Pattern Recognition 98 (2020) 107065

Contents lists available at ScienceDirect

Pattern Recognition

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

3D shape reconstruction from multifocus image fusion using a multidirectional modified Laplacian operator

Tao Yan[a,b,*], Zhiguo Hu[a,b], Yuhua Qian[b], Zhiwei Qiao[a,b], Linyuan Zhang[c]

[a] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
[b] Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China
[c] Beijing Zhongchao Banknote Designing and Plate-making Co.,Ltd., Beijing 100070, China

ARTICLE INFO

Article history:
Received 3 February 2019
Revised 10 September 2019
Accepted 23 September 2019
Available online 24 September 2019

Keywords:
3D shape reconstruction
Image fusion
Shape-from-focus
Microscopic imaging
Nonsubsampled shearlet transform

ABSTRACT

Multifocus image fusion techniques primarily emphasize human vision and machine perception to evaluate an image, which often ignore depth information contained in the focus regions. In this paper, a novel 3D shape reconstruction algorithm based on nonsubsampled shearlet transform (NSST) microscopic multifocus image fusion method is proposed to mine 3D depth information from the fusion process. The shift-invariant property of NSST guarantees the spatial corresponding relationship between the image sequence and its high-frequency subbands. Since the high-frequency components of an image represent the focus level of the image, a new multidirectional modified Laplacian (MDML) as the focus measure maps the high-frequency subbands to images of various levels of depth. Next, the initial 3D reconstruction result is obtained by using an optimal level selection strategy based on the summation of the multiscale Laplace responses to exploit these depth maps. Finally, an iterative edge repair method is implemented to refine the reconstruction result. The experimental results show that the proposed method has better performance, especially when the source images have low-contrast regions.

1. Introduction

Digital imaging of real scenes is usually a two-dimensional (2D) discretization process of three-dimensional (3D) scenes. Since the depth information is lost in this process, mining 3D information from 2D images is crucial in computer vision. In recent years, various 3D information recovery techniques have been proposed, and they are roughly divided into three categories: multiview stereo vision, depth camera and monocular vision. Three-dimensional reconstruction based on multiview stereo vision tends to fuse features from different views of the same scene. Some representative examples are multiview information fusion [1] and convolutional neural network [2]. Depth cameras provide explicit 3D information that can significantly reduce the difficulty of depth information acquisition. Lately, some advanced techniques such as multitask learning [3,4] and 3D convolutional neural networks [5] have been introduced into 3D scene recognition. In addition, depth maps can be recovered by capturing one or several images with a monocular vision system. These methods are referred to as shape from x [6], where x denotes a crucial factor of the depth information acquisition, such as shading [7], contours [8] and motion [9]. The focus level of the image is another important element for depth information recovery, on which a passive method named shape from focus (SFF) [10] has been developed for 3D shape reconstruction from 2D images.

SFF methods obtain the depth map from a sequence of images captured by changing the focus setting of the single-view camera. Because various focus settings correspond to different depth of field, only a part of the area is focused during the imaging process of the object in the 3D scene, while the rest is defocused. If the image sequence is sufficient to cover the entire depth range in the 3D scene by adjusting the distance between the camera and the object, the depth image can be obtained by the position in the image sequence in which each pixel has the highest focus level. Most SFF methods are run by mapping image sequences through an appropriate focus measure (FM) function followed by a reconstruction scheme to obtain 3D reconstruction results. FM is a kind of function evaluating the focus level of each pixel in an image. Some FM functions, such as 3D Laplacian [11], curvelet [12] and optical-based [13] have been proposed in relevant studies. The initial depth map can be obtained from maximizing these focus measure values. After that, some reconstruction schemes, such as the Gaussian model [14] and genetic programming [15], will make use of the focus information in the image sequence to find the accurate depth map.

* Corresponding author.
E-mail address: hongyanyutian@126.com (T. Yan).

https://doi.org/10.1016/j.patcog.2019.107065
0031-3203/

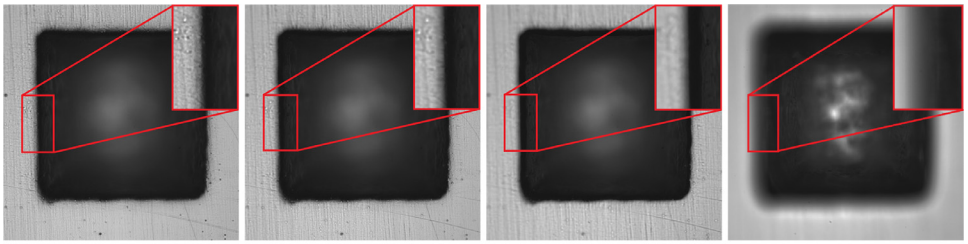


Fig. 1. Columns represent images with various levels of focus.

Optical microscopy is an important tool for high precision analysis and measurement of microscale objects due to its high magnification. However, as the resolution increases, the depth of field becomes shallower, leaving more regions out of focus, subsequently leading to inefficiencies when capturing the 3D structural information of the observed objects. Therefore, it is very necessary to develop a 3D reconstruction method based on optical microscopy. Multiview stereo vision has a complex image registration process, which may have higher complexity and increase the probability of incorrect reconstruction results in 3D reconstruction of optical microscopy [16]. Depth cameras are susceptible to the interference of ambient light in the process of microscopic imaging, leading to noise, occlusion and loss of depth information [17]. Fortunately, no special hardware assistance is needed in SFF. Low hardware complexity makes this algorithm easy to apply to 3D shape reconstruction in optical microscopy [18]. The SFF methods are undeniably efficient, but the following issues deserve further investigation.

(1) The effect of noises in the real scene on the reconstruction results. Typically, experiments examine the effectiveness of the SFF algorithms on simulated objects, and then these algorithms are transferred onto real cases. However, the reconstruction results always suffer from performance deterioration due to noises that are difficult to reproduce in a simulation. For example, a large number of highly reflective areas in the images are inevitably produced in the optical imaging process of reflective objects.
(2) The effect of weak contrast regions on reconstruction results. When imaging a concave object with a great depth, it is difficult for light to enter the inside of the object, resulting in low-contrast and low-texture areas. Fig. 1 shows image sequences of a concave object with different lens settings. It is clear that the details are difficult to detect and that the contrast declines in regions with small red rectangular windows due to the weak light. The traditional SFF methods might fail to recognize changes in focus in low-contrast areas, leading to a significant deviation in estimating the depth information [12].
(3) Additionally, in many application scenarios, for example in microscopy for printed circuit board defects detection, depth information of the object obtained by SFF methods cannot accurately judge the type of defect, and auxiliary gray information is needed as well. Unfortunately, capturing gray information from source images leads to fusion discreteness and then produces more discontinuous regions in fused images, ultimately resulting in low-quality fusion results.

Mitigation of the noises in the real scene will be achieved by an image acquisition device in Section 2. In this paper, for amending the shortcomings of (2) and (3), multifocus image fusion techniques are considered.

Like in SFF, the key step in multifocus image fusion is to select an effective focus measure [19]. Therefore, we expect multifocus image fusion algorithms to provide more potential solutions for SFF. There are two types of multifocus image fusion algorithms: spatial domain methods and transform domain methods. Spatial domain methods directly pick pixels, blocks or regions to construct fused images linearly or nonlinearly. Nevertheless, these methods rely heavily on the accuracy of the pixels, blocks and regions [20]. Therefore, when these spatial domain methods are applied to SFF, they might cause spatial distortions and ghosting artifacts, which result in shape inconsistency in reconstruction results. To overcome the above disadvantages of spatial domain fusion methods, transform domain fusion methods should be considered.

Transform domain fusion methods decompose images into high and low frequency coefficients and employ the fusion rules to handle these coefficients. In recent years, decomposition tools, such as wavelet transform [21] and discrete cosine harmonic wavelet transform [22], have been applied in multifocus image fusion. However, poor directionality and shift variance that lead to unsatisfactory solutions is a common deficiency in these tools. Nonsubsampled contourlet transform (NSCT) [23] is regarded as an expedient remedy to these problems, but it suffers from large computational burdens. The recently proposed nonsubsampled shearlet transform (NSST) overcomes the above mentioned shortcomings [24]. By avoiding downsampling operations, NSST has better performance in terms of shift invariance. It yields decomposition subimages all having the same size of source images and facilitates tracking the depth information during the decomposition process. In addition, compared to other transform domain methods, NSST not only has excellent properties such as anisotropy and multiscale but also provides better direction sensitivity to capture the intrinsic geometrical structure of images. The abundant detailed information obtained by using these properties can provide a more accurate basis for depth information evaluation. Therefore, it is feasible to capture depth information in the process of NSST-based multifocus image fusion.

The main contributions of this paper include three aspects. (1) This paper proposes a new framework for embedding SFF method into multifocus image fusion algorithm, in which the fused image and depth map of an object in the 3D scene can be obtained simultaneously. (2) A new multidirectional modified Laplacian (MDML) as a focus measure to realize the mapping from high-frequency subbands to depth maps is analyzed and discussed. (3) An iterative edge repair method that can automatically detect and repair the error areas in the depth maps is proposed, and in the end, this method can effectively improve reconstruction accuracy of the low contrast regions.

The structure of this paper is as follows. Section 2 describes the problems inherent in applying the SFF technique to the 3D reconstruction of microscales and designs an image capture device. Section 3 shows the proposed algorithm in detail, and

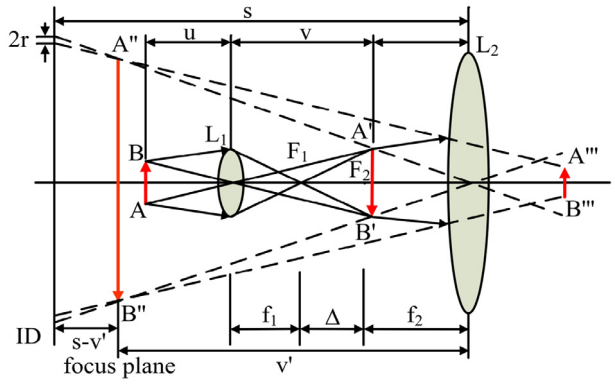


Fig. 2. Schematic illustration of microscopic imaging.

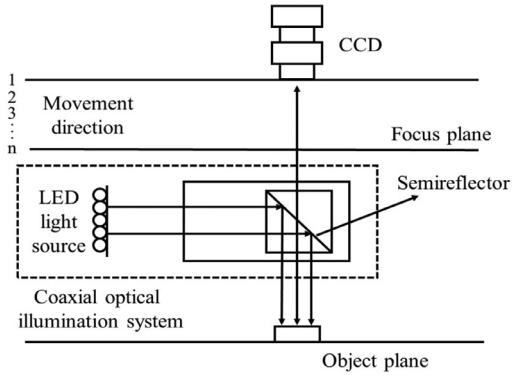


Fig. 3. Image acquisition device with the coaxial optical illumination system.

Section 4 presents several experimental settings, comparative analysis. Conclusions and future work are discussed in Section 5.

2. Background

To generate the depth image of an object, it is necessary to estimate the distance between every point of the object and the camera. Fig. 2 illustrates the schematic illustration of microscopic imaging, where f_1 and f_2 indicate the focal lengths of the lenses L_1 and L_2, u is the distance of the object from the lens, and v is the distance of the magnified image from the lens. The relationship between f_1, u and v is given by the Gaussian lens law: $1/f_1 = 1/u + 1/v$. The object AB is located near the front focus of the objective lens. A magnified image $A'B'$ is formed near the front focus of the eyepiece, which is in turn enlarged into the virtual image $A''B''$ as an object of the eyepiece.

If the image detector (ID) is placed at exact distance v', a well-focused image $A''B''$ will be formed. Otherwise, if the ID is at a distance $s - v'$ from the focus plane, point A'' will be projected onto a blurry circle of diameter $2r$. Therefore, only some parts of the image are in focus during imaging, whereas other parts are blurred.

Presume there is a microscale object with larger depth, it is impossible to have a focal plane to cover the entire depth of the object. Therefore, it is necessary to design an image acquisition device that can retain different focal points in an image sequence. The image sequence of a given object was obtained by a moving camera. As shown in Fig. 3, the top parts of the object are well focused when the focus plane is at level 1, whereas the other regions are blurred due to their distance far from the focus plane. The camera then moves downward until the bottom parts of the object are focused and the focus plane reaches the level n. The two reasons we avoid moving the object are: (a) the object's vibration when moving results in inaccuracy of focus information, and (b) for 3D reconstruction of small regions in large-format samples, the expense of moving the objects is huge. In addition, for metal object imaging in this paper, it is easy to produce a large number of exposure areas in the use of the traditional optical imaging method, and the loss of detailed information in these exposure areas will lead to inaccurate judgment of the focus level. Therefore, an optical microscopy imaging device based on a coaxial optical illumination system is designed in Fig. 3. This illumination system not only provides more uniform lighting than traditional optical illumination systems, but it also effectively overcomes the impact of object reflectivity.

3. The proposed approach

Most established multifocus image fusion methods provide better visual perception and quality. However, these methods usually use a large number of image processing techniques, which are not conducive to accurately extract the depth information of image sequences. In this paper, a new 3D shape reconstruction scheme is proposed to search the depth maps that represent the best-focused image during the image fusion process. Briefly, it takes three steps to implement the proposed method. First, we apply NSST to an image sequence to obtain the low-frequency subbands and high-frequency subbands. Then, the low-frequency fused coefficients are obtained by an averaging method, and the high-frequency fused coefficients are obtained by a new MDML fusion approach. The all-in-focus image is finally obtained by inverse NSST on these fused coefficients. Second, pixels in the same position of the image sequence that exhibit the highest value of MDML are mapped to various levels of depth maps. However, not all levels of depth maps are suitable to reconstruct the depth image. Then, the optimal level selection principle is used to choose the appropriate level for initial 3D reconstruction. Finally, in order to remedy the less accurate depth points, an iterative edge repair algorithm is implemented to obtain more accurate 3D reconstruction results. The real 3D reconstruction results can be obtained by joining fused images and 3D reconstruction results. The schematic diagram of the proposed method is given in Fig. 4.

3.1. Fusion rules

Fusion rules are an important part of NSST-based multifocus image fusion. Each source image can be decomposed into a low-frequency subband and a series of high-frequency subbands. In this paper, the high-frequency subbands have a key influence on the performance of the 3D reconstruction. Therefore, in order to reduce the computational cost, we employ the weighted-average based method on low frequency subbands. The fused low-frequency subbands $L_F(x, y)$ are given by the following:

$$L_F(x,y) = \frac{1}{n}\sum_{i=1}^{n} C_i(x,y) \tag{1}$$

where $C_i(x, y)$ is the coefficients of the position (x, y) in the i subband, and $L_F(x, y)$ represents the fused low-frequency subbands.

As a counterpart, significant coefficients of high-frequency subbands should be selected for fusion. To make full use of the strong correlation among adjacent pixels and to mine detail information from various directions accurately, we propose a new multidirectional modified Laplacian (MDML), in which the standard stencil of Laplacian is rotated on eight angles and the maximum value from all eight directions of the current position is chosen as the focus measure and high frequency fusion rule.

$$\mathrm{MDML}_i^{j,l} = \underset{(\xi,\eta)\in U(x,y))}{\mathrm{MAX}} \{\cos((n-1)\theta)L_x + \sin((n-1)\theta)L_y\}_{n=1}^{8}$$

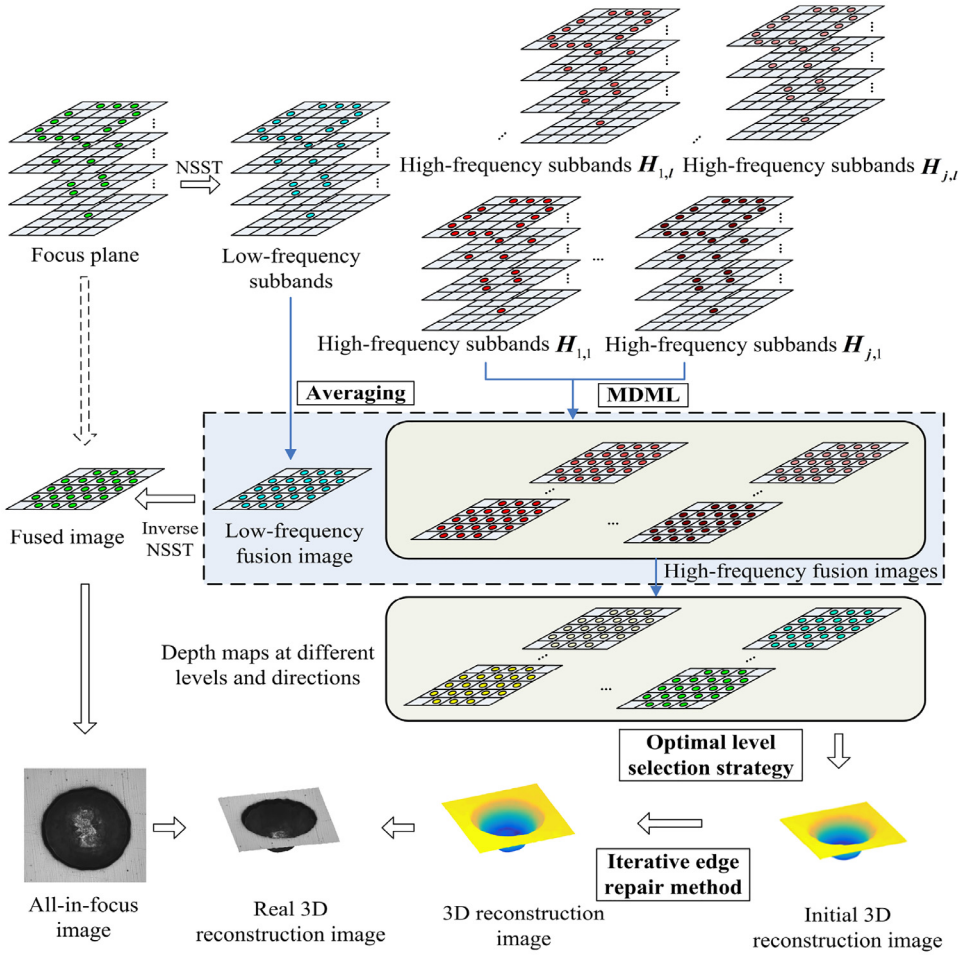


Fig. 4. Schematic diagram of the proposed 3D shape reconstruction approach.

$$
\begin{aligned}
L_x &= |C_i^{j,l}(\xi-2s,\eta)+4C_i^{j,l}(\xi-s,\eta)-10C_i^{j,l}(\xi,\eta)\\
&\quad +4C_i^{j,l}(\xi+s,\eta)+C_i^{j,l}(\xi+2s,\eta)|\\
L_y &= |C_i^{j,l}(\xi,\eta-2s)+4C_i^{j,l}(\xi,\eta-s)-10C_i^{j,l}(\xi,\eta)\\
&\quad +4C_i^{j,l}(\xi,\eta+s)+C_i^{j,l}(\xi,\eta+2s)|
\end{aligned} \tag{2}
$$

where $U(x, y)$ is a square pixel neighborhood of (x, y) and where the parameter s denotes the characteristic size of the square window. θ as the step of angle is set 22.5° in this paper. The coefficients with maximum MDML values are selected as the fused high-frequency coefficients:

$$
H_F^{j,l}(x,y)=C_k^{j,l}(x,y),\ k=\arg\max_{1\le i\le n}\mathrm{MDML}_i^{j,l}(x,y) \tag{3}
$$

where j and l denote the level and direction, respectively. Thus, the fused image is obtained by taking inverse NSST on the fused high-frequency subbands $H_F^{j,l}$ and fused low-frequency subbands L_F.

3.2. Relationship between image sequence and depth maps

In this section, we will find the relationship between the image sequence and the depth map using NSST. The process of NSST in image sequence decomposition consists of the following four steps.

First, consider an image sequence $\{f_i(x,y)\}_{x,y=0}^{M-1,M-1}, 1\le i\le N$ consists of N images with the size of $M\times M$. For the level j, perform the nonsubsampled Laplacian pyramid transform [23] on the approximation of the image sequence $\{f_{ai}^{j-1}(x,y)\}_{i=1}^{N}$ with a low-pass filter $h_{j-1}(1)$ and a highpass filter $h_{j-1}(0)$ to obtain their low frequency subbands and high frequency subbands. Where $\left(h_{j-1}(1),h_{j-1}(0)\right)$ are the pyramid filters for the 2D nonsubsampled filter bank:

$$
\begin{aligned}
\left\{f_{ai}^{j}(x,y)\right\}_{i=1}^{N} &= \left\{f_{ai}^{j-1}(x,y)*h_{j-1}(1)\right\}_{i=1}^{N}\\
\left\{f_{di}^{j}(x,y)\right\}_{i=1}^{N} &= \left\{f_{ai}^{j-1}(x,y)*h_{j-1}(0)\right\}_{i=1}^{N}
\end{aligned} \tag{4}
$$

where $*$ denotes the convolution operator. It is noticeable that both low frequency subband $f_{ai}^{j}(x,y)$ and high frequency subband $f_{di}^{j}(x,y)$ have the same image size $f_{di}^{j-1}(x,y)$, and the shift invariant is implemented by upsampling on the filters $\left\{h_{j-1}(0),h_{j-1}(1)\right\}$ instead of image downsampling. In particular, $f_{ai}^{0}(x,y)=f_i(x,y)$.

Second, compute the 2D discrete Fourier transform of high frequency subband on a pseudo-polar grid [25]. This gives by the matrix $Pf_{di}^{j}(x,y)=[\hat{f}_1(k_1,k_2),\hat{f}_2(k_1,k_2)]^T$, where $\hat{f}_1(k_1,k_2),\hat{f}_2(k_1,k_2)$ are given by the following:

$$
\begin{aligned}
\hat{f}_1(k_1,k_2) &= \sum_{x=-M/2}^{M/2-1}\sum_{x=-M/2}^{M/2-1} f_{di}^{j}(x,y)e^{-ix\frac{\pi k_1}{M}}e^{-iy\frac{\pi k_1}{M}\frac{2k_2}{M}}\\
\hat{f}_2(k_1,k_2) &= \sum_{x=-M/2}^{M/2-1}\sum_{x=-M/2}^{M/2-1} f_{di}^{j}(x,y)e^{-iy\frac{\pi k_2}{M}}e^{-ix\frac{\pi k_2}{M}\frac{2k_1}{M}}
\end{aligned} \tag{5}
$$

Next, apply a bandpass filtering to the $Pf_{di}^{j}(x,y)$, let $w_{j,l}(y)$ be the sequence whose discrete Fourier transform gives by window function $\overline{W(2^jk-l)}$, we have the following:

$$
Pf_{di}^{j}(x,y)\overline{W(2^jk-l)}=Pf_{di}^{j}(x,y)\mathcal{F}_1(w_{j,l}(y)) \tag{6}
$$

where $\mathcal{F}_1$ is the one-dimensional discrete Fourier transform, and l, k are direction and shift parameters.

Finally, the nonsubsampled shearlet coefficients [26] can be obtained by the following:

$$f_{di}^{j,l}(x,y) = \mathcal{F}^{-1}\left(Pf_{di}^{j}(x,y) \cdot \mathcal{F}_1(w_{j,l}(y))\right) \tag{7}$$

where $\mathcal{F}^{-1}$ represent the inverse pseudo-polar discrete Fourier transform.

As a consequence, each pixel (x, y) of the image corresponds to a low frequency subband and to a sequence of high frequency subbands of various levels and directions by NSST decomposition:

$$\{f_i(x,y)\}_{i=1}^{N} \underset{\text{NSST}}{\Rightarrow} \left\{ f_{ai}^{J}(x,y), \left\{ f_{di}^{j,l}(x,y) \right\}_{j,l=1}^{J,L} \right\}_{i=1}^{N} \tag{8}$$

where J and L are the maximum values of the level j and direction l.

From the above analysis, each image in image sequence generates corresponding high frequency subbands. Researchers have figured out that changes in focus mostly affect the high frequency components of an image [12]. Therefore, it is reasonable to establish the relationship between high frequency subbands and depth maps via a mapping function, and depth maps can be deduced by maximizing the mapping function values as below:

$$D_{j,l}(x,y) = \underset{1 \le i \le N}{\arg\max}(\mathrm{M}(f_{di}^{j,l}(x,y))) \tag{9}$$

where $D_{j,l}(x, y)$ denotes the best focus frame number of the point (x, y) at the level j and direction l. $\mathrm{M}(\cdot)$ is a mapping function that can be regarded as a focus measure in SFF. This mapping function can also be used as a fusion rule in high frequency subbands as shown in Section 3.1.

Finally, the depth map at a certain level is obtained by merging depth maps of different directions:

$$D_j(x,y) = \mathrm{Merge}\left(\left\{ D_{j,l}(x,y) \right\}_{l=1}^{L} \right) \tag{10}$$

where the $\mathrm{Merge}(\cdot)$ represents the operators such as averaging or choosing max methods.

However, some levels are not significant in a representation, and their removal has no real impact on the performances of the representation of the depth map. In the next section, we explain how the optimal depth levels are chosen.

3.3. Optimal level selection strategy

A larger scale results in blurring of the local details in the image, whereas a smaller scale cannot reflect the structural relationships in the image [27]. It is very important to choose the optimal level so that the 3D reconstruction result reflects the properties of the real object. To achieve this goal, first, 3D reconstruction results with different levels are obtained, and then, a search for the suitable level is conducted using different scales.

A Butterworth filter (a maximally flat filter) is a typical scaling function of NSST [26]. The two-dimensional Butterworth filter function is defined by the following:

$$H(x,y,s_0) = \frac{1}{1 + [s_0/s(x,y)]^2} \tag{11}$$

where s_0 denotes the scale parameter and $s(x, y)$ represents the distance between coordinates $s(x, y)$ and image origin. The Laplace transformation of the formula is expressed as follows:

$$\nabla^2 H = \frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2} \tag{12}$$

To remove the influence of scale parameters on the characteristic response, a normalized Butterworth Laplace transform is given by the following:

$$G(s_0) = \nabla^2_{\text{norm}} H = s_0^2 \left(\frac{\partial^2 H}{\partial x^2} + \frac{\partial^2 H}{\partial y^2} \right) \tag{13}$$

thus

$$G(s_0) = \frac{4s_0^2(s_0^2 - x^2 - y^2)}{(x^2 + y^2 + s_0^2)^3} \tag{14}$$

When the structure in the image is consistent with the shape of the Butterworth Laplace function, the Laplace response of the image reaches the maximum. Therefore, we can get the extreme value points for the normalized Butterworth Laplace function as follows:

$$\frac{\partial G(s_0)}{\partial s_0} = 0 \Rightarrow x^2 + y^2 = 2s_0^2 \tag{15}$$

we assume that $x^2 + y^2 = r^2$. Hence, when $s_0 = r/\sqrt{2}$, the structure with r radius in the image will reach the peak of the response of Laplace at scale s_0.

Reliable reconstruction results take into account the information from most of the structures in the object. In this paper, we apply the maximum method to merge the depth maps of all directions on the same level into a single depth map at first. Then, by calculating the total response of Laplace of different levels, we obtain depth images in the scale range of $s_{\min}$ to $s_{\max}$. The level L_{optimal} that gets the total of the maximum value is selected as the initial 3D reconstruction result.

$$L_{\text{optimal}} = \underset{1 \le j \le J}{\arg\max} \sum_{s=s_{\min}}^{s_{\max}} G_j(s) \tag{16}$$

3.4. The iterative edge repair method

To make the reconstruction results more accurate and reliable, an iterative edge repair algorithm to gradually export substitute pixels from the neighboring correct pixels is proposed in this section. A generally accepted hypothesis is that error regions tend to be in the regions with intense changes in the depth image. Hence, first the Sobel operator is used to pick up the edge of the error regions. Then, the Otsu method is used to calculate the initial binary threshold T_{otsu}, and the minimum percentage $p_{\min}$ of the edge detection results I_{sobel} is obtained by Eq. (17) as follows:

$$p_{\min} = \frac{1}{MN} \sum_{k=0}^{T_{\text{otsu}}} n_k \tag{17}$$

where M and N are height and width of the image I_{sobel}, and n_k is the number of pixels with a gray level of k.

Next, increment percentage p_{step} is calculated by maximum percentage $p_{\max}$ and minimum percentage $p_{\min}$ as shown in Eq. (18), where IN is the maximum number of the iterations. To ensure that the repaired region gradually increases as the iteration process, the maximum percentage $p_{\max}$ is set to 0.99.

$$p_{\text{step}} = \frac{1}{IN}(p_{\max} - p_{\min}) \tag{18}$$

The binary threshold of the current iteration T_{current} for generating the inpainting region is searched by Eq. (19) as follows:

$$\sum_{k=0}^{T_{\text{current}}} n_k > (p_{\min} + p_{\text{step}})MN \tag{19}$$

Finally, the repaired image of the current iteration is obtained by the fast marching method (FMM) [28]. It should be noted that the repaired image serves as the input image to the next iteration until the maximum iterations are reached. This method uniqueness is that erroneous regions can be repaired gradually through a binary method based on percentage. Fig. 5 shows a schematic diagram of the proposed algorithm.

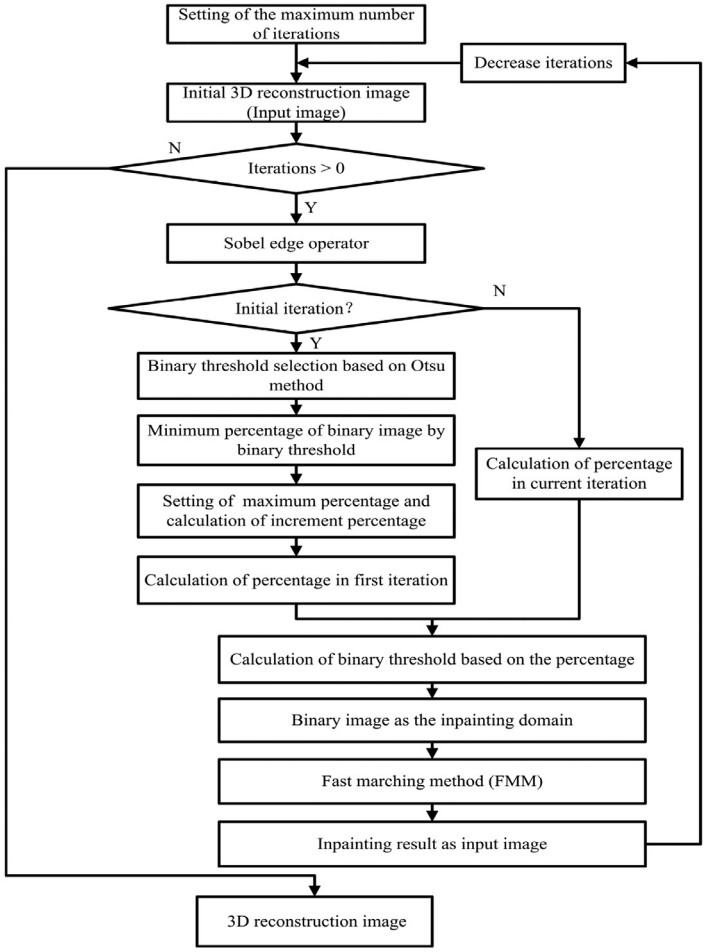


Fig. 5. Iterative edge repair algorithm for postprocessing of 3D shape reconstruction.

4. Experiments

4.1. Datasets

Both simulated and real objects are used to verify the effectiveness of the proposed 3D shape reconstruction method. The first simulated sequence consists of 100 images of 360×360 pixels, and the corresponding image generation algorithm of this sequence can be found in the previous study [16]. There are three other image sequences with respect to microscopic objects with varying textures and structures. The second and third sequences, which have concave structures, come from the intaglio plate, and the fourth sequence comes directly from the surface of a coin. These three sequences, respectively, contain 100 images at varying focus planes of the microscopy objects. The size of each image is 1024×1024 pixels, and the resolution is about 0.25 μm. The vertical interval distance between two consecutive image frames is approximately 1 μm. A portion of the test image sequences is shown in Fig. 6(a)–(d) represent various image sequences of a simulated object and three microscopy objects.

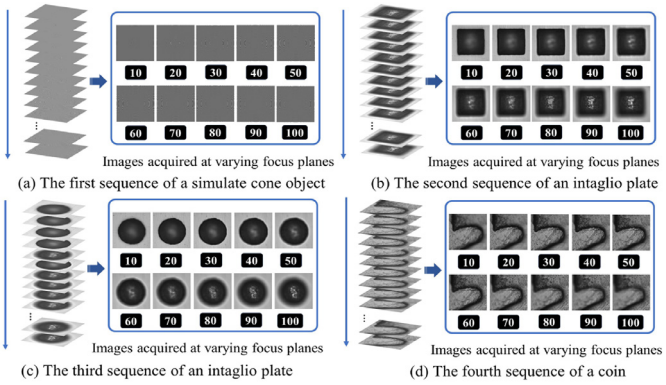


Fig. 6. A portion of the test image sequences.

4.2. Parameter setting

To choose the optimal level for analyzing the 3D shape reconstruction, the maximum decomposition level of NSST is set from 3 to 6 on the simulated object, and from 3 to 7 on the real objects. The number of directions from coarser scales to finer scales is set to 4. Depth maps with different decomposition levels obtained by selecting the best depth result with the largest value at the same pixel position of different directional subimages. The response of the normalized Butterworth Laplace results is in the scale range of $s_{\min} = 1.5$ to $s_{\max} = W$ ($s_{\min}, s_{\max}$ denote minimum and maximum scale, W denotes the width of the depth map). Four objective evaluation criteria are employed to evaluate the performance of our algorithm on the simulated object: RMSE (RMSE measures the discrepancy between the evaluated depth map and the ground-truth depth map), PSNR (the value of PSNR is defined as a ratio between the maximum possible power of a signal and the power of the noise), correlation (the correlation coefficient quantifies the lin-

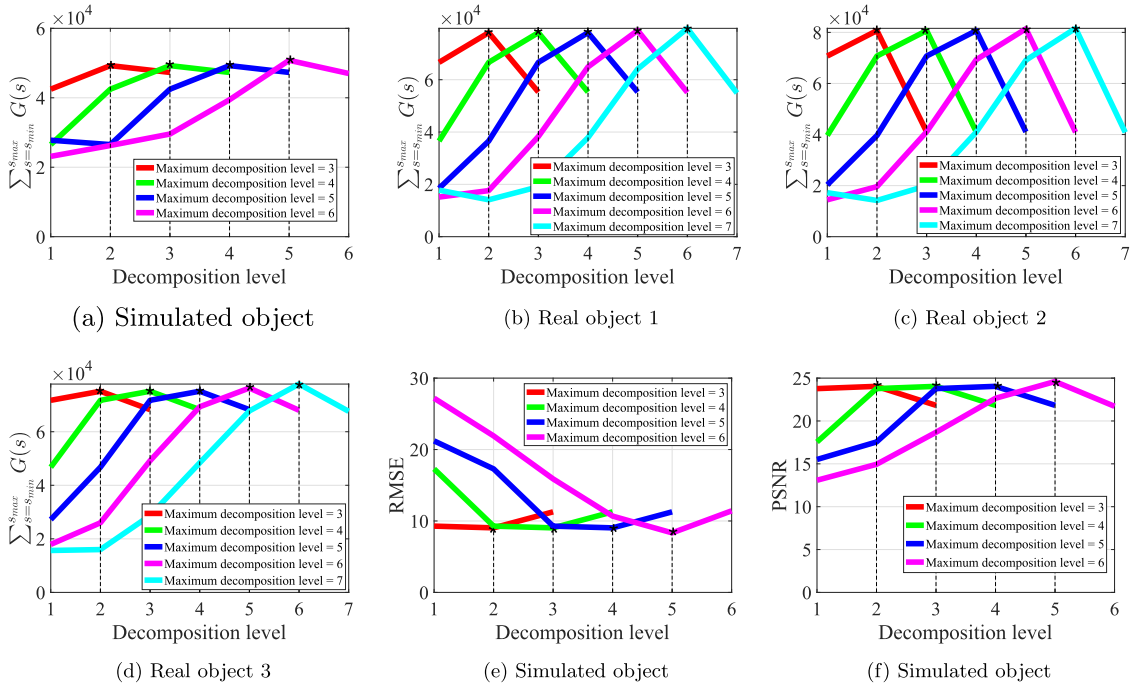


Fig. 7. Performance analysis of different depth maps in total response of the normalized Butterworth Laplace results ((a)–(d)) and the RMSE and PSNR metrics of depth maps with different decomposition levels of the simulated object ((e) and (f)).

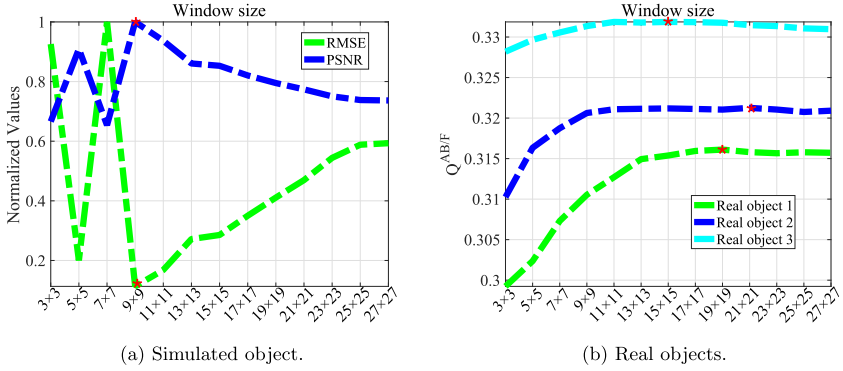


Fig. 8. Performance analysis of proposed method on different window size.

ear correlation between the evaluated depth map and the ground-truth map) and SSIM (SSIM is a measure of similarity between the evaluated depth map and the ground-truth depth map). As shown in Fig. 7(a)–(d), with an increase of the maximum decomposition level, the strongest response always appears at the former level of the maximum decomposition level, and this phenomenon is consistent with the RMSE and PSNR of the simulated object (Fig. 7(e) and (f)). Furthermore, it is noticeable that different maximum decomposition levels produce comparable results to each other. Thus, to decrease the computational time of the algorithm, in the following experiments, the maximum decomposition level is fixed at 3, and the number of directions is set to [1,4,1]. The maximum number of iterations of the postprocess method is set to 10.

To find an optimum window size to balance the robustness to the noise and detail preservation. For the real objects, we found that a pixel of depth map corresponds to a frame of the input image sequence. Therefore, the fused image can be obtained by the depth image, and the gradient-based metric $Q^{AB/F}$, which evaluates the preservation of spatial information from the source images to the fused image, is used as the evaluation criterion. Fig. 8(a) shows that the proposed method performs better for simulated object utilizing the window size 9×9 (i.e., smaller RMSE and higher PSNR). In Fig. 8(b), it can be seen that the proposed method preserves more important information within the 15×15 to 21×21 size. Therefore, in the following experiment, the focus evaluation by other SFF methods is computed by using a fixed window size 9×9 for simulated object and 17×17 for real objects, and the Gaussian model is used to estimate the depth of the SFF methods. The obtained depth maps are denoised by a 9×9 mean filter.

4.3. Comparison with SFF methods

The proposed method is implemented on simulated and real objects, and its performance is compared with some SFF methods including the Laplacian-based operators (FM_{LAPE} [29], FM_{LAPD} [30], FM_{LAP3D} [11]), wavelet-based operators (FM_{CURVELET} [12], FM_{WAVELET} [31], FM_{WAVV} [32]), gradient-based operators (FM_{GDER} [33], FM_{TEN} [34], FM_{SFIL} [35]) and the other miscellaneous operators (FM_{SF}[36], FM_{DCT} [37]). These methods are the most widely used focus measure operators to estimate image depth. The related parameters are set to the recommended values from the corresponding studies.

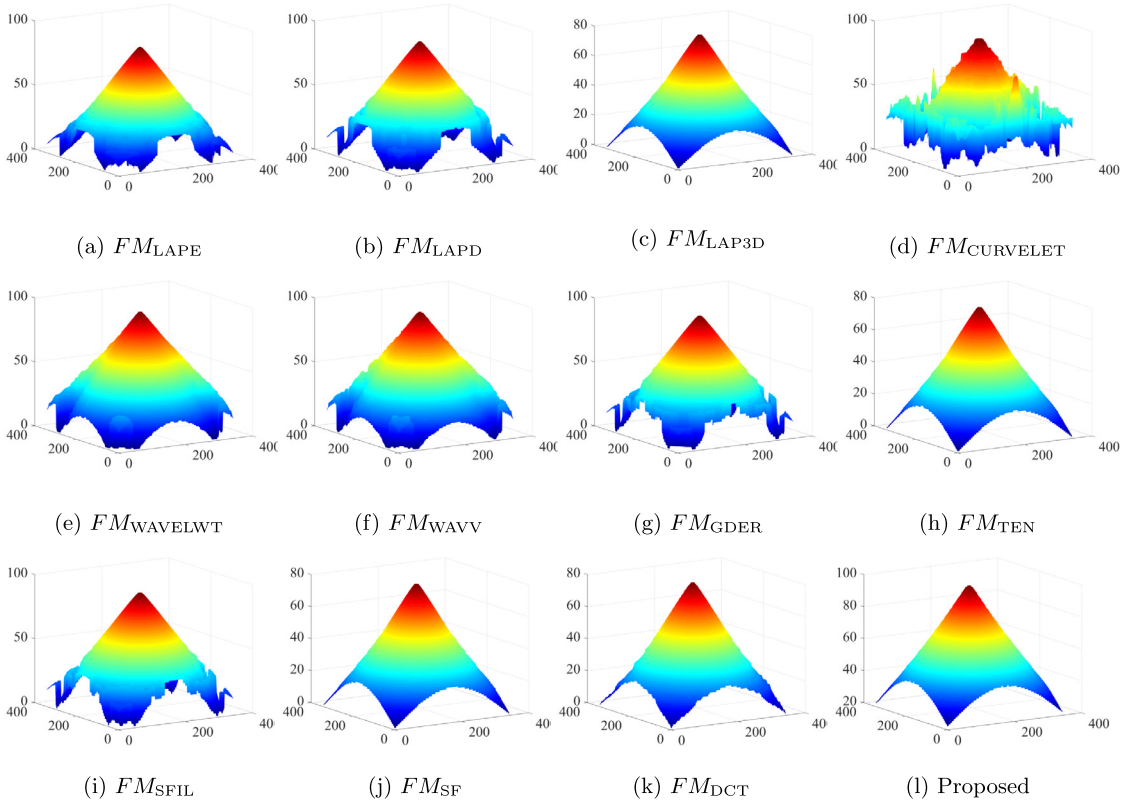


Fig. 9. Depth images computed using various SFF methods for the simulated object.

Table 1
Performance comparison of different SFF methods with four objective metrics.

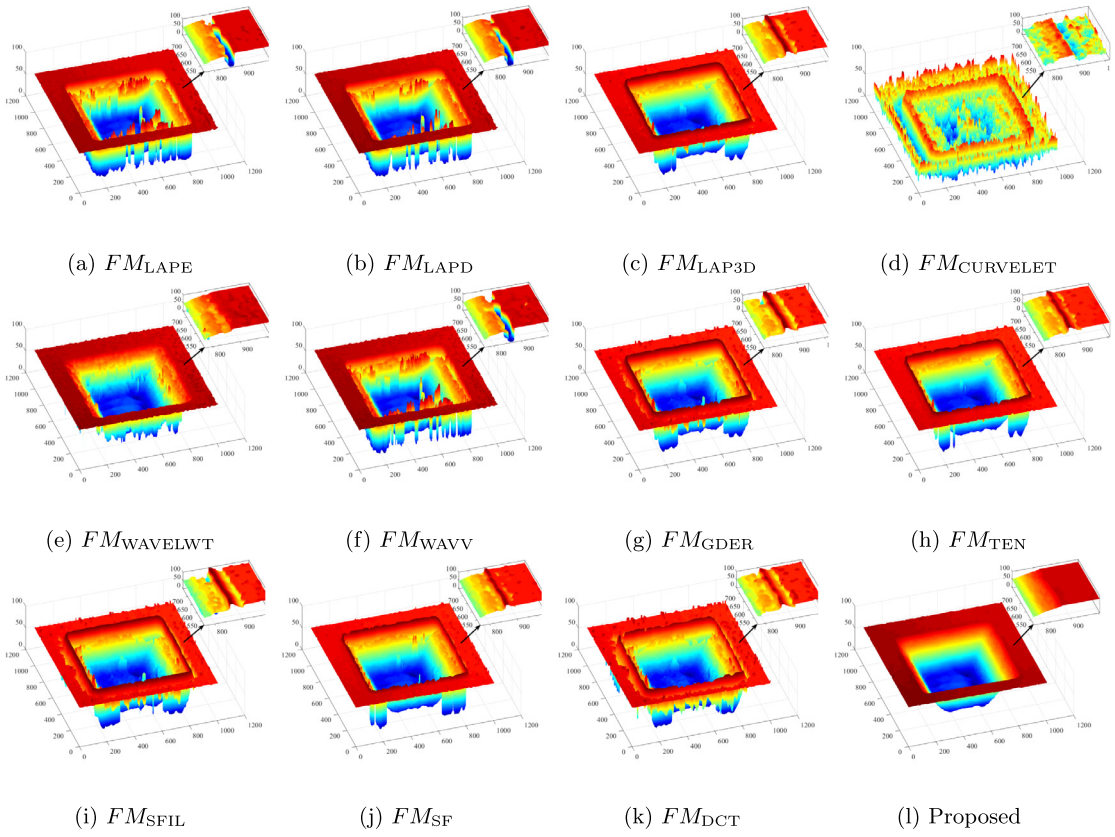
	RMSE	PSNR	Correlation	SSIM
FM_{LAPE}	6.3871	32.0247	0.9626	0.9129
FM_{LAPD}	12.3869	26.2716	0.9936	0.9449
FM_{LAP3D}	6.0448	32.5031	0.9558	0.9125
FM_{CURVELET}	16.1156	23.9859	0.6899	0.7179
FM_{WAVELET}	9.2984	28.7626	0.9175	0.8926
FM_{WAVV}	9.6178	28.4693	0.9205	0.8692
FM_{GDER}	6.2077	32.2722	0.9653	0.9189
FM_{TEN}	12.1329	26.4515	0.9944	0.9512
FM_{SFIL}	5.8818	32.7405	0.9660	0.9220
FM_{SF}	12.2128	26.3945	0.9809	0.9530
FM_{DCT}	4.3304	35.4002	0.9887	0.9879
Proposed	**2.3172**	**39.7153**	**0.9971**	**0.9896**

Three-dimensional shape reconstruction results of simulated object generated using the proposed method and other SFF methods are shown in Fig. 9. The result of FM_{CURVELET} generates a large number of inaccurate depth points (Fig. 9(d)), and the results of FM_{LAPE}, FM_{LAPD}, FM_{GDER} and FM_{SFIL} produce some incorrect depth estimation results near the boundary regions of the depth maps (Fig. 9(a), (b), (g) and (i)). The FM_{WAVELET} and FM_{WAVV} fail to preserve the smooth surface of the four corners of the depth maps. The reconstruction results of FM_{LAP3D}, FM_{SF} and FM_{DCT} extract the depth details very well, since the horizontal and vertical gradient information calculated by these methods coincides with the circular shape of the simulated object. The proposed algorithm is successful in the selection of the focus pixels from the image sequence, and less discontinuous depth information can be found in the reconstruction result (Fig. 9(l)).

Table 1 makes a quantitative analysis of the reconstructed 3D shapes for different algorithms in terms of RMSE, PSNR, correlation and SSIM. As seen, the proposed algorithm outperforms the compared methods for all four metrics. It is also observed that FM_{DCT} and FM_{SFIL} produce comparable results for the RMSE and PSNR metrics, and FM_{LAPD} and FM_{TEN} obtain first place for the correlation metric, and FM_{DCT} achieves better results for the SSIM metric.

The results of the second sequences are displayed in Fig. 10. In our experiments, the robustness to noise of the focus measure operators is assessed by performing 3D shape reconstructions under weak light conditions. It can be observed that the 3D shape of the second object recovered through the proposed method outperforms the compared methods. In addition, FM_{WAVELET} shows a good response to the reconstruction of low contrast regions, such as the boundary between the plane and the rectangular region. However, this operator is sensitive to the uniform-distributional noise. By comparing FM_{LAP3D}, FM_{GDER}, FM_{TEN}, FM_{SFIL}, FM_{SF} and FM_{DCT}, it is evident that these operators have a high sensitivity to marginal noise. From the small window, the depth image obtained by these operations is supposed to be a flat area, but it is actually a discontinuous region. The reconstructed image obtained by using the proposed method is shown in Fig. 10(l). It can be observed that the depth image is smooth, and it contains fewer discontinuities and effectively overcomes the interference of weak light regions and halo phenomena.

The reconstructed images of the third sequences, using SFF methods and the proposed method, are shown in Fig. 11. It is notable that the shape reconstructed by the proposed method performs better than the shapes reconstructed by the other methods. In particular, small changes in the plane are observed from the image boundary to the ring structure. Additionally, noise is smoothed, and the edge structure information is preserved in the results of the proposed method. It is better adapted to the human visual system and closely follows the real structure of the object. On the other hand, it is important to note that the FM_{LAPD} operator ex-

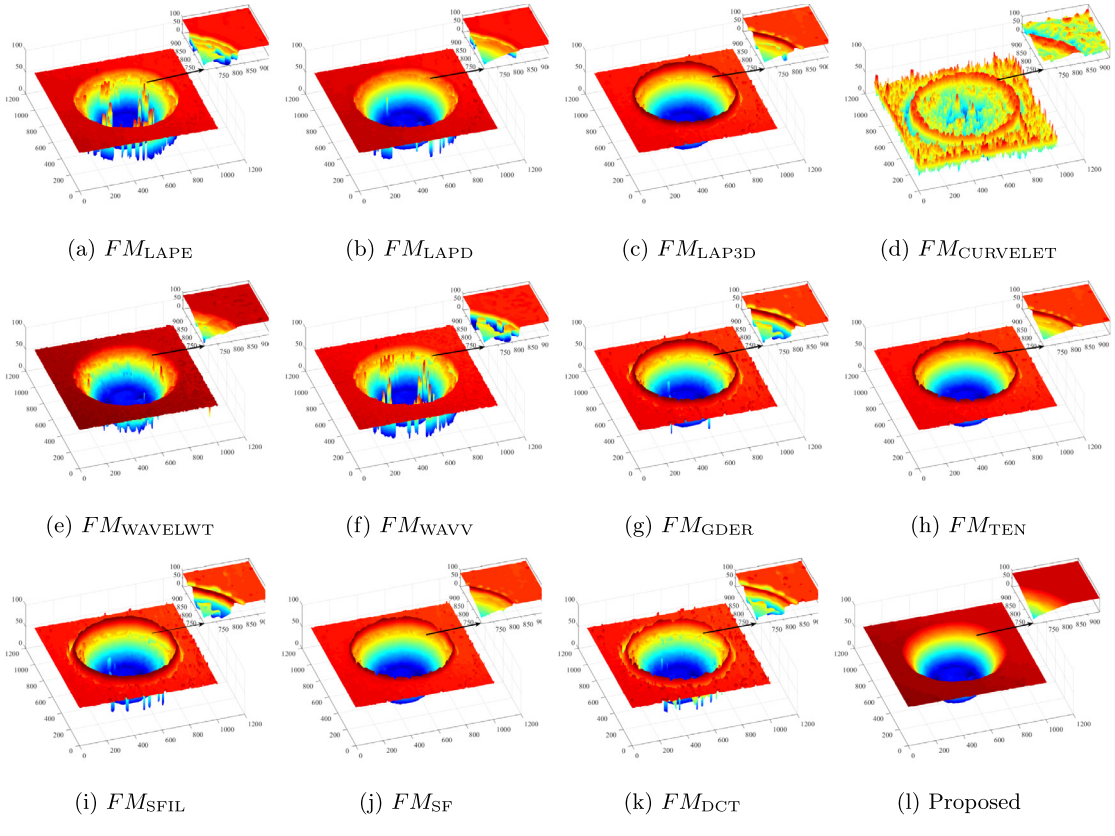


(a) FM_{LAPE} (b) FM_{LAPD} (c) FM_{LAP3D} (d) FM_{CURVELET}

(e) FM_{WAVELWT} (f) FM_{WAVV} (g) FM_{GDER} (h) FM_{TEN}

(i) FM_{SFIL} (j) FM_{SF} (k) FM_{DCT} (l) Proposed

Fig. 10. Depth images computed using various SFF methods for the second image sequence.



(a) FM_{LAPE} (b) FM_{LAPD} (c) FM_{LAP3D} (d) FM_{CURVELET}

(e) FM_{WAVELWT} (f) FM_{WAVV} (g) FM_{GDER} (h) FM_{TEN}

(i) FM_{SFIL} (j) FM_{SF} (k) FM_{DCT} (l) Proposed

Fig. 11. Depth images computed using various SFF methods for the third image sequence.

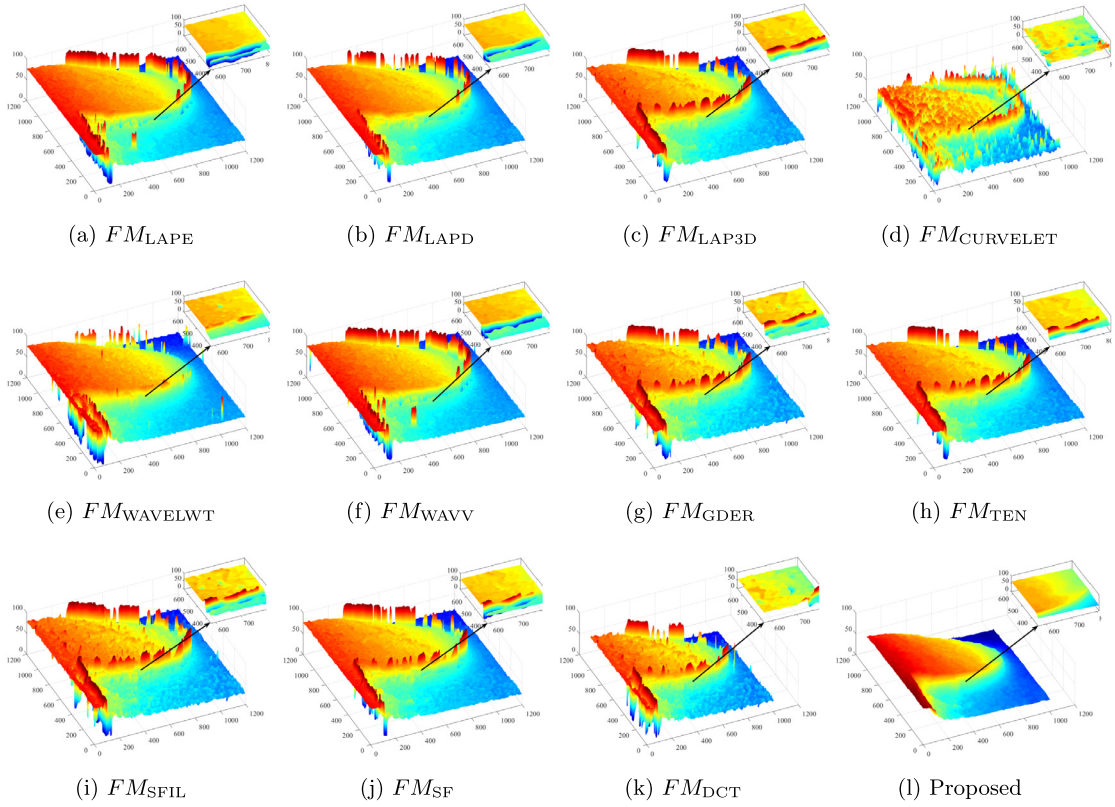


Fig. 12. Depth images computed using various SFF methods for the fourth image sequence.

hibits good performance on this sequence, but the region of the slope is still not well reconstructed. The other methods generate depth images with high fluctuations and low noise suppression, indicating their unstable behavior in noise and weak-light environments.

Fig. 12 shows the reconstructed shapes of the fourth sequence using the proposed method and the eleven SFF focus measures. It is noticeable that the 3D shape recovered by the proposed method is better than those recovered by the other methods. In addition, the depth image obtained from FM_{WAVELET} presents good results in this sequence. One explanation is that wavelet-based operators work well on rich-textured image sequences. On the other hand, errors in depth images are introduced by noise, and more spikes can be observed in the 3D shapes of the other focus measures. Therefore, reconstruction of objects with high texture by these methods produces inconsistent surfaces due to poor tradeoffs between robustness to noise and accuracy.

In general, although some traditional SFF methods have achieved better reconstruction results for the simulated object, they cannot effectively reconstruct the low contrast regions in real objects. Meanwhile, it is observed not only that the proposed method can be an effective suppression of noise, but that it can also better preserve the details of the depth information. Therefore, the proposed method can yield a good tradeoff between noise suppression and detail preservation.

4.4. Comparison with multifocus image fusion methods

The seven popular multifocus image fusion methods, which are deep convolutional neural network (CNN) [38], guided filtering (GFF) [39], convolutional sparse representation (CSR) [40], multiscale weighted gradient (MWGF) [41], NSCT [42], sparse representation (SR) [43], and dual-tree complex wavelet transform (DTCWT) [44], are used to verify the advantages of the proposed method. In addition, the related parameters in these fusion methods are set to the recommended values in the corresponding studies. In our experiments, all fusion methods map the focus image to the corresponding depth image. This operator can verify whether these image fusion results correctly reflect their true depth results.

Fig. 13 is a comparison between the reconstruction results of applying the proposed method on the simulated object and those of the other seven methods. In Fig. 13(d), the shape of the NWGF method produces many erroneous depth estimation results in the central regions. For the CSR and SR methods, a large number of discontinuous reconstruction results can be found around the bottom and slope surface. The CNN and GFF methods fail to retain the sharp tip in the depth maps. In general, the NSCT and DTCWT methods work better than other fusion methods, but the proposed method provides a sharper tip of the depth map.

The more 3D reconstruction results are close to the real depth results, the smaller the RMSE values and the larger the PSNR, correlation and SSIM values will be. The proposed method shows the lowest RMSE and the highest PSNR, correlation and SSIM values in the simulated object (Fig. 14). Additionally, the result of the NWGF method shows the negative correlation with the ground-truth. The above analysis shows that it is feasible to estimate the depth map by applying image fusion methods.

Fig. 15 displays the second image sequence of depth images obtained with different fusion methods. The depth result of the CNN method contains some undesirable results on the boundary regions between the rectangle and the plate (Fig. 15(a)). Moreover, the GFF method suffers from oversmoothness (Fig. 15(b)). In the depth result of the CSR method, we observe false depth results in some flat regions, indicating that this fusion method cannot extract the right depth for more than two source multifocus images. The same result can be observed in the SR and DTCWT methods. The MWGF method produces some artifacts near the boundaries and the NSCT method fails to reconstruct the bottom of the object.

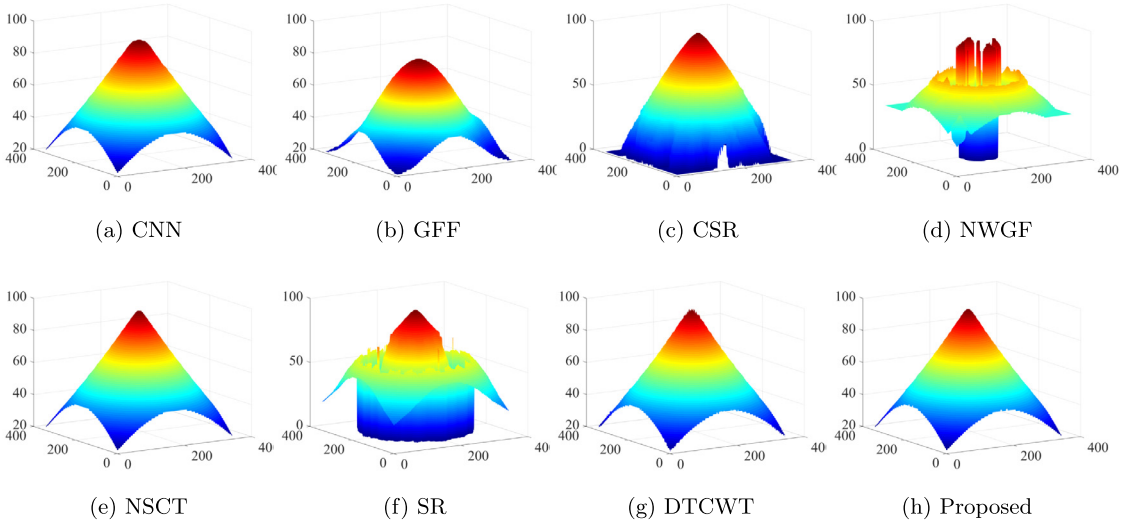


Fig. 13. Depth images of the simulated object obtained by different fusion methods.

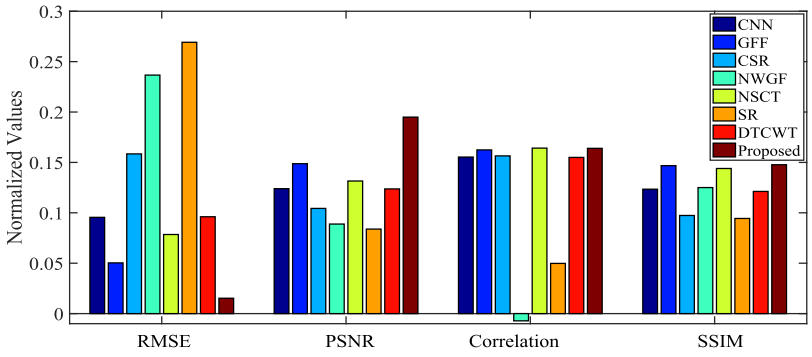


Fig. 14. Performance analysis of various depth maps using RMSE, PSNR, correlation and SSIM metrics.

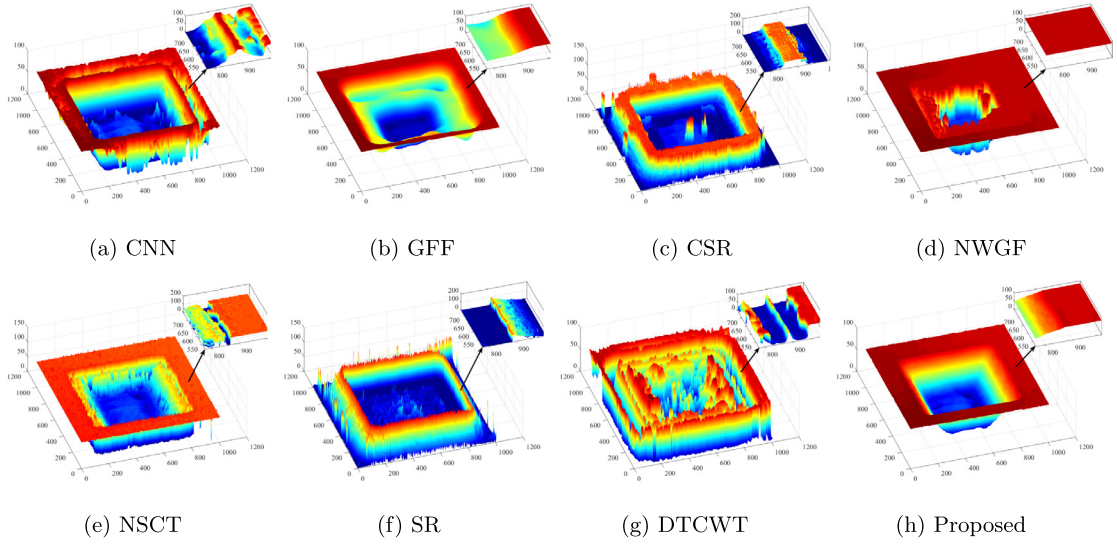


Fig. 15. Depth images of the second sequence obtained by different fusion methods.

The depth results of the third sequences are shown in Fig. 16. For the CNN method, the artifacts on the slope areas are caused by insufficient focus on low contrast images (Fig. 16(a)). For the GFF method, the overmoothness at the edges of the ring are produced, resulting in the loss of most details of the reconstruction results (Fig. 16(b)). The CSR and SR methods are different from other fusion methods in flat areas, in which the depth information of these areas is not located accurately. From Fig. 16(d), we observe that the sloped areas are reconstructed as plane areas because the MWGF method is sensitive to the obvious gradient changes. In addition,

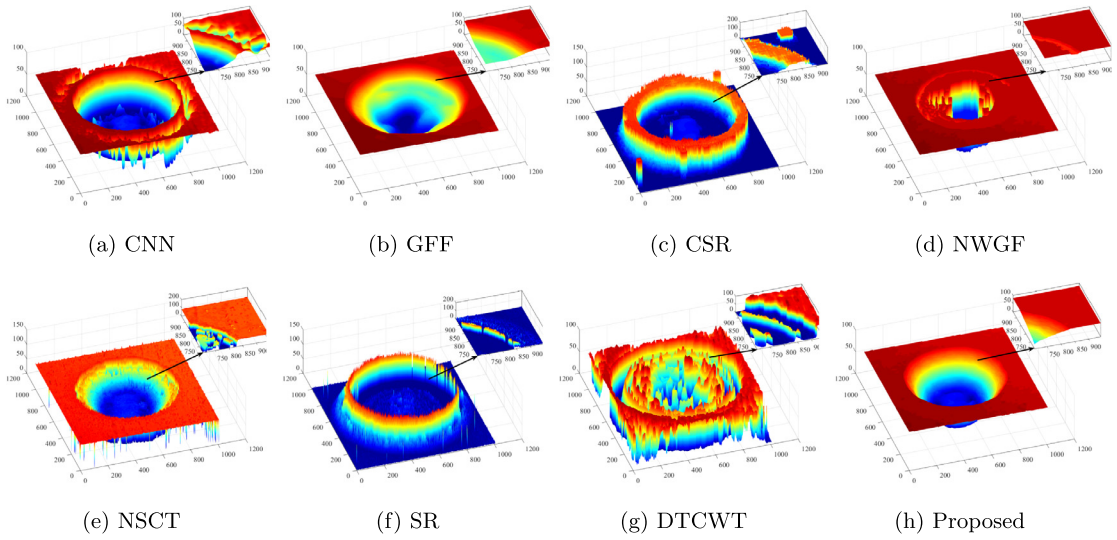


Fig. 16. Depth images of the third sequence obtained by different fusion methods.

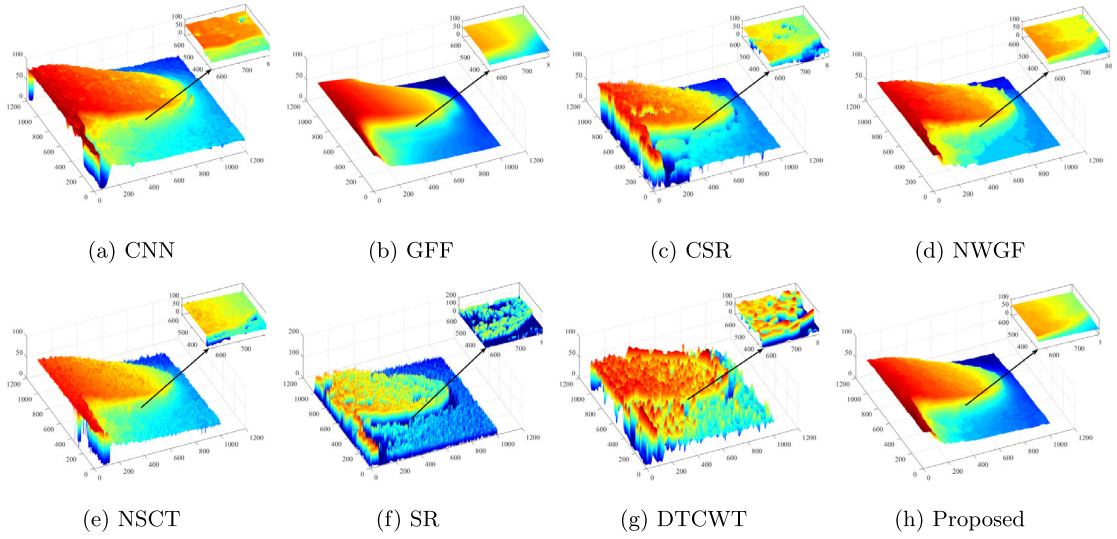


Fig. 17. Depth images of the fourth sequence obtained by different fusion methods.

the DTCWT method cannot effectively suppress the disturbance of discrete noise that derives from the halo artifacts around the contours of the object (Fig. 16(g)).

A real coin image sequence is presented in Fig. 17. The surface details of the object are rich in this sequence. Many details on the surface of an object are lost in the CNN and GFF methods (Fig. 17(a) and (b)). The depth images obtained by the CSR and SR methods contain many false depth results (Fig. 17(c) and (f)). From Fig. 17(d), MWGF is more robust for richly detailed objects since the gradient-based saliency measure method performs better at finding focus areas. NSCT has a high sensitivity to noise because image noise mostly corresponds to high subbands (Fig. 17(e)). As shown in Fig. 17(g), the DTCWT method had the worst depth result performance among these methods.

From the above analysis, although the CNN method satisfactorily preserves the focus results in small depth regions such as the plane of the object, the 3D reconstruction results produce inaccurate depth estimates in low contrast areas such as slopes and contours. This poor performance might be due to the lack of training examples for low contrast images. The guided filter of the GFF method results in oversmoothness of the depth image, and it can be explained that the guided filter blurs the source image details. The CSR and SR methods cause spatial inconsistency in flat areas. A possible reason is that the defocused regions have plenty of high-frequency noise, which could impact the accuracy of the reconstruction results. As a result, the depth information in flat areas might originate from false image frames, which lead to a large deviation between the 3D reconstruction results and the true depth results. The depth map of the MWGF method demonstrates that the reconstruction quality in the halo regions is lower than that of the other methods, but the performance in the richly detailed regions improves significantly, since a weighted gradient-based operator is more suitable for catching the saliency of local image structures. In addition, the designed strategies in the NSCT and DTCWT methods tend to confuse local details and noise. In summary, experiments from this section have shown that whether in the continuity of depth images or in detail preservation, the proposed method has the best overall performance for different objects.

Computational complexity is one of the important indices for evaluating an algorithm′s validity. In our experiments, all algorithms are implemented in MATLAB R2016a using an Intel Core

Table 2
Computational complexity of different methods (in seconds).

	Simulated object	Real object 1	Real object 2	Real object 3
FM_{LAPE}	2.13	13.27	13.28	13.28
FM_{LAPD}	3.36	25.13	24.99	26.09
FM_{LAP3D}	3.07	23.13	23.19	23.24
FM_{CURVELET}	21.18	245.84	239.48	226.11
FM_{WAVELET}	6.68	75.12	79.42	77.17
FM_{WAVV}	5.18	64.65	71.04	69.98
FM_{GDER}	3.64	25.55	22.41	27.98
FM_{TEN}	1.89	20.87	18.84	17.39
FM_{SFIL}	4.95	61.60	63.14	63.34
FM_{SF}	2.53	24.95	25.20	29.63
FM_{DCT}	$1.09e+03$	$9.08e+03$	$8.59e+03$	$8.56e+03$
CNN	$5.26e+03$	$4.85e+04$	$4.05e+04$	$4.17e+04$
GFF	8.26	106.01	104.90	105.31
CSR	$3.05e+03$	$2.77e+04$	$2.77e+04$	$2.19e+04$
NWGF	119.02	$1.24e+03$	$1.22e+03$	$1.18e+03$
NSCT	230.53	$1.91e+03$	$1.99e+03$	$1.89e+03$
SR	$3.35e+03$	$2.88e+04$	$2.25e+04$	$2.19e+04$
DTCWT	15.48	178.41	175.78	176.29
Proposed	38.18	315.80	318.07	317.15

i5 processor of 2.5GHz with 8GB RAM. The proposed algorithm mainly includes two parts: NSST decomposition and postprocessing, in which the complexity of NSST is $O(N^2 \log N)$ [26], and the complexity of postprocessing is $O(N^2)$, so the complexity of the proposed algorithm is $O(N^2 \log N)$, where $N \times N$ denotes the size of the image. Each algorithm runs 5 times independently to get the average computational time. From Table 2, it is evident that CNN is the most time-consuming among the compared algorithms, whereas FM_{LAPE} has the lowest computational burden. Among the SFF methods, the computational complexity of the proposed method is equivalent to that of FM_{CURVELET}. Compared to other image fusion algorithms, the computational complexity of our method is much lower than the majority except for those of GFF and DTCWT. Although the computational complexity of our proposed method is slightly higher than those of other traditional SFF methods, the results of the proposed method, as described in this paper, not only achieve higher reconstruction accuracy but also obtain the fused image of the object.

5. Conclusion

The NSST-based microscopic multifocus fusion method for 3D shape reconstruction has been introduced in this paper. It is based on the existence of depth information in the fused image. Due to the combination of depth estimates and fusion processes by using a novel multidirectional modified Laplacian operator, the proposed method can directly identify the depth map and fused image simultaneously. Future work on this topic will focus on three aspects:

(1) *SFF scheme based on multifactor collaborative optimizaion.* In this paper, the focus level can be evaluated using fixed window size, which may sacrifice some accuracy of the reconstruction results. However, the accuracy of the reconstruction results in the actual application scenario is affected by many factors such as the sampling interval of image sequence, the window size, the noise level, the image resolution, the focus measure operators and the reconstruction schemes. A feasible method is to convert multiple factors into different dimensions of an optimization problem that heuristic algorithms may solve. In this way, the optimal mapping from multifactor to 3D reconstruction results can be established.

(2) *Multifocus image fusion methods ensemble.* Embedding depth estimation into a NSST-based multifocus image fusion, our method can directly identify the depth map and fused image simultaneously. Therefore, it is feasible to mine depth information from the image fusion process. In recent years, many state-of-the-art multifocus image fusion algorithms, such as multiscale decomposition-based, sparse representation-based and deep learning-based methods, have emerged. Each type of fusion method has its own advantages. First, we take depth maps from different fusion algorithms as input. Next, new depth maps are generated by using replication, crossover, mutation, competition, selection and reorganization operators of genetic algorithms. It is expected to obtain higher accuracy reconstruction results through iterative optimization.

(3) *Future applications.* In the future, our method will offer important reference for increasing the accuracy of the depth information and reducing consumption in the field of intaglio printing in which a full-size plate is usually characterized by a large format. A 3D scan of the whole plate is impossible because the image file is too large. Therefore, some critical reference points should be defined for quality checks. The marked areas on the reference points in each plate will obtain some 3D results compared with theoretical designed values. In the end, different check results, such as whether the plate is ready for production or should be rejected, are obtained through the above data analysis.

The main innovation of this paper is a new attempt to bridge the gap between SFF and multifocus image fusion. A wide variety of applications in the field of quality inspection of micromanufacturing processes can benefit from this proposed 3D shape reconstruction method.

Acknowledgements

This work is supported by National Key R&D Program of China (No. 2018YFB1004300), the National Natural Science Foundation of China (Nos. 61672332 and 61872226), the Key R&D Program of Shanxi Province, China (No. 201803D421012), the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi, China (No. 2019L0100)

References

[1] C. Hong, J. Yu, J. You, X. Chen, D. Tao, Multi-view ensemble manifold regularization for 3D object recognition, Inf. Sci. 320 (2015) 395–405.
[2] C. Lin, A. Kumar, Contactless and partial 3D fingerprint recognition using multi-view deep representation, Pattern Recognit. 83 (2018) 314–327.
[3] Y. Yang, C. Deng, D. Tao, S. Zhang, W. Liu, X. Gao, Latent max-margin multitask learning with skelets for 3D action recognition, IEEE Trans. Cybern. 47 (2) (2017) 439–448.
[4] Y. Yang, C. Deng, S. Gao, W. Liu, D. Tao, X. Gao, Discriminative multi-instance multitask learning for 3D action recognition, IEEE Trans. Multimed. 19 (3) (2017) 519–529.
[5] L. Ge, H. Liang, J. Yuan, D. Thalmann, Real-time 3D hand pose estimation with 3D convolutional neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 41 (4) (2019) 956–970.
[6] E. Stoykova, A.A. Alatan, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, X. Zabulis, 3DTV: 3D time-varying scene capture technologiesa survey, IEEE Trans. Circuits Syst. Video Technol. 17 (11) (2007) 1568–1586.
[7] J. Hu, W. Zheng, X. Xie, J. Lai, Sparse transfer for facial shape-from-shading, Pattern Recognit. 68 (2017) 272–285.
[8] M. Grum, A.G. Bors, 3D modeling of multiple-object scenes from sets of images, Pattern Recognit. 47 (1) (2014) 326–343.
[9] M.T. Mahmood, T.S. Choi, 3D shape recovery from image focus using kernel regression in eigenspace, Image Vis. Comput. 28 (4) (2010) 634–643.
[10] S.K. Nayar, Y. Nakagawa, Shape from focus, IEEE Trans. Pattern Anal. Mach. Intell. 16 (8) (1994) 824–831.
[11] Y. An, G. Kang, I.J. Kim, H.S. Chung, J. Park, Shape from focus through laplacian using 3D window, in: International Conference on Future Generation Communication and Networking, volume 1, 2008, pp. 517–520.
[12] R. Minhas, A.A. Mohammed, Q.M.J. Wu, Shape from focus using fast discrete curvelet transform, Pattern Recognit. 44 (4) (2011) 839–853.

[13] A.S. Malik, T.S. Choi, A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise, Pattern Recognit. 41 (7) (2008) 2200–2225.
[14] S. Pertuz, D. Puig, M.A. Garcia, Analysis of focus measure operators for shape-from-focus, Pattern Recognit. 46 (5) (2013) 1415–1432.
[15] M.T. Mahmood, A. Majid, T.S. Choi, Optimal depth estimation by combining focus measures using genetic programming, Inf. Sci. 181 (7) (2011) 1249–1263.
[16] S. Pertuz, D. Puig, M.A. Garcia, Reliability measure for shape-from-focus, Image Vis. Comput. 31 (10) (2013) 725–734.
[17] K. Chen, Y.K. Lai, Y.X. Wu, R. Martin, S.-M. Hu, Automatic semantic modeling of indoor scenes from low-quality RGB-d data using contextual information, ACM Trans. Graph. 33 (6) (2014) 208:1–208:12.
[18] M. Muhammad, T. Choi, Sampling for shape from focus in optical microscopy, IEEE Trans. Pattern Anal. Mach. Intell. 34 (3) (2012) 564–573.
[19] M. Nejati, S. Samavi, N. Karimi, S.M.R. Soroushmehr, S. Shirani, I. Roosta, K. Najarian, Surface area-based focus criterion for multi-focus image fusion, Inf. Fusion 36 (2017) 284–295.
[20] S. Li, X. Kang, L. Fang, J. Hu, H. Yin, Pixel-level image fusion: a survey of the state of the art, Inf. Fusion 33 (2017) 100–112.
[21] G. Pajares, J.M. de la Cruz, A wavelet-based image fusion tutorial, Pattern Recognit. 37 (9) (2004) 1855–1872.
[22] B.K.S. Kumar, Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform, Signal Image Video Process. 7 (6) (2013) 1125–1143.
[23] A.L. Da Cunha, J. Zhou, M.N. Do, The nonsubsampled contourlet transform: theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.
[24] W. Lim, The discrete shearlet transform: a new directional transform and compactly supported shearlet frames, IEEE Trans. Image Process. 19 (5) (2010) 1166–1180.
[25] A. Averbuch, R. Coifman, D. Donoho, M. Israeli, J. Walden, Fast slant stack: a notion of radon transform for data in a cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible, SIAM J. Sci. Comput. (2001) 1–40.
[26] G. Easley, D. Labate, W.Q. Lim, Sparse directional image representations using the discrete shearlet transform, Appl. Comput. Harmon. Anal. 25 (1) (2008) 25–46.
[27] T. Lindeberg, Feature detection with automatic scale selection, Int. J. Comput. Vis. 30 (2) (1998) 79–116.
[28] A. Telea, An image inpainting technique based on the fast marching method, J. Graph. Tools 9 (1) (2004) 23–34.
[29] C.Y. Wee, R. Paramesran, Measure of image sharpness using eigenvalues, Inf. Sci. 177 (12) (2007) 2533–2552.
[30] A. Thelen, S. Frey, S. Hirsch, P. Hering, Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood–size, and height value interpolation, IEEE Trans. Image Process. 18 (1) (2009) 151–157.
[31] H. Xie, W. Rong, L. Sun, Wavelet-based focus measure and 3D surface reconstruction method for microscopy images, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007, pp. 229–234.
[32] G. Yang, B. Nelson, Wavelet-based autofocusing and unsupervised segmentation of microscopic images, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, volume 3, 2003, pp. 2143–2148.
[33] M.J. Russell, T.S. Douglas, Evaluation of autofocus algorithms for tuberculosis microscopy, in: International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, pp. 3489–3492.
[34] R. Rahmat, A.S. Malik, N. Kamel, H. Nisar, 3D shape from focus using LULU operators and discrete pulse transform in the presence of noise, J. Vis. Commun. Image Represent. 24 (3) (2013) 303–317.
[35] R. Minhas, A.A. Mohammed, Q.M.J. Wu, An efficient algorithm for focus measure computation in constant time, IEEE Trans. Circuits Syst. Video Technol. 22 (1) (2012) 152–156.
[36] W. Huang, Z. Jing, Evaluation of focus measures in multi-focus image fusion, Pattern Recognit. Lett. 28 (4) (2007) 493–500.
[37] S. Lee, J. Yoo, Y. Kumar, S. Kim, Reduced energy-ratio measure for robust autofocusing in digital camera, IEEE Signal Process. Lett. 16 (2) (2009) 133–136.
[38] Y. Liu, X. Chen, H. Peng, Z. Wang, Multi-focus image fusion with a deep convolutional neural network, Inf. Fusion 36 (2017) 191–207.
[39] S. Li, X. Kang, J. Hu, Image fusion with guided filtering, IEEE Trans. Image Process. 22 (7) (2013) 2864–2875.
[40] Y. Liu, X. Chen, R.K. Ward, Z. Jane Wang, Image fusion with convolutional sparse representation, IEEE Signal Process. Lett. 23 (12) (2016) 1882–1886.
[41] Z. Zhou, S. Li, B. Wang, Multi-scale weighted gradient-based fusion for multi–focus images, Inf. Fusion 20 (2014) 60–72.
[42] Q. Zhang, B.-l. Guo, Multifocus image fusion using the nonsubsampled contourlet transform, Signal Process. 89 (7) (2009) 1334–1346.
[43] B. Yang, S. Li, Multifocus image fusion and restoration with sparse representation, IEEE Trans. Instrum. Meas. 59 (4) (2010) 884–892.
[44] Y. Liu, S. Liu, Z. Wang, A general framework for image fusion based on multi--scale transform and sparse representation, Inf. Fusion 24 (2015) 147–164.

Tao Yan (corresponding author) was received the Ph.D. degree from Chengdu Institute of Computer Applications, Chinese Academy of Science. He is now a lecturer at Shanxi University. His research interests include image processing and evolutionary computation.