

# Improved shape-from-focus reconstruction for high dynamic range freedom surface

Zhoumiao He<sup>a</sup>, Pei Zhou<sup>a,b</sup>, Jiangping Zhu<sup>a,b,\*</sup>, Jianwei Zhang<sup>a,b</sup>

<sup>a</sup> College of Computer Science, Sichuan University, Chengdu 610065, China

<sup>b</sup> National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China



## ARTICLE INFO

Dataset link: <https://github.com/abaoannasleep/shape-from-focus.git>

### Keywords:

Shape-from-focus  
High dynamic range  
Guided filtering  
Depth map optimization  
Markov random field

## ABSTRACT

The 3D reconstruction of freedom surface has been a major challenge for computer vision as well as metrology. Complex surface structures such as large undulations and deep grooves, along with surface reflection properties like high dynamic range (HDR) pose many difficulties for 3D reconstruction. Shape from focus (SFF) is widely studied for its low cost, ease of deployment, and effectively avoiding occlusion. However, its inherent disadvantages, such as inaccuracy and sensitivity to noise, cannot be ignored. This article demonstrates an improved depth map optimization method to solve these problems. Firstly, a more comprehensive and reliable depth recovery evaluation method is proposed. Then an efficient optimization algorithm is applied to the filtered depth maps. Comparable experiments confirm our method's effective suppression of noise and the reliable compensation of missing signals when measuring freedom surfaces.

## 1. Introduction

3D topography recovery of freedom surface has many applications, including aerospace, medical devices, intelligent manufacturing, heritage archaeology, etc. [1,2]. SFF is a type of shape reconstruction method that utilizes focus cues from the scene. With few application restrictions, SFF is a credible alternative to 3D reconstruction in various scenes [3–6], especially in microscopic scenes, which are naturally rich in texture information. What's more, SFF faces less occlusion compared to triangulation when measuring undulating surfaces.

To recover a surface profile with SFF, a set of differently focused images usually called an image sequence is captured. With a suitable focus measure operator (FMO), the volume of focus (FV) can be determined, which contains the sharpness of each pixel. And then depth can be estimated by fitting the FV pixel by pixel [7–9].

To date, many improved SFF techniques have been proposed to recover high-quality 3D topography. These can be divided into three main categories as follows: (i) improvement of FV, (ii) reduction of time-consuming, and (iii) optimization of the depth map.

For the first category, one of the common means is to select a suitable neighborhood for the FMO. Lee et al. [10] adopted adaptive windows to decide the neighborhood, while Saxena et al. [11] employed a nested window, and Shim [12] further extended the search to 3D space.

Additionally, Nayar projected a chessboard grid pattern to solve focus ambiguity in textureless regions [13]. Mutahira et al. [14] mathematically modeled jitter noise and used recursive least squares to remove it, and Zhou et al. proposed a dual camera schema to improve the depth of field and measurement accuracy [15].

To reduce time consumption, studies have been conducted from both the hardware platform and the algorithmic perspectives. In terms of hardware-level improvements, Gladines et al. combined SFF with fringe projection profilometry (FPP) to reduce measurement time [16], and Haessig et al. [17] used event-based silicon retinas and neural processing devices to speed up SFF image processing. While in terms of algorithm-level optimization, one is to reduce the number of images required by solving the constrained Gaussian equation of focus [18–21]. Instead, Tseng et al. [22] utilized maximum a posteriori (MAP) to reduce the image volumes. In recent years, the emergence of deep learning has made the recovery of depth from a single image a hot research topic [23–29]. He et al. improved the accuracy of the VGG model by adding focus length to the dataset [30]. Regretfully, the current applications of deep learning are mainly dedicated to large scenes, instead of a single object.

Optimization of the depth map can be divided into two steps. The first is guided filtering of the depth map and the next step is the optimization of the depth map. Existing evaluations are based on the

\* Corresponding author at: College of Computer Science, Sichuan University, Chengdu 610065, China.  
E-mail address: [zjp16@scu.edu.cn](mailto:zjp16@scu.edu.cn) (J. Zhu).

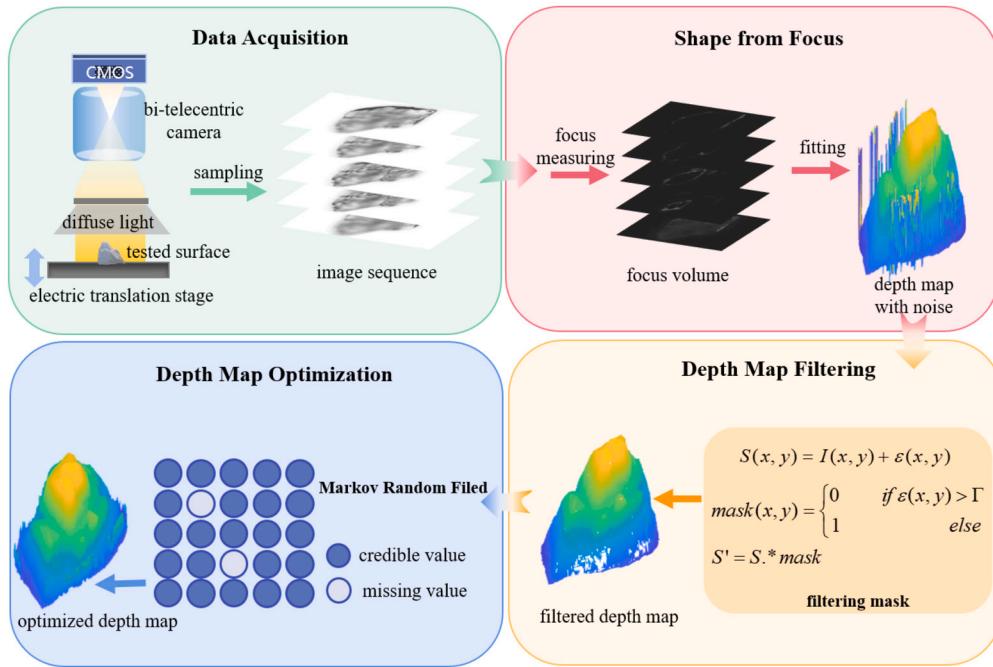


Fig. 1. The workflow overview.

statistical properties of the fitted focusing curve. For example, [31] evaluated the reliability of the fit by the Kolmogorov–Smirnov test. Ali et al. compared different guided maps in [32] and proposed to generate a guided map through the symmetric correlation of the focusing curve [33]. Furthermore, Trouve-Peoux et al. proposed a general depth error evaluation model that adopts the sensor point spread function at a given depth [34]. As for the optimization of the depth map, Moller et al. [35] optimized the depth map with the alternate directions method of multipliers (ADMM) to solve the energy function. Unlike Moller, Ali et al. used majorize-minimization [36] and tridiagonal matrix algorithms [37] to solve similar energy functions, separately. However, all these optimization methods are too complex. Markov random field (MRF) modeling is also a typical approach to achieving depth map recovery [38–41]. All these MRF methods are developed for the optimization of discrete depth values. It is challenging to optimize the continuous depth required in measurement for solutions including graph cut [38,39] and the closed-form solution [41]. Sukla proposed the Gauss–MRF model and the gradient descent algorithm was used to solve the problem [40], which gave us great inspiration because the gradient descent method is naturally applicable to the solution of continuous values.

Our work enhances SFF 3D reconstruction for HDR freedom surfaces by introducing two novel contributions:

(1) We propose a more comprehensive and effective method of removing error points. Compared to existing methods, our method achieves an almost 50% increase in the removal rate of error points.

(2) We present a more efficient and robust method for filtered depth map restoration. Our approach outperforms traditional filters, delivering higher precision and more reasonable restoration results.

## 2. Methodology

The overview of our workflow is illustrated in Fig. 1. It involves four steps: data acquisition, shape form focus, depth map filtering, and depth map optimization. Firstly, we use a dual-telecentric camera to acquire differently focused images. Then we obtain a depth map with noise based on the principle of focused imaging. With a well-designed filtering mask, the error data can be filtered out. Finally, we optimize the depth map through MRF.

### 2.1. Shape from focus

Shape from focus is a technique that derives the height information of the measured structure based on the relationship between the focus and depth. The basic principle is the point spread function (PSF), as shown in Fig. 2(a). When imaging with an optical lens, only if the object distance, image distance, and focal length satisfy the relationship of Eq. (1), the object point can be perfectly focused as a point on the image plane (e.g.,  $P$ ). Otherwise, the image point will diffuse into a dispersion circle (e.g.,  $P_1$ ,  $P_2$ ). The radius of the dispersion circle can be calculated with Eq. (2).

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{z}. \quad (1)$$

$$r = 2Rs \left( \frac{1}{f} - \frac{1}{s} - \frac{1}{z} \right). \quad (2)$$

$f$  denotes the focal length,  $z$  stands for the object distance, and  $s$  for the image distance.  $r$  is the radius of the dispersion circle, and  $R$  is the lens aperture radius. Fig. 2(b) represents the trend of the response intensity of the diffusion circle centers for  $P$ ,  $P_1$ , and  $P_2$  with the increasing axial position of the lens. When the object point is exactly in the focus state, the diffusion circle center has the largest contrast. The whole variation trend is approximate to a Gaussian distribution.

The process of SFF for 3D recovery is illustrated in Fig. 3. Once the image sequence is obtained, the FV can be calculated with FMO. The depth can be determined by fitting the focus curve.

The commonly used FMOs are Laplacian, Modified Laplacian (ML), Tenenbaum Focus Measure (TFM), Modified Gray level variance (MGLV), etc. [42,43] Here, we choose MGLV, as shown in Eq. (3). The window size of MGLV is set to 9\*9 in the following experiments. And fitting methods include Gaussian fitting and polynomial fitting [44]. In our experiment, Gaussian fitting is taken as the fitting method.

$$MGLV(x_0, y_0) = \frac{1}{N-1} \sum_{p(x,y) \in U(x_0, y_0)} (g(x, y) - \mu_{U(x_0, y_0)})^2, \quad (3)$$

where  $g(x, y)$  represents the current pixel value,  $U(x_0, y_0)$  denotes the neighboring pixels of  $(x_0, y_0)$ ,  $N$  is the number of neighboring pixels,  $p(x, y)$  is the neighboring pixel value.  $\mu_{U(x_0, y_0)}$  means the average value of neighboring pixels. Since this part is not the core of this paper, we do

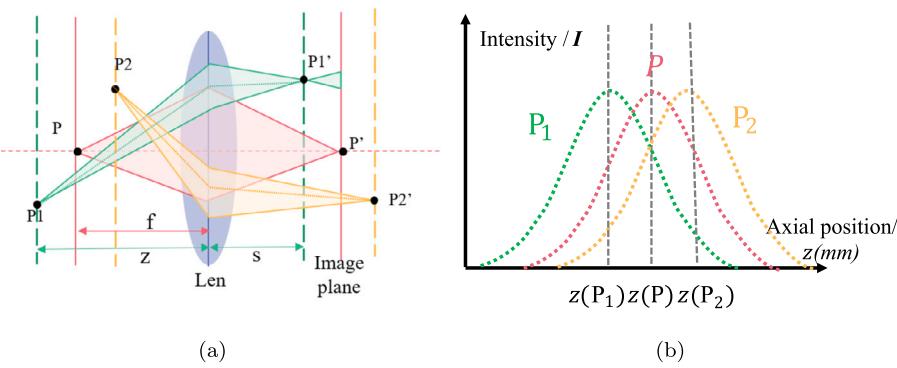


Fig. 2. The principle of SFF.

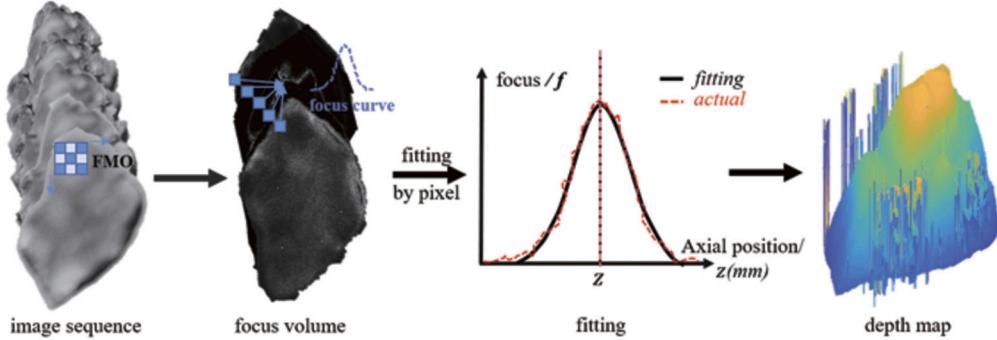


Fig. 3. The workflow of SFF.

not elaborate on it. The specific mathematical principles are detailed in [44,45].

## 2.2. Depth map filtering

In practical measurement activities, low SNR inevitably appears in the area where the light is shielded by the groove structure or the surface with complex surface properties such as weak texture and HDR(the ratio between the brightest and darkest pixel values in a captured image is extremely high.). Measurement errors will be introduced in the low SNR region, which needs to be judged and located.

If the ideal signal is defined as  $I(x, y)$ , then the captured signal  $S(x, y)$  can be expressed as Eq. (4), with  $\epsilon(x, y)$  denoting the noise due to the environment and other factors.

$$S(x, y) = I(x, y) + \epsilon(x, y). \quad (4)$$

When the noise  $\epsilon(x, y)$  is greater than the set threshold  $\Gamma$ , the SNR of the pixel at that location is determined to be too low and should be repaired, as in Eq. (5).

$$\text{mask}(x, y) = \begin{cases} 0 & \text{if } \epsilon(x, y) > \Gamma \\ 1 & \text{else} \end{cases}, \quad S' = S \cdot \text{mask}. \quad (5)$$

$S'$  is a reliable depth map with noise filtered and  $I$  is the ideal depth that we need to predict based on  $S'$ .

Based on the distribution of the focus curve, we provide a summary of the causes for two major types of error points, as shown in Fig. 4. The situation in Fig. 4(a) commonly exists in the low SNR region, where the entire focus sequence is subject to large noise interference, resulting in non-Gaussian distribution. The case in Fig. 4(b) is because the focus peak deviates from the original position. Noise interference during imaging, abrupt changes caused by large height drops, and the effect of diffusion from highly reflective spots to surrounding dark spots can all cause a shift in the focus peak.

Consequently, we use two filtering masks in Eq. (6) to comprehensively assess both types. One evaluates the Gaussian fitting errors to

avoid the error in Fig. 4(a), and the other filters out the outliers caused by the error in Fig. 4(b).

$$\text{mask}(x, y) = \text{mask 1}(x, y) \cap \text{mask 2}(x, y). \quad (6)$$

The sum of squared distances given in Eq. (7) is adopted as the metric to evaluate the fitting error. Then we can get the first mask, as expressed in Eq. (8).  $F(i)$  is the focus curve and  $G(i)$  is the Gaussian fit curve.  $\Gamma_1$  is set to  $12.5N$  in our experiments.

$$\text{SSD} = \frac{\sum_{i=1}^N (F(i) - G(i))^2}{N}. \quad (7)$$

$$\text{mask 1}(x, y) = \begin{cases} 0 & \text{if } \text{SSD} > \Gamma_1 \\ 1 & \text{else} \end{cases}. \quad (8)$$

The case in Fig. 4(b) is hard to judge by fitting tests, so we need the neighboring depth information to tell whether it is an outlier. Outliers can be detected with the Laplacian kernel in Eq. (9). Besides, hole filling of the isolated points is also required because the outliers are usually clusters of points. The hole filling is done by conditional dilation presented in Eq. (10), where  $I$  is the mask of isolated points,  $X_0$  is a zero matrix with the same size as  $I$ , and  $B$  is the dilation template. When  $X_k = X_{k-1} - 1$ , all holes are filled and  $\text{mask2} = X_k$ .  $\text{mask2}$  will also be useful in the following depth map optimization.

$$\text{Laplacian} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad (9)$$

$$X_k = (X_{k-1} \oplus B) \cap I, k = 1, 2, 3 \dots. \quad (10)$$

## 2.3. Depth map optimization

Our optimization method is based on MRF theory. Sec. 1 has introduced related studies of MRF to formulate the energy function for depth map optimization, so the derivation of the MRF energy function will be briefly described.

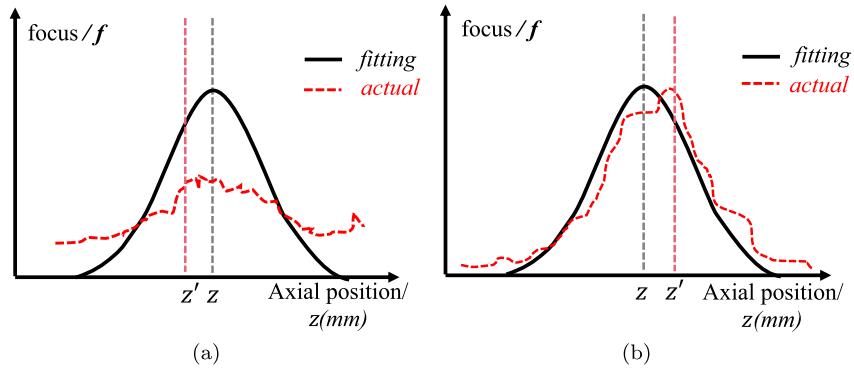


Fig. 4. Fitting errors.

Considering  $I$  as a Markov random field, according to the maximum a posteriori (MAP), we can deduce that

$$I_{MAP} = \arg \min_I \left\{ -\ln(P(S' | I)) - \ln(P(I)) \right\}. \quad (11)$$

In this case, prior in Eq. (11) can be factorized using the Hammersley-Clifford theorem, and the minimum of Eq. (11) is equivalent to

$$E(I) = V(I) + \lambda \sum_{n \in N} U_n(I). \quad (12)$$

$E(I)$  is MRF energy function,  $V(I)$  is the data term,  $U_n(I)$  means the smoothness term,  $\lambda$  is responsible for the regularization of the depth map estimated, and

$$V(I) = \|S' - I\|_2^2, \quad (13)$$

$$\sum_{n \in N} U_n(I) = \sum_{q \in N(p)} (\varpi_{(p,q)} I(p) - I(q))^2. \quad (14)$$

$\|S' - I\|_2^2$  represents the difference between observed depth  $S'$  and ideal data  $I$ .  $N(p)$  is  $p$ 's first order neighborhood, and  $\varpi_{(p,q)}$  is the weight of different neighbors. Eq. (15) is the derivative of  $\sum_{n \in N} U_n(I)$ . When the  $mask$  is 0,  $V(I) = 0$ ,  $E(I) = \sum_{n \in N} U_n(I)$ . Thus the gradient of  $E(I)$  can be expressed by Eq. (16).

$$\tau^{(n)}(p) = 2\omega_{(p,q)} [I^{(n)}(p) - I^{(n)}(q)]. \quad (15)$$

$$g^{(n)} = 2(\text{mask}.* (I^{(n)} - S' + \lambda \tau^{(n)}) + (1 - \text{mask}).* \tau^{(n)}). \quad (16)$$

The computation of  $\tau^{(n)}$  equals to convolving  $I^{(n)}$  with the kernel in Eq. (17).

$$\varpi = \begin{bmatrix} -0.5 & -1 & -0.5 \\ -1 & 6 & -1 \\ -0.5 & -1 & -0.5 \end{bmatrix}. \quad (17)$$

Since randomly initialized variables may cause local non-convergence, it's necessary to update the trusted depth map and mask after each iteration of convergence. Multiple measurement experiments prove that global convergence can be achieved by repeating 3 times. The procedure of depth map optimization is outlined in Algorithm 1.  $\lambda$  is 0.2 in our experiments.

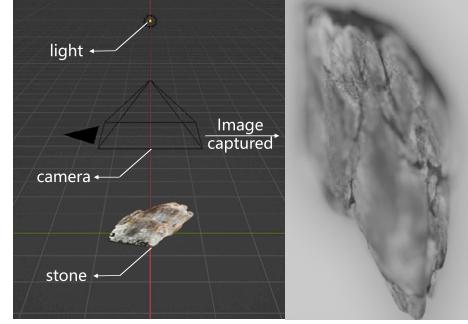


Fig. 5. Simulation configurations.

---

**Algorithm 1** Depth map optimization with MRF.

---

```

Input: observed depth map =  $S'$ , occlusion mask =  $mask$ , tolerance =  $\epsilon$ , gradient step
size =  $\alpha$ , iterations =  $t$ 
1: Initialization:  $I^{(0)}, i \leftarrow 0$ 
2: Compute  $g^{(0)}$  based on Eq. (16)
3: repeat
4:    $i \leftarrow i + 1$ 
5:    $I^{(1)} = I^{(0)} - \partial g^{(0)}$ 
6:   while  $\|I^{(n)} - I^{(n-1)}\|_2 \leq \epsilon$  do
7:     Compute  $g^{(n-1)}$  based on Eq. (16)
8:      $I^{(n)} = I^{(n-1)} - \partial g^{(n-1)}$ 
9:   end while
10:   $S' = S' * mask + I^{(n)} * (1 - mask)$ 
11:  Update  $mask$  with Eq. (9) and Eq. (10)
12: until  $i \geq t$ 
Output:  $S'$ 

```

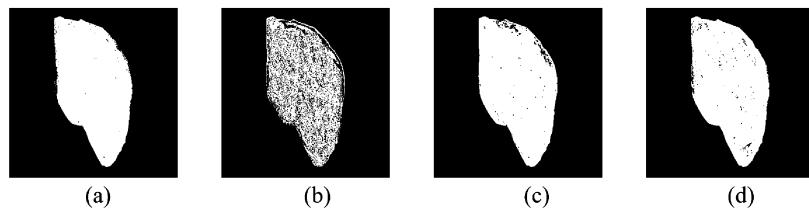
---

### 3. Experiments

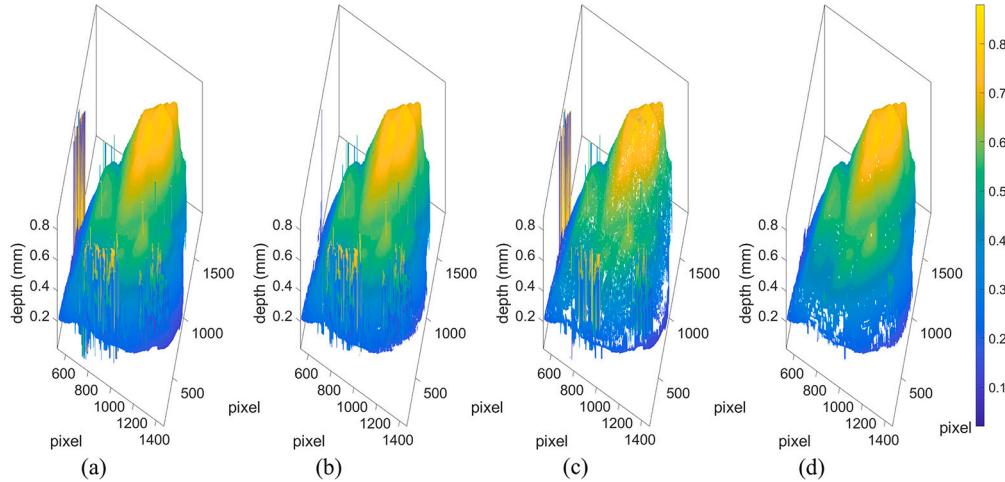
To verify our theory, both simulation and physical experiments are conducted, and the following are our experimental environment and results.

#### 3.1. Simulation

We establish simulation experiments by virtue of the 3D rendering tool *Blender 3.0.0*, which has been applied to generate simulation data because of its open source and good simulation of real scenes [46]. The experimental configuration is illustrated in Fig. 5. The tested sample is an undulating stone with deep grooves. A point light source is used as the illumination, and the camera is set as the orthographic mode to simulate a bi-telecentric imaging model. The camera aperture's F-Stop is 1.0 and the ratio is 1.0. The image sequence is sampled from 0.02 mm to 0.88 mm at an interval of 0.02 mm, and the image resolution is 2000 × 2000 pixels. To be consistent with the real environment, 1% random noise is added when rendering.



**Fig. 6.** Different masks (a) the mask generated with Method1 [31], (b) the mask generated with Method2 [33], (c) our mask1, (d) our mask2 (Black indicates irrelevant or erroneous points.)



**Fig. 7.** Filtered depth maps (a) the original unprocessed depth map, (b) the depth map filtered by Method1 [31], (c) the filtering results of Method2 [33], (d) the processing results of our method.

**Table 1**  
Accuracy of depth maps (mm).

type	Unprocessed	Method1 [31]	Method2 [33]	Ours
Valid points	92.96%	91.26%	72.75%	<b>90.81%</b>
Error point	6.23%	5.15%	5.53%	<b>3.32%</b>
Accuracy	–	92.07%	74.64%	<b>92.32%</b>

Firstly, the accuracy of the depth map evaluation is verified. Masks generated by different methods are presented in Fig. 6. The reconstructed 3D shape is visualized in Fig. 7.

It's not difficult to find that our method not only removes the outliers at the edges well but also drops the cluster of error points in the central region, compared with Method1 [31] and Method2 [33].

In addition to qualitative experiments, we conducted quantitative experiments to make our work more convincing. A deviation of 2% ( $\pm 0.172$  mm) between the measurement result and the standards is considered as the confidence interval. The measured values falling within this range are considered valid data points, while any measured values outside this range are considered erroneous. Accuracy is defined as the percentage of points that are judged correctly. The comparison results are presented in Table 1. It can be seen that Method1 [31] fails to filter out all error points, while Method2 [33] performs unstably because the edge error is not cleanly removed, and some reliable signals are excluded. So the accuracy of our method is better than that of Method1 [31] and Method2 [33].

Furthermore, two error criteria, root mean square error (RMSE) and maximum error (ME) given in Eq. (18) and Eq. (19), respectively, were evaluated for the filtered and optimized depth maps, and the results are listed in Table 2.

$$RMSE = \sqrt{\sum_{j=1}^M (dmap_j - DMAP)^2 / M}, \quad (18)$$

**Table 2**  
RMSE and ME between GT and other depth maps (mm).

Error	Unprocessed	Method1 [31]	Method2 [33]	Ours
RMSE	0.0379	0.0336	0.0334	<b>0.0156</b>
ME	0.7884	0.7854	0.7884	<b>0.3569</b>

**Table 3**  
RMSE and ME between GT and other restored depth maps (mm).

Error	Unprocessed	MF [47]	WGIF [48]	Ours
RMSE	0.0379	0.0529	0.0641	<b>0.0273</b>
ME	0.7884	0.7396	0.7611	<b>0.3169</b>

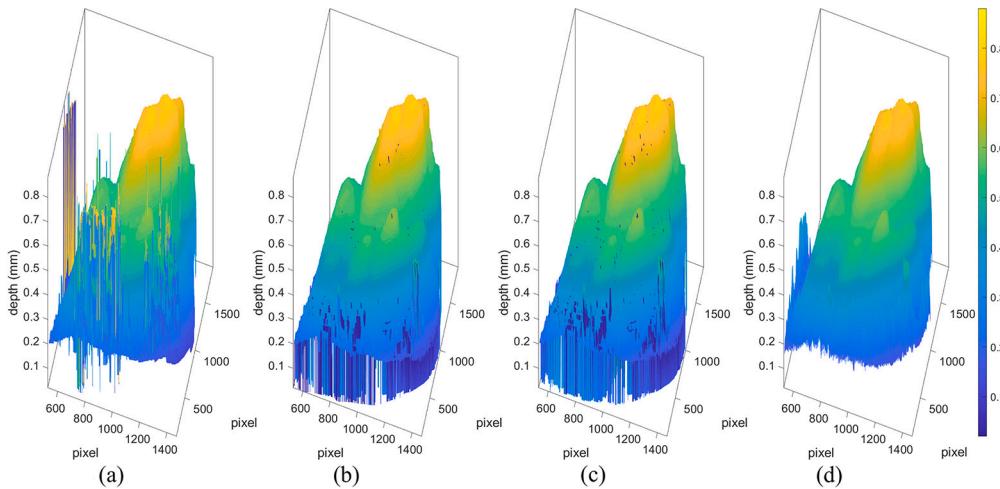
$$ME = \max(|dmap_i - DMAP|), \quad (19)$$

where  $DMAP$  is the ground truth (GT) exported by *Blender*,  $dmap_i$  means the depth map to be evaluated, and  $M$  is the number of filtered points.

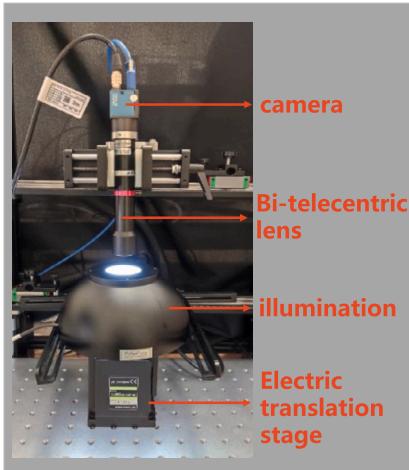
It is obvious from the data in Table 2 that the ME of the results of Method1 [31] and Method2 [33] does not differ from that of the unfiltered depth map, suggesting that they fail to handle the noise extremes. Besides, the RMSE of our method also has an advantage over them. So, it is sufficient to prove that our filtering method can effectively eliminate the noise.

For the depth map optimization, we compared our method with the median filter (MF) [47] on the filtered depth map and weighted guided filter (WGIF) [48]. The window size of MF is the same as FMO's window size (9\*9). The guided map of WGIF is the mask we generated in Sec. 2.2. Fig. 8 shows the comparison of the restored depth map. Table 3 shows the RMSE and the ME result of different methods.

Based on both the visualization and quantification results, it can be observed that our method differs from the traditional filtering method. The guided map of WGIF is the mask we generated in Sec 2.2. Fig. 8



**Fig. 8.** Simulation object (a) the original depth map, (b) the depth map restored with MF [47], (c) the depth map restored with WGIF [48], (d) the depth map restored by ours.



**Fig. 9.** Experimental environment.

shows the comparison of the restored depth map. Table 3 shows the RMSE and the ME result of different methods. Some noise at the edges is not perfectly filtered, which is the reason why the marginal area in Fig. 8(d) is not repaired effectively.

### 3.2. Objects' measurement

To further support the validity of our method, we performed physical measurements with the experimental setup in Fig. 9. The measurement system consists of a bi-telecentric camera, an electric translation stage, and diffuse illumination. The stage completes the rapid scanning of the tested surface while triggering the bi-telecentric camera to sample with constant magnification. Diffuse illumination is adopted to supplement the illumination and avoid local overexposure, thereby increasing the captured image sequence's signal-to-noise ratio (SNR). The electric translation stage is OptoSigma OSMS80-20ZF-0B with a maximum resolution of  $0.1 \mu\text{m}/\text{pulse}$ . Daheng MER2-301-125U3M is employed for the camera with a double telecentric lens WWH20-110AT-G. The double telecentric lens' aperture ( $F$ ) is 18.2 and the depth of field is 0.4. The captured image size is  $2048 \times 1536$  pixels.

First, we measured a set of metal blocks to quantify the performance of the filter. The block set consists of two stacked metal standards of 1 mm and 3 mm, given in Fig. 10(a). Due to the large height gap between the two standards and the reflective properties of the metal cross-section as shown in Figs. 10(b) and (c), the highlighted or dif-

**Table 4**  
Accuracy of depth maps (mm).

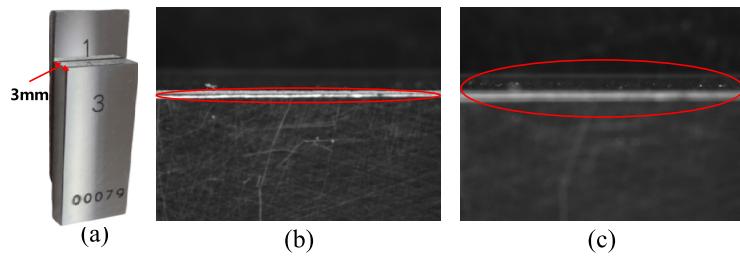
type	Unprocessed	Method1 [31]	Method2 [33]	Ours
Valid points	90.36%	90.32%	77.63%	<b>89.02%</b>
Error point	9.64%	9.16%	8.52%	<b>4.01%</b>
Accuracy	90.36%	90.80%	78.75%	<b>94.65%</b>

**Table 5**  
RMSE and ME between GT and other depth maps(mm).

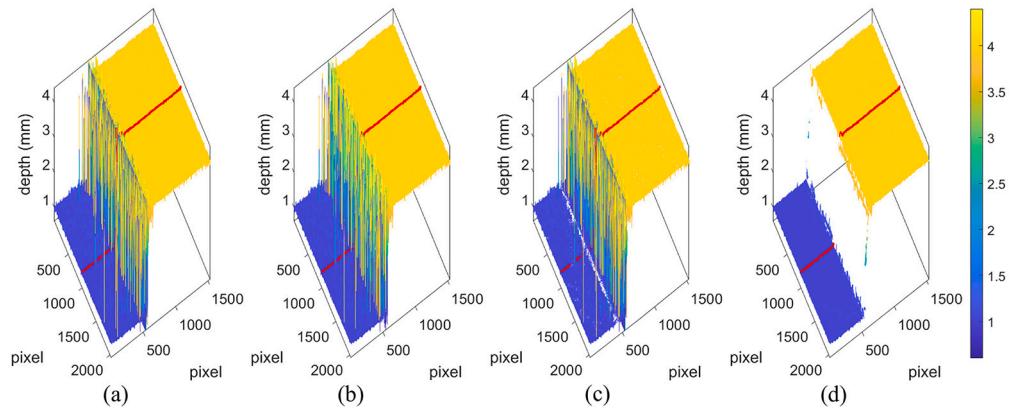
Error	Unprocessed	Method1 [31]	Method2 [33]	Ours
RMSE	0.2245	0.2269	0.2411	<b>0.0906</b>
ME	3.4000	3.4000	3.4000	<b>2.0231</b>

fuse edges inevitably cause troubles in the measurement. Fig. 11 shows the measurement and filtered results. Fig. 12 corresponds to the height information of the cross-section of the two standards. As can be seen, our method does a good job of retaining reliable results while filtering out the noise at the edges. A deviation of  $\pm 0.03$  mm between the measurement result and the standards is considered as the confidence interval. Accuracy results are presented in Table 4. The RMSE and ME of different depth maps are given in Table 5. The experimental results demonstrate that Method1 [31] is of little help in the removal of error points, although Method1 [31] retains the most valid points. While Method2 [33] improves over Method1 [31] in error points removal, it also filters out many valid points. Our method is slightly inferior to Method1 [31] in the retention of valid points but far superior to Method1 [31] and Method2 [33] in the removal of error points. In a comprehensive view, our method is more accurate than the other two methods.

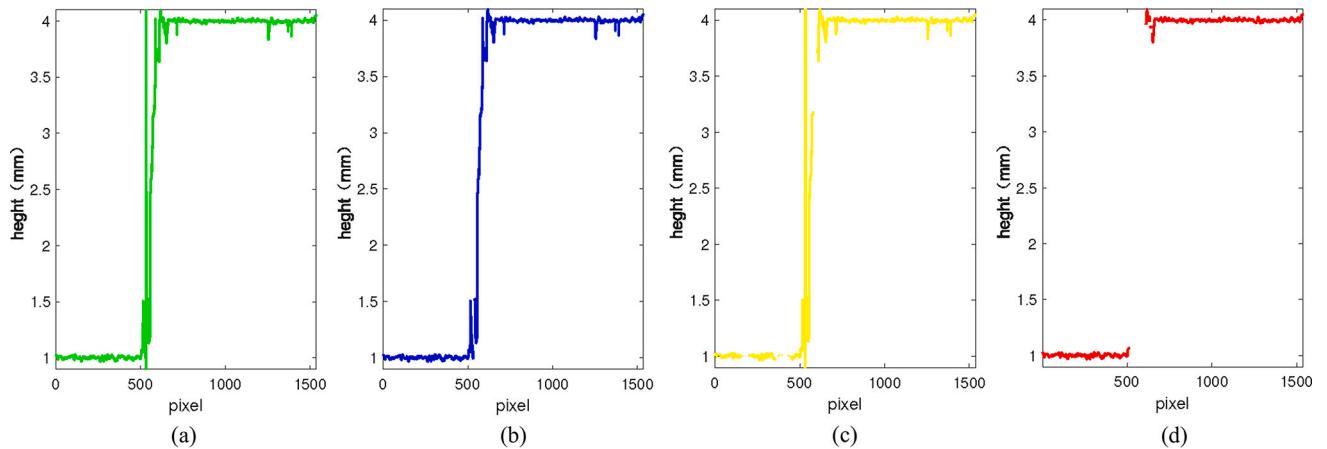
We further tested the local topography of two mechanical watch movements (red-boxed area). Since both are made of metal, the HDR properties are evident. As shown in Fig. 13, the yellow dashed box indicates the area is easily overexposed, while the red solid circle means it's too dark. Fig. 14 shows masks generated by different methods for the two movements. Fig. 15 illustrates the shape of the recovered movements and the results of the three filtering methods. The highlight and dark areas are disturbed by noise, yet our method can suppress the noise properly. The filtered depth map is restored and the comparison is given in Fig. 16 and Fig. 17. The lower half of each of the two figures corresponds to the partial enlargement of the red box. Compared with MF [47] and WGIF [48], our method restores depth data that is more natural and reasonable.



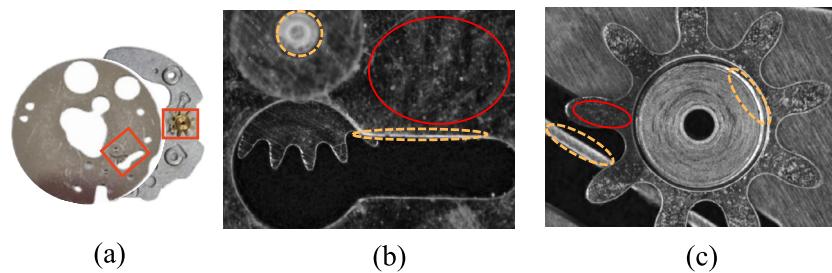
**Fig. 10.** Mental standards (a) stacked standards, (b) the highlighted edge, (c) the diffuse edge.



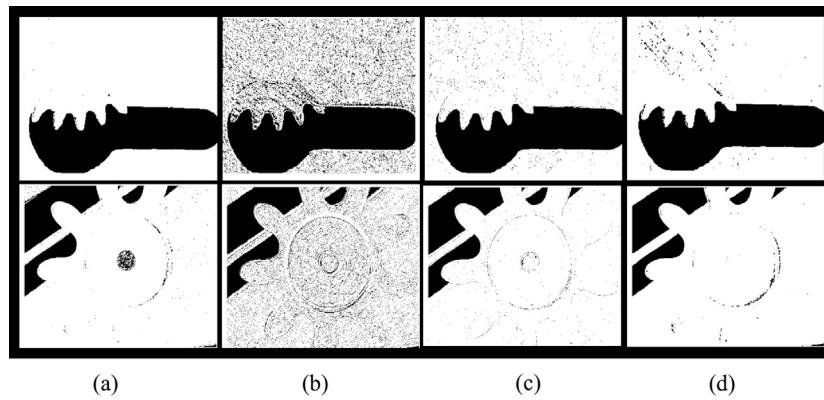
**Fig. 11.** Measurement and filtering results (a) depth maps filtered, (b) the depth map filtered with Method1 [31], (c) the depth map filtered with Method2 [33], (d) the depth map filtered with ours.



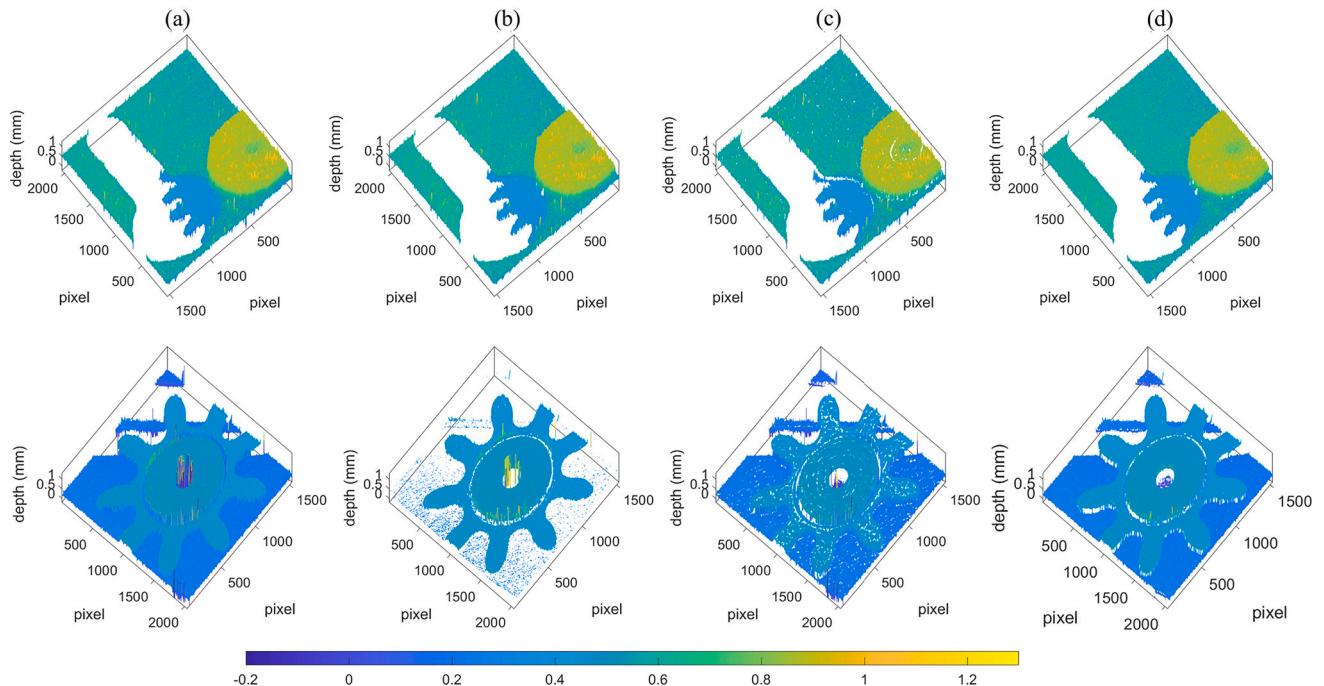
**Fig. 12.** Heights of the transects, numbered as in Fig. 11.



**Fig. 13.** (a) Watch movements; the measured area is marked with red boxes, (b) movement part-1, (c) movement part-2.



**Fig. 14.** Different masks for movements (a) the mask generated with Method1 [31], (b) the mask generated with Method2 [33], (c) our mask1, (d) our mask2 (Black indicates irrelevant or erroneous points.)



**Fig. 15.** Measurement and filtering results of watch movements (a) the depth map unfiltered, (b) the depth map filtered with Method1 [31], (c) the depth map filtered with Method2 [33], (d) the depth map filtered with ours.

Finally, we measured a more complex component, the orthodontic bracket. The bracket size is extremely small, as can be seen in Fig. 18. The complex HDR surface has deep grooves (the depth-to-width ratio reaches 1:2, as shown in Fig. 18(b) and (c)). Light source obstruction and mutual reflection greatly increase the difficulty of measurement. Masks for the bracket can be seen in Fig. 19. Fig. 20 shows a comparison of the filtered depth maps. The complex morphology and reflective properties of the brackets greatly affect the SNR of the depth maps, especially at the top and bottom of the bracket's four protrusions. The results in Fig. 20 indicate that our method is still able to filter out noise as much as possible. The optimized depth map of the bracket is presented in Fig. 21, demonstrating that our method does achieve noise suppression and whole-field optimization.

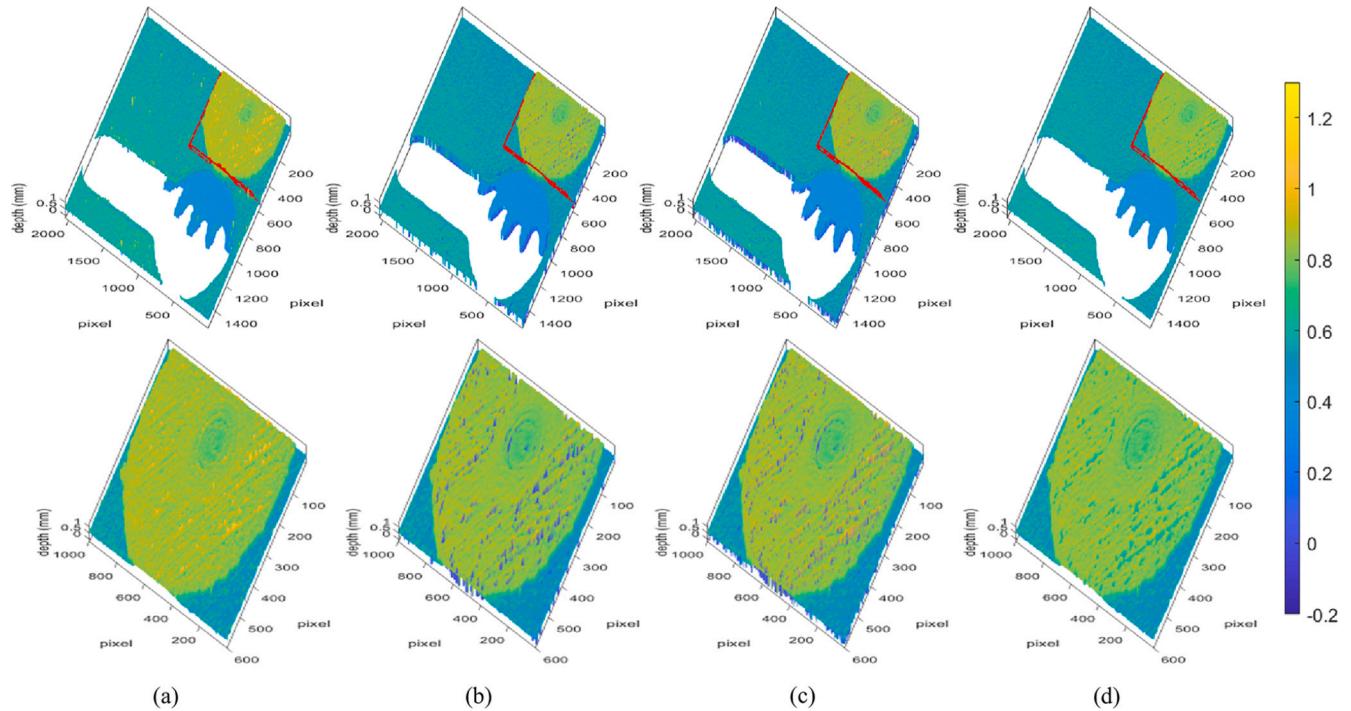
#### 4. Conclusions

To summarize, we improve the SFF 3D reconstruction for the free-form surface. Considering factors such as the complex morphological structure and the HDR properties, we design a more comprehensive evaluation method to reasonably screen out outliers to ensure the reli-

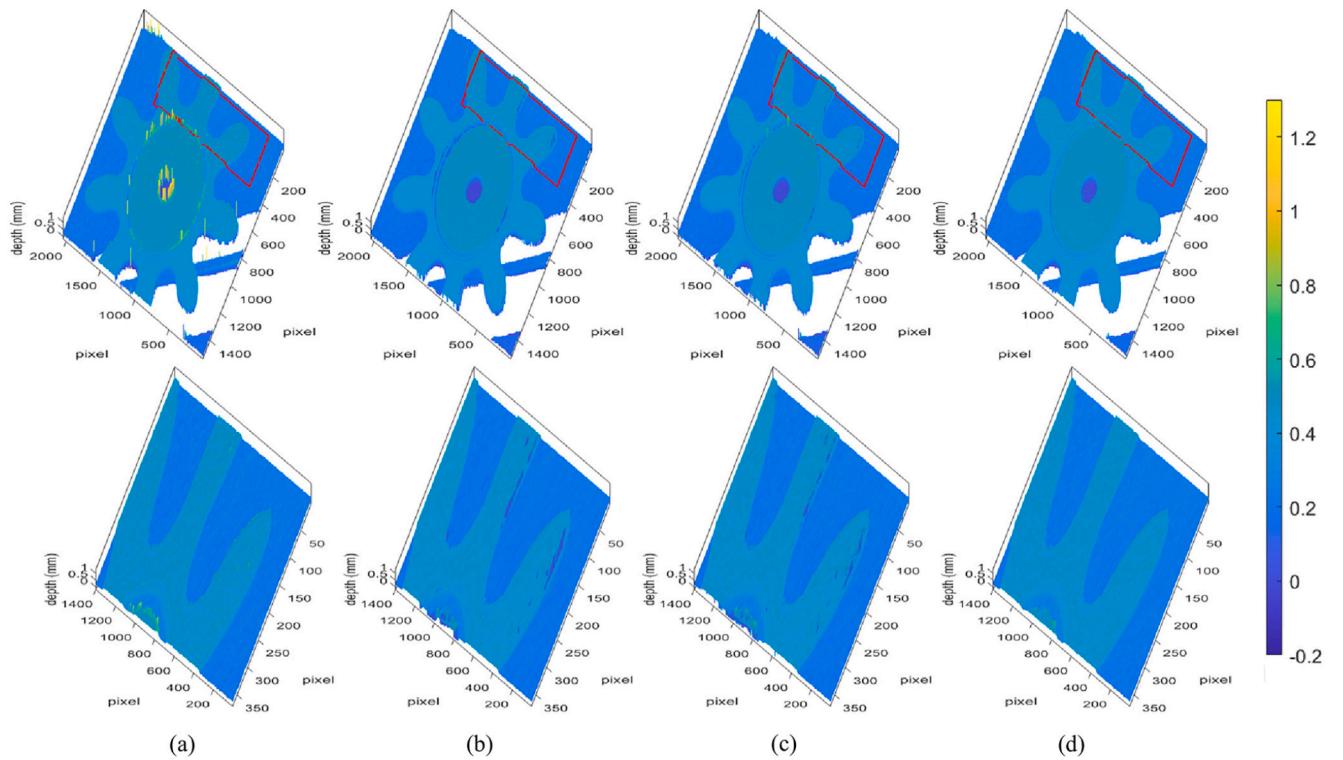
bility of the measured results. Moreover, the gradient descent algorithm is employed to solve the depth maps' optimization function based on MRF, which improves the optimization speed while ensuring accuracy. The comparison of experimental results also proves the effectiveness of the proposed method. However, the filtering method is not perfect, the threshold value needs to be adjusted according to the actual measured objects. In addition, the depth map optimization objective function needs to be improved. The MRF objective function does not consider the structural information of the object, and the restoration is not very reasonable at locations with large height drops. Constraints such as the structure information of the object may be added as an aid.

#### CRediT authorship contribution statement

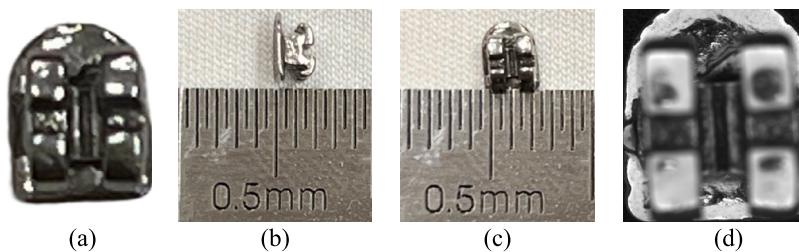
**Zhoumiao He:** Investigation, Methodology, Software, Writing – original draft. **Pei Zhou:** Validation, Writing – review & editing. **Jiangping Zhu:** Conceptualization, Resources. **Jianwei Zhang:** Project administration, Supervision.



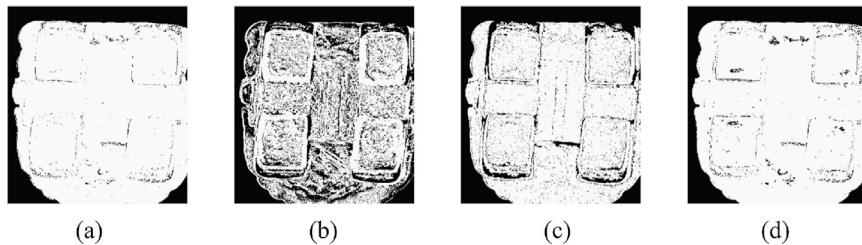
**Fig. 16.** Depth maps of watch movements-1 (a) the original depth map, (b) the depth map restored with MF [47], (c) the depth map restored with WGIF [48], (d) the depth map restored by ours.



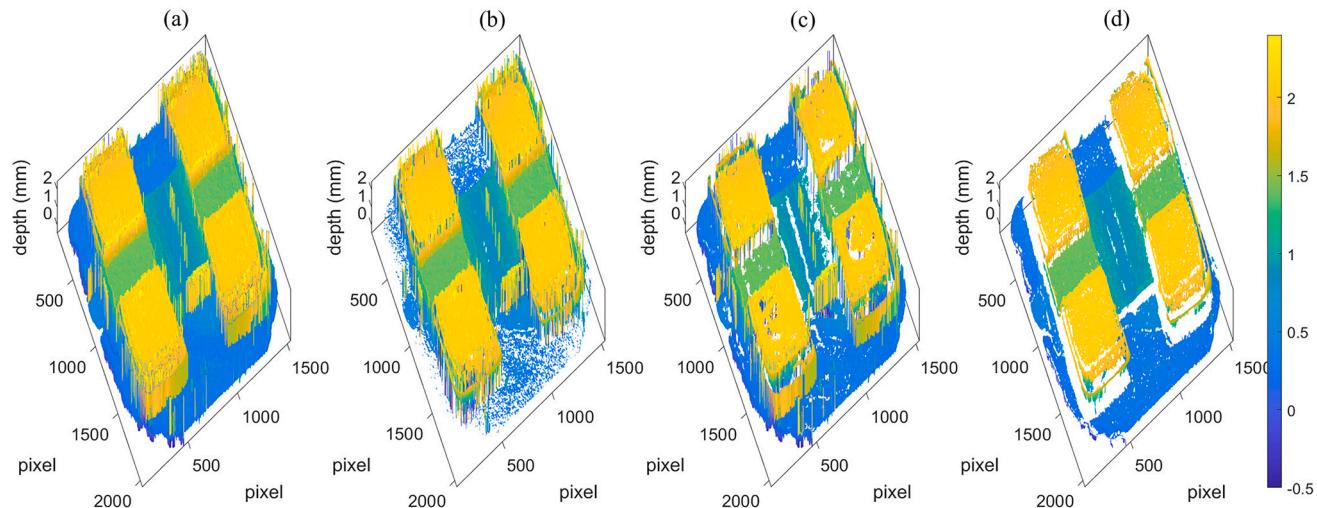
**Fig. 17.** Depth maps of watch movements-2 (a) the original depth map, (b) the depth map restored with MF [47], (c) the depth map restored with WGIF [48], (d) the depth map restored by ours.



**Fig. 18.** (a) Orthodontic bracket, (b) the depth of the bracket, (c) the scale of the bracket, (d) the image of the bracket.



**Fig. 19.** Different masks for the bracket (a) the mask generated with Method1 [31], (b) the mask generated with Method2 [33], (c) our mask1, (d) our mask2 (Black indicates irrelevant or erroneous points.)



**Fig. 20.** Filtered depth maps of the orthodontic bracket (a) the depth map unfiltered, (b) the depth map filtered with Method1 [31], (c) the depth map filtered with Method2 [33], (d) the depth map filtered with ours.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Data availability

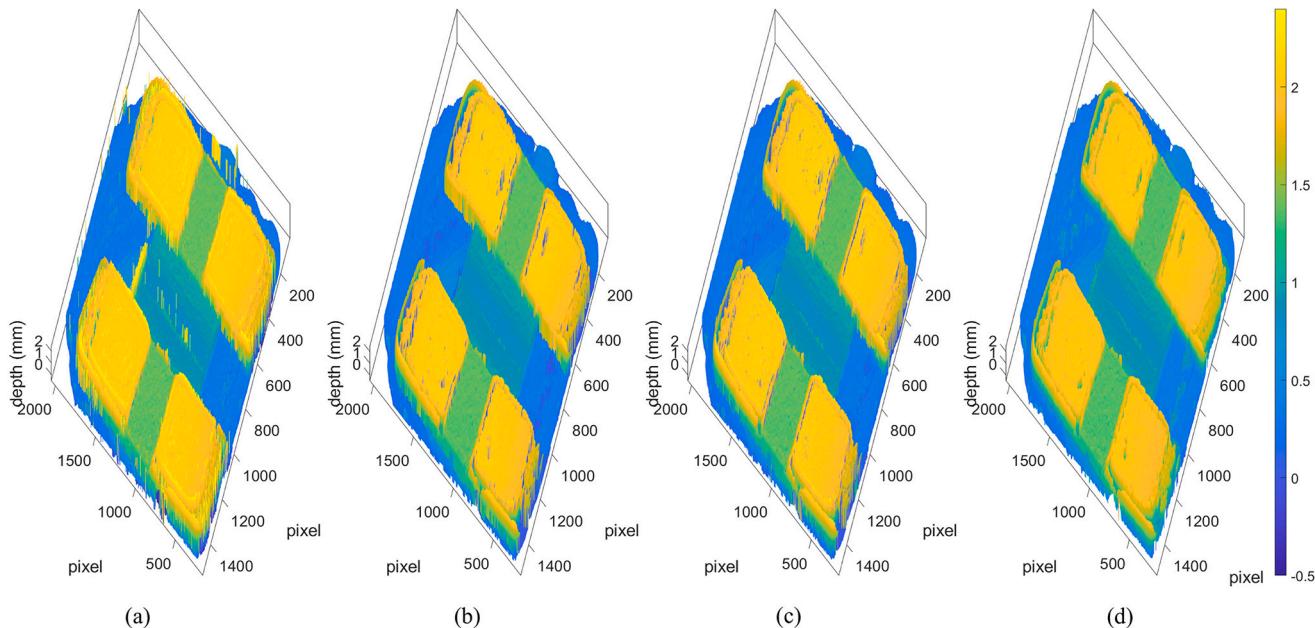
Data presented in this paper and the code will be uploaded to GitHub at <https://github.com/abaowannasleep/shape-from-focus.git>.

### Funding

This work was supported by the National Natural Science Foundation of China (grant number 62101364); The central government guides local funds for science and technology development (2022ZYD0111); Key research and development project of Sichuan province (grant numbers 2022YFG0053, 2021YFG0195); China Postdoctoral Science Foundation (2021M692260).

### References

- [1] Gomes L, Bellon ORP, Silva L. 3d reconstruction methods for digital preservation of cultural heritage: a survey. *Pattern Recognit Lett* 2014;50:3–14.
- [2] Wang H, Ma J, Yang H, Sun F, Wei Y, Wang L. Development of three-dimensional pavement texture measurement technique using surface structured light projection. *Measurement* 2021;185:110003.
- [3] Nakajima Y, Tanigaki N, Sugino T, Kawase T, Onogi S. A small and high-speed driving mechanism for 3d shape measurement in monocular endoscopy. *Sensors* 2021;21:4887.
- [4] Dinh VQ, Munir F, Sheri AM, Jeon M. Disparity estimation using stereo images with different focal lengths. *IEEE Trans Intell Transp Syst* 2019;21:5258–70.
- [5] Tang J, Qiu Z, Li T. A novel measurement method and application for grinding wheel surface topography based on shape from focus. *Measurement* 2019;133:495–507.
- [6] Jeon H-G, Surh J, Im S, Kweon IS. Ring difference filter for fast and noise robust depth from focus. *IEEE Trans Image Process* 2019;29:1045–60.
- [7] Favaro P, Mennucci A, Soatto S. Observing shape from defocused images. *Int J Computut Vis* 2003;52:25–43.
- [8] Pavliček P, Hamarová I. Shape from focus for large image fields. *Appl Opt* 2015;54:9747–51.
- [9] Nakagawa Y, Nayar S. Shape from focus. *IEEE Trans Pattern Anal Mach Intell* 1994;16:824–31.



**Fig. 21.** Restored depth maps of the orthodontic bracket (a) the original depth map, (b) the depth map restored with MF [47], (c) the depth map restored with WGIF [48], (d) the depth map restored by ours.

- [10] Lee I-H, Shim S-O, Choi T-S. Improving focus measurement via variable window shape on surface radiance distribution for 3d shape reconstruction. *Opt Lasers Eng* 2013;51:520–6.
- [11] Saxena A, Chung S, Ng A. Learning depth from single monocular images. *Adv Neural Inf Process Syst* 2005;18.
- [12] Shim S-O. Multidirectional focus measure for accurate three-dimensional shape recovery of microscopic objects. *Microsc Res Tech* 2022;85:940–7.
- [13] Nayar SK, Watanabe M, Noguchi M. Real-time focus range sensor. *IEEE Trans Pattern Anal Mach Intell* 1996;18:1186–98.
- [14] Mutahira H, Shin V, Park U, Muhammad MS. Jitter noise modeling and its removal using recursive least squares in shape from focus systems. *Sci Rep* 2022;12:1–20.
- [15] Zhou W, Zhang Y, Chen B, Tropea C, Xu R, Cai X. Sensitivity analysis and measurement uncertainties of a two-camera depth from defocus imaging system. *Exp Fluids* 2021;62:1–14.
- [16] Gladines J, Sels S, Blom J, Vanlanduit S. A fast shape-from-focus-based surface topography measurement method. *Sensors* 2021;21:2574.
- [17] Haessig G, Berthelon X, Ieng S-H, Benosman R. A spiking neural network model of depth from defocus for event-based neuromorphic vision. *Sci Rep* 2019;9:1–11.
- [18] Zhuo S, Sim T. Defocus map estimation from a single image. *Pattern Recognit* 2011;44:1852–8.
- [19] Ye M, Chen X, Li Q, Zeng J, Yu S. Depth from defocus measurement method based on liquid crystal lens. *Opt Express* 2018;26:28413–20.
- [20] Pradeep K, Rajagopalan A. Improving shape from focus using defocus cue. *IEEE Trans Image Process* 2007;16:1920–5.
- [21] Pradeep K, Rajagopalan A. Improving shape from focus using defocus information. *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 1. IEEE; 2006. p. 731–4.
- [22] Tseng C-Y, Wang S-J. Shape-from-focus depth reconstruction with a spatial consistency model. *IEEE Trans Circuits Syst Video Technol* 2014;24:2063–76.
- [23] Carvalho M, Le Saux B, Trouvé-Peloux P, Almansa A, Champagnat F. Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks? In: Computer Vision-ECCV 2018 Workshops. Proceedings, Part I, vol. 15. Springer; 2019. p. 307–23.
- [24] Anwar S, Hayder Z, Porikli F. Depth estimation and blur removal from a single out-of-focus image. *BMVC*, vol. 1. 2017. p. 2.
- [25] Anwar S, Hayder Z, Porikli F. Deblur and deep depth from single defocus image. *Mach Vis Appl* 2021;32:1–13.
- [26] Zhang A, Sun J. Joint depth and defocus estimation from a single image using physical consistency. *IEEE Trans Image Process* 2021;30:3419–33.
- [27] Ikoma H, Nguyen CM, Metzler CA, Peng Y, Wetzstein G. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. In: 2021 IEEE International Conference on Computational Photography (ICCP). IEEE; 2021. p. 1–12.
- [28] Eigen D, Puhrsch C, Fergus R. Depth map prediction from a single image using a multi-scale deep network. *Adv Neural Inf Process Syst* 2014;27.
- [29] Gur S, Wolf L. Single image depth estimation trained via depth from defocus cues. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2019. p. 7683–92.
- [30] He L, Wang G, Hu Z. Learning depth from single images with deep neural network embedding focal length. *IEEE Trans Image Process* 2018;27:4676–89.
- [31] Pertuz S, Puig D, Garcia MA. Reliability measure for shape-from-focus. *Image Vis Comput* 2013;31:725–34.
- [32] Ali U, Lee IH, Mahmood MT. Guided image filtering in shape-from-focus: a comparative analysis. *Pattern Recognit* 2021;111:107670.
- [33] Ali U, Mahmood MT. Guidance-based improvement in 3d shape recovery. In: 2021 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). IEEE; 2021. p. 1–4.
- [34] Trouvé-Peloux P, Champagnat F, Le Besnerais G, Druart G, Idier J. Performance model of depth from defocus with an unconventional camera. *JOSA A* 2021;38:1489–500.
- [35] Moeller M, Benning M, Schönlieb C, Cremers D. Variational depth from focus reconstruction. *IEEE Trans Image Process* 2015;24:5369–78.
- [36] Ali U, Mahmood MT. Robust focus volume regularization in shape from focus. *IEEE Trans Image Process* 2021;30:7215–27.
- [37] Ali U, Mahmood MT. Energy minimization for image focus volume in shape from focus. *Pattern Recognit* 2022;126:108559.
- [38] Namboodiri VP, Chaudhuri S. Recovery of relative depth from a single observation using an uncalibrated (real-aperture) camera. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE; 2008. p. 1–6.
- [39] Gaganov V, Ignatenko A. Robust shape from focus via Markov random fields. In: Proceedings of graphicon conference; 2009. p. 74–80.
- [40] Satapathy S, Sahay RR. Robust depth map inpainting using superpixels and non-local Gauss-Markov random field prior. *Signal Process Image Commun* 2021;98:116378.
- [41] Ma Z, Kim D, Shin Y-G. Shape-from-focus reconstruction using nonlocal matting laplacian prior followed by mrf-based refinement. *Pattern Recognit* 2020;103:107302.
- [42] Malik AS, Choi TS. A novel algorithm for estimation of depth map using image focus for 3d shape recovery in the presence of noise. *Pattern Recognit* 2008;41:2200–25.
- [43] Pertuz S, Puig D, Garcia MA. Analysis of focus measure operators for shape-from-focus. *Pattern Recognit* 2013;46:1415–32.
- [44] Xiong Y, Shafer SA. Depth from focusing and defocusing. In: Proceedings of IEEE conference on computer vision and pattern recognition. IEEE; 1993. p. 68–73.
- [45] Martíšek D, Mikulášek K. Mathematical principles of object 3d reconstruction by shape-from-focus methods. *Mathematics* 2021;9:2253.
- [46] Zheng Y, Wang S, Li Q, Li B. Fringe projection profilometry by conducting deep learning from its digital twin. *Opt Express* 2020;28:36568–83.
- [47] Wiener N, Kates M. Extrapolation, interpolation, and smoothing of stationary time series. John Wiley & Sons, Inc.; 1949.
- [48] Kou F, Chen W, Wen C, Li Z. Gradient domain guided image filtering. *IEEE Trans Image Process* 2015;24:4528–39.