

Research on Image Stitching Technology for Road Surface Cracks

Chunlin He^{1,*}, Jiaye Wu², Yujie Yang³

¹Sichuan University of Science & Engineering, Zigong, Sichuan, 643000, China

²Southwest Petroleum University, Chengdu, 610500, China

³Sichuan University of Science & Engineering, Zigong, Sichuan, 643000, China

ABSTRACT

Image stitching refers to the combination of multiple images with a certain overlapping area rate acquired by the image acquisition device into one image, which has a full picture of each image. Image stitching technology is an important part in the field of digital image processing. With the increasing requirements for image quality and the limited resolution of ordinary cameras, it is impossible to obtain large angle of view and clear image at the same time. In this paper, the key technologies involved in pavement crack image stitching are intensively studied, including image feature point detection and matching, internal and external camera parameter estimation and optimization, and image fusion. The work done in this paper and the main innovations include: (1) Two main feature detection and description algorithms, SIFT algorithm and ORB algorithm, are studied. In view of the fact that the SIFT algorithm has redundancy when the Gaussian pyramid is established, this paper appropriately reduces the number of groups and maintains the number and stability of feature point detection. Combined with the advantages and disadvantages of the improved SIFT algorithm and ORB algorithm, and the performance of the two algorithms in the detection of pavement cracks, the improved SIFT algorithm is used as the feature point detection algorithm, and rBRIEF algorithm for describing the feature points in the ORB algorithm is used as the feature point description algorithm. The description algorithm is called the BSIFT algorithm. Experiments show that the robustness of BSIFT algorithm is similar to that of SIFT algorithm, which is better than ORB algorithm and faster than SIFT algorithm. (2) Image fusion algorithms are studied, including best seam-line search and image fusion. Aiming at the ghost phenomenon that may occur in overlapping regions, this paper proposes a best seam-line search algorithm based on distance transform.

KEYWORDS

Pavement Crack Image, Feature Point Detection, Image Fusion

1. INTRODUCTION

In the past 20 years, with the rapid development of computer technology, digital image processing technology has become an important research field. Due to the limitations of the camera's field of view, the content displayed in an image is not sufficient to represent the scene that the human eye can see. Therefore, in order to increase the field of view captured by the camera, people consider combining multiple images into a large photo through stitching processing. Image stitching technology has emerged as a result. Image stitching technology is one of the earliest and longest existing problems in the fields of computer vision and photogrammetry. The task of stitching includes aligning multiple overlapping images in the scene to a common reference plane to generate seamless high-resolution images. In the past decade, due to the development and popularization of digital

cameras, smartphones, and various image acquisition devices, a large amount of rich and diverse image data has been generated, and image stitching has achieved unprecedented development.

The digital image representation of road cracks can prepare for subsequent observation and research of cracks, but there is a contradiction between the field of view size and image resolution during the acquisition process: if the field of view is relatively large, the details of the cracks will be relatively blurred, which is not conducive to subsequent research; If the focal length of the camera is increased, due to the curved shape of road cracks, some cracks are very narrow, and only local cracks can be obtained. It is not possible to obtain the full view of the entire crack, which will affect the accuracy of subsequent crack length measurement. Therefore, for a field of view containing cracks, using a large focal length to capture multiple images and then concatenating them can obtain a high-precision large field of view crack image, which can be used to detect and analyze the cracks. This provides good support for future research and related applications.

In image stitching algorithms, image registration is the process of finding two or more images with overlapping areas for matching and then performing geometric transformations. It is the most important and critical step in the process of image stitching. According to the methods of image registration, there are currently two main types of registration methods - based on image grayscale and based on image features.

The registration method based on image grayscale utilizes the grayscale information between two images to measure the similarity of grayscale. This method does not require extracting local features of the image. The basic idea is to find whether there are similar grayscale values or areas of the same size in the overlapping areas of two images. The specific steps are: based on the internal information of the image, use search detection methods to find the point with the highest similarity, and determine the transformation parameters between the two images. Among them, the representative method proposed by Zhang[1] is to collect and divide the corresponding windows for the template image and the image to be registered, and calculate the mean and covariance within the windows. The correlation between the two is divided to obtain a similar score, with a score interval of (-1, 1), where -1 represents a complete mismatch between the two correlation windows and 1 represents a complete match. The registration method based on image features has been the research direction with the most attention, fastest development, and most remarkable achievements in the field of image registration in recent years. Feature based image registration methods have good robustness to image rotation, perspective transformation, grayscale changes, uneven lighting, and noise, making them widely used in real-world problems. Lowe was the first to propose a feature detection method based on scale invariance, the SIFT algorithm, in 1999 [2], which pioneered scale invariance detection and has a milestone significance in the field of digital image processing. He improved it in 2004 [3], making it very robust. The SIFT algorithm is invariant to rotation, scaling, and scale between images, and has strong noise resistance.

2. RESEARCH ON IMAGE REGISTRATION TECHNOLOGY

2.1. Image transformation model

The transformation of two images can be represented by the following equation:

$$f_2(x, y) = g[f_1(h(x, y))] \quad (1)$$

Among them, $f(x, y)$ and (x, y) represent the two images to be transformed, g represents the grayscale value transformation, and h represents the spatial coordinate transformation. The spatial coordinate transformation can be represented in matrix form specifically:

$$P' = HP \Leftrightarrow \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

Among them, $P(x, y)$ represents a point on the reference image, $P'(x', y')$ represents a point on the target image, and H represents the transformation matrix of the two images, also known as the homography matrix.

(1) Translational transformation

Translation transformation is the simplest geometric transformation of an image, involving only two parameters t_x and t_y ; If the horizontal movement is t_x and the vertical movement is t_y , the transformation relationship of coordinates is as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3)$$

(2) Rigid body transformation

Rigid body transformation refers to the transformation of an image that remains unchanged in size before and after transformation, mainly involving rotation and translation. If the rotation angle is set to θ , the transformation relationship of coordinates is as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta & t_x \\ \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4)$$

(3) Similarity transformation

The similarity transformation of images adds a scaling factor on the basis of rigid body transformation, and its coordinate transformation relationship is as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (5)$$

(4) Affine transformation

The affine transformation still maintains its original parallel relationship, but it is no longer a uniform scaling. After affine transformation, a rectangle will become a parallelogram, and its coordinate transformation relationship is as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & t_x \\ l_{21} & l_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (6)$$

(5) Perspective transformation

Perspective transformation is the most commonly used image geometric transformation. In perspective transformation, all 8 parameters are unknown variables (the last parameter h_{33} is fixed to 1), and the transformation relationship of its coordinates is expressed as equation (2).

2.2. Image feature detection algorithm

There are many manifestations of image features, such as points, edges, contours, and statistical features. The main focus of this section is on feature points. The so-called feature points, literally interpreted, can be considered as points in an image that have certain special features and feature structures, but currently there is no clear definition. Feature detection, which uses mathematical modeling to identify salient features from images, is the first and most important step in the image registration process, as its quality directly affects the subsequent stitching effect.[4] This section provides a detailed study and analysis of the SIFT feature detection algorithm and the ORB feature detection algorithm.

2.2.1. SIFT feature detection algorithm

The SIFT algorithm was proposed by David Lowe from the University of British Columbia in Canada in 1999 and improved in 2004.[5] It is an algorithm for detecting and describing local features in images. The feature points detected by this algorithm can maintain invariance to the scale and rotation of the image, and have strong robustness and stability to changes in image brightness and noise. Before using deep learning methods for image recognition, the SIFT algorithm was a state of the art method in the field of object recognition. In ISVRC2010 (ImageNet Large Scale Visual Recognition Challenges 2010), a team used the SIFT algorithm and LBP algorithm, combined with SVM, to win the championship, which is enough to demonstrate the strength of the SIFT algorithm. The SIFT algorithm mainly consists of four steps, namely detecting extreme values in scale space, accurately locating feature points, determining the direction and angle of feature points, and generating feature point descriptors.[6]

2.2.2. ORB feature detection algorithm

The ORB algorithm is based on the FAST corner detection algorithm and the BRIEF feature point description algorithm, and by adding scale pyramid and centroid method to the FAST algorithm to make it scale invariant and rotation invariant, it is called the OFAST algorithm; By rotating the coordinate axis of feature points in the BRIEF algorithm to achieve rotational consistency, it is called the BRIEF algorithm. The improved ORB algorithm is fast and suitable for most scenarios, especially in embedded fields such as smartphones, drones, and surveillance cameras, enabling real-time detection.[7]

The idea of the FAST corner detection method is very simple: the candidate point P is compared with the pixel values of 16 points on its radius of 3.[8] If there are a certain number of pixel points and the absolute difference between the pixel values of point P is within a certain range, then P is recognized as a feature point. The specific method is as follows:

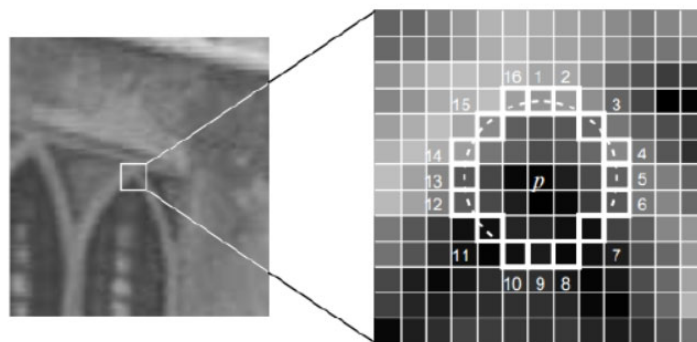


Figure 1. Schematic diagram of FAST corner detection

- (1) Select a pixel point P in the image. Assuming the intensity of the pixel is I_p , this point is the pixel to be recognized as a point of interest.
- (2) Set the threshold intensity value T. (For example, 20% of the pixel value of point P)

- (3) Obtain 16 pixel points on the circumference of pixel P.
- (4) If the pixel value of N (N is taken as 12) out of the 16 pixels on its circumference is higher or lower than the threshold T, then point P is considered a corner.
- (5) To speed up the algorithm, you can first compare the intensity I of pixels numbered 1, 5, 9, and 13 on the circumference.
- (6) If at least three pixels from the four pixel values I1, I5, I9, and I13 have an absolute difference from the pixel values of point P that does not exceed the threshold T, then P is not a corner.[9] Otherwise, if there are at least three pixels whose absolute difference from the pixel value of point P exceeds the threshold T, then check all 16 pixels and check if there are 12 consecutive pixels that meet this condition.
- (7) Repeat this process for all pixels in the image. The FAST corners obtained through the above method do not have scale invariance. In the ORB algorithm, by constructing a scale image pyramid, FAST corner detection is performed on each layer of the pyramid image. Due to the strong response of FAST corner detection to edges, feature points on the edges are unstable and therefore need to be removed. The FAST algorithm also uses Harris corner response, which uses a lower threshold Harris response to obtain more than N corner points. Then, by sorting the Harris response values of each corner point, the top N corner points are obtained.

2.2.3. BSIFT feature detection algorithm

The establishment of Gaussian difference pyramid relies on Gaussian pyramid, and the establishment of Gaussian pyramid requires determining multiple parameters. Firstly, the number of groups O in the Gaussian pyramid and the number of layers S in each group. The resolution of all images in each group of the Gaussian pyramid is the same, and the scale increases sequentially.[10] The next set of images is downsampled from the previous set, and the length and width of the images are reduced to half of the original. The number of groups O can be determined by the following formula:

$$O = \lfloor \log_2 \min(X, Y) - 2 \rfloor \quad (7)$$

The reason why the number of groups in the Gaussian pyramid is calculated as $O = \lfloor \log_2 \min(X, Y) - 2 \rfloor = 10$ from equation (7) is that the author of the SIFT algorithm, Lowe, suggests doubling the length and width of the input image before creating the Gaussian pyramid. The resolution sizes of each group in the Gaussian pyramid are: 2880 * 1920 * 1440 * 960 * 720 * 480 * 360 * 240 * 180 * 120, 90 * 60, 45 * 30, 22 * 15, 11 * 7, 5 * 3. According to the Gaussian pyramid created by the SIFT algorithm, the resolution of the last few layers of images is too small, such as images with resolutions of 5 * 3 and 11 * 7, whose features have almost disappeared. Searching for feature points is a redundant operation. Therefore, an equation is proposed here to determine the number of Gaussian pyramid systems: $O = \lfloor \ln \min(X, Y) - 2 \rfloor$. The original equation used a logarithmic function with a base of 2. Here, a logarithmic function with a natural constant e as the base is used to calculate the number of groups, resulting in 8 groups.

Taking into account the stability of the SIFT algorithm in feature point detection and the speed of the ORB algorithm in feature point detection, the improved SIFT algorithm is used for feature point detection, while the IBRIEF algorithm in the ORB algorithm is used for feature point description, which is called the BSIFT algorithm.

2.3. Research on Image Fusion Technology

For concatenated images, adjacent images can be aligned well in the overlapping boundary area. For example, if a crack passes through several images, the crack maintains its original integrity and continuity in the concatenated image. However, in terms of visual effects, due to exposure differences, the colors at both ends of the boundaries of each image are discontinuous, giving people a feeling of

block by block. So it is necessary to perform fusion processing on the spliced images to obtain a spliced image with uniform colors and smooth transitions.[11]

The best stitching line search is to find a stitching line in the overlapping area of the image, which is the line connecting the most similar pixels between two images in the overlapping area. After finding this seam line, there is no need to directly cover the overlapping area as before. Instead, the overlapping area is divided into two parts. On one side of the seam line, only the pixels on that side are selected in the overlapping area, while on the other side of the line, the pixels on the other side are selected. After finding the optimal stitching line, image fusion was performed to solve the ghosting problem.

3. EXPERIMENTAL RESULTS AND ANALYSIS

This experiment used road crack images captured in the field, as shown in the four images in Figure 2. These four images all contain cracks, among which the cracks in Figures (b) and (c) are relatively obvious, while the cracks in Figures (a) and (d) are relatively fine and have a low contrast with the surrounding background.

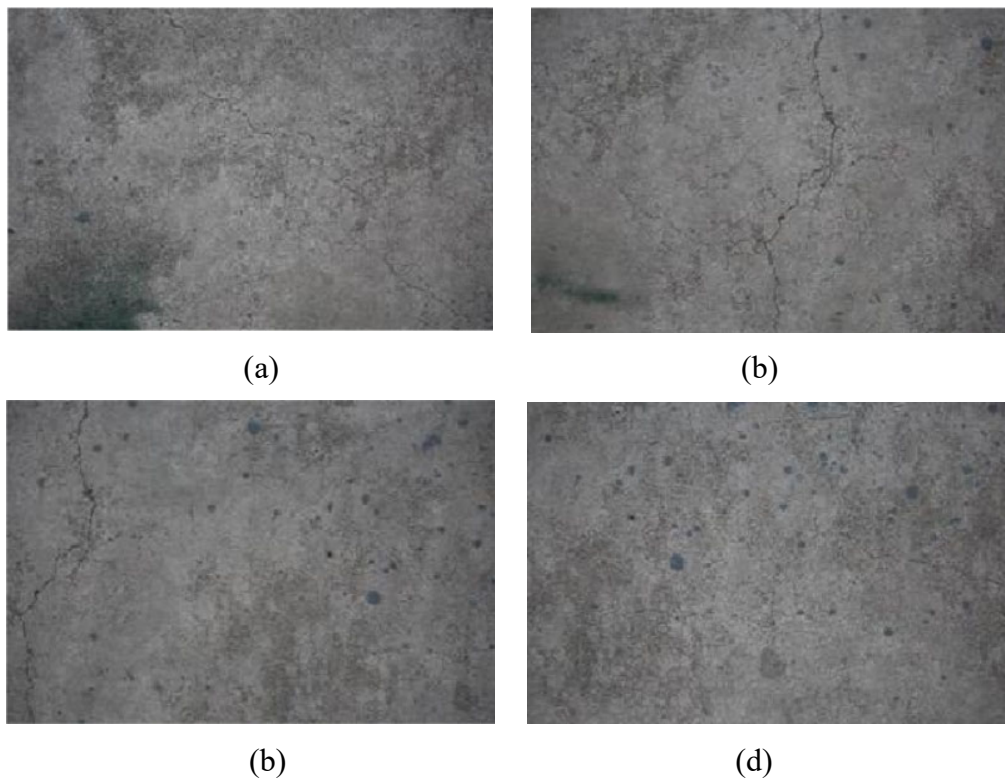


Figure 2. Original image of SIFT algorithm feature point detection experiment

Figure (a) in Figure 2 is adjacent to Figure (b), and there is a certain overlap area between the lower right part of Figure (a) and the upper left part of Figure (b). Figure 3 shows the paired point results of detecting and describing feature points using the SIFT algorithm in Figures (a) and (b). Figure 4 is obtained through the ORB algorithm for Figures (a) and (b), while Figure 5 is obtained through the BSIFT algorithm proposed in Section 2.2.3 for Figures (a) and (b). Due to the fact that the two images were captured by a translational camera, the rotation between images (a) and (b) is very small, and the lines connecting all matching point pairs should be approximately parallel. Among them, there are two mismatches in the result graph using the SIFT algorithm, but the vast majority are correct, with a total of 1430 pairs of inliers. Therefore, the homography matrix obtained through subsequent RANSAC algorithms is relatively correct. The results of using the ORB algorithm show that the paired lines in the graph are somewhat cluttered compared to Figure 3, with many mismatches. There

are only 607 pairs of inlier points in total, indicating that the correlation between the feature points obtained by the ORB algorithm is relatively high, making it easy for different points to match as well. And Figure 5 is almost the same as Figure 3, with two mismatches, a total of 1294 pairs of inliers, indicating that the BSIFT algorithm has high robustness and good performance in feature point matching.

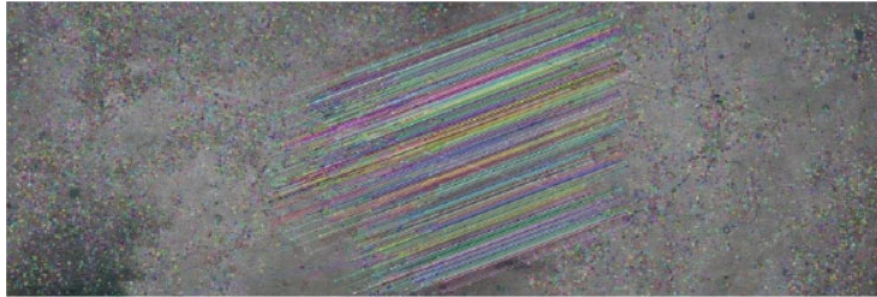


Figure 3. Schematic diagram of SIFT algorithm detection and matching

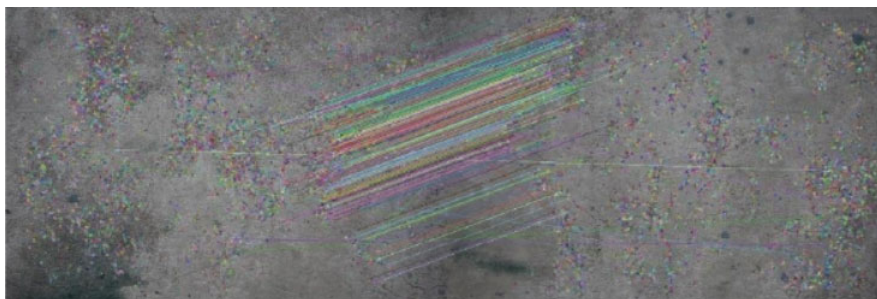


Figure 4. Schematic diagram of ORB algorithm detection and matching

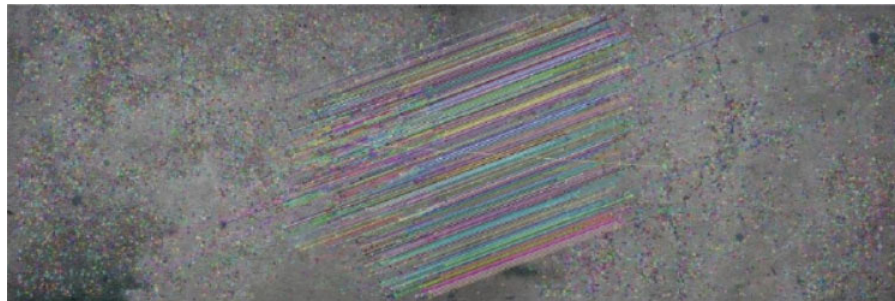


Figure 5. Schematic diagram of BSIFT algorithm detection and matching

For the four images in Figure 2, SIFT, ORB, and BSIFT algorithms were used for feature point detection, and the detection results are shown in the table below.

Table 1. Comparison of SIFT algorithm before and after improvement

		(a)	(b)	(c)	(d)
Original SIFT algorithm	Time consumption(s)	3.54	3.95	4.16	3.69
	Number of feature points	11658	11223	14785	12562
Improved SIFT algorithm	Time consumption(s)	3.02	3.62	3.28	3.22
	Number of feature points	11658	11223	14785	12562

Table 2. ORB algorithm feature point detection results

		(a)	(b)	(c)	(d)
ORB	Time consumption(s)	0.1757	0.1762	0.1408	0.2016

algorithm	Number of feature points	9286	9263	9182	9318
-----------	--------------------------	------	------	------	------

From Table 2, it can be seen that the detection speed of the ORB algorithm is nearly two orders of magnitude faster than the SIFT algorithm. Due to the maximum number of detected feature points being limited to 10000, the number of detected feature points in all four images is below 10000. However, due to the limitations of the ORB algorithm in detection accuracy, the SIFT algorithm still outperforms the SIFT algorithm in terms of stability and robustness in feature point detection.

Table 3. BSIFT algorithm feature point detection results

		(a)	(b)	(c)	(d)
BSIFT algorithm	Time consumption(s)	1.73	1.41	1.47	1.38
	Number of feature points	10664	10334	13754	11465

The detection time of the BSIFT algorithm is more than twice that of the original SIFT algorithm. Due to the fact that the BRIEF algorithm filters the detected feature points to a certain extent, the feature points around the image are removed, so the number of detected feature points is less than that of the SIFT algorithm.

Using the BSIFT algorithm, several sets of multiple crack images were tested for stitching. The original image and the stitching effect are shown below:

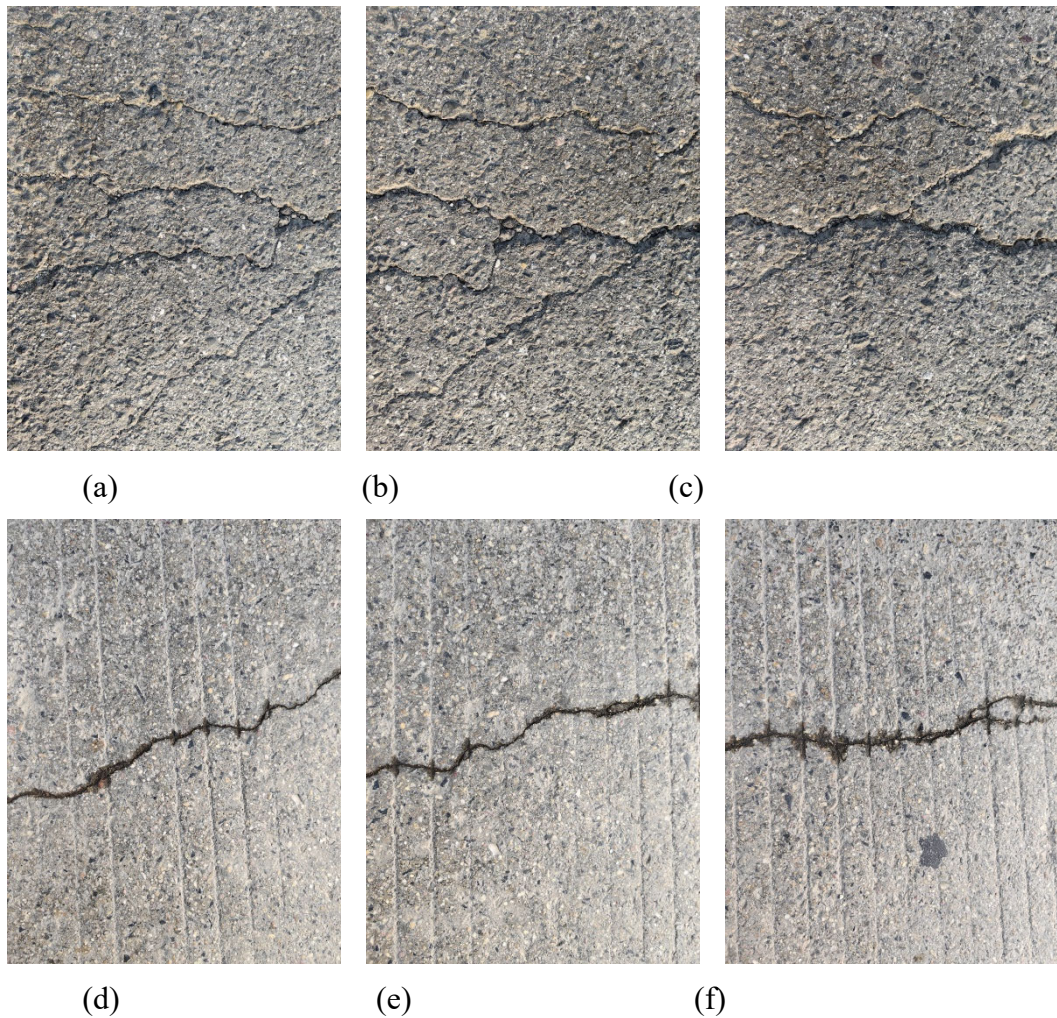


Figure 6. Crack image before splicing



Figure 7.The splicing effect of (a), (b), and (c) in Figure 6



Figure 8.The splicing effect of (d), (e), and (f) in Figure 6

4. SUMMARY

Image stitching technology has always been a popular research direction in the field of digital image processing. This article analyzes in detail the various steps in the image stitching process - image registration and image fusion. The image registration stage mainly includes feature point detection, coarse and fine matching of feature points, and obtaining a homography matrix based on pairing points. Due to the involvement of camera internal and external parameters in multiple image stitching, it is necessary to solve the camera internal and external parameters based on the homography matrix, and then use the beam adjustment algorithm to globally optimize the camera internal and external parameters. Then, through projection transformation, optimal stitching line search, and image fusion, the final stitching image is obtained. The summary of this article's work is as follows:

(1) This article elaborates on the background and significance of road crack image stitching, introduces the current status of image stitching technology at home and abroad, and outlines the overall steps of image stitching.

(2) This article provides a detailed explanation of two commonly used image feature point detection algorithms - SIFT algorithm and ORB algorithm. An analysis was conducted on the number of groups that need to be established in the traditional SIFT algorithm for detection. The resolution of the top few groups of images in the Gaussian pyramid is very small, and feature points can no longer be detected. Therefore, an improved method was proposed by modifying the equation that determines the number of groups and appropriately reducing the number of groups. We compared the performance of SIFT algorithm and ORB algorithm in detecting feature points in road crack images,

and combined their respective advantages to propose improvements. The improved SIFT algorithm was used for feature point detection, while the rBRIEF feature description algorithm, also known as BSIFT algorithm, was used to describe feature points. The experiment showed that the detection performance of BSIFT algorithm was comparable to SIFT algorithm, superior to ORB algorithm, and the processing speed was faster than SIFT algorithm.

ACKNOWLEDGEMENTS

This work was supported in part by Sichuan Central Inspection Technology Inc.

REFERENCES

- [1] Zhang,Z.,Deriche,R.,Faugeras,O, et al. 1995. A robust technique for matching twouncalibrated images through the recovery of the unknown epipolar geometry[C]. *ArtificialIntelligence*,78:87-119
- [2] Lowe, D. G.. Object recognition from local scale-invariant features[P]. *Computer Vision, 1999The Proceedings of the Seventh IEEE International Conference on*,1999.
- [3] D. Lowe. Distinctive image features from scale-invariant keypoints]]. *International Journal of Computer Vision*, 60(2):91-110,2004.
- [4] Simo-Serra, E., Trulls, E., Ferraz, L., et al. Moreno-Noguer, F.: Discriminative Learning of DeepConvolutional Feature Point Descriptors[C].In: *ICCV*.(2015)
- [5] Yi, K., Verdie, Y., Lepetit, V., et al. Learning to Assign Orientations to Feature Points[C]. In:*CVPR*.(2016)
- [6] Han Chao, Fang Lu, Zhang Sheng. An optimized image registration algorithm. *Journal of electronic measurement and instrumentation*,2017,31(2):178-184.
- [7] Zhang J, Chen G, Jia Z. An image stitching algorithm based on Histogram matching and SIFT algorithm_]. *International Journal of Pattern Recognition &. Artificial Intelligence*, 2017, 31(04):381-395.
- [8] Horun M. H. Multilevel minimum cross entropy threshold selection based on the honeybee mating optimization[1]. *Expert System with Applications*, 2010,37(6): 4580-4592.
- [9] Toh K K V, Isa N.A.M. Noise adaptive fuzzy switching median filter for salt-and-peppernoise reduction11. *Signal Processing Letters, IEEE*, 2010, 17(3): 281-284.
- [10] LI WJ, ZHANG M, SHEN Z H, et al. Track crack detection meth-od in complex environment [C]//11th International Symposium onComputational Intelligence and Design (ISCID). New York: IEEE,2018:356-359.
- [11] YUSOF' N A M, OSMAMN M K, NOOR M H M, et al. Crack de-tection and classification in asphalt pavement images using deepconvolution neural network [C]//8th IEEE International Conference on Control System, Computing and Engineering(ICCSCE)New York: EEE, 2018:227-232.