Multi-focus Fusion Using Image Matting and Geometric Mean of DCT-Variance

Manali Roy[(✉)] and Susanta Mukhopadhyay

Indian Institute of Technology (ISM), Dhanbad, India

Abstract. Optical lenses installed in most imaging devices suffer from the limited depth of field due to which objects get imaged with varying sharpness and details, thereby losing essential information. To cope with the problem, an effective multi-focus fusion technique is proposed in this paper based on a focus measure obtained from the statistical properties of an image and its DCT-bandpass filtered versions. The focus information obtained is eventually converted into a decision trimap using fundamental image processing operations. This trimap largely discriminates between the focussed and defocussed pixels and expedites the fusion process by acting as an input to a robust image matting framework to obtain a final decision map. Experimental results illustrate the superiority of the proposed approach over other competing multi-focus methods in terms of visual quality and fusion metrics.

Keywords: Multifocus fusion · DCT · Geometric mean · Focus measure · Robust image matting

1 Introduction

Multi focus image fusion happens to be a widely accepted solution to solve the defocus problem for digital images where objects at different focal depths appear to be out-of-focus. It aims to integrate several partially focused images of a similar scene captured under different focal settings to produce an all-in-focus image [5]. The homogeneously focussed fused image contains more information compared to the source images. Fusion approaches are mainly performed in these domains, i.e., a) Spatial, b) Transform, and c) Hybrid. Unlike the transform domain (which tends to introduce brightness or contrast distortion in the result), the spatial domain (pixel-based and region-based) methods exploit the spatial consistency between the pixel values for better focus analysis. The performance of conventional window-based spatial methods depends on a fixed block size, which often generates blocking artifacts on object boundaries due to block-based computation of focus measure. Moreover, the fixed block size may not produce equally good results for all the source images. Region-based methods are introduced to solve the above problem, which splits/segments the image into focussed and defocused regions. However, the accuracy of the segmentation greatly influences the final result. To achieve a perfect segmentation, optimization approaches are adopted to generate several image matting algorithms [14].

Multi-focus fusion algorithms are being developed, employing image matting techniques to refine the focussed region maps [2,4]. In this paper, a multi-focus fusion algorithm is proposed where a matting algorithm based on color sampling is used to segment the background from the foreground. The focussed and defocussed regions are roughly obtained by using a focus measure derived from statistical properties lying within the bandpass filtered versions of an image. The primary highlights of the paper are:

– The focus measure is derived using geometric mean from bandpass filtered versions of the source image.
– The focus information is incorporated into an image matting model to obtain accurate weight decision maps.
– The algorithm performs equally well for both registered, unregistered and artificial multi-focus image pairs.

The course of the paper is organized as follows: Sect. 2.1 briefly introduces the preliminary concepts used in the proposed work. Section 3 outlines the proposed algorithm in detail. Experimental results and discussions are presented in Sect. 4. Lastly, the paper is concluded in Sect. 5.

2 Preliminaries

2.1 Image Matting in Multi-focus Fusion

For natural images, the majority of the pixels either belong to the definite foreground or definite background. However, pixels lying on the junction separating the foreground and background cannot be assigned completely to either of them. The image matting technique is used for accurate separation of the foreground from the background using the foreground opacity as a parameter. In this framework, an image $I(x,y)$ can be viewed as a linear combination of foreground $I_f(x,y)$ and background $I_b(x,y)$ as presented in Eq. 1.

$$I(x,y) = \alpha(x,y)I_f(x,y) + (1-\alpha)(x,y)I_b(x,y) \qquad (1)$$

where $\alpha(x,y)$ is the foreground opacity or alpha matte and lies between [0,1]. Constraining $\alpha = 0$ or 1 reduces the matting problem into a classic problem of binary segmentation. Matting method deals with appropriate estimation of alpha values (α) for indefinite or mixed pixels (lying at the intersection). Equation 1 is clearly an under-constrained problem as it contains three unknown variables i.e., α, I_f and I_b to be determined from single input image, I. So, a user needs to provide a 3-pixel image known as 'trimap', which clearly divides the input image into three regions: definite foreground, definite background, and unknown regions. This trimap fastens the process of solving the unknown variables to obtain the alpha matte. The alpha matte obtained in Fig. 1(c) uses the trimap in Fig. 1(b) to separate the foreground from the background.
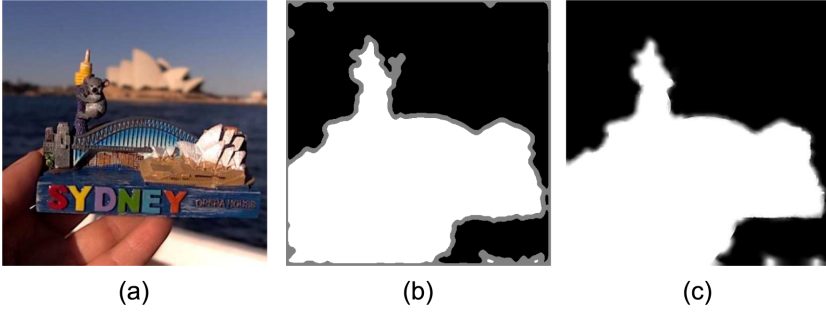
Fig. 1. (a) Source image (Focus on Foreground); (b) Trimap of (a); (c) Alpha Matte;

2.2 Variance as Focus Measure

Operators based on grayscale pixel intensity statistics such as variance, standard deviation, and other variants are widely used in extracting focus information from images [7]. Statistics based operators take advantage of meaningful patterns existing within the images. A well-focused image is expected to have high variance due to a wide-spread edge or non-edge like responses, whereas for blurry images, the value of variance is lower. Theoretically, variance effectively gives us an idea about the spread of pixel intensity values about the mean value and mathematically defined as,

$$\sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n - 1} \tag{2}$$

3 Proposed Algorithm

This section contains detailed description of the proposed scheme divided into three subsections, a) focus map generation, b) trimap formation and c) fusion based on image matting. The process flowchart is pictorially given in Fig. 3.

3.1 Generation of Focus Maps

The focus information from the source images is extracted using an efficient in-focus measure, which exploits the consistent statistical properties exhibited by the source image along with its bandpass filtered versions. It is to be noted that instead of using a single bandpass filter, a bank of filters is applied to retain the whole of data in a different format without loss of any information. To obtain the bank of filters, block-based (8×8) DCT is chosen due to a) excellent energy compaction properties, b) reduced computation cost, and c) good estimator of image sharpness. The 2-D block based DCT transform of an $N \times N$ block of an image, $I(x, y)$ is given by,

$$F(u, v) = \frac{2}{N} C(u) C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \cos \frac{(2x+1)u\pi}{2N} \cos \frac{(2y+1)v\pi}{2N} I(x, y) \tag{3}$$

where $u, v = 0, 1, 2, \ldots N - 1$.

$$C(u) = \begin{cases} \frac{1}{\sqrt{2}}, & u = 0 \\ 1, & otherwise \end{cases}, \quad C(v) = \begin{cases} \frac{1}{\sqrt{2}}, & v = 0 \\ 1, & otherwise \end{cases}$$

Eq. 3 gives N^2 (i.e., 64) number of basis images (Fig. 2b). Excluding the top-left basis image (i.e., F(0,0) representing mean intensity, marked by the red cross), the rest of the $N^2 - 1$ basis images constitute the bank of bandpass filters (bf_m where $m = 1 \ldots N^2 - 1$, marked by green square) which retains frequencies lying within the desired range. A set of bandpass filtered images, I_m where $m = 1 \ldots N^2 - 1$ is obtained after filtering the source image with each of the filters, thereby producing 63 feature images in total (Fig. 2c,d,e). As stated earlier, defocus regions have smaller values of variance (σ^2), but for a blurry patch with a complex pattern, a large value of variance causes it to be falsely marked as focussed. To avoid such outliers, variance calculated using Eq. 3 for all the feature images are combined using geometric mean. Theoretically, the geometric mean is not affected by data fluctuations, gives more weight to smaller values, and is easy to compute. Mathematically, the geometric mean of variances is defined as follows,

$$g(x) = \sqrt[n]{\prod_{n=1}^{N^2-1} \sigma_n^2(x)} \tag{4}$$

Now, for a pair of source images (say I_i and I_j), the corresponding focus maps (fm_i and fm_j) are obtained by carrying out simple difference operation between their geometric variance maps as given below,

$$fm_i = g_i(x) - g_j(x), \quad fm_j = g_j(x) - g_i(x) \tag{5}$$

where $g_i(x)$ and $g_j(x)$ are the variance maps obtained from Eq. 4 for i^{th} and j^{th} source image respectively.

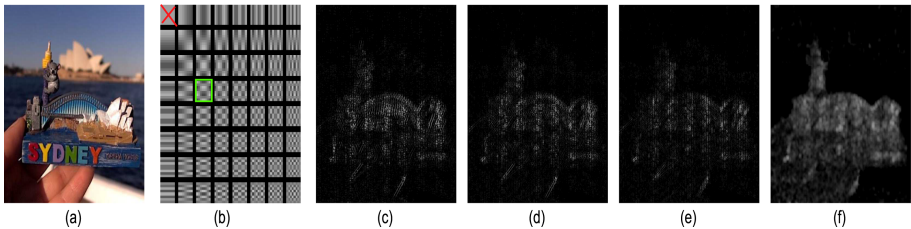(a)   (b)   (c)   (d)   (e)   (f)

Fig. 2. (a) Source image (Focus on Foreground); (b) DCT basis functions; (c),(d),(e) Filtered versions of (a) using (b) (Contrast-corrected for better visualization in the manuscript); (f) Focus Map (f_i)

3.2   Trimap Generation

After obtaining the focus information using Eq. 5, this section outlines the steps involved in trimap generation. The individual focus maps are binarized in two stages, a) the first stage of binarization involves comparing each pixel value with the global maximum intensity value using Eq. 6 and b) the second stage of binarization is performed using a threshold parameter T to obtain the definite focussed pixels as expressed in Eq. 7.

$$f_b^1 = \begin{cases} 1, & fm > \max\{fm\} \\ 0, & otherwise \end{cases} \quad (6) \qquad f_b^2 = \begin{cases} 1, & fm - \max\{fm\} > T \\ 0, & otherwise \end{cases} \quad (7)$$

Equation 6 gives a coarse yet believable focussed binary region which requires further iterations of processing to remove isolated pixels, small holes generated within the focussed region and fragments of regions caused due to image noise. Therefore, further processing of the binary map obtained at Eq. 6 is carried out by a hole filling operation which fills up existing holes within the definite region followed by application of window based median filter to remove isolated pixels and fragmented regions. The next stage of processing involves iterative skeletonization followed by a final median filtering to effectively remove the scattered pieces. So, mathematically the entire processing equation can be presented as,

$$(f_b^1)' = Med[Skel\{Med(f_b^1, w), i\}] \tag{8}$$

where w and i refers to the sliding window used for median filter and iterations of skeletonization respectively. From Eq. 7 and Eq. 8, the definite focussed region of a source image can be defined as,

$$f_d(x, y) = \begin{cases} 1, & (f_b^1)' = 1 \text{ or } f_b^2 = 1 \\ 0, & otherwise \end{cases} \tag{9}$$

The trimap is generated as follows,

$$t_{map} = \begin{cases} 1, & \text{if } f_d = 1 \text{ and } max\{f_d\} = 0 \\ 0, & \text{if } f_d = 0 \text{ and } max\{f_d\} = 1 \\ 0.5 & otherwise \end{cases} \tag{10}$$

where $t_{map} = 0$ or $t_{map} = 1$ signifies definite focus or defocus and $t_{map} = 0.5$ denote unknown regions.

3.3   Fusion Using Matting

The final step of the proposed work is to generate the alpha matte(α) for the source images using the trimap generated in the previous section. To obtain

appropriate alpha values for indefinite pixels, a matting based approach based on color sampling is adopted [13]. The matting algorithm estimates the initial value of alpha for unknown pixels using 'good' sample pairs of foreground and background pixels from the neighborhood. The good samples can express the color of the unknown pixel as a convex combination of themselves. To select the 'best' sample pair from the probable candidates, a '*distance ratio*' is defined, which associates each pair with a confidence value. The initial matte is further refined by minimizing the following energy function by solving a graph labeling problem as a random walk.

$$E = \sum_{z \in I} [\hat{f}_z(\alpha_z - \hat{\alpha_z})^2 + (1 - \hat{f}_z)(\alpha_z - \delta(\hat{\alpha_z} > 0.5))^2] + \lambda \cdot J(\alpha, a, b) \qquad (11)$$

Here $\hat{\alpha_z}$ and $\hat{f}_z$ are the initial alpha and confidence value estimated at the sampling step respectively, δ stands for boolean function and $J(\alpha, a, b)$ denotes the additional neighbourhood energy term around 3×3 pixels for further improvement. Minimizing the energy term, $J(\alpha, a, b)$ means finding suitable values for constants α, a and b for which the energy term is optimized.

Now, let α_f be the final alpha matte obtained after solving Eq. 11, the fused image can be represented as linear combination of the source images with respect to the alpha matte (α_f).

$$I_f = \alpha_f I_i(x, y) + (1 - \alpha_f)I_j(x, y) \qquad (12)$$

4   Experimental Results and Discussion

This section discusses and compares the experimental performance of the proposed approach in terms of objective and subjective evaluation with other fusion approaches.

4.1   Experimental Setup

Table 1. Experimental Parameters

Name	Value
Median filter (Window size, w)	8×8
Skeletonization (Iterations, i)	5
Binarization (Threshold, T)	128

The experimental setup adopted for testing the proposed approach includes datasets, execution environment, fusion metrics and methods for visual comparison. The algorithm is tested using two multi-focus datasets, a) Lytro Dataset
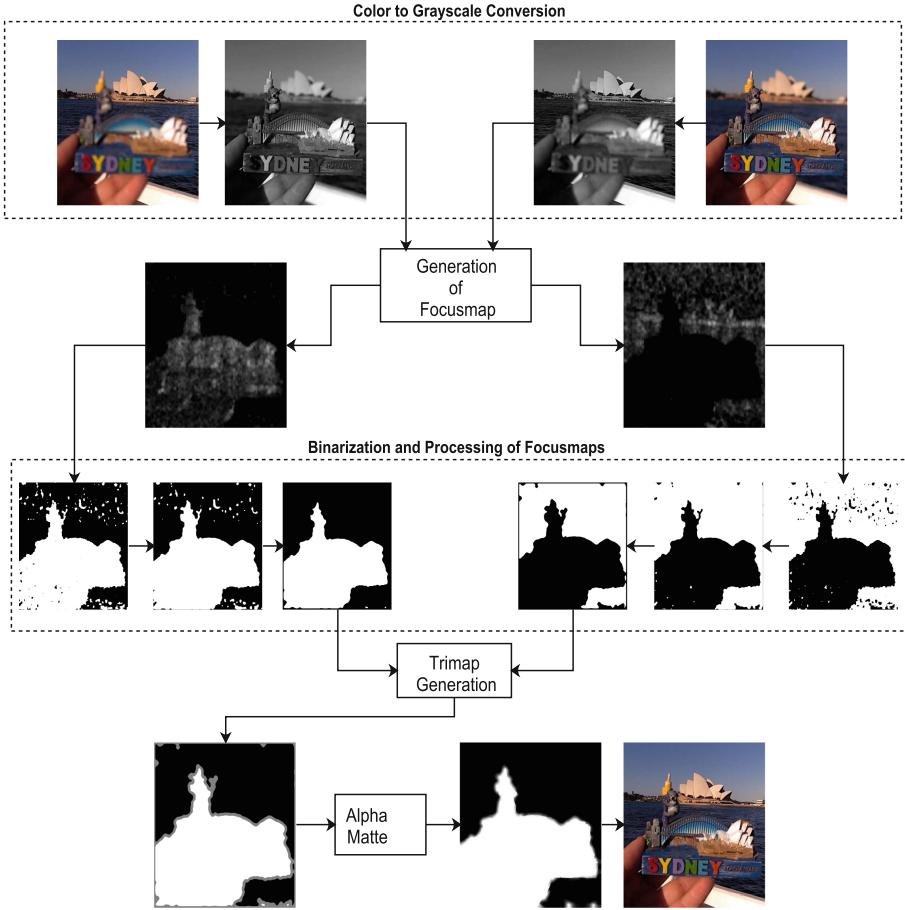
Fig. 3. Process flowchart for the proposed approach

[8] consisting of registered multifocus pairs and b) Pxleyes Dataset [1] containing unregistered multi-focus image pairs submitted as a part of photography contest. The experiments are performed over MATLAB R2015a, installed over 64-bit windows platform having Intel 2.30 GHz Core i5 CPU and 4GB RAM. The algorithm has been quantitatively compared with four equivalent fusion methods, GD [9], DWT-AB [6], MWGF [17] and GCF [12] using the following fusion metrics, mutual information (MI_{AB}^{F}) [11], Piella's metric (Q_o) [10], feature mutual information (FMI) [3], Xydeas's metric (Q_{AB}^{F}) [15] and Zhao's metric (P'_{blind}) [16]. The parameters used for the experiment and average value for individual metrics for both the datasets are presented in Table 1 and Table 2 respectively.

4.2 Subjective Evaluation

A good fusion algorithm should not introduce or enhance extra features, artifacts, or inconsistencies in the fused result. In this section, the visual quality of the results from the proposed method is compared with other standard fusion algorithms to study their relative performances. Figure 4 and Fig. 5 presents the results for registered source pairs from the Lytro dataset, whereas Fig. 6 and Fig. 7 portrays the same using unregistered source pairs from the Pxleyes dataset. For both the datasets, GD based method produces fusion results with increased contrast and brightness (as marked by red boxes in Fig. 4c and Fig. 5c). Besides, it also blurs out the foreground (Fig. 6c) and creates visible shadows around light (dark) color objects against a darker (lighter) background. This distortion can be attributed to the poor wavelet-based reconstruction of fused coefficients in the gradient domain. Similarly, results from the DWT-AB method suffer from color distortion, as evident from random color spots scattered over the fused result (green-blue spots in Fig. 6d-red box and Fig. 7d-black box). GCF method extracts the salient regions using a gaussian curvature filter followed by a focus criterion that combines spatial frequency and local variance. The performance of the method is illustrated in Fig. 4e, 5e (red box) where the fused result has picked up pixels from the source image with defocussed foreground. For unregistered source pair (Fig. 6e-red box and Fig. 7e-black box), staircase/block effect can be observed along the boundary of the objects. For MWGF based method, the fusion performance for registered source pair is visually appealing without visible distortions (Fig. 4f, 5f, 7f), yet for some unregistered pair, it creates a blurred effect along the border (Fig. 6). Results from the proposed method (Fig. 4h, 5h, 6h, 5h) are best in terms of visual quality with no pixel/color distortion, maximum preservation of source image information. The superiority of the approach is further validated by the highest value of fusion metric obtained for individual datasets, as listed in Table 2.

Table 2. Average objective evaluation on multifocus image sets

| Images	Metrics	GD [9]	DWT-AB [6]	MWGF [17]	GCF [12]	Proposed
	Methods					
Lytro (512×512)	$MI_{\mathrm{AB}}^{\mathrm{F}}$	0.5322	1.0773	1.1872	1.1115	**1.2118**
	Q_o	0.8981	0.9401	0.9305	0.9399	**0.9435**
	FMI	0.8938	0.9050	0.9059	**0.9062**	0.9061
	$Q_{\mathrm{AB}}^{\mathrm{F}}$	0.7079	0.7230	0.7245	0.7309	**0.7340**
	P_{blind}'	0.7640	0.8290	0.7749	0.8441	**0.8513**
Pxyeles	$MI_{\mathrm{AB}}^{\mathrm{F}}$	0.3747	0.6325	0.8303	0.8929	**0.9579**
	Q_o	0.7910	0.8643	0.8586	0.8701	**0.8839**
	FMI	0.8812	0.8927	0.8897	0.9026	**0.9028**
	$Q_{\mathrm{AB}}^{\mathrm{F}}$	0.5911	0.6128	0.6808	0.7024	**0.7027**
	P_{blind}'	0.5191	0.5731	0.5629	0.7222	**0.7281**
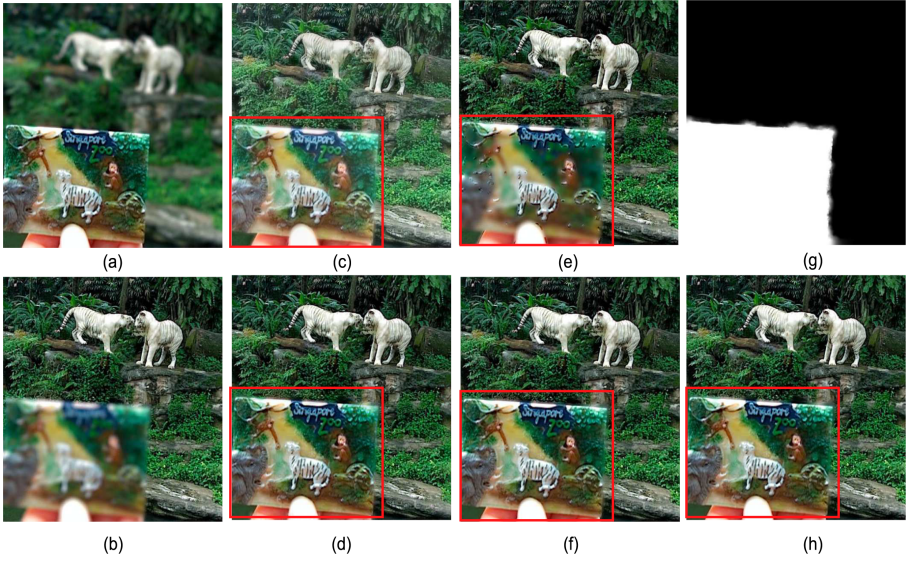
Fig. 4. Registered source image and fusion results: (a) Focus on the foreground; (b) Focus on background; (c) GD based result; (d) DWT-AB result; (e) GCF based result; (f) MWGF based result; (g) Generated Alpha Matte; (h) Result using proposed method
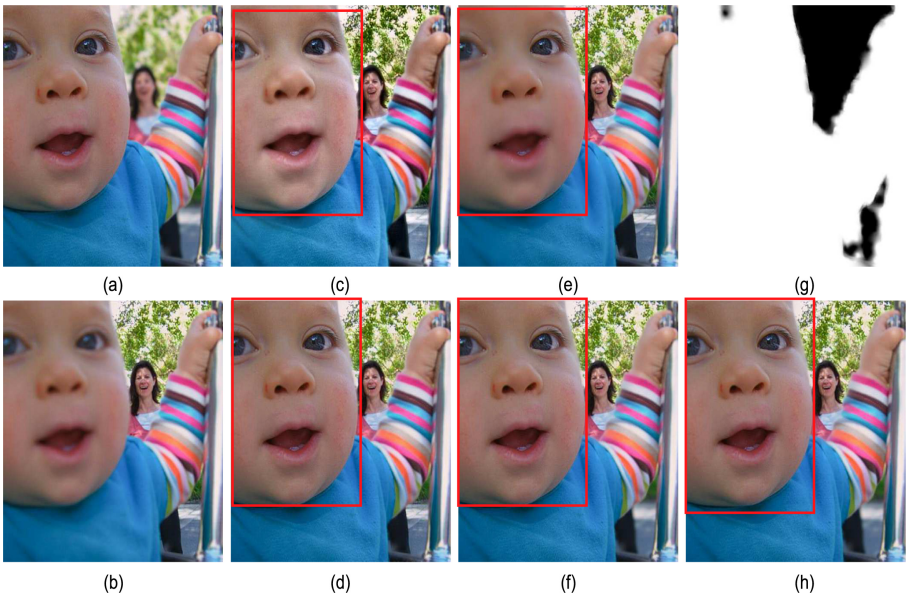

Fig. 5. Registered source image and fusion results: Same order as in Fig. 4

Fig. 6. Unregistered source image and fusion results: Same order as in Fig. 4

Fig. 7. Unregistered source image and fusion results: Same order as in Fig 4

4.3   Performance on Artificial Source Images

The performance of the algorithm for artificial multi-focus source images has been studied in this section. In case of artificial source images, availability of the groundtruth all-in-focus image helps us to compare the quality of fusion achieved by the proposed algorithm. Figure 8 presents the results for two artificial source images. Additionally, a reference based image quality metric, i.e., SSIM is calculated which gives us a reasonably higher accuracy with repsect to the ground truth image. Irrespective of the source images being real or artifical, the defocussed blurred patches around a pixel is expressed as gaussian convolution where the standard deviation (σ) is considered as the blur kernel. For a defocus region, the value of the square of standard deviation, i.e., variance will have smaller values in comparison to a focussed region. It uses statistical averaging on the variances of DCT filtered bandpass responses to measure the degree of focus.

Fig. 8. Performance on artificial images: (a),(a1) Focus on Foreground; (b),(b1) Focus on Background; (c),(c1) Fused by the proposed algorithm; (d),(d1) Groundtruth image

5   Conclusion

This paper proposes a multi-focus fusion algorithm that exploits first-order statistics to extract the salient regions from a partially focused image. The source images are subjected to a bank of bandpass filters to obtain multiple filtered versions, which is statistically combined using the geometric mean of variance. The focus map is gradually converted into a 3-pixel trimap, which acts as an input

to an image matting model to achieve perfect fusion results. The qualitative and quantitative efficacy of the proposed method is confirmed by performing experiments using suitable multi-focus datasets and fusion metrics.

References

1. www.pxleyes.com
2. Chen, Y., Guan, J., Cham, W.K.: Robust multi-focus image fusion using edge model and multi-matting. IEEE Trans. Image Process. **27**(3), 1526–1541 (2017)
3. Haghighat, M.B.A., Aghagolzadeh, A., Seyedarabi, H.: A non-reference image fusion metric based on mutual information of image features. Comput. Electric. Eng. **37**(5), 744–756 (2011)
4. Li, S., Kang, X., Hu, J., Yang, B.: Image matting for fusion of multi-focus images in dynamic scenes. Inf. Fusion **14**(2), 147–162 (2013)
5. Liu, Y., Wang, L., Cheng, J., Li, C., Chen, X.: Multi-focus image fusion: a survey of the state of the art. Information Fusion (2020)
6. Liu, Y., Wang, Z.: Multi-focus image fusion based on wavelet transform and adaptive block. J. Image Graph. **18**(11), 1435–1444 (2013)
7. Maruthi, R., Sankarasubramanian, K.: Multi focus image fusion technique in spatial domain using an image variance as a focus measure. i-Manager's J. Fut. Eng. Technol. **5**(3), 24 (2010)
8. Nejati, M., Samavi, S., Shirani, S.: Multi-focus image fusion using dictionary-based sparse representation. Inf. Fus. **25**, 72–84 (2015)
9. Paul, S., Sevcenco, I.S., Agathoklis, P.: Multi-exposure and multi-focus image fusion in gradient domain. J. Circ. Syst. Comput. **25**(10), 1650123 (2016)
10. Piella, G., Heijmans, H.: A new quality metric for image fusion. In: Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429). vol. 3, pp. III-173. IEEE (2003)
11. Qu, G., Zhang, D., Yan, P.: Information measure for performance of image fusion. Electronics lett. **38**(7), 313–315 (2002)
12. Tan, W., Zhou, H., Rong, S., Qian, K., Yu, Y.: Fusion of multi-focus images via a gaussian curvature filter and synthetic focusing degree criterion. Appl. Optics **57**(35), 10092–10101 (2018)
13. Wang, J., Cohen, M.F.: Optimized color sampling for robust matting. In: 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2007)
14. Wang, J., Cohen, M.F.: Image and video matting: a survey. Now Publishers Inc. (2008)
15. Xydeas, C., Petrovic, V.: Objective image fusion performance measure. Electronics Lett. **36**(4), 308–309 (2000)
16. Zhao, J., Laganiere, R., Liu, Z.: Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement. Int. J. Innov. Comput. Inf. Control **3**(6), 1433–1447 (2007)
17. Zhou, Z., Li, S., Wang, B.: Multi-scale weighted gradient-based fusion for multi-focus images. Inf. Fusion **20**, 60–72 (2014)