Sequence Image Registration for Large Depth of Microscopic Focus Stacking

ANG WU1,2, YANQIONG SHI3, RONGSHENG LU1, AND ZILONG ZHANG1
1School of Instrument Science and Opto-Electronics Engineering, Hefei University of Technology, Hefei 230009, China
2College of Mechanical and Electrical Engineering, Henan Agricultural University, Zhengzhou 450002, China
3School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei 232038, China

Corresponding authors: Yanqiong Shi (syq1105@qq.com) and Rongsheng Lu (rslu@hfut.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB2003801, in part by the National Natural Science Foundation of China under Grant 51875164, and in part by the National Key Scientific Instrument and Equipment Development Project of China under Grant 2013YQ220749.

ABSTRACT The large depth of images for microscopic measurement can be achieved by using focus stacking techniques with a small depth of field of objective lens. It is implemented by fusing the image sequences of short depth images. However, the non-linear movement of the objective imaging system or the measured object caused by the moving stage straightness error brings the misalignment of the image sequences, such as transversal translation, rotation, and tilting. All of these interferences, as well as the image brightness variation must be corrected by image registration before fusing the image sequences. In this paper, a fast-automatic registration method based on the scale invariant feature transform (SIFT) is proposed. It is achieved by firstly segmenting the focal regions of the image sequences through fast edge detection. Then the image features are extracted within the small segmented focal areas. It greatly reduces the computational cost of feature extraction and the following steps of image correction, and alignment.

In the process, the random sampling consistency (RANSAC) algorithm is also used to remove the mistake features. The Laplacian pyramid method is adopted for the large depth of image fusion. The experimental results show that the proposed method is more efficient than the traditional SIFT algorithm. Its registration efficiency is improved by about 60%. This method facilitates the high-precision and real-time imaging of a monocular three-dimensional focus stacking.

INDEX TERMS Focus stacking, microscopic measurement, sequence image registration, focus segmentation, scale invariant feature transform.

I. INTRODUCTION

Focus/defocus imaging technology [1] has been widely used in the field of three-dimensional (3D) microscopic measurement, such as 3D measurement with the focus variation method [2], 3D imaging through focus stacking [3], and extending depth of field from different focused images [4]. Monocular 3D focus stacking microscopic measurement method obtains the optimal focal plane of each point on the target surface utilizing the limited depth of field and vertical scanning method of the optical system, and then obtains the target 3D structure through focus stacking technique to realize 3D measurement. This method is an inexpensive way to extend the depth of focus and expand the dynamic range of imaging for microscopic measurement. Its lateral resolution depends on the size of the camera pixel and the view field, and the depth resolution reaches micrometer degree. The method can overcome the occlusion problem and has a unique advantage in detecting small slope and complex micromachined parts which are difficult to detect.

The microscopic measurement of monocular 3D focus stacking is subject to various disturbances, such as vibration, in the process of collecting 2D image sequences. These disturbances might cause the attitude change of the camera’s photosensitive surface, and a certain degree of spatial rotation of the focal plane of the camera object at each shooting position during the shooting process. Therefore, the obtained image sequence might undergo rotation, translation, and projective transformation, as shown in Fig. 1. If the image registration technology is not adopted to align the image sequences, it is difficult to obtain accurate 3D measurement results. Therefore, considering the great significance
to reduce the complexity of image fusion algorithm and improve the robustness to displacement and vibration, it is necessary to study the registration problem of microscopic image sequences.

There are two major types of image registration algorithms: gray-based method and feature-based method. The gray-based registration method has the disadvantages of large computational complexity, rotation and scaling variance. The feature-based image registration algorithm [5] is fast, and has strong robustness to image scale, rotation, and translation, so it has been increasingly investigated and applied. At present, the most popular feature-based image registration algorithm is the well-known scale invariant feature transform (SIFT) algorithm [6], [7]. SIFT algorithm has high detection accuracy and robustness in image scale transformation, rotation, scaling, and illumination change. However, it has the disadvantages of high computational complexity and time-consuming, which makes it difficult to meet the requirements of real-time detection.

To meet the requirement of real-time and high accuracy in engineering for monocular 3D focus stacking microscopic measurement, the key problem is to improve the registration speed and accuracy. The different degrees of focus and defocus of two adjacent images measured by the focus stacking microscopic system with the registration algorithm has strong feature detection capability.

The method proposed in this paper is different from the traditional method for registering the whole image. It first extracts the focal areas of the image, and then registers the focal areas, to improve the efficiency and accuracy of the measurement. Because of the small depth of field of the camera in the focus stacking microscopic image measurement system, the image focal areas occupy a very small proportion in the whole image. If the focus area is extracted for registering the image, the calculation efficiency can be greatly improved. On the other hand, according to the characteristics of the microscopic imaging of focus stacking, when the measured object is in the out-of-focus position, the pixel position of the object is imaged with a certain offset, and the registration based on focal areas has a certain feasibility in improving the registration accuracy.

The main contributions of this work include:

1. According to the characteristics of the focus area of the focus stacking microscopic image, the focus area can be searched and segmented self-adaptively;

2. The proposed fast registration method based on the focus area greatly improves the registration speed. The method can meet the real-time detection requirements of engineering and simultaneously ensure the accuracy of registration;

3. Good fused results are obtained based on the Laplacian pyramid image fusion method.

The rest of this paper is organized as follows. Section 2 introduces the related work. Section 3 presents the principle of focus stacking imaging system. Section 4 describes the proposed registration algorithm in detail. Section 5 shows the experimental results. Finally, conclusions are given in Section 6.

II. RELATED WORK
In the past 20 years, scholars have done a lot of research on how to improve the speed and effect of image registration and put forward some representative methods. Gradient location-orientation histogram (GLOH) algorithm [8] uses a circular sector to construct a gradient position histogram descriptor and then reduces the dimensions of descriptor from 272-dimensional to 128-dimensional using principle component analysis (PCA) method. It has good robustness and distinctiveness, but the calculation has a high time cost. Features from accelerated segment test (FAST) [9] feature detector is used to detect corners, which has fast detection speed but has no scale invariance and weak robustness. Speed up robust features (SURF) [10] method improves the SIFT descriptor by using image integration instead of convolution for fast operations and Haar wavelets to approximate gradient operations in the SIFT method. The SURF feature descriptor is a 64-dimensional vector, which is faster than SIFT, but it is inferior to SIFT in terms of scale invariance, rotation invariance, and detection accuracy. Binary robust independent elementary features (BRIEF) descriptor [11] uses a binary string as the feature point descriptor. It is fast, but it is sensitive to image noise and other degradation factors, and has low rotation invariance and robustness. Oriented FAST and rotated BRIEF (ORB) [12] detector adds the main direction to each feature point, which has more rotationally invariant than FAST detector. It is ten times faster than SURF, but it does not have scale invariance. It has good detection effect for texture-rich images and poor detection effect for blur-degraded images. Binary robust invariant scalable keypoints (BRISK) feature detector [13] uses FAST feature detector for corner detection in each scale space. It has scale invariance and rotation invariance, fast speed, and
strong robustness. Fast retina keypoint (FREAK) [14] descriptor is similar to ORB feature descriptor in fast speed, rotation invariance, and scale invariance, but it is less robust to noise and blurred degraded images.

Although these methods have the fast matching speed, especially binary descriptor methods, such as BRIEF [11], ORB [12], BRISK [13], and FREAK [14], it is difficult to meet the requirements of high matching accuracy and speed at the same time. SIFT algorithm has high registration accuracy. To meet the requirements of fast matching, selected improved SIFT methods have been proposed. Some scholars used PCA method to reduce the dimension of SIFT descriptors from 128-dimensional to 36-dimensional or even lower [15], [16]. Although this method improves the registration speed, its generality and matching accuracy are not as good as SIFT. Independent component analysis (ICA) SIFT method [17] apply ICA to remove redundant information from the original SIFT descriptor to reduce runtime. Affine-SIFT (ASIFT) algorithm [18] achieves the full affine performance and four parameters invariants of an affine transform of SIFT by simulating zooms out and normalizing translation and rotation, but the algorithm complexity is more than twice that of the original SIFT algorithm. Literature [19] constrains image matching by Voronoi diagram regions to implement fast and accurate image stitching using the SIFT algorithm. In general, the performance and complexity of these feature-based image registration methods are not satisfactory at the same time. In this paper, a fast SIFT registration method based on self-adaptive focus region segmentation is proposed to achieve high speed and accurate registration of monocular 3D focus stacking microscopic measurement.

III. PRINCIPLE OF FOCUS STACKING IMAGING SYSTEM

Monocular focus stacking microscopic measurement is realized by using a small depth of field objective lens and a vertical moving scanning technique. The surface topography measurement is realized by searching the position of focus point on the surface. The simplified optical imaging system model is shown in Fig. 2. It satisfies the Gaussian imaging law:

\[
\frac{f'}{l'} + \frac{f}{l} = 1 \tag{1}
\]

![FIGURE 2. Geometry of focus stacking imaging.](image)

where \(f\) and \(f'\) are the object focal length and the image focal length of the imaging system, respectively. \(l\) and \(l'\) are the object distance and image distance of the imaging system, respectively.

In Fig. 2, \(A'B'\) is the image of the object \(AB\). The imaging sensor can obtain a clear image of \(AB\) when it is also in the position \(P\), which is called the in-focus state. If the imaging sensor is in the position \(P'\) or \(P''\), the image formed on the imaging sensor is a confusion circle, which is called the defocused state. Defocus degree is related to the distance of imaging sensor moving away from the position \(AB\).

For a small depth of field focus stacking imaging system, the imaging of object surface of different depth undergoes the process of focusing to focusing, and then to defocusing with the relative motion between the object and the camera. The process of 3D focus stacking microscopic measurement includes obtaining image sequences, image registration and fusion, and forming a full-focus 3D image with a large depth of field.

IV. THE PROPOSED METHOD

An image registration and fusion method based on approximate focus region segmentation is proposed. This method can not only improve the efficiency of image registration, but also improve the accuracy of image registration.

A. SELF-ADAPTIVE FOCUS AREA SEGMENTATION METHOD

The aim of focus area segmentation is to improve the speed and accuracy of registration [20], [21]. Because of the complexity and timeliness of precise focus region segmentation algorithm, it cannot improve image registration efficiency. In this study, an approximate adaptive focus region segmentation method is used to obtain the location information of the approximate focus region from the reference image, which is also used as the location information of the focus region of the registered image. This method can quickly complete the segmentation of the focus region.

The detail information of the focus area of the image is abundant and the gray gradient value of the pixel is significantly different from that of the out-of-focus area. According to the threshold of the gray gradient, the focus area can be segmented. Roberts gradient operator [22] is used to segment the focus area quickly and adaptively. The specific steps are described as follows:

Step 1: Calculate the gradient \(G\) of each pixel of the reference image. The formula is listed as follows:

\[
G(x, y) = \left| I(x, y) - I(x+1, y+1) \right| + \left| I(x+1, y) - I(x, y+1) \right| \tag{2}
\]

where \(I(x, y)\) is the pixel grey value at image coordinate \((x, y)\).

Step 2: Segment the image and calculate the gradient value of image sub-blocks. The source image of size \((N, M)\) is divided into sub-blocks with the same size. The smaller the sub-block \(W\) is, the more accurate the segmentation of the focus area is, but the amount of calculation is larger. The \(W\) size chosen in this work is 128 \(\times\) 128 and the step size is 64.
The gradient mean \( \eta \) of \( W \) is calculated by

\[
\eta = \frac{1}{128^2} \sum_{x \in W} \sum_{y \in W} G(x_i, y_i) 
\]  

(3)

Step 3: Determine the threshold \( T \) of focus region segmentation.

The segmentation threshold \( T \) of the image focus area is presented through the experimental analysis of the measurement system. It contains the following steps:

1. Calculate the gradient mean of each column of the reference image, and subsequently obtain a \( 1 \times M \) gradient sequence \( P \):

\[
P(y) = \frac{1}{N} \sum_{x=1}^{N} G(x, y), \quad y = 1, 2, \ldots, M
\]  

(4)

2. Sort the sequence \( P \) from large to small, and obtain the sequence \( P' \);

3. Calculate the segmentation threshold \( T \) of the image sub-block to be judged as the focus area.

It can be seen from the statistical results of a large number of experiments that the percentage of the focus area in the focus stacking image is 10-15%, and it's appropriate to take 10% in this work, \( |0.1 \times M| P' \) so the focus region segmentation threshold \( T = P' \times [0.1 \times M] \).

Step 4: Obtain the polygon \( A_p \) of image focus area. If \( r_i > T \), the image sub-block is judged as the focus area \( s_i \). The union set \( A_p \) of the focus image sub-blocks is a polygonal area, in other words, \( A_p = \cup s_i \).

Step 5: The minimum outer truncated rectangle of \( A_p \) is the final segmented focus area \( S \).

B. FEATURE DETECTION IN SCALE SPACE

Gauss kernel is the only kernel that can generate multiscale space [23]. The scale space \( L(x, y, \sigma) \) can be expressed as the convolution of Gauss kernel \( G(x, y, \sigma) \) and an image \( I(x, y) \) [6], which can be calculated by the following formulas:

\[
L(x, y, \sigma) = G(x, y, \sigma) \ast I(x, y) 
\]  

(5)

\[
G(x, y, \sigma) = \frac{1}{2\pi \sigma^2} e^{-\frac{x^2 + y^2}{2\sigma^2}}
\]  

(6)

where \((x, y)\) are the spatial coordinates, \( \sigma \) is the scale space factor, \( G(x, y, \sigma) \) is a scale-variable Gaussian function.

The difference of Gaussian (DoG) is employed to detect the extrema in the scale space [7], which is defined as follows:

\[
D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] \ast I(x, y)
\]  

(7)

where \( k \) is the scale ratio of two adjacent Gauss convolution images.

DOG operator is used to approximating the scale normalized Laplace-Gauss (LOG) operator. Gauss pyramid is obtained by Gaussian smoothing and downsampling, and then DOG pyramid is generated by the subtraction of adjacent scale images \( k \) and \( k-1 \). In scale space, each sampling point is compared with its adjacent points on the same scale and its adjacent points on the upper and lower scales, to ensure it is a partial extremum point in image space and scale space.

To improve the anti-noise ability and stability, the keypoints with low contrast are removed by setting threshold, and the unstable edge response points are removed by calculating the threshold of principal curvature with Hessian matrix.

C. FEATURE DESCRIPTOR

The size and direction of the gradient are calculated as follows:

\[
m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2}
\]  

(8)

\[
\theta(x, y) = \arctan \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)}
\]  

(9)

The feature region around the keypoint is divided into some small blocks, and the gradient histogram of each block is calculated to generate the vector descriptor of the keypoints. There are 4 \( \times 4 \) small blocks around the key point, and the size of each small block is \( 4 \times 4 \). Each small block has eight direction gradient cumulative values, which constitutes the feature descriptor. The descriptor is normalized to enhance the invariance to illumination changes [24]. The final keypoints contain the information of location, scale, and direction.

D. FEATURE MATCHING

The nearest neighbor distance algorithm (NN) is used to feature match. The feature points of the nearest neighbor and the next nearest neighbor in the registered image are sought, corresponding to the feature point in the reference image. If the ratio of Euclidean distances is less than a certain threshold, the matching point pair is correct.

E. IMAGE TRANSFORMATION

The spatial transformation model parameters between the reference image and the registered image are calculated based on the matching feature points. Affine transformation can be used to describe the translation, rotation, and scaling transformation. This is the main transformation in focus stacking microscopic measurement system. The linear geometric relationship between matching point pairs is expressed as:

\[
\begin{bmatrix}
x' \\
y' \\
1
\end{bmatrix} = \begin{bmatrix}
h_{11} & h_{12} & h_{13} \\h_{21} & h_{22} & h_{23} \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
x \\
y \\
1
\end{bmatrix} = H \begin{bmatrix}
x \\
y \\
1
\end{bmatrix}
\]  

(10)

where \( H \) is the homograph matrix of affine transformation, \([x' \ y' \ 1]^T\) and \([x \ y \ 1]^T\) are the pixel coordinates of matching feature point pairs.

Equation (10) can be converted into

\[
\begin{cases}
h_{11}x + h_{12}y + h_{13} - x' = 0 \\
h_{21}x + h_{22}y + h_{23} - y' = 0
\end{cases}
\]  

(11)
Three matching feature points can calculate the homograph matrix $H$ in (11). In practical application, due to the existence of mismatching, more matching feature point pairs are needed to solve $H$ to improve the accuracy of the solution.

Supposing the matching point pairs is $n$, a matrix equation is obtained as

$$Ah = 0$$  \hspace{1cm} (12)$$

where,

$$A = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x'_1 \\ 0 & 0 & x_1 & y_1 & 1 & -y'_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x'_2 \\ 0 & 0 & x_2 & y_2 & 1 & -y'_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x'_n \\ 0 & 0 & x_n & y_n & 1 & -y'_n \end{bmatrix}$$ \hspace{1cm} (13)$$

$$h = [h_{11} \ h_{12} \ h_{13} \ h_{21} \ h_{22} \ h_{23} \ 1]^T$$ \hspace{1cm} (14)$$

If the neighborhood information of the feature points is particularly similar, the possibility of mismatching will increase. When erroneous matches in the matching point set appear, the minimum least squares method to solve the transformation parameter $h$ is easy to fail. To improve the image matching accuracy, random sample consensus (RANSAC) algorithm [25], [26] is employed to eliminate the mismatched points.

First three pairs of matching points are randomly selected in the initial matching set, and the homograph matrix $H$ is solved by (12). Subsequently, the homo-order expression $\begin{bmatrix} \bar{x}_i \ \bar{y}_i \ 1 \end{bmatrix}$ of $(x_i, y_i)$ is calculated by

$$\begin{bmatrix} \bar{x}_i \\ \bar{y}_i \\ 1 \end{bmatrix} = H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$ \hspace{1cm} (15)$$

If the difference value between $(\bar{x}_i, \bar{y}_i)$ and $(x'_i, y'_i)$ satisfies the following equations:

$$\sqrt{(\bar{x}_i - x'_i)^2 + (\bar{y}_i - y'_i)^2} < \eta$$ \hspace{1cm} (16)$$

$(x_i, y_i)$ is considered as the inside point. Otherwise, it is considered as the outside point. In this work, $\eta$ is the preset threshold value (normally, $\eta = 3$) [27].

$H$ is calculated by the inside points. The coordinates $(x, y)$ in the registered image is transformed to the coordinates $(x', y')$ in reference image coordinates for alignment. The mapping relationship is listed as follows:

$$\begin{cases} x' = h_{11}x + h_{12}y + h_{13} \\ y' = h_{21}x + h_{22}y + h_{23} \end{cases}$$ \hspace{1cm} (17)$$

The bilinear interpolation method is used to solve the problem of decimal coordinates, and the registered image is obtained.

**V. EXPERIMENTAL RESULTS AND ANALYSIS**

**A. EXPERIMENT SETTING OF FOCUS STACKING MICROSCOPIC MEASUREMENT**

The experimental system for acquiring image sequences of 3D focus stacking measurement is shown in Fig. 4.

**F. FOCUS STACKING IMAGE FUSION**

An image with large depth of field and full clarity is formed to expand the depth of field of microscopic objective by accurately finding the image number and position of the greatest focusing value and extracting the clearest pixels. As a multi-scale decomposition method of image fusion, pyramid transform [28], [29] can highlight the feature information of the image and obtain better fusion effect. Moreover, the salient features in different scales and resolutions are much more in line with human visual characteristics. Laplacian pyramid method [30] can extract and preserve the details of the source image and fuse them into the fused image to achieve a better fusion effect. In this work, the Laplacian pyramid method is used for image fusion. The main steps are described as follows:

Step 1: Laplacian pyramid decomposition is performed on the registered image sequences to obtain a series of pyramid images consisting of low-frequency approximate images and high-frequency detail images.

Step 2: Calculate the weight map of the original image. Sobel operator is used to calculate the gradient value of image sequences, and the gradient mean of each pixel is calculated to form the weighted map of the original image.

Step 3: Gaussian pyramid decomposition is performed on the weighted map.

Step 4: A fused image is reconstructed by weighting the layered source images and the corresponding weights.

The flow chart of image fusion is shown in Fig. 3.

**FIGURE 3. Flowchart of image fusion based on Laplacian pyramid method.**
A. Wu et al.: Sequence Image Registration for Large Depth of Microscopic Focus Stacking

It consists of the following components: an optical imaging system with small depth of field, a motion control system, and an image acquisition system. The motion control system consists of a longitudinally adjustable platform, a screw, a precise stepper motor (PK545, Japan Oriental Motor Co., Ltd.), a driver (DFC5107P), and a motion control card (DMC2210, Shenzhen Leadshine Control Technology Co., Ltd.). The motion control system is used to precisely control the uniform motion of the camera and the objective lens, to realize the stratified scanning imaging with different focal depth. The illumination system is composed of a ring LED light source and a blue light source [31]. The ring LED light source (OPT-RI7000) has the advantages of uniform and large area illumination. It can highlight the edge and is suitable for illuminating the measured object with a complex structure [32]. The CCD area array camera (MVC3000F, Beijing Microview Science & Technology Co., Ltd.) has a pixel size of $3.2 \mu m \times 3.2 \mu m$ and a maximum resolution of $1536 \times 2048$. The frame rate is 30 fps when the resolution is $768 \times 1024$. The magnification of the objective is five times.

The PC with Intel Corei5-4300 CPU (2.6GHz), 12GB RAM and Windows 7 (64 bit) operating system is employed in the experiment. The programming software is MATLAB 2019a. The image size collected by the experimental system is $768 \times 1024$.

**B. IMAGING SEQUENCES ACQUISITION OF FOCUS STACKING**

The proposed method was validated by acquiring focus stacking image sequences of a threaded hole and a screw with a diameter of approximately 1 mm. The thread hole and screw were placed on the experimental platform and tilted by $\theta$ angle, and the stepper motor controlled the image acquisition and illumination system to move up and down to collect hierarchically focus stacking image sequences. A total of 106 threaded hole images and 127 screw images were collected. The acquisition schematic diagram is shown in Fig. 5.

**C. FOCUS AREA SEGMENTATION OF SOURCE IMAGE**

The segmentation process of focus area of focus stacking image sequences is shown in Fig. 6. The focus polygon area is first segmented from the original image and then converted into an approximate focus rectangle area as shown by the red box area. The size and position of the extracted focus area are automatically adjusted according to the actual focus stacking images. The focus area segmentation method has good applicability for this experimental system.

**D. TIME AND EFFICIENCY ANALYSIS OF FEATURE DETECTION**

Two consecutive images are individually selected from the two image sequences to analyze the time and efficiency of feature point detection and matching, as shown in Fig. 7. The proposed method including extraction of focus area and feature detection and registration, was compared with the SIFT, SURF, BRISK [13], [33], and ORB [12] algorithms for image registration. The experimental results in Table 1 is the average value of experimental results measured repeatedly for three times. The results show that the average registration time of the proposed method decreases by more than 60%, and the number of extracted feature points decreases by only 5%, when compared with the SIFT algorithm. SURF algorithm is very fast, however, it can only extract very few feature points in the threaded hole images, which is not suitable for image registration in this work. BRISK algorithm has poor ability to detect and match feature points, and fails to match the threaded holes. There are some mismatches in matching the threaded hole images using the ORB algorithm.
The registration results of these methods are shown in Fig. 8.

E. COMPARISON OF IMAGE REGISTRATION ACCURACY
Image registration accuracy analysis was performed using SIFT, SURF, BRISK, ORB, and the proposed method, respectively. The source image was translated, rotated, and scaled, and then the results were registered with the source image. The simulation results in Table 2 show that the proposed method has the highest accuracy, and the error is within 0.2 pixels. The proposed method has greatly improved the accuracy and efficiency of image registration.

| Images   | Method  | keypoints number1 | keypoints number 2 | Matching Number | Registration time (s) |
|----------|---------|-------------------|--------------------|-----------------|-----------------------|
| Threaded hole | SIFT | 81                 | 91                 | 44              | 5.131                 |
|           | SURF   | 3                  | 7                  | 3               | 0.192                 |
|           | BRISK  | 1                  | 1                  | 1               | 1.013                 |
|           | ORB    | 189                | 227                | 98              | 0.118                 |
|           | Proposed method | 77                | 81                 | 36              | 1.72                  |
| Screw    | SIFT   | 202                | 241                | 79              | 5.142                 |
|           | SURF   | 36                 | 34                 | 22              | 0.210                 |
|           | BRISK  | 28                 | 19                 | 3               | 1.009                 |
|           | ORB    | 503                | 591                | 123             | 0.145                 |
|           | Proposed method | 197               | 224                | 74              | 1.361                 |
Moreover, it can solve the fast registration problem of two images with offset, rotation and scale change, and the result achieves the registration accuracy of sub-pixel level.

### TABLE 2. Comparison of image registration accuracy.

| Images | Method                  | Real parameter | SIFT    | SURF    | BRISK   | ORB     | Proposed method |
|--------|-------------------------|----------------|---------|---------|---------|---------|----------------|
| Threaded hole | Horizontal offset/pixel | 20             | 19.52   | 20.00   | Failure | 19.95   | 19.96          |
|         | Vertical offset/pixel   | 15             | 15.30   | 15.24   | Failure | 15.18   | 15.18          |
|         | Rotation(°)             | 10             | 9.99    | 10.48   | Failure | 9.97    | 10.05          |
|         | Scale/pixel             | 0.85           | 0.85    | 0.85    | Failure | 0.85    | 0.85           |
| Screw  | Horizontal offset/pixel | 20             | 20.45   | 20.00   | 19.99   | 20.04   | 20.16          |
|         | Vertical offset/pixel   | 15             | 14.97   | 15.01   | 14.88   | 14.89   | 15.02          |
|         | Rotation(°)             | 10             | 10.02   | 9.78    | 10.49   | 10.15   | 10.01          |
|         | Scale/pixel             | 0.85           | 0.85    | 0.85    | 0.85    | 0.85    | 0.85           |

**F. RESULT OF SEQUENCE IMAGE FUSION**

Using the Laplacian pyramid fusion algorithm based on Sobel gradient proposed in this paper, the registered screw hole and
screw image sequences are fused. The number of layers in the pyramid is set to four layers. The fully focused images of the threaded hole and the screw with the large depth of field obtained after image fusion is shown in Fig. 9. The blurred part of the lower part of Fig. 9 (a) is always in an out-of-focus state during the layered scanning imaging process, so it is normal for the blurred region to exist in the fused image. Overall, the images have high clarity and abundant information. The results show that the proposed image registration and fusion method is effective.

VI. CONCLUSION

A monocular 3D focus stacking microscopic measurement system platform was built. A fast-automatic and high-precision registration method based on SIFT was proposed. The image registration method based on self-adaptive and fast segmentation of the focus area improves dramatically the registration speed under the premise of ensuring the registration accuracy. RANSAC algorithm was used to filter the mismatched points and obtain more accurate matched point pairs. The registered image is re-projected to eliminate the phenomenon of homonymous pixel offset, rotating, and scaling by solving the transformation matrix. The Laplacian pyramid method was used to fuse the image sequences. The experimental results show that the image registration method based on focus area segmentation is 60% faster than the traditional SIFT method, which improves the accuracy and speed of image registration. The results of sequence image registration and fusion meet the field testing requirements. The proposed method can achieve better registration effect for the image with relatively concentrated or continuous focus areas. However, it cannot show its advantages for the image registration with the focus area scattered throughout the image.

REFERENCES

[1] P. Trouvé-Peloux, F. Champagnat, G. Le Besnerais, and J. Idier, “Theoretical performance model for single image depth from defocus,” J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 31, no. 12, pp. 2650–2662, Dec. 2014.

[2] F. Helmlt, R. Danzl, M. Prantl, and M. Grabner, “Ultra high speed 3D measurement with the focus variation method,” Fringe. Berlin, Germany: Springer, 2014, pp. 617–622.

[3] J. Włodek, K. J. Gofron, and Y. Q. Cai, “Achieving 3D imaging through focus stacking,” in Proc. 15th Int. Conf. Synchrotron Radiat. Instrum. (SRM), 2019.

[4] Q. Qian and B. K. Gunturk, “Extending depth of field and dynamic range from differently focused and exposed images,” Multidimensional Syst. Signal Process., vol. 27, no. 2, pp. 493–509, Apr. 2016.

[5] G. Wang, J. Yu, and Y. Hou, “An image registration algorithm based on features point and cluster,” in Proc. 5th Int. Cong. Image Signal Process., Oct. 2012, pp. 836–840.

[6] D. G. Lowe, “Distinctive image features from scale–invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, Nov. 2004.

[7] D. Lowe, “Object recognition from local scale-invariant features,” in Proc. 7th IEEE Int. Conf. Comput. Vis., vol. 2, 1999, pp. 1150–1157.

[8] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2, Oct. 2002.

[9] E. Rosten and T. Drummond, “Machine learning for high-speed corner detection,” in Computer Vision. Berlin, Germany: Springer, 2006, pp. 430–443.

[10] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359, Jun. 2008.

[11] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, “BRIEF: Binary robust independent elementary features,” in Proc. Eur. Conf. Comput. Vis. Berlin, Germany: Springer, 2010, pp. 778–792.

[12] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 2564–2571.

[13] S. Leutenegger, M. Chli, and R. Y. Siegwart, “BRISK: Binary robust invariant scalable keypoints,” in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 2548–2555.

[14] M. Kashif, T. M. Deserno, D. Haak, and S. Jonas, “Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment,” Comput. Biol. Med., vol. 68, pp. 67–75, Jan. 2016.

[15] Y. Ke and R. Sukthankar, “PCA–SIFT: A more distinctive representation for local image descriptors,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), vol. 2, Nov. 2004, pp. 506–513.

[16] J. Dai-Hong, D. Lei, L. Dan, and Z. San-You, “Moving–object tracking algorithm based on SIFT and optimization for underground coal mines,” IEEE Access, vol. 7, pp. 35556–35563, 2019.

[17] C. Duan, X. Meng, C. Tu, and C. Yang, “How to make local image features more efficient and distinctive,” IET Comput. Vis., vol. 2, no. 3, p. 178, 2008.

[18] J.-M. Morel and G. Yu, “ASIFT: A new framework for fully affine invariant image comparison,” SIAM J. Imaging Sci., vol. 2, no. 2, pp. 438–469, Jun. 2009.

[19] A. Laraqui, A. Baataoui, A. Saaidi, A. Jarrar, M. Masrar, and K. Satori, “Image mosaicing using Voronoi diagram,” Multimedia Tools Appl., vol. 76, no. 6, pp. 8803–8829, Mar. 2017.

[20] Q. Zhou, W. Yang, G. Gao, W. Ou, H. Lu, J. Chen, and L. J. Latecki, “Multi-scale deep context convolutional neural networks for semantic segmentation,” World Wide Web, vol. 22, no. 2, pp. 555–570, Mar. 2019.

[21] D. Cong, Q. Zhou, J. Cheng, X. Wu, S. Zhang, W. Ou, and H. Lu, “CAN: Contextual aggregating network for semantic segmentation,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), May 2019, pp. 1892–1896.

[22] L. G. Roberts, “Machine perception of three-dimensional solids,” in Optical and Electro-Optimal Information Processing. Cambridge, MA, USA: MIT Press, 1965.

[23] T. Lindeberg, “Scale-space theory: A basic tool for analyzing structures at different scales,” J. Appl. Statist., vol. 21, nos. 1–2, pp. 225–270, Jan. 1994.

[24] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, “Low illumination underwater light field images reconstruction using deep convolutional neural networks,” Future Gener. Comput. Syst., vol. 82, pp. 142–148, May 2018.

[25] M. A. Fischler and R. C. Bolles, “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” in Readings in Computer Vision: Issues, Problem, Principles, and Paradigms, M. A. Fischler and O. Firschein, Eds. San Francisco, CA, USA: Morgan Kaufmann, 1987, pp. 726–740.

[26] J. Matas and O. Chum, “Randomized RANSAC with sequential probabil-ity ratio test,” in Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV), Nov. 2005, pp. 1727–1732.

[27] G. Cao, S. Ruan, Y. Peng, S. Huang, and N. Kwok, “Large-complex-surface defect detection by hybrid gradient threshold segmentation and image registration,” IEEE Access, vol. 6, pp. 30235–30246, 2018.
J. Sun, Q. Han, L. Kou, L. Zhang, K. Zhang, and Z. Jin, “Multi-focus image fusion algorithm based on Laplacian pyramids,” J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 35, no. 3, pp. 480–490, 2018.

X. Hao, H. Zhao, and J. Liu, “Multifocus color image sequence fusion based on mean shift segmentation,” Appl. Opt., vol. 54, no. 30, p. 8982, Oct. 2015.

P. Burt and E. Adelson, “The Laplacian pyramid as a compact image code,” IEEE Trans. Commun., vol. 31, no. 4, pp. 532–540, Apr. 1983.

S. Serikawa and H. Lu, “Underwater image dehazing using joint trilateral filter,” Comput. Elect. Eng., vol. 40, no. 1, pp. 41–50, Jan. 2014.

R. Lu, A. Wu, T. Zhang, and Y. Wang, “Review on automated optical (visual) inspection and its applications in defect detection,” Acta Opt. Sinica, vol. 38, no. 8, pp. 23–58, 2018.

P. P. Rebouças Filho, F. D. L. Moreira, L. B. Marinho, and J. S. Almeida, “Novel and powerful mosaic constructor for territorial analysis using mobile robots via binary robust invariant scalable keypoints,” Measurement, vol. 114, pp. 448–470, Jan. 2018.

ANG WU received the B.S. degree from Anhui Engineering University, in 2000, and the M.S. degree from the Hefei University of Technology, in 2005. He is currently pursuing the Ph.D. degree with the School of Instrument Science and Opto-Electronics Engineering, Hefei University of Technology. He is currently the Lecturer of the College of Mechanical and Electrical Engineering, Henan Agriculture University. His research interests are machine vision and optical metrology.

YANQIONG SHI received the Ph.D. degree in precision engineering from the Hefei University of Technology, China, in 2012. She is currently an Associate Professor with the School of Mechanical and Electrical Engineering, Anhui Jianzhu University, China. Her research interests include 3D machine vision and optical metrology.

RONGSHENG LU received the Ph.D. degree in precision instrument engineering from the Hefei University of Technology, China. He then joined the Key Laboratory Precision Measuring Technology and Instrument, Tianjin University, China, as a Postdoctoral Researcher, and later an Associate Professor in image processing and 3D machine vision inspection. From 2001 to 2002, he was a Senior Research Associate in robot vision with the City University of Hong Kong. From 2002 to 2006, he worked as an Associate Researcher at Imperial College London and as a Research Fellow at the University of Huddersfield, U.K., in automated optical inspection. His research interests include machine vision and optical metrology. He is currently a Professor with the School of Instrument Science and Opto-electronics Engineering, Hefei University of Technology, and the Chairman of the Metrology Society of Anhui, China.

ZILONG ZHANG received the B.S. degree in photoelectric information engineering from the Hefei University of Technology, China, in 2015, where he is currently pursuing the Ph.D. degree with the School of Instrument Science and Opto-Electronics Engineering. His research interests are machine vision and optical metrology.