms have block artifacts and reduced contrast. The framelet transform can overcome these problems using tight frame filter banks, which are symmetry- and shift-invariant [15]. The theory of shearlet transform is based on composite wavelets, which are the optimal approximation of 2D functions [126]. The support of the shearlet filter is smaller than that of the directional filter of the contourlet transform, which has the advantage of directional selectivity and computational efficiency [107]. The tetrolet transform is an adaptive Haar wavelet transform, which has tetromino support and geometric adaptability [110]. This transform can capture the geometric features of images well. Hence, the tetrolet transform has been applied in infrared and visible image fusion [110,111]. The top-hat transform is a multi-scale transform that is based on mathematical morphology. This transform extracts regions of interest while preserving image details [115]. The directionlet transform was proposed by Velisavljevic et al. as a multi-scale, multi-directional anisotropic transform [127] and can preserve the geometric features of source images. This transform has also been successfully applied in infrared and visible image fusion [117]. Empirical mode decomposition was proposed by Huang et al. [128] and emerged as a non-stationary adaptive signal decomposition method. This model is now applied in image fusion and many other applications. Empirical mode decomposition has the advantages of extracting low-frequency oscillations and fully exploiting the original images [128].

Multi-scale transform methods have also been successfully applied in infrared and visible image fusion due to the abovementioned advantages. For example, Huang et al. proposed an infrared and visible image fusion method that was based on curvelet transformation and visual attention mechanisms; their model could elevate the signal-to-noise ratio of fused images and highlight dim targets [101]. Zhu et al. proposed an infrared and visible image fusion method that was based on an improved multi-scale top-hat transform model; this model could highlight the target of infrared images and preserve details of visible images better than certain conventional multi-scale transform image methods could [115]. Song et al. adopted multi-resolution SVD to decompose source images into smooth and detailed sub-images [123].

### 2.1.2. Fusion rule

#### (1) Coefficient combination method

When decomposing the source images into a series of multi-scale representations using the multi-scale transform, we need to fuse the multi-scale representations of the infrared and visible images according to a given fusion rule. The coefficient combination method is the most commonly used rule in infrared and visible image fusion and uses at least two schemes to combine the multi-scale representation coefficients. These schemes are the choose-max [18,55,67,68,70,81,83,99,105,107,110,112–115] and weighted average [16,32,34,45–47,49,51,55,56,66–69,71,79–82,84–88,90,91,98–100,109,112,113,115,119–121,129] models.

The choose-max scheme selects the max coefficient as the multi-scale representation coefficient. For example, Chai et al. fused high-frequency subbands with the choose-max rule on the basis of energy

and contrast [55]. Hu et al. performed the choose-max fusion rule on directional detail subbands [83]. Yan et al. adopted the max absolute rule to fuse the coefficients of detail subbands [81]. Zhu et al. adopted the max absolute rule to merge high-pass bands [114]. Meanwhile, the weighted average scheme combines the multi-scale representation coefficients of the source images by weighting. Obtaining the weight of the weighted average scheme can be done in a number of ways, the simplest of which is the averaging rule [68,86,99,113]. Recently, saliency analysis has emerged as a powerful tool for obtaining weight maps [16,79,80,85,87,88,129]. Certain weighted average fusion rules are based on this method [80,85,88]. Furthermore, saliency analysis is typically combined with a filter to obtain weight maps [16,79,87,129]. Fuzzy logic is usually integrated into the weighted average fusion scheme for infrared and visible image fusion [32,49,66].

Approaches to obtaining the weight of the weighted average scheme apart from the above are available. For example, Vanmali et al. adopted local entropy, local contrast, and visibility to obtain a weight map [34]. Yan et al. used the local neighborhood gradient to fuse approximate subbands [81]. Hu et al. obtained the weight maps of base and detail layers using entropy and the cumulative distribution of gray levels [90]. Liu et al. constructed a weight map on the basis of scene consistency [109], thus possibly taking advantage of not only the pixel domain but also the spatial domain. Sharma et al. adopted local variance to obtain adaptive weight [120].

#### (2) Region-based fusion rule

Many fusion rules are pixel-level image fusion. However, in several practical applications, our attention is focused on the objects of images at the region level. Hence, region-level information should be considered during image fusion [130]. Region-based rules are based on spatial and inter- and intra-scale dependencies of images and can consider the low- and intermediate-level structures of images. Consequently, region-based rules have been widely used in infrared and visible image fusion [46,50,51,67–69,105,107,110–113,115,119].

The most representative method is based on the salient region [16,79,80,85,87,88,107,129]. The saliency method aims to identify regions that are more salient than their neighbors. This model has been used to extract visually salient regions of images, which can be used to obtain saliency maps of multi-scale sub-images. Hence, the saliency method can be used for infrared and visible image fusion. Zhang et al. adopted the super-pixel-based saliency method to extract salient target regions; then, the fused coefficient could be obtained by the extracted target region using a morphological method [107].

Other region-based rules, such as feature region extraction [112,114,115], regions of interest [68,113], region segmentation [50,119], regional uniformity [67,69], regional energy [46,105], local region gradient [110,111], region-based activity measurement [51], and multi-judgment fusion rule [70] have also been proposed for infrared and visible image fusion. Bright and dim region methods extract feature and detail information first. Then, the extracted feature regions can be integrated into the fusion process [112,114,115]. Regional energy is used to obtain the fused low-frequency coefficients of infrared and visible images [46,105]. Chen et al. proposed a multi-judgment fusion rule by comparing regional energy, clarity, and variance contrast [70].

#### (3) Other fusion rules

Other types of infrared and visible image fusion rules, such as activity-level measurement [18,51,72], coefficient grouping method [131], optimization-based rule [19,38–40,47], Otsu method and top-bottom-hat morphological transform [41], mean gradient and PCA [61], NMF [15], and spiking cortical model [106], have been proposed.

The activity level of the multi-scale representation coefficient measures the local energy of corresponding multi-scale sub-images. The activity-level measurement can be roughly categorized into three classes, i.e., coefficient-, window-, and region-based measures [10]. Liu et al. adopted the max absolute values of coefficients as activity-level

measurement to obtain fused high-pass coefficients [18]. Zou et al. adopted a region-based measure to obtain low-frequency coefficients [51]. The coefficient grouping method associates multi-scale representation coefficients with each other and can be classified into three, *i.e.*, no-, single-scale, and multi-scale grouping schemes [10,28]. Xue et al. proposed an infrared and visible image fusion method that was based on the multi-scale grouping scheme [131] and could detect weapons concealed in clothing.

Optimization schemes are adopted to optimize fusion coefficients for infrared and visible image fusion [19,38–40,47]. The basic concept is to formulate a rule for obtaining the fused coefficients as an optimization problem. Jin et al. proposed three infrared and visible image fusion methods that were based on clonal selection [38] and teaching-learning-based optimization [39] and a multi-objective evolutionary algorithm [40]. Moreover, Li et al. adopted the mean gradient and PCA to fuse high- and low-frequency sub-images, respectively [61]. Kong et al. used fast NMF to fuse high- and low-frequency sub-images; their model could not only preserve image geometry but also achieve computational efficiency higher than that of traditional multi-scale geometry methods [15].

## 2.2. Sparse representation-based methods

Sparse representation is an effective tool for characterizing the human visual system [132] and has been successfully applied in different fields, such as image analysis, computer vision, pattern recognition, and machine learning [133–135]. Unlike infrared and visible image fusion methods based on multi-scale transform with prefixed basis functions, sparse representation image fusion methods aim to learn an over-complete dictionary from a large number of high-quality natural images. Then, the source images can be sparsely represented by the learned dictionary, thereby potentially enhancing the representation of meaningful and stable images [26]. Besides, misregistration or noise can bring bias to fused multi-scale representation coefficients, thus leading to visual artifacts in the fused image. Meanwhile, sparse representation-based fusion methods divide source images into several overlapping patches using a sliding window strategy, thereby potentially reducing visual artifacts and improving robustness to misregistration [18]. In general, sparse representation-based infrared and visible image fusion schemes comprise four steps [26], as shown in Fig. 3. First, each source image is decomposed into several overlapping patches using a sliding window strategy. Second, an over-complete dictionary is learned from many high-quality natural images, and sparse coding is performed on each patch to obtain the sparse representation coefficient using the learned over-complete dictionary. Third, the sparse representation coefficients are fused according to a given fusion rule. Finally, the fused image is reconstructed using the fused coefficients using the learned over-complete dictionary. The key

in sparse representation-based fusion schemes lies in the over-complete dictionary construction, sparse coding, and fusion rule. In the following sections, we review related techniques on the basis of these three aspects.

### 2.2.1. Over-complete dictionary construction

Constructing a good over-complete dictionary is important in infrared and visible image fusion and can be roughly categorized into two classes, *i.e.*, fixed-basis and learning-based methods [26]. Methods in the first category adopt fixed bases [136,137] are simple and computationally efficient. Yang et al. constructed two over-complete dictionaries using a fixed basis, *i.e.*, an over-complete discrete cosine transform dictionary and a hybrid dictionary consisting of four different types of bases [136]. Bin et al. adopted a fixed over-complete discrete cosine transform dictionary to represent infrared and visible images [137].

The second category uses learning methods to obtain over-complete dictionaries. These methods are often more flexible and effective than fixed-basis methods [26]. According to the manner in which they adopt the training set, learning-based dictionary construction methods can be roughly categorized into two classes, *i.e.*, globally and adaptively trained dictionary-based methods [26]. Methods in the former class use public training sets to construct over-complete dictionaries [138–144]. These public training images typically have high resolution and quality. Chang et al. used a training set that comprised 100,000 8 × 8 patches from 40 high-quality images to construct an over-complete dictionary [143]. Meanwhile, adaptively trained dictionary-based methods adopt source images to construct over-complete dictionaries [145–154]. For instance, Liu et al. used overlapping patches of source infrared and visible images for training to construct the over-complete dictionary [153].

Learning-based dictionary construction methods, such as K-means generalized SVD (K-SVD) [139,140,142–144,146,149,151,153,154], adaptive sparse representation [138], optimal directions [145], online dictionary learning [147], multi-scale dictionary learning [148], and clustering and PCA [150,152], have been successfully applied in infrared and visible image fusion. The most representative of these dictionary learning methods is K-SVD [155], which alternates between sparse coding and dictionary updating. Only one column of the dictionary is updated at each iteration, and this update is completed by the SVD. Liu et al. proposed an adaptive sparse representation model to learn a series of additionally compact sub-dictionaries from high-resolution and high-quality image patches; their method could reduce visual artifacts and computational cost [138]. Zhang et al. proposed a new dictionary learning algorithm with the Landweber technique and adopted the method of optimal directions for initialization; the computational cost of their model was lower than that of K-SVD [145]. Yao et al. adopted the online dictionary learning algorithm to train their dictionary [147], and their method was considerably faster than batch

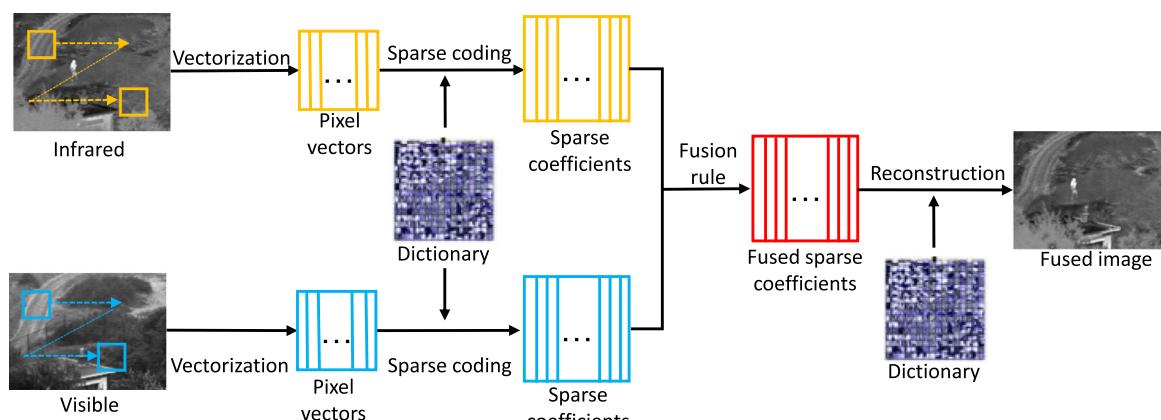


Fig. 3. Sparse representation based infrared and visible image fusion scheme (Credit to [27]).

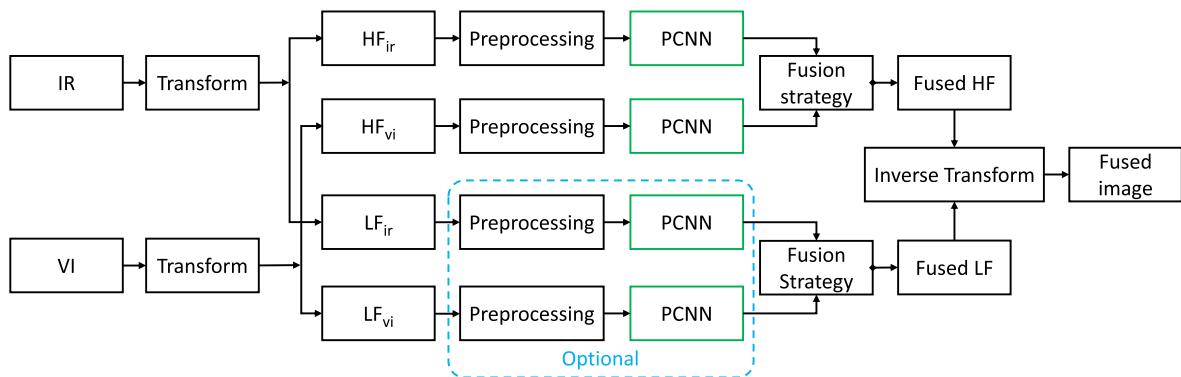


Fig. 4. PCNN based infrared and visible image fusion scheme.

alternatives on millions of training datasets. Yin et al. proposed a multi-scale dictionary learning method by integrating the wavelet into dictionary learning, thus potentially taking advantage of multi-scale representation and dictionary learning [148]. Kim et al. proposed a new dictionary learning method that was based on image patch clustering and PCA [150]; the proposed model could not only remove the redundancy of the learned dictionary but also maintain the infrared and visible image fusion performance.

#### 2.2.2. Sparse coding for obtaining coefficients

After constructing an over-complete dictionary, the representation coefficients can be obtained using sparse coding techniques, such as orthogonal matching pursuit [138,140,142–144,146,148,152], simultaneous orthogonal matching pursuit [136,149–151], joint sparse representation model [139,145,147,153,154], approximate sparse representation with multi-selection strategy [137], and convolutional sparse representation [141].

Orthogonal matching pursuit is a greedy pursuit algorithm created to identify the nonzero sparse coding coefficient [156]. This model has the advantages of computational efficiency and easy implementation and hence has been widely used for sparse coding. For example, Lu et al. adopted orthogonal matching pursuit to obtain the sparse coding of a training sample [146]. Simultaneous orthogonal matching pursuit is a generalization of orthogonal matching pursuit; it considers the coherent geometric structure of input signals [157]. For instance, Yang et al. adopted simultaneous orthogonal matching pursuit to obtain the sparse coefficient of source image patches [136]. Bin et al. proposed an approximate sparse representation with a multi-selection strategy to obtain the possible sparest coefficients [137]. Liu et al. adopted convolutional sparse representation to obtain the sparse coding coefficient [141], this method was a translation and shift-invariant sparse representation that was robust to misregistration.

#### 2.2.3. Fusion rule

In general, most of the fusion rules of sparse representation-based infrared and visible image fusion methods are coefficient combining models or their variants, such as choose-max [136–138,140–143,148,151,152] and weighted average [137,139,141,144–150, 153,154]. For example, Wang et al. adopted the Max- $L_1$  fusion rule to obtain the fused sparse coding coefficient [151]. Alternatively, Liu et al. adopted the weighted average fusion rule to achieve the same goal [153].

### 2.3. Neural network-based methods

A neural network usually consists of many neurons and can thus imitate the perception behavior mechanism of the human brain to deal with neuron information. The interactions among neurons characterize the transmission and processing of neuron information, and the neural network has the advantages of strong adaptability and fault tolerance

and anti-noise capabilities [158]. These merits have caused neural networks to be successfully applied in various fields, such as infrared and visible image fusion [12,13,37,52,53,62,63,75–77,97,104,122, 146,159–161].

Most neural network-based infrared and visible image fusion methods adopt the pulse-coupled neural network (PCNN) or its variants. PCNN was proposed by Eckhorn et al. in the late 1980s [162]; they introduced a neural approach called Eckhorn's model to simulate the synchronous pulse and coupling of neurons in the visual cortices of cats. These discoveries motivated the development and generation of PCNN; i.e., Johnson modified Eckhorn's model and called the modified approach the PCNN [163]. In 1996, Broussard and Rogers applied PCNN to image fusion on the basis of physiologically inspired theoretical models for the first time, thus demonstrating the feasibility and advantages of PCNN [164]. Since then, many infrared and visible image fusion methods that are based on PCNN and its variants have been proposed.

PCNN-based infrared and visible image fusion schemes are usually combined with multi-scale transforms and comprise three steps, as shown in Fig. 4. First, the source images are decomposed into low- and high-frequency subbands. Then, various strategies are utilized to obtain the fused low- and high-frequency subbands. Finally, inverse transform methods are applied to generate the final fused results. Different image fusion methods may adopt different multi-scale transform methods, such as steerable pyramid [37], lifting stationary wavelet transform [52], redundant-lifting non-separable wavelet multi-directional analysis [53], contourlet [62,63], NSCT [12,75–77], curvelet [97], shearlet [104], nonsubsampled shearlet transform [13], and internal generative mechanism [122].

The key in PCNN-based infrared and visible image fusion schemes lies in the use of PCNN in fusion strategies, and many researchers have utilized PCNNs in high-frequency subbands [12,37,52,53,62,63,75, 77,97,104,122,159]. For example, Li et al. presented a modified PCNN model for adjusting high-frequency coefficients in the NSCT domain while adopting the regional variance integration rule to fuse the low-frequency subband coefficients [75]. Xi et al. adopted the PCNN in the high-frequency subband, which was generated by the fast discrete curvelet transform, to simulate the biological activity of the human visual system [97]. In addition, the PCNN can also be adopted to fuse the low- and high-frequency subband coefficients [12,52,57,62,77,104,159]. For example, Chai et al. adopted the lifting stationary wavelet transform to obtain a flexible multi-scale shift-invariant representation of original images and used the dual-channel PCNN to fuse the high- and low-frequency subband coefficients [52].

Approaches other than the combined multi-scale transform and PCNN-based infrared and visible image fusion scheme can be adopted to apply PCNN in image fusion. For example, Chen et al. adopted the PCNN in the compressed domain [160], and Lu et al. adopted the PCNN to segment the source image and fuse the target coefficient [146].

Furthermore, certain fusion methods are based on the modified PCNN model [13,76]. Kong et al. proposed a unit-fast-linking PCNN to fuse high- and low-frequency subband coefficients, thereby exploiting the advantages of unit- and fast-linking PCNNs [76]; they also proposed a spatial-frequency PCNN to fuse high- and low-frequency subband coefficients, which were decomposed using the nonsubsampled shearlet transform [13].

In recent years, deep learning has demonstrated state-of-the-art performance in various fields, such as computer vision, pattern recognition, and image processing. Deep learning has a strong capability of modeling complicated relationships among data. Moreover, deep learning can automatically extract distinctive features from data without manual intervention [165]. Thus, deep learning has also been successfully applied to image fusion, such as multi-focus image fusion [166,167], multi-modality image fusion [141,168,169], remote sensing image fusion [170,171], as well as infrared and visible image fusion [172]. To the best of our knowledge, there is only one deep learning-based fusion method specialized for infrared and visible image fusion. Specifically, Liu et al. proposed a method based on convolutional neural networks, which can deal with activity level measurement and weight assignment in infrared and visible image fusion as a whole to overcome the difficulty of manual design [172]. Nevertheless, there are several general image fusion methods that can be applied to infrared and visible image fusion. For example, Liu et al. introduced the convolutional sparse representation for image fusion [141]; they derived the idea from deconvolutional networks [173], which aim to build a hierarchy of layers that each comprise an encoder and a decoder. Zhong et al. proposed a joint image fusion and super-resolution method that was based on a convolutional neural network (CNN) trained using 90 selected images from the ImageNet database [168].

#### 2.4. Subspace-based methods

Subspace-based methods aim to project high-dimensional input images into low-dimensional spaces or subspaces. For most natural images, redundant information exists and low-dimensional subspaces can help capture the intrinsic structures of the original images. In addition, processing low-dimensional subspace data consumes less time and memory than high-dimensional input images. Moreover, low-dimensional subspace representation can be used to improve generalization [174]. Thus, subspace-based methods, including PCA, ICA, and NMF, have been successfully applied in infrared and visible image fusion.

##### 2.4.1. Principal component analysis

PCA aims to transform possibly correlated variables into uncorrelated variables called principal components, thereby possibly reducing dimensionality and maintaining the information of the original data simultaneously. In addition, PCA can reduce redundant information and highlight similarities and differences [175]. Thus, PCA has been applied in infrared and visible image fusion [14,61,176–179]. For example, Li et al. adopted PCA to fuse low-frequency images decomposed by morphology-hat transform [61]; energy information exists mainly in low-frequency images, and PCA was adopted to maintain the brightness information of the original images. Bavirisetti et al. used PCA to fuse detail images decomposed by fourth-order partial differential equations [14]; their method could obtain the optimal weights for transferring the detail information to the final fused image.

Although PCA has been widely used in different fields, it is sensitive to gross errors, which are ubiquitous in many applications. Robust PCA (RPCA) overcomes the abovementioned problems by decomposing a matrix into low-rank and sparse components [180]. In image fusion, RPCA can be used to retain useful information and remove sparse noise [181–183]. For example, Fu et al. adopted the RPCA to obtain the sparse components of infrared and visible images, which were used to obtain the fused low- and high-frequency subband coefficients [182].

Wang et al. used the RPCA to determine the fused low-frequency subband coefficients in their work [183].

##### 2.4.2. Independent component analysis

ICA is an extension of PCA and aims to transform possibly correlated variables into uncorrelated and independent variables [174]; it has also been successfully applied in infrared and visible image fusion [184–187]. ICA-based image fusion methods usually train a set of bases using several natural images that have similar contents as the images to be fused. After obtaining a set of pretrained bases, they can be fused with those with similar contents.

Cvejic et al. proposed a region-based ICA image fusion method [184] where the source images are segmented into different regions and the ICA coefficients are obtained for each region using pretrained bases; the ICA coefficients can be subsequently weighted with the Piella fusion metric according to the maximizing fused image quality criterion. Mitianoudis et al. adopted ICA and topographical ICA bases for image fusion under pixel- and region-based fusion schemes [185]. The topographical ICA bases demonstrated a better directional selectivity than the traditional ICA bases and hence captured image features better. In addition, the authors proposed an image fusion method that was based on self-trained ICA bases; this proposed approach adopted k-means clustering and hierarchical grouping to improve ICA image fusion performance under the region-based fusion scheme [186]. Omar et al. proposed a new image fusion method by combining ICA and the Chebyshev polynomial approximation [187]; the authors used ICA to capture the salient feature and the Chebyshev polynomial to remove noise, especially in seriously noise-degraded images.

##### 2.4.3. Non-negative matrix factorization

NMF aims to decompose the original data matrix into a product of two non-negative matrices. This approach is a part-based object representation model [188] that is consistent with the human perception mechanism. Given these merits, NMF has been applied in infrared and visible image fusion [15,189–191].

Mou et al. adopted the NMF to extract the feature bases of source images [191], thereby maintaining the details in the visible image, highlighting the features in the infrared image, and removing the noise. However, traditional NMF is time consuming and has low efficiency. To overcome these problems, Kong et al. adopted the improved NMF, which could overcome the drawback of SVD-based random initialization and hence converge faster than traditional NMF with random initialization [190]; moreover, they adopted fast NMF to fuse low- and high-frequency sub-images decomposed by the nonsubsampled shearlet transform; fast NMF is less complex and more computationally efficient than traditional NMF [15].

#### 2.5. Saliency-based methods

Saliency attracts human visual attention in a bottom-up manner [192], this attention is often captured by objects or pixels that are more significant than their neighbors. According to the mechanism of the human visual system, saliency-based fusion methods can maintain the integrity of the salient object region and improve the visual quality of the fused image. Thus, saliency-based methods have been widely used for infrared and visible image fusion.

In recent years, researchers have adopted saliency-based methods for infrared and visible image fusion in two ways, *i.e.*, weight calculation [16,19,79,80,85,87,88,109,129] and salient object extraction [65,107,153,193–195]. In the first case, saliency is adopted for the reconstruction of the fused image. Infrared and visible images are decomposed into a base layer and a detail layer through a multi-scale transform, and saliency extraction models are applied in the base or/and detail layer of the source image to obtain the saliency map. Then, the weight map can be obtained by the saliency map to obtain the fused

base image or/and the detail image, and the fused image can be subsequently constructed through the fused base image and the detail image. For example, Gan et al. adopted phase congruency to obtain the saliency maps of the base and detail layers of the source images and then applied guided filtering on the saliency maps to obtain the weighting maps [87]. Ma et al. initially adopted the rolling guidance and Gaussian filters to decompose the infrared and visible images into the base and detail layers and then used the improved visual saliency map and weighted least square optimization to fuse the base and then detail layers, respectively [19].

Saliency can also be used to extract the significant object regions of source images, thereby obtaining good visual effects and preserving image details. Thus, this model is highly suitable for surveillance applications, such as target detection and recognition. For example, certain research works in [65,107,193] adopted saliency models to extract the salient regions of infrared images. Meng et al. adopted the saliency detection method to extract the salient object region, which can be mapped to the region of the final fused image. Zhang et al. used super-pixel-based saliency models to obtain the salient regions of infrared images; the models could preserve the target information of the infrared images well [107]. In addition, several research works [153,194,195] have adopted saliency models to extract the salient objects and information from infrared and visible images. Liu et al. integrated saliency detection into the sparse representation framework for infrared and visible image fusion, which adopted global and local saliency maps to obtain the weight for fused image reconstruction [153]. Shibata extracted salient information from infrared and visible images, thereby preserving their salient information [195].

## 2.6. Hybrid methods

The abovementioned infrared and visible image fusion methods all have their advantages and disadvantages, and their advantages should be combined to improve image fusion performance. The most common hybrid infrared and visible image fusion methods integrate other types of methods into the multi-scale transform framework, such as hybrid multi-scale transform and saliency [16,19,65,79,80,85,87,88,101,103,107], hybrid multi-scale transform and sparse representation [18,48,73,148,196,197], hybrid multi-scale transform and neural network [13,53,57,62,63,104], and hybrid multi-scale transform, sparse representation, and neural network methods [74,108,198].

Hybrid multi-scale transform and saliency methods aim to integrate salient region detection into the multi-scale transform image fusion framework, which can enhance the region of interest and preserve the details. For example, Zhang et al. adopted the local edge-preserving filter to decompose the source images into the base and detail layers and used the detected salient targets of the infrared image to determine the weight of the base layer [16]. In addition, sparse representation and neural network are often integrated into the multi-scale transform infrared and visible image fusion framework. For example, Liu et al. proposed a general image fusion framework by combining multi-scale transform with sparse representation [18] and adopted sparse representation to obtain the fused low-frequency subband coefficients. Lin et al. proposed a new infrared and visible image fusion method that was based on the contourlet transform and an improved PCNN; the contourlet transform was used to decompose the source images into multi-scale sub-images, and the improved PCNN was utilized to fuse the high- and low-frequency subband coefficients [62]. Yin et al. proposed a new hybrid multi-scale transform, sparse representation, and neural network fusion method, where a novel shift-invariant dual-tree complex shearlet transform is adopted to decompose the source images into multi-scale sub-images, sparse representation is adopted to obtain the fused low-frequency subband coefficients, and the adaptive dual-channel PCNN was adopted to obtain the fused high-frequency subband coefficients [108].

Certain hybrid infrared and visible image fusion methods are

classified under the sparse representation framework, such as hybrid sparse representation and saliency methods [139,153] and hybrid sparse representation and neural network methods [143,146]. For example, Liu et al. integrated saliency detection into the sparse representation framework [153], constructed the saliency map on the basis of the sparse representation coefficients, and then obtained the fused sparse representation coefficients through the saliency map. Chang et al. proposed a new image fusion method that was based on a combination of sparse representation and a deep neural network [143] to combine the external and internal cartoon and texture sparse representation results.

## 2.7. Other infrared and visible image fusion methods

Other infrared and visible image fusion methods can inspire new ideas and perspectives for image fusion. Ma et al. proposed a series of infrared and visible image fusion methods that was based on total variation [20,199,200]. The main idea of a previous work [20,199] was to retain the intensity information of infrared images and preserve the appearance information of visible images simultaneously. However, the authors ignored the intensity information of the visible image, which could lead to a low dynamic range and detail loss. In [200], Guo et al. improved fusion performance by not only keeping the intensity information of both infrared and visible images but also preserving the appearance information of the visible image.

High-quality image fusion is a fuzzy problem [201], and fuzzy theory is an effective tool for solving the fuzzy problem through fuzzy logic or fuzzy based methods. Therefore, fuzzy theory has been successfully applied in infrared and visible image fusion [25,49,66,202–204]. This theory is often adopted to determine the fusion rules for infrared and visible image fusion [25,49,66,202]. For example, Yin et al. adopted fuzzy logic to obtain the adaptive weighted average rule [66]. The fuzzy logic system is also applied to infrared and visible image fusion [203], and the key concept of the fuzzy logic system is the membership function that ranges from 0 to 1, which can be adopted to measure the uncertainty of image fusion performance. Bai et al. also proposed a new infrared and visible image fusion rule based on the fuzzy measure weight strategy [204].

Other types of infrared and visible image fusion methods exist, such as entropy [21,205], Markov random field [193,206], morphology [207,208], and infrared feature extraction and visual information preservation [209]. Entropy is an objective metric used to measure the amount of information transferred from the source images into the fused image. Zhao et al. adopted the global maximum entropy criterion to transfer as much information from the source images into the fused image [21]. Markov random field is also an effective tool for infrared and visible image fusion. For example, Han et al. adopted the Markov random field for saliency detection, which could be formulated as a pixel labeling problem [193]. Shibata et al. used the Markov random field to preserve the visible information and spatial coherence of the source images [206]. Morphological methods have also been applied to image fusion. Specifically, Bai et al. adopted morphological methods to extract multi-scale morphological features for infrared and visible image fusion [207,208].

Most of the above mentioned infrared and visible image fusion methods are performed at pixel-level, which combine each pixel in source images to obtain the fused image directly. However, image fusion can also be performed at feature- or decision-level. The feature-level image fusion methods usually extract the features of source images at first, and then obtain the fused image based on the extracted features for some specific purposes, which have the advantages of designing more intelligent semantic fusion rules based on the extracted features compared with pixel-level fusion strategies [210]. Therefore, feature-level fusion methods have also been adopted for infrared and visible image fusion [72,191,207,209,211,212]. For example, Li et al. proposed a fusion scheme based on nonsubsampled contourlet

transform and low-level visual features [72]. Due to the fact that the human visual system perceives image quality mainly according to some low-level features, they designed two activity measures of lowpass and highpass subbands decomposed by nonsubsampled contourlet transform based on the low-level features. Alternatively, Bai et al. proposed a fusion method through feature extraction by morphological sequential toggle operator [207]. Sequential toggle operator is used to extract the features of source images, and then multi-scale features of the original infrared and visible images are transferred into the fused image based on multi-scale morphological theory.

Decision-level image fusion is the highest level of the above mentioned three levels, which can also be called interpretation- or symbol-level image fusion. In decision-level image fusion, the information is extracted from each source image, and then decision rules are made to fuse the extracted information according to specific criteria [213,214], which can help to better understand the objects under observation. The research about decision-level image fusion is the least among the above mentioned three image fusion levels, and the main application of decision-level infrared and visible image fusion is face recognition [215–217]. For example, Neagoie et al. proposed a real-time face recognition system using the decision fusion, which can combine the recognition scores generated from visual channels with a thermal infrared neural classifier [215]. Zhao et al. proposed a decision-level method for face recognition, in which the principal component analysis and linear discriminant analysis were adopted to extract the face features, and the decision-level fusion was implemented with the infrared and visible image recognition results and their confidence measures [216].

## 2.8. Image registration

The integration of multi-sensor images can provide a complex and detailed scene representation, thereby increasing the accuracy of decision-making in subsequent tasks. However, a successful image fusion requires the images to be fused to be strictly geometrically aligned [218,219]. This problem is known as image registration [220,221], which occurs because the images captured by different sensors, such as infrared and visible sensors, are usually different in size, perspective, and field of view. Such an alignment can be achieved through hardware (e.g., placing a beam-splitter) [222]; however, a special imaging device for generating aligned image pairs may not be practical due to its high cost and low availability, especially for real-world applications, where large-scale deployments are required [219]. Therefore, using an image registration algorithm based on off-the-shelf low-cost visible and infrared cameras may be appropriate.

Image registration has been widely used in many fields, including computer vision [223,224], medical image analysis [225,226], and remote sensing [227,228]. The images to be aligned may be captured at different times from different perspectives or with different modalities, which typically refer to the problems of temporal, spatial, and multimodal registrations. Specifically, the registration of infrared and visible images refers to the problem of multimodal registration. A broad overview of the registration methods can be found in the literature [220,224,225]. In general, registration methods can be classified into two categories, i.e., area-based and feature-based methods. Next, we briefly present an overview these two types of techniques, particularly in the context of infrared and visible image registration.

### 2.8.1. Area-based methods

Area-based methods deal directly with the intensity values of entire original images; for example, minimizing the total distance between the pixel correspondences under a certain metric. These methods are preferable in cases of few prominent details in images where distinctive information is provided by pixel intensities rather than by salient structures, but they suffer from image distortions, illumination changes, and heavy computational complexities. Area-based methods come in

three major types, i.e., correlation-like methods [229], Fourier methods [230], and mutual information (MI) methods [231].

Cross-correlation is a classic correlation-like method [229] that calculates the cross-correlation between window pairs in two images, and the one with the maximum value is considered a correspondence. To address the multimodal registration problem, rather than using standard image statistics, Roche et al. introduced a similarity measure that was based on probability theory, i.e., the correlation ratio, which had perceptible practical improvements in registration accuracy and robustness [232]. Correlation-like methods are frequently used in real-time applications because of their easy implementation in hardware. However, they have disadvantages when the image contents in windows suffer from complex transformations or contain a relatively smooth area without prominent details.

Fourier methods exploit the Fourier representation of images in the frequency domain [230]. One representative of such method is the phase correlation method based on the Fourier shift theorem, which was originally proposed to deal with image translation and later extended to consider rotation and scaling [233]. The applications in infrared and visible image registration were described in [234], where the Fourier-Mellin transform was used to address the multimodal nature together with spatial distortion issues. In comparison with correlation-like methods, Fourier methods have the advantages of computational efficiency and strong robustness against varying imaging conditions and frequency-dependent noise.

Mutual information methods utilize mutual information as the metric to maximize the statistical dependency between two given images and then establish pixel correspondences accordingly [231]. These models work well in domains where edge or gradient magnitude-based methods have difficulty and hence are particularly suitable for registering multimodal images. However, for infrared and visible image pairs, the textures are usually significantly different; mutual information may then work well on only a small part rather than on the entire image [235]. Therefore, some salient regions, such as regions with a similar edge density or a detected foreground, are selected before applying mutual information [236]. Moreover, mutual information can be operated on extracted features instead of image intensities (e.g., points sampled from shape contours [237]).

### 2.8.2. Feature-based methods

Feature-based methods first extract two sets of salient structures (e.g., feature points) and then determine the correct correspondence between them and estimate the spatial transformation accordingly, which is further used to align the given image pair. Compared with area-based methods, feature-based methods are more robust against typical appearance changes and scene movements and are potentially faster if implemented correctly [218]. In general, feature-based methods comprise two major steps, i.e., feature extraction and feature matching. The former refers to the detection of salient and distinctive objects, whereas the latter refers to the establishment of correspondences between the detected features.

The first step of feature-based methods is to extract robust common features that can represent the original images. For infrared and visible images that manifest two different phenomena, the appearance features with global statistical dependence, including gray levels/colors, textures (e.g., Gabor filters [238]), and gradient histograms (e.g., SIFT [239] and HOG [240]), are unlikely to match. Instead, the features that represent salient structures are preferred, such as corner points [241], edge maps [218], silhouettes [242], trajectories [235], visual salient features [212], and hybrid visual features [243]. Among these features, edge information is one of the most popular selections for infrared and visible image registration [218,242,244] as its magnitudes and orientations may be preserved well across different modalities. A typical example of infrared and visible image registration based on edge information is presented in Fig. 5.

After obtaining the salient features from two images, we need to

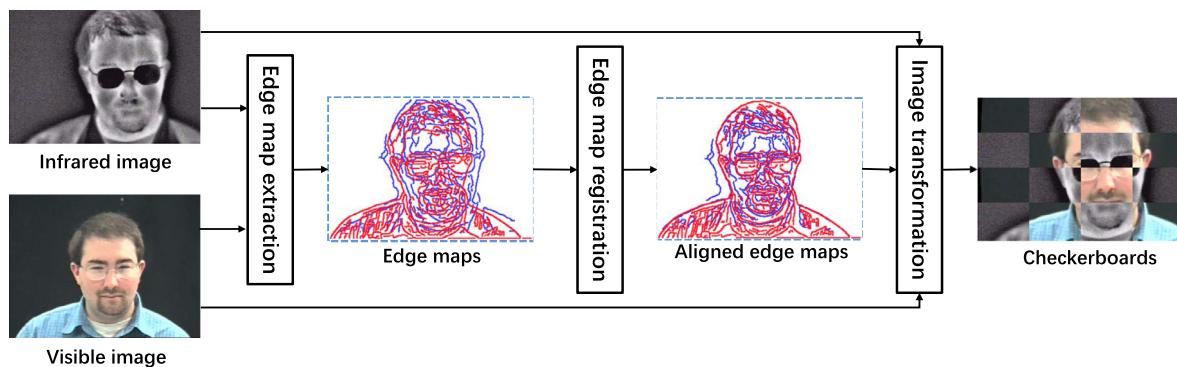


Fig. 5. Infrared and visible image registration based on edge maps.

establish correct correspondences between them. In general, high-level features, such as edge maps and surfaces, can be discretized as point sets. Therefore, the registration problem is lessened to determine the correspondence and estimate the transformation between two sets of point features [218]. A popular strategy for solving the point matching problem involves two steps [245]: (i) computing a set of putative correspondence, and then (ii) removing the outliers via geometrical constraints. Putative correspondence instances are obtained in the first step by pruning the set of all possible point correspondences. This scenario is achieved by computing the feature descriptors [239,246] at the points and eliminating the matches between points whose descriptors are excessively dissimilar. However, the use of local feature descriptors alone will inevitably result in numerous false matches. In the second step, robust estimators based on geometrical constraints are used to detect and remove the outliers.

To remove false matches from putative sets, numerous methods have been proposed over the last decades. One of the most widely used robust estimator is random sample consensus (RANSAC) [247], which adopts a hypothesize-and-verify approach and attempts to obtain the smallest possible outlier-free subset to estimate a given parametric model through resampling. The RANSAC algorithm has achieved success in the registration of infrared and visible images [212,235,243]. However, this model relies on a predefined parametric model, which becomes less efficient when the underlying image transformation is non-rigid; it also tends to degrade severely when the proportion of false matches becomes large. To address these issues, several non-parametric interpolation methods [245,248] have been introduced recently, especially in the context of infrared and visible image registration [218,242]. Specifically, to deal with the expression and pose changes in face recognition, Ma et al. proposed a regularized Gaussian fields criterion to align infrared and visible faces in a non-rigid manner [218]. Furthermore, the authors introduced a sparse approximation for non-rigid transformation estimation that could reduce the computational complexity from cubic to linear without sacrificing registration accuracy [249].

### 3. Applications

Fusion images produced by combining infrared and visible source information have many complementary characteristics. For example, these images make the infrared image more vivid through the transfer of certain textures. Therefore, fusion algorithms can be embedded into many applications and improve the original methods. Main applications that use fusion techniques include recognition [118,250,251], object detection [4,252,253], tracking [254,255], surveillance [6,7,256], color vision [257,258], and remote sensing [259,260]. In this section, the algorithms for applications in these techno-spheres are introduced for a comprehensive understanding of their peculiarities and performances.

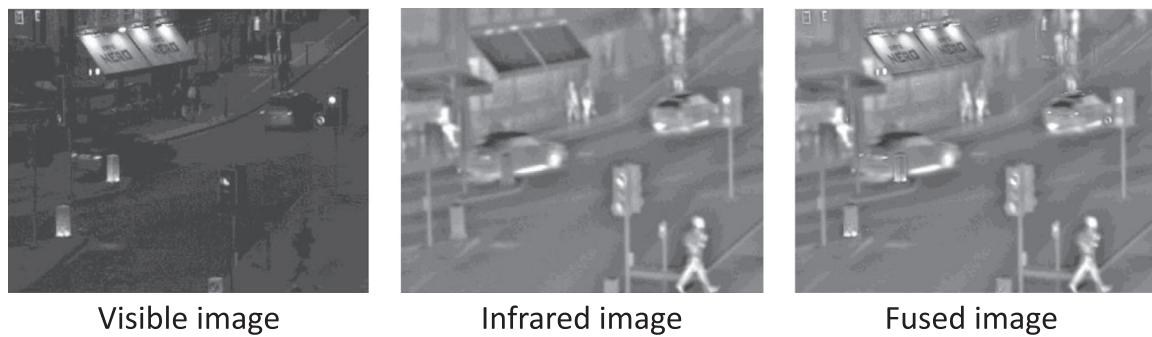
#### 3.1. Fusion for recognition

Given an image that contains one or some objects, recognition aims to specify the category of the target object. Face recognition is one of the major applications in recognition that employs infrared and visible fusion. Recognition algorithms that use infrared and visible image fusion come in two types. The first class is two-stage recognition: fusing first and then recognizing with the fused result. In the second category, the fusion algorithm is embedded into the recognition process in which distinguishing the boundary between the two processes is difficult. The algorithm proposed by Faten et al. is a representative of the first class [261]. It solves the problem of near infrared and visible image fusion in face recognition, in which identifying a person at a distance rather than at close range is often necessary. This model uses the wavelet transform to decompose the original face images and then applies SVD and PCA in the fusion process. This method can improve recognition precision by using narrow-band images to enhance the fusion result. Other methods of this class usually change the decomposition algorithms or the fusion rules to achieve enhanced visibility for a specific situation. For example, Heo et al. employed weighted average [251], and in [118] intrinsic mode functions were used instead of the discrete wavelet transform to decompose the original images. Similar methods include [250,262,263]. Meanwhile, Richa et al. proposed a fusion algorithm with classifiers [3]. In this method, 2v-granular SVMs are used to learn the global and local features of the original images at several resolutions, which are used to decide the relative weights between the infrared and visible images. Then, the fused image is simultaneously processed by 2D log polar Gabor feature matching and local binary pattern feature matching. Finally, the results are judged by the Dezert-Smarandache match score fusion, which determines whether the candidate is identified successfully. Compared with other algorithms, this method demonstrates significant superiority even when visible and infrared images provide conflicting results under severe interference.

Other biological recognition applications that are related to infrared and visible image fusion exist. A person identification algorithm based on periocular region [264] was proposed to solve the sever occlusion problem in surveillance. In this method, an image set consists of infrared and visible images instead of single-band data and feature score fusion is used. This algorithm can identify a person with only a periocular region available. An algorithm for iris recognition [265] was developed with the Haar wavelet method; in this approach, fusion is performed at the matching score level.

#### 3.2. Fusion for detection and tracking

Detection based on infrared and visible image fusion is widely used in many real-life applications, such as general object detection [266], pedestrian detection [267], and fruit detection [32]. Different from recognition, it should determine the accurate position of the objects at



**Fig. 6.** Example of surveillance with infrared and visible image fusion technique.

the same time. According to the implementation process, detection methods based on fusion algorithms can be mainly divided into two classes: (1) those that fuse before detecting and (2) those that detect before fusing. Many algorithms adopt the former style. Emna et al. proposed a two-stage method that includes multi-scale decomposition, low-level fusion, and background modeling detection [268]. The main idea of this algorithm is to detect objects through the inter-frame differences in the background model of the fused image. This algorithm is fast because of the simplicity of the fusion process; however, it suffers from degradation under the condition of moving objects, which may cast a shadow and introduce noise from the changes in illumination. He et al. proposed an algorithm with multi-level fusion and enhancement for target detection [213]. In this method, both pixel- and feature-level fusion are considered and the relationship between low-frequency information and high-frequency component, which is usually ignored in the wavelet transform fusion, is explored to enhance target visibility. This model exhibits superiority in terms of fusion details and detection precision. In addition, weak infrared targets that cannot be found with common algorithms can be detected with this method. As for the second class, the people detection algorithm based on the INT3 framework [269] proposed by Jose et al. is a typical example [270]. In this framework, the input source images are segmented and then fused at the pixel level, thereby generating possible candidates for the target objects. Following this step, a detection algorithm, such as deformable part-based model, is performed to search the targets. Although detection is behind fusion superficially, the segmentation process for candidates that is implemented before fusion is an important part of detection. Besides detection and recognition, popular vision tasks, such as tracking or event interpretation, can be embedded in the method. Therefore, the framework provides a unified solution for multi-task implementation that is based on the fusion results of infrared and visible images.

Target tracking algorithms share some similarities with detection, it should determine the relationship between frames and detect the target object in a time sequence. In target tracking, the algorithms should be implemented quickly because tracking is performed with video or images of time sequence in most situations. Therefore, multi-scale decomposition-based methods are usually used in the fusion process. Stephen et al. proposed an enhanced target tracking approach through infrared-visible image fusion by applying a PCA-weighted fusing rule instead of a simple combination, such as average or maximum selection after pyramid decomposition [271]. A target tracking system based on infrared and visible image fusion and used in unmanned aerial vehicles was designed with wavelet transform and fuzzy inference fusion [272]. In addition to the multi-scale decomposition-based method, algorithms that use sparse representation [255,273] to achieve fusion are important in target tracking. The key to this kind of method is to find an appropriate dictionary with a powerful representation capability for the specific task.

In most algorithms for recognition, detection, and tracking, fusion is a relatively independent part that is designed to improve the visibility

or textures of the original images. Different from some end-to-end methods, such as CNNs, suitable recognition or detection algorithms should be selected to produce superior fused images in a specific target task.

### 3.3. Fusion for surveillance

As shown in Fig. 6, surveillance need clear, robust, and expressive images or videos. The complementary properties of infrared and visible image make them a good fusion sources for this task. Infrared and visible images provide object information in various respects given their different imagery characteristics. Infrared images capture the thermal radiation of objects that is uninfluenced by the changes in illumination, weather, and other disturbances. However, infrared images have low spatial resolution and are therefore hardly distinguish details. By contrast, visible images have high spatial resolution and are rich in color, which make objects in them vivid and easily identified. The fusion of these two images significantly improves visibility and contributes to several real-life applications. Surveillance methods usually deal with videos, and they are required to be implemented fast. Therefore, fusion algorithms designed for surveillance applications are usually developed in a multi-scale fashion. Liu et al. used an infrared and visible image fusion technique to improve the visibility of unmanned aerial vehicles with discrete wavelet transform and region segmentation [45]. Framelet transform was implemented in [274], NSCT was applied to the decomposition process for a night-vision surveillance system [275], and cascade wavelet transform and Robert operator were used in a surveillance system with multiple sensors using hybrid fusion algorithms [276]. Fusion models designed from other respects also exist. Ma et al. solved the fusion problem in another perspective and proposed a fusion method [20] that could transform a low-resolution infrared image to a high-resolution one by employing total variation and gradient transfer. The algorithm could be applied to images without registration and worked efficiently in face image enhancement. An improved version [200] of this model was subsequently proposed by providing additive information about intensity from visible images. The fused image generated by this algorithm was natural and rich in details. The algorithm could be implemented fast by solving the fusion problem in the framework of alternating direction method of multipliers.

### 3.4. Fusion for color vision

The human visual system can perceive only visible light. Therefore, information of thermal radiation from target objects or scenes must be converted into images with pseudo-color. No suitable rendering scheme has been used for a long time, and fusion results have been represented by gray-level images. This method degrades surveillance performance because the human visual system can distinguish more than a thousand colors but can only discern approximately 100 levels of gray intensity [277]. The color transfer [5] proposed by Reinhard et al. was a

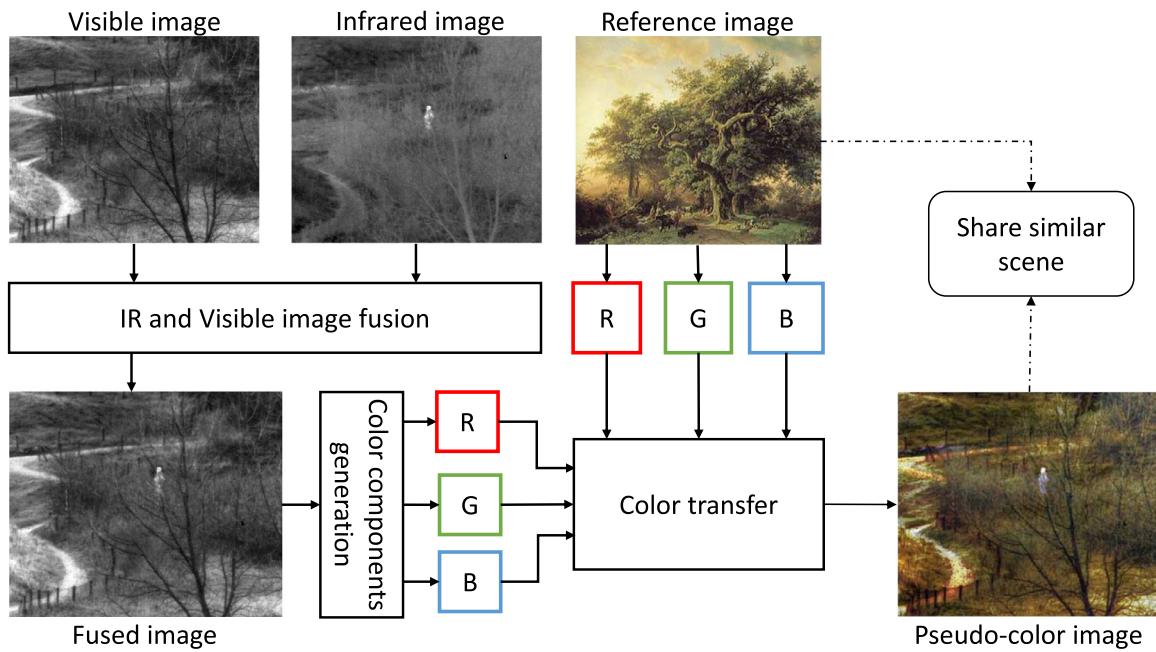


Fig. 7. Pseudo-color rendering in image fusion.

breakthrough in this field. In the method, statistics of natural images in the principal component space is transferred to a multi-band image to make the gray image have a natural appearance. The multi-band image is required to show a similar scene to the source image lest it exhibits a strange appearance. The process of this method is presented in Fig. 7. As for infrared and visible image fusion, the original images used are from the same scene; therefore, such a fusion can generate satisfactory results. Subsequent research improving this transfer scheme has been conducted. Toet et al. converted the RGB tristimulus into XYZ tristimulus and used  $la\beta$  transform to achieve color transferring while optimizing luminance contrast [278,279]. Tsagaris et al. used NMF instead of PCA to find an additive part-based representation in the fusion method [189]. Cheng et al. proposed a pseudo-color rendering algorithm by developing a deep neural network that was totally different from all previous works [280]. The algorithm combined high-level features extracted by the neural network and adaptive image clustering to improve colorization performance and was better than previous algorithms developed with color transferring. However, this method depended on a large training dataset to learn a powerful representation. Good results were difficult to obtain in applications without sufficient training samples.

### 3.5. Fusion for other applications

Infrared and visible image fusion techniques are also widely used in other fields. Typical areas include remote sensing, culture relic analysis, and augmented reality.

Remote sensing that uses infrared and visible image fusion can be used under numerous conditions. Li et al. proposed a medium-altitude unmanned aerial vehicle remote sensing system using image registration and fusion [260]. Particle swarm optimization was introduced to improve efficiency, and the fusion method was achieved by a PCNN. An adaptive fusion method for infrared and visible remote sensing images was developed with multi-contourlet transform [281]. This method was efficient in direction selectivity, and considerable information about details could be captured in the fusion process. A similar method [282] was proposed to solve the fusion problem in astronomical images. Instead of multi-contourlet undecimated dual-tree complex wavelet transform, which exhibited shift-invariance, was applied in the fusion process to reduce noise. Lu et al. proposed a synergistic classification

method for long-wave infrared hyperspectral and visible images [283]. In this method, a new feature extraction method named semi-supervised local discriminant analysis was proposed in infrared image classification. Unlike other methods, which fuse at the pixel level, this method fused at the decision level for infrared and visible images. Applications developed on remote sensing with fusion techniques, such as urban object detection [284] and geostationary meteorological satellite image fusion [285], are also presented.

Infrared and visible image fusion is also used in areas other than remote sensing. Liang et al. proposed a polarimetric dehazing method for visibility improvement that was based on visible and infrared image fusion [286]. The method used a multi-scale directional nonlocal means filter to achieve high robustness and minimal error points. However, the algorithm implementation consumed considerable time and was hence inappropriate for certain real-time applications. Li et al. proposed a fusion method for deterioration risk analysis of ancient frescoes [273]. A fusion technique for color and thermal images was developed for augmented reality in rescue robotics [287]. In the agriculture field, techniques for predicting apple fruit firmness and soluble solid content have been proposed in recent years [288]. Fusion algorithms for infrared and visible videos in the field of medical imaging have helped improve the clinical intraoperative assessment of critical biliary structures [289].

Current infrared and visible image fusion algorithms are mostly developed at the pixel level. Although fused images can provide complementary information about scenes, distortions, such as halos, still occur in many algorithms because of inappropriate fusion rules. These undesired changes limit or degrade the performances of applications built on the corresponding fusion method. Methods with good fusion principles and powerful representations that combine pixel-level semantic features will significantly improve current applications and contribute to inventions of advanced systems in different areas.

### 3.6. Implementations

Many fusion algorithms involve large computation and require considerable time to be implemented on machines for general computation purposes. However, numerous real-life applications, such as video surveillance, that are based on infrared and visible image fusion require that the algorithm be efficient to achieve a satisfactory

performance. Thus, implementation methods involving special purpose computers have been designed to accelerate the process. Implementation schemes come in two main types: field-programmable gate array (FPGA)-based architecture and compute unified device architecture (CUDA)-based implementation.

FPGA has many advantages over general purpose computers; it has a low power/dissipation ratio and a massive processing throughput and is therefore energy efficient and can be designed for parallel computing [290]. The programmable characteristic of a chip also enables the development of a compact and portable fusion system that can be used in such platforms as unmanned aerial vehicles. Only a few attempts in designing fusion systems using FPGA have been documented. Uninhabited airborne vehicles that use the wavelet transform are probably the first systems developed on FPGA [291]. Currently, a common fusion system developed on FPGA has been designed with three-level pyramid transform and fast and adaptive bidimensional empirical mode decomposition (FABEMD) [290,292]. This design was pursued because the characteristics of decomposition and fusion in pyramid transform and FABEMD make the model suitable for parallel computing in FPGA. In addition to modules related to decomposition and fusion, a local buffer for image input has been designed to meet the requirement of immediate access to data. A fusion scheme that uses a linear transformation named Cholesky decomposition has also been developed. This method has been implemented in a system developed with ALTERA Cyclone II FPGA in VHSIC hardware description language [293]. Another typical method for accelerating implementation is through CUDA. Literature about related architecture is limited. A multi-scale method based on discrete wavelet transform was introduced in [294]. However, similar architectures of other fusion algorithms implemented with graphic processing units can be good references [295–297].

#### 4. Performance evaluation of image fusion

Infrared and visible image fusion has been attracting considerable attention in information fusion in the past few decades, and several image fusion methods have been proposed. Infrared and visible image fusion techniques have been widely used in different applications, such as object recognition, tracking, and surveillance, due to the progress of image fusion techniques. However, different fusion methods have different characteristics, and the performances of fusion methods in real-life applications greatly rely on the quality of the fused images. Therefore, the fusion performances of different methods should be evaluated qualitatively and quantitatively [298].

Numerous quality evaluation methods have been proposed to evaluate the performances of different infrared and visible image fusion methods, and they can be classified as subjective and objective evaluation methods [299]. Quality evaluation methods can be used to compare the performances of different fusion methods and adopted as a guide to select fusion methods in real applications. Evaluation methods can also be used to tune the parameters of fusion methods.

Subjective evaluation methods assess the quality of fused images on the basis of the human visual system and play an important role in fusion quality evaluation. Subjective human inference can consistently compare different fusion methods according to several criterions, such as image details, object completeness, and image distortion; hence, subjective evaluation methods are popular, reliable, and direct in the quality evaluation of infrared and visible image fusion. Nevertheless, the simplest subjective evaluation method is to score fused images by trained observers, which has the drawbacks of human intervention, considerable time consumed, high cost, and irreproducibility [299]. Objective evaluation methods that can measure the qualities of fused images quantitatively and automatically have been proposed to overcome these problems.

Unlike subjective evaluation methods, objective evaluation methods can assess fused image quality quantitatively; they are highly consistent

with the human visual perception and not easily biased by observers or interpreters. Objective evaluation methods come in different types, which are based on information theory, structure similarity, image gradient, statistics, and human visual system. In this section, we briefly introduce some representative image quality evaluation measures, which are listed in Table 1. We use  $A$ ,  $B$ , and  $F$  to represent the infrared, visible, and fused images, respectively, and  $X$  to represent the source image variables that have a size of  $M \times N$ .

##### (1) Entropy

Entropy (EN) measures the amount of information contained in a fused image on the basis of information theory [313]. EN is mathematically defined as follows:

$$EN = - \sum_{l=0}^{L-1} p_l \log_2 p_l, \quad (1)$$

where  $L$  denotes the number of gray levels and  $p_l$  is the normalized histogram of the corresponding gray level in the fused image. The larger the EN, the more information is contained in the fused image and the better the performance of the fusion method. However, EN may be influenced by noise; the more noise the fused image contains, the larger the EN. Therefore, EN is usually used as an auxiliary metric.

##### (2) Mutual information

The mutual information (MI) metric is a quality index that measures the amount of information that is transferred from source images to the fused image [314]. MI is a fundamental concept in information theory and measures the dependence of two random variables. The MI fusion metric is defined as follows:

$$MI = MI_{A,F} + MI_{B,F}, \quad (2)$$

where  $MI_{A,F}$  and  $MI_{B,F}$  denote the amount of information that is transferred from infrared and visible images to the fused image, respectively. The MI between two random variables can be calculated by the Kullback–Leibler measure, which is defined as follows:

$$MI_{X,F} = \sum_{x,f} p_{X,F}(x,f) \log \frac{p_{X,F}(x,f)}{p_X(x)p_F(f)}, \quad (3)$$

where  $p_X(x)$  and  $p_F(f)$  denote the marginal histograms of source image  $X$  and fused image  $F$ , respectively.  $p_{X,F}(x,f)$  denotes the joint histogram of source image  $X$  and fused image  $F$ . A large MI metric means that considerable information is transferred from source images to the fused image, which indicates a good fusion performance.

##### (3) Feature mutual information

In many image processing tasks, an image is often represented by its features and most of the information related to fused image quality comprise certain features, such as edges, details, and contrast. Therefore, measuring the amount of feature information that is transferred from source images to the fused image is reasonable. Haghhighat et al. proposed the feature mutual information (FMI) quality index, which is based on MI and feature information, to measure the amount of feature information that is transferred from source images to the fused image [300]. The definition of the FMI metric is as follows:

$$FMI = MI_{\dot{A},\dot{F}} + MI_{\dot{B},\dot{F}}, \quad (4)$$

where  $\dot{A}$ ,  $\dot{B}$ , and  $\dot{F}$  denote the feature maps of infrared, visible, and fused images, respectively. A large FMI metric generally indicates that considerable feature information is transferred from source images to the fused image.

##### (4) Structural similarity index measure

The human visual system is sensitive to structure loss and distortion. Wang et al. proposed a universal quality index called structural similarity index measure (SSIM) to model image loss and distortion [315]. The index mainly consists of three components: loss of correlation and luminance and contrast distortion. The product of the three components is the assessment result of the fused image and defined as follows:

$$SSIM_{X,F} = \sum_{x,f} \frac{2\mu_x\mu_f + C_1}{\mu_x^2 + \mu_f^2 + C_1} \cdot \frac{2\sigma_x\sigma_f + C_2}{\sigma_x^2 + \sigma_f^2 + C_2} \cdot \frac{\sigma_{xf} + C_3}{\sigma_x\sigma_f + C_3}, \quad (5)$$

where  $SSIM_{X,F}$  denotes the structural similarity between source image  $X$  and fused image  $F$ ;  $x$  and  $f$  denote the image patches of source and fused images in a sliding window, respectively;  $\sigma_{xf}$  denotes the covariance of source and fused images;  $\sigma_x$  and  $\sigma_f$  denote the standard deviation (SD); and  $\mu_x$  and  $\mu_f$  denote the mean values of source and fused images, respectively.  $C_1$ ,  $C_2$ , and  $C_3$  are the parameters used to make the algorithm stable; when  $C_1 = C_2 = C_3 = 0$ , the SSIM is reduced to the universal image quality index [316]. Thus, the structural similarities between all source images and the fused image can be written as follows:

$$SSIM = SSIM_{A,F} + SSIM_{B,F}, \quad (6)$$

where  $SSIM_{A,F}$  and  $SSIM_{B,F}$  denote the structural similarities between infrared/visible and fused images.

(5)  $Q^{AB/F}$

$Q^{AB/F}$  measures the amount of edge information that is transferred from source images to the fused image and is based on the assumption that the edge information in the source images is preserved in the fused image.  $Q^{AB/F}$  is defined as follows:

$$Q^{AB/F} = \frac{\sum_{i=1}^N \sum_{j=1}^M Q^{AF}(i,j)w^A(i,j) + Q^{BF}(i,j)w^B(i,j)}{\sum_{i=1}^N \sum_{j=1}^M (w^A(i,j) + w^B(i,j))}, \quad (7)$$

where  $Q^{XF}(i,j) = Q_g^{XF}(i,j)Q_a^{XF}(i,j)$ ,  $Q_g^{XF}(i,j)$  and  $Q_a^{XF}(i,j)$  denote the edge strength and orientation values at location  $(i,j)$ , respectively.  $w^X$  denotes the weight that expresses the importance of each source image to the fused image. A large  $Q^{AB/F}$  means that considerable edge information is transferred to the fused image.

(6)  $Q_W$

$Q_W$  can be determined as a variant of the universal image quality index [317]. This measure assigns high weights to salient visual regions in an image; these weights can consider the locations and magnitudes of distortions. The definition of  $Q_W$  is as follows:

$$Q_W(A, B, F) = \sum_{w \in W} c(w)(\lambda(w)Q_0(A, F|w) + (1 - \lambda(w)Q_0(B, F|w)), \quad (8)$$

where  $w$  denotes the local window and  $|W|$  denotes the number of windows in an image.  $Q_0(A, F|w)$  and  $Q_0(B, F|w)$  can be calculated by the universal image quality index in a sliding window  $w$ ,  $\lambda(w)$  denotes the saliency weight in each window, and  $c(w)$  normalizes saliency among all windows.

(7)  $Q_E$

$Q_E$  was designed by modifying  $Q_W$  [317] and can consider the edge information of the human visual system. The definition is as follows:

$$Q_E = Q_W(A, B, F) \cdot Q_W(\dot{A}, \dot{B}, \dot{F})^\alpha, \quad (9)$$

where  $\dot{A}$ ,  $\dot{B}$ , and  $\dot{F}$  denote the edges of images  $A$ ,  $B$ , and  $F$ , respectively, and  $\alpha$  is a variable that determines the contribution of edge images.  $Q_E$  is consistent with the human perception due to the use of the original and corresponding edge images. A high  $Q_E$  indicates a good fusion performance.

(8) Standard deviation

The standard deviation (SD) metric is based on the statistical concept that reflects the distribution and contrast of the fused image. SD is mathematically defined as follows:

$$SD = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i,j) - \mu)^2}, \quad (10)$$

where  $\mu$  denotes the mean value of the fused image. Regions with high contrast always attract human attention due to the sensitivity of the human visual system to contrast. Therefore, a fused image with high contrast often results in a large SD, which means that the fused image

achieves a good visual effect.

(9) Spatial frequency

Spatial frequency (SF) is an image quality index based on gradients [318], i.e., horizontal and vertical gradients, which are also called spatial row frequency (RF) and column frequency (CF), respectively. The SF metric can measure the gradient distribution of an image effectively, thereby revealing the detail and texture of an image. This metric is defined as follows:

$$SF = \sqrt{RF^2 + CF^2}, \quad (11)$$

where  $RF = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i,j) - F(i,j-1))^2}$  and  $CF = \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i,j) - F(i-1,j))^2}$ . A fused image with a large SF is sensitive to human perception according to the human visual system and has rich edges and textures.

(10) Average gradient

The average gradient (AG) metric quantifies the gradient information of the fused image [88] and represents its detail and texture. The AG metric is defined as follows:

$$AG = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{\nabla F_x^2(i,j) + \nabla F_y^2(i,j)}{2}}, \quad (12)$$

where  $\nabla F_x(i,j) = F(i,j) - F(i+1,j)$  and  $\nabla F_y(i,j) = F(i,j) - F(i,j+1)$ . The larger the AG metric, the more gradient information the fused image contains and the better the performance of the fusion algorithm.

(11) Mean gradient

The mean gradient (MG) metric is similar to the AG metric and defined as follows:

$$MG = \frac{1}{(M-1)(N-1)} \times \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} \sqrt{((F(x,y) - F(x-1,y))^2 + (F(x,y) - F(x,y-1))^2)/2}. \quad (13)$$

A large MG metric indicates that the fused image contains rich edges and textures and hence has a good fusion performance.

(12) Mean squared error

The mean squared error (MSE) computes the error of the fused image in comparison with those in the source images and hence measures the dissimilarity between the fused image and source images. The MSE is defined as follows:

$$MSE = \frac{MSE_{AF} + MSE_{BF}}{2}, \quad (14)$$

where  $MSE_{XF} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X(i,j) - F(i,j))^2$ ,  $MSE_{AF}$  and  $MSE_{BF}$  denote the dissimilarity between the fused and infrared/visible images. A small MSE metric indicates a good fusion performance, which means that the fused image approximates to the source images and minimal error occurs in the fusion process.

(13) Root mean squared error

The root mean squared error (RMSE) metric is similar to the MSE metric and defined as follows:

$$RMSE = \frac{RMSE_{AF} + RMSE_{BF}}{2}, \quad (15)$$

where  $RMSE = \sqrt{\frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (X(m,n) - F(m,n))^2}$ ,  $MSE_{AF}$  and  $MSE_{BF}$  denote the dissimilarity between the fused and infrared/visible images. A small RMSE metric indicates that the fused image has a small amount of error and distortion.

(14) Peak signal-to-noise ratio

The peak signal-to-noise ratio (PSNR) metric is the ratio of peak value power and noise power in the fused image and thus reflects the distortion during the fusion process. The PSNR metric is defined as follows:

$$PSNR = 10 \log_{10} \frac{r^2}{MSE}, \quad (16)$$

where  $r$  denotes the peak value of the fused image. The larger the PSNR, the closer the fused image is to the source images and the less distortion the fusion method produces.

#### (15) Visual information fidelity

The visual information fidelity (VIF) metric measures the information fidelity of the fused image [298], which is consistent with the human visual system. VIF aims to build a model to compute the distortion between the fused and source images, and this process can be achieved by four steps. First, the source images and fused image are filtered and divided into different blocks. Second, the visual information of each block with and without distortion is evaluated. Third, the VIF for each subband is calculated. Finally, the overall metric based on VIF is calculated. The VIF metric demonstrated a good performance on Petrovic's subjective test database [298] and showed higher computational efficiency than several conventional evaluation metrics.

#### (16) Correlation coefficient

The correlation coefficient (CC) measures the degree of linear correlation of the fused image and source images and is defined as follows:

$$CC = \frac{(r_{AF} + r_{BF})}{2}, \quad (17)$$

where  $r_{XF} = \frac{\sum_{i=1}^M \sum_{j=1}^N (X(i,j) - \bar{X})(F(i,j) - \mu)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N (X(i,j) - \bar{X})^2} \sqrt{\sum_{i=1}^M \sum_{j=1}^N (F(i,j) - \mu)^2}}$  and  $\bar{X}$  denotes the mean value of source image  $X$ . The larger the CC, the more similar the fused image is to the source images and the better the fusion performance.

#### (17) Nonlinear correlation coefficient

The nonlinear correlation coefficient (NCC) denotes the nonlinear correlation between the fused image and source images [68] and can be computed as follows:

$$NCC(X; F) = 2 + \sum_{i=1}^{b^2} \frac{n_i}{N} \log_b \frac{n_i}{N}, \quad (18)$$

where  $n_i$  denotes the number of samples distributed in the  $i$ th rank,  $b$  denotes the total number of ranks, and  $N$  denotes the total number of sample pairs.

#### (18) Other metrics

Other image fusion metrics include localized MI [319], normalized MI [320], Weber MI [321],  $Q_P$  [322], average pixel intensity [323], and sum of correlation differences [324]. Hossny et al. proposed the localized MI for image fusion; this metric adopts the MI on each block of source and fused images, and each block is obtained by EN-driven quadtree decomposition [319]. Luo et al. proposed the Weber MI for evaluating infrared and visible image fusion; this metric groups the pixels in infrared and visible images according to the local Weber components to capture image structures [321]. Aslantas et al. proposed the sum of correlation differences for image fusion quality evaluation [324]. This metric is based on the correlation among different images and obtained from the differences between source images and the fused image.

## 5. Experiments

In this section, we select 18 representative methods and 9 assessment metrics to conduct experiments, which can provide an objective performance reference for different infrared and visible image fusion methods and hence support relative engineering with credible and solid evidence. The representative infrared and visible image fusion methods are as follows: LP<sup>1</sup> [31], Wavelet<sup>2</sup> [44], NSCT<sup>3</sup> [69], dual-tree multi-

resolution discrete cosine transform (DTMDCT)<sup>4</sup> [116], cross bilateral filter (CBF)<sup>5</sup> [84], hybrid multi-scale decomposition (HMSD)<sup>6</sup> [91], guided filtering-based fusion (GFF)<sup>7</sup> [79], anisotropic diffusion-based fusion (ADF)<sup>8</sup> [82], ASR<sup>9</sup> [138], LP and sparse representation (LPSR)<sup>10</sup> [18], orientation information-motivated PCNN (OI-PCNN)<sup>11</sup> [161], SF-motivated PCNNs in NSCT domain (NSCT-SF-PCNN)<sup>12</sup> [159], directional discrete cosine transform and PCA (DDCTPCA)<sup>13</sup> [179], FPDE<sup>14</sup> [14], two-scale image fusion based on visual saliency (TSIFVS)<sup>15</sup> [80], local edge-preserving LC (LEPLC) [16], gradient transfer fusion (GTF)<sup>16</sup> [20], and IFEVIP<sup>17</sup> [209]. The LP, Wavelet, NSCT, DTMDCT, CBF, HMSD, GFF, and ADF are representative multi-scale transform-based methods, ASR and LPSR are representative sparse representation-based methods, OI-PCNN and NSCT-SF-PCNN are representative neural network-based methods, DDCTPCA and FPDE are representative subspace-based methods, TSIFVS and LEPLC are representative saliency-based methods, and GTF and IFEVIP belong to other method classes. Many of these 18 methods, including GFF, LPSR, NSCT-SF-PCNN, DDCTPCA, TSIFVS, and LEPLC, can be considered hybrid models because they combine different approaches. LPSR is a hybrid multi-scale transform and sparse representation method, NSCT-SF-PCNN is a hybrid multi-scale transform and neural network model, DDCTPCA is a hybrid multi-scale transform and subspace approach, and GFF, TSIFVS, and LEPLC are hybrid multi-scale transform and saliency methods. The experiments are conducted on a desktop with 3.3 GHz Intel Core CPU, 8 GB memory, and MATLAB codes. The codes of the 18 representative image fusion methods are all publicly available, and their parameters are all set in accordance with those in the original studies.

We test the 18 representative methods on the surveillance images from TNO Human Factors, which contain multispectral nighttime imagery of different military relevant scenarios, registered with different multiband camera systems<sup>18</sup>. We conduct experiments on 12 representative image infrared/visible pairs and two infrared/visible image sequences for qualitative and quantitative comparisons. We select nine frequently used assessment metrics, i.e., EN [313], MI [314], SSIM<sup>19</sup> [316],  $Q^{AB/FP}$ <sup>20</sup> [314],  $Q_W$  [317], SD [325], SF [318], VIF<sup>19</sup> [298], and CC [301], to evaluate the performances of the different infrared and visible image fusion methods. Large assessment metric values indicate good model performances.

Fig. 8 shows the visual fusion performances of the 18 representative methods on the 12 infrared and visible image pairs. The representative multi-scale-transform-based methods except DTMDCT obtain similar visual fusion performances. Fusion performance using DTMDCT is

<sup>4</sup> Available at <https://cn.mathworks.com/matlabcentral/fileexchange/43566-novel-image-fusion-techniques-using-dct>.

<sup>5</sup> Available at <https://cn.mathworks.com/matlabcentral/fileexchange/43781-image-fusion-based-on-pixel-significance-using-cross-bilateral-filter>.

<sup>6</sup> Available at <https://github.com/bitzhouq/Context-Enhance-via-Fusion>.

<sup>7</sup> Available at <http://xudongkang.weebly.com/>.

<sup>8</sup> Available at <https://cn.mathworks.com/matlabcentral/fileexchange/63591-fusion-of-infrared-and-visible-sensor-images-based-on-anisotropic-diffusion-and-kl-transform>.

<sup>9</sup> Available at [http://quxiaobo.8866.org/project/OI\\_PCNN/OI\\_PCNN.zip](http://quxiaobo.8866.org/project/OI_PCNN/OI_PCNN.zip).

<sup>10</sup> Available at [http://csrc.xmu.edu.cn/project/Image%20fusion/ImageFusion\\_index.html](http://csrc.xmu.edu.cn/project/Image%20fusion/ImageFusion_index.html).

<sup>11</sup> Available at <https://cn.mathworks.com/matlabcentral/fileexchange/46169-directional-discrete-cosine-transform-and-principal-component-analysis-based-image-fusion>.

<sup>12</sup> Available at <https://cn.mathworks.com/matlabcentral/fileexchange/63570-multi-sensor-image-fusion-based-on-fourth-order-partial-differential-equations>.

<sup>13</sup> Available at <https://sites.google.com/view/durgaprasadbavirasetti/publications>.

<sup>14</sup> Available at <http://www.escience.cn/people/jiayima/cxdm.html>.

<sup>15</sup> Available at <https://github.com/uzeful/Infrared-and-Visual-Image-Fusion-via-Infrared-Feature-Extraction-and-Visual-Information-Preservation>.

<sup>16</sup> Available at [https://figshare.com/articles/TNO\\_Image\\_Fusion\\_Dataset/1008029](https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029).

<sup>17</sup> Available at [http://www.cns.nyu.edu/~zwang/files/research/quality\\_index/demo.html](http://www.cns.nyu.edu/~zwang/files/research/quality_index/demo.html).

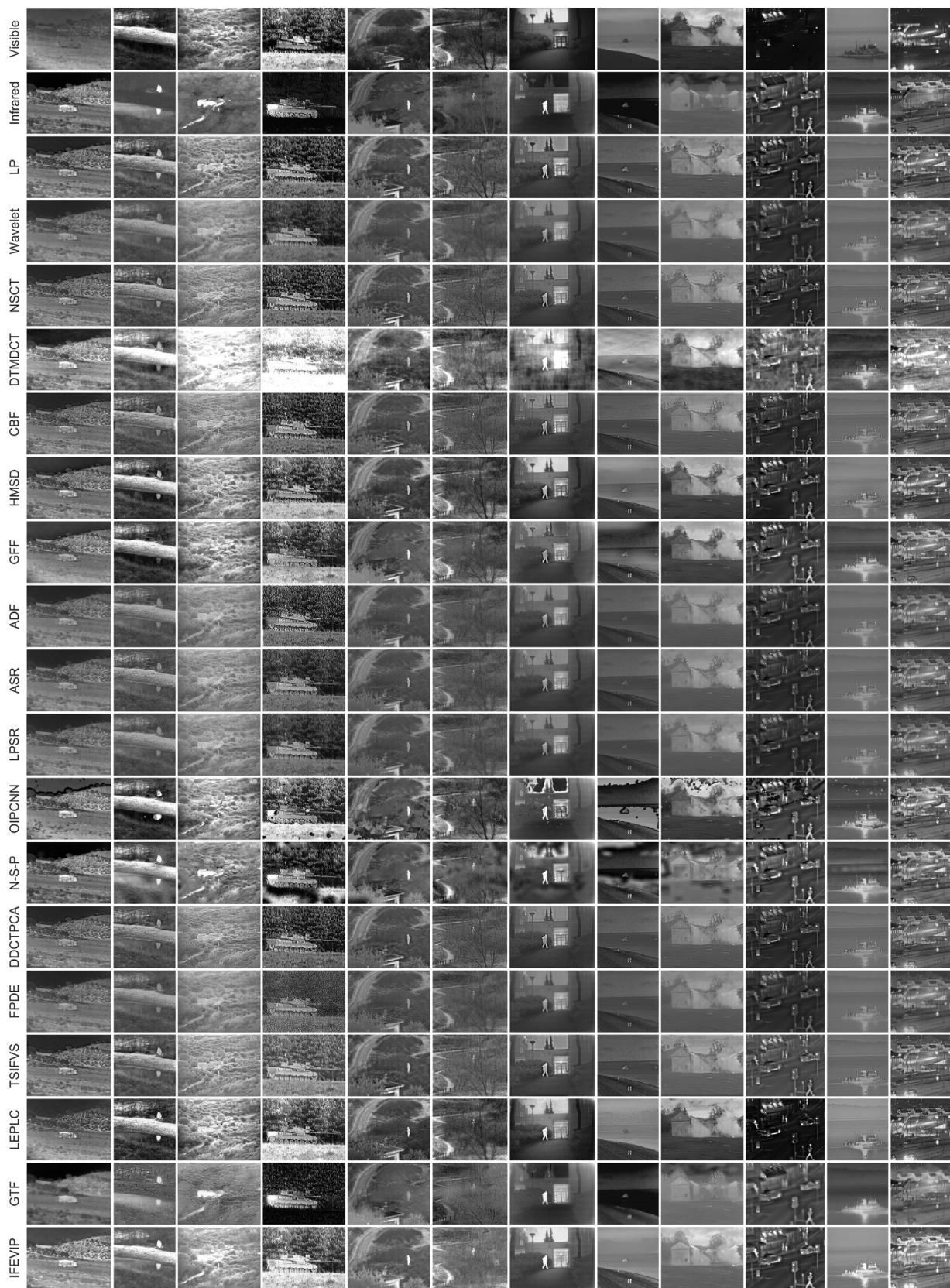
<sup>18</sup> Available at <https://cn.mathworks.com/matlabcentral/fileexchange/18213-objective-image-fusion-performance-measure>.

<sup>19</sup> Available at <http://hansy.weebly.com/image-fusion-metric-ifm.html>.

<sup>1</sup> Available at <http://www.metapix.de/toolbox.htm>.

<sup>2</sup> Available at <https://cn.mathworks.com/help/wavelet/ref/wfusimg.html>.

<sup>3</sup> Available at <http://www.escience.cn/people/liuyu1/Codes.html>.



**Fig. 8.** Fusion performance of eighteen representative methods on twelve infrared and visible image pairs. From left to right: *Athena*, *Bench*, *Bunker*, *Tank*, *Nato\_camp*, *Sandpath*, *Kaptein*, *Kayak*, *Octec*, *Street*, *Steamboat* and *Road*.

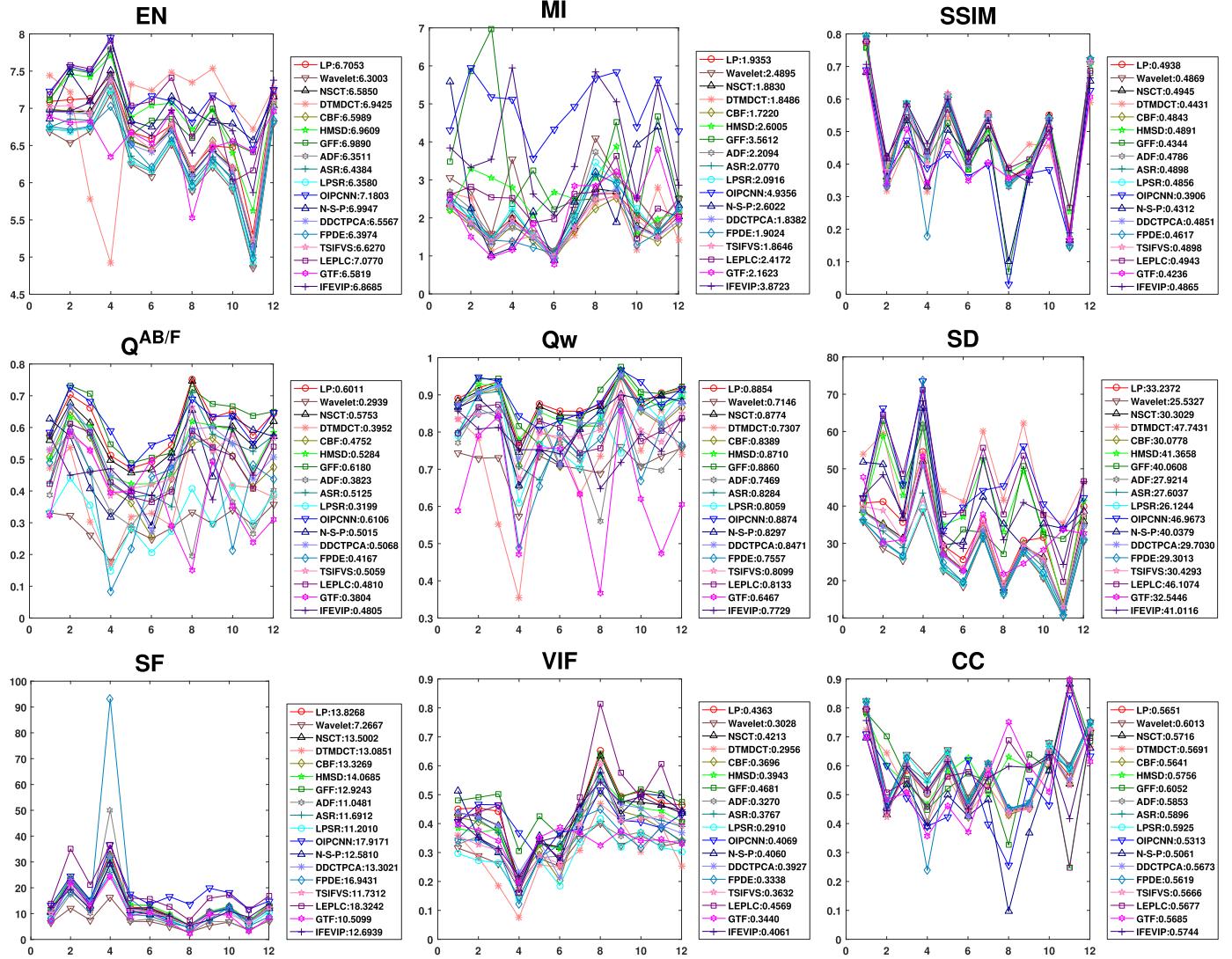


Fig. 9. Quantitative comparisons of the nine metrics using eighteen representative methods on the twelve infrared and visible image pairs shown in Fig. 8. From 1 to 12: *Athena*, *Bench*, *Bunker*, *Tank*, *Nato\_camp*, *Sandpath*, *Kaptein*, *Kayak*, *Octec*, *Street*, *Steamboat* and *Road*.

generally brighter than those of the other multi-scale transform-based methods, and the fusion performances of *Bunker* and *Tank* using DTMDCT are noticeably worse than those of the other multi-scale transform-based methods. The two representative sparse representation-based methods, ASR and LPSR, obtain comparable fusion performances on the 12 image pairs. By contrast, the fusion performances of the two representative neural network-based methods, OI-PCNN and NSCT-SF-PCNN, greatly vary for the different image pairs. The fused results on *Kaptein*, *Kayak*, and *Octec* using OI-PCNN and NSCT-SF-PCNN show many serious distortions in the form of non-smooth color lumps, whereas the nine other image pairs using these two approaches can obtain comparable fusion performances. Similar to the sparse representation-based methods, the two representative subspace-based methods, DDCTPCA and FPDE, can also produce comparable fusion performances for all the image pairs. The two representative saliency-based methods, TSIFVS and LEPLC, can transfer salient information, such as targets, to the fused image better than the other methods can. The images fused using GTF can preserve the thermal radiation intensity information in infrared images and the detail texture information in visible images. The images fused using IFEVIP can highlight

important infrared objects and have achieve visual quality. In addition, we report the results of the nine metrics using the 18 representative methods on the 12 image pairs, as shown in Fig. 9, which can give quantitative comparisons of the fusion performances of the different methods (The quantitative results are discussed subsequently.).

We also conduct experiments on two infrared and visible image sequences, *Nato\_camp* and *Dune*, to evaluate the fusion performances of the selected methods; these sequences contain 32 and 23 image pairs, respectively. Note that the image sequence fusion here is analogous to the video fusion problem. The results are reported in Figs. 10 and 11. Each metric can evaluate image quality in only one aspect, which means that all scores should be considered when making an accurate and comprehensive decision. The results of the quantitative experiments show that each algorithm has its advantages and disadvantages and different approaches show superiority in different aspects. The multi-scale transform-based methods show large values in  $Q_w$  and CC; these values mean that a large amount of gradients is transferred from the source images to the final results. The sparse representation-based methods lead in the SSIM; thus, the algorithms introduce minimal distortion in the fusion process. The PCNN-based methods achieve good

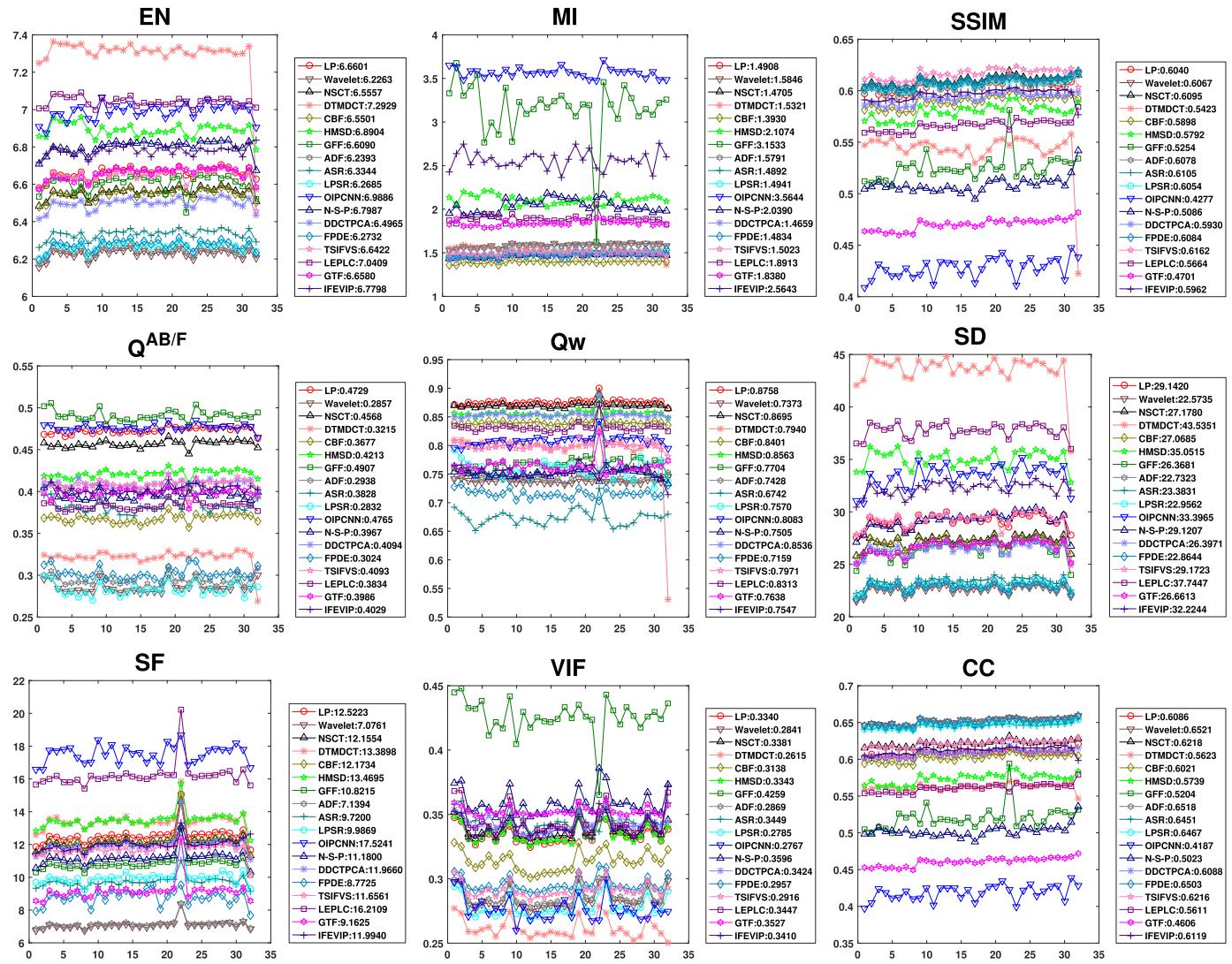


Fig. 10. Quantitative comparisons of the nine metrics using eighteen representative methods on the *Nato\_camp* sequence.

performances in MI but exhibit low values in SSIM. These results indicate severe distortions in the qualitative experiments. The subspace-based methods show poor performances in MI; these values indicate that only a small amount of source information is transferred to the fused images. The saliency-based methods show advantages in EN and SD; these figures indicate richness in information and large contrast. Other approaches exhibit relative mediocre performances with scores ranking in the middle.

The run time comparison of the 18 algorithms on the two sequences are given in Table 2, where the images all have a size of  $270 \times 360$  and each value denotes the mean and SD of the run time of the corresponding method on a sequence. The results show that the performances of the multi-scale transform-based methods come in a wide range. Most of these models are fast and stable and have small means and standard variances. However, some approaches, such as NSCT and CBF, are slow. ASR, NSCT\_SF\_PCNN, and DDCTPCA are also relatively slow and may be unsuitable for surveillance and tracking applications, which need to achieve fusion in real time.

## 6. Conclusion and future trend

Infrared and visible image fusion has attracted considerable attention and made significant progress in the past few decades. Thus, we comprehensively survey existing infrared and visible image fusion methods and applications. These approaches can be divided into seven categories: multi-scale decomposition-based, sparse representation-based, neural network-based, subspace-based, saliency-based, and hybrid methods and other models. Each category is briefly introduced and summarized according to core idea, theory, procedure, and progress. Image registration significantly affects fusion performance; thus, we briefly review extant registration techniques. Infrared and visible image fusion methods have been widely used in different applications, such as object detection and recognition, tracking, image enhancement, surveillance, and remote sensing, due to their advantages and progress. We then summarize several frequently used infrared and visible image fusion evaluation metrics to evaluate the performances of the studied infrared and visible image fusion methods. Furthermore, we conduct extensive experiments to evaluate the performances of different representative methods to potentially provide an objective performance

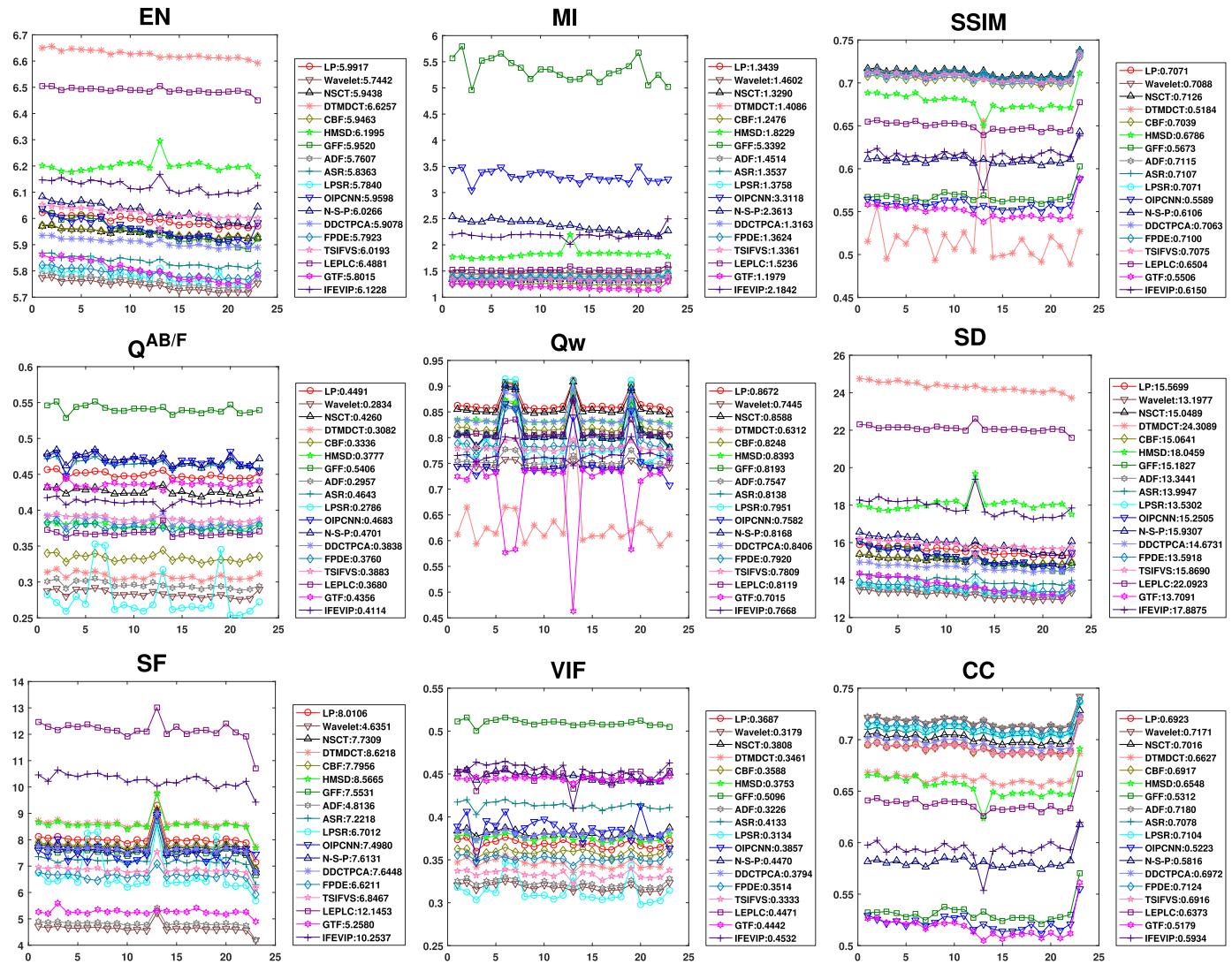


Fig. 11. Quantitative comparisons of the nine metrics using eighteen representative methods on the *Dune* sequence.

reference for researchers in the field of infrared and visible image fusion and consequently support relative engineering with credible and solid evidence.

Despite the considerable progress that has been achieved in infrared and visible image fusion, several issues remain for future work.

- Multi-scale transform-based infrared and visible image fusion methods usually fix the basis functions and decomposition levels. It is still challenging to select flexible basis functions that allow data-driven choice of the best representation of source images, and the adaptive selection of decomposition levels still remains to be solved.
- Edge-preserving filtering-based methods have been widely used for image fusion, and efficient, robust, and general filters (*e.g.*, extensions of the original filters) can be designed for infrared and visible image fusion.
- Traditional sparse representation-based fusion methods often adopt patch-based procedures, which ignore the correlation among different patches and lead to the loss of detail information. In the

future, the correlation among different patches can be considered to improve fusion performance. Convolutional sparse coding can be used to enhance detail information. This idea is motivated by deconvolutional networks, which aim to build a hierarchy of sparse representations of source images.

- Few deep learning-based fusion methods have been specialized for infrared and visible image fusion. In the future, state-of-the-art deep learning techniques can be applied in infrared and visible image fusion. For example, more effective deep models can be designed to perform joint activity-level measurement and weight assignment.
- The imaging mechanisms of infrared and visible sensors are different; thus, we should design niche-targeting saliency detection methods for infrared and visible images separately. The thermal information of objects is reflected by the pixel intensity in infrared images, and regions with serious noise are easily regarded as targets when using saliency detection methods. Therefore, anti-noise saliency detection methods should be designed in future work.
- Different infrared and visible image fusion methods have their own

**Table 1**  
Statistics of some representative fusion evaluation measures and references.

evaluation measure	reference
entropy (EN)	[8,12,13,15,16,18,20,21,25,32,34,36–41,45–51,53–57,61,62,65,67,69,70,73,75–77,81,85,88–90,97–99,101,102–105,106,110–112,115,119,121,131,140,141,143,146,158,160,169,175,178,190,191,194,198,199,200,204]
mutual information (MI)	[1,2,11,13,15,20,22,26,36,37,45,46,52,54,55,61,62,66,70–72,76,77,81,83,84,86,87,90,98,100,102,1–04,106–109,115,119,121–123,129,136,138,140,142–144,146–153,158,160,167,175,178,181–183,191,198–200]
feature mutual information (FMI)	[20,169,199,300,301]
structural similarity index measure (SSIM)	[8,11,26,48,53,71,73,77,90,101,102,105,109,142,146,167,178,182,198]
$Q_{BF/F}$	[2,8,11,12,15,18,20,22,26,47,48,50–52,54–57,70–72,77,79,81–84,86–87,90,100,102–104,107,108,120,122,129,136,137,139–142,144–145,147,148,150,153,154,168,197]
$Q_w$	[2,8,11,18,47,71,73,77,83,103,107,122,136,137,139,142,144,145,147,148,153,154,169,182,197]
$Q_E$	[2,8,11,12,21,71,77,83,103,107,122,136,142,144,145,147,148,153,154,169,182,197]
standard deviation (SD)	[9,12,16,18,24,32,38–41,46,50,54–57,61,62,66,75,76,81,84,90,98,99,105,108,110,111,115,121,131,140,143,149,154,178,190,191,198,200]
spatial frequency (SF)	[13,15,16,21,40,41,53–56,61,65,73,81,84,86,88,98,102,106,113,115,121,158,178,179,204,208]
average gradient (AG)	[37–41,46,48,50,54–55,62,65–67,69,73,75,76,81,84,88,97,99,105,108,110,115,117,121,149,183,190,198]
mean gradient (MG)	[61,86,113,142,204,208,302,303]
mean squared error (MSE)	[304–307]
root mean squared error (RMSE)	[1,35,47,120,131,301,308,309]
peak signal to noise ratio (PSNR)	[25,48,61,101,131,142,158,175,181–183,185,277,304,305,310]
visual information fidelity (VIF)	[34,71,72,74,80,83,86,109,119,114,122,141,142,150–152,169,277,298,311]
correlation coefficient (CC)	[9,24,84,109,183,305,312]
nonlinear correlation coefficient (NCC)	[25,68]

**Table 2**

Run time comparison of eighteen algorithms on the *Nato\_camp* and *Duine* sequences, where each value denotes the mean and standard deviation of run time of a certain method on a sequence (unit: second).

Method	<i>Nato_camp</i>	<i>Duine</i>
LP	$0.0043 \pm 0.0007$	$0.0044 \pm 0.0002$
Wavelet	$0.1550 \pm 0.0382$	$0.1592 \pm 0.0018$
NSCT	$1.4385 \pm 0.0092$	$1.4402 \pm 0.0096$
DTMDCT	$0.0347 \pm 0.0018$	$0.0337 \pm 0.0019$
CBF	$6.1431 \pm 0.0213$	$6.1211 \pm 0.0304$
HMSD	$0.5441 \pm 0.0558$	$0.5492 \pm 0.0328$
GFF	$0.0871 \pm 0.0067$	$0.0927 \pm 0.0091$
ADF	$0.1769 \pm 0.0031$	$0.1730 \pm 0.0075$
ASR	$94.638 \pm 0.3782$	$94.638 \pm 0.3199$
LPSR	$0.0111 \pm 0.0026$	$0.0087 \pm 0.0005$
OIPCNN	$0.4000 \pm 0.0021$	$0.3995 \pm 0.0018$
NSCT_SF PCNN	$72.047 \pm 0.2027$	$72.028 \pm 0.1884$
DDCTPCA	$36.901 \pm 0.1771$	$37.102 \pm 0.1162$
FPDE	$0.0922 \pm 0.0040$	$0.0925 \pm 0.0043$
TSIFVS	$0.0102 \pm 0.0019$	$0.0102 \pm 0.0014$
LEPLC	$0.1494 \pm 0.0085$	$0.1575 \pm 0.0056$
GTF	$0.9917 \pm 0.0609$	$1.3706 \pm 0.1052$
IFEVIP	$0.0517 \pm 0.0012$	$0.0542 \pm 0.0015$

strengths and weaknesses, and the advantages of different methods should be combined to obtain efficient hybrid image fusion methods.

- The spatial resolutions of infrared and visible image sensors sometimes differ; hence, joint image super-resolution and fusion can be performed in the future.
- Traditional infrared and visible image fusion methods typically ignore noise. Joint image denoising and fusion can be pursued in the future work.
- Image pairs without strict alignments lead to ghosting in fused images. The inconsistency of salient structures between the original and fused images can provide a guide for image registration. Therefore, spatial transformation can be considered a variable when designing fusion rules, and registration and fusion can be conducted simultaneously to eliminate ghosting.
- Traditional fusion methods mainly focus on designing appropriate principles to keep useful information in source images as much as possible. In the future, the develop of fusion methods should also be focus on the application. For example, rather than producing an informative fused image, the fusion should highlight the valuable information in a specific application.
- Fusion performance evaluation methods usually use edge and intensity to compute the contributions of infrared and visible images to the fused image, but these features do not conform to the human visual system significantly. One trend is to design features that represent visual information that conforms to the human visual system. Another trend involves proposing a new model to represent the information transformation between fused and source images.
- Since the qualitative performance of a fusion algorithm may not always match with its quantitative performance, it is desirable to design new metrics, which can match with the qualitative performance of fused images.

## Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant nos. 61773295 and 61503288.

## References

- [1] G. Piella, A general framework for multiresolution image fusion: from pixels to regions, *Inf. Fusion* 4 (4) (2003) 259–280.
- [2] 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.
- [3] 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.
- [4] J. Han, B. Bhanu, Fusion of color and infrared video for moving human detection, *Pattern Recognit.* 40 (6) (2007) 1771–1784.
- [5] E. Reinhard, M. Adhikmin, B. Gooch, P. Shirley, Color transfer between images, *IEEE Comput. Graph. Appl.* 21 (5) (2001) 34–41.
- [6] P. Kumar, A. Mittal, P. Kumar, Fusion of thermal infrared and visible spectrum video for robust surveillance, *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, (2006), pp. 528–539.
- [7] G. Simone, A. Farina, F.C. Morabito, S.B. Serpico, L. Bruzzone, Image fusion techniques for remote sensing applications, *Inform. Fusion* 3 (1) (2002) 3–15.
- [8] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transforms for image fusion, *Inform. Fusion* 12 (2) (2011) 74–84.
- [9] G. Pajares, J.M. De La Cruz, A wavelet-based image fusion tutorial, *Pattern Recognit.* 37 (9) (2004) 1855–1872.
- [10] Z. Zhang, R.S. Blum, 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.
- [11] 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.
- [12] T. Xiang, L. Yan, R. Gao, A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NCST domain, *Infrared Phys. Technol.* 69 (2015) 53–61.
- [13] W. Kong, L. Zhang, Y. Lei, Novel fusion method for visible light and infrared images based on nst-sf-pcnn, *Infrared Phys. Technol.* 65 (2014) 103–112.
- [14] D.P. Bavrisetti, G. Xiao, G. Liu, Multi-sensor image fusion based on fourth order partial differential equations, *Proceedings of the International Conference on Information Fusion*, (2017), pp. 1–9.
- [15] 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.
- [16] 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, *JOSA A* 34 (8) (2017) 1400–1410.
- [17] 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.
- [18] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi-scale transform and sparse representation, *Inform. Fusion* 24 (2015) 147–164.
- [19] J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, *Infrared Phys. Technol.* 82 (2017) 8–17.
- [20] J. Ma, C. Chen, C. Li, J. Huang, Infrared and visible image fusion via gradient transfer and total variation minimization, *Inform. Fusion* 31 (2016) 100–109.
- [21] J. Zhao, G. Cui, X. Gong, Y. Zang, S. Tao, D. Wang, Fusion of visible and infrared images using global entropy and gradient constrained regularization, *Infrared Physics & Technology* 81 (2017) 201–209.
- [22] Y. Liu, J. Jin, Q. Wang, Y. Shen, X. Dong, Region level based multi-focus image fusion using quaternion wavelet and normalized cut, *Signal Process.* 97 (2014) 9–30.
- [23] A. Toet, Image fusion by a ratio of low-pass pyramid, *Pattern Recognit. Lett.* 9 (4) (1989) 245–253.
- [24] M. Choi, R.Y. Kim, M.-R. Nam, H.O. Kim, Fusion of multispectral and panchromatic satellite images using the curvelet transform, *IEEE Geosci. Remote Sensing Lett.* 2 (2) (2005) 136–140.
- [25] S. Rajkumar, P.C. Mouli, Infrared and visible image fusion using entropy and neuro-fuzzy concepts, *Proceedings of the Annual Convention of Computer Society of India*, (2014), pp. 93–100.
- [26] Q. Zhang, Y. Liu, R.S. Blum, J. Han, D. Tao, Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review, *Inform. Fusion* 40 (2018) 57–75.
- [27] S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: A survey of the state of the art, *Inform. Fusion* 33 (2017) 100–112.
- [28] A. Dogra, B. Goyal, S. Agrawal, From multi-scale decomposition to non-multi-scale decomposition methods: A comprehensive survey of image fusion techniques and its applications, *IEEE Access* 5 (2017) 16040–16067.
- [29] D.L. Donoho, A.G. Flesia, Can recent innovations in harmonic analysis ‘explain’ key findings in natural image statistics? *Network: Comput. Neural Syst.* 12 (3) (2001) 371–393.
- [30] H. Lin, Y. Tian, R. Pu, L. Liang, Remotely sensing image fusion based on wavelet transform and human vision system, *Int. J. Signal Process. Image Process. Pattern Recognit.* 8 (7) (2015) 291–298.
- [31] P. Burt, E. Adelson, The laplacian pyramid as a compact image code, *IEEE Trans. Commun.* 31 (4) (1983) 532–540.
- [32] D. Bulanov, T. Burks, V. Alchanatis, Image fusion of visible and thermal images for fruit detection, *Biosyst. Eng.* 103 (1) (2009) 12–22.
- [33] X. Yu, J. Ren, Q. Chen, X. Sui, A false color image fusion method based on multi-resolution color transfer in normalization ycbcr space, *Optik-Int. J. Light Electron Opt.* 125 (20) (2014) 6010–6016.
- [34] A.V. Vannali, V.M. Gadre, Visible and nir image fusion using weight-map-guided Laplacian–Gaussian pyramid for improving scene visibility, *Sādhānā* 42 (7) (2017) 1063–1082.
- [35] Z. Liu, K. Tsukada, K. Hanasaki, Y.-K. Ho, Y. Dai, Image fusion by using steerable pyramid, *Pattern Recognit. Lett.* 22 (9) (2001) 929–939.
- [36] G. Liu, Z. Jing, S. Sun, J. Li, Z. Li, H. Leung, Image fusion based on expectation maximization algorithm and steerable pyramid, *Chinese Opt. Lett.* 2 (7) (2004) 386–389.
- [37] H. Deng, Y. Ma, Image fusion based on steerable pyramid and PCNN, *Proceedings of the International Conference on the Applications of Digital Information and Web Technologies*, (2009), pp. 569–573.
- [38] H. Jin, L. Jiao, F. Liu, Y. Qi, Fusion of infrared and visual images based on contrast pyramid directional filter banks using clonal selection optimizing, *Opt. Eng.* 47 (2) (2008) 027002.
- [39] H. Jin, Y. Wang, A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization, *Infrared Phys. Technol.* 64 (2014) 134–142.
- [40] H. Jin, Q. Xi, Y. Wang, X. Hei, Fusion of visible and infrared images using multiobjective evolutionary algorithm based on decomposition, *Infrared Phys. Technol.* 71 (2015) 151–158.
- [41] H. Xu, Y. Wang, Y. Wu, Y. Qian, Infrared and multi-type images fusion algorithm based on contrast pyramid transform, *Infrared Phys. Technol.* 78 (2016) 133–146.
- [42] E.P. Simoncelli, W.T. Freeman, The steerable pyramid: A flexible architecture for multi-scale derivative computation, *Proceedings of the International Conference on Image Processing*, (1995), pp. 444–447.
- [43] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 11 (7) (1989) 674–693.
- [44] L.J. Chipman, T.M. Orr, L.N. Graham, Wavelets and image fusion, *Proceedings of the International Conference on Image Processing*, (1995), pp. 248–251.
- [45] Y. Niu, S. Xu, L. Wu, W. Hu, Airborne infrared and visible image fusion for target perception based on target region segmentation and discrete wavelet transform, *Math. Probl. Eng.* 2012 (2012) 732–748.
- [46] L. Zhan, Y. Zhuang, L. Huang, Infrared and visible images fusion method based on discrete wavelet transform, *J. Comput.* 28 (2) (2017) 57–71.
- [47] K. Madheswari, N. Venkateswaran, Swarm intelligence based optimisation in thermal image fusion using dual tree discrete wavelet transform, *Quant. InfraRed Thermography* 14 (1) (2017) 24–43.
- [48] Z. Liu, H. Yin, B. Fang, Y. Chai, A novel fusion scheme for visible and infrared images based on compressive sensing, *Optics Commun.* 335 (2015) 168–177.
- [49] J. Saeedi, K. Faez, Infrared and visible image fusion using fuzzy logic and population-based optimization, *Appl. Soft Comput.* 12 (3) (2012) 1041–1054.
- [50] Y. Zuo, J. Liu, G. Bai, X. Wang, M. Sun, Airborne infrared and visible image fusion combined with region segmentation, *Sensors* 17 (5) (2017) 1127.
- [51] Y. Zou, X. Liang, T. Wang, Visible and infrared image fusion using the lifting wavelet, *Indonesian J. Electr. Eng. Comput. Sci.* 11 (11) (2013) 6290–6295.
- [52] Y. Chai, H. Li, J. Qu, Image fusion scheme using a novel dual-channel pcnn in lifting stationary wavelet domain, *Optics Commun.* 283 (19) (2010) 3591–3602.
- [53] C. Zhao, G. Shao, L. Ma, X. Zhang, Image fusion algorithm based on redundant-lifting nswmda and adaptive pcnn, *Optik-Int. J. Light Electron Opt.* 125 (20) (2014) 6247–6255.
- [54] X. Yan, H. Qin, J. Li, H. Zhou, J.-g. Zong, Infrared and visible image fusion with spectral graph wavelet transform, *JOSA A* 32 (9) (2015) 1643–1652.
- [55] P. Chai, X. Luo, Z. Zhang, Image fusion using quaternion wavelet transform and multiple features, *IEEE Access* 5 (2017) 6724–6734.
- [56] L. Xu, J. Du, Z. Zhang, Infrared-visible video fusion based on motion-compensated wavelet transforms, *IET Image Process.* 9 (4) (2014) 318–328.
- [57] Z. Wang, C. Gong, A multi-faceted adaptive image fusion algorithm using a multi-wavelet-based matching measure in the PCNN domain, *Appl. Soft Comput.* 61 (2017) 1113–1124.
- [58] I.W. Selesnick, R.G. Baraniuk, N.C. Kingsbury, The dual-tree complex wavelet transform, *IEEE Signal Process. Mag.* 22 (6) (2005) 123–151.
- [59] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multi-resolution image representation, *IEEE Trans. Image Process.* 14 (12) (2005) 2091–2106.
- [60] R.H. Bamberger, M.J. Smith, A filter bank for the directional decomposition of images: Theory and design, *IEEE Trans. Image Process.* 40 (4) (1992) 882–893.
- [61] H. Li, L. Liu, W. Huang, C. Yue, An improved fusion algorithm for infrared and visible images based on multi-scale transform, *Infrared Phys. Technol.* 74 (2016) 28–37.
- [62] Y. Lin, L. Song, X. Zhou, Y. Huang, Infrared and visible image fusion algorithm based on contourlet transform and penn, *Proceedings of the Infrared Materials, Devices, and Applications*, (2008), p. 683514.
- [63] S. Liu, Y. Piao, M. Tahir, Research on fusion technology based on low-light visible image and infrared image, *Opt. Eng.* 55 (12) (2016) 123104.
- [64] 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.
- [65] F. Meng, M. Song, B. Guo, R. Shi, D. Shan, Image fusion based on object region detection and non-subsampled contourlet transform, *Computers & Electrical Engineering* 62 (2017) 375–383.
- [66] S. Yin, L. Cao, Q. Tan, G. Jin, Infrared and visible image fusion based on nst and fuzzy logic, *Proceedings of the International Conference on Mechatronics and Automation*, (2010), pp. 671–675.
- [67] J. Adu, M. Wang, Z. Wu, J. Hu, Infrared image and visible light image fusion based on nonsubsampled contourlet transform and the gradient of uniformity, *International Journal of Advancements in Computing Technology* 4 (5) (2012) 114–121.
- [68] H.-X. Liu, T.-H. Zhu, J.-J. Zhao, Infrared and visible image fusion based on region of interest detection and nonsubsampled contourlet transform, *Journal of Shanghai Jiaotong University (Science)* 18 (5) (2013) 526–534.

- [69] J. Adu, J. Gan, Y. Wang, J. Huang, Image fusion based on nonsubsampled contourlet transform for infrared and visible light image, *Infrared Phys. Technol.* 61 (2013) 94–100.
- [70] Y. Chen, J. Xiong, H.-l. Liu, Q. Fan, Fusion method of infrared and visible images based on neighborhood characteristic and regionalization in nsct domain, *Optik-International Journal for Light and Electron Optics* 125 (17) (2014) 4980–4984.
- [71] C. Zhao, Y. Guo, Y. Wang, A fast fusion scheme for infrared and visible light images in nsct domain, *Infrared Phys. Technol.* 72 (2015) 266–275.
- [72] H. Li, H. Qiu, Z. Yu, Y. Zhang, Infrared and visible image fusion scheme based on nsct and low-level visual features, *Infrared Phys. Technol.* 76 (2016) 174–184.
- [73] J. Cai, Q. Cheng, M. Peng, Y. Song, Fusion of infrared and visible images based on nonsubsampled contourlet transform and sparse k-svd dictionary learning, *Infrared Phys. Technol.* 82 (2017) 85–95.
- [74] X. Wang, L. Yao, R. Song, H. Xie, A new infrared and visible image fusion algorithm in nsct domain, *Proceedings of the International Conference on Intelligent Computing*, (2017), pp. 420–431.
- [75] Y. Li, G.-h. Song, S.-c. Yang, Multi-sensor image fusion by nsct-pcnn transform, *Proceedings of the IEEE International Conference on Computer Science and Automation Engineering*, (2011), pp. 638–642.
- [76] W. Kong, Y. Lei, Y. Lei, S. Lu, Image fusion technique based on non-subsampled contourlet transform and adaptive unit-fast-linking pulse-coupled neural network, *IET Image Process.* 5 (2) (2011) 113–121.
- [77] G.S. El-taweeel, A.K. Helmy, Image fusion scheme based on modified dual pulse coupled neural network, *IET Image Process.* 7 (5) (2013) 407–414.
- [78] Z. Farbman, R. Fattal, D. Lischinski, Edge-preserving decompositions for multi-scale tone and detail manipulation, *ACM Trans. Graph.* 27 (3) (2008) 1–10.
- [79] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, *IEEE Trans. Image Process.* 22 (7) (2013) 2864–2875.
- [80] D.P. Bavirisetti, R. Dhuli, Two-scale image fusion of visible and infrared images using saliency detection, *Infrared Phys. Technol.* 76 (2016) 52–64.
- [81] X. Yan, H. Qin, J. Li, H. Zhou, J.-g. Zong, Q. Zeng, Infrared and visible image fusion using multiscale directional nonlocal means filter, *Appl. Opt.* 54 (13) (2015) 4299–4308.
- [82] 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.
- [83] J. Hu, S. Li, The multiscale directional bilateral filter and its application to multisensor image fusion, *Inform. Fusion* 13 (3) (2012) 196–206.
- [84] B.S. Kumar, Image fusion based on pixel significance using cross bilateral filter, *Signal Image Video Process.* 9 (5) (2015) 1193–1204.
- [85] J. Zhao, Q. Zhou, Y. Chen, H. Feng, Z. Xu, Q. Li, Fusion of visible and infrared images using saliency analysis and detail preserving based image decomposition, *Infrared Phys. Technol.* 56 (2013) 93–99.
- [86] Y. Jiang, M. Wang, Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter, *IET Image Process.* 8 (3) (2014) 183–190.
- [87] W. Gan, X. Wu, W. Wu, X. Yang, C. Ren, X. He, K. Liu, Infrared and visible image fusion with the use of multi-scale edge-preserving decomposition and guided image filter, *Infrared Phys. Technol.* 72 (2015) 37–51.
- [88] G. Cui, H. Feng, Z. Xu, Q. Li, Y. Chen, Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition, *Opt. Commun.* 341 (2015) 199–209.
- [89] A. Toet, M.A. Hogervorst, Multiscale image fusion through guided filtering, *Proceedings of the SPIE Security + Defense*, (2016), p. 99970J.
- [90] H.-M. Hu, J. Wu, B. Li, Q. Guo, J. Zheng, An adaptive fusion algorithm for visible and infrared videos based on entropy and the cumulative distribution of gray levels, *IEEE Trans. Multim.* 19 (12) (2017) 2706–2719.
- [91] Z. Zhou, B. Wang, S. Li, M. Dong, Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters, *Inform. Fusion* 30 (2016) 15–26.
- [92] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (7) (1990) 629–639.
- [93] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, *Proceedings of the International Conference on Computer Vision*, (1998), pp. 839–846.
- [94] B. Gu, W. Li, M. Zhu, M. Wang, Local edge-preserving multiscale decomposition for high dynamic range image tone mapping, *IEEE Trans. Image Process.* 22 (1) (2013) 70–79.
- [95] C.-T. Shen, F.-J. Chang, Y.-P. Hung, S.-C. Pei, Edge-preserving image decomposition using l1 fidelity with l0 gradient, *Proceedings of the SIGGRAPH Asia*, (2012), p. 6.
- [96] K. He, J. Sun, X. Tang, Guided image filtering, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (6) (2013) 1397–1409.
- [97] C. Xi, Z. Wei, G. Fei, Image fusion algorithm based on adaptive pulse coupled neural networks in curvelet domain, *Proceedings of the International Conference on Signal Processing*, (2010), pp. 845–848.
- [98] Z. Shao, J. Liu, Q. Cheng, Fusion of infrared and visible images based on focus measure operators in the curvelet domain, *Appl. Optics* 51 (12) (2012) 1910–1921.
- [99] S. Quan, W. Qian, J. Guo, H. Zhao, Visible and infrared image fusion based on curvelet transform, *Proceedings of the International Conference on Systems and Informatics*, (2014), pp. 828–832.
- [100] Q. Li, J. Du, L. Xu, Visible and infrared video fusion using uniform discrete curvelet transform and spatial-temporal information, *Chinese J. Electron.* 24 (4) (2015) 761–766.
- [101] Y. Huang, K. Gao, C. Gong, L. Han, Y. Guo, Infrared and visible image fusion with the target marked based on multi-resolution visual attention mechanisms, Proceedings of the Selected Papers of the Chinese Society for Optical Engineering Conferences, (2017), p. 102552V.
- [102] G. Bhatnagar, Q.J. Wu, An image fusion framework based on human visual system in framelet domain, *Int. J. Wavelets, Multiresolution Inform. Process.* 10 (01) (2012) 1250002.
- [103] C. Wei, B. Zhou, W. Guo, A three scale image transformation for infrared and visible image fusion, *Proceedings of the International Conference on Information Fusion*, (2017), pp. 1–6.
- [104] P. Geng, Z. Wang, Z. Zhang, Z. Xiao, Image fusion by pulse couple neural network with shearlet, *Opt. Eng.* 51 (6) (2012) 067005.
- [105] W. Kong, Technique for gray-scale visual light and infrared image fusion based on non-subsampled shearlet transform, *Infrared Phys. Technol.* 63 (2014) 110–118.
- [106] W. Kong, B. Wang, Y. Lei, Technique for infrared and visible image fusion based on non-subsampled shearlet transform and spiking cortical model, *Infrared Phys. Technol.* 71 (2015) 87–98.
- [107] B. Zhang, X. Lu, H. Pei, Y. Zhao, A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled shearlet transform, *Infrared Phys. Technol.* 73 (2015) 286–297.
- [108] M. Yin, P. Duan, W. Liu, X. Liang, A novel infrared and visible image fusion algorithm based on shift-invariant dual-tree complex shearlet transform and sparse representation, *Neurocomputing* 226 (2017) 182–191.
- [109] Z. Liu, Y. Feng, H. Chen, L. Jiao, A fusion algorithm for infrared and visible based on guided filtering and phase congruity in nsst domain, *Opt. Lasers Eng.* 97 (2017) 71–77.
- [110] X. Zhou, W. Wang, Infrared and visible image fusion based on tetrolet transform, *Proceedings of the International Conference on Communications, Signal Processing, and Systems*, (2016), pp. 701–708.
- [111] Y. Huang, D. Zhang, B. Yuan, J. Kang, Fusion of visible and infrared image based on stationary tetrolet transform, *Proceedings of the Youth Academic Annual Conference of Chinese Association of Automation*, (2017), pp. 854–859.
- [112] X. Bai, F. Zhou, B. Xue, Fusion of infrared and visual images through region extraction by using multi scale center-surround top-hat transform, *Opt. Express* 19 (9) (2011) 8444–8457.
- [113] X. Bai, X. Chen, F. Zhou, Z. Liu, B. Xue, Multiscale top-hat selection transform based infrared and visual image fusion with emphasis on extracting regions of interest, *Infrared Phys. Technol.* 60 (2013) 81–93.
- [114] P. Zhu, Z. Huang, A fusion method for infrared-visible image and infrared-polarization image based on multi-scale center-surround top-hat transform, *Optical Rev.* 24 (3) (2017) 370–382.
- [115] P. Zhu, X. Ma, Z. Huang, Fusion of infrared-visible images using improved multi-scale top-hat transform and suitable fusion rules, *Infrared Phys. Technol.* 81 (2017) 282–295.
- [116] V. Naidu, Novel image fusion techniques using dct, *Int. J. Comput. Sci. Bus. Inf.* 5 (1) (2013) 1–18.
- [117] X. Zhou, X. Yin, R.-A. Liu, W. Wang, Infrared and visible image fusion technology based on directionlets transform, *EURASIP J. Wireless Commun. Netw.* 2013 (1) (2013) 42.
- [118] H. Hariharan, A. Koschan, B. Abidi, A. Gribok, M. Abidi, Fusion of visible and infrared images using empirical mode decomposition to improve face recognition, *Proceedings of the IEEE International Conference on Image Processing*, (2006), pp. 2049–2052.
- [119] W. Liang, Z. Liu, Region-based fusion of infrared and visible images using bidimensional empirical mode decomposition, *Proceedings of the International Conference on Educational and Information Technology*, (2010), p. 358.
- [120] J. Sharma, K. Sharma, V. Sahula, Hybrid image fusion scheme using self-fractional fourier functions and multivariate empirical mode decomposition, *Signal Process.* 100 (2014) 146–159.
- [121] P. Zhu, Z. Huang, H. Lei, Fusion of infrared and visible images based on bemd and nsfdb, *Infrared Phys. Technol.* 77 (2016) 82–93.
- [122] X. Zhang, X. Li, Y. Feng, H. Zhao, Z. Liu, Image fusion with internal generative mechanism, *Expert Syst. Appl.* 42 (5) (2015) 2382–2391.
- [123] Y. Song, J. Xiao, J. Yang, Z. Chai, Y. Wu, Research on mr-svd based visual and infrared image fusion, *Proceedings of the International Symposium on Optoelectronic Technology and Application*, (2016), p. 101571.
- [124] E.J. Candès, D.L. Donoho, Curvelets - a surprisingly effective nonadaptive representation for objects with edges, *Astron. Astrophys.* 283 (3) (1999) 1051–1057.
- [125] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, *Multiscale Model. Simul.* 5 (3) (2006) 861–899.
- [126] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, *SIAM J. Math. Anal.* 39 (1) (2007) 298–318.
- [127] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, Space-frequency quantization for image compression with directionlets, *IEEE Trans. Image Process.* 16 (7) (2007) 1761–1773.
- [128] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.T. Chi, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. Math. Phys. Eng. Sci.* 454 (1971) (1998) 903–995.
- [129] Y. Yang, Y. Que, S. Huang, P. Lin, Multiple visual features measurement with gradient domain guided filtering for multisensor image fusion, *IEEE Trans. Instrum. Measur.* 66 (4) (2017) 691–703.
- [130] G. Piella, A region-based multiresolution image fusion algorithm, *Proceedings of the International Conference on Information Fusion*, (2002), pp. 1557–1564.
- [131] Z. Xue, R.S. Blum, Y. Li, Fusion of visual and ir images for concealed weapon detection, *Proceedings of the International Conference on Information Fusion*, (2002), pp. 1198–1205.
- [132] B.A. Olshausen, et al., Emergence of simple-cell receptive field properties by

- learning a sparse code for natural images, *Nature* 381 (6583) (1996) 607–609.
- [133] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, *Proc. IEEE* 98 (6) (2010) 1031–1044.
- [134] Z. Zhang, Y. Xu, J. Yang, X. Li, D. Zhang, A survey of sparse representation: algorithms and applications, *IEEE Access* 3 (2015) 490–530.
- [135] Y. Gao, J. Ma, A.L. Yuille, Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples, *IEEE Trans. Image Process.* 26 (5) (2017) 2545–2560.
- [136] B. Yang, S. Li, Pixel-level image fusion with simultaneous orthogonal matching pursuit, *Inform. Fusion* 13 (1) (2012) 10–19.
- [137] Y. Bin, Y. Chao, H. Guoyu, Efficient image fusion with approximate sparse representation, *Int. J. Wavelets, Multiresolut. Inform. Process.* 14 (04) (2016) 1650024.
- [138] Y. Liu, Z. Wang, Simultaneous image fusion and denoising with adaptive sparse representation, *IET Image Process.* 9 (5) (2014) 347–357.
- [139] B. Yang, S. Li, Visual attention guided image fusion with sparse representation, *Optik-Int. J. Light Electron Opt.* 125 (17) (2014) 4881–4888.
- [140] R. Wang, L. Du, Infrared and visible image fusion based on random projection and sparse representation, *Int. J. Remote Sens.* 35 (5) (2014) 1640–1652.
- [141] 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.
- [142] N. Aishwarya, C.B. Thangammal, An image fusion framework using novel dictionary based sparse representation, *Multimed. Tools Appl.* 76 (21) (2017) 21869–21888.
- [143] L. Chang, X. Feng, R. Zhang, H. Huang, W. Wang, C. Xu, Image decomposition fusion method based on sparse representation and neural network, *Appl. Opt.* 56 (28) (2017) 7969–7977.
- [144] Z. Gao, M. Yang, C. Xie, Space target image fusion method based on image clarity criterion, *Opt. Eng.* 56 (5) (2017) 053102.
- [145] Q. Zhang, Y. Fu, H. Li, J. Zou, Dictionary learning method for joint sparse representation-based image fusion, *Opt. Eng.* 52 (5) (2013) 057006.
- [146] X. Lu, B. Zhang, Y. Zhao, H. Liu, H. Pei, The infrared and visible image fusion algorithm based on target separation and sparse representation, *Infrared Phys. Technol.* 67 (2014) 397–407.
- [147] Y. Yao, P. Guo, X. Xin, Z. Jiang, Image fusion by hierarchical joint sparse representation, *Cognit. Comput.* 6 (3) (2014) 281–292.
- [148] H. Yin, Sparse representation with learned multiscale dictionary for image fusion, *Neurocomputing* 148 (2015) 600–610.
- [149] Y. Li, F. Li, B. Bai, Q. Shen, Image fusion via nonlocal sparse k-svd dictionary learning, *Appl. Opt.* 55 (7) (2016) 1814–1823.
- [150] M. Kim, D.K. Han, H. Ko, Joint patch clustering-based dictionary learning for multimodal image fusion, *Inform. Fusion* 27 (2016) 198–214.
- [151] K. Wang, G. Qi, Z. Zhu, Y. Chai, A novel geometric dictionary construction approach for sparse representation based image fusion, *Entropy* 19 (7) (2017) 306.
- [152] Z. Zhu, H. Yin, Y. Chai, Y. Li, G. Qi, A novel multi-modality image fusion method based on image decomposition and sparse representation, *Inform. Sci.* (2017) 1–14.
- [153] C. Liu, Y. Qi, W. Ding, Infrared and visible image fusion method based on saliency detection in sparse domain, *Infrared Phys. Technol.* 83 (2017) 94–102.
- [154] Z. Gao, C. Zhang, Texture clear multi-modal image fusion with joint sparsity model, *Optik-Int. J. Light Electron Opt.* 130 (2017) 255–265.
- [155] M. Aharon, M. Elad, A. Bruckstein, k-svd: An algorithm for designing overcomplete dictionaries for sparse representation, *IEEE Trans. Image Process.* 15 (11) (2006) 4311–4322.
- [156] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory* 53 (12) (2007) 4655–4666.
- [157] J.A. Tropp, A.C. Gilbert, M.J. Strauss, Algorithms for simultaneous sparse approximation. part i: Greedy pursuit, *Signal Process.* 86 (3) (2006) 572–588.
- [158] H. Jiang, Y. Tian, Fuzzy image fusion based on modified self-generating neural network, *Expert Syst. Appl.* 38 (7) (2011) 8515–8523.
- [159] X.-B. Qu, J.-W. Yan, H.-Z. Xiao, Z.-Q. Zhu, Image fusion algorithm based on spatial frequency-motivated pulse coupled neural networks in nonsubsampled contourlet transform domain, *Acta Automatica Sinica* 34 (12) (2008) 1508–1514.
- [160] Y. Chen, Z. Qin, Penn-based image fusion in compressed domain, *Math. Probl. Eng.* 2015 (2015) 536215.
- [161] X. Qu, C. Hu, J. Yan, Image fusion algorithm based on orientation information motivated pulse coupled neural networks, *Proceedings of the World Congress on Intelligent Control and Automation*, (2008), pp. 2437–2441.
- [162] R. Eckhorn, H.J. Reitbock, M. Arndt, P. Dicke, A neural network for feature linking via synchronous activity: Results from cat visual cortex and from simulations, *Can. J. Microbiol.* 46 (8) (1989) 759–763.
- [163] J.L. Johnson, Pulse-coupled neural nets: translation, rotation, scale, distortion, and intensity signal invariance for images, *Appl. Optics* 33 (26) (1994) 6239–6253.
- [164] R.P. Broussard, S.K. Rogers, Physiologically motivated image fusion using pulse-coupled neural networks, *Proceedings of the Applications and Science of Artificial Neural Networks II*, (1996), pp. 372–384.
- [165] 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, *Information Fusion* 42 (2018) 158–173.
- [166] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, *Inform. Fusion* 36 (2017) 191–207.
- [167] C. Du, S. Gao, Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network, *IEEE Access* 5 (2017) 15750–15761.
- [168] J. Zhong, B. Yang, Y. Li, F. Zhong, Z. Chen, Image fusion and super-resolution with convolutional neural network, *Proceedings of the Chinese Conference on Pattern Recognition*, (2016), pp. 78–88.
- [169] Y. Liu, X. Chen, J. Cheng, H. Peng, A medical image fusion method based on convolutional neural networks, *Proceedings of the International Conference on Information Fusion*, (2017), pp. 1–7.
- [170] F. Palsson, J.R. Steinsson, M.O. Ulfarsson, Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network, *IEEE Geosci. Remote Sens. Lett.* 14 (5) (2017) 639–643.
- [171] G. Masi, D. Cozzolino, L. Verdoliva, G. Scarpa, Pansharpening by convolutional neural networks, *Remote Sens.* 8 (7) (2016) 594.
- [172] Y. Liu, X. Chen, J. Cheng, H. Peng, Z. Wang, Infrared and visible image fusion with convolutional neural networks, *Int. J. Wavelets, Multiresolut. Inform. Process.* 16 (2) (2018) 1850018.
- [173] M.D. Zeiler, D. Krishnan, G.W. Taylor, R. Fergus, Deconvolutional networks, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2010), pp. 2528–2535.
- [174] H.B. Mitchell, *Image Fusion: Theories, Techniques and Applications*, Springer Science & Business Media, 2010.
- [175] Y. Zhou, A. Mayyas, M.A. Omar, Principal component analysis-based image fusion routine with application to automotive stamping split detection, *Res. Nondestruct. Eval.* 22 (2) (2011) 76–91.
- [176] Y. Zheng, E.A. Essock, B.C. Hansen, An advanced image fusion algorithm based on wavelet transform: incorporation with pca and morphological processing, *Proceedings of SPIE*, (2004), pp. 177–187.
- [177] S.S. Kumar, S. Muttan, Pca-based image fusion, *Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, (2006), p. 62331T.
- [178] U. Patil, U. Mudugudi, Image fusion using hierarchical PCA, *Proceedings of the International Conference on Image Information Processing*, (2011), pp. 1–6.
- [179] V. Naidu, Hybrid ddct-pca based multi sensor image fusion, *J. Opt.* 43 (1) (2014) 48–61.
- [180] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis? *J. ACM* 58 (3) (2011) 11.
- [181] R. Ibrahim, J. Alirezai, P. Babyn, Pixel level jointed sparse representation with RPCA image fusion algorithm, *Proceedings of the International Conference on Telecommunications and Signal Processing*, (2015), pp. 592–595.
- [182] Z. Fu, X. Wang, J. Xu, N. Zhou, Y. Zhao, Infrared and visible images fusion based on RPCA and NSCT, *Infrared Phys. Technol.* 77 (2016) 114–123.
- [183] Z. Wang, J. Deller, B.D. Fleet, Pixel-level multisensor image fusion based on matrix completion and robust principal component analysis, *J. Electron. Imaging* 25 (1) (2016) 013007.
- [184] N. Cvejic, D. Bull, N. Canagarajah, Region-based multimodal image fusion using ICA bases, *IEEE Sens. J.* 7 (5) (2007) 743–751.
- [185] N. Mitianoudis, T. Stathaki, Pixel-based and region-based image fusion schemes using ICA bases, *Inform. Fusion* 8 (2) (2007) 131–142.
- [186] N. Mitianoudis, S.-A. Antonopoulos, T. Stathaki, Region-based ICA image fusion using textural information, *Proceedings of the International Conference on Digital Signal Processing*, (2013), pp. 1–6.
- [187] Z. Omar, N. Mitianoudis, T. Stathaki, Region-based image fusion using a combinatorial chebyshev-ica method, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, (2011), pp. 1213–1216.
- [188] X. Li, G. Cui, Y. Dong, Graph regularized non-negative low-rank matrix factorization for image clustering, *IEEE Trans. Cybern.* 47 (11) (2017) 3840–3853.
- [189] V. Tsagaris, V. Anastassopoulos, Fusion of visible and infrared imagery for night color vision, *Displays* 26 (4) (2005) 191–196.
- [190] W. Kong, Y. Lei, Y. Lei, J. Zhang, Technique for image fusion based on non-subsampled contourlet transform domain improved nmf, *Sci. China Inform. Sci.* 53 (12) (2010) 2429–2440.
- [191] J. Mou, W. Gao, Z. Song, Image fusion based on non-negative matrix factorization and infrared feature extraction, *Proceedings of the International Congress on Image and Signal Processing*, (2013), pp. 1046–1050.
- [192] A. Toet, Computational versus psychophysical bottom-up image saliency: a comparative evaluation study, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (11) (2011) 2131–2146.
- [193] J. Han, E.J. Pauwels, P. De Zeeuw, Fast saliency-aware multi-modality image fusion, *Neurocomputing* 111 (2013) 70–80.
- [194] J. Zhao, X. Gao, Y. Chen, H. Feng, D. Wang, Multi-window visual saliency extraction for fusion of visible and infrared images, *Infrared Phys. Technol.* 76 (2016) 295–302.
- [195] T. Shibata, M. Tanaka, M. Okutomi, Visible and near-infrared image fusion based on visually salient area selection, *Proceedings of the Digital Photography*, (2015), p. 94040G.
- [196] Q. Zhang, X. Mal dague, An infrared-visible image fusion scheme based on nsct and compressed sensing, *Proceedings of the SPIE Defense + Security*, (2015), p. 94740Y.
- [197] G. He, D. Dong, Z. Xia, S. Xing, Y. Wei, An infrared and visible image fusion method based on non-subsampled contourlet transform and joint sparse representation, *Proceedings of the IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, (2016), pp. 492–497.
- [198] H. Yin, Z. Liu, B. Fang, Y. Li, A novel image fusion approach based on compressive sensing, *Opt. Commun.* 354 (2015) 299–313.
- [199] Y. Ma, J. Chen, C. Chen, F. Fan, J. Ma, Infrared and visible image fusion using total variation model, *Neurocomputing* 202 (2016) 12–19.
- [200] H. Guo, Y. Ma, X. Mei, J. Ma, Infrared and visible image fusion based on total variation and augmented lagrangian, *JOSA A* 34 (11) (2017) 1961–1968.
- [201] Q. Jiang, X. Jin, S.-J. Lee, S. Yao, A novel multi-focus image fusion method based

- on stationary wavelet transform and local features of fuzzy sets, *IEEE Access* (2017) 20286–20302.
- [202] Z. Fu, X. Dai, Y. Li, H. Wu, X. Wang, An improved visible and infrared image fusion based on local energy and fuzzy logic, *Proceedings of the International Conference on Signal Processing*, (2014), pp. 861–865.
- [203] S.K. Kashyap, Ir and color image fusion using interval type 2 fuzzy logic system, *Proceedings of the International Conference on Cognitive Computing and Information Processing*, (2015), pp. 1–4.
- [204] X. Bai, Infrared and visual image fusion through fuzzy measure and alternating operators, *Sensors* 15 (7) (2015) 17149–17167.
- [205] Z. Wang, B. Hong, Entropy based multi-resolution visible-infrared image fusion, *Proceedings of the IEEE Symposium on Electrical & Electronics Engineering*, (2012), pp. 639–643.
- [206] T. Shibata, M. Tanaka, M. Okutomi, Versatile visible and near-infrared image fusion based on high visibility area selection, *J. Electron. Imaging* 25 (1) (2016). 013016–013016.
- [207] X. Bai, Infrared and visual image fusion through feature extraction by morphological sequential toggle operator, *Infrared Phys. Technol.* 71 (2015) 77–86.
- [208] X. Bai, Morphological center operator based infrared and visible image fusion through correlation coefficient, *Infrared Phys. Technol.* 76 (2016) 546–554.
- [209] Y. Zhang, L. Zhang, X. Bai, L. Zhang, Infrared and visual image fusion through infrared feature extraction and visual information preservation, *Infrared Phys. Technol.* 83 (2017) 227–237.
- [210] S.S. Malik, B. Shivprasad, G. Maruthi, J. Kuriakose, V. Amruth, Feature level image fusion, *Proceedings of CCSO*, (2013), pp. 42–46.
- [211] J. Adu, S. Xie, J. Gan, Image fusion based on visual salient features and the cross-contrast, *J. Visual Communun. Image Represent.* 40 (2016) 218–224.
- [212] F. Wu, B. Wang, X. Yi, M. Li, J. Hao, H. Qin, H. Zhou, Visible and infrared image registration based on visual salient features, *J. Electron. Imaging* 24 (5) (2015) 053017.
- [213] W. He, W. Feng, Y. Peng, Q. Chen, G. Gu, Z. Miao, Multi-level image fusion and enhancement for target detection, *Optik-Int. J. Light Electron. Opt.* 126 (11) (2015) 1203–1208.
- [214] R. Wang, C. Tao, Symbol-based image fusion for object recognition, *Proceedings of the International Conference on Multisource-Multisensor Information Fusion*, (1998), pp. 509–515.
- [215] V.-E. Neagoie, A.-D. Ropot, A.C. Mugioiu, Real time face recognition using decision fusion of neural classifiers in the visible and thermal infrared spectrum, *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*, (2007), pp. 301–306.
- [216] Y. Zhao, Y. Yin, D. Fu, Decision-level fusion of infrared and visible images for face recognition, *Proceedings of the Chinese Control and Decision Conference*, (2008), pp. 2411–2414.
- [217] A. Wang, J. Jiang, H. Zhang, Multi-sensor image decision level fusion detection algorithm based on ds evidence theory, *Proceedings of the International Conference on Instrumentation and Measurement, Computer, Communication and Control*, (2014), pp. 620–623.
- [218] 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.
- [219] 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.
- [220] B. Zitova, J. Flusser, Image registration methods: a survey, *Image Vis. Comput.* 21 (11) (2003) 977–1000.
- [221] J. Ma, W. Qiu, J. Zhao, Y. Ma, A.L. Yuille, Z. Tu, Robust  $l_2$  estimation of transformation for non-rigid registration, *IEEE Trans. Image Process.* 63 (5) (2015) 1115–1129.
- [222] X. Chen, G. Zhai, J. Wang, C. Hu, Y. Chen, Color guided thermal image super resolution, *Proceedings of the Visual Communications and Image Processing*, (2016), pp. 1–4.
- [223] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, *Proceedings of the International Joint Conference on Artificial Intelligence*, (1981), pp. 674–679.
- [224] L.G. Brown, A survey of image registration techniques, *ACM Comput. Surv.* 24 (4) (1992) 325–376.
- [225] A. Sotiras, C. Davatzikos, N. Paragios, Deformable medical image registration: a survey, *IEEE Trans. Med. Imaging* 32 (7) (2013) 1153–1190.
- [226] J. Ma, J. Jiang, C. Liu, Y. Li, Feature guided gaussian mixture model with semi-supervised em and local geometric constraint for retinal image registration, *Inform. Sci.* 417 (2017) 128–142.
- [227] A. Wong, D.A. Clausi, ARRSI: Automatic registration of remote-sensing images, *IEEE Trans. Geosci. Remote Sens.* 45 (5) (2007) 1483–1493.
- [228] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, J. Tian, Robust feature matching for remote sensing image registration via locally linear transforming, *IEEE Trans. Geosci. Remote Sens.* 53 (12) (2015) 6469–6481.
- [229] F. Zhao, Q. Huang, W. Gao, Image matching by normalized cross-correlation, *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, (2006), pp. 729–732.
- [230] R.N. Bracewell, R.N. Bracewell, *The Fourier Transform and its Applications*, McGraw-Hill, New York, 1986.
- [231] P. Viola, W.M. Wells III, Alignment by maximization of mutual information, *Int. J. Comput. Vis.* 24 (2) (1997) 137–154.
- [232] A. Roche, G. Malandain, X. Pennec, N. Ayache, The correlation ratio as a new similarity measure for multimodal image registration, *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, (1998), pp. 1115–1124.
- [233] B.S. Reddy, B.N. Chatterji, An fft-based technique for translation, rotation, and scale-invariant image registration, *IEEE Trans. Image Process.* 5 (8) (1996) 1266–1271.
- [234] G. Rabatel, S. Labb  , Registration of visible and near infrared unmanned aerial vehicle images based on fourier-mellin transform, *Precision Agri.* 17 (5) (2016) 564–587.
- [235] G.-A. Bilodeau, A. Torabi, F. Morin, Visible and infrared image registration using trajectories and composite foreground images, *Image and Vision Computing* 29 (1) (2011) 41–50.
- [236] S.J. Kratosky, M.M. Trivedi, Mutual information based registration of multimodal stereo videos for person tracking, *Comput. Vision Image Understand.* 106 (2) (2007) 270–287.
- [237] A. Rangarajan, H. Chui, J.S. Duncan, Rigid point feature registration using mutual information, *Med. Image Anal.* 3 (4) (1999) 425–440.
- [238] S.E. Grigorescu, N. Petkov, P. Kruizinga, Comparison of texture features based on gabor filters, *IEEE Trans. Image Process.* 11 (10) (2002) 1160–1167.
- [239] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vis.* 60 (2) (2004) 91–110.
- [240] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2005), pp. 886–893.
- [241] T. Hrka  , Z. Kalafati  , J. Krapac, Infrared-visual image registration based on corners and hausdorff distance, *Proceedings of the Scandinavian Conference on Image Analysis*, (2007), pp. 383–392.
- [242] T. Tian, X. Mei, Y. Yu, C. Zhang, X. Zhang, Automatic visible and infrared face registration based on silhouette matching and robust transformation estimation, *Infrared Phys. Technol.* 69 (2015) 145–154.
- [243] J. Han, E.J. Pauwels, P. De Zeeuw, Visible and infrared image registration in man-made environments employing hybrid visual features, *Pattern Recognit. Lett.* 34 (1) (2013) 42–51.
- [244] W. Yang, X. Wang, B. Moran, A. Wheaton, N. Cooley, Efficient registration of optical and infrared images via modified sobel edging for plant canopy temperature estimation, *Comput. Electr. Eng.* 38 (5) (2012) 1213–1221.
- [245] 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.
- [246] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (4) (2002) 509–522.
- [247] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Commun. ACM* 24 (6) (1981) 381–395.
- [248] J. Ma, J. Zhao, J. Tian, Z. Tu, A.L. Yuille, Robust estimation of nonrigid transformation for point set registration, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2013), pp. 2147–2154.
- [249] J. Ma, J. Zhao, J. Tian, X. Bai, Z. Tu, Regularized vector field learning with sparse approximation for mismatch removal, *Pattern Recognit.* 46 (12) (2013) 3519–3532.
- [250] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition review, *Comput. Vis. Image Understand.* 97 (1) (2005) 103–135.
- [251] J. Heo, S.G. Kong, B.R. Abidi, M.A. Abidi, Fusion of visual and thermal signatures with eyeglass removal for robust face recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop*, (2004), p. 122.
- [252] I. Ulusoy, H. Yuruk, New method for the fusion of complementary information from infrared and visual images for object detection, *IET Image Processing* 5 (1) (2011) 36–48.
- [253] S. Gao, Y. Cheng, Y. Zhao, Method of visual and infrared fusion for moving object detection, *Opt. Lett.* 38 (11) (2013) 1981–1983.
- [254] H. Liu, F. Sun, Fusion tracking in color and infrared images using joint sparse representation, *Sci. China Inform. Sci.* 55 (3) (2012) 590–599.
- [255] D. Smith, S. Singh, Approaches to multisensor data fusion in target tracking: a survey, *IEEE Trans. Knowl. Data Eng.* 18 (12) (2006) 1696–1710.
- [256] C. O’Conaire, N.E. O’Connor, E. Cooke, A.F. Smeaton, Comparison of fusion methods for thermo-visual surveillance tracking, *Proceedings of the International Conference on Information Fusion*, (2006), pp. 1–7.
- [257] S. Yin, L. Cao, Y. Ling, G. Jin, One color contrast enhanced infrared and visible image fusion method, *Infrared Phys. Technol.* 53 (2) (2010) 146–150.
- [258] T. Stathaki, *Image Fusion: Algorithms and Applications*, Academic Press, 2011.
- [259] M.-Y. Fu, C. Zhao, Fusion of infrared and visible images based on the second generation curvelet transform, *J. Infrared Millimeter Waves* 28 (4) (2009) 254–258.
- [260] H. Li, W. Ding, X. Cao, C. Liu, Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing, *Remote Sens.* 9 (5) (2017) 441.
- [261] F. Omri, S. Foufou, M. Abidi, Nir and visible image fusion for improving face recognition at long distance, *Proceedings of the International Conference on Image and Signal Processing*, (2014), pp. 549–557.
- [262] S. Singh, A. Gyaourova, G. Bebis, I. Pavlidis, Infrared and visible image fusion for face recognition, *Proceedings of SPIE*, 5404 (2004), pp. 585–596.
- [263] G. Bebis, A. Gyaourova, S. Singh, I. Pavlidis, Face recognition by fusing thermal infrared and visible imagery, *Image Vis. Comput.* 24 (7) (2006) 727–742.
- [264] M. Uzair, A. Mahmood, A. Mian, C. McDonald, Periocular region-based person identification in the visible, infrared and hyperspectral imagery, *Neurocomputing* 149 (2015) 854–867.
- [265] F. Shamsafar, H. Seyedarabi, A. Aghagolzadeh, Fusing the information in visible light and near-infrared images for iris recognition, *Mach. Vis. Appl.* 25 (4) (2014)

- 881–899.
- [266] T. Elguebaly, N. Bouguila, Finite asymmetric generalized gaussian mixture models learning for infrared object detection, *Comput. Vis. Image Understand.* 117 (12) (2013) 1659–1671.
- [267] A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, J. Xu, A.M. López, Pedestrian detection at day/night time with visible and fir cameras: A comparison, *Sensors* 16 (6) (2016) 820.
- [268] E. Fendri, R.R. Boukhriss, M. Hammami, Fusion of thermal infrared and visible spectra for robust moving object detection, *Pattern Anal. Appl.* 20 (4) (2017) 907–926.
- [269] A. Fernández-Caballero, J.C. Castillo, M.T. López, J. Serrano-Cuerda, M.V. Sokolova, Int 3-horus framework for multispectrum activity interpretation in intelligent environments, *Expert Syst. Appl.* 40 (17) (2013) 6715–6727.
- [270] J.C. Castillo, A. Fernández-Caballero, J. Serrano-Cuerda, M.T. López, A. Martínez-Rodrigo, Smart environment architecture for robust people detection by infrared and visible video fusion, *J. Ambient Intel. Humanized Comput.* 8 (2) (2017) 223–237.
- [271] S.R. Schnelle, A.L. Chan, Enhanced target tracking through infrared-visible image fusion, *Proceedings of the International Conference on Information Fusion*, (2011), pp. 1–8.
- [272] K.S. Kumar, G. Kavitha, R. Subramanian, G. Ramesh, Visual and thermal image fusion for UAV based target tracking, *MATLAB-A Ubiquitous Tool for the Practical Engineer*, InTech, 2011.
- [273] H. Li, J. Zhang, J. Sun, A visual analytics approach for deterioration risk analysis of ancient frescoes, *J. Visual.* 19 (3) (2016) 529–542.
- [274] G. Bhatnagar, Q.J. Wu, B. Raman, Navigation and surveillance using night vision and image fusion, *Proceedings of the IEEE Symposium on Industrial Electronics and Applications*, (2011), pp. 342–347.
- [275] G. Bhatnagar, Z. Liu, A novel image fusion framework for night-vision navigation and surveillance, *Signal, Image Video Process.* 9 (1) (2015) 165–175.
- [276] N. Paramanandham, K. Rajendiran, Multi sensor image fusion for surveillance applications using hybrid image fusion algorithm, *Multim. Tools Appl.* (2017) 1–32.
- [277] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, J. Hai, K. He, A survey of infrared and visual image fusion methods, *Infrared Phys. Technol.* 85 (2017) 478–501.
- [278] A. Toet, Natural colour mapping for multiband nightvision imagery, *Inform. Fus.* 4 (3) (2003) 155–166.
- [279] A. Toet, E.M. Franken, Perceptual evaluation of different image fusion schemes, *Displays* 24 (1) (2003) 25–37.
- [280] Z. Cheng, Q. Yang, B. Sheng, Deep colorization, *Proceedings of the IEEE International Conference on Computer Vision*, (2015), pp. 415–423.
- [281] X. Chang, L. Jiao, F. Liu, F. Xin, Multicontourlet-based adaptive fusion of infrared and visible remote sensing images, *IEEE Geosci. Remote Sens. Lett.* 7 (3) (2010) 549–553.
- [282] A. Ahmad, M.M. Riaz, A. Ghafoor, T. Zaidi, An improved infrared/visible fusion for astronomical images, *Adv. Astron.* 2015 (2015) 203872.
- [283] X. Lu, J. Zhang, T. Li, G. Zhang, Synergetic classification of long-wave infrared hyperspectral and visible images, *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 8 (7) (2015) 3546–3557.
- [284] M. Eslami, A. Mohammadzadeh, Developing a spectral-based strategy for urban object detection from airborne hyperspectral tir and visible data, *IEEE J. Selected Top. Appl. Earth Observ. Remote Sens.* 9 (5) (2016) 1808–1816.
- [285] L. Han, B. Wulie, Y. Yang, H. Wang, Direct fusion of geostationary meteorological satellite visible and infrared images based on thermal physical properties, *Sensors* 15 (1) (2015) 703–714.
- [286] J. Liang, W. Zhang, L. Ren, H. Ju, E. Qu, Polarimetric dehazing method for visibility improvement based on visible and infrared image fusion, *Appl. Opt.* 55 (29) (2016) 8221–8226.
- [287] L. Zalud, P. Kocmanova, F. Burian, T. Jilek, Color and thermal image fusion for augmented reality in rescue robotics, *Proceedings of the International Conference on Robotic, Vision, Signal Processing & Power Applications*, (2014), pp. 47–55.
- [288] F. Mendoza, R. Lu, H. Cen, Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content, *Postharvest Biol. Technol.* 73 (2012) 89–98.
- [289] B.V. Hanna, A.M. Gorbach, F.A. Gage, P.A. Pinto, J.S. Silva, L.G. Gilfillan, A.D. Kirk, E.A. Elster, Intraoperative assessment of critical biliary structures with visible range/infrared image fusion, *J. Am. College Surgeons* 206 (6) (2008) 1227–1231.
- [290] M. Bartys, B. Putz, A. Antoniewicz, L. Zbrzezny, Real-time single fpga-based multimodal image fusion system, *Proceedings of the IEEE International Conference on Imaging Systems and Techniques*, (2012), pp. 460–465.
- [291] M.D. Jasiusas, D.A. Kearney, J. Hopf, G.B. Wigley, Image fusion for uninhabited airborne vehicles, *Proceedings of the IEEE International Conference on Field-Programmable Technology*, (2002), pp. 348–351.
- [292] A. Antoniewicz, Fpga implementation of decomposition methods for real-time image fusion, *Image Process. Commun. Chall.* 4 (2013) 163–170.
- [293] D. Besiris, V. Tsagaris, N. Fragoulis, C. Theoharatos, An fpga-based hardware implementation of configurable pixel-level color image fusion, *IEEE Trans. Geosci. Remote Sens.* 50 (2) (2012) 362–373.
- [294] S.-H. Yoo, J.-H. Park, C.-S. Jeong, Accelerating multi-scale image fusion algorithms using cuda, *Proceedings of the International Conference of Soft Computing and Pattern Recognition*, (2009), pp. 278–282.
- [295] Y. Zhang, Y. Li, K. Zhang, H. Wang, Fusion of infrared and visible images based on nonsubsampled contourlet transform, *Comput. Eng. Appl.* 47 (3) (2011) 196–198.
- [296] G. Kiss, J.P. Asen, B. Amundsen, P. Claus, J. D'hooge, H.G. Torp, Multi-modal cardiac image fusion and visualization on the GPU, *Proceedings of the IEEE International Ultrasonics Symposium*, (2011), pp. 254–257.
- [297] J. Lu, B. Zhang, H. He, H. Zhang, The high-pass filtering fusion based on GPU, *Proceedings of the International Symposium on Computer Science and Society*, (2011), pp. 122–125.
- [298] Y. Han, Y. Cai, Y. Cao, X. Xu, A new image fusion performance metric based on visual information fidelity, *Inform. Fus.* 14 (2) (2013) 127–135.
- [299] Y. Chen, R.S. Blum, A new automated quality assessment algorithm for image fusion, *Image Vis. Comput.* 27 (10) (2009) 1421–1432.
- [300] 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.
- [301] M. Deshmukh, U. Bhosale, Image fusion and image quality assessment of fused images, *Int. J. Image Process.* 4 (5) (2010) 484–508.
- [302] P. Pradham, N.H. Younan, R.L. King, *Concepts of Image Fusion in Remote Sensing Applications*, Academic Press: London, UK, 2008.
- [303] B.S. Kumar, Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform, *Signal Image Video Process.* 7 (6) (2013) 1125–1143.
- [304] A.R. Pal, A. Singha, A comparative analysis of visual and thermal face image fusion based on different wavelet family, *Proceedings of the International Conference on Innovations in Electronics, Signal Processing and Communication*, (2017), pp. 213–218.
- [305] M. Kumar, et al., A local correlation and directive contrast based image fusion, *Proceedings of the International Conference on Computer Vision and Image Processing*, (2017), pp. 419–430.
- [306] V. Aslantas, R. Kurban, A comparison of criterion functions for fusion of multi-focus noisy images, *Opt. Commun.* 282 (16) (2009) 3231–3242.
- [307] S. Krishnamoorthy, K. Soman, Implementation and comparative study of image fusion algorithms, *Int. J. Comput. Appl.* 9 (2) (2010) 10–20.
- [308] Y. Zheng, E.A. Esock, B.C. Hansen, A.M. Haun, A new metric based on extended spatial frequency and its application to dwt based fusion algorithms, *Inform. Fus.* 8 (2) (2007) 177–192.
- [309] V. Petrović, Subjective tests for image fusion evaluation and objective metric validation, *Inform. Fus.* 8 (2) (2007) 208–216.
- [310] S. Vekkot, P. Shukla, A novel architecture for wavelet based image fusion, *World Acad. Sci. Eng. Technol.* 57 (2009) 372–377.
- [311] H.R. Sheikh, A.C. Bovik, Image information and visual quality, *IEEE Trans. Image Process.* 15 (2) (2006) 430–444.
- [312] W. Shi, C. Zhu, Y. Tian, J. Nichol, Wavelet-based image fusion and quality assessment, *Int. J. Appl. Earth Observ. Geoinform.* 6 (3) (2005) 241–251.
- [313] J.W. Roberts, J. Van Aardt, F. Ahmed, Assessment of image fusion procedures using entropy, image quality, and multispectral classification, *J. Appl. Remote Sens.* 2 (1) (2008) 023522.
- [314] G. Qu, D. Zhang, P. Yan, Information measure for performance of image fusion, *Electron. Lett.* 38 (7) (2002) 313–315.
- [315] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (4) (2004) 600–612.
- [316] Z. Wang, A.C. Bovik, A universal image quality index, *IEEE Signal Process. Lett.* 9 (3) (2002) 81–84.
- [317] G. Piella, H. Heijmans, A new quality metric for image fusion, *Proceedings of the International Conference on Image Processing*, (2003), pp. 173–176.
- [318] A.M. Eskicioglu, P.S. Fisher, Image quality measures and their performance, *IEEE Trans. Commun.* 43 (12) (1995) 2959–2965.
- [319] M. Hossny, S. Nahavandi, D. Creighton, A. Bhatti, Image fusion performance metric based on mutual information and entropy driven quadtree decomposition, *Electron. Lett.* 46 (18) (2010) 1266–1268.
- [320] M. Hossny, S. Nahavandi, D. Creighton, Comments on ‘information measure for performance of image fusion’, *Electron. Lett.* 44 (18) (2008) 1066–1067.
- [321] X. Luo, S. Wang, D. Yuan, Weber-aware weighted mutual information evaluation for infrared-visible image fusion, *J. Appl. Remote Sens.* 10 (4) (2016) 045004.
- [322] Z. Liu, D.S. Forsyth, R. Laganière, A feature-based metric for the quantitative evaluation of pixel-level image fusion, *Comput. Vis. Image Understand.* 109 (1) (2008) 56–68.
- [323] V. Petrović, C. Xydeas, Objective image fusion performance characterisation, *Proceedings of the IEEE International Conference on Computer Vision*, (2005), pp. 1866–1871.
- [324] V. Aslantas, E. Bendas, A new image quality metric for image fusion: The sum of the correlations of differences, *AEU-Int. J. Electron. Commun.* 69 (12) (2015) 1890–1896.
- [325] Y.-J. Rao, In-fibre bragg grating sensors, *Measurem. Sci. Technol.* 8 (4) (1997) 355.