



Novel and fast EMD-based image fusion via morphological filter

Qi Xie^{1,2} · Jianping Hu² · Xiaochao Wang³ · Daochang Zhang² · Hong Qin⁴

Accepted: 6 June 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

This paper presents a novel and fast EMD-based (empirical mode decomposition-based) image fusion approach via morphological filter. Firstly, we develop a multi-channel bidimensional EMD method based on morphological filter to conduct image fusion. It uses the morphological expansion and erosion filters to compute the upper and lower envelopes of a multi-channel image in the sifting processing, and can decompose the input source images into several intrinsic mode functions (IMFs) with different scales and a residue. It significantly improves the computation efficiency of EMD for multi-channel images while maintaining the decomposition quality. Secondly, we adopt a patch-based fusion strategy with overlapping partition to fuse the IMFs and residue instead of the pixel-based fusion way usually used in EMD-based image fusion, where an energy-based maximum selection rule is designed to fuse the IMFs, and the feature information extracted by IMFs is used as a guide to merge the residue. Such strategy can extract the salient information of the source images well and can also reduce the spatial artifacts introduced by the noisy characteristics of the pixel-wise maps. A large number of comparative experiments on the fusion of several commonly used image data sets with multi-focus and multi-modal images, show that our newly proposed fusion method can obtain much better results than the existing EMD-based image fusion approaches. Furthermore, it is very competitive with the state-of-the-art image fusion methods in visualization, objective metrics, and time performance. The code of the proposed method can be downloaded from: <https://github.com/neepuhjp/MFMBEMD-ImageFusion>.

Keywords Image fusion · Empirical mode decomposition · Morphological filter · Intrinsic mode functions

1 Introduction

With the rapid progress of data acquiring technology, a variety of sensors capturing images are emerging. Although each sensor has irreplaceable advantages in its appropriate working condition and range, the information obtained according to a single imaging sensor is incomplete. Image fusion is an important technique to integrate image data of the same target collected by multi-sensors, so as to extract the favorable information in each sensor to the greatest extent and synthesize it into high-quality images. They are more suitable for human visual perception and further image processing tasks [1]. Image fusion has been extensively used in computer vision, military, remote sensing, etc. [2].

The classical empirical mode decomposition (EMD) proposed by Huang et al. [3] is a powerful tool for processing non-stationary and nonlinear one-dimensional signals, which can adaptively decompose a 1D time series signal into several intrinsic mode functions (IMFs) and a residue through an iterative sifting processing. It has also been extended to 2D images called bidimensional empirical mode decomposition (BEMD) and can adaptively represent the feature informa-

✉ Jianping Hu
neduhjp307@163.com

✉ Xiaochao Wang
wangxiaochao18@163.com

Qi Xie
xieqi_19820302@126.com

Daochang Zhang
daochangzhang@126.com

Hong Qin
qin@cs.stonybrook.edu

¹ School of Mathematics, Jilin University,
Changchun 130012, China

² School of Sciences, Northeast Electric Power University,
Jilin 132012, China

³ School of Mathematical Sciences, Tiangong University,
Tianjin 300387, China

⁴ Department of Computer Science, Stony Brook University
(SUNY), Stony Brook, NY 11794-2424, USA

tion with different scales of input source images. Many researchers have applied EMD into image fusion to overcome the artifacts of predefined basis functions in the fusion methods based on classical transforms such as Fourier transform and wavelet transform. Unfortunately, the existing EMD-based image fusion approaches [4–11] have limited impact in image fusion since there are still some limitations that need to be addressed. Firstly, the computation efficiency of such methods is unsatisfactory since the envelope generation in the sifting processing of 2D images becomes very time-consuming as the image size rapidly increasing especially for the surface interpolation-based EMD methods. Secondly, they mainly employ the pixel-based fusion strategy to merge each component of EMD, which may introduce some spatial artifacts generated by the noisy characteristics of the pixel-wise maps. Furthermore, the fusion rules in such methods cannot capture the salient information of input source images well, therefore they generate fuzzy fusion results.

To improve the performance of EMD-based image fusion methods, we propose a novel and fast EMD-based image fusion method in this paper (Fig. 1). We firstly develop a multi-channel bidimensional EMD method based on morphological filter (we name it as MF-MBEMD) to decompose the input source images into several IMFs with different scales and a residue, which can accelerate the computation

of EMD for multi-channel images significantly. We then adopt a patch-based fusion strategy with overlapping partition instead of the pixel-based fusion way usually used in EMD-based image fusion to fuse the IMFs and residue, where the energy-based rules are designed to extract the salient information of input source images. Finally, we obtain the final result by adding the fused IMFs and the fused residue together. The main contributions of this paper can be summarized as follows:

- We develop a multi-channel bidimensional EMD method based on morphological filter (MF-MBEMD) to conduct image fusion. It uses the morphological expansion and erosion filters to compute the upper and lower envelopes of a multi-channel image, where the size of the morphological filter window is determined by the minimal average extremum distance of all image channels. It significantly improves the computation efficiency of EMD for multi-channel images while maintaining the decomposition quality.
- We adopt a patch-based fusion strategy with overlapping partition to replace the pixel-based fusion way commonly used in EMD-based image fusion, where an energy-based maximum selection rule is designed to fuse the IMFs, and the feature information extracted by IMFs is used

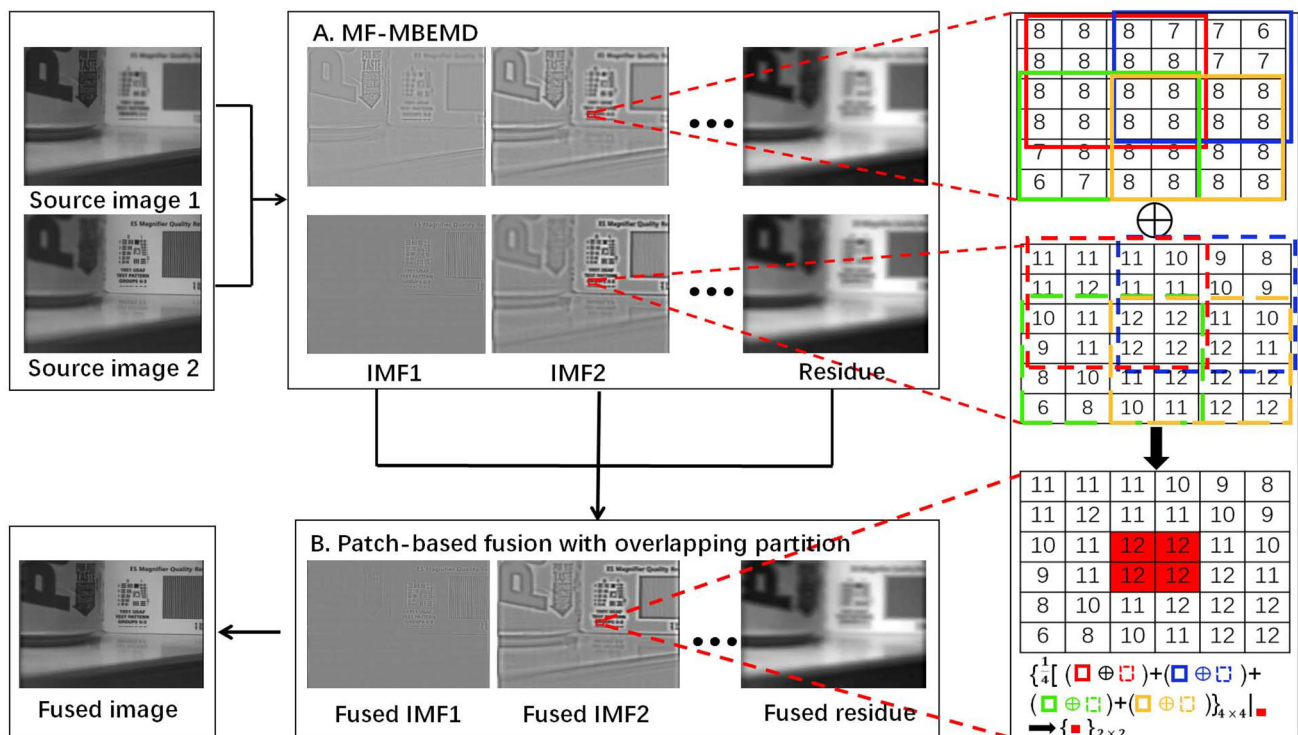


Fig. 1 The pipeline of the proposed novel and fast EMD-based image fusion method for two multi-focus images, where the input source images are decomposed into several IMFs and a residue, the patch

size of the partition is set into 4×4 , and the overlapping number of rows/columns is set into 2 in order to better illustrate our method

as a guide to merge the residue. It can capture the salient information of the source images well and also reduce the spatial artifacts introduced by the noisy characteristics of the pixel-wise maps.

- We evaluate the fusion performance of several state-of-the-art EMD methods for multi-channel images under the popular fusion strategies used in the EMD-based image fusion methods on some commonly used data sets with multi-focus and multi-modal images. The evaluation results demonstrate that our proposed MF-MBEMD obtains better time performance, and our proposed patch and energy-based fusion strategy achieves better visualization and objective metrics.

2 Related work

2.1 Bidimensional empirical mode decomposition

The original 1D EMD [3] employs a sifting processing to extract IMFs with different scales, which needs to generate the upper and lower envelopes of all local extrema points (maxima and minima) by cubic splines. To generalize it to 2D images, Nunes et al. [12] generated envelope surfaces by radial basis function interpolation. Al-Baddai et al. [13] proposed a novel method of envelope surface interpolation based on Green's function to improve the stability of BEMD. The approach using bi-Laplacian operator interpolation to compute the upper and lower envelopes [14–16] is developed to perform the decomposition for signals defined on 3D surfaces and can be applied to 2D images naturally [17,18]. To avoid the computation of the envelope surfaces in the sifting processing, Pan and Yang [8] generated the mean surface by interpolating the centroid points of neighbour extremum points in its Delaunay triangulation, while Qin et al. [6] used a window function as an alternative to compute the mean surface and obtained a window BEMD method. Wang et al. [9] presented a bidimensional ensemble empirical mode decomposition method (BEEMD) to optimize the decomposition by averaging the modes of all noise-added images, which needs very expensive time costs. In order to improve the computation efficiency of the surface interpolation-based BEMD methods, a direct envelope estimation method based on order-statistics filters [19] is proposed to generate the envelope surfaces of an input image, where the distance map between the adjacent maxima/minima is computed to determine the filter size. Trusiak et al. developed an enhanced fast empirical mode decomposition method [20] to further accelerate the computation of envelope surfaces by using morphological operations instead of order-statistics filters. In addition, it generates the filter size only by the extremum number and can avoid the time-consuming computation of the distance map between the adjacent extrema.

The aforementioned BEMD methods are designed to decompose 2D images with one data channel (single-channel images). Although they can also be applied separately to decompose each channel of a bivariate/multivariate 2D signal (two/multiple data channel), they suffer from a lot of problems (e.g. mode alignment and nonuniqueness) as illustrated in [21]. These issues limit the application of EMD in data fusion which requires the same-index IMFs containing the same-scale information. To make the extracted modes from each of the multiple 2D signals match in feature scales, Yeh [5] proposed the complex BEMD to decompose a bivariate (complex) 2D signal, which applies the standard BEMD based on surface interpolation [12] to four real-valued 2D signals to generate the complex-valued IMFs. Rehman et al. [7] reshaped a multi-channel 2D image into a multivariate 1D signal and adopted the 1D multivariate EMD [22] to obtain the multi-scale decomposition. Xia et al. [10] improved such decomposition by using n -dimensional surface projections to estimate the mean surface of a multi-channel 2D image and can reserve its spatial information. Bhuiyan et al. [23] used a direct envelope estimation method based on order-statistics filters [19] to generate the upper and lower envelopes of color images, which is less time-consuming than the surface interpolation-based BEMD methods for multi-channel images. In this paper, we obtain the envelope surfaces of a multi-channel image by using the morphological filters so as to further reduce the decomposition time, which is inspired by the enhanced fast empirical mode decomposition [20] developed for single-channel images. It can not only generate good decomposition results, but also can significantly accelerate the computation.

2.2 Image fusion

With the development of signal processing and analysis, a large number of image fusion methods have been proposed such as transform domain methods, spatial domain methods, deep learning methods, etc. Here we mainly discuss the transform domain methods which are more relevant with EMD-based image fusion. More comprehensive survey can be found in [2,24,25].

The transform domain methods merge the transformed coefficients of the input source images obtained by a transform, and generate the fused image by a reconstruction step with the corresponding inverse transform. Undoubtedly, the selection of transform domain plays an important role in these methods. So far, a lot of transforms have been involved to conduct image fusion, which include Laplacian pyramid [26], multi-scale geometric analysis [27–30], fast Fourier transform [31], wavelet transform [32,33], empirical mode decomposition (EMD) [4,5,7,10], etc.

Different from many classical transforms such as Fourier transform and wavelet transform using predefined basis

functions, EMD is fully adaptive and data-driven. Pan and Yang [8] used the mean approximation-based BEMD to decompose the source images and then employed a standard deviation-based weighted averaging rule to fuse each component pixel by pixel. Wang et al. [9] proposed a pixel-based image fusion method based on BEEMD and an entropy-based weighted averaging rule to conduct the fusion. Qin et al. [6] used the window BEMD to obtain the decomposition and fused all IMFs and the residue by the maximum selection rule based on two saliency measures in a pixel-based manner, which may introduce some artifacts. In order to make the decompositions of different source images match in number and property, the multivariate 1D EMD [22] are used to decompose source images and each component can be merged by a variance-based weighted averaging rule pixel by pixel [7]. Xia et al. [10] adopted the multivariate BEMD based on surface projection to obtain the multi-scale decomposition, which can improve the fusion quality of the multivariate 1D EMD-based fusion method [7]. Yeh [5] proposed a pixel-based multi-focus image fusion approach according to the complex BEMD. Ahmed and Manic [4] used the order-statistics filter-based EMD [19,23] to decompose the source images and perform the fusion by using a variance-based maximum selection rule pixel by pixel. Zhu et al. [11] employed bivariate BEMD and sparse representation to conduct the fusion of infrared-visible images. Although many attempts have been made in the research of EMD-based image fusion, they are limited by the computation efficiency of the used EMD methods and the fusion schemes. In this paper, we propose a novel and fast EMD-based image fusion method based on a fast multi-channel bidimensional EMD method and a patch and energy-based fusion strategy, which obtains better performance than the existing EMD-based image fusion approaches on several commonly used image data sets with multi-focus and multi-modal images.

3 Multi-channel bidimensional EMD based on morphological filter

To our knowledge, the enhanced fast empirical mode decomposition method [20] obtains the best time performance for the decomposition of a greyscale image in the existing BEMD methods. It adopts morphological filters to implement the direct envelope estimation method based on order-statistics filters [19] and employs the mean extremum distance as the filter size to avoid the computation of the distance map between the adjacent maxima/minima. However, it is only designed for single-channel images to conduct fringe pattern processing. In this paper, we present a multi-channel bidimensional EMD method based on morphological filter (MF-MBEMD), where the relevant modifications of the enhanced fast empirical mode decomposition method [20]

are given to make it be suitable for more general images with multiple channels.

The proposed MF-MBEMD employs the morphological filters with the same widow size for each channel to generate the envelope surfaces of a multi-channel image, which can extract similar spatial scale of each channel image during the decomposition. Specifically, given a multi-channel image $I = (I_1, \dots, I_n)$ with the size $W \times H$, its upper (lower) envelope $U = (U_1, \dots, U_n)$ ($D = (D_1, \dots, D_n)$) can be generated by

$$\begin{aligned} U_k(x, y)|_{k=1, \dots, n} &= (I_k \oplus b)(x, y) = \max_{(s, t) \in Z_{xy}} I_k(s, t), \\ D_k(x, y)|_{k=1, \dots, n} &= (I_k \ominus b)(x, y) = \min_{(s, t) \in Z_{xy}} I_k(s, t), \end{aligned} \quad (1)$$

where \oplus denotes the morphological expansion filter, \ominus denotes the morphological corrosion filter, Z_{xy} represents the set of pixels in the window $w \times w$ centered on the pixel (x, y) , and b is a binary set indicator function on Z_{xy} . The average filter is used to get much smoother envelopes by

$$\begin{aligned} U'_k(x, y)|_{k=1, \dots, n} &= \frac{1}{w \times w} \sum_{(s, t) \in Z_{xy}} U_k(s, t), \\ D'_k(x, y)|_{k=1, \dots, n} &= \frac{1}{w \times w} \sum_{(s, t) \in Z_{xy}} D_k(s, t). \end{aligned} \quad (2)$$

To consider the feature extraction of all data channels of the input image, we set the window size w in Eq. (1) and Eq. (2) into the following minimal average extremum distance of all image channels,

$$w = \min\{w_1, \dots, w_n\}, \quad (3)$$

where $w_k (k = 1, \dots, n)$ denotes the average extremum distance of the k -th channel image I_k and is computed by

$$w_k = \sqrt{\frac{W \times H}{N_k}}, \quad (4)$$

where N_k is the average value of the numbers of all local maxima and minima of I_k . We find all local maxima (minima) of I_k by comparing the values of each pixel and its neighbours in the 3×3 window centered on it in each iteration. It is different from the extraction of local extrema in the enhanced fast empirical mode decomposition method [20], where the size of extremum window is equal to the average extremum distance of the previous iteration and the number of extracted extrema is reduced. In contrast, our strategy obtains more extrema in each iteration and can extract much-finer feature scale of each channel image.

The proposed MF-MBEMD can extract the IMFs with different scales iteratively by a sifting processing based on the above envelope computation method until the specified

Algorithm 1 Multi-channel bidimensional EMD via morphological filter

Input: a multi-channel image $I = (I_1, \dots, I_n)$

Initialization: set the initial residue $R_0 = I$, $i = 1$;

1: **repeat**

2: find all local maxima (minima) of each channel image of R_{i-1} by comparing the values of each pixel and its neighbours in the 3×3 window centered on it;

3: compute the average extremum distance w_k^{i-1} ($k = 1, \dots, n$) for each channel image of R_{i-1} by Eq. (4), and get the window size w_{i-1} of morphological filter by Eq. (3);

4: obtain the upper (lower) envelope U_{i-1} (D_{i-1}) of R_{i-1} by Eq. (1), and smooth them by Eq. (2);

5: obtain the i -th residue $R_i = \frac{1}{2}(U_{i-1} + D_{i-1})$ and the i -th IMF $F_i = R_{i-1} - R_i$;

6: $i = i + 1$;

7: **until** the residue R_i is a constant or a monotonic function, or the number of IMFs is more than a given threshold

Output: IMFs F_i , $i = 1, \dots, K$ and the corresponding residue R_K .

number of IMFs is achieved or the residue is a constant or a monotonic function (see Algorithm 1). Figures 1 and 2 give the decomposition results of some multi-focus images using MF-MBEMD. As can be seen that the leading IMFs extract much finer-scale features and the trailing IMFs describe much coarser-scale features. Furthermore, the extracted feature scales of two multi-focus images are matched very well for the same-index IMFs, which is very important for the application of EMD in image fusion.

Table 1 gives the time performance comparison in generating three IMFs for the image in Fig. 2a with different sizes between our proposed MF-MBEMD and several state-

Table 1 Time performance comparison (in seconds) between our proposed MF-MBEMD and several state-of-the-art multi-channel bidimensional EMD methods in generating three IMFs for the image in Fig. 2a with different sizes

Size	SP-MBEMD [10]	BL-MEMD [16]	OSF-CBEMD [23]	MF-MBEMD Ours
128×128	102.32	15.73	0.11	0.05
256×256	704.64	83.22	0.42	0.17
512×512	9488.67	516.15	3.84	0.60
1024×1024	> 72000	3695.53	38.51	2.40

of-the-art multi-channel bidimensional EMD methods. The compared methods include the multivariate BEMD based on surface projection [10], the multi-channel extension of the EMD based on bi-Laplacian operator interpolation [14] [16], and the color bidimensional EMD method based on order-statistics filters [23], which will be called SP-MBEMD, BL-MEMD, and OSF-CBEMD in this paper, respectively. Benefited from the aforementioned direct envelope estimation method via morphological filters with the mean extremum distance-based filter size, MF-MBEMD obtains the best time performance among the compared EMD methods for multi-channel images as shown in Table 1. Furthermore, as the size of image increasing, the time of MF-MBEMD increases slowly, while the time of the other methods increases very fast. The worst case is SP-MBEMD [10] cannot obtain the decomposition for the image with the size 1024×1024 in twenty hours. Therefore MF-MBEMD

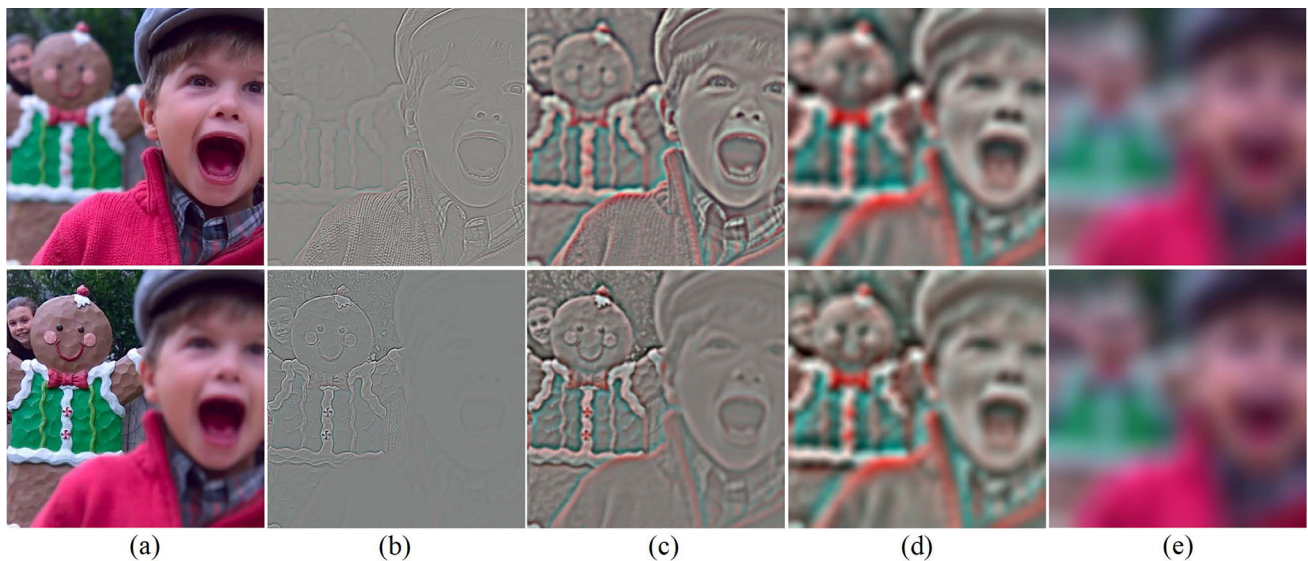


Fig. 2 Decompositions of two color multi-focus images using MF-MBEMD. **a** Source images. **b** IMF 1. **c** IMF 2. **d** IMF 3. **e** Residue

can be used to improve the efficiency of EMD-based image fusion algorithms significantly.

4 Proposed EMD-based image fusion

Firstly, our EMD-based image fusion method decomposes the input source images into several IMFs and a residue by the aforementioned MF-MBEMD. Secondly, it fuses the IMFs and the residue separately by a patch-based fusion strategy with overlapping partition. An energy-based maximum selection rule is developed for the fusion of IMFs, and two different rules based on the feature information extracted by IMFs are devised for the fusion of the residue according to the image types (multi-focus image or multi-modal images) in order to get high-quality fusion results. Finally, the fused image are reconstructed by the fused IMFs and the fused residue. The whole pipeline of our method is shown in Fig. 1.

4.1 Multi-scale decomposition by MF-MBEMD and overlapping patch partition

Given two source images I_1 and I_2 , we combine them into a two-channel image $I = (I_1, I_2)$, which can be decomposed into K IMFs with the scale from fine to coarse and a residue by MF-MBEMD,

$$I = \sum_{i=1}^K F_i + R_K, \quad (5)$$

where $F_i = (F_{i1}, F_{i2}) (i = 1, \dots, K)$ is the i -th IMF and $R_K = (R_{K1}, R_{K2})$ is the corresponding residue.

We divide all IMFs and the residue into many patches of size $M \times M$ with N overlapping rows/columns. Figure 3 gives the illustration of this overlapping patch partition scheme for a 3×6 image, where $M = 3$ and N is set into different values. If some pixels are not contained into the partition as shown in Fig. 3c, we need to add some patches to cover such pixels (Fig. 4b). This overlapping patch strategy is designed to reduce the artifacts which may be produced around the partition boundary in the patch-based fusion methods.

4.2 Patch-based fusion for IMFs and residue

Fusion of IMFs. As for the j -th patch $F_i^j = (F_{i1}^j, F_{i2}^j)$ of the i -th IMF F_i , we adopt an energy-based maximum selection rule to obtain the fused patch G_i^j by comparing the energy of two corresponding patches as follows,

$$G_i^j = \begin{cases} F_{i1}^j, & E(F_{i1}^j) \geq E(F_{i2}^j) \\ F_{i2}^j, & E(F_{i1}^j) < E(F_{i2}^j) \end{cases}. \quad (6)$$

The energy of each patch in Eq. (6) is computed by

$$E(F_{ip}^j) = \sum_{(s,t) \in Z_j} F_{ip}^j(s,t)^2, \quad p = 1, 2, \quad (7)$$

where Z_j represents the set of pixels in the j -th patch. This rule is used to capture the salient information of the input source images.

Fusion of residue. As for the j -th patch $R_K^j = (R_{K1}^j, R_{K2}^j)$ of the residue R_K , we design two different fusion rules based on the feature information extracted by IMFs to obtain the fusion residue patch H_K^j according to the image types (multi-focus images or multi-modal images).

The first one is an energy-based maximum selection rule which is used to fuse multi-focus images. It compares the energy of two corresponding patches in the first IMF as

$$H_K^j = \begin{cases} R_{K1}^j, & E(F_{11}^j) \geq E(F_{12}^j) \\ R_{K2}^j, & E(F_{11}^j) < E(F_{12}^j) \end{cases}, \quad (8)$$

where $E(F_{1p}^j) (p = 1, 2)$ denotes the energy of the first IMF and can capture the finest-scale features. This fusion scheme can describe the focused region of multi-focus images effectively.

The second one is an energy-based weighted averaging rule which is used to merge multi-modal images. It uses the mean of the residue patch as a guide to fuse the brightness of each modal image, and adopts the feature information extracted by IMFs as a guide to merge the mean-separated part of the residue. The fusion formula is given by

$$\begin{aligned} H_K^j &= \sum_{p=1}^2 a_p^j (R_{Kp}^j - \mu_{Kp}^j) + \sum_{p=1}^2 b_p^j \mu_{Kp}^j, \\ a_p^j &= \left| \sum_{i=1}^K E(F_{ip}^j) \right|^l / \left(\left| \sum_{i=1}^K E(F_{i1}^j) \right|^l + \left| \sum_{i=1}^K E(F_{i2}^j) \right|^l \right), \\ b_p^j &= \left| \mu_{Kp}^j \right|^m / \left(\left| \mu_{K1}^j \right|^m + \left| \mu_{K2}^j \right|^m \right), \end{aligned} \quad (9)$$

where μ_{Kp}^j is the mean of the j -th patch of the residue R_{Kp}^j , $p = 1, 2, l$ and m are two nonnegative exponent parameters to control the feature guide intensity and the brightness fusion intensity, respectively. If we set $l = 0$ and $m = 0$, it becomes a simple averaging of the residue of two multi-modal images. If we set $l > 0$ and $m > 0$, it adopts the feature information extracted by IMFs as a guide to merge the mean-separated part of the residue of each modal image, and adopts the mean of the residue patch as a guide to fuse the brightness of each modal image. A larger value of l means much stronger features are transmitted to the fusion result, and a bigger value of m means more bright targets are involved into the merged result. In our experiments, both l and m are set into 6 and can generate good results.

4.3 Multi-scale image reconstruction

Once all patches of the IMFs and residue are fused, we first get the value at each pixel (x, y) of the fused IMFs and residue by averaging the values of the pixel (x, y) in all overlapping patches by

$$\begin{aligned} G'_i(x, y) &= \frac{1}{S(x, y)} \sum_j G_i^j(x, y), \\ H'_K(x, y) &= \frac{1}{S(x, y)} \sum_j H_K^j(x, y), \end{aligned} \quad (10)$$

where $S(x, y)$ denotes the overlapping patch number at the pixel (x, y) , and then we add the fused IMFs and the fused residue together to obtain the final fused image I' by

$$I'(x, y) = \sum_{i=1}^K G'_i(x, y) + H'_K(x, y). \quad (11)$$

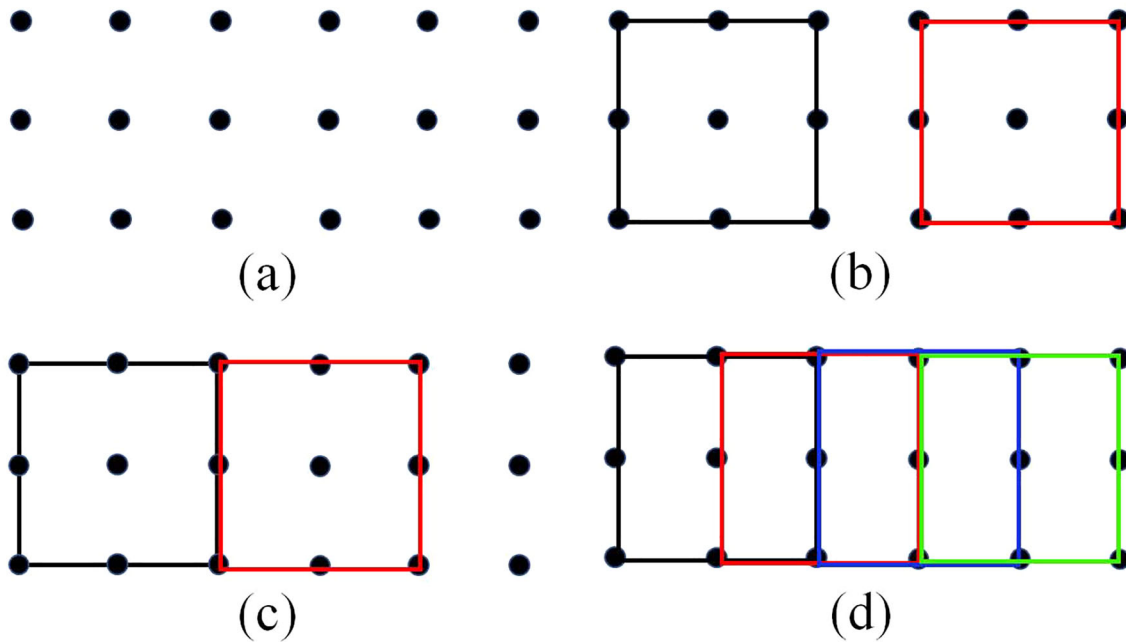


Fig. 3 Illustration of overlapping patch partitions for a 3×6 image (a), where the patch size M is set into 3 and the overlapping row/column number N is set into 0 in (b), 1 in (c), and 2 in (d), respectively

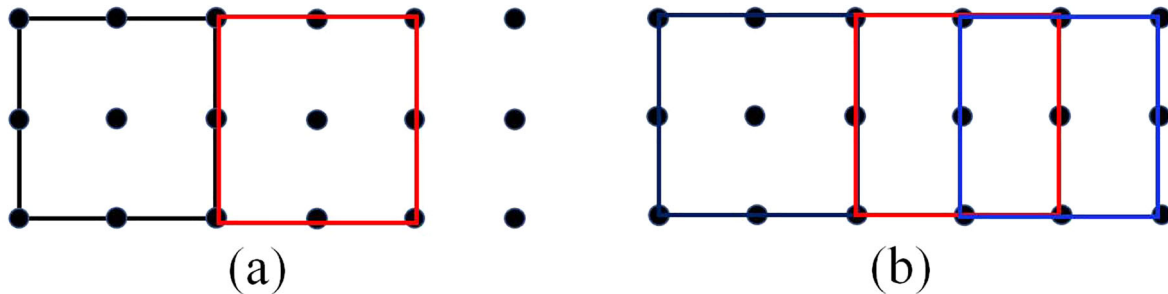


Fig. 4 Processing of the boundary pixels not contained in the partition. a The partition in Fig. 3c. b The partition with an additional blue patch covering these boundary pixels

4.4 Implementation

We implement our newly proposed EMD-based image fusion method effectively by accumulating the fused value at each pixel patch by patch (Algorithm 2). Specifically, we first set the initial value at each pixel (x, y) of the fused IMFs, the fused residue, and the overlapping patch number by $G'_i(x, y)|_{i=1, \dots, K} = 0$, $H'_K(x, y) = 0$, and $S(x, y) = 0$. After each patch is merged, we update the fused values by

$$G'_i(x, y) \rightarrow G'_i(x, y) + G_i^j(x, y), \quad (12)$$

and

$$H'_K(x, y) \rightarrow H'_K(x, y) + H_K^j(x, y), \quad (13)$$

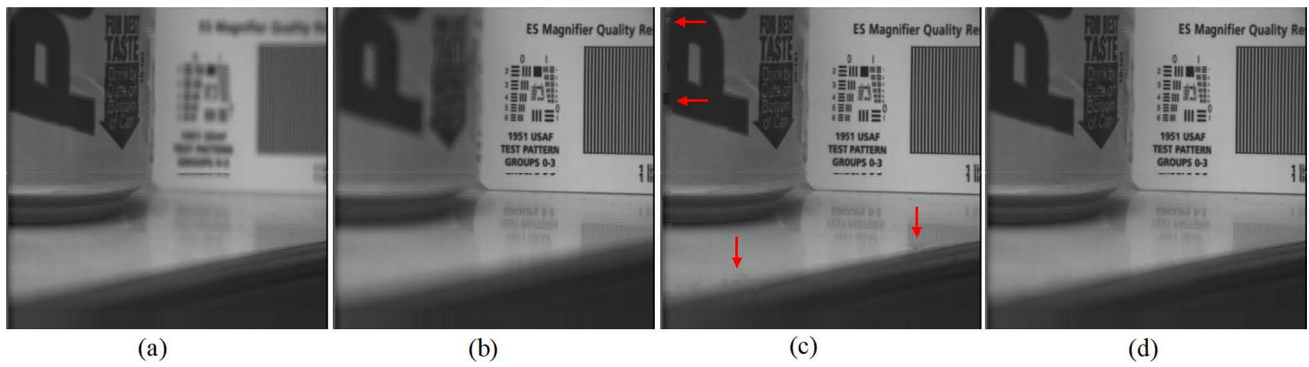


Fig. 5 Fusion results of two greyscale multi-focus images generated by the pixel-based strategy and our patch-based strategy. **a** Source image 1. **b** Source image 2. **c** The pixel-based strategy based on our energy-based maximum selection rule. **d** Our fusion strategy based on overlapping patch division

and update the overlapping patch number at each pixel of the j -th patch by

$$S(x, y) \rightarrow S(x, y) + 1. \quad (14)$$

Once all patches of K IMFs and the residue are fused, we get the final fused IMFs and residue at each pixel by

$$\begin{aligned} G'_i(x, y) &\rightarrow \frac{1}{S(x, y)} G'_i(x, y), \\ H'_K(x, y) &\rightarrow \frac{1}{S(x, y)} H'_K(x, y). \end{aligned} \quad (15)$$

Our fusion strategy based on overlapping patch partition can reduce spatial artifacts introduced by the noisy characteristics of the pixel-wise maps in the pixel-based fusion methods. Figure 5 gives a comparison of these two fusion strategies based on our MF-MBEMD, where the pixel-based scheme obtains the fused IMFs and the fused residue pixel by pixel by comparing the IMFs' energy of two patches centered at each pixel according to Eqs. (6) and (8), respectively. It can be seen that some spatial artifacts have been produced by this pixel-based fusion scheme in Fig. 5c, while they cannot be observed in our fusion strategy based on overlapping patch division (Fig. 5d).

5 Experiments and discussions

In this section, we first introduce the fusion image data sets used in our experiments and give the used objective metrics for evaluating the algorithm. Then, we discuss how to select the main parameters of our method. Afterwards, we make various comparison to further illustrate the effectiveness of the proposed algorithm. All experimental results shown in this paper are carried out by MATLAB 2019 on a Laptop with Inter Core (TM) i7 CPU and 16.0 GB RAM.

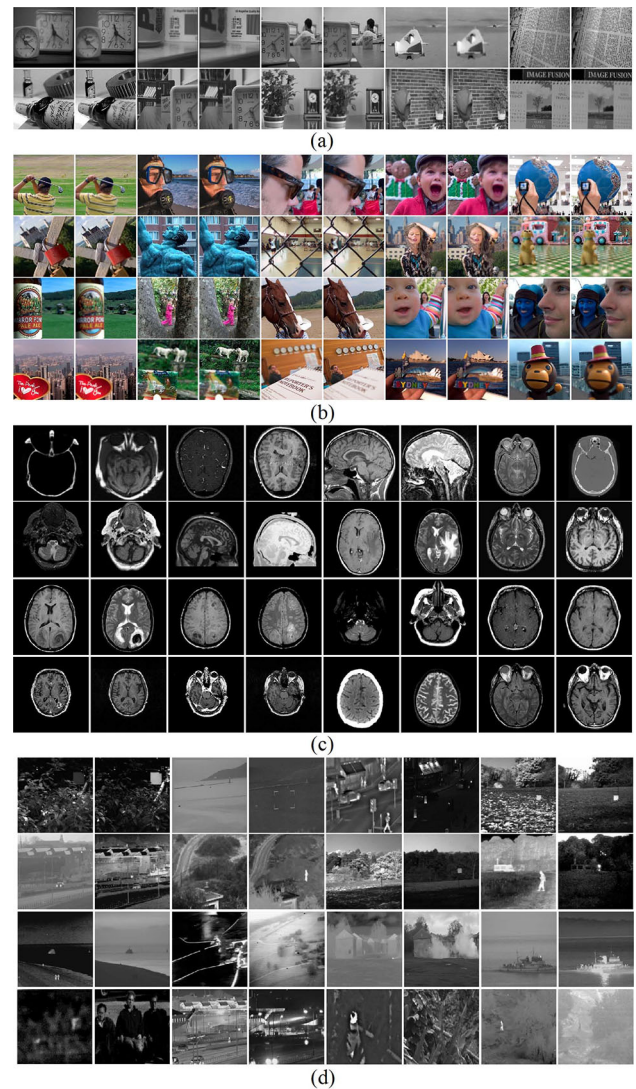


Fig. 6 Four data sets in our experiments. **a** Greyscale multi-focus set. **b** Color multi-focus set. **c** Medical multi-modal set. **d** Infrared-visible multi-modal set

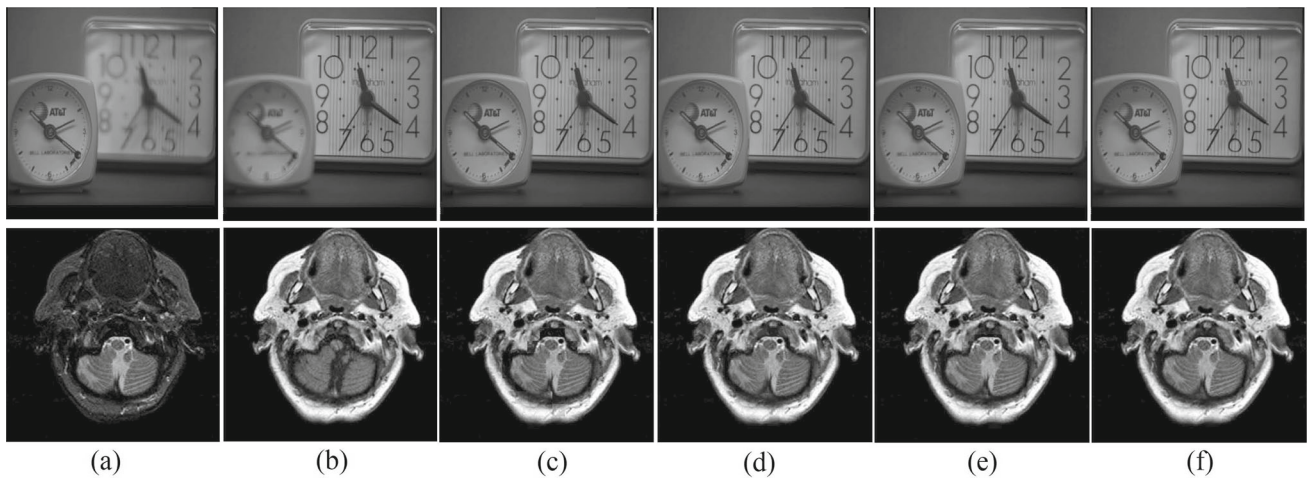


Fig. 7 Fusion results with different overlapping number N of rows/columns. **a** Source image 1. **b** Source image 1. **c** $N = 0$. **d** $N = M/6$. **e** $N = M/2$. **f** $N = M - 2$. M is set into 31 in the first row and is set into 15 in the second row

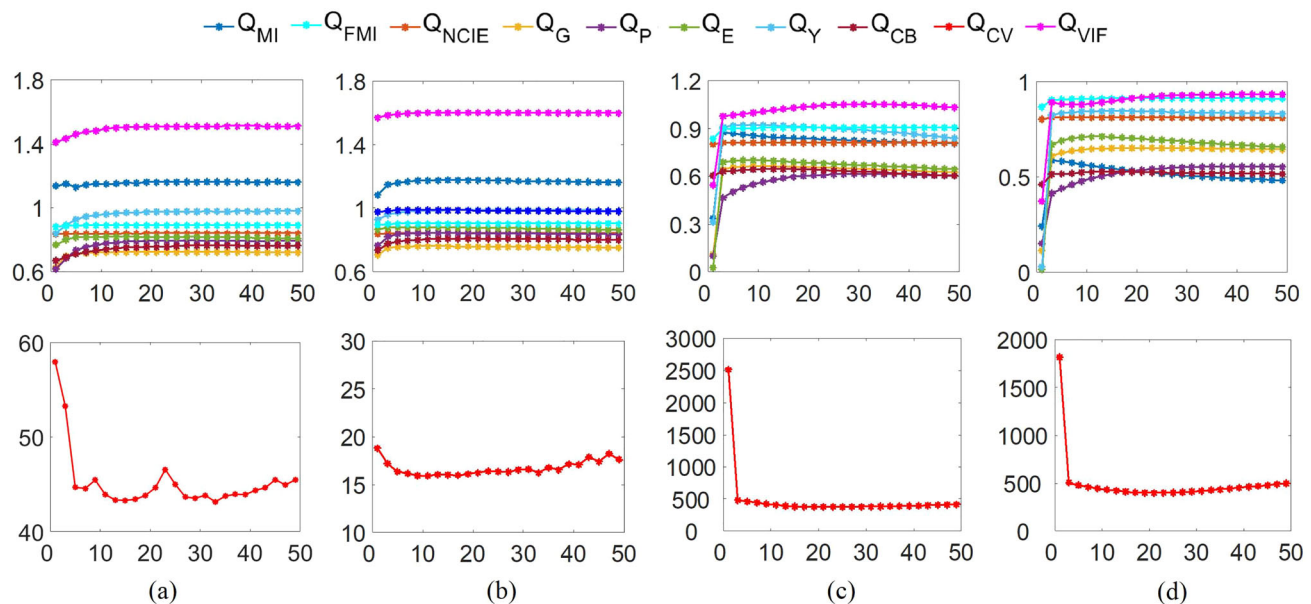


Fig. 8 The average values of ten fusion metrics for four data sets with the patch size varying from 1 to 50. **a** Greyscale multi-focus set. **b** Color multi-focus set. **c** Medical multi-modal set. **d** Infrared-visible multi-modal set

5.1 Data sets and objective metrics

Many widely-used multi-focus images¹ and multi-modal images^{2,3} in the literature are selected to assess the performance of image fusion (Fig. 6). The multi-focus images are divided into two data sets including a greyscale set with 10 pairs of greyscale images and a color set with 20 pairs of color images, which are used to evaluate the performance of the state-of-the-art methods in a recent survey of multi-

focus image fusion [25]. The multi-modal images are also divided into two data sets including a medical image set with 16 pairs of medical images and an infrared-visible image set with 16 pairs of infrared-visible images, which almost cover all tested multi-modal images in [26,30,34].

In order to measure the fusion quality of our method objectively, ten widely used fusion metrics are employed including mutual information Q_{MI} [25], feature mutual information Q_{FMI} [35], the nonlinear correlation information entropy Q_{NCIE} [25], the gradient-based metric Q_G [36], the phase congruency-based metric Q_P [25], the structural similarity-based metric Q_E [25] and Q_Y [37], the human perception

¹ <https://github.com/yuliu316316/MFIF>.

² <https://github.com/uzeful/IFCNN>.

³ <https://github.com/arbabsufyan/Proposed-Code.git>.

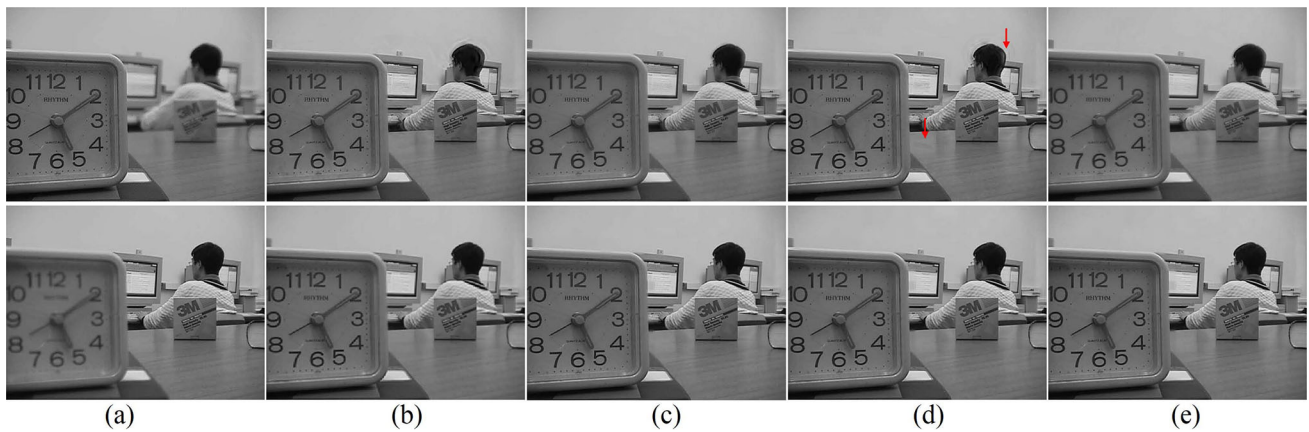


Fig. 9 Comparison of EMD-based image fusion methods on two multi-focus images (a). The top results in (b), (c), (d) and (e) are generated in a pixel-based manner using SP-MBEMD [10] with the fusion rule in [10], BL-MEMD [16] with the fusion rule in [8], OSF-CBEMD [23]

with the fusion rule in [4], and the proposed MF-MBEMD with the fusion rule in [9], respectively. The bottom results in (b), (c), (d) and (e) are obtained in a patch-based manner by the corresponding EMD methods in the top row and our fusion strategy

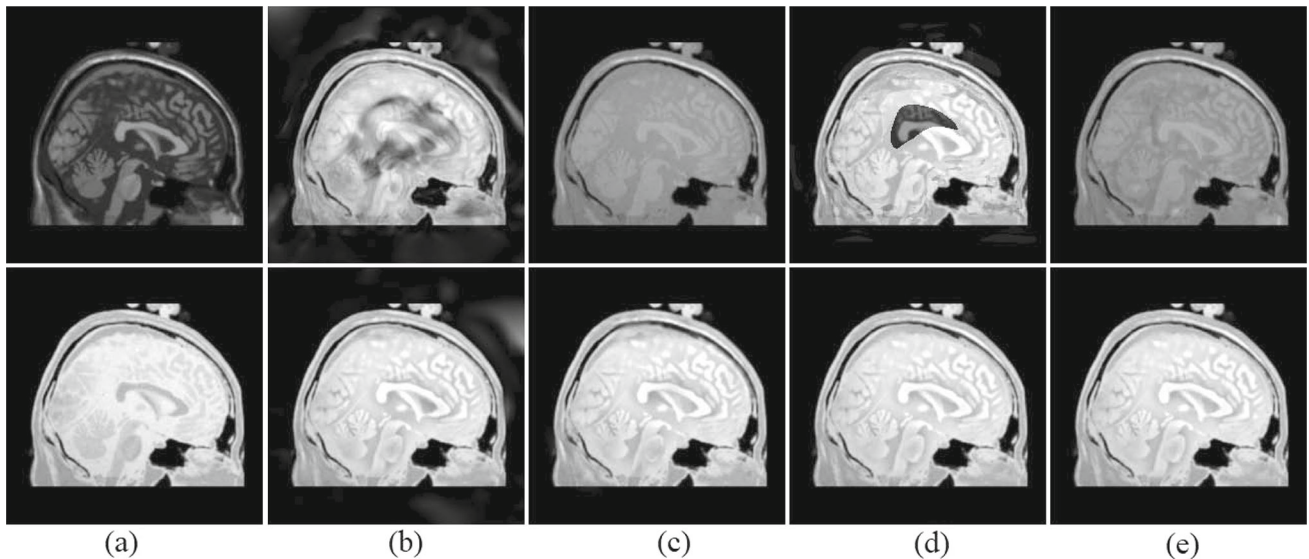


Fig. 10 Comparison of EMD-based image fusion methods on two multi-focus images (a). The top results in (b), (c), (d) and (e) are generated in a pixel-based manner using SP-MBEMD [10] with the fusion rule in [10], BL-MEMD [16] with the fusion rule in [8], OSF-CBEMD

[23] with the fusion rule in [4], and the proposed MF-MBEMD with the fusion rule in [9], respectively. The bottom results in (b), (c), (d) and (e) are obtained in a patch-based manner by the corresponding EMD methods in the top row and our fusion strategy

inspired metric Q_{CB} and Q_{CV} [25], and visual information fidelity Q_{VIF} [38]. Larger values represent better fusion qualities for Q_{MI} , Q_{FMI} , Q_{NCIE} , Q_G , Q_P , Q_E , Q_Y , Q_{CB} and Q_{VIF} , while smaller values represent better fusion qualities for Q_{CV} .

5.2 Selection of main parameters

Decomposition level K of MF-MBEMD. We use the energy-based maximum selection rule to fuse all IMFs in order to extract the salient information of the source images.

The decomposition level K of MF-MBEMD is set into 1 for multi-focus images since the first IMF can effectively capture the focused region of the multi-focus images, while K is set into 2 for multi-modal images considering that the salient information of the source images concentrates on the leading two IMFs of MF-MBEMD in the tested multi-modal image sets.

Overlapping number N of rows/columns. Usually, a much larger value of the overlapping number N of rows/columns in the patch division can reduce the artifacts of our method as shown in Fig. 7, but it increases the com-

Algorithm 2 EMD-based image fusion via morphological filter

Input: two source images I_1 and I_2 , decomposition level K , patch size M , overlapping row/column number N , and exponent parameters l and m in the fusion of multi-modal images;

Initialization: set the initial values at each pixel (x, y) by $G'_i(x, y)|_{i=1,\dots,K} = 0$, $H'_K(x, y) = 0$, and $S(x, y) = 0$;

1: decompose $I = (I_1, I_2)$ into K IMFs $F_i, i = 1, \dots, K$ and a residue R_K by MF-MBEMD (Algorithm 1);

2: **for** $i = 1$ to K **do**

3: compute the energy of all patches of F_k by Eq. (7);

4: **end for**

5: **repeat**

6: **for** $i = 1$ to K **do**

7: merge the j -th patch of the i -th IMF by Eq. (6), and update the fused values by Eq. (12);

8: **end for**

9: merge the j -th patch of the residue by Eq. (8) for multi-focus images or by Eq. (9) for multi-modal images, and update the fused values by Eq. (13);

10: update the overlapping patch number by Eq. (14);

11: **until** all patches of K IMFs and the residue are fused

12: get the fused IMFs and the fused residue by Eq. (15);

13: add the fused IMFs and the fused residue together to obtain the final fused image by Eq. (11);

Output: fused image I' .

putation cost at the same time. In our experiments, N is set into $M/6$ for multi-focus data sets, and $M - 2$ for multi-modal data sets, where M is the block size of the division. A larger number of experiments demonstrate such selections can produce satisfactory fusion results.

Block size M of the division. In order to select a proper M , we conduct the fusion of each data set with M varying from 1 to 50 to evaluate the fusion performance by the average values of ten fusion metrics (Fig. 8), where the overlapping number N of rows/columns is set into $M/6$ for multi-focus data sets, and $M - 2$ for multi-modal data sets. In our experiments, we have obtained good results by selecting $M = 31$ for the grayscale multi-focus set, $M = 15$ for the color multi-focus set, $M = 15$ for the medical multi-modal set, and $M = 33$ for the infrared-visible multi-modal set.

5.3 Comparisons

Comparison with EMD-based image fusion methods.

There are two factors to determine the fusion quality of EMD-based image fusion methods which include the used EMD method and the used fusion strategy. In order to showcase the advantage of our newly proposed EMD-based image fusion method, we compare the fusion results generated by SP-MBEMD [10], BL-MEMD [16], OSF-CBEMD [23], and the proposed MF-MBEMD with the popular fusion strategies in a pixel-based manner [4] [8] [9] [10] and a patch-based manner in this paper (Fig. 9, Fig. 10 and Table 2). The weighted averaging rule based on variance [10], standard

deviation [8] and entropy [9] in a pixel-wise manner cannot capture the salient information of input source images well and produce fuzzy fusion results as shown in the top of Fig. 9b, c, e. In contrast, the variance-based maximum selection rule [4] in a pixel-wise manner can obtain much clearer results, but it also generate some artifacts, as shown in the top of Fig. 9d and Fig. 10d. Compared with the pixel-based fusion strategies in the existing EMD-based image fusion methods [4] [8] [9] [10], the fusion strategy based on overlapping patch division proposed in this paper can improve the fusion quality of each EMD method in visualization and objective metrics (Fig. 9, Fig. 10 and Table 2), since the energy-based maximum selection rule for the fusion of each IMF can capture more salient information, and the patched-based strategy reduces the artifacts introduced by the pixel-wise fusion manner at the same time. Furthermore, as for the fusion of the residue, the activity level generated by the first IMF can more effectively capture the focused region of the multi-focus images, and the energy-based weighted averaging rule generated by the extracted IMFs can capture more structure information of the multi-modal images. The visual results of four BEMD methods under our fusion strategy are close according to the observation in Figs. 9 and 10. When considering all fusion objective metrics together (Table 2), MF-MBEMD and OSF-CBEMD [23] generate better results than SP-MBEMD [10] and BL-MEMD [16]. In terms of time performance, SP-MBEMD [10] is extremely time-consuming, while MF-MBEMD obtains the fastest speed among the four EMD methods as reported in Table 2.

Comparison with non-EMD-based image fusion methods. We also compare our method with many state-of-the-art non-EMD-based methods including six transform domain methods (CVT [29], DTCWT [33], NSCT [28], NSCT-SR [30], NSCT-PCLE [41], IMA [26]), three spatial domain methods (BRW [39], MISF [42], MSID [43]) and two deep learning method (MADCNN [40], IFCNN [34]). The source code of these methods are publicly available online. Among them, CVT [29], DTCWT [33], NSCT-SR [30] and IFCNN [34] can be used for various types of images. NSCT [28], BRW [39] and MADCNN [40] are designed for multi-focus images. NSCT-PCLE [41], MISF [42] IMA [26] and MSID [43]) are developed for multi-modal images. Therefore, we conduct our comparison experiments to take care of the preference of each method in the image types. Fig. 11 shows the fusion results of the compared methods for two color multi-focus images. As can be found that CVT [29], DTCWT [33], NSCT [28], NSCT-SR [30] and BRW [39] suffer from the visible ghosting artifact around the hat, while MADCNN [40], IFCNN [34] and our method can reduce this effect. Figure 12 gives the fusion results of the compared methods for an infrared image and a visible image. It can be seen that NSCT-SR [30], NSCT-PCLE [41] and MISF [42] produce some obvious artifacts. CVT [29] and DTCWT [33] cannot

Table 2 Comparison of EMD-based image fusion methods on the objective metrics and the average running time for the data sets in Fig. 6

Data set	EMD method	Fusion strategy	Q_{MI}	Q_{FMI}	Q_{NCIE}	Q_G	Q_P	Q_E	Q_Y	Q_{CB}	Q_{CV}	Q_{VIF}	Time (s)
Greyscale multi-focus set	SP-MBEMD [10]	[10]	0.817/0.218	0.885/0.049	0.822/0.008	0.660/0.063	0.691/0.120	0.809/0.087	0.874/0.044	0.638/0.058	54.649/34.039	1.358/0.285	4315.238
	BL-MEMD [16]	[8]	0.854/0.232	0.882/0.049	0.823/0.008	0.659/0.079	0.725/0.108	0.775/0.101	0.886/0.032	0.638/0.064	57.564/35.663	1.396/0.260	51.324
	OSF-CBEMD [23]	[4]	0.827/0.215	0.887/0.050	0.822/0.008	0.682/0.028	0.710/0.097	0.808/0.072	0.895/0.030	0.672/0.042	50.421/33.097	1.382/0.243	14.074
	MF-MBEMD	[9]	0.853/0.231	0.878/0.052	0.823/0.008	0.607/0.109	0.700/0.142	0.680/0.129	0.851/0.055	0.623/0.079	82.197/62.520	1.381/0.289	45.989
	SP-MBEMD [10]	Ours	1.166/0.131	0.891/0.050	0.840/0.007	0.724/0.028	0.795/0.066	0.816/0.068	0.979/0.011	0.765/0.037	44.176/31.669	1.510/0.155	4426.325
	BL-MEMD [16]	Ours	1.151/0.126	0.891/0.050	0.838/0.006	0.719/0.031	0.789/0.061	0.804/0.068	0.976/0.011	0.760/0.037	48.497/35.357	1.501/0.156	46.578
	OSF-CBEMD [23]	Ours	1.161/0.130	0.891/0.050	0.840/0.007	0.724/0.028	0.793/0.067	0.815/0.067	0.978/0.012	0.763/0.037	43.750/31.080	1.509/0.156	2.607
	MF-MBEMD	Ours	1.163/0.130	0.891/0.050	0.840/0.007	0.724/0.028	0.795/0.066	0.816/0.067	0.978/0.013	0.764/0.036	43.840/31.138	1.510/0.155	0.146
	SP-MBEMD [10]	[10]	0.894/0.146	0.879/0.029	0.828/0.008	0.739/0.050	0.812/0.051	0.878/0.066	0.946/0.054	0.712/0.044	24.390/22.241	1.509/0.173	12783.256
	BL-MEMD [16]	[8]	0.878/0.152	0.897/0.028	0.827/0.009	0.715/0.035	0.794/0.040	0.835/0.039	0.926/0.017	0.663/0.058	45.507/32.789	1.507/0.188	167.137
Color multi-focus set	OSF-CBEMD [23]	[4]	0.918/0.149	0.900/0.028	0.829/0.009	0.737/0.033	0.818/0.045	0.876/0.031	0.951/0.021	0.733/0.041	23.392/17.820	1.539/0.155	43.414
	MF-MBEMD	[9]	0.872/0.154	0.896/0.027	0.826/0.009	0.663/0.045	0.784/0.037	0.737/0.058	0.888/0.022	0.644/0.070	75.139/64.678	1.502/0.201	163.530
	SP-MBEMD [10]	Ours	1.178/0.148	0.879/0.029	0.845/0.008	0.760/0.048	0.843/0.050	0.878/0.060	0.986/0.043	0.806/0.038	16.240/13.394	1.581/0.142	12779.124
	BL-MEMD [16]	Ours	1.144/0.103	0.901/0.028	0.842/0.007	0.753/0.032	0.832/0.043	0.867/0.041	0.982/0.008	0.796/0.043	23.503/16.113	1.586/0.136	153.293
	OSF-CBEMD [23]	Ours	1.176/0.103	0.902/0.028	0.845/0.008	0.759/0.030	0.843/0.042	0.878/0.031	0.986/0.006	0.805/0.04	16.195/14.609	1.601/0.135	6.197
	MF-MBEMD	Ours	1.177/0.103	0.902/0.028	0.845/0.008	0.759/0.030	0.843/0.042	0.878/0.031	0.986/0.006	0.806/0.040	16.016/14.589	1.601/0.135	0.651
	SP-MBEMD [10]	[10]	0.524/0.133	0.885/0.028	0.807/0.002	0.558/0.072	0.337/0.143	0.610/0.136	0.584/0.125	0.476/0.062	694.808/608.431	0.696/0.219	993.661
	BL-MEMD [16]	[8]	0.791/0.150	0.889/0.027	0.809/0.001	0.530/0.093	0.436/0.134	0.563/0.134	0.828/0.052	0.598/0.105	624.537/414.262	0.848/0.233	15.195
	OSF-CBEMD [23]	[4]	0.722/0.117	0.887/0.029	0.809/0.002	0.609/0.053	0.370/0.141	0.649/0.120	0.727/0.103	0.578/0.066	534.803/404.559	0.794/0.205	7.441
	MF-MBEMD	[9]	0.784/0.128	0.886/0.029	0.810/0.002	0.527/0.098	0.428/0.127	0.546/0.127	0.833/0.063	0.592/0.110	728.266/418.418	0.845/0.243	23.535
Infrared-visible multi-modal set	SP-MBEMD [10]	Ours	0.712/0.128	0.902/0.023	0.810/0.002	0.644/0.063	0.541/0.160	0.682/0.111	0.758/0.118	0.538/0.068	453.104/413.590	0.969/0.248	976.413
	BL-MEMD [16]	Ours	0.807/0.118	0.900/0.023	0.811/0.003	0.642/0.065	0.525/0.146	0.664/0.131	0.899/0.051	0.637/0.114	593.092/714.721	0.972/0.239	12.404
	OSF-CBEMD [23]	Ours	0.843/0.122	0.905/0.023	0.811/0.003	0.663/0.061	0.590/0.142	0.695/0.105	0.919/0.039	0.646/0.116	380.691/302.896	1.023/0.258	3.863
	MF-MBEMD	Ours	0.842/0.124	0.905/0.023	0.811/0.003	0.662/0.062	0.589/0.143	0.695/0.105	0.918/0.040	0.646/0.115	383.739/307.492	1.024/0.257	0.658
	SP-MBEMD [10]	[10]	0.301/0.104	0.890/0.033	0.805/0.002	0.521/0.096	0.287/0.185	0.641/0.132	0.698/0.103	0.489/0.083	707.960/808.103	0.642/0.170	1444.208
	BL-MEMD [16]	[8]	0.342/0.137	0.903/0.035	0.806/0.003	0.546/0.111	0.351/0.216	0.655/0.141	0.756/0.106	0.522/0.101	813.711/1213.655	0.713/0.204	71.513
	OSF-CBEMD [23]	[4]	0.529/0.157	0.869/0.034	0.810/0.005	0.593/0.076	0.290/0.163	0.619/0.117	0.725/0.076	0.495/0.082	844.478/728.467	0.728/0.178	8.667
	MF-MBEMD	[9]	0.367/0.159	0.894/0.036	0.806/0.004	0.471/0.100	0.327/0.223	0.515/0.117	0.695/0.111	0.502/0.101	919.556/1447.256	0.663/0.222	49.861
	SP-MBEMD [10]	Ours	0.497/0.195	0.911/0.034	0.810/0.009	0.641/0.081	0.542/0.170	0.673/0.142	0.834/0.105	0.515/0.155	555.554/817.634	0.918/0.216	1328.412
	BL-MEMD [16]	Ours	0.502/0.197	0.911/0.034	0.811/0.009	0.640/0.083	0.535/0.171	0.673/0.138	0.836/0.106	0.517/0.148	504.806/669.505	0.931/0.213	69.111
Medical multi-modal set	OSF-CBEMD [23]	Ours	0.505/0.197	0.912/0.034	0.811/0.009	0.651/0.077	0.560/0.166	0.679/0.136	0.838/0.105	0.519/0.152	445.301/581.433	0.927/0.217	3.151
	MF-MBEMD	Ours	0.502/0.196	0.911/0.035	0.810/0.009	0.649/0.077	0.553/0.164	0.679/0.138	0.837/0.105	0.519/0.153	427.899/566.891	0.931/0.218	1.662

Each metric is reported by the mean and standard deviation (mean/std) over all the samples in each data set. For each item, the bold is the best result in every data set

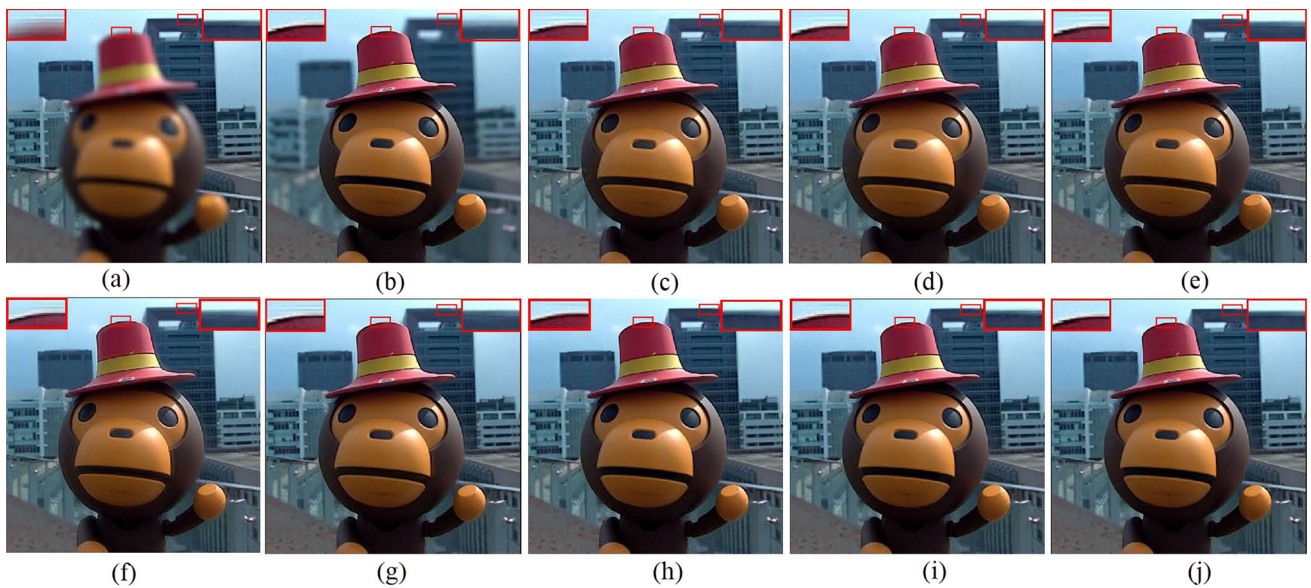


Fig. 11 Comparison with non-EMD-based methods for the fusion of two color multi-focus images. **a** Source image 1. **b** Source image 2. **c** CVT [29]. **d** DTCWT [33]. **e** NSCT [28]. **f** NSCT-SR [30]. **g** BRW [39]. **h** MADCNN [40]. **i** IFCNN [34]. **j** Ours

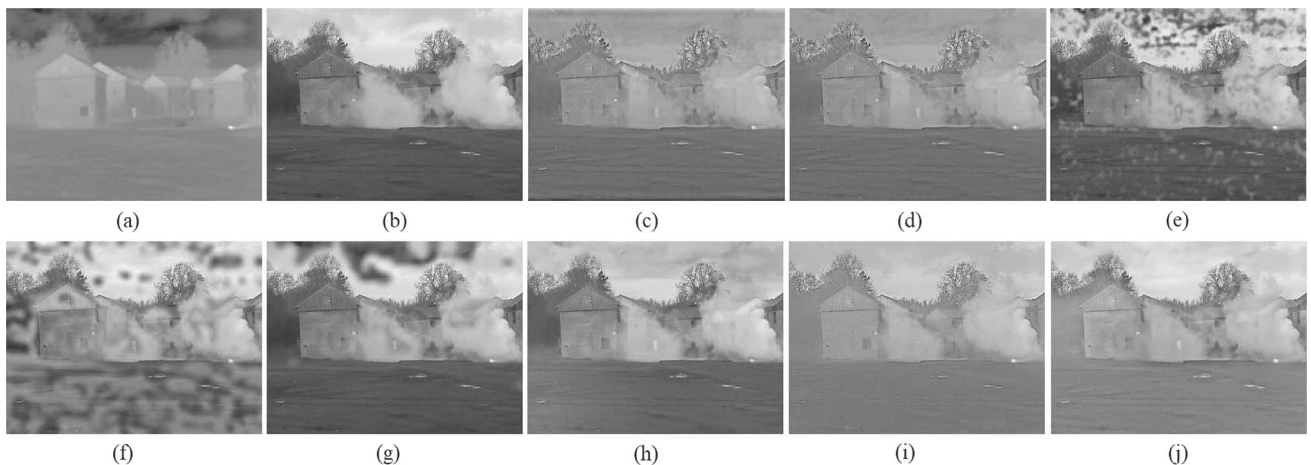


Fig. 12 Comparison with non-EMD-based methods for the fusion of an infrared image and a visible image. **a** Source image 1. **b** Source image 2. **c** CVT [29]. **d** DTCWT [33]. **e** NSCT-SR [30]. **f** NSCT-PCLE [41]. **g** MISF [42]. **h** MSID [43]. **i** IFCNN [34]. **j** Ours

incorporate the bright target of the infrared image into the fusion result well, and IFCNN [34] loses some detail features of the visible image. In contrast, MSID [43] and our method obtain better visual results than the other methods. Table 3 lists the objective metrics of the compared fusion methods on the tested data sets in Fig. 6. It can be seen that our method and BRW [39] have some advantages on most objective metrics in greyscale and color multi-focus set, respectively. As for multi-modal images, MISF [42] obtains the best performance on most objective metrics, while our method can compare favorably with the other methods when considering all metrics together. In terms of time performance, we observe our method is much faster than the methods listed the running time in the fusion of multi-focus images as shown in Table 3. Although the efficiency of our method is reduced in the fusion

of multi-modal images since a larger overlapping number are used, it only needs one second around to complete the fusion of two input images on average.

6 Conclusion

In this paper, we have presented a novel and fast EMD-based image fusion method via morphological filter in order to generate high-quality fusion images. A multi-channel bidimensional EMD method based on morphological filter (MF-MBEMD) is first developed to decompose the input images into several IMFs with different scales and a residue, which uses the morphological expansion and erosion filters to compute the upper and lower envelopes of a multi-channel

Table 3 Comparison with non-EMD-based image fusion methods on the objective metrics and the average running time for the data sets in Fig. 6

Data set	Method	Q_{MI}	Q_{FMI}	Q_{NCIE}	Q_G	Q_P	Q_E	Q_Y	Q_{CB}	Q_{CV}	Q_{VIF}	Time (s)
Greyscale multi-focus set	CVT [29]	0.826/0.207	0.888/ 0.050	0.822/ 0.007	0.678/0.039	0.736/0.091	0.818/0.072	0.899/0.041	0.683/0.037	46.002/28.120	1.411/0.211	0.605
	DTCWT [33]	0.858/0.211	0.889/ 0.050	0.824/0.008	0.694/0.033	0.761/0.085	0.820/0.070	0.961/0.112	0.708/ 0.033	44.564/27.657	1.435/0.193	0.164
	NSCT [28]	0.869/0.227	0.888/ 0.050	0.824/0.008	0.695/0.042	0.757/0.106	0.826 /0.072	0.974/0.151	0.707/0.043	45.165/29.127	1.437/0.249	1.914
	NSCT-SR [30]	1.022/0.179	0.890/ 0.050	0.832/0.008	0.707/0.032	0.775/0.076	0.825/0.069	0.935/0.027	0.717/0.056	39.507/25.182	1.490/0.178	40.798
	BRW [39]	1.138/0.142	0.891/0.050	0.838/ 0.007	0.724/0.029	0.804/0.071	0.822/0.068	0.977/ 0.013	0.763/0.034	42.735/30.664	1.515 /0.156	0.276
	MADCNN [40]	1.066/0.220	0.888/0.051	0.835/0.010	0.705/0.052	0.774/0.101	0.818/0.076	0.952/0.022	0.736/0.044	46.911/34.583	1.482/0.212	–
	IFCNN [34]	0.869/0.237	0.883/ 0.050	0.824/0.008	0.664/0.065	0.731/0.125	0.812/0.090	0.910/0.031	0.683/0.052	58.437/42.514	1.424/0.275	–
	Ours	1.163/0.130	0.891/0.050	0.840/0.007	0.724/0.028	0.795/ 0.066	0.816/ 0.067	0.978/0.013	0.764/0.036	43.840/31.138	1.510/ 0.155	0.146
	CVT [29]	0.904/0.143	0.900/ 0.028	0.828/0.008	0.731/0.035	0.813/0.047	0.878/0.030	0.948/0.021	0.728/0.047	20.539/ 14.295	1.535/0.167	3.061
	DTCWT [33]	0.940/0.148	0.901/ 0.028	0.830/0.009	0.745/0.033	0.833/0.041	0.880/0.030	0.965/0.014	0.751/0.044	20.524/15.159	1.548/0.160	0.803
Color multi-focus set	NSCT [28]	0.953/0.140	0.901/ 0.028	0.831/0.008	0.745/0.032	0.834/0.039	0.884 /0.029	0.962/0.014	0.748/0.043	21.937/15.855	1.567/0.160	10.963
	NSCT-SR [30]	1.119/ 0.101	0.901/ 0.028	0.841/ 0.007	0.758/0.031	0.846/0.039	0.883/0.030	0.977/0.010	0.793/0.043	15.974/15.144	1.600/0.140	159.032
	BRW [39]	1.166/0.107	0.901/ 0.028	0.844/0.008	0.762/0.029	0.850/0.039	0.882/0.030	0.988 /0.006	0.808 /0.040	15.875 /14.434	1.604/0.134	1.508
	MADCNN [40]	1.165/0.105	0.901/ 0.028	0.844/0.008	0.762/0.029	0.847/ 0.038	0.881/0.031	0.986/ 0.005	0.806/ 0.036	16.234/14.566	1.602/ 0.134	–
	IFCNN [34]	0.938/0.131	0.896/ 0.028	0.830/0.008	0.730/0.031	0.818/0.040	0.879/ 0.028	0.951/0.017	0.729/0.046	19.956/15.510	1.566/0.155	–
	Ours	1.177/0.103	0.902/0.028	0.845/0.008	0.759/0.030	0.843/0.042	0.878/0.031	0.986/0.006	0.806/0.040	16.016/14.589	1.601/0.135	0.651
	CVT [29]	0.582/0.117	0.887/0.028	0.807/ 0.001	0.558/0.065	0.371/0.144	0.624/0.115	0.568/0.112	0.498/ 0.048	55.692/447.449	0.717/0.223	0.327
	DTCWT [33]	0.627/0.126	0.893/0.026	0.807/ 0.001	0.588/0.061	0.430/0.158	0.638/0.109	0.700/0.081	0.552/0.064	535.972/399.413	0.790/0.247	0.070
	NSCT-SR [30]	0.835/0.142	0.899/0.025	0.811/0.003	0.646/0.073	0.473/0.188	0.684/0.108	0.910/0.044	0.647/0.107	537.728/432.502	0.916/0.289	11.845
	NSCT-PCLE [41]	0.745/0.126	0.894/0.027	0.809/0.002	0.621/0.080	0.472/0.188	0.671/0.121	0.837/0.074	0.622/0.096	492.476/521.492	0.988/0.250	1.785
Infrared-visible multi-modal set	MISF [42]	1.003 /0.166	0.904/0.024	0.814/0.004	0.687/0.059	0.639/0.136	0.677/0.106	0.957/0.023	0.679 /0.121	530.208/517.127	1.098 /0.255	0.029
	IMA [26]	0.758/0.121	0.900/0.024	0.810/0.002	0.669/0.061	0.679/0.115	0.637/0.144	0.769/0.161	0.538/0.123	636.985/632.224	1.030/0.239	0.165
	IFCNN [34]	0.740/ 0.107	0.884/0.029	0.809/ 0.001	0.596/0.069	0.440/0.146	0.669/0.099	0.851/0.044	0.609/0.110	502.167/408.97	0.851/ 0.201	–
	Ours	0.842/0.124	0.905/0.023	0.811/0.003	0.566/0.086	0.589/0.143	0.695 /0.105	0.918/0.040	0.646/0.115	383.739/307.492	1.024/0.257	0.658
	CVT [29]	0.326/ 0.101	0.905/0.035	0.805/ 0.003	0.566/0.086	0.394/0.175	0.673/ 0.106	0.765/0.078	0.505/ 0.090	488.432/ 441.044	0.683/0.199	0.560
	DTCWT [33]	0.347/0.108	0.902/0.039	0.805/ 0.003	0.611/0.084	0.454/0.181	0.692/0.107	0.788/0.101	0.506/0.104	682.680/812.790	0.721/ 0.188	0.254
	NSCT-SR [30]	0.647/0.191	0.901/ 0.031	0.816/0.009	0.627/0.105	0.398/0.241	0.690/0.148	0.828/0.092	0.583/0.100	1188.000/1590.272	0.835/0.264	30.372
	NSCT-PCLE [41]	0.424/0.142	0.903/0.039	0.808/0.004	0.582/0.097	0.382/0.224	0.695/0.119	0.793/0.111	0.532/0.131	613.915/746.474	0.858/0.228	4.140
	MISF [42]	0.787 /0.186	0.915 /0.034	0.823 /0.012	0.673/0.072	0.542/0.186	0.708 /0.111	0.917/0.054	0.644 /0.094	835.008/801.133	1.011 /0.192	0.076
	MSID [43]	0.419/0.115	0.902/0.038	0.807/ 0.003	0.650/0.075	0.506/0.159	0.696/0.128	0.813/0.065	0.508/0.114	472.907/602.834	0.987/0.219	0.102
	IFCNN [34]	0.431/0.113	0.893/0.040	0.807/ 0.003	0.576/0.087	0.426/0.190	0.670/0.117	0.813/0.070	0.502/0.119	518.969/614.769	0.807/0.206	–
	Ours	0.502/0.196	0.911/0.035	0.810/0.009	0.649/0.077	0.553/0.164	0.679/0.138	0.837/0.105	0.519/0.153	427.899 /566.891	0.931/0.218	1.662

Each metric is reported by the mean and standard deviation (mean/std) over all the samples in each data set. For each item, the bold is the best result in every data set. The running time of MADCNN [40] and IFCNN [34] are not listed since they need be computed efficiently via GPU

image. It can significantly improve the computation efficiency of EMD-based image fusion techniques. And then, a patch-based fusion strategy with overlapping partition is adopted to instead the pixel-based fusion method commonly used in EMD-based image fusion, where an energy-based maximum selection rule is designed to fuse the IMFs, and the feature information extracted by IMFs is used as a guide to merge the residue. Finally, the final result is generated by adding all fused IMFs and fused residue together. Our newly proposed EMD-based image fusion method can be implemented effectively by accumulating the fused values at each pixel patch by patch. The performance evaluation of the EMD-based image fusion methods on several commonly used data sets with multi-focus and multi-modal images shows that our newly proposed image fusion method obtains better results. Furthermore, a large number of comparative experiments have also demonstrated our method is very competitive with the state-of-the-art image fusion methods in visualization, objective metrics, and time performance.

It is still possible to continue to improve the performance of our method in the following aspects. Our method can reduce the artifacts in the fusion of boundary regions of multi-focus images as illustrated in Fig. 11. However, the fusion of boundary regions is still a challenge especially for the regions with irregular shapes, which is an open problem in the multi-focus image fusion as introduced in a recent survey of this topic [25]. In addition, our method treats each patch separately in the fusion processing, which results in the loss of detail features in the fusion of multi-modal images. The relationship of different patches should be considered in order to improve the fusion quality.

Acknowledgements We would like to thank the anonymous reviewers for their helpful comments. This work is supported in part by National Science Foundation of USA (IIS-1812606, IIS-1715985); National Natural Science Foundation of China (No. 61672149, 61532002, 61602344, 61802279); the Open Project Program of the State Key Lab of CAD&CG(No. A2105), Zhejiang University.

Declaration

Conflict of interest The authors declare no conflict of interest in this paper.

References

- Goshtasby, A.A., Nikolov, S.: Image fusion: advances in the state of the art. *Inf. Fus.* **8**(2), 114–118 (2007)
- Ma, J., Ma, Y., Li, C.: Infrared and visible image fusion methods and applications: a survey, *Inf. Fus.* pp 153–178 (2019)
- Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **454**(1971), 903–995 (1998)
- Ahmed, M.U., Mandic, D.P.: Image fusion based on fast and adaptive bidimensional empirical mode decomposition, In: 2010 13th International Conference on Information Fusion, 2010, pp. 1–6
- Yeh, M.H.: The complex bidimensional empirical mode decomposition. *Signal Process.* **92**(2), 523–541 (2012)
- Qin, X., Zheng, J., Hu, G., Wang, J.: Multi-focus image fusion based on window empirical mode decomposition. *Infrared Phys. Technol.* **85**, 251–260 (2017)
- Rehman, N., Ehsan, S., Abdullah, S., Akhtar, M., Mandic, D., McDonald-Maier, K.: Multi-scale pixel-based image fusion using multivariate empirical mode decomposition. *Sensors* **15**(5), 10923–10947 (2015)
- Pan, J., Tang, Y.Y.: A mean approximation based bidimensional empirical mode decomposition with application to image fusion. *Digital Signal Process.* **50**, 61–71 (2016)
- Wang, P., Fu, H., Zhang, K.: A pixel-level entropy-weighted image fusion algorithm based on bidimensional ensemble empirical mode decomposition. *Int. J. Distrib. Sens. Netw.* **14**(12), 1–16 (2018)
- Xia, Y., Zhang, B., Pei, W., Mandic, D.P.: Bidimensional multivariate empirical mode decomposition with applications in multi-scale image fusion. *IEEE Access* **7**, 114261–114270 (2019)
- Zhu, P., Liu, L., Zhou, X.: Infrared polarization and intensity image fusion based on bivariate bemd and sparse representation. *Multimed. Tools Appl.* **80**, 4455–4471 (2021)
- Nunes, J.C., Bouaoune, Y., Delechelle, E., Niang, O., Bunel, P.: Image analysis by bidimensional empirical mode decomposition. *Image Vis. Comput.* **21**(12), 1019–1026 (2003)
- Al-Baddai, S., Al-Subari, K., Tom, A.M., Sol-Casals, J., Lang, E.W.: A green function-based bi-dimensional empirical mode decomposition. *Inf. Sci.* **348**, 305–321 (2016)
- Hu, J., Wang, X., Qin, H.: Improved, feature-centric emd for 3d surface modeling and processing. *Graph. Models* **76**(5), 340–354 (2014)
- Hu, J., Wang, X., Qin, H.: Novel and efficient computation of hilbert-huang transform on surfaces. *Comput. Aided Geom. Design* **43**, 95–108 (2016)
- Wang, X., Hu, J., Guo, L., Zhang, D., Hong, Q., Hao, A.: Feature-preserving, mesh-free empirical mode decomposition for point clouds and its applications. *Comput. Aided Geomet. Design* **59**, 1–16 (2018)
- Wang, X., Hu, K., Hu, J., Du, L., Ho, A.T.S., Qin, H.: Robust and blind image watermarking via circular embedding and bidimensional empirical mode decomposition. *Vis. Comput.* **36**(19), 2201–2214 (2020)
- Wang, X., Hu, K., Hu, J., Du, L., Ho, A.T.S., Qin, H.: A novel robust zero-watermarking algorithm for medical images. *Vis. Comput.* **37**, 2841–2853 (2021)
- Bhuiyan, S., Adhami, R.R., Khan, J.F.: Fast and adaptive bidimensional empirical mode decomposition using order-statistics filter based envelope estimation. *Eurasip J. Adv. Signal Process.* **2008**(164), 1–18 (2008)
- Trusiak, M., Wielgus, M., Patorski, K.: Advanced processing of optical fringe patterns by automated selective reconstruction and enhanced fast empirical mode decomposition. *Opt. Lasers Eng.* **52**, 230–240 (2014)
- Mandic, D.P., ur Rehman, N., Wu, Z., Huang, N.E.: Empirical mode decomposition-based time-frequency analysis of multivariate signals: the power of adaptive data analysis. *IEEE Signal Process. Mag.* **30**(6), 74–86 (2013)
- Rehman, N., Mandic, D.P.: Multivariate empirical mode decomposition. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **466**(2117), 1291–1302 (2010)

23. Bhuiyan, S., Khan, J.F., Alam, M.S., Adhami, R.R.: Color image trend adjustment using a color bidimensional empirical mode decomposition method. *J. Electron. Imaging* **21**(3), 234–242 (2012)
24. Liu, Y., Chen, X., Wang, Z., Wang, Z.J., Ward, R.K., Wang, X.: Deep learning for pixel-level image fusion: recent advances and future prospects. *Inf. Fus.* **42**, 158–173 (2018)
25. Yu, L., Lei, W., Juan, C., Chang, L., Xun, C.: Multi-focus image fusion: a survey of the state of the art. *Inf. Fus.* **64**, 71–91 (2020)
26. Sufyan, A., Imran, M., Shah, S.A., Shahwani, H., Wadood, A.A.: A novel multimodality anatomical image fusion method based on contrast and structure extraction. *Int. J. Imag. Syst. Technol.* **32**(1), 324–342 (2022)
27. Li, X., Li, H., Yu, Z., Kong, Y.: Multifocus image fusion scheme based on the multiscale curvature in nonsubsampling contourlet transform domain. *Opt. Eng.* **54**(7), 1–15 (2015)
28. Li, H., Chai, Y., Li, Z.: Multi-focus image fusion based on nonsub-sampled contourlet transform and focused regions detection. *Optik Int. J. Light Electron Opt.* **124**(1), 40–51 (2013)
29. Nencini, F., Garzelli, A., Baronti, S., Alparone, L.: Remote sensing image fusion using the curvelet transform. *Inf. Fus.* **8**(2), 143–156 (2007)
30. Liu, Y., Liu, S., Wang, Z.: A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fus.* **24**, 147–164 (2015)
31. Naidu, V.: Multi-resolution image fusion by fft, In: International Conference on Image Information Processing **2011**, 1–6 (2011)
32. Li, H., Manjunath, B.S., Mitra, S.: Multisensor image fusion using the wavelet transform. *Gr. Models Image Process.* **57**(3), 235–245 (1995)
33. Lewis, J.J., OCallaghan, R., Nikolov, S.G., Bull, D.R., Canagarajah, N.: Pixel- and region-based image fusion with complex wavelets. *Inf. Fus.* **8**(2), 119–130 (2007)
34. Yu, Z.A., Yu, L.B., Peng, S.C., Han, Y.A., Xz, D., Li, Z.A.: Ifcnn: a general image fusion framework based on convolutional neural network. *Inf. Fus.* **54**, 99–118 (2020)
35. Haghighat, M.B.A., Aghagolzadeh, A., Seyedarabi, H.: A non-reference image fusion metric based on mutual information of image features. *Comput. Electr. Eng.* **37**(5), 744–756 (2011)
36. Xydeas, C.S., Petrovic, V.: Objective image fusion performance measure. *Electron. Lett.* **36**(4), 308–309 (2000)
37. Yang, C., Zhang, J.-Q., Wang, X.-R., Liu, X.: A novel similarity based quality metric for image fusion. *Inf. Fus.* **9**(2), 156–160 (2008)
38. Han, Y., Cai, Y., Cao, Y., Xu, X.: A new image fusion performance metric based on visual information fidelity. *Inf. Fus.* **14**(2), 127–135 (2013)
39. Ma, J., Zhou, Z., Wang, B., Miao, L., Zong, H.: Multi-focus image fusion using boosted random walks-based algorithm with two-scale focus maps. *Neurocomputing* **335**, 9–20 (2019)
40. Lai, R., Li, Y., Guan, J., Xiong, A.: Multi-scale visual attention deep convolutional neural network for multi-focus image fusion. *IEEE Access* **7**, 114385–114399 (2019)
41. Zhu, Z., Zheng, M., Qi, G., Wang, D., Xiang, Y.: A phase congruency and local laplacian energy based multi-modality medical image fusion method in nsct domain. *IEEE Access* **7**, 20811–20824 (2019)
42. Zhan, K., Kong, L., Liu, B., He, Y.: Multimodal image seamless fusion. *J. Electron. Imaging* **28**(2), 1–9 (2019)
43. Li, H., Qi, X., Xie, W.: Fast infrared and visible image fusion with structural decomposition. *Knowl.-Based Syst.* **204**, 106182 (2020)



Qi Xie is a Ph.D. student in School of Mathematics at Jilin University. She received her M.S. in Computational Mathematics at Jilin University in 2007. She is also a lecturer in School of Sciences at Northeast Electric Power University. Her research interests include digital image processing and geometry processing.



Jianping Hu Jianping Hu is a professor in School of Sciences at Northeast Electric Power University. He obtained his B. S. and M. S. in Mathematics at Jilin University. He received his Ph. D. in Computational Mathematics at Dalian University of Technology in 2009. His research interests include computer graphics, digital geometry processing, and image processing.



Xiaochao Wang is an associate professor in School of Mathematical Sciences at Tiangong University. He received his PhD degree in Computational Mathematics from Dalian University of Technology in 2014. His research interests include computational geometry, computer graphics, and image processing.



Daochang Zhang is an associate professor in School of Sciences at Northeast Electric Power University. He received his PhD degree in Mathematics from Jilin University in 2015. His research interests include matrix analysis, matrix algebra, and image processing.



Hong Qin is a full professor of Computer Science in Department of Computer Science at Stony Brook University (State University of New York). He received the BS and MS degrees from Peking University, and the PhD degree from the University of Toronto, all in computer science. His research interests include geometric and solid modeling, graphics, physics-based modeling and simulation, computer-aided geometric design, visualization, and scientific computing.