



A scheme for edge-based multi-focus color image fusion

Manali Roy¹  · Susanta Mukhopadhyay¹

Received: 23 March 2019 / Revised: 15 April 2020 / Accepted: 27 May 2020 /

Published online: 17 June 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

In this paper, a novel region-based multi-focus color image fusion method is proposed, which employs the focused edges extracted from the source images to obtain a fused image with better focus. At first, the edges are obtained from the source images, using two suitable edge operators (Zero-cross and Canny). Then, a block-wise region comparison is performed to extract out the focused edges which have been morphologically dilated, followed by the selection of the largest component to remove isolated points. Any discontinuity in the detected edges is removed by consulting with the edge detection output from the Canny edge operator. The best reconstructed edge image is chosen, which is later converted into a focused region. Finally, the fused image is constructed by selecting pixels from the source images with the help of a prescribed color decision map. The proposed method has been implemented and tested on a set of real 2-D multi-focus image pairs (both gray-scale and color). The algorithm has a competitive performance with respect to the recent fusion methods in terms of subjective and objective evaluation.

Keywords Multi-focus fusion · Canny edge operator · Morphological dilation · Morphological edge reconstruction · Euler number · Decision map

1 Introduction

The motivation for the study and research of multi-focus image fusion is to provide a practical approach to solve the problem of adaptive focusing ability of an imaging device. Normal imaging devices generally contain converging (convex) lenses, which captures a bundle of light rays originating from a specific point on the object and converges them to a single point in the focal plane. Such lenses are unable to produce a homogeneous focused capture of an object or scene due to its limited range of focus or Depth of Field (DoF). In optics,

✉ Manali Roy
manalir66@gmail.com

Susanta Mukhopadhyay
msusanta2001@gmail.com

¹ Department of Computer Science and Engineering, Indian Institute of Technology (ISM), Dhanbad, India

the range of focus is defined as the distance between the nearest and the farthest objects in a scene that appears to be acceptably sharp in an image. Objects lying within the focal range appear to be more sharp and clear compared to the objects which are away from it. Focused capture of a scene comprises of more feature details and spatial information such as edges, color, contour, texture, intensity. So, it is a necessary task to achieve uniform focus throughout the image for better human understanding and machine perception. This can be achieved by a multi-focus fusion algorithm which takes one or more input images with different levels of focus to produce a more interpretive all-in-focus image. The focussed image can be utilized in solving many image processing problems such as feature extraction, contour detection, object segmentation, and recognition. Image fusion is largely applicable in the field of computer vision, geographic information systems, biomedical research, robotics, navigation and, surveillance systems [2].

2 Related work

Fusion has been a widely researched domain in the field of image processing literature, and yet it continues to be an active area of research. Image fusion algorithms generally take place at different (a) levels of abstraction/information representation (pixel, feature, and decision), (b) domains (spatial, spectral and hybrid) and (c) types of source images (multi-focus, multi-sensor, multi-exposure and multispectral) [15]. Several algorithms for multi-focus image fusion have been introduced for both gray-scale and color images in the literature. Pixel level fusion methods can be further divided into three groups, i.e., coefficient based, window-based, and region-based. Coefficient based methods in pixel domain follow a general three-step procedure, 1) application of an appropriate transform operation on the image pixels to obtain the domain-specific coefficients, 2) combining the coefficients so obtained by employing suitably devised rules, 3) reconstruction of the fused image by taking an inverse transform. Some notable transforms that are used in the literature are discrete cosine transform (DCT), discrete wavelet transform (DWT), dual-tree complex wavelet transform (DTCWT), log-Gabor transform, discrete cosine harmonic wavelet transform (DCHWT), curvelet, contourlet and ripplelet [7] transform, non-subsampled contourlet transform (NSCT), scale-invariant feature transform (SIFT), non-subsampled shearlet transform etc. [12]. The choice of transform depends on the type of the source images to be fused. Transform domain methods modify the pixel values and require perfect inversion to the spatial domain for proper visualization. Pixel-based methods in spatial domain directly manipulates the pixel intensity or integrates spatial features (both global and local) extracted from the source images to achieve a uniformly focused image. Few of the simplest approaches include simple or weighted average and selection of maximum or minimum pixel. The evaluation of focus measures (e.g., spatial frequency, sum-modified-laplacian, Tenengrad measure, energy of gradient, index of fuzziness and moment-based measure) on a block of pixel to measure/rank the activity level of a pixel/block has been popularly used in this domain [10, 30]. Spatial level image fusion is also performed using different (a) colour spaces (HSI, RGB, LUV) , (b) dimensionality reduction techniques (principal component analysis (PCA) [27], independent component analysis (ICA) [20]) and (c) specialized filters (fast filter [38], guided filter [13], edge-preserving filters like bilateral filter [25]). The methods discussed above are also applied in one or more combination at multiple resolution levels (multi-resolution methods) to extract the coarse and fine features from the image where the number of resolution levels largely determines the quality of the fused images.

The concepts of machine learning and deep learning [36] have also been utilized to develop image fusion algorithms which uses pulse-coupled neural network (PCNN) [11], convolutional neural network (CNN) [31], multi-scale CNN [4] etc. Although these methods require the network to be trained with numerous training data prior to testing [18]. Zhang et al. has presented a detailed review of multi-focus fusion techniques based on sparse representation [23, 40]. Sparse coding mechanism which simulates the behaviour of human vision are also used to develop algorithms using dictionary and sub-dictionaries like convolutional sparse representation (CSR) [34],[37] adaptive sparse representation (ASR) [1, 33] etc. However, computational complexity still happens to be a major issue in these algorithms [5, 16].

Window/coefficient based methods in pixel domain often introduce intensity variation, blocking artifacts, blurring effect, sensitivity to noise in the fused images resulting in the introduction of region-based fusion approaches. In such techniques, the irregular semantic regions are segmented/extracted at first, leading to the creation of a joint/separate segmentation map prior to fusion [19]. Morphology has been repeatedly used to distinguish between focussed and defocussed pixels. Li et al. has proposed a matting based fusion approach where the focused region is roughly obtained by morphological filtering followed by image matting to distinguish the foreground from the background [14]. A novel algorithm to obtain a boundary between the focused and defocused regions within the image using a multi-scale morphological focus measure is proposed in [35]. Baohua et al. have used a graph-based visual saliency (GBVS) algorithm followed by morphological watershed transform to extract the focused regions by [39]. Similar region partitioning strategies proposed for image fusion can be found in [6, 17]. In this paper, a novel region-based algorithm for multi-focus images is proposed where the focused region is obtained from the edges of the source images assuming the fact that the focused portion of an image contains more prominent and clear edge features. Hence, the edge images are used as a starting point to form a region for the fusion process. The efficacy of the proposed method has been evaluated qualitatively and quantitatively with appropriate fusion metrics and compared with the state-of-art.

The contribution of the paper is as follows:

- It is a region-based fusion method which uses edge features as a basis for focus/saliency detection.
- The threshold values used in the edge detection procedure do not require any tuning.
- The focused edges are completely separated from the defocused ones using morphological edge by reconstruction, which are gradually converted into region.
- It produces significantly good results in less execution time and works for multiple focus situation, as evident from experimental results.

The rest of the paper is organized as follows: Section 3 consists of a detailed description of the proposed fusion algorithm. Section 4 presents the experimental results along with subjective and objective evaluations with some future directions. The concluding remarks are drawn in Section 5.

3 Proposed method

A multi-focus fusion algorithm should be able to isolate and extract the maximum amount of focused information from the source images so as to construct the resultant fused image by combining them. Accordingly, the proposed algorithm first separates the more focused

features from the ill-focused features (Fig. 1a and c). The constituting color images, are first converted into YUV format using the weighted sum of red, green and blue channel, $I = 0.299 * R + 0.587 * G + 0.114 * B$ prior to finding the edge features [8]. The edge features from the constituting images are then extracted employing the zero crossing edge detector using (1). The focused edges in the component images are those which have a significant concentration of foreground pixels in the edge image, as obtained in Fig. 1b and d.

$$LoG(x, y) = \left[\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right] e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (1)$$

For each of the source image, the edge detection is performed using two edge operators with varying detection strength, i.e., (a) Zero-crossing with LoG and (b) Canny edge operator [3]. The edge output obtained from (a) is used in Section 3.1 as a starting edge map to detect the focused edge pixels. The edge image obtained from (b) has been used for two different purposes: 1) It is used as a reference image for correcting the discontinuities that may arise as a result of block-wise operation (Section 3.3). 2) It is used in Section 3.3.1 to perform XOR operation to find the number of edge pixels that lie in the focussed region. The reason behind choosing two separate edge detectors is explained as follows: Being a weak edge detector, zero-cross edge operator detects the edges belonging to the focused region of the image alone (Fig. 2b and c). On the other hand, the Canny operator performs an optimal edge detection and produces extra edges belonging to the poorly focused regions that may be connected with the edges from the focussed regions (Fig. 2f). Such extra edges are undesirable here, and to mitigate the problem, the upper and lower threshold values used in the Canny edge detection algorithm could have been fine-tuned to control the amount of edges to be detected. But, this approach becomes time consuming and image dependent because the focused features in individual source images are present at different locations.

3.1 Blockwise Region Comparison

The binary edge images obtained from the zero-cross edge operator has separated the set of focused features from those of defocused ones to a significant extent. The concentration of the foreground pixels (white) in the focused regions are more than that corresponding to the regions out of focus. However, due to false alarm in the edge detection process, some of the non-focussed pixels will also have a contribution in the same, which are subsequently removed by an additional block-based region comparative algorithm [29]. In this algorithm, we compare the corresponding blocks (non-overlapping) of $n \times n$ pixels of the candidate images and construct the focused edge image as:

$$E_{FA}(m, n) = \begin{cases} 1, & \text{if } S_A > S_B \text{ and } E_{RA}(m, n) = 1 \\ 0, & \text{others} \end{cases} \quad (2)$$

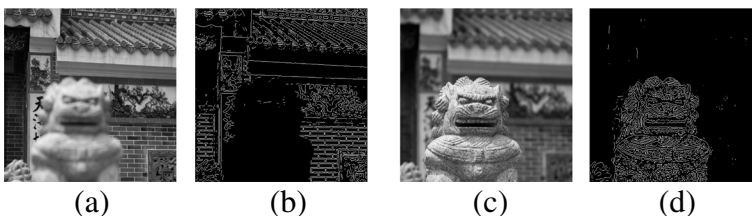


Fig. 1 **a** Focus on Background; **b** Edge Image of **(a)**; **c** Focus on Foreground; **d** Edge Image of **c**

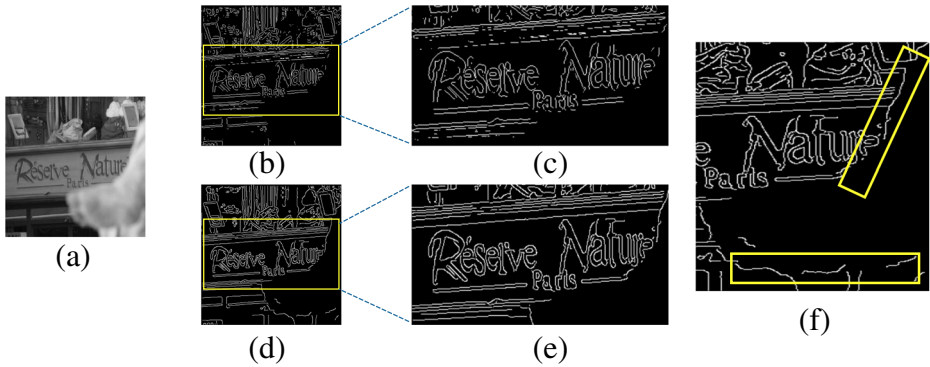


Fig. 2 **a** y component (Focus on the Background); **b** Edge image using Zero-cross; **c** Local magnified region of **(b)**; **d** Edge image using Canny; **e** Local magnified region of **(d)**; **f** Extra edges captured from ill-focused portion (Yellow Box)

$$E_{FB}(m, n) = \begin{cases} 1, & \text{if } S_A < S_B \text{ and } E_{RB}(m, n) = 1 \\ 0, & \text{others} \end{cases} \quad (3)$$

where, E_{FA} and E_{FB} are the edge pixels corresponding to focused edge features in the source images A and B respectively,

$$S_A = \sum_{(m,n) \in W_{n \times n}} E_{RA}$$

$$S_B = \sum_{(m,n) \in W_{n \times n}} E_{RB}$$

and S_A and S_B denotes the total contribution of the white pixels from $n \times n$ block from the binary edge images E_{RA} and E_{RB} obtained using zero-cross operator.

$$W_{n \times n} = \text{window of } n \times n \text{ pixels}$$

This method eliminates the maximum number of foreground pixels belonging to the out-of-focus edges of the image. It retains the pixels belonging only to the focused area of the image, as shown in Fig. 3b and d. Nonetheless, this process comes along with two drawbacks, as mentioned below:

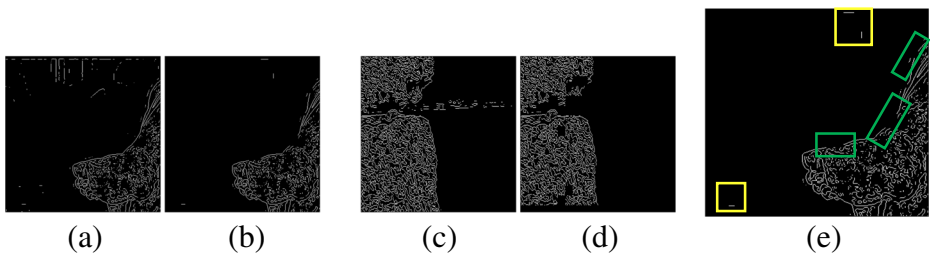


Fig. 3 **a, c** Before region comparison; **b, d** After region comparison; **e** Isolated pixels(yellow box) and introduced breaks(green box)

- As shown in Fig. 3e (Yellow box), in spite of the maximal removal of the blurred edge pixels, though a number of white pixels still remain in the picture. The block-wise area comparison algorithm is unable to remove the white pixels as they are *isolated* in nature.
- The algorithm discussed above makes a comparison between the local region of two source images, and this comparison tends to introduce breaks and dis-connectivity in the outer edges of the main structure in the focused portion, as shown in the Fig. 3e (Green Box).

The above two drawbacks are removed by performing the following steps (Sections 3.2 and 3.3).

3.2 Removal of isolated points

The isolated white foreground points against the dark background, as discussed above for both the source images, are removed by performing binary morphological dilation (4) followed by the largest connected component selection. The binary edge images obtained until now consists of edge pixels in a disconnected and distributed manner resulting in more than one connected component. Components that are not connected but exhibit a kind of coherence in terms of size and local orientation are more likely to be parts of bigger components. The dilation operation aims at doing justice to such coherent yet isolated components by connecting them, provided they are in close proximity as assessed by the size of the structuring element used in dilation. Binary morphological dilation is a procedure which thickens the object boundaries and causes a growth in such reasons. Therefore, it reduces the number of connected components in the edge image by merging/unifying a significant number of isolated components. The number of connected components before and after dilation for some of the source images are shown in the Table 1. The reduction in the number of components keeps the edge structure of the image intact. After dilation, the largest component is kept discarding all the smaller components (which also includes the isolated pixels). For all the input images, this combined approach of dilation followed by the largest component selection has removed all the isolated pixels, thereby leaving the focused edge region alone. However, binary dilation operation saturates after a certain number of iterations and fails to restore any wide gaps/breaks that exist in the outer boundary of the images. To recover the original edge structure of the image, morphological reconstruction is performed as briefly discussed in the next section. This is where the canny edge counterpart of the source image comes into play.

$$\text{Dilation: } I \oplus s = \{z | [(s')_z \cap I] \subseteq s\} \tag{4}$$

where, I and s are the image and structuring element respectively.

Table 1 Number of connected components: before and after dilation

The bold value indicates significant reduction in the number of components after dilation operation for background image

Image	$I_1(\text{Background})$		$I_2(\text{Foreground})$	
	Before	After	Before	After
Leg	2535	81	1846	51
Temple	5297	106	4504	18
Pot	1698	30	2801	113
Calender	319	3	261	4

3.3 Morphological edge reconstruction

Morphological reconstruction is a process in which an image known as ‘marker’ image gets repeatedly morphologically processed/modified based on the characteristics/features of another image called ‘mask’ [26]. The reconstruction takes place on the basis of specified connectivity. The marker and mask images should preferably be same in size and the number of elements in the marker image should be less than or equal to that in the mask image. In this context, the respective edges images after dilation (obtained in Section 3.2) formed after selecting the largest component becomes the **marker image** which gets morphologically reconstructed with respect to canny edge image chosen as the **mask image**. Canny edge image provides us with strongly connected edges containing the maximum amount of edge information. This reconstruction procedure successfully restores the wide breaks that get introduced as a result of the block-wise comparison. Figure 4 shows the reconstruction taking place for the source images, as mentioned. It is obvious that the number of pixels in the reconstructed image is less than its respective canny edge image counterpart for all the source images. To simplify the process of region conversion, only one of the reconstructed image is considered for region conversion. Hence, the next task would be to identify the reconstructed image that is structurally complete enough to convert into a region. This is obtained with the help of a focus measure, as discussed in the next section.

3.3.1 Choosing the best reconstructed image

For selecting the best-reconstructed image, we use a focus measure based on morphological filtering used in [14]. It uses the bright and dark top-hat transforms to extract the high-frequency details of the image. The initial focus information map (Fig. 5b, b1) is generated by taking the maximum value among the two transforms at each pixel location using (5):

$$F_I(x, y) = \max[T_I^b(x, y), T_I^d(x, y)] \tag{5}$$

where,

$$T_I^b(x, y) = \text{Bright top-hat transform of } I(x,y)$$

$$T_I^d(x, y) = \text{Dark top-hat transform of } I(x,y)$$

The image obtained is binarized using the condition given by (6). It provides better visualization of the blurred and prominent pixels for further processing (Fig. 5c, c1).

$$R_I(x, y) = \begin{cases} 1, & F_{I_1}(x, y) > F_{I_2}(x, y) \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

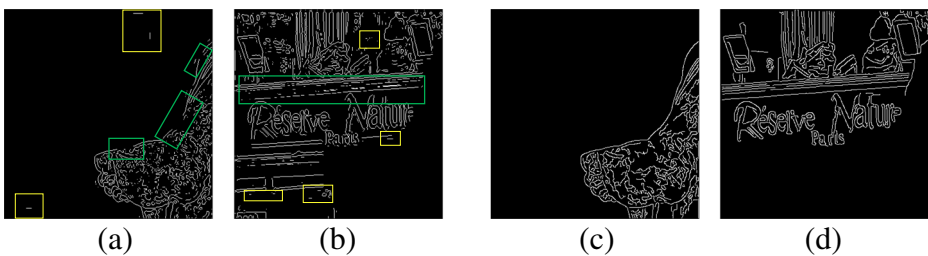


Fig. 4 **a** Foreground and **b** Background edge image with breaks (green box) and isolated pixels (yellow box); **c** Foreground and **d** Background edge images after reconstruction

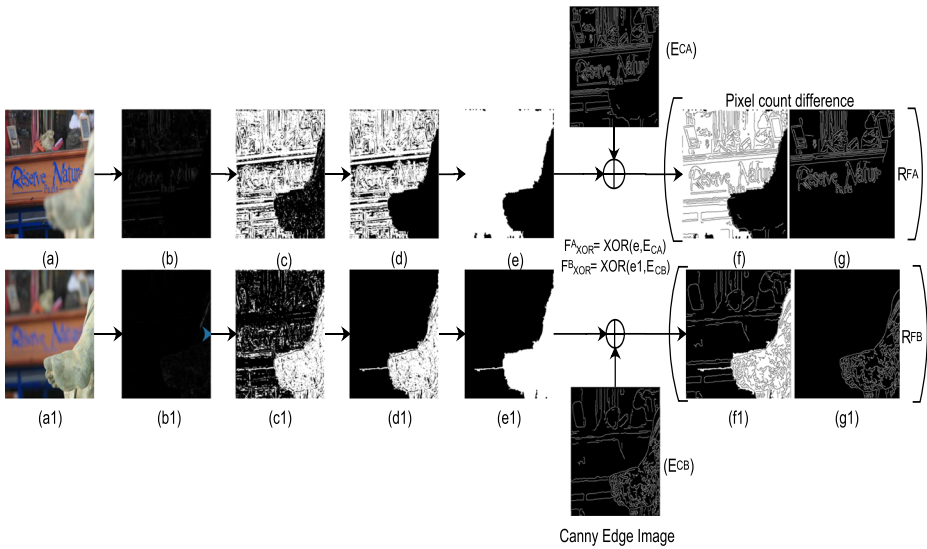


Fig. 5 Process representing the selection of best reconstructed image: **a, (a1)** Color/Gray-scale image; **b, (b1)** Initial focus map; **c, (c1)** Binary thresholded focus map; **d, (d1)** Cleaned focus map; **e, (e1)** Rough focus region; **f, (f1)** Resultant XOR image; **g, (g1)** Reconstructed Image

The output from (6) provides better demarcation of focused features from the blurred ones. The images obtained from (6) are further cleaned by sequential operation of morphological closing, $I \bullet s = (I \oplus s) \ominus s$ followed by hole-filling for obtaining the focused regions in a more prominent form that can serve the purpose (Fig. 5e, e1). Now, an exclusive OR operation is performed between this binary image and the canny edge image to count the number of edge pixels lying within the focused regions. To select the best reconstructed image obtained in Section 3.3, a simple rule is followed which uses the difference in the count of the dark pixels against bright background and vice-versa. Let P_d be the number of dark pixels within the white area of the XOR image (Fig. 5f, f1) and P_b denote the bright pixel count in the reconstructed image (R^{FA}, R^{FB}). Let D_{IA} and D_{IB} be the absolute differences in the pixel count given by, $D_{IA} = P_d(I_A) - P_b(I_A)$, $D_{IB} = P_d(I_B) - P_b(I_B)$. Then, the best reconstructed edge image (I_R) is given by,

$$I_R = \begin{cases} R_{FA}, & \text{if } D_{IA} < D_{IB} \\ R_{FB}, & \text{if } D_{IB} < D_{IA} \end{cases} \quad (7)$$

Table 2 Difference in Pixel Count

Image	Pixel counts					
	I_A (Background)			I_B (Foreground)		
	P_b	P_d	D_{IA}	P_b	P_d	D_{IB}
Leg	9552	12288	2736	5471	5527	56
Temple	10801	23606	12805	14474	14482	8
Pot	7083	8359	1276	4839	5195	356
Calender	479	1339	860	859	1255	396

The bold value indicates least absolute difference in the pixel count for the best reconstructed image

The image with less absolute difference in the pixel count is considered to be the best-reconstructed image because it signifies that the image has more number of pixels. The observations are presented in Table 2 below for some of the source images, which show that the near focused (foreground) object is better reconstructed as compared to the far-focused (background).

Algorithm 1 Algorithm for the best reconstructed image (I_R).

- 1: **Input:** Pair of a) source images (I_A, I_B); b) reconstructed edge images (R_{FA}, R_{FB}); c) canny edge images (E_{CA}, E_{CB})
 - 2: **Output:** Best reconstructed image (I_R)
 - 3: Focus maps generated using morphological filtering approach are cleaned to obtain rough focus regions, say F'_{IA} and F'_{IB} .
 - 4: $F_{XOR}^A = F'_{IA} \oplus E_{CA}$
 - 5: $F_{XOR}^B = F'_{IB} \oplus E_{CB}$
 - 6: The black pixels lying within the focussed regions are counted in F_{XOR}^A and F_{XOR}^B , denoted by $P_d(I_A)$ and $P_d(I_B)$.
 - 7: The white pixels are counted in R_{FA} and R_{FB} , denoted by $P_b(I_A)$ and $P_b(I_B)$.
 - 8: Calculate the absolute differences of the pixel count as, $D_{IA} = P_d(I_A) - P_b(I_A)$ and $D_{IB} = P_d(I_B) - P_b(I_B)$.
 - 9: The image with the less absolute difference is the best reconstructed image.
-

3.4 Region conversion

The objects in the edge images are generally disconnected in nature with more than one connected edge component. The best-reconstructed edge image selected by using (7) is converted into an approximate binary region by carrying out an iterative step combining a morphological closing operation and hole filling. For binary images, the holes are defined as a set of its regional minima, which are not connected to the image border. The holes are filled by removing all minima, which are not connected to the image border, by using morphological reconstruction by erosion. The marker image is set to the maximum value except along its border, where the values of the original images are kept. Closing operation in a binary image enlarges the foreground regions keeping the original boundary intact while the hole filling operation helps fasten the process of region conversion by reducing the number of iterations (Fig. 6). To fill in the gaps between the edges, a square matrix is chosen as the structuring element for the closing operation, the dimension of which changes with every iteration. The number of iterations required till saturation differs for each input edge image. The procedure is summarized in Algo. 2. It is observed that two different types of regions are hence formed:

- a) *Type I: Region touching the image boundary:* For regions touching the borders, the borders are identified as left, right, top or bottom (Fig. 6a and b). The small gaps between the objects and the identified image borders if any, are filled up with bright pixels.
- b) *Type II: Regions away from image boundary:* In this type, the region images obtained after execution of Algo. 2 is perfect for carrying out the fusion procedure and do not require any further processing.

An extra post-processing step is carried out for all the rough region images (Type I or Type II) to detect and fill leftover holes (if any). **Post-Processing of the images:** For certain

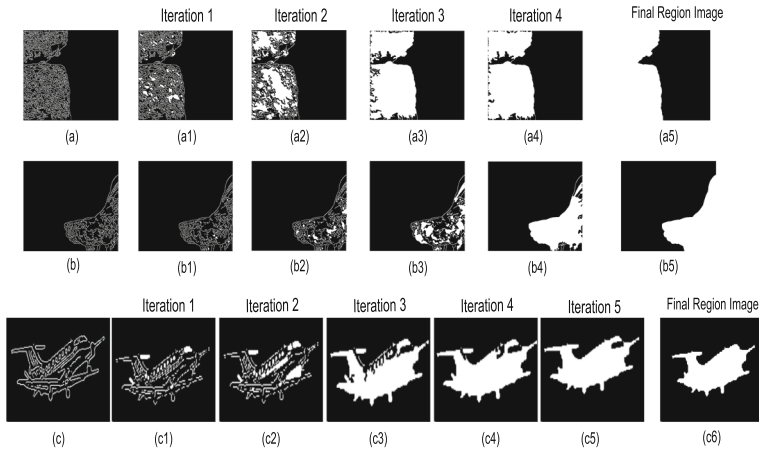


Fig. 6 Region conversion: **a** and **b** Type-I image; (a1-a4,b1-b4) Intermediate Images; (a4,b4) Partial region image; (a5,b5) Final region image after border detection and filling; **c** Type-II image; (c1-c5) Intermediate Images; (c5) Partial region image; (c6) Final region image after post-processing

region images, the hole filling scheme described above is not capable of vanishing all types of holes from the regions. For example, there may be holes in the regions which are connected to the background through narrow constriction (Fig. 6(c5)). To fill up such holes, we have first dilated the image so as to fill up narrow constrictions, which leaves bigger holes isolated from the background. Subsequently, we detect such holes by computing the **Euler Number** corresponding to all such regions.

b) *Detection of holes using Euler Number (E_u):* It is defined as the difference between the number of components and the number of holes in a binary image. Since all the region images obtained consists of a single component as of now, image consisting holes will have either 0, or negative Euler number. Hence, the Euler number is calculated after dilation to detect the presence of holes.

$$E_u = \begin{cases} 1, & \text{,no hole is present} \\ 0, / \text{less than } 1 & \text{,holes are present} \end{cases} \quad (8)$$

After hole detection and filling, the resulting image is again eroded by the same structuring element to bring the regions back to their original configuration, but with the holes disappearing (Fig. 6(c6)). The region image formed at this step itself acts as an initial focus information map for the fusion procedure.

3.5 Construction of decision map

A decision map (D) corresponding to the input image- color or gray-scale needs to be constructed after the region formation described above.

Gray-scale Images: For gray-scale images, the decision map is simply the output from the previous step, i.e., $D_{IG} = R_{IF}$.

Color Images: For color input images, the binary region image from the previous step is converted to a 2-color decision map as per (9). The color decision map is created as follows:

$$D_{IC}(x, y) = \begin{cases} C_1, & \text{if } R_I(x, y) = 0 \\ C_2, & \text{if } R_I(x, y) = 1 \end{cases} \quad (9)$$

where C_1 and C_2 refers to different colors.

Algorithm 2 Region conversion algorithm (I_R).

```

1: Input: Best reconstructed edge image ( $I_R$ )
2: Output: Final region image ( $R_{IF}$ )
3:  $N_c \leftarrow$  Number of components in the edge image ( $N_c > 1$ )
4:  $s_i \leftarrow$  Structuring element
5:  $R_I \leftarrow$  Intermediate region image
6: while  $N_c > 1$  do
7:   for  $i = 1 \rightarrow n$  do
8:      $s_i = i \times i$  matrix of logical 1's.
9:     Perform closing operation,  $I_R' = I_R \bullet s_i$ 
10:    Perform hole-filling of  $I_R'$ 
11:    Count and update  $N_c$ 
12:    if  $N_c = 1$  then break;
13:    end if
14:  end for
15: end while
16: Categorize  $R_I$  as Type-I or Type-II
17: if  $R_I$  belongs to Type-I then
18:   Identify the borders as left, right, top or bottom and fill the gaps with white pixels.
19:   Perform post processing to obtain  $R_{IF}$ .
20: end if
21: if  $R_I$  belongs to Type-II then
22:   Perform post processing to obtain  $R_{IF}$ .
23: end if

```

Algorithm 3 Post Processing algorithm (R_I).

```

1: Input: Intermediate region image ( $R_I$ )
2: Output: Final Region Image ( $R_{IF}$ )
3: Perform morphological dilation
4: Detect holes using Euler Number( $E_u$ )
5: Fill the detected holes
6: Perform Erosion by same amount

```

3.6 Image fusion

The fused image is formed by utilizing the decision map obtained in the previous section for gray-scale as well as color images. Depending on the decision map, the pixels from the respective source images are selected to construct the final fused image (Fig. 7). The fusion process is executed by using the fusion rule as specified in (10) and (11).

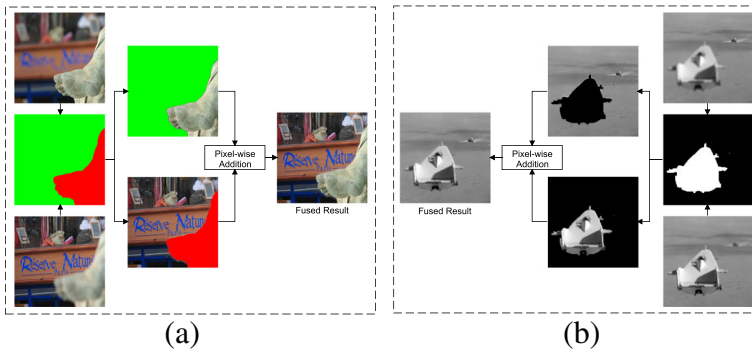


Fig. 7 Diagram representing the fusion process using the decision map for **(a)** color image; **(b)** grayscale image

Gray-scale images:

$$I_{FG}(x, y) = \begin{cases} I_1(x, y), & \text{if } D_{IG}(x, y) = 0 \\ I_2(x, y), & \text{if } D_{IG}(x, y) = 1 \end{cases} \tag{10}$$

where D_{IG} is the binary decision map.

Color Images:

$$I_{FC}(x, y) = \begin{cases} I_1(x, y), & \text{if } D_{IC}(x, y) = C_1 \\ I_2(x, y), & \text{if } D_{IC}(x, y) = C_2 \end{cases} \tag{11}$$

where D_{IC} is the colored decision map. The overall process flowchart of the fusion algorithm is given in Fig. 8.

4 Experiment results and discussion

4.1 Parameters

The various parameters associated with this experiment are: block size (n), standard deviation (σ_1) of the Gaussian filter used in Canny, (σ_2) of LoG filter, threshold values (t_1, t_2) and (t_3) used in Canny and zero-cross edge detection respectively and structuring element (s) used for morphological operations. The block size chosen for region comparison to separate the focussed edges are chosen by trial and error such that it produces visually superior results. The threshold values for edge detection are heuristically chosen by the edge detectors depending on the features present in the source images. However, all the threshold values lie between 0 and 1, some of which are presented in Table 3. The optimal value of σ in the Gaussian kernel depends on image factors such as the resolution of the image and size of the objects in it. For all such image dependent parameters, the default values are chosen without any tuning. For binary morphological operations, disk-shaped structuring element is chosen because of its isotropic property, which retains the image details without introducing block effects. Here, we have adopted the minimum value for the size of the structuring element. The values used for the above-mentioned parameters are provided in Table 4.

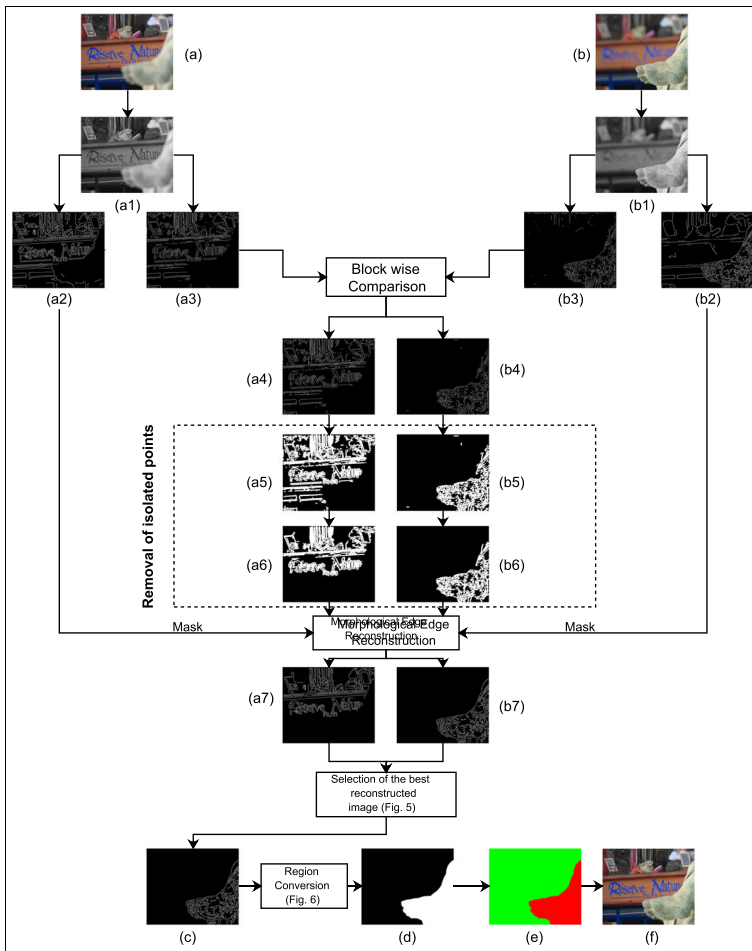


Fig. 8 Flowchart representing the fusion process: **a** and **b** Color Image; (a1,b1) Y-Component; (a2,b2) Canny edge image; (a3,b3) Zero-cross edge image; (a4,b4) Block-wise compared image; (a5,b5) Dilated images; (a6,b6) Removed isolated points; (a7,b7) Reconstructed image; **c** Best reconstructed image; **d** Region image; **e** Color decision map; **f** Fused image

Table 3 Threshold values used in edge detection

Image	Source 1		Source 2	
	Canny	Log	Canny	Log
Leg	(0.0500,0.1250)	0.0022	(0.0313,0.0041)	0.0055
Seascape	(0.0688,0.1719)	0.0066	(0.0781,0.0563)	0.1406
Calender	(0.0250,0.0625)	0.0025	(0.0250,0.0625)	0.0023
Jar	(0.0250,0.0625)	0.0024	(0.0250,0.0625)	0.0020
Temple	(0.0688,0.1719)	0.0059	(0.0563,0.1406)	0.1719

Table 4 Values of the parameters used in the experiment

Parameters	Used value
Block Size ($n \times n$)	18
Threshold value $[(t_1, t_2), (t_3)]$	(0,1)
Standard Deviation (σ_1, σ_2)	$(\sqrt{2}, 2)$
Structuring Element (s)	Disk of Radius 2

4.2 Comparison methods

The proposed method has been tested on 30 pairs of 2-D colored and grayscale multi-focus images, and the results are compared with other representative image fusion algorithms in terms of visual/subjective perception and objective evaluation. In the paper, we have presented the results for 16 pairs. All the algorithms selected for the purpose of comparison are CPU executable pixel/region based fusion methods implemented using gradient domain (GD) [21], adaptive block-based discrete wavelet transform (DWT-AB) [32], image matting (IM) [14] and Gaussian curvature filter (GCF) [24]. Additionally, to establish the efficacy of the proposed method with the state-of-art, it has also been compared with two learning-based methods, e.g., adaptive sparse representation (ASR) [33], and convolutional neural networks (CNN) [31]. In GD based method, ‘weighted sum’ is adopted to fuse the chrominance channels while the fused luminance channel is obtained by a gradient reconstruction technique based on ‘Haar’ wavelets. DWT-AB employs discrete wavelet transform with three levels of decomposition, in which the low-frequency coefficients are fused using adaptive block method, and high-frequency components are combined using local wavelet energy. IM is a spatial method which applies image matting technique on a roughly segmented result obtained by using morphological bright and dark top-hat transforms as focus measure. GCF initially uses a gaussian curvature filter to obtain the salient (sharpest) regions from the source images followed by a course fusion map using a synthetic focusing degree criterion combining spatial frequency and local variance. ASR proposes a learning-based fusion approach using sub-dictionaries, whereas CNN method uses convolutional neural networks for the same. The proposed algorithm, as well as the comparison methods, are implemented using Matlab programming language on a computer with 2.4 GHz CPU and 4GB RAM.

4.3 Objective evaluation

The quality of the fusion results can be quantitatively assessed by using several fusion metrics, which may or may not take reference (ground truth) image into account. In most of the cases, no-reference based fusion metric is adopted due to unavailability of perfect ground truth image. The concept of a perfectly fused image is subjective in nature because it is quite challenging to measure the ‘perfectness’ of a fused image. Moreover, the quality of source images, mis-registration, and illumination defects directly impact the results as well. Also the process of combining information from the source images might create additional effects such as contrast enhancement in the fused image, which is not desirable. So, for objective evaluation of the fused results, fusion quality indices are extensively used by researchers. Here, the results of the proposed algorithm are numerically evaluated using two groups of quality metrics, a) image quality metrics, (b) fusion quality metrics. Average pixel intensity (API), standard deviation (SD), average gradient (AG) judges the quality of fused image

beyond fusion (for e.g., enhancement of features) and are independent of the source images. On the other hand, feature mutual information (FMI), gradient-based metric (Q_{AB}^F), Piella’s edge-based metric (Q_e), and Zhao’s metric (P'_{blind}) measures the degree of fusion with respect to the source images. Each of the metrics mentioned above is briefly defined below. To maintain the generality, the source images are denoted by ‘A’, ‘B’ whereas ‘F’ denotes the fused image having a dimension of $M \times N$.

- 1) *Average pixel intensity (API)*: It serves as an index to measure the overall brightness in an image and is given by (12):

$$API = \frac{\sum_{i=1}^M \sum_{j=1}^N F(i, j)}{MN} \tag{12}$$

where $F(i, j)$ is the pixel intensity.

- 2) *Standard Deviation (SD)*: It is defined as the square root of variance and measures the spread of the data from the mean value (13).

$$SD = \sqrt{\frac{\sum_{i=1}^M \sum_{j=1}^N (F(i, j) - \text{line}F)^2}{MN}} \tag{13}$$

- 3) *Average Gradient (AG)*: It measures the degree of clarity and sharpness which is given by (14):

$$AG = \sqrt{\frac{(F(i, j) - F(i + 1, j))^2 + (F(i, j) - F(i, j + 1))^2}{MN}} \tag{14}$$

- 4) *Feature Mutual Information (FMI)* [9]: It is a non-reference objective image fusion metric proposed by Haghghat et al. which measures the transfer of features from the source image to the fused image. It is based on mutual information, and the uniqueness of the algorithm lies in the choice of gradient map as an information feature. A gradient map contains information about pixel neighborhoods, edge strength and directions, texture, contrast, and other region-based features. The authors have proved that a JPDF (joint probability distribution function) can be constructed with a given marginal probability distribution function (MPDF). The normalized values of the gradient magnitude at each pixel location in the feature images are used in marginal distributions. The amount of feature information transferred to F from A and B are individually measured by (15) and (16):

$$I_{FA} = \sum_{f,a} p_{FA} \log_2 \frac{p_{FA}}{p_F \cdot p_A} \tag{15}$$

$$I_{FB} = \sum_{f,b} p_{FB} \log_2 \frac{p_{FB}}{p_F \cdot p_B} \tag{16}$$

The FMI metric is expressed as:

$$FMI_F^{AB} = I_{FA} + I_{FB} \tag{17}$$

The normalized FMI is formulated as:

$$FMI_F^{AB} = \frac{I_{FA}}{H_F + H_A} + \frac{I_{FB}}{H_F + H_B} \tag{18}$$

where H_A , H_B and H_F are histogram based entropies of the images A,B and F respectively. It lies in the interval of [0,1].

- 5) *Gradient-Based fusion metric (Q_{AB}^F)* [28]: Xydeas and Petrovic proposed a pixel-level fusion metric which measures the amount of transfer of edge information from source

images to the fused image. It employs the Sobel edge detector and calculates the edge strength and orientation at each edge pixel. This metric is widely used in analyzing the edge strength and quality of fusion results. It is mathematically expressed in (19):

$$Q_{AB}^F = \frac{\sum_{m=1}^M \sum_{n=1}^N Q_{m,n}^{AF} w_{m,n}^{AF} + Q_{m,n}^{BF} w_{m,n}^{BF}}{\sum_{m=1}^M \sum_{n=1}^N w_{m,n}^{AF} + w_{m,n}^{BF}} \tag{19}$$

where $Q_{m,n}^{AF}$ and $Q_{m,n}^{BF}$ are edge preservation values weighted by $w_{m,n}^{AF}$ and $w_{m,n}^{BF}$ at coordinate (x,y) , respectively. The value of this metric lies within $[0,1]$ where larger value indicates better performance.

- 6) *Piella’s metric (Q_E)* [22]: This metric proposed by Piella and Heijmans quantifies the transfer of salient information from the source images to the fused images. Three different indices, namely, fusion quality index (Q_O), weighted fusion quality index (Q_W), and edge dependent fusion quality index (Q_E) are evaluated separately. Keeping up with the context of this paper, only Q_E is evaluated for this algorithm because it measures the transfer of edges in the fused results. The mathematical representation of this metric is given in (20):

$$Q_E(A, B, F) = Q_W(A, B, F) \cdot Q_W(A', B', F')^\alpha \tag{20}$$

where $\alpha \in [0, 1]$. The parameter α denotes the relative contribution of the edge images by the original images. The mathematical derivation of all the indices are elaborated in the original paper.

- 7) *Zhao’s Metric (P'_{blind})* [41]: Zhao et al. have relied on image features based on phase congruency and principal moments to design a pixel level quantitative fusion metric. It is defined as the product of three separate correlation coefficients.

$$P'_{blind} = (P_p)^\alpha (P_M)^\beta (P_m)^\gamma \tag{21}$$

where p, M, m denotes the phase congruency, maximum and minimum moments respectively. α, β, γ are the tunable parameters used in the algorithm. For local analysis, a block-based approach is adopted, where the final value of P'_{blind} is obtained by averaging all the values computed over the number of blocks of the entire image.

$$P'_{blind} = \frac{1}{K} \sum_{k=1}^K P'_{blind}(k) \tag{22}$$

The above mentioned metrics have been evaluated on sets of colored and gray-scale multi-focus image pairs as shown in Fig. 9. The values of API, SD, AG are provided in Table 6 and $FMI_F^{AB}, Q_{AB}^F, Q_E$ and P'_{blind} for all the images are presented in Table 5.

4.4 Subjective evaluation and fusion results

A good fusion algorithm should be able to produce accurate and reliable results by simultaneously removing the redundant information and integrating the complementary information. Besides, it should not introduce/enhance extra features, artifacts, or inconsistencies which are not a part of the source images. Below, we present a comparative discussion on the results produced by the proposed method and the other six methods. Some of the results are zoomed separately so as to illustrate the quality of results produced by the algorithm (See Figs. 10, 11, 12, 13 and 14). The quality of the source images determines the quality of the fusion results. All the input source images are assumed to be pre-registered.



Fig. 9 Sets of multi-focus images used in experiment

As evident from Fig. 10a and b, the features are not identically focused in both the pairs. The results of the proposed fusion method, along with other methods for the first source image, are presented in Fig. 10(c-i). It may be observed that the fused image obtained in GD based method (Fig. 10c) produces an image with highly increased contrast and brightness, which is an undesirable effect and modifies the intensity values taken from the source images. In addition, it creates a strong shadow around lighter objects within the scene (Fig. 10(c1)). In DWT-AB based approach, there is unnecessary distortion and distribution of color components in the vicinity. In Fig. 10d, the blue color gets unevenly spread around the text ‘Nature’ and in Fig. 10(d1), colors get introduced along the fold in the cloth within the leg, which was not present earlier. Figure 10e presents the results of IM based approach, which clearly shows the lack of clarity along the edges of the leg (Fig. 10(e1)). Figure 10f presents the results of ASR based method which also casts shadows around lighter objects (Fig. 10(f1)) similar to GD based method. In GCF based method, the details present in the stone leg are not captured in the fused image (Fig. 10g). The results obtained from the CNN (Fig. 10h) as well as the proposed method (Fig. 10i) are quite comparable. The respective local magnified regions are presented in Fig. 10(h1) and (i1). The proposed method has produced better results with no noticeable distortion and artifacts in less amount of time.

The second experiment has been performed on the ‘Seascape’ source images, as shown in Fig. 11a and b. Originally, there is a color disparity in the input images, the reasons for which could be many but not known to us in particular. This has in fact affected the fusion results in all the methods except ours. Results from GD, DWT-AB, and ASR based method are shown in Fig. 11c, d and f, respectively, clearly shows that there is a gross change/distortion in color after the sea surface is brought to focus. It confirms the fact that these fusion algorithms have taken color values from the foreground object in the far-focussed image, which is not

Table 5 Performance evaluation using fusion quality metrics

Images	Methods							Proposed
	Metrics	GD	DWT-AB	IM	ASR	GCF	CNN	
Leg (378×378)	Q^F_{AB}	0.7278	0.6992	0.7667	0.7226	0.7440	0.7667	0.7710
	FMI	0.9110	0.9166	0.9182	0.9181	0.9185	0.9191	0.9172
	Q_e	0.9982	0.9984	0.9992	0.9992	0.9994	0.9993	0.9989
	P'_{blind}	0.7645	0.7588	0.8296	0.7806	0.6793	0.8402	0.8251
Seascape (414×414)	Q^F_{AB}	0.6500	0.4862	0.8007	0.6682	0.6582	0.7012	0.7021
	FMI	0.8046	0.7767	0.8307	0.8161	0.8345	0.8322	0.8292
	Q_e	0.9931	0.9868	0.9939	0.9940	0.9939	0.9936	0.9937
	P'_{blind}	0.6730	0.4879	0.8361	0.7129	0.8496	0.8421	0.8331
Calendar (225×162)	Q^F_{AB}	0.6831	0.6223	0.7258	0.6789	0.7239	0.7127	0.7302
	FMI	0.8947	0.8941	0.9016	0.8990	0.9012	0.9020	0.8977
	Q_e	0.9969	0.9983	0.9979	0.9989	0.9990	0.9991	0.9976
	P'_{blind}	0.8049	0.8035	0.8897	0.8592	0.8250	0.8965	0.8835
Jar (512×512)	Q^F_{AB}	0.7285	0.7726	0.7863	0.7566	0.7732	0.7856	0.7864
	FMI	0.9259	0.9305	0.9306	0.9297	0.9296	0.9308	0.9294
	Q_e	0.9975	0.9989	0.9998	0.9995	0.9997	0.9998	0.9999
	P'_{blind}	0.6877	0.8075	0.8011	0.7870	0.6777	0.8056	0.8083
Temple (480×480)	Q^F_{AB}	0.7372	0.7259	0.7529	0.7099	0.7619	0.7578	0.7622
	FMI	0.8184	0.8216	0.8202	0.8198	0.8224	0.8221	0.8199
	Q_e	0.9978	0.9978	0.9974	0.9984	0.9986	0.9985	0.9985
	P'_{blind}	0.7207	0.7237	0.7843	0.6616	0.7810	0.7857	0.7879
Hill (400×400)	Q^F_{AB}	0.6846	0.7190	0.7401	0.7322	0.7396	0.7418	0.7377
	FMI	0.8434	0.8653	0.8670	0.8658	0.8671	0.8670	0.8672
	Q_e	0.9926	0.9964	0.9968	0.9970	0.9963	0.9965	0.9969
	P'_{blind}	0.8158	0.8551	0.8633	0.8646	0.8701	0.8736	0.8582
Dog (700×400)	Q^F_{AB}	0.6585	0.6727	0.6957	0.6922	0.6753	0.6960	0.6960
	FMI	0.8883	0.8971	0.8970	0.8983	0.8995	0.8998	0.9019
	Q_e	0.9892	0.9957	0.9955	0.9959	0.9954	0.9956	0.9956
	P'_{blind}	0.7668	0.8305	0.8386	0.8437	0.7761	0.8534	0.8575

Table 5 (continued)

Images	Methods Metrics	GD	DWT-AB	IM	ASR	GCF	CNN	Proposed
Pot (700×300)	Q_{AB}^F	0.6735	0.6717	0.6924	0.6985	0.6877	0.7133	0.7134
	FMI	0.8983	0.9004	0.9079	0.9063	0.9046	0.9073	0.9065
	Q_e	0.9939	0.9974	0.9977	0.9978	0.9978	0.9979	0.9976
	P_{blind}	0.6970	0.7959	0.7280	0.7797	0.6431	0.8431	0.8451
Child (520×520)	Q_{AB}^F	0.6933	0.7106	0.7353	0.7158	0.7242	0.7391	0.7333
	FMI	0.8736	0.8814	0.8808	0.8797	0.8808	0.8819	0.8810
	Q_e	0.9920	0.9985	0.9994	0.9991	0.9990	0.9991	0.9986
	P'_{blind}	0.7784	0.8361	0.8530	0.8366	0.8180	0.8488	0.8466
Car (500×500)	Q_{AB}^F	0.8017	0.7596	0.8456	0.8279	0.8335	0.8480	0.8360
	FMI	0.8275	0.8258	0.8323	0.8315	0.8323	0.8325	0.8327
	Q_e	0.9970	0.9976	0.9997	0.9997	0.9971	0.9977	0.9985
	P'_{blind}	0.8248	0.7966	0.8470	0.8286	0.8422	0.8530	0.8314
Clock (512×512)	Q_{AB}^F	0.6859	0.6987	0.7170	0.6899	0.7106	0.7170	0.7172
	FMI	0.9150	0.9235	0.9256	0.9212	0.9216	0.9255	0.9215
	Q_e	0.9984	0.9987	0.9984	0.9988	0.9987	0.9987	0.9989
	P_{blind}	0.6674	0.7803	0.7826	0.7411	0.7221	0.8169	0.7573

The bold values indicates the highest value obtained in the metric

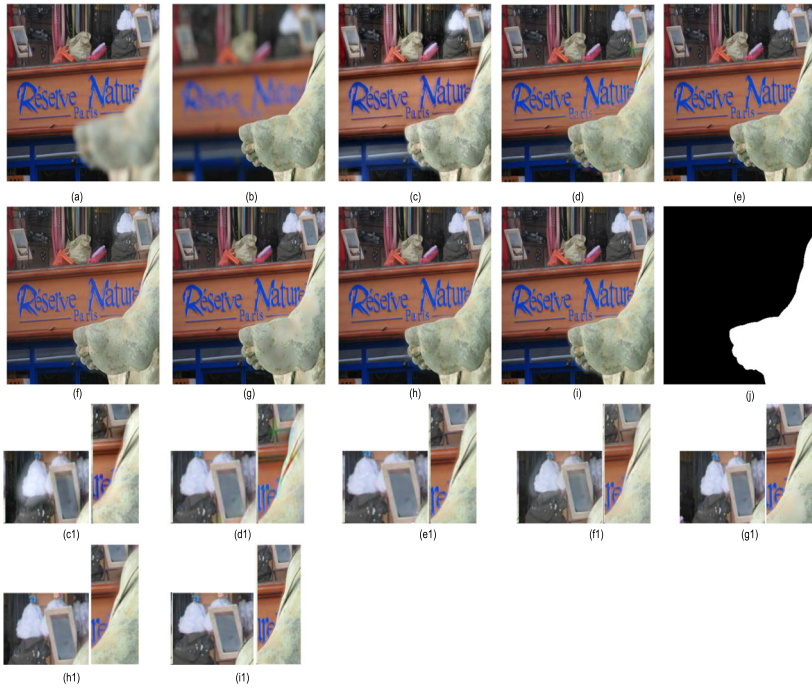


Fig. 10 Source image “Leg” and fusion results: **a** Source image with focus on the background; **b** Second source image with focus on foreground; **c** GD based result; **d** DWT-AB result; **e** IM based result; **f** ASR based result; **g** GCF based result; **h** CNN based result; **i** Result using proposed method; **j** Region image obtained by our method; (c1), (d1), (e1), (f1), (g1), (h1) and (i1) are the local magnified regions of **c**, **d**, **e**, **f**, **g**, **h** and **i** respectively

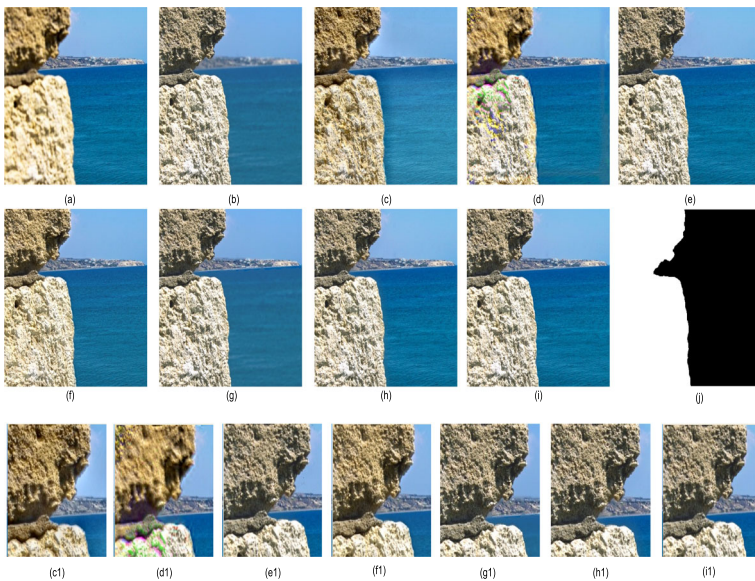


Fig. 11 Source image “Seascape” and fusion results: Same order as in Fig. 10



Fig. 12 Source image “Calender” and fusion results: Same order as in Fig. 10



Fig. 13 Source image “Jar” and fusion results: Same order as in Fig. 10

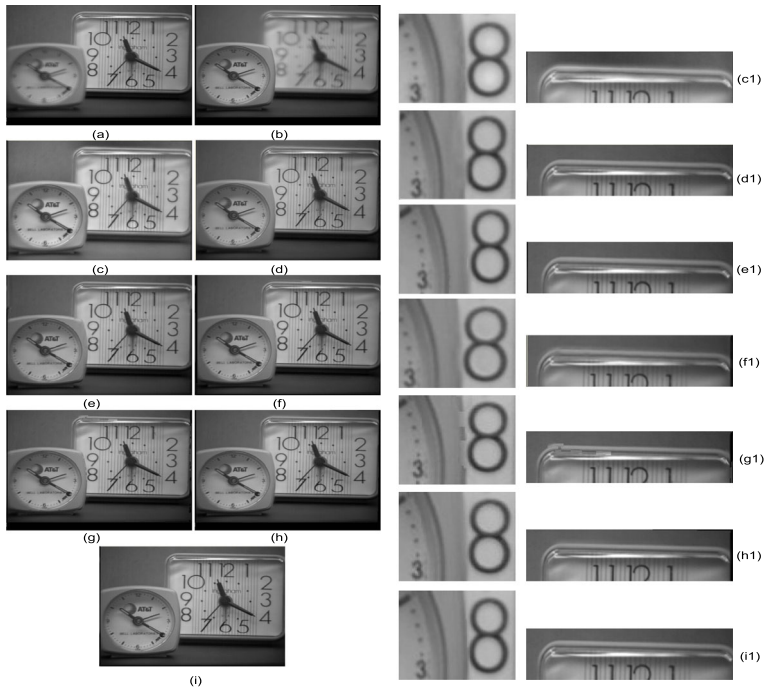


Fig. 14 Source image “Clock” and fusion results: Same order as in Fig. 10

desirable. Moreover, in the magnified regions (Fig. 11(c1 and d1)), the edges of the rock do not appear to be prominent. The magnified regions of IM based approach (Fig. 11(e1)) shows the presence of unnecessary pixels along the edges of the upper portion of the rock. The GCF based method has fused the blurred version of the sea from the near focussed image instead of capturing the details present in the far focussed source image (Fig. 11(g)). Results from CNN based method lacks clarity along the curvature of the rock, as shown in Fig. 11(h, h1). The results of the proposed algorithm and its magnified images presented in Fig. 11(i and i1), respectively, have outdone the other methods in terms of visual clarity. It has perfectly combined the complementary features from the source images with adequate edge clarity, no color distortion, shadow or blurriness.

In ‘Calendar’ source image, one of the source image is focused on the book (Fig. 12(a)) with the blue background while the other image has its focus on the table calendar (Fig. 12(b)). GD based result has led to an increase in the brightness as well as contrast, thereby creating strong shadows around the text (Fig. 12(c1)). DWT-AB based results show a slight spillover of blue color along the edges of the table calendar (Fig. 12(d1)). The results of IM based technique shown in Fig. 12(e1) are satisfactory in terms of visual quality. Results using ASR have introduced shadow around the text, ‘IMAGE’ in the book, and the top-left corner of the table calendar (Fig. 12(f1)). GCF based method has also produced artifacts over the same text (Fig. 12(g1)). Taking into account all the defects mentioned above, the CNN based method (Fig. 12(h, h1)), and the proposed algorithm has produced better results. (Fig. 12(i, i1)).

The multi-focus image pair presented in Fig. 13(a and b) consist of blue jar focused at foreground and orange jar focused at background, respectively. The results obtained by

different fusion methods are presented in Fig. 13c-i. For a minute comparison, the magnified images are presented in Fig. 13(c1)-(i1). From the results, it is perceived that the GD based method has intensified the color and brightness and introduced shadows in the results (Fig. 13(c1)). DWT-AB results are slightly distorted in terms of color around the text ‘Flora’ in the orange jar (Fig. 13(d1)). Likewise, the GCF method has produced box effects, which is clearly visible in Fig. 13(g, g1). IM based result is shown in Fig. 13(e1) has some white noise/dots spread over the bottom portion of the blue jar (Fig. 13(e1)). This is certainly a drawback of the algorithm and could be attributed to ill measures of focus. Apart from the presence of shadows around the dark text/objects, ASR based results are quite encouraging (Fig. 13(f1)). The results from the CNN algorithm presented in Fig. 13(h, h1) and the proposed method in Fig. 13(i, i1) are quite satisfactory.

For gray-scale images, the visual comparison has been demonstrated using only one of the six images used in this experiment. The gray-scale image, ‘Clock’, consists of two clocks of varying sizes where the focus is on the larger and smaller clock, respectively (Fig. 14a and b). The observations regarding the results obtained from various methods are quite similar to that of the color images. The results from the different fusion methods are provided in Fig. 14c-i. GD based method suffers from a similar problem as discussed previously

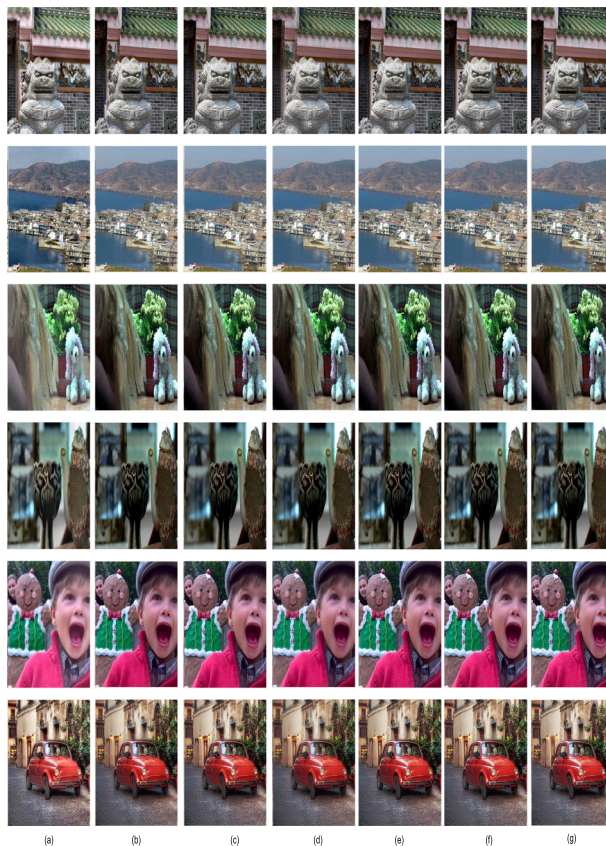


Fig. 15 Results for other source images: **a** using GD; **b** using DWT-AB; **c** using IM; **d** using ASR; **e** using GCF; **f** using CNN; **g** using proposed method

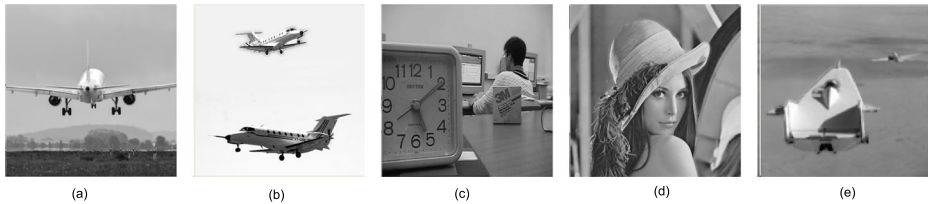


Fig. 16 Results for other gray-scale images using the proposed method

(Fig. 14(c1)). In DWT-AB based method, the edge of the smaller clock lacks sharpness at the region of overlap over the larger clock (Fig. 14(d1)). The edges are comparatively better expressed in IM based method, but the edge of the smaller clock slightly bends along the number ‘8’ of the larger clock (Fig. 14(e1)) which is more prominent in the results obtained using ASR (Fig. 14(f1)). The GCF based result (Fig. 14(g1)) and CNN based result (Fig. 14(h1)) clearly produces distortion and lacks clarity along the edges respectively. The results of the proposed method presented in Fig. 14(i1) shows the superiority in terms of quality the fused results.

The results on rest of the color and gray-scale source image pairs are provided in Figs. 15 and 16 respectively. Summarizing the observation in the obtained results, we can say the gradient domain (GD) based method enhances the brightness as well as the contrast for all the images, which is also established by the metric values from Table 6.

It also creates shadows around light objects in the dark background and vice-versa. DWT-AB based method leads to incorrect scattering and spread of color components from adjacent pixels. For certain color images, the focus measure used in image matting (IM) based method fails to distinguish the focused and defocused pixel, thereby creating slight noise along the edges. The GCF based method introduces block effects and fails to distinguish between focussed and defocussed pixels. Again, ASR based approach suffers from similar problem i.e., shadows and construction of dictionaries increases the computational complexity as well. The results from the CNN approach and the proposed method are quite comparable in terms of fusion quality; however training a deep neural network requires a significant amount of computation time and processing speed (Table 7).

4.5 Results using other source images

The proposed approach has been tested using source images obtained under different (a) complex background environment, (b) illumination conditions, and (c) time to perform rigorous analysis of the algorithm. Figure 17 illustrates the results obtained after fusing source images that satisfy the above criteria.

Table 6 Average values of image quality metrics

Metrics	GD	DWT-AB	IM	ASR	GCF	CNN	Proposed
API	0.4700	0.4087	0.4078	0.0492	0.3509	0.4705	0.4079
SD	0.2153	0.2087	0.2101	0.2078	0.2093	0.2109	0.2104
AG	0.0556	0.0453	0.0522	0.0509	0.0518	0.0519	0.0523

The bold value indicates the highest value obtained in the metric

Table 7 Average execution time of algorithms (in seconds)

GD	DWT-AB	IM	ASR	GCF	CNN	Proposed
1.1514	28.6434	4.4746	1048.838	550.3413	198.9899	1.7305

The bold value indicates the maximum time taken by the algorithm

4.6 Fusion with multiple source images

Generally, to restore focus within a scene, we may need to deal with multiple partially focussed source images. Though the algorithm demonstrates the two focus situation, but it can be extended to process multiple source images by fusing them one by one. This is illustrated by Fig. 18 which contains three source images.

4.7 Results on noisy source images

To study the performance of the proposed algorithm in the presence of noise, experiments are performed on noisy image pairs generated by adding noise externally to the multi-focus source images. The degree of noise degradation is an important factor exceeding which highly reduces the quality and the focus information/content from the source images, thereby producing results of less practical value. The source image pairs are corrupted with two types of noise, a) Salt-and-pepper and b) Gaussian white noise prior to the application of the proposed fusion algorithm. In case of salt-and-pepper, the maximum noise density (d) that is acceptable to retain the focus content is 0.2, i.e., 20% of pixels. For Gaussian white noise, keeping the variance constant, the acceptable mean (m) is found out to be 0.50. Fig. 19 illustrates the results produced by the proposed algorithm in presence of noise.



Fig. 17 1st Row, 2nd Row: Image datasets with different (a1),(a2) complex background; (b1),(b2) artificial lighting; c night time; d poor lighting; 3rd Row: Results obtained by proposed method



Fig. 18 Results obtained by the proposed algorithm for multiple source images, **a** Focus on front seal; **b** Focus on middle seal; **c** Focus on distant scenery; **d** Fused result of **(a)** and **(b)**; **e** Fused result of **(c)** and **(d)**

4.8 Limitations of the algorithm

In case the source images chosen for fusion happen to contain weak edges, the proposed algorithm does not perform well. The source image pair provided in Fig. 20 consists of a sea beach focused on the left and right side, respectively. The proposed algorithm fails to work on such source images due to (a) absence of strong and prominent edge content, (b) absence of prominent structures/object. This is illustrated from the respective edge images depicted alongside where the sky does not produce any strong edges due to smooth regions. Also, we cannot differentiate between the edge pixels belonging to objects and pixels due to noise. As the edge detectors are unable to locate the focused edge pixels primarily, features depending solely on spatial intensity values of an image are not sufficient to use in such cases. In addition to this, the proposed algorithm cannot deal with multi-modal or multi-sensor image sets because features expressed in such images are partly complementary and partly redundant.



Fig. 19 Results on noisy multi-focus source pairs, **a** and **b** Salt-and-pepper, $d = 0.2$; **c** and **d** Gaussian white noise, $m = 0.50$; **(a1)**,**(b1)**,**(c1)**,**(d1)** Results produced by the proposed approach

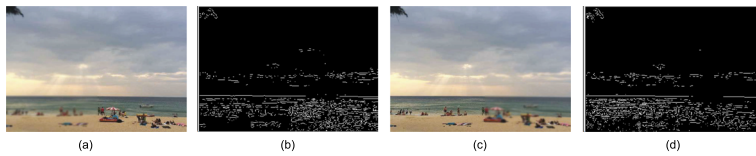


Fig. 20 **a** Focus on the right; **b** Edge image of (a); **c** Focus on left; **d** Edge image of (c)

4.9 Future works

- Edge detectors used in the algorithm are based on brightness gradients alone, which may give strong responses for irrelevant regions and weak for the relevant ones, thus producing improper edge maps to start with. Moreover, it may not work with natural, real-life images. Deep learning methods can be used to devise edge detector models, which will directly respond to the relevant focussed edges, which can automate the process to some extent.
- For real-time application, we can explore the possibility of FPGA (Field Programmable Gate Arrays) implementation, which would further reduce the overall execution time.

5 Conclusion

This paper presents a multi-focus image fusion method which uses edges of focused features from the source images as a basis for selecting the focused regions prior to constructing the fusion result. A block-wise region comparison followed by morphological dilation operation is performed to enhance the clarity of the focused edges further. Morphological edge by reconstruction is performed to restore and reconstruct any broken edge to maintain the continuity. The best reconstructed source image is chosen based on a focus measure, which in turn is used in constructing a binary region (initial decision map). For color images, the binary region is further converted to a colored decision map for the ease of fusion procedure. It is to be noted that the decision map is formed by using the best reconstructed edge image out of the two source edge images. To form the fused image, the binary (or colored decision map) is used to combine pixels from the gray-scale (or color) source images. The method has been compared with other similar methods, both in terms of quantitative and qualitative evaluation. It is observed that the visual quality of the outputs is superior in comparison to the results produced by other methods. The values of the fusion metrics obtained as a part of quantitative evaluation is as good as the other state-of-the-art methods.

References

1. Aishwarya N, Bennila Thangammal C (2017) An image fusion framework using novel dictionary based sparse representation. *Multimed Tools Appl* 76(21):21869–21888
2. Ardeshir Goshtasby A, Nikolov S (2007) Guest editorial: Image fusion: Advances in the state of the art. *Inf Fusion* 8(2):114–118
3. Canny J (1987) A computational approach to edge detection. In: *Readings in computer vision*. Elsevier, pp 184–203
4. Du C, Gao S (2017) Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. *IEEE Access* 5:15750–15761
5. Du J, Li G, Sun Y, Kong J, Bo T (2019) Gesture recognition based on skeletonization algorithm and cnn with asl database. *Multimed Tools Appl* 78(21):29953–29970

6. Duan J, Meng G, Xiang S, Pan C (2014) Multifocus image fusion via focus segmentation and region reconstruction. *Neurocomputing* 140:193–209
7. Geng P, Huang M, Liu S, Feng J, Bao P (2016) Multifocus image fusion method of Ripplet transform based on cycle spinning. *Multimed Tools Appl* 75(17):10583–10593
8. Gonzalez RC, Woods RE (1977) *Digital Image Processing* 28(4)
9. Haghghat MBA, Aghagolzadeh A, Seyedarabi H (2011) A non-reference image fusion metric based on mutual information of image features. *Comput Electr Eng* 37(5):744–756
10. Huang W, Jing Z (2007) Evaluation of focus measures in multi-focus image fusion. *Pattern Recogn Lett* 28(4):493–500
11. Huang W, Jing Z (2007) Multi-focus image fusion using pulse coupled neural network. *Pattern Recogn Lett* 28(9):1123–1132
12. Li S, Yang B, Hu J (2011) Performance comparison of different multi-resolution transforms for image fusion. *Inf Fusion* 12(2):74–84
13. Li S, Kang X, Hu J (2013) Image fusion with guided filtering. *IEEE Trans Image process* 22(7):2864–2875
14. Li S, Kang X, Hu J, Yang B (2013) Image matting for fusion of multi-focus images in dynamic scenes. *Inf Fusion* 14(2):147–162
15. Li S, Kang X, Fang L, Hu J, Yin H (2017) Pixel-level image fusion: a survey of the state of the art. *Inf Fusion* 33:100–112
16. Li G, Zhang L, Sun Y, Kong J (2019) Towards the semg hand: internet of things sensors and haptic feedback application. *Multimed Tools Appl* 78(21):29765–29782
17. Luo X, Zhang J, Dai Q (2012) A regional image fusion based on similarity characteristics. *Signal Process* 92(5):1268–1280
18. Ma R, Zhang L, Li G, Jiang D, Xu S, Chen D (2020) Grasping force prediction based on semg signals. *Alexandria Engineering Journal*
19. Meher B, Agrawal S, Panda R, Abraham A (2019) A survey on region based image fusion methods. *Inf Fusion* 48:119–132
20. Mitianoudis N, Stathaki T (2007) Pixel-based and region-based image fusion schemes using ICA bases. *Inf Fusion* 8(2):131–142
21. Paul S, Sevcenco IS, Agathoklis P (2016) Multi-exposure and multi-focus image fusion in gradient domain. *J Circ Syst Comput* 25(10):1650123
22. Piella G, Heijmans H (2003) A new quality metric for image fusion. In: *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, vol 3. IEEE, pp III–173
23. Sun Y, Xu C, Li G, Xu W, Kong J, Jiang D, Tao B, Chen D (2020) Intelligent human computer interaction based on non redundant emg signal. *Alexandria Engineering Journal*
24. Tan W, Zhou H, Rong S, Qian K, Yu Y (2018) Fusion of multi-focus images via a gaussian curvature filter and synthetic focusing degree criterion. *Appl Opt* 57(35):10092–10101
25. Tian J, Li C, Ma L, Yu W (2011) Multi-focus image fusion using a bilateral gradient-based sharpness criterion. *Opt Commun* 284(1):80–87
26. Vincent L (1993) Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE Trans Image Process* 2(2):176–201
27. Wan T, Zhu C, Qin Z (2013) Multifocus image fusion based on robust principal component analysis. *Pattern Recogn Lett* 34(9):1001–1008
28. Xydeas CS, Petrovic V (2000) Objective image fusion performance measure. *Electron Lett* 36(4):308–309
29. Yajie W, Ye Y, Xiaoyan R, Wu Y, Xiangbin S (2015) A multi-focus color image fusion method based on edge detection. In: *The 27th Chinese Control and Decision Conference (2015 CCDC)*. IEEE, pp 4294–4299
30. Yao Y, Roui-Abidi B, Doggaz N, Abidi MA (2006) Evaluation of sharpness measures and search algorithms for the auto focusing of high-magnification images. In *Visual Information Processing*
31. Yu L, Chen X, Hu P, Wang Z (2017) Multi-focus image fusion with a deep convolutional neural network. *Inf Fusion* 36:191–207
32. Yu L, Wang Z (2013) Multi-focus image fusion based on wavelet transform and adaptive block. *J Image Graph* 18(11):1435–1444
33. Yu L, Wang Z (2014) Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Process* 9(5):347–357
34. Yu L, Chen X, Ward RK, Jane Wang Z (2016) Image fusion with convolutional sparse representation. *IEEE Signal Process Lett* 23(12):1882–1886
35. Yu Z, Bai X, Wang T (2017) Boundary finding based multi-focus image fusion through multi-scale morphological focus-measure. *Inf Fusion* 35:81–101

36. Yu L, Chen X, Wang Z, Jane Wang Z, Ward RK, Wang X (2018) Deep learning for pixel-level image fusion Recent advances and future prospects. *Inf Fusion* 42:158–173
37. Yu L, Chen X, Ward RK, Jane Wang Z (2019) Medical image fusion via convolutional sparsity based morphological component analysis. *IEEE Signal Process Lett* 26(3):485–489
38. Zhan K, Xie Y, Wang H, Min Y (2017) Fast filtering image fusion. *J Electron Imaging* 26(6):063004
39. Zhang B, Lu X, Pei H, He L, Zhao Y, Zhou W (2016) Multi-focus image fusion algorithm based on focused region extraction. *Neurocomputing* 174:733–748
40. Zhang Q, Liu Y, Blum RS, Han J, Tao D (2018) Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Inf Fusion* 40:57–75
41. Zhao J, Laganieri R, Liu Z (2007) Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement. *Int J Innov Comput Inf Control* 3(6):1433–1447

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Manali Roy is pursuing her PhD degree in Image fusion under Dr. Susanta Mukhopadhyay, Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology (ISM), Dhanbad, India. She has completed her B-tech from West Bengal University of Technology and M-tech from IIT(ISM), Dhanbad in 2014 and 2017 respectively.



Susanta Mukhopadhyay did his B.Sc. with Honours in Physics from Presidency College, Calcutta, B.Tech and M.Tech in Radiophysics and Electronics from the University of Calcutta and Ph.D. in Image Processing from the Indian Statistical Institute, Calcutta in 1988, 1992 and 2003 respectively. During 2001–2003 and 2004–2007 he worked at the Sanford Burnham Medical Research Institute, La Jolla, California, USA as post-doctoral researcher and Nanyang Technological University, Singapore, as research fellow respectively. His research area and interest include image processing, fMRI, image compression, image encryption and image watermarking. He has published articles in internationally reputed journals like *IEEE Transactions on Image Processing*, *Pattern Recognition*, *Signal Processing* etc. He is currently working as Associate Professor in the Department of Computer Science and Engineering, Indian Institute of Technology (ISM), Dhanbad, India.