ROUTE PANORAMA ACQUISITION AND RENDERING
FOR HIGH-SPEED RAILWAY MONITORING

Shengchun Wang1,2, Jiang Yu Zheng2, Siwei Luo1, Xiaoyue Luo3, Yaping Huang1, Dalong Gao1

1School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, China
Email: scwang.bjtu@gmail.com; {swluo, yphuang, 11120435 }@bjtu.edu.cn
2Department of Computer and Information Science, Indiana University Purdue University Indianapolis, IN 46202, USA
Email: jzheng@cs.iupui.edu
3Department of Mathematics, Linfield College, McMinnville, OR 97128, USA
Email: xiaoyue_luo@hotmail.com

ABSTRACT

With the rapid development of high-speed railway, the safety running of trains becomes extremely urgent. Video is a direct and effective manner for railway environment monitoring. In this paper, we propose a route panoramic representation to produce a virtual environment of the railway scene from a train-borne video, which provides an all-round display for the entire length of railway. We generate the route panorama from the forward motion video and then render 3D scene. The experimental results show many advantages of the representation in the large data storage, browsing, and examination. It will be used for railway safety checking, railway facility inspection and virtual sightseeing from the train in the future.

Index Terms — route panorama, scene tunnel, environment monitoring, high-speed railway, image based rendering

1. INTRODUCTION

The high-speed railway is growing rapidly all over the world in recent years. High-speed trains run in closed environments fenced by guardrails for security assurance, and any unexpected matters such as missing communication units and bolts on the track, broken fences, unpredictable objects falling into the rail area or hanging on wires on top of the train will lead to disastrous consequences. It is an extremely urgent task to ensure the high-speed trains free from accidents. Static cameras have installed widely along the railway for surveillance [1,2]. Taking the Beijing-Shanghai High-speed Railway as an example, more than 400 video surveillance devices are installed at the railway roadbed, road junctions, bridges, and critical railway sections to meet the needs of railway maintenance and safety management. The managers can detect unexpected matters timely by analyzing the video data sent from the railway, and then carry out proper emergency treatments. However, because the cameras cannot cover complete route sections and their costs in maintaining the systems are high, a mobile surveillance system along the railway is effective. The moving camera installed in the font of a high-speed patrol train [3,4,5] is used for railway inspection tasks. Compared with the fixed-point monitoring, it is more economic, easier to implement, and more comprehensive in data acquisition. Because multiple video will be captured from the entire railway during the daily patrol, the challenge is how to archive recorded videos in the small storage space. In addition, the post video analysis is still infeasible other than manual searching and discrimination. The goal of this research is to acquire critical information from such large capacity videos automatically for proper early warning.

In this paper, we propose a framework to achieve the route panorama acquisition and rendering from the train borne video. By comparing and analyzing the existing methods, we introduce a fast and concise model to generate route panorama or scene tunnel from the video captured by a moving camera on a high-speed train. Variable width strips are extracted from video frames for a rectangular tunnel along the railway, and then the panoramic scenes are rendered in a 3D virtual environment.

The rest of the paper is organized as follows. Related works on the panorama development are described in Section 2. Section 3 formulates the construction of route panorama images. Section 4 is devoted to panorama rendering. Experimental results are demonstrated in Section 5. Section 6 concludes the paper and gives the further work.

2. RELATED WORKS AND OUR METHOD

Panorama stitching technology is widely used in video conferencing, aerial photography, military monitoring and virtual view rendering. Panorama was first put forward by Irish painter Robert Barker [6], while digital panoramas were generated in 1990s [7]. There are different versions on this concept according to applications and production
methods. For example, local panoramas with wide fields of view are generated from an imaging device with a single optical center. They can be composed from a rotating video or stitched from images to yield 360-degree cylindrical or spherical views. Another panorama is generated by moving a camera along a path. Such a route panorama [13] extends the image unlimitedly in space. It can be extracted from a video by connecting pixel lines from consecutive frames with a fixed slit (a pixel line) as push-broom [7, 8, 9, 13] and adaptive manifold [10, 11], or with a dynamic X-slit [12, 14]. The fixed slit achieves parallel-perspective projection while the dynamic slit generates a near-perspective projection depending on the depth from the camera path. Zheng et al further produced a complete scene tunnel to show the entire scenes around the path [15]. Such a method requires a high frame rate of camera at 300fr/sec [14] or a slow vehicle speed at 30–40km/h [13]. An alternative approach for the route panorama is to stitch wider strips [16] (strips panorama as in Fig. 1) or strips with a dynamic strip width [19]. It requires a front/dominant depth layer for the matching and stitching of consecutive patches. The camera uses a short exposure time to avoid motion blur but may introduce noises. In addition, it requires sufficient amount of features in the scenes for matching frames. In addition, Zhang [17] constructed a manifold model to tackle the panorama acquisition from forward or backward motion video.

Fig. 1. Diagram of the route panorama generation.

Our work to acquire route panorama is similar as the scene tunnel. The camera faces forward to record a video with lower image velocities than that from a side viewing camera at an ordinary frame rate (25fr/sec), to satisfy the fast movements of high-speed trains and reduce motion blur in the video. We determine the strip width in video frames according to the train speed and known geometric constraints to avoid matching frames, because images may not have sufficient salient features for stitching due to the simplicity of railway scenes. Generating a route panorama over a long distance is very challenging. The longest video of a high-speed train runs 2000km in 8hrs. Hence, the sampling process has to be robust. The method we propose here is fast and concise for a desirable route panorama without the performing the complex image matching and stitching.

To construct a compact scene tunnel in the 3D virtual space, we capture a forward motion video sequence and sample a rectangular ring with a dynamic width to cover four side scenes of the rail as colored in Fig. 2. Four route panoramas including areas of sky-wires, left and right fence-poles, and ground-rails are obtained for constructing a box tube along the rail direction.

To visualize railway scenes from a long video, we render a box tube that allows for a global view, random access of spots, free speed maneuver and fast streaming shorter than real running time of train. The fast rendering serves as a visual index for further zoom into a particular frame for examining details.

Fig. 2. Structural diagram of the framework

3. OBTAINING PANORAMA ALONG RAILWAY

3.1. The structure of the railway framework

The railway infrastructure provides the properties as follows: (1) the train moves at almost a constant speed locally on a smooth track; (2) the monitoring environments such as poles, fences, and track in the rail area have almost standard depths, intervals, and structure; and (3) landscapes outside rail area and weather conditions are less important than rails, but provide additional information for reference. In addition, the
camera FOV is sufficiently large to contain information surrounding the rail. The camera parameters including focal length and image resolution are known or calibrated.

The discrete video frames contain overlapping scenes with non-uniformed resolutions at different depths. Highly repeated rail patterns in consecutive frames confuse the current monitoring part with the examined part of track. Searching video itself for finding suspect spots on the track and surrounding is inefficient. Therefore, it is a feasible solution to generate the panorama from the video and render it to 3D virtual scene for further examination.

![Diagram of panorama acquisition from the stitching area in every frame.](image)

**Fig. 3.** Panorama acquisition from the stitching area in every frame.

### 3.2. Panorama acquisition model

As depicted in Fig. 3, the camera is fixed on the train. From a close range of the track and fences from the camera, a vanishing point, \( Q \), or accordingly Focus of Expansion (FOE) of optical flow is detected in the video frame from the extension of track lines and road edges, even if the track may be curved at distance ahead on a turning rail. This vanishing point detection can be performed with a long video sequence and the accurate position of \( Q \) can be voted from the results of all the frames.

Through the point \( Q \), four radial lines which pass through the top and bottom point of the poles at both sides separately are located to divide four regions in the video frames corresponding to two vertical side planes and two horizontal planes of ground and sky in the 3D space respectively as depicted in Fig. 4.

An *outer rectangular coil* with corners located on the four radial lines is set maximally in the frame to obtain the best resolution of scenes. The position of the *inner rectangle coil* must guarantee that the strips sampled from consecutive frames have a perfect coverage at the rail area neither overlapping nor missing scenes, i.e., *just-sampling*. The image velocity \( v \) on the rail track is obtained from train speed \( V \) to ensure this just-sampling. The four strips between two rectangles, which are denoted as \( S_1, S_2, S_b, \) and \( S_r \), can be sampled from the frames as shown in Fig. 4.

The horizontal and vertical lines on be sampling coil are mapped onto the real box of scene tunnel (Fig. 3) are parallel lines. The angles of these lines with the path direction depend on the installed camera orientation panned and tilted from the exact translation direction. Therefore, the mapping of strips onto the real box tube form skew box rings connected along the train path as depicted in Fig. 3.

The sampled strips from each frame are transformed through a *homography* change to obtain the patches on the route panoramas (Fig. 5). The composed route panoramas can thus form a box tube along the path direction for visualizing the scene tunnel.

![Diagram of stripe correction using homography transform for panorama generation from the left side of rail environment.](image)

**Fig. 4.** The construction of the just-sampling stitching area in the \( n^{th} \) frame.

**Fig. 5.** Stripe correction using homography transform for panorama generation from the left side of rail environment.
3.3. Image velocity estimation for just-sampling of coil

As shown in Fig. 6, O-XYZ is the Train Coordinates System where the Z axis is in the train moving direction moving at speed $V$ and the Y axis is perpendicular to the ground. The camera coordinates system is o-xyz with tilt, pan, and roll changes from O-XYZ. Usually, the camera axis my direct slightly to a side with another rail track to obtain wider or larger view of the rail bed. We capture a video segment as the train moves on straight and smooth tracks; the camera roll can be considered as zero under such an ideal situation. We will determine the image velocity, $v$, at the bottom strip on rail track for the just-sampling span of the coils. This will further guarantee the just-sampling spans of other sides around rail.

![Fig. 6.](image)

Fig. 6. The relationship between the image velocity $v$ and the train speed $V$ to realize the just-sampling requirement on the ground.

From vanishing point $Q(x_0, y_0, f)$ of the rail (or Focus of Expansion of train motion in the video frame), the camera directions including tilt $\phi$ and pan $\theta$ from the train moving direction are calculated as

$$\phi = \arctan \frac{y_0}{f}, \quad \theta = \arctan \frac{x_0}{\sqrt{f^2 + y_0^2}}$$

(1)

where $f$ is the calibrated camera focal length and $C(0, 0, f)$ is the image center. The outer sampling line $l_1$ and inner sampling line $l_2$ are determined as

$$l_1: A_1(X_{x1}, Y_{y1}, Z_{z1}) B_1(X_{x3}, Y_{y3}, Z_{z3});$$

$$l_2: A_2(X_{x2}, Y_{y2}, Z_{z2}) B_2(X_{x2}, Y_{y2}, Z_{z2})$$

(2)

under the camera coordinate system o-xyz. The two image lines in the train coordinates system O-XYZ are denoted as

$$l_1: A_1(X_{x1}, Y_{y1}, Z_{z1}) B_1(X_{x3}, Y_{y3}, Z_{z3});$$

$$l_2: A_2(X_{x2}, Y_{y2}, Z_{z2}) B_2(X_{x2}, Y_{y2}, Z_{z2})$$

(3)

which can be obtained via their transformation from system o-xyz to system O-XYZ through

$$
\begin{align*}
A_1: (X_{x1}, Y_{y1}, Z_{z1}) &= (x_{x1}, y_{y1}, f) M(\phi)M(\theta) \\
B_1: (X_{x1}, Y_{y1}, Z_{z1}) &= (x_{x1}, y_{y1}, f) M(\phi)M(\theta) \\
A_2: (X_{x2}, Y_{y2}, Z_{z2}) &= (x_{x2}, y_{y2}, f) M(\phi)M(\theta) \\
B_2: (X_{x2}, Y_{y2}, Z_{z2}) &= (x_{x2}, y_{y2}, f) M(\phi)M(\theta)
\end{align*}
$$

(4)

where $M(\phi)$ and $M(\theta)$ are rotation matrices as

$$M(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix} \quad M(\theta) = \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix}$$

(5)

The plane of sight through $l_1$ is thus determined from its normal $N_i = OA_i \times OB_i = (X_{ai}, Y_{yi}, Z_{zi})(X_{bi}, Y_{yi}, Z_{zi})$ and camera focus $O$. The plane has an intersection line $l_1$ with the rail surface on the ground, which can be obtained by enforcing $Y= -H$, where $H$ is the height of the camera from ground.

$$
\begin{align*}
[X \ Y \ Z] &= \left[ X_{ai}Y_{yi}Z_{zi} - X_{bi}Y_{yi}Z_{zi} \right] = 0; \\
Y &= -H.
\end{align*}
$$

(6)

This is further detailed by computing (4) from the image coordinates of $l_1$ as

$$
\begin{align*}
[X \ Y \ Z] &= \left[ \begin{array}{l} y_i \cos \phi - f \sin \phi \sin \theta \\ -\left( y_i \sin \phi + f \cos \phi \right) \\ \left( y_i \cos \phi - f \sin \phi \cos \theta \right) \end{array} \right] = 0; \\
Y &= -H.
\end{align*}
$$

(7)

which can be expended as the following to yield its direction and intercept with the rail track $(X=0)$.

$$Z = -\tan \theta \cdot X + \frac{f \cos \phi + y_i \sin \phi}{f \sin \phi - y_i \cos \phi} \cdot H$$

(8)

In the same way as (8), the inner rectangle has the image distance $v$ from the outer rectangle (i.e., $Y = y_i + v$) at the bottom lines, which is the image velocity to shift between consecutive frames when the train moves at speed $V$. The bottom line $l_2$ is projected to rail and ground surface as $L_2$,

$$
\begin{align*}
[X \ Y \ Z] &= \left[ \begin{array}{l} y_i \cos \phi - f \sin \phi \sin \theta \\ -\left( y_i \sin \phi + f \cos \phi \right) \\ \left( y_i \cos \phi - f \sin \phi \cos \theta \right) \end{array} \right] = 0; \\
Y &= -H.
\end{align*}
$$

(9)

The 3D distance that the train moves between two successive frames is $V/R$, where $R$ is the frame rate of the camera (25fps). This must be equivalent to the intercept difference $D_{l_2}$, of lines $L_1$ and $L_2$ on the Z axis, and is computed as,

$$D_{l_2} = \frac{H}{f \sin \phi - (y_i \cos \phi)}$$

(10)

From this relation, the image velocity at $l_1$ can be calculated from the known train speed $V$ as

$$v = \frac{(f \sin \phi - y_i \cos \phi)^2}{fHR + (f \sin \phi - y_i \cos \phi) \cos \phi \cos \theta \cdot V}$$

(11)
The camera height, $H$, is a constant converted from the rail widths both in 3D (nationwide standard) and in the image after the camera with the focal length $f$ is set.

After bottom line $l_2$ is determined, corners $A_2$ and $B_2$ of the inner rectangle are further determined at its crossing points with the radial lines from the vanishing point as in Fig. 4. Side strips, $S_r$ and $S_l$ of the coil are thus determined in width from the corners, which guarantee the just-sampling properties at the rail side infrastructures including fence, pole, and so on. Beyond the depths of rail side planes, landscapes have overlapped-sampling [18]. However, it is not our current focus to represent landscapes outside the railway areas. Although the train passing space closer than the side planes is under-sampled, there are no obstacles to be inspected in the space within the box tube. The top strip, $S_t$, is also determined after the side strips are located precisely to scan wires above the train.

The relation between the $v$ and train speed $V$ is obtained using a sample video with a varied speed. The resulting data are stored in a lookup table for the coil selection in the video scanning.

4. PANORAMA IMAGE RENDERING

Virtual scene generation is an essential part for the fast data browsing and interactive display. We implement the virtual scene observation by allowing speed control and perspective transform via rendering four panoramas on the tube box. The steps to render panorama images are as follows: (1) Copy and transform strips in each frame to four panoramas consecutively. (2) Use four polygons to construct a tube box scene model; (3) Set a view point and field of view in the box model for scene observation; (4) Perform the texture mapping from the panoramas to the scene model and display pixels on the computer screen.

The panorama is constructed offline and stored as a specific big-data file which encapsulates the image data by the bit stream using C++ Programming. The big-data file appends an indexing file for fast location retrieval.

The virtual scene can be rendered based on the panorama image very fast and receive a considerable effect with real sensation, since panorama has a small amount of data and very broad extension of perspective. The distance information can be obtained using the known conditions of the train and environments. We can even divide the scene according to the depth and render them onto the different plane layers for a stereoscopic scene.

5. EXPERIMENTS AND DISCUSSION

This work aims at obtaining a complete archive of the route scenes along a high-speed railway lasting over 100km between two major cities. The entire video has been taken from a patrol train in the forward direction with the headlight of the train always on. Such videos may include various illumination conditions and weathers. As shown in Fig. 7, a frame of the forward motion video is captured at a relatively constant speed of 156km/h and ordinary frame rate (25fps) on a smooth path. The result of route panorama acquisition is robust and the success rate for stitching is high.

The result of panorama acquisition is displayed in Fig. 8. Our method generates the desirable panorama, even if the original video quality is poor. Taking left side panorama as an example, we can observe that the fences and poles close to the camera very clearly; no information is lost and the distortion is also small. The sights more distant are suffered from an apparent distortion, which is consistent to previous analysis. In general, the result meets the current monitoring requirements. The distortion of the distant scenery can be further solved for virtual sightseeing from the train in the future. The unevenness of the track shown in the image is due to the fact that the train as well as the camera experiences a pitch shaking at a non-smooth connection spot on the rail track. Such a location needs to be examined further to ensure the safety of the rail.
A virtual interactive environment can be rendered easily after generating the four parts of panorama. We have given a panoramic scene as shown in Fig.9 for virtual browsing using OpenGL in which the viewer can move back and forth freely, and rotate left or right through keyboard interaction. It is easy to examine the rail back and forth with a fast scrolling and a zoom-in function for the long scene. This facilitates a coarse-to-fine investigation further down to video frames. On the other hand, an automatic method has been developed to scan the railway instruments appearing at a regular interval. The challenge now is the structure extraction under various illuminations and background scenes outside the track area, which has not been explored on the generated scene.

Fig.9. Panoramic scene rendered from four generated panoramas

6. CONCLUSION

The panoramic representation of the route scenes is first generated for a long distance railway in this work. We solved the problem of fast indexing and profiling of a large volume of video data. We proposed a fast and concise method to generate a desirable panorama using simple geometric computation with some known priors. In our method, a rectangular ring is located according to the railroad structure to better reflect the shapes of interest for monitoring, and the width is adjusted according to the known train speed. The panoramic virtual scene successfully archived the route scenes in a compact format.

Our future works will be the automatic comparison of panorama image generated at different days and periods to find differences related to weather, illumination and background scenes. The differences on the railway will be detected at spots for further investigation and repairing. In addition, the panorama rendering would be improved to make a more realistic scene for sightseeing from the train through appending lights, blending, shading, anti-aliasing as well as exploring depth information.

REFERENCES