

Received October 12, 2021, accepted November 2, 2021, date of publication November 8, 2021, date of current version November 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3125645

Railway Panorama: A Fast Inspection Method for High-Speed Railway Infrastructure Monitoring

XINLAN JIANG^{1,2} AND SHENGCHUN WANG³

¹Department of Computer Teaching and Research, University of Chinese Academy of Social Sciences, Beijing 102488, China

²Research Center for Computational Social Sciences, University of Chinese Academy of Social Sciences, Beijing 102488, China

³Infrastructure Inspection Research Institute, China Academy of Railway Sciences Corporation Ltd., Beijing 100081, China

Corresponding author: Shengchun Wang (wangshengchun@rails.cn)

This work was supported in part by the Research and Development Plan of University of Chinese Academy of Social Sciences under Grant 2021-KYLX03005, and in part by the Major Research and Development Plan of China State Railway Group Company Ltd., under Grant K2020G024.

ABSTRACT At present, the comprehensive inspection trains of all countries in the world are equipped with an environmental video inspection system. The running environment is monitored by manual analysis of the video data, which is labor intensive and inefficient. In this paper, we propose a fast and intelligent inspection method for high-speed railway infrastructure monitoring. A railway environment video is acquired by an imaging device installed on a high-speed integrated detection train, and the forward-looking video is used to establish a panoramic imaging model by constructing the rectangular sampling area having a dynamic width from each frame of the video sequence, which covers the four side scenes of the railway and generate four panoramic images through image stitching technology. The experimental analysis indicates that proposed method makes the examination and analysis of video faster, simpler and more intelligent, which not only simplifies video content and reduces the time cost for users to browse video; but is also more feasible to run automatic identification algorithm on the compressed panoramic image for realizing the automatic intelligent detection of railway environment anomalies.

INDEX TERMS Image stitching, panorama, railway monitoring, forward-looking video, high-speed railway.


I. INTRODUCTION

At high speeds, any foreign matter intrusion or loss in the railway environment will seriously affect the safety of the train running. Intrusion of foreign objects caused by damage to the fence and sound barrier, and foreign objects in the track and power contact wire will pose a threat to the safety of the railway. It is required that the high-speed railway running environment must be closed throughout. Therefore, through the high-speed railway environmental monitoring to grasp the unsafe factors affecting the safety of high-speed rail, it is crucial to timely check the railway condition and ensure the safe driving of high-speed train. At present, the fixed-line video equipment is mainly used for environment monitoring of high-speed railway. The camera equipment is in a fixed static state, and the key parts along the railway are fixedly monitored. The video captured is transmitted to the control center through wireless transmission, so that the management

personnel can grasp the railway safety status at any time. However, the fixed-point monitoring method is limited by the acquisition field of view, and it is impossible to control the entire line and all the conditions along the line. Therefore, it's an effective method to monitor the state of the line in front of the train and the state of the fences on both sides of the line by using the train-borne environmental monitoring equipment installed in the high-speed comprehensive detection train.

High-speed integrated detection trains at home and abroad are equipped with video equipment for obtaining environmental information along the road, as a record of environmental conditions, providing environmental information for later manual inspection. However, how to quickly obtain information that affects environmental closeness and abnormal intrusion of foreign objects or equipment within the line from a large number of video images acquired by high-speed integrated detection trains, and correct warning is a problem that needs to be solved.

Aiming at the shortcomings of traditional video surveillance systems, we propose a fast panorama generation

The associate editor coordinating the review of this manuscript and approving it for publication was Victor Sanchez .

method based on the video stitching method, and converts the video into a more suitable format for the computer processing and analyzing. The redundant information between the image sequences is removed, reducing the storage overhead of the video data. On this basis, the generated railway environment panorama is used as the detection data, and then intelligent algorithm can be used to realize the automatic inspection of the foreign matter intrusion in the high-speed railway environment.

For traditional video stitching methods, the problems in time-consuming image matching and complicated optical flow computation have always given rise to the bottleneck of real-time panorama stitching. Starting from the human visual storage mode, we converted the storage mode of video frames into the storage mode of panoramic images, and proposed URS-PIM, a panoramic imaging model with uneven rectangular sampling for forward moving video.

The video images taken by forward-looking video consist of a series of two-dimensional frames, which usually include the road surface, sky and scenery on both sides of the three-dimensional European space. For this kind of scene, pixels on the image plane do not all belong to the same two-dimensional Euclidian space, and the adjacent pixels in the image may be the scene points far apart in the actual scene. From the perspective of optical flow, adjacent pixels in the image may have different optical flow velocity and direction, which is manifested as non-consistency of pixel movement. The reflection of objects from far to near in the image is the amplification of objects.

The URS-PIM model proposed in this paper can effectively solve the above problems. As shown in Figure 1, we elaborate on the Non-uniform Rectangular Sampling Panoramic Imaging Model (URS-PIM) based on the forward-looking video. The URS-PIM is used to construct the stitching area of forward-looking video by making full use of the geometric structure of the physical railway scene and the motion information of Comprehensive Inspection Train (CIT). The URS-PIM enables the alignment of stitching area. The entire panorama stitching simplifies the image processing, and realizes the real-time acquisition of railway panorama. Meanwhile, it presents a panorama index for fast video browsing, and saves the time and space costs of video browsing and storing. It also facilitates the free visual inspection by railway staff so that the safe operation of high-speed railway is ensured.

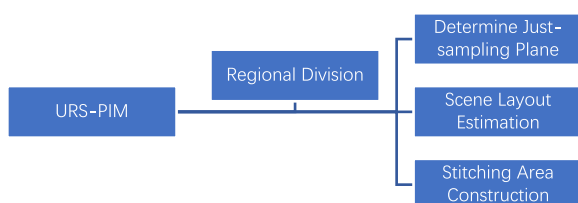


FIGURE 1. Block diagram of URS-PIM.

II. RELATED WORKS

A. FORWARD-LOOKING VIDEO STITCHING TECHNOLOGY AND APPLICATION

The forward-looking video stitching technology has been widely applied to target detection, environmental monitoring and scene reconstruction in a wide range of motion scenes. The forward-looking video falls into different categories by the motion carrier, and varies in the application target. For example, the infrared imager installed on the nose of aircraft is used to track flying objects in the air [1]. The vehicle-borne forward-looking fisheye lens enable the virtual reconstruction of urban road scenes [2]. The cameras installed on the front end of train make it possible to achieve the dynamic monitoring of high-speed railway operating environment and the automatic identification of signals and signs along the railway line [3]–[5]. The forward-looking sonar devices for underwater robots or submarines function to acquire sound wave images under underwater low-light environments, and enable the reconstruction of underwater scenes through the image stitching method [6]. The miniature camera installed on the endoscope can record the intestinal image information of living organisms and generate a panorama of the entire intestine along the forward motion path [7]. The imaging equipment installed on the percussion bit can be used to generate a panorama of the interior of geological rock formation, which is conducive to analyzing the composition and structure of rock formation [8]. Some robot motion units equipped with forward-looking imaging equipment make it possible to realize the visual inspection of inner walls of pipelines, such as oil pipelines or gun bores [9], [10] by collecting and analyzing the related images.

The above-mentioned forward-looking imaging methods and applications, as a way of video capture under specific motion conditions, differs greatly from the traditional image or video stitching models. The computational complexity of massive forward-looking video data makes it hard to realize real-time detection, thus bringing difficulties to the storage and retrieval of video data. There are few researches on the algorithm of forward-looking video stitching, and the representative researches mainly focus on:

1) Optimizing the algorithms to reduce the computational time and complexity;

2) Making use of the information redundancy between video sequences, and removing redundant information of adjacent image sequences through the multi-view panorama stitching technology, so as to better compress and extract information, retain key video information and reduce the cost of video data storage and retrieval.

B. MULTI-VIEW PANORAMIC STITCHING TECHNOLOGY

Peleg *et al.* [11] proposed a strip stitching method, which obtained a group of thin strips from video sequence images and stitched these strips into a panorama after image warping. In this method, video is regarded as spatio-temporal scene, and image strip is constructed to simulate scanning imaging. Strip content is the scene that

the virtual scan line sweeps through the video frame. The process of constructing strips introduces artificial interpolation deformation operations, so the resulting panorama is a projection of the scene content on an irregular surface. The shape of the surface is affected by many factors such as scene model, camera motion and deformation formula, and there is no regular expression form, so this method is also called manifold projection method.

Based on the theoretical analysis of manifold geometry and topological changes, Zhang [12] systematically explained the stitching method of forward or backward motion video. A group of splicing strips covering the whole scene space is extracted from the video sequence with forward motion. Each stripe is corrected into regular rectangular stitching strips after “Möbius transformation”, and then merged them together after alignment to generate a panorama. The key of extracting stitching strips from video frames is how to calculate the homography matrix which describes the relation of similarity change. Two effective methods to calculate homography matrix are summarized, that is, feature matching and optical flow calculation. However, feature matching does not always work in detection scenarios with single and repeated targets. On the other hand, in the actual detection task, the optical flow calculation fails because the assumption condition of optical flow cannot be guaranteed, and the time cost of optical flow estimation algorithm will affect the real-time performance of visual detection in practical application.

Zheng *et al.* [2] applied push-broom model to streetscape reconstruction research to generate Route Panorama images. Based on push-broom model, the algorithm continuously shoots buildings on both sides of the city through vehicle-mounted cameras. A vertical pixel line is extracted from each frame image of side-view video and arranged and combined along the time axis to form a multi-view puzzle. Since the push-broom model has a fixed sampling position for pixel lines, this method requires a uniform speed of the vehicle; otherwise, the proportion of object length in the generated panorama will be distorted. Therefore, the algorithm needs to locally adjust the generated panorama according to the location information obtained by GPS, which affects the real-time performance of the Mosaic algorithm. In order to obtain the Scene information in a larger field of vision, they further improved the algorithm and proposed the “Scene tunnel” method, which generates multi-view panoramic images with fish-eye videos from multiple angles. In the algorithm, post-processing method was used to correct the tilted and jitter panoramic images to improve the image clarity. However, the algorithm requires that the main lines in urban buildings should be straight lines parallel to the ground.

Wang [3] proposed a panoramic sampling model based on the geometric structure of detection region, and introduced the multi-view stitching technology into railway scene panorama acquisition for the first time, which achieved state-of-the-art stitching results for railway scene. Different from the previous video mosaic models and image alignment methods, the proposed panoramic sampling method based on the

detection region geometry structure can generate panoramic images from forward-moving videos quickly and succinctly. Has been in the camera’s internal parameters calibration, just rely on the camera motion information and the geometry of space structure is a priori, completed the stitching area construction and alignment, did not perform time-consuming image matching and the complicated optical flow calculation, realized the railway environment video real-time stitching, to ensure that the subsequent visual detection based on panorama of real time.

C. PROBLEMS EXISTING IN RAILWAY SCENE STITCHING

Image mosaic technology is widely used and is becoming mature in many fields, but the mode of video capture through the train-borne camera for the CIT in the forward/backward direction is significantly different from the video capture mode with fixed viewpoint, the mode of video capture through the UAV, and the mode of video capture through the train-borne camera in the lateral direction.

Specifically, the video stitching in the railway scene has the following characteristics:

- The railway scene is motionless, and the train-borne camera is relatively motionless, but the viewpoint is moving forward;
- The fields of view for image capture are different. The forward-looking camera is suitable for collecting environmental information of the entire railway line with a large field of view;
- The capture frame rate is several times that under previous video capture modes. The high-speed rail forward-looking video capture frame rate reaches up to 117 frames/sec. However, the rate of video frame capture by the camera usually cannot match, and the velocity of movement is relatively low.
- There is still no mature theory and technology for the forward-looking video stitching method. The previous multi-view panoramic stitching methods are not applicable. For example, the stitching method based on optical flow estimation [13], [14] and image matching [15], [16] has high time complexity of algorithm, and the problems such as optical flow estimation failure and matching misalignment are likely to occur.
- For the previous panorama stitching algorithms, it is assumed that there is no significant change in the depth of detection scene (closed detection areas such as pipes, intestines and stomach, gun bores and other scene targets are all located on the same plane). There is no motion parallax (MP) between adjacent images. The stitching correspondence can be obtained by selecting the image transformation model. For such open scenes as railway, the depth of the scene from the pin-hole imaging system is not constant, the scene targets are located in different depth layers, and the relative positional relationship between the targets (occlusion relationship) changes with the movement of camera. Although the relationship of homography transformation exists between

adjacent stitching images, direct acquisition is impossible through the image transformation model.

- The forward-looking video, containing massive data of unique encoding format, can't be played through all existing players, particularly mobile phones and other portable devices. Therefore, it is hard for railway inspection workers to visually detect the problems of railway operating environment. Only after format conversion through the dedicated player of China Academy of Railway Sciences can the forward-looking video be decoded and played through common players.
- The motion blur of single pixel in each frame of forward-looking video capture image sequence is not the same as that generated in other video capture modes.

Detection of high-speed railway environment through the forward-looking video is regular and normalized. There emerge some new problems regarding the use of video information, fast retrieval, environmental change monitoring, environmental anomaly detection, and reduction of storage space occupancy arising out of the repeated acquisition of video images, etc.

To solve the aforesaid problems, we propose the URS-PIM based on the forward-looking video by estimating the geometric structure of the physical railway scene in the video and balancing the relationship between the motion blur and the spatial resolution, which extracts the nondestructive information of the massive video data, obtains the lightweight panorama, and reduces the storage and browsing overhead of the video data. In addition, the video, through the URS-PIM, can be converted into an image format more suitable for manual retrieval or automatic inspection.

III. REGIONAL DIVISION OF FORWARD-LOOKING VIDEO IMAGES

There is often only a single depth layer in the occasions without scene change in depth, such as tunnels, animal intestines and gun bores. In other words, the scene stitching target objects are all located in the same scene depth layer, and the pixel correspondence between adjacent stitching frames can be acquired through the appropriate image transformation model. The high-speed train runs in an enclosed (not closed) environment, but the target objects in the open scene range are located at different scene depth layers. Although the model of image transformation between two adjacent frames meets the homography requirements, the mapping relationship between pixels cannot be obtained directly.

The shooting direction of train-borne forward-looking camera is parallel to the direction of motion. The collected video sequence is composed of a series of two-dimensional images, usually including the road surface, sky and objects on both sides of the three-dimensional Euclidean space. However, the pixels on such object image plane do not all belong to the same three-dimensional Euclidean space, and the adjacent pixels in the image may be the points far apart in the actual scene. From the perspective of optical flow, adjacent pixels in the image might have different optical flow speeds

and directions, which somewhat reflects the inconsistency of pixel motion. The reflection of objects in the image from far to near is an enlargement of objects. Among the train-borne forward-looking video images, the video image captured through the camera for the high-speed railway in the forward direction is representative in the application.

The URS-PIM based on the high-speed railway forward-looking video is mainly used to divide the two-dimensional perspective projection image according to the location of the scene image in the three-dimensional world coordinate system. As shown in Figure 2, the scene layout is estimated for the image of any frame of forward-looking video sequence, and the scene is divided into the areas of sky, left, ground and right by the detected vanishing point and electric pole.

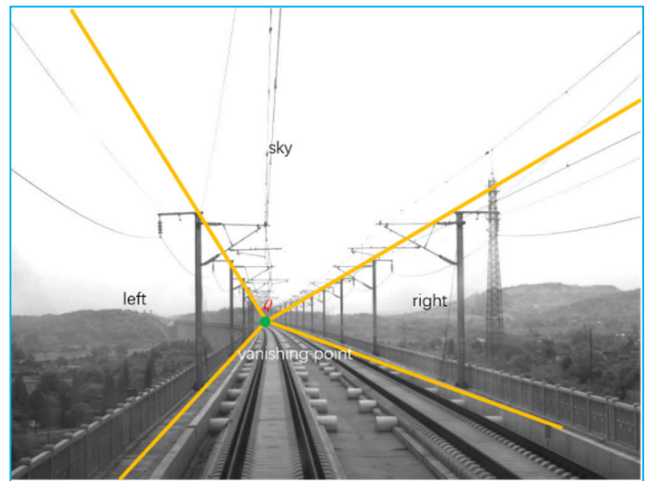


FIGURE 2. Areas of forward-looking video: sky, left, ground and right areas.

The images in different areas are stitched through the strip stitching method. The strip stitching of high-speed railway forward-looking video image is shown in Figure 3.

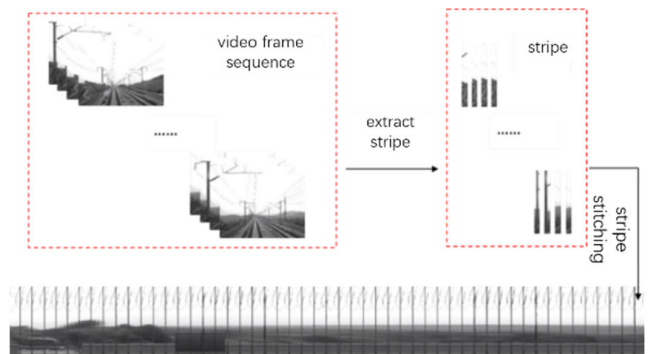


FIGURE 3. Non-uniform strip stitching.

The panorama stitching method proposed for the high-speed railway environment is brand-new. The so-called "rectangular sampling" means that the scene images are located in different positions of three-dimensional world coordinate system. To meet the needs of detection of

high-speed railway’s enclosed environment, four planes, namely the left and right guardrail (or sound barrier), sky and ground, were selected. The outer border of strip is rectangular. The panoramas of rectangular planes on the left, right, top and bottom were drawn in the cuboid geometry pipeline after sampling and stitching. Nonuniformity means that the widths of stitching strips along the left and right guardrails (or sound barrier), sky and ground are not uniform during sampling. The specific widths can be calculated according to the estimated scene layout, motion blur, camera pose, and speed of CIT. The frame diagram of URS-PIM based on the forward-looking video is shown in Figure 4.

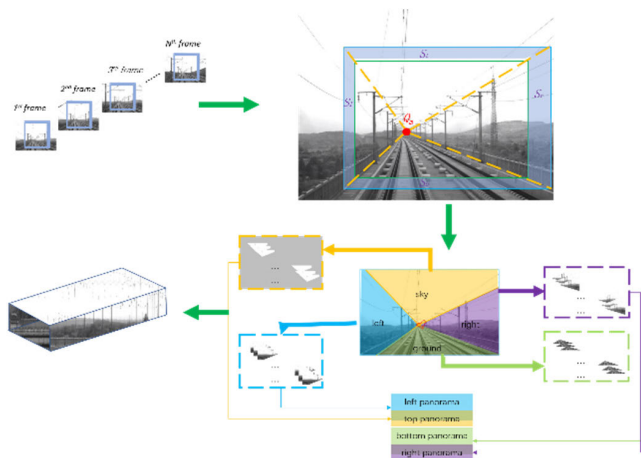


FIGURE 4. URS-PIM (left, right, top, bottom).

IV. RAILWAY PANORAMA USING URS-PIM

A. JUST-SAMPLING AND MAIN PLANE OF THE SCENE

As shown in Figure 5, $S_l^t, S_r^t, S_u^t, S_d^t$ are four uneven rectangular strips extracted from the k^{th} frame image of forward-looking video with a frame number of N . Each strip is composed of external stitching rectangle ESR_k position (green) and the internal stitching rectangle ISR_k position (red). $R_l^k, R_r^k, R_u^k, R_d^k$ denote the spatial sampling areas

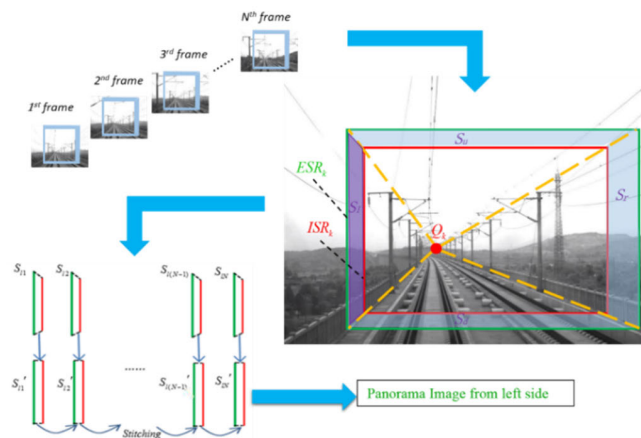


FIGURE 5. Left panorama acquired from the left strip.

corresponding to $S_l^k, S_r^k, S_u^k, S_d^k$, as shown in Figure 6. The irregular rectangular strip S_{k-1} is formed after sampling in the four directional areas R_{k-1} of the physical scene.

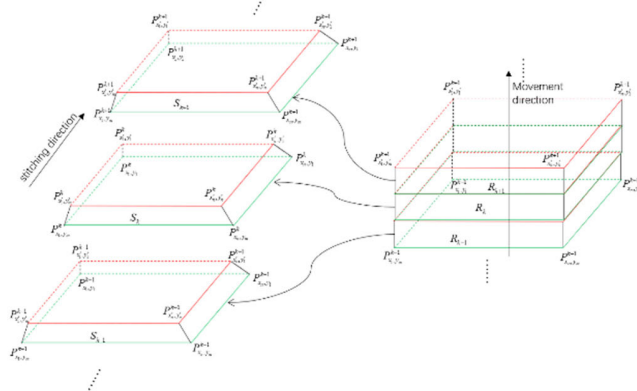


FIGURE 6. Spatial sampling area R_k and irregular rectangular strip S_k .

According to the aforementioned definition, the stitching area should satisfy certain conditions. To be specific, spatial sampling areas of two adjacent frames should not overlap (to avoid overlapped sampling), without gap (to avoid under-sampling). If there is neither overlapping nor gap between the sampling areas, the space sampling by the stitching strips is regarded as just-sampling, as shown in Figure 7.

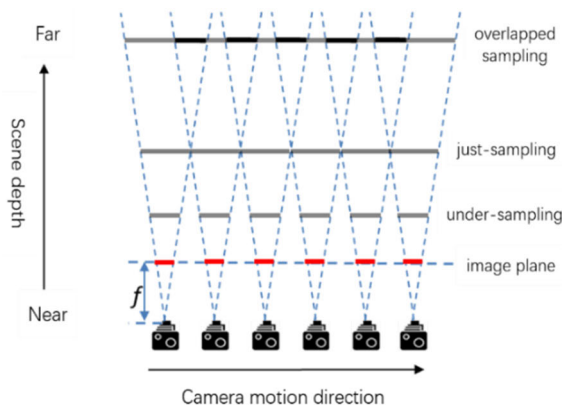


FIGURE 7. Overlapped sampling, under-sampling and just-sampling.

Panorama stitching can be recorded as follows:

$$\begin{cases} M_l = \bigcup_{i=1}^N T(S_{li}) \\ M_r = \bigcup_{i=1}^N T(S_{ri}) \\ M_u = \bigcup_{i=1}^N T(S_{ui}) \\ M_d = \bigcup_{i=1}^N T(S_{di}) \end{cases} \quad (1)$$

where, $T(\cdot)$ refers to the image stretching transformation.

In actual situations, the scene depth varies. According to the principle of pin-hole imaging, sampling distances for

different depth layers in the scene of the same strip are different, which are gradually enlarged farther away. As a result, it is impossible to achieve just-sampling of the entire scene for the extracted strip. For the distant scene, sampling areas overlap due to overlapped sampling; for the nearby scene, the scene information is lost due to under-sampling. Therefore, most applications based on mobile video stitching require the minor changes in the scene depth. All the objects are roughly on the same plane called as the dominant plane. In the railway scene, the guardrails and electric poles on both sides, the railway track at the bottom, and the catenary above are all located on the same plane. Each plane is closest to the camera, and is the key area of railway environment detection. According to the three principles, the dominant plane was selected as the detection area of railway scene. The panoramic sampling loop should ensure just-sampling on the detection area plane.

Based on the aforesaid definition, the panoramic stitching of forward-looking video can be described as follows: First, a set of 4-directional strip sequences $\{S_1, S_2, S_3, \dots, S_N\}$ are extracted from the video sequence. There is neither overlapping nor gap between any adjacent spatial sampling areas, which is defined as the just-sampling of the physical scene with a strip sequence. The plane for just-sampling is called the dominant plane of the scene. The panoramic image will be generated after geometric correction and combination of conditional strip sequences. The panorama acquisition method is shown in Figure 3.4, and the specific panorama acquisition algorithm is shown in Algorithm 1.

Algorithm 1 Panorama Acquisition Algorithm

Algorithm: Panorama acquisition

Input: high-speed railway forward-looking video

Output: Four-sided panorama

1: for $k = 1; k \leq N; k = k + 1$ (N is the total number of video frames)

1.1 Constructing the image stitching areas $S_{lk}, S_{dk}, S_{rk}, S_{uk}$

1.1.1 Calculating the ESR_k position according to the high-speed railway scene layout and motion blur

1.1.2 Calculating the ISR_k position according to the train speed and scene layout

1.2 Dividing the loop formed by ESR_k and ISR_k into four strips S_l, S_d, S_r, S_u

1.3 Unifying irregular strips S_l, S_d, S_r, S_u to regular rectangular strips S'_l, S'_d, S'_r, S'_u through homography transformation

2: end for

3: Combining all the regular rectangular strips to acquire the corresponding four-sided panoramas

Therefore, the key to generating the forward-looking video panorama lies in identifying the positions of ESR_k (green) and ISR_k (red) in four directions.

B. SCENE LAYOUT ESTIMATION BASED ON VANISHING POINT DETECTION

The motion direction of the train-borne imaging device is parallel to the extension direction of track, so the FOE of pixel motion direction will be concentrated at the intersection of two rails on the image, that is, the vanishing point. As shown in Figure 8, the line segment detection algorithm is used

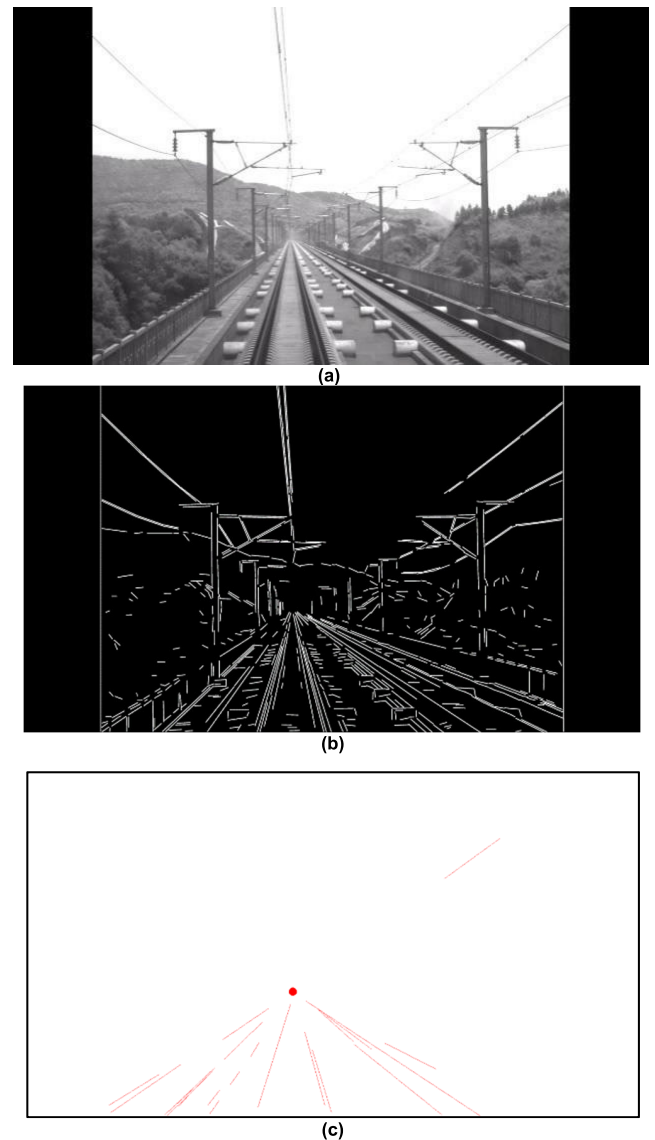


FIGURE 8. Locating the vanishing point based on straight line segment detection. (a) One frame randomly chosen from the video; (b) straight line segment detection result; (c) vanishing point detection result.

to extract line segment components such as rails, poles and edges in the railway scene so as to locate the vanishing point of the scene [17]. In addition, the common location of vanishing point is screened out according to voting rules, and is marked as a constant global vanishing point.

As shown in Figure 9(a), four rays which are drawn through the vanishing point and pass through the top and bottom of two poles on the left and right respectively, divide the image into four areas, namely the areas on the left and right in vertical direction, and the sky and ground areas in the horizontal direction. The stitching area is enclosed by the internal stitching rectangle and the external stitching rectangle. The location of external rectangle should be selected according to the impact of image resolution and motion blur on the panorama stitching effect. As shown in Figure 9(b), the external rectangle delineated by the dashed box will lead to

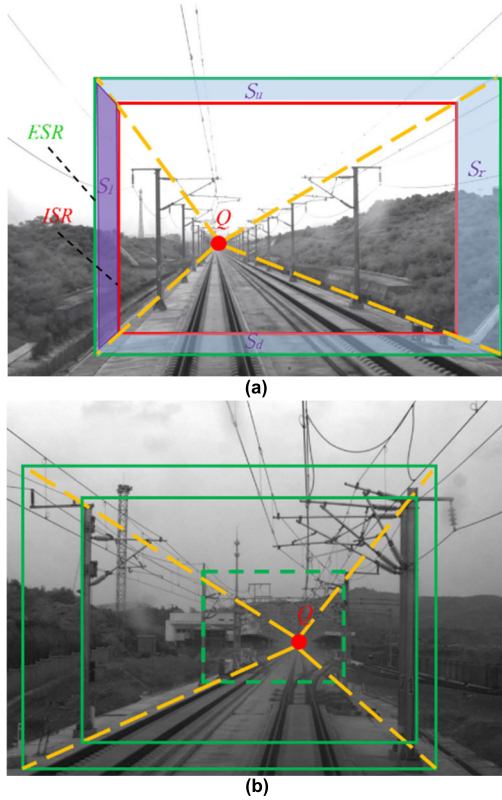


FIGURE 9. Rectangular stitching area in the railway scene: (a) closed railway environment; (b) wide railroad switch.

the too low resolution of panorama generated. For the internal rectangle, the stitching area formed by it and the external rectangle must allow just-sampling of the space scene, i.e., there should be no overlap or gap of sampling.

C. STITCHING AREA CONSTRUCTION BASED ON SCENE LAYOUT AND MOTION BLUR

According to previous research requirements, the sampling line should be perpendicular to the optical flow of pixel in order to maximize the spatial sampling. However, it is time-consuming, and also affects the resolution of generated panorama. More importantly, the calculation of optical flow is susceptible to environmental interference. For the railway scene with changeable environments, calculation result is usually distorted due to failure to meet the assumed conditions of optical flow, namely spatial consistency, brightness constancy and local small movements.

As shown in Figure 10, most practical application scenarios have a fixed geometric structure. The railway scene is composed of guardrails, electric poles on both sides, the catenary above and the ground track, which can be regarded as a cuboid. The tunnel is a semi-cylindrical geometry consisting of an arched top, the walls on both sides, and the road surface. The petroleum transportation pipeline is a cylinder surrounded by arc-shaped tube walls. In contrast, the fixed scene layout of high-speed railway comprises the guardrails and electric poles on both sides, the catenary above and the

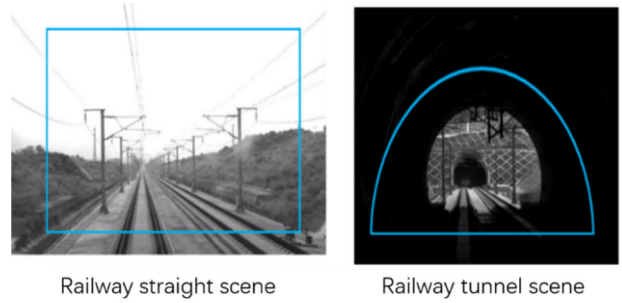


FIGURE 10. Geometry of detection area.

ground track. The external stitching rectangle of nonuniform rectangular sampling is constructed according to the detection area of the scene and the specific position in the world coordinate system, which avoids complicated optical flow calculation and improves the stitching efficiency and stability. In order to ensure the best resolution of panorama, the external stitching rectangle should be located at the edge of image with a larger spatial resolution. However, when the train runs at high speed, significant motion blur is often found in the collected images (when the camera moves but the objects in the railway scene keep motionless, global motion blur will be caused). In addition, out-of-focus blur may occur at the infinity of rail extension. To solve the problem, the optimal external stitching rectangle of nonuniform rectangular sampling should be located through balancing the image resolution and motion blur, as shown in Figure 4 (green dotted line on the left).

Take the positioning of left stitching strip S_l as an example. First, the *ESR* and the *ISR* should be located. The positioning methods for stitching strips in the other three directions are similar.

$$\tilde{P}'_{x_l, y_u} = \mathbf{H} \cdot \tilde{P}_{x_l, y_u} \quad (2)$$

where, \tilde{P}_{x_l, y_u} represents a three-dimensional object point observed in the physical scene under the world coordinate system, \tilde{P}'_{x_l, y_u} refers to the same three-dimensional object point observed in the physical scene under the world coordinate system from another viewpoint, and \mathbf{H} denotes the homography, which is expressed as follows:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (3)$$

For the image plane, the homogeneous coordinates are expressed as $P'_{x'_l, y'_u} = [x'_l, y'_u, 1]^T$ and $P_{x_l, y_u} = [x_l, y_u, 1]^T$, which are like the two points as shown in Figure 11. If some points are all located on a plane in the image, the direct mapping between the points on the image plane can be directly expressed as

$$P'_{x'_l, y'_u} = \mathbf{H} \cdot P_{x_l, y_u} \quad (4)$$

This formula can be directly used to correct and align the image, without the need to adjust the three-dimensional

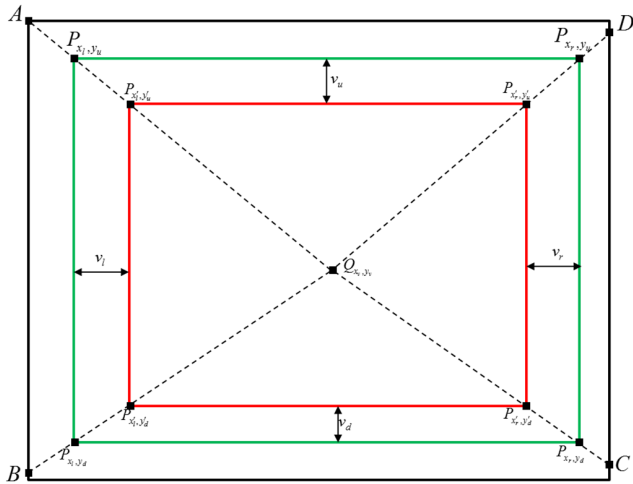


FIGURE 11. Diagram of external stitching rectangle and internal stitching rectangle.

coordinate points. The formula can be expressed as

$$\begin{cases} x'_l = h_{11}x_l + h_{12}y_u + h_{13} \\ y'_l = h_{21}x_l + h_{22}y_u + h_{23} \\ 1 = h_{31}x_l + h_{32}y_u + h_{33} \end{cases} \quad (5)$$

Further converted as

$$\begin{cases} x'_l = \frac{h_{11}x_l + h_{12}y_u + h_{13}}{h_{31}x_l + h_{32}y_u + h_{33}} \\ y'_l = \frac{h_{21}x_l + h_{22}y_u + h_{23}}{h_{31}x_l + h_{32}y_u + h_{33}} \end{cases} \quad (6)$$

Conversion form:

$$\begin{cases} h_{31}x_l x'_l + h_{32}y_u x'_l + h_{33}x'_l - (h_{11}x_l + h_{12}y_u + h_{13}) = 0 \\ h_{31}x_l y'_l + h_{32}y_l y'_l + h_{33}y'_l - (h_{21}x_l + h_{22}y_u + h_{23}) = 0 \end{cases} \quad (7)$$

Matrix form:

$$\begin{bmatrix} -x_l & -y_u & -1 & 0 & 0 & 0 & x_l x'_l & y_u x'_l & x'_l \\ 0 & 0 & 0 & -x_l & -y_u & -1 & x_l y'_l & y_u y'_l & y'_l \end{bmatrix} \cdot [h_{11} \ h_{12} \ h_{13} \ h_{21} \ h_{22} \ h_{23} \ h_{31} \ h_{32} \ h_{33}]^T = \mathbf{0} \quad (8)$$

$$\mathbf{A}_1 \bar{\mathbf{H}} = \mathbf{0}$$

$$\mathbf{A}_1 = \begin{bmatrix} -x_l & -y_u & -1 & 0 & 0 & 0 & x_l x'_l & y_u x'_l & x'_l \\ 0 & 0 & 0 & -x_l & -y_u & -1 & x_l y'_l & y_u y'_l & y'_l \end{bmatrix}$$

$$\bar{\mathbf{H}} = [h_{11} \ h_{12} \ h_{13} \ h_{21} \ h_{22} \ h_{23} \ h_{31} \ h_{32} \ h_{33}]^T \quad (9)$$

Similarly,

$$\mathbf{A}_2 \bar{\mathbf{H}} = \mathbf{0}; \quad \mathbf{A}_3 \bar{\mathbf{H}} = \mathbf{0}; \quad \mathbf{A}_4 \bar{\mathbf{H}} = \mathbf{0}.$$

A is expressed as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \\ \mathbf{A}_4 \end{bmatrix}$$

$$= \begin{bmatrix} -x_l & -y_l & -1 & 0 & 0 & 0 & x_l x'_l & y_u x'_l & x'_l \\ 0 & 0 & 0 & -x_l & -y_u & -1 & x_l y'_l & y_u y'_l & y'_l \\ -x_l & -y_d & -1 & 0 & 0 & 0 & x_l x'_l & y_d x'_l & x'_l \\ 0 & 0 & 0 & -x_l & -y_d & -1 & x_l y'_d & y_d y'_d & y'_d \\ -x_r & -y_d & -1 & 0 & 0 & 0 & x_r x'_r & y_d x'_r & x'_r \\ 0 & 0 & 0 & -x_r & -y_d & -1 & x_r y'_d & y_d y'_d & y'_d \\ -x_r & -y_u & -1 & 0 & 0 & 0 & x_r x'_r & y_u x'_r & x'_r \\ 0 & 0 & 0 & -x_r & -y_u & -1 & x_r y'_u & y_u y'_u & y'_u \end{bmatrix} \quad (10)$$

Abbreviated as

$$\mathbf{A} \bar{\mathbf{H}} = \mathbf{0} \quad (11)$$

The optimal ESR is located based on the vanishing point and scene geometry through the homography matrix, according to the effect of motion blur and the consideration of maintaining the maximum spatial sampling resolution, as shown in Figure 12.

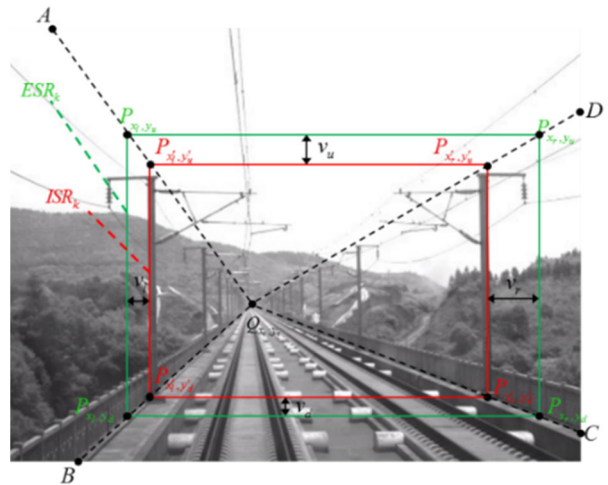


FIGURE 12. Stitching area construction.

Assuming that the scene layout has been estimated as shown in Figure 12, that is, the vanishing point Q_{x_v, y_v} and the line segment of the longest pole are detected, and four rays are drawn from the point Q_{x_v, y_v} , the linear equation of rays can be expressed as

$$\begin{cases} Q_{x_v, y_v} A : y = k_A x + b_A \\ Q_{x_v, y_v} B : y = k_B x + b_B \\ Q_{x_v, y_v} C : y = k_C x + b_C \\ Q_{x_v, y_v} D : y = k_D x + b_D \end{cases} \quad (12)$$

where, k_A, k_B, k_C and k_D represent the straight-line slopes of rays, while b_A, b_B, b_C and b_D denote the straight-line intercepts. Since the vanishing point, the top and bottom points of electric pole are known, the slope and intercept can be calculated, so they are also deemed to be known. The stitching area constructed by the vanishing point and the formula (12) is surrounded by external stitching

rectangles $P_{x_l, y_l} P_{x_l, y_d} P_{x_r, y_d} P_{x_r, y_u}$ (expressed as *ESR*) and internal stitching rectangles $P_{x'_l, y'_u} P_{x'_l, y'_d} P_{x'_r, y'_d} P_{x'_r, y'_u}$ (expressed as *ISR*).

We assume that origin (0, 0) is on the upper left corner of image coordinate system, *X* axis is the horizontal axis, and *Y* axis is the vertical axis. Four stitching lines will be derived and acquired through the left stitching lines $P_{x_l, y_u} P_{x_l, y_d}$ of external stitching rectangle *ESR*. If the *x*-coordinates of $P_{x_l, y_u} P_{x_l, y_d}$ are known, that is, x_l is known, x_l should meet two conditions to ensure the quality of panorama:

$$\begin{cases} x_l \rightarrow 0 & x_l \in [0, W] \\ D_l(P_{x_l, y_u} P_{x_l, y_d}) \rightarrow 1 & D_l(\cdot) \in [0, 1] \end{cases} \quad (13)$$

The first condition in the formula means the maximum spatial resolution of panorama generated after spatial sampling, and *W* is the image width. $D_l(\cdot)$ in the second condition denotes the high definition of stitching line on the left, with the value range of [0, 1].

Assuming x_l is known, and the resolution of panorama generated by stitching is $m \times n$, first obtain the vertex coordinates of *ESR* and *ISR*, namely $x'_l, y_u, y'_d, x_r, x'_r, y_d$ and y'_d .

The following is obtained from Figure 11.

$$\begin{cases} x'_l = x_l + v_l \\ y'_u = y_u + v_l \cdot \frac{y_v - y_u}{x_v - x_l} \end{cases} \quad (14)$$

To obtain x'_l , first calculate the width v_l of stitching strip on the left.

As shown in Figure 13, the point P_{x_l, y_u}^w in the world coordinate system in the physical space is expressed as P_{x_l, y_u}^d in the dominant plane of the scene, and projected as P_{x_l, y_u} on the image plane. The vertical line of P_{x_l, y_u} is the *ESR*_{*m*} of the *m*th frame video image. v_l is the width of the left stitching strip, the red vertical line represents the *ISR*_{*m*} of the *m*th frame video image, and *G* is the distance between the camera's optical axis and the dominant plane of the scene. Without loss of generality, each angle, including each attitude angle, is set as 0, in order to simplify the analysis. The camera is completely parallel to the imaging plane.

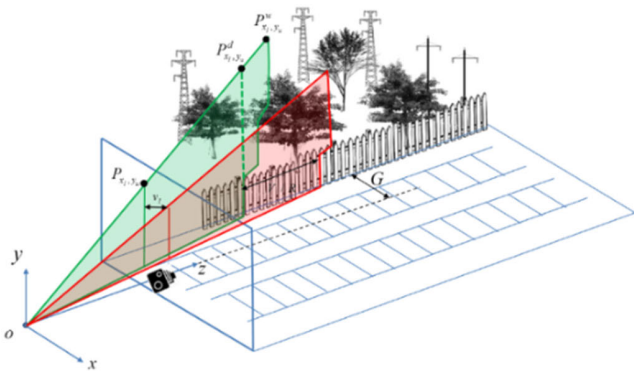


FIGURE 13. Sketch map of left-side stitching lines of two adjacent frames in the image plane arranged in the physical space and the dominant plane of the scene.

The following equation can be deduced [18]:

$$v_l = \frac{V \cdot x_l^2}{V \cdot x_l - fGR} \quad (15)$$

where, *V* is the train speed, *f* refers to the focal length of camera, and *R* denotes the video capture frame rate.

As shown above, if x_l and v_l are known, then x'_l is known.

The following can be obtained from Figure 12:

$$\begin{aligned} P_{x_l, y_u} \in Q_{x_v, y_v} A &\Rightarrow y_u = k_A x_l + b_A \\ P_{x_l, y_d} \in Q_{x_v, y_v} B &\Rightarrow y_d = k_B x_l + b_B \\ P_{x_r, y_d} \in Q_{x_v, y_v} C &\Rightarrow x_r = (y_d - b_C) / k_C \\ &= (k_B x_l + b_B - b_C) / k_C \\ P_{x'_l, y'_u} \in Q_{x_v, y_v} A &\Rightarrow y'_u = k_A x'_l + b_A = k_A (x_l + v_l) + b_A \\ P_{x'_l, y'_d} \in Q_{x_v, y_v} B &\Rightarrow y'_d = k_B x'_l + b_B = k_B (x_l + v_l) + b_B \\ P_{x'_r, y'_d} \in Q_{x_v, y_v} C &\Rightarrow x'_r = (y'_d - b_C) / k_C \\ &= [k_B (x_l + v_l) + b_B - b_C] / k_C \end{aligned} \quad (16)$$

y_u and y_d can be known from the above formulas, and then x_r, y'_u, y'_d and x'_r can be obtained through combination of those formulas (14) and (16). Then, straight-line slopes k_A, k_B, k_C and k_D and intercepts b_A, b_B, b_C and b_D of four rays can be obtained based on the vanishing point (x_v, y_v) and vertices at both ends of pole, so that the scene layout of image can be acquired finally. If x_l and the scene layout are known, eight vertex positions of external stitching rectangle and internal stitching rectangle, namely $(x_l, y_u), (x'_l, y'_u), (x_l, y_d), (x'_l, y'_d), (x_r, y_d), (x'_r, y'_d), (x_r, y_u)$ and (x'_r, y'_u) can be obtained through combination of those formulas (14) and (16).

The following can be derived from the corresponding formula and the scene layout shown in Figure 12.

$$\begin{aligned} v_r &= x_r - x'_r = -\frac{k_B}{k_C} v_l \\ v_u &= y'_u - y_u = k_A v_l \\ v_d &= y_d - y'_d = -k_B v_l \end{aligned} \quad (17)$$

v_l, v_r, v_u and v_d , the widths of stitching strips in four directions, are all known now. “Nonuniform” in “URS-PIM” means that the widths of four stitching strips S_l, S_r, S_u and S_d change dynamically for the train speed *V*, left stitching line position x_l of external stitching rectangle *ESR*, and scene layout structure.

The panorama resolution can be derived from the scene layout shown in Figure 12:

$$\begin{aligned} m &= \lceil y_d - y_u + 1 \rceil = \lceil (k_B - k_A) x_l + (b_B - b_A) + 1 \rceil \\ n &= \lceil x_r - x_l + 1 \rceil = \lceil (k_B/k_C - 1) x_l + (b_B - b_C)/k_C + 1 \rceil \end{aligned} \quad (18)$$

The prerequisite for derivation through the above formula is that the *x*-coordinate x_l of the left stitching line $P_{x_l, y_u} P_{x_l, y_d}$ of *ESR* is known. According to the above analysis, x_l must meet two conditions in the formula. The optimal position of x_l is obtained below.

The innermost red box in Figure 14 represents the ESR_{k-1} of the $k - 1$ th frame image f_{k-1} . The green box refers to the ESR_k of the k th frame image f_k , while the orange box denotes the ESR_{k+1} of the $k + 1$ th frame image f_{k+1} . The pixel point P_{x_j, y_u} in the k th frame image is affected by the previous frame ($k - 1$ th frame) and the next frame ($k + 1$ th frame). The pixel (x_j, y_u) displacement of the camera from the frame f_{k-1} to the frame f_{k+1} through the frame f_k can be obtained as follows:

$$\begin{aligned} \Delta l_{k-1,j} &= d(\mathbf{H}_{k-1}^{-1}(x_j, y_u, 1)^T, (x_j, y_u, 1)) \\ \Delta l_{k,j} &= d(\mathbf{H}_k(x_j, y_u, 1)^T, (x_j, y_u, 1)) \end{aligned} \quad (19)$$

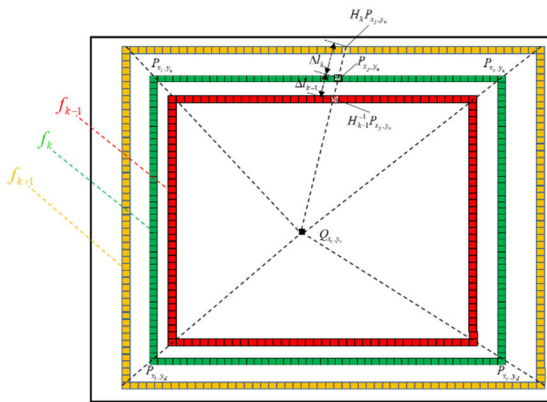


FIGURE 14. Pixel displacement between three consecutive frames.

The homography matrix \mathbf{H} above maps a pixel location to another pixel location. $d(\cdot)$ is the Euclidean distance between two pixel points. For three consecutive frames, the following condition makes sense.

$$\mathbf{H}_{k-1} \approx \mathbf{H}_k \quad (20)$$

In order to satisfy the two conditions described above, the definition of pixels in the video frame is used to describe the absolute displacement of pixels between three adjacent frames. For pixel (x_j, y_u) of the frame f_k , the definition is expressed as [19]:

$$\alpha_k^u(x_j, y_u) = \exp\left(-\frac{\Delta l_{k-1,j}^2 + \Delta l_{k,j}^2}{2\sigma^2}\right) \quad (21)$$

where, $\Delta l_{k-1,j}$ denotes the displacement from the frame f_{k-1} to the frame f_k , $\Delta l_{k,j}$ represents the displacement from the frame f_k to the frame f_{k+1} , and σ is a constant. When the motion of (x_j, y_u) between frames is small, \mathbf{H}_{k-1} and \mathbf{H}_k approach the unit matrix \mathbf{I} , and $\alpha_k^u(x_j, y_u)$ approximates 1, indicating that the image block centered by (x_j, y_u) is most likely to be sharpened. Likewise, if $\alpha_k^u(x_j, y_u)$ is very small, it means that the image block may contain large motion blur.

The definitions of stitching lines $P_{x_l, y_u} P_{x_r, y_u}$, $P_{x_l, y_d} P_{x_r, y_d}$, $P_{x_l, y_u} P_{x_l, y_d}$ and $P_{x_r, y_u} P_{x_r, y_d}$ in the four directions are expressed as follows:

$$D_u = \sum_{j=1}^n \alpha_k^u(x_j, y_u)$$

$$\begin{aligned} D_d &= \sum_{j=1}^n \alpha_k^d(x_j, y_d) \\ D_l &= \sum_{i=1}^m \alpha_k^l(x_l, y_i) \\ D_r &= \sum_{i=1}^m \alpha_k^r(x_r, y_i) \end{aligned} \quad (22)$$

The following mathematical model conditions should be met.

$$\begin{aligned} &\arg \max_{x_l, y_u, x_r, y_d} (D_u + D_d + D_l + D_r) \\ &s.t. \begin{cases} (x_l, y_u) \in Q_{x_v, y_v} A \\ (x_l, y_d) \in Q_{x_v, y_v} B \\ (x_r, y_d) \in Q_{x_v, y_v} C \\ (x_r, y_u) \in Q_{x_v, y_v} D \end{cases} \end{aligned} \quad (23)$$

When the constraint conditions of the scene geometry are met, the sum of definitions of four stitching lines of external stitching rectangle is maximized. When the sum is maximized, x_l, y_u, x_r and y_d , the optimal position of external stitching rectangle can be located. If y_u, x_r and y_d are related to x_l , the mathematical model can be rewritten as follows:

$$\begin{aligned} &\arg \max_{x_l} (D_u + D_d + D_l + D_r) \\ &s.t. \begin{cases} (x_l, y_u) \in Q_{x_v, y_v} A \\ (x_l, y_d) \in Q_{x_v, y_v} B \\ (x_r, y_d) \in Q_{x_v, y_v} C \\ (x_r, y_u) \in Q_{x_v, y_v} D \end{cases} \end{aligned} \quad (24)$$

As shown in Figure 15, as x_l moves towards the vanishing point on the right side of image, the number of pixels on the left stitching line shows a monotonous decreasing trend. On the contrary, the definition of the single point on the left stitching line shows a monotonous increasing trend, indicating that the motion blur tends to be smaller and the definition becomes higher as x_l approaches the vanishing point. The change curve of definition D_l of the entire left stitching line shows that the optimal position of external stitching rectangle can be located on the premise of maximizing the spatial sampling resolution and panorama resolution, taking into account motion blur and ensuring the definition of the left stitching strip. As shown in Figure 14(c), the optimal value of x_l is at the intersection between the red dashed line and the curve.

V. EXPERIMENTS AND DISCUSSIONS

A. DATASET

In order to verify the accuracy and effectiveness of our algorithm, panoramas were generated according to the experimental video data of two railway lines. Figure 16(a) and Figure 16(c) show the video data collected by the CIT, with the video capture frame rate of 117 frames/sec. The scenes that the high-speed rail passes the straight/curve road section are recorded respectively. Figure 16(b) shows the

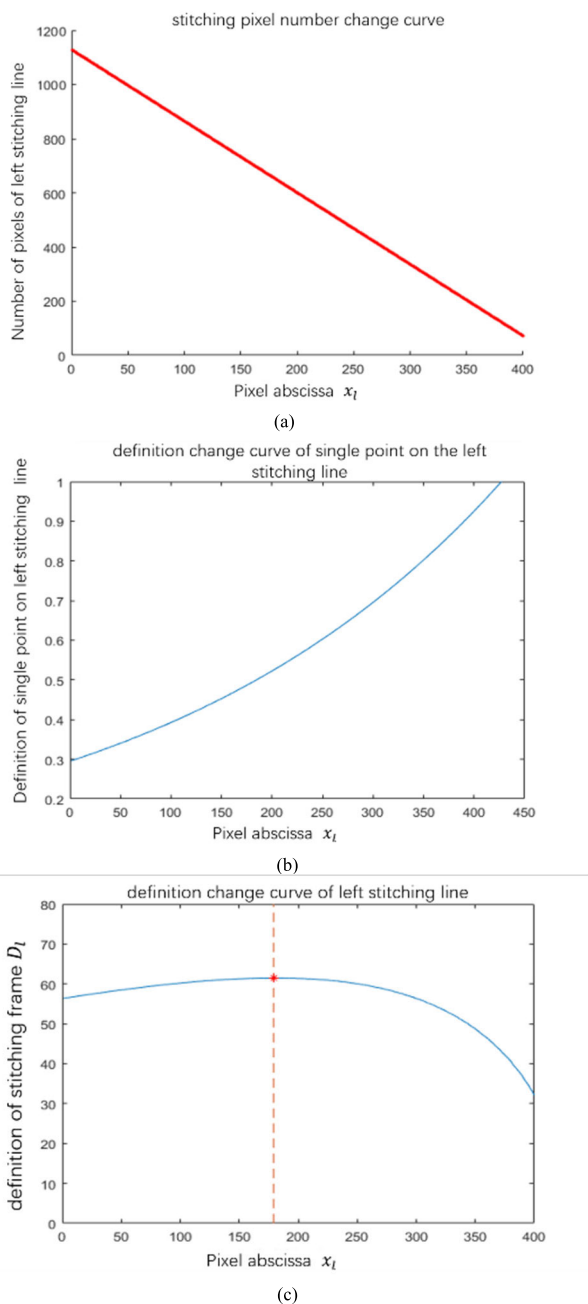


FIGURE 15. Optimal stitching line position: (a) stitching pixel number change curve; (b) definition change curve of single point on the left stitching line; (c) definition change curve of left stitching line.

experimental video data collected through the train-borne camera on the train turning at a relatively constant speed. The video capture frame rate is 25 frames/sec. See TABLE 1 for details.

The experimental platform configuration is as follows: Intel i74.0GHz processor, Windows 8 operating system, and 32 GB memory. There are totally 55,577 frames of experimental video data. Low-quality video data was acquired through the camera mounted on the train running at a speed of less than 50 km/h, with the video capture frame rate of 25 frames/sec. High-quality video data was acquired

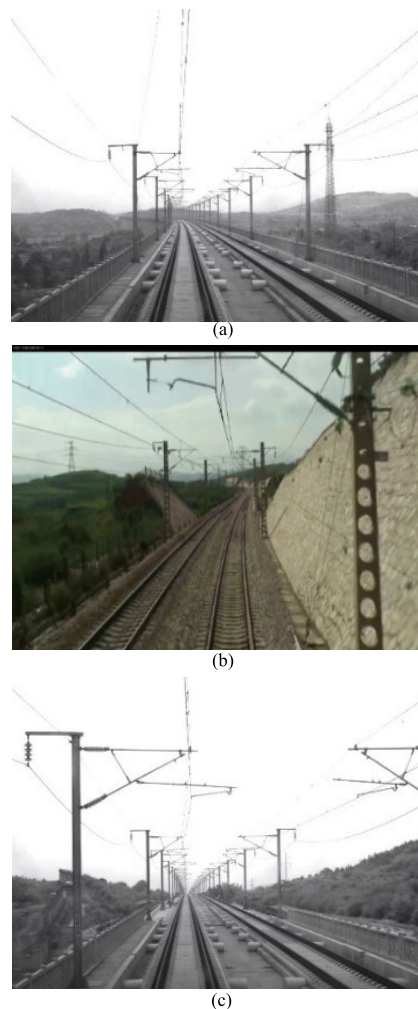


FIGURE 16. Video frames captured in the forward direction: (a) curve road section of high-speed railway; (b) wide railway switch; (c) straight road section of high-speed railway.

through the camera mounted on the train running at a speed of 350 km/h, with the video capture frame rate of 117 frames/sec. Two railway lines were selected to generate panoramas. The specific scenes include: the train passed the curve road section of Loudi-Zhuzhou section; the CIT passed the straight/curve road section of Changsha South-Hangzhou East section.

B. RESULTS OF THE PANORAMA

Figure 17 shows the panorama of the scene that the train passed the curve road section of Changsha South-Hangzhou East Section. The experimental data of video shot when the train ran at high speed shows that satisfactory stitching effect was achieved. Although the distant trees in the video are distorted horizontally because the overlapped sampling affects the optical flow estimation, the information of railway targets monitored such as the electric poles and sound barriers nearby is not lost due to just-sampling, with the little distortion. Figure 17(a) shows that a section of sound barrier is interrupted. Figure 17(c) shows that the uniform stitching effect

TABLE 1. Details of video data.

Section	Speed (<i>km / h</i>)	Resolution (<i>pixel</i> × <i>pixel</i>)	Code rate (<i>kbps</i>)	Capacity (<i>AVI, MB</i>)
Loudi-Zhuzhou(50km)	50	1280 × 720	3072	1145
Changsha South-Hangzhou East	350	1024 × 768	52505	320

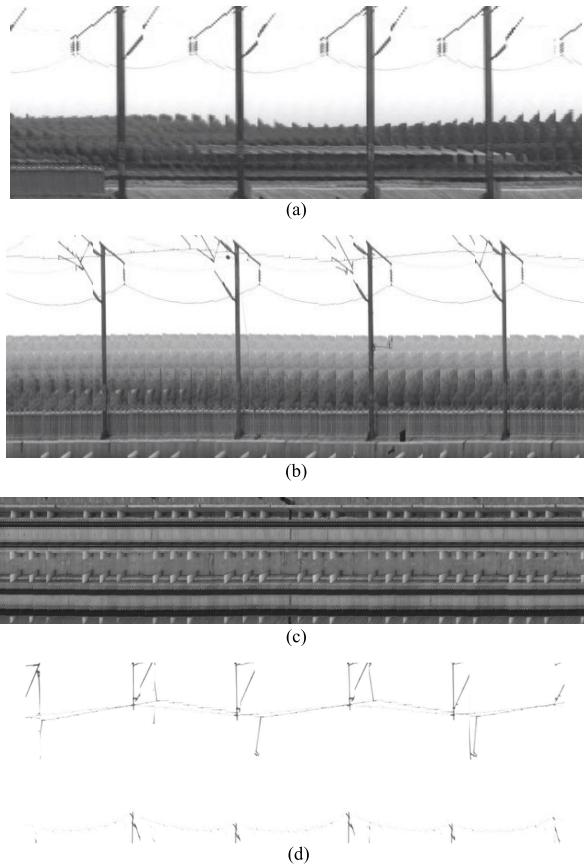


FIGURE 17. Four panoramas generated from the curve road scene of high-speed railway: (a) left panorama; (b) right panorama; (c) bottom panorama; (d) top panorama.

is achieved, and the railway track bed and other facilities are arranged at equal intervals. If the monitored target is located in a distant scene, the distortion of image can be avoided through adjusting the width of stitching strip.

Figure 18 shows the panoramas generated when the train passed the curve road section of Loudi-Zhuzhou section. Since the curve road in the image is still like the straight road at the stitching area, the image alignment method for the straight road scene also applies to get the proper stitching result, as shown in Figure 18(a) and Figure 18(b). For the track plane, the stitching result also reflects the shape of the curve road, as shown in Figure 18(c).

Figure 19 shows the panoramas generated when the CIT passed the straight road section of Changsha South-Hangzhou East section at a speed of less than 350 *km/h*, with the video frame capture rate of 117 *frames/ sec*. The stitching is very regular.

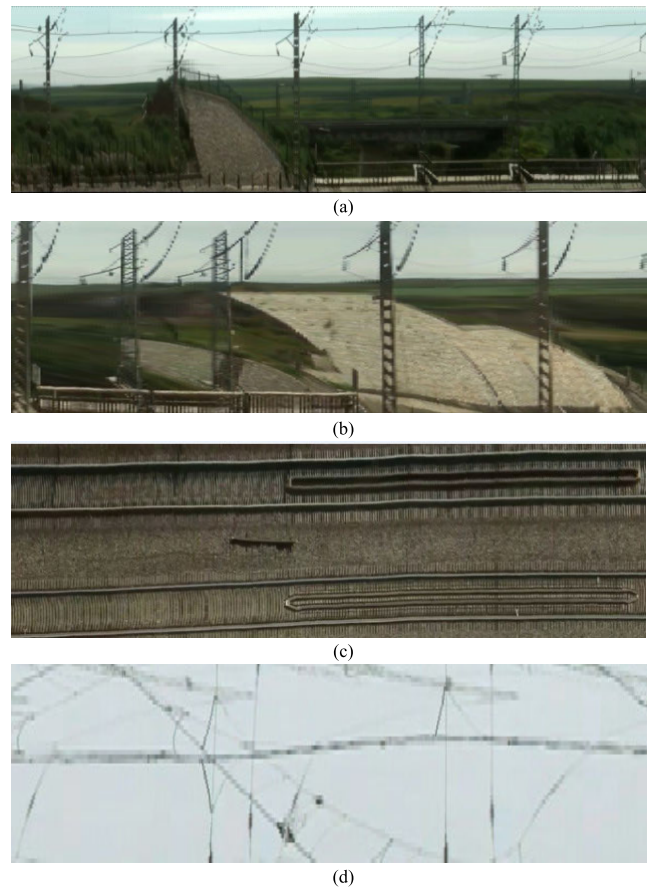


FIGURE 18. Four panoramas generated from curve road scene: (a) left panorama; (b) right panorama; (c) bottom panorama; (d) top panorama.

The literature [11] proposes to locate the external stitching line on the principle of maximizing the resolution, that is, the external stitching line should be located at the edge of image as much as possible while the stitching rectangle is kept. This method is effective, through which the panorama with the largest image resolution can be generated. However, large motion blur occurs at the edge of image under high-speed motion conditions, so obvious motion blur exists in the generated panorama, as shown in Figure 20(a).

The stitching effect of SIFT feature matching [20] is shown in Figure 20(b). Due to the lack of significant features in repetitive and simple scenes, the Mosaic images cannot be well aligned, resulting in serious distortion.

Figure 20(c) is the representative method in the current railway panorama generation [3], and the stitching effect is acceptable without obvious motion blur or image distortion. However, due to improper selection of stitching location, the

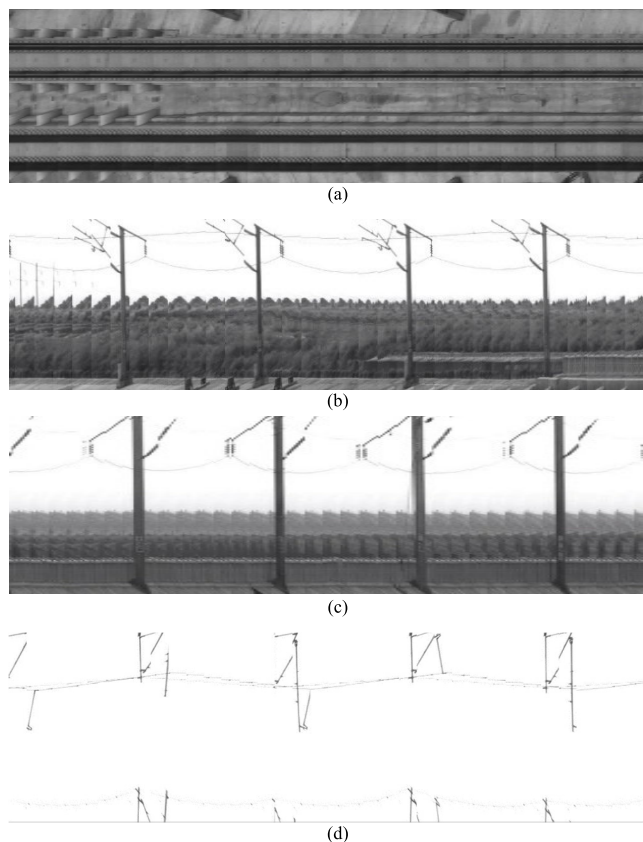


FIGURE 19. Four panoramas generated from straight road scene of high-speed railway: (a) bottom panorama; (b) right panorama; (c) left panorama; (d) top panorama.

generated panoramic image has low resolution and unclear image.

The stitching area construction method based on the maximized definition of stitching wireframe improves the quality of panorama generation because the impact of image resolution and motion blur on the stitching effect is comprehensively considered, as shown in Figure 20(d).

In TABLE 2, original video data is compared with that in the panorama. The size of video data is directly proportional to the acquisition time and the video code stream rate. The panorama stitching is similar to linear array push-broom. The size of image captured by the line scan camera is only related to the resolution and scan length of linear array CCD, rather than the acquisition time and video code stream rate. Therefore, the size of panorama, in principle, should be only related to the length of stitching strip (pixel unit) and the mileage.

The URS-PIM is specific to the scene of video capture by the railway camera in the forward direction. The original panorama stitching model and video stitching model are not applicable. In order to verify the effectiveness and timeliness of the URS-PIM (25 frames/sec), different image alignment methods (SIFT Matching [20], L-K Optical Flow [14], MRF Optimization [8], Fourier-based Registration [6], and the method of alignment based on the geometry of detection area [3]) were compared at the stitching speed.

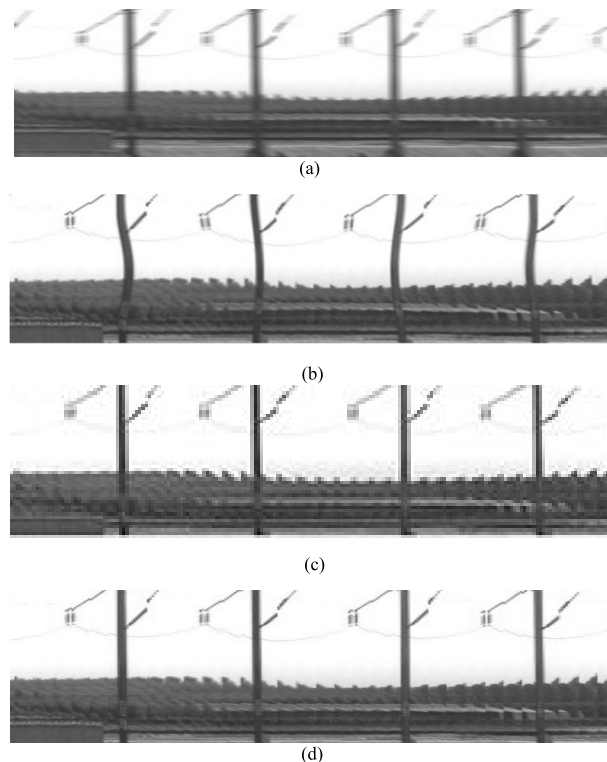


FIGURE 20. Comparison of stitching results: (a) Maximum resolution; (b) Sift matching; (c) Geometric structure detection; (d) Optimal stitching position.

TABLE 2. Data size comparison between video and panorama.

Section	Capacity (MB)	Panorama capacity (MB)	Compression ratio
Loudi-Zhuzhou	1145	17.3	66:1
Changsha South-Hangzhou East	320	6.23	51:1

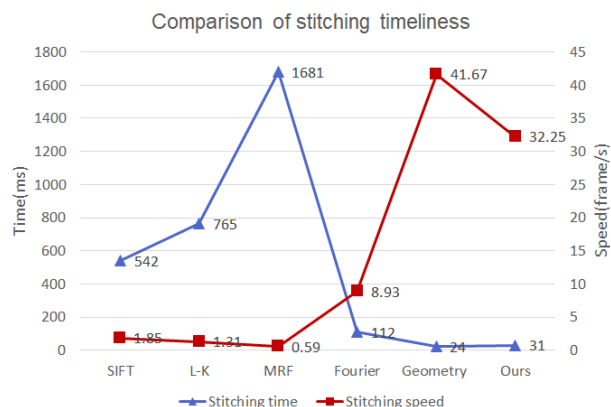


FIGURE 21. Comparison of stitching timeliness for stitching time and stitching speed.

As shown in Figure 21, four different image alignment methods are compared in the stitching speed. The blue poly-line represents the stitching time, and the red poly-line denotes

the stitching speed. Through the URS-PIM, the stitching speed can reach 42 frames/sec, much higher than the video capture rate (25 frames/sec). The URS-PIM can align the image according to the information of physical scene geometry and camera motion, and avoid the time-consuming feature estimation and optical flow computation. In addition, the globally invariant vanishing point for the stitching area of just-sampling, image speed and other key parameters have been stored offline in the query table, making image alignment similar to simulation of fast query in the table.

VI. CONCLUSION

In this paper, for the requirements of railway environment video storage and later monitoring, aiming at the problems that the traditional video mosaic theory is not applicable, and the timeliness of mosaic method can't meet the application requirements, a fast panorama generation model based on high-speed forward motion video is established, and an uneven rectangular sampling panorama imaging model URS-PIM for vehicle forward motion video is proposed, which avoids the complex and time-consuming image analysis, only uses the prior geometric structure of the railway scene and the motion information of the train, constructs the video mosaic region based on the maximum definition constraint of the mosaic region, realizes the alignment of the mosaic image, and provides a lightweight video panoramic index method for railway environment video. Railway panorama not only reduces the cost of video data storage and access, but also transforms the video into a more suitable form for manual inspection or computer automatic analysis, so as to achieve the purpose of using panorama instead of video image to realize scene image information compression, and discovering the differences of railway operation environment through the later panorama comparison.

REFERENCES

- [1] A. Yilmaz, K. Shafiq, and M. Shah, "Target tracking in airborne forward looking infrared imagery," *Image Vis. Comput.*, vol. 21, no. 7, pp. 623–635, Jul. 2003.
- [2] J. Y. Zheng, Y. U. Zhou, and P. Mili, "Scanning scene tunnel for city traversing," *IEEE Trans. Vis. Comput. Graphics*, vol. 12, no. 2, pp. 155–167, Mar. 2006.
- [3] W. Shengchun, "Research and application of computer vision detection technology based on forward motion video," Beijing Jiaotong Univ., Beijing, China, 2015.
- [4] W. Shengchun, L. Siwei, H. Yaping, W. Liang, and D. Peng, "Stereo panorama generation from the train-borne video," *Acta Optica Sinica*, vol. 34, no. 12, pp. 210–218, 2014.
- [5] J. Xinlan, W. Shengchun, L. Siwei, H. Yaping, and D. Peng, "Real-time panoramic imaging method for train-borne forward motion video," *Acta Optica Sinica*, vol. 37, no. 5, pp. 187–198, 2017.
- [6] N. Hurtós, D. Ribas, X. Cufí, Y. Petillot, and J. Salvi, "Fourier-based registration for robust forward-looking sonar mosaicing in low-visibility underwater environments," *J. Field Robot.*, vol. 32, no. 1, pp. 123–151, Jan. 2015.
- [7] M. Ou-Yang, "Image stitching and image reconstruction of intestines captured using radial imaging capsule endoscope," *Opt. Eng.*, vol. 51, no. 5, May 2012, Art. no. 057004.
- [8] C. Niu, F. Zhong, S. Xu, C. Yang, and X. Qin, "Cylindrical panoramic mosaicing from a pipeline video through MRF based optimization," *Vis. Comput.*, vol. 29, no. 4, pp. 253–263, Apr. 2013.
- [9] P. Hansen, H. Alismail, P. Rander, and B. Browning, "Towards a visual perception system for pipe inspection: Monocular visual odometry," Robot. Inst., Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-RI-TR-10-22/CMU-CS-QTR-104, 2010.
- [10] Z. Geyi, W. Guohui, G. Jiaying, and C. Yanhua, "Research on image jointing algorithm for the inner surface of the artillery bore," *J. Acad. Armored Force Eng.*, vol. 21, no. 4, pp. 28–31, 2007.
- [11] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1144–1154, Oct. 2000.
- [12] H. Zhang, T. Li, and J. Geng, "Manifold modeling and its application to tubular scene manifold mosaicing algorithm," *J. Math. Imag. Vis.*, vol. 44, no. 1, pp. 80–98, Sep. 2012.
- [13] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1385–1392.
- [14] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *Int. J. Comput. Vis.*, vol. 56, no. 3, pp. 221–255, 2004.
- [15] W. Xianxiang, G. Baolong, and W. Juan, "Automatic stitching algorithm for cylindrical panoramic images based on phase correlation," *Acta Optica Sinica*, vol. 29, no. 7, pp. 1824–1829, 2009.
- [16] L. Li, J. Yao, R. Xie, M. Xia, and W. Zhang, "A unified framework for street-view panorama stitching," *Sensors*, vol. 17, no. 12, p. 1, Dec. 2016.
- [17] X. Yu, S. Luo, T. Xu, L. Wang, and S. Wang, "Attitude estimation of the train-borne camera from the railway-environment," *Opt. Technique*, vol. 41, no. 3, pp. 197–203, 2015.
- [18] P. Dai, S. Wang, Y. Huang, H. Wang, X. Du, and Q. Han, "Visual indexing of large scale train-borne video for rail condition perceiving," *IEICE Trans. Inf. Syst.*, vol. E100.D, no. 9, pp. 2017–2026, 2017.
- [19] S. Cho, J. Wang, and S. Lee, "Video deblurring for hand-held cameras using patch-based synthesis," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 1–9, Aug. 2012.
- [20] L.-N. Li and N. Geng, "Algorithm for sequence image automatic mosaicing based on SIFT feature," in *Proc. WASE Int. Conf. Inf. Eng.*, vol. 1, Aug. 2010, pp. 203–206.



XINLAN JIANG was born in 1976. She received the Ph.D. degree in computer science and technology from Beijing Jiaotong University, in 2018. She is currently an Associate Professor with the University of Chinese Academy of Social Sciences. Her research interests include computer vision, machine learning, and optical inspection.



SHENGCHUN WANG was born in 1985. He received the B.S. and Ph.D. degrees from Beijing Jiaotong University, Beijing, China, in 2008 and 2015, respectively. He worked as an Associate Researcher at the Railway Infrastructure Research Institute, China Academy of Railway Science, Beijing. His research interests include scene representation for railway environment, computer vision, and structured light measurement.