Tunnel image stitching based on geometry and features

ZhaoYuan Wang¹, Li He²,⁴, Taosheng Li³, Jian Tao¹, Chengxue Hu¹ and Mochuan Wang¹

¹School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, China
²College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China
³Anqing Vocational and Technical College, Anqing, China

⁴Email: heli@szu.edu.cn

Abstract. In the fast tunnel detection of multi-camera, a disease distributed in multiple images is prone to be misidentified as multiple diseases, which affects the evaluation of the status of the tunnel. This paper proposes a high-precision stitching method driven by data and scene based on multi-camera sequence images. Firstly, geometric rough calculation is performed to generate the theoretical stitching mode using the geometric positional relationship between the cameras in the scene and the image relationship is renewed after theoretical stitching. Secondly, feature points are extracted and matched for adjacent overlapping images by SURF algorithm. Therefore, pixel-level data registration is performed to achieve the image stitching. Finally, an integrated stitching mode is proposed based on the theoretical stitching and pixel-level data registration, which utilizes the stitched sequence images with high physical resolution to extract cross-section information. Practical results show that the method can achieve image stitching of the tunnels with high accuracy and good reliability.

1. Introduction

Due to factors such as climate, hydrology and geology, different diseases will gradually appear in tunnel, which affects the use of the tunnel [1]. At present, traditional detection methods based on eyes or auxiliary instrument have low efficiency and high labour cost, and are affected by objective conditions such as skylight time, lighting, visual distance, etc. [2].

In the tunnel detection, the displacement between the cameras due to the change of the carrier poses and the calibration accuracy of the camera's collinear and coplanarity, cracks and other diseases are misaligned in multiple pictures. Without high-precision stitching, one disease will be recognized as multiple diseases and it will affect the evaluation of the tunnel status. Yang et al. studied a sequence images registration method combining spatial and frequency domains, improved the estimation of translation parameters, solved the problem of automatic image sequencing, and reduced the amount of algorithm calculations [3]; Wang et al. studied a sequence image stitching algorithm suitable for light field depth images, which solved the problem of low efficiency of light field sequence images stitching [4]; Ren et al. studied the sequence image registration method based on cylindrical surface mapping, with the camera rotating at a constant speed when collecting data [5].

Current common image stitching methods which can achieve good results are based on feature points. The tunnel lining has few features and some images are of poor quality due to the collection environment. It is difficult to solve the tunnel image stitching problem in actual engineering by only
relying on image features. Therefore, it is necessary to study the stitching algorithm for the tunnel image. Because the images collected by the tunnel automatic inspection equipment are regular, the positional relationship between the tunnel lining images is regular. The position of the lining image in the panorama is regular, and it can be positioned by geometric relationship. In addition, some good-quality images have enough feature points. Image stitching can be achieved based on image features.

This paper proposes a tunnel image stitching method based on geometry and features. Image stitching based on geometry is called scene-driven registration. Image stitching based on features is called image data-driven registration. When the tunnel lining image cannot be detected enough feature points, the relationship between the cameras is used to achieve image registration; When the image quality is good, the image stitching algorithm based on feature points is used to extract feature points for image registration.

2. Related work

At present, the usual tunnel disease detection methods are directly using discrete lining images for disease detection. However, when assessing the condition of the tunnel, the results of an independent assessment of a large crack are different from the assessment of its division into several small cracks. Therefore, it is necessary to stitch the lining images in a large area. Although some tunnel inspection institutions use manual methods such as ZOYON, manual stitching has low efficiency and poor accuracy, which is not suitable for large-scale tunnel inspection. At the same time, some institutions only stitch the adjacent images with cracks, which cannot meet the detection task well.

The research data comes from the tunnel fast detection equipment developed by ZOYON Technology Co., Ltd. This equipment integrates 16 high-resolution CCD cameras on a rigid support. According to the distance from each camera to the lining, cameras with different focal lengths were selected to ensure the physical resolution. In order to provide object distance and positioning estimation for each camera, the equipment integrates a high-frequency, high-precision LIDAR, and uses a static calibration method to determine the object distance calculation method and the camera measuring angle, as shown in Figure 1.

![Figure 1. Testing equipment.](image)

The measurement model of the detection system is shown in Figure 2. The lining image of half tunnel is collected every time. Suppose the clear height of the tunnel is \( H \), the height from the sensor platform to the top surface is \( H_0 \), the maximum distance from the sensor to the lining level is \( L \), the angle of view of the camera is \( \alpha \), and the angle between the LIDAR and the centre line of the camera is \( \theta \).

![Figure 2. Principles of tunnel detection.](image)
In the Figure 2, the focal length $f$ and resolution of the camera is given, $h$ is measured by LIDAR, and $b$ is pre-calibrated, the object distance $d$ corresponding to the camera can be calculated.

$$\cos(\theta) = \frac{b^2 + h^2 - d^2}{2bh}$$

In the camera imaging triangle AOB:

$$AB = 2d \tan \alpha$$

If the camera resolution is $r \times r$, the object resolution $R$ is:

$$R = \frac{AB}{r}$$

The multi-camera centre line is collinear to ensure the maximum utilization of the multi-camera data, and the multi-camera coplanarity can ensure that the multi-camera data has a common measurement attitude, which is conducive to the processing of data. At the same time, the calculation of camera images needs to measure the object distance. The calibration of the relationship between the cameras or the cameras and the LIDAR can effectively solve this problem [6].

3. Tunnel lining image stitching

Because there are few feature points of the tunnel lining and there may be no matching feature points between adjacent images, there is a mismatch by relying on the image features for stitching. In order to ensure cross-sectional image stitching, the rough location of images is obtained by using the geometry calibration information. On the basis of positioning, the relationship of adjacent images is calculated, feature point matching is performed on the two intersecting images, and registration is performed to achieve high-precision stitching of images.

3.1. Scene-driven registration

The measurement multi-sensor is installed on a rigid support, and the relationship between the camera and the relationship between the camera and the LIDAR are calibrated with high precision in advance. The LIDAR data is used for 3D tunnel modelling to obtain the object distance of cameras. Each camera selects a different focal length lens to ensure the physical resolution of the data.

In order to model the tunnel, the noise in the point cloud needs to be removed firstly. The auxiliary structures such as ventilation and lighting will cause the auxiliary structure noise. When the fast tunnel detection system collects point cloud data, the LIDAR rotates and scans and the detection system travels forward to obtain spiral point cloud data. The point cloud data is stored in units of circles and is referred to a meta tunnel. Since the detection system cannot maintain an absolute straight line during driving, there is sway between the laps. In order to solve the problem, this paper uses the ICP algorithm [7] to register multi-circle meta tunnels so that multi-circle point clouds can overlap. In order to effectively reduce the noise of certain auxiliary structures, the number of overlapping circles should be determined by the interval of the auxiliary structures in the tunnel and the scanning speed of the LIDAR. Because the auxiliary structures appear at intervals and different auxiliary structures are distributed in different locations in tunnel, after overlapping multiple circles of point clouds, there will be intact parts of other circles to fill the depression which causes by the auxiliary structures. Since the point cloud of the auxiliary structure will only sag inwards, finding the outermost periphery of the multi-circle point clouds that are overlapped together, so that the point cloud after noise removal can be obtained. In order to get the noise-removed tunnel, the tunnel is segmented, a circle of standard meta tunnels is found in each segment, then the standard meta tunnels are repeated to rebuild the tunnel, and all the segments are recombined.

In order to perform 3D reconstruction of the tunnel, the point cloud is flattened firstly. Then use the Delaunay triangulation algorithm [8] to construct a triangular relationship network of the flattened point cloud, and apply the triangular relationship network to the 3D point cloud to realize the reconstruction of
the tunnel. For a circle, rolling it along the horizontal plane can find the position where each point on the circumference is pressed on the horizontal plane, as shown in Figure 3.a.

\[ P(x, y) \] is on the circle with radius \( r \). The flattened coordinate becomes \( Q(l, 0) \), then the coordinate conversion relationship is:

\[
\begin{align*}
     l &= r \times \arccos\left(1 - \frac{x^2 + y^2}{2r^2}\right), (0 \leq \theta \leq \pi) \\
     l &= 2\pi - r \times \arccos\left(1 - \frac{x^2 + y^2}{2r^2}\right), (\pi < \theta \leq 2\pi)
\end{align*}
\] (4)

For discrete point cloud, similar ideas can also be used to flatten it. In order to expand a circle of discrete arc-shaped point cloud, you can use line segments to connect the points between the discrete points, and then roll it along the x-axis to flatten the point cloud, as shown 3.b. The three-dimensional model of the tunnel is shown in Figure 3.c. On this basis, the geometric calibration data is used to establish the camera scene model.

The LIDAR adopts point-by-point scanning at a set angular interval. A laser line that almost coincides with the direction of the camera's main optical axis can be extracted in one scan cycle. By binding each camera to a laser line, the intersection of the main optical axis and the tunnel lining can be determined which is the centre of the image. Therefore, it can be determined that the centre of each image corresponds to the position in the tunnel lining. At the same time, the object distance of the
camera can be solved by the distance between the LIDAR and the lining and the positional relationship between the LIDAR and the camera. The spatial resolution of each image can be calculated through object distance to evaluate the image quality. According to the positional relationship between the centre points of the images, the theoretical position of the images in the panorama can be located, and the spatial relationship between adjacent images can be calculated that is separated, intersected or tangent. Thus, the rough stitching of all the images is obtained. The principle is shown in Figure 4. The $(x_i, y_i)$ is the position of the centre point of the image.

The CCD sensor will be calibrated by the manufacturer before it is sold, so there is an error in the sensor when it collects images. The detection system moves fast during image acquisition, and the sensor platform will generate vibration. Vibration can cause errors in image positioning. The detection equipment uses INS/GPS to ensure that the driving track of the equipment is kept as straight as possible. Shaking of the trajectory during the driving of the device will cause image positioning errors. If the adjacent images do not intersect, the two images stop stitching. The number of pixels that two images overlap in the panorama can be calculated using the following formula.

$$s = \frac{d_1}{2} + \frac{d_2}{2} - d$$  \hspace{1cm} (5)

In the formula 5, the $s$ represents the overlapping width of the two images in the panorama, the $d_1$ represents the width of the previous image in the panorama, the $d_2$ represents the width of the next image in the panorama, the $d$ represents the horizontal distance between the centre points of two adjacent images.

3.2. Image data-driven registration
The rough calculation can locate the sequence images, but due to the error of the sensor platform and the calibration parameters, there is always a misalignment of the adjacent camera images. For overlapping image data, feature points matching can be used for data registration if there are features. If the image quality is good, a high-precision image registration transformation matrix can be obtained. This paper uses the SURF algorithm [9] to solve the transformation matrix.

The SURF algorithm uses Hessian matrix to extract feature points [10]. To speed up and simplify the calculation process, a box filter is used instead of a Gaussian filter, and an integrated image is used in the calculation. A box filter template is used to establish the scale space, and the image scale is equal to the box filter scale [11]. The smallest scale image is filtered by a 9×9 box filter. And the other filters are expanded by this filter.

By setting the threshold, pixels with low determinant values and pixels with non-maximum values of the Hessian matrix will be removed. In order to meet the scale invariance, it is necessary to compare not only with the neighbour pixels of the scale image, but also with the image pixels of the adjacent scale. Then use interpolation to determine the feature point location.

The SURF algorithm descriptor contains several sub-regions, each sub-region is represented by a 4-dimensional feature vector, and 16 sub-regions form a 64-dimensional feature vector. Based on the descriptor, a determinant trace that approximates the Hessian matrix is calculated, and if the traces have same sign, it is preliminarily determined that the two points match. The Euclidian distance between the feature descriptors of the two matching points was further calculated to detect the degree of similarity. For the matching points with high degree of similarity, RANSAC algorithm [12] was used to correct the matching results, and finally the matching success point was determined.

After the registration of cross-sectional image is completed, the positional relationship between the centre points of the images are re-determined. Due to the fast movement of the detection equipment during the detection process, and the camera bracket is not an ideal rigid body, the centre points of the images will move slightly, so the centre points of the images are not ideally arranged on a straight line. This is where the image should be in the real position in the panorama. As shown in Figure 5.
Based on the above stitching results, the position relation of the sequence images in the panoramic image is calculated to obtain the effective interval in the high direction. Then, the overlapping relation and physical resolution of adjacent images are calculated. In tunnel detection, high-resolution data should be maximized. Therefore, the overlapping part of adjacent images with high resolution is reserved. The stitching result of the cross-section image is shown in Figure 6.

**Figure 6.** Example of actual stitching.

4. **Experiment**

In this paper, the tunnel fast detection system (ZOYON-TFS) is used to collect experimental data. The image resolution is greater than 0.3mm, and the detection system adopts 16 cameras. A total of 3 tunnel data are used in this experiment. The experiment extracts data from it, and uses different algorithms to test the stitching of 16 images.

Because the 16 cameras are fixed on a common rigid support, as long as the tunnel cross section does not change, the sensor has no relative displacement, and the field of view of each camera is fixed on the cross section. Ideally, the coordinates of the cross-sectional image can be located using the camera's fixed parameters and LIDAR measurements. The LIDAR continuously measures to obtain a complete three-dimensional point cloud of the tunnel. The three-dimensional point cloud can be used to produce a tunnel cross-section. The curved surface is flattened and used as a reference coordinate space to locate the sequence image on the cross-section. The positioning test results are shown in the following example as shown in Figure 7.

**Figure 7.** Geometry stitching test.

In the cross-sectional image, two adjacent images are a group which is numbered A–O, and a total of 15 groups of stitching results based on features are shown in the following table.

| Sample number | A   | B   | C   | D   | E   | F   | G   | H   | I   | J   | K   | L   | M   | N   | O   |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| SIFT          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| I             | 3551| 5332| 11594| 7357| 822 | 2016| 450 | 6259| 5868| 8702| 9257| 5322| 3389| 1549| 1070|
| II            | 80% | 86% | 89% | 91% | 91% | 92% | N/A | N/A | 86% | 100%| 78% | 86% | 82% | 98% | 100%|
| SURF          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| I             | 580 | 352 | 682 | 1119| 50  | 84  | 73  | 251 | 315 | 593 | 1478| 1341| 895 | 334 | 241 |
| II            | 77% | N/A | 67% | 83% | 22% | N/A | N/A | N/A | 68% | 55% | 77% | 77% | 51% | 58% |     |
| FAST          |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| I             | 13  | N/A | N/A | 49  | 67  | 202 | 20  | 364 | 399 | 723 | 298 | 55  | 2   | 2   | N/A |
| II            | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | 51% | 58% | N/A | N/A | N/A | N/A | N/A |

In the Table 1, the line of “I” is the number of feature points, The line of “II” is the matching accuracy of feature points. For the test data, Table 1 shows the number of feature points detected using different algorithms and the registration accuracy. The accuracy rate represents the number of feature
points which are excluded mismatching through RANSAC divided by the total number of feature points. “N/A” means registration failed based on feature points. As can be seen from Table 1, both SURF and SIFT algorithms can obtain a large number of feature points, and the number of feature points solved by the FAST algorithm is small, which is not suitable for solving the transformation matrix. In addition, when there are many feature points, the SIFT algorithm has two images that match incorrectly which results in the inability to solve the correct transformation matrix; the SURF algorithm has five images transformation matrices that cannot be solved; the FAST algorithm only can solve transformation matrices of two images. For images that can correctly solve the transformation matrix, the matching accuracy of the SURF algorithm is less than 50% for one image, and higher than 50% for the rest.

At the same time, Counting the running time of the algorithm can find that the SITF algorithm requires 241.2219s to solve the transformation matrix of the cross-sectional image, while the SURF algorithm only needs 5.7680s. Comparing the running time and efficiency of the algorithm, it can be found that the SIFT algorithm takes about 42 times the SURF algorithm. Due to the large amount of tunnel image data and the similar effect of SURF and SIFT for tunnel images with good quality, the SURF algorithm is more suitable for tunnel lining image stitching.

5. Discussion
By comparing the different stitching results, it can be received that for the overlapping area of the image, due to the different texture characteristics and overlapping area, the accuracy of the image registration is different. The more obvious the texture is and the larger the overlap area is, the higher the registration accuracy of feature points will be; otherwise, the lower the accuracy will be. And comparing the two stitching methods could find that the two stitching methods are different. For scene-driven stitching, there is a deviation at the seam. From the observable line type, point type and other characteristics, this deviation varies from a few pixels to a dozen pixels.

6. Conclusions
In this paper, for the fast detection of tunnel automation, an image stitching method based on geometry and features is designed. The scene-driven registration mode needs not the specific images data and is only related to the geometric relationship. It is not affected by the images quality and provides a reliable guarantee for the cross-sectional image stitching; At the same time, using image feature points matching to achieve data-driven registration, the data-driven registration mode relies on images data, which provides the possibility of realizing high-precision registration for certain images. Finally, the transformation relationship obtained by the two methods is fused to achieve the stitching of tunnel lining images. In the process of image stitching, adjacent images with overlapping parts and good features can achieve high-precision stitching. Adjacent images without overlapping parts or without good features can be stitched by using a scene-driven approach.

Acknowledgement
This research is funded by the Wuhan Science and Technology Plan Project (No.2016070204020160), and the Key Projects of Natural Science Research in Universities of Anhui Education Department (No.KJ2019A1193), supported this work. The authors are grateful to the editor and the anonymous reviewers for their insightful comments and suggestions.

References
[1] Wang L H 2019 Talking about the maintenance management of highway tunnel during operation J. Shandong Industrial Technology 7 96
[2] Li J C, Zhang C B, Chai X S et al. 2018 Research on Crack Detection System of Tunnel Lining Based on Image Recognition Technology J. Railway Engineering 58 20-24
[3] Yang L, Cao J Z, Tang L N et al. 2014 Optimized design of panoramic image automatic stitching algorithm J. Infrared and Laser Engineering 43 985-90
[4] Wang W F, Zhang Y X, Chen Y et al. 2018 Large-field image stitching algorithm based on light
field depth sequence J. Acta Optica Sinica 38 268-75

[5] Ren G, Peng D L and Gu Y 2017 Fast image stitching algorithm based on cylindrical surface mapping J. Application Research of Computers 34 3473-76

[6] Zhang Z Y 2000 A flexible new technique for camera calibration J. IEEE Transactions on pattern Analysis and Machine Intelligence 22 1330-34

[7] Chen L, Wu Y, Tong L H et al. 2015 Hierarchical Registration Method for Airborne and Vehicle LiDAR Point Cloud J. Remote Sensing 7 13921-42

[8] Cohen-steiner D and Da F 2004 A Greedy Delaunay-based surface Reconstruction Algorithm J. The Visual Computer: International Journal of Computer Graphics 20 4-16

[9] Bay H, Ess A, Tuytelaars T et al. 2008 Speeded-Up Robust Features(SURF) J. Computer Vision and Image Understanding 110 346-59

[10] Lu L J, Hu G P and Bao W X 2019 Image Fusion Technology Based on Improved SURF Algorithm J. Computer & Digital Engineering 47 1684-87

[11] Zhang M H, Yang Y Q and Jin B W 2019 Research on SURF Feature Matching Algorithm Based on Image Enhancement Technology J. Automation and Instrumentation 34 98-102

[12] Mehran F, Hamid H, kashani N et al. 2018 SC-RANSAC: Spatial consistency on RANSAC J. Multimedia Tools and Applications 78 9429-61