Visual Indexing of Large Scale Train-Borne Video for Rail Condition Perceiving

Peng DAI[†], *Nonmember*, Shengchun WANG^{†a)}, *Member*, Yaping HUANG^{††b)}, Hao WANG[†], Xinyu DU[†],
and Qiang HAN[†], *Nonmembers*

SUMMARY Train-borne video captured from the camera installed in the front or back of the train has been used for railway environment surveillance, including missing communication units and bolts on the track, broken fences, unpredictable objects falling into the rail area or hanging on wires on the top of rails. Moreover, the track condition can be perceived visually from the video by observing and analyzing the train-swaying arising from the track irregularity. However, it's a time-consuming and labor-intensive work to examine the whole large scale video up to dozens of hours frequently. In this paper, we propose a simple and effective method to detect the train-swaying quickly and automatically. We first generate the long rail track panorama (RTP) by stitching the stripes cut from the video frames, and then extract track profile to perform the unevenness detection algorithm on the RTP. The experimental results show that RTP, the compact video representation, can fast examine the visual train-swaying information for track condition perceiving, on which we detect the irregular spots with 92.86% recall and 82.98% precision in only 2 minutes computation from the video close to 1 hour.

key words: route panorama, video indexing, train-borne video, train-swaying, track irregularity

1. Introduction

Track irregularity arises from the geometrical deviations between the rails' actual location and the ideal location at the horizontal or vertical direction, which produce an external shock disturbance to the rolling stock system, and are the major cause of locomotives vibration. The violent shaking will badly affect the ride comfort of the passenger, and even will derail the train or leads to collision if the fault further takes a change for the worse. According to statistics on train accidents by Office of Safety of US Federal Railroad Administration, 25027 accidents are mainly caused by Track Geometry from January 1975 to September 2014, which result in 24705 (98.7%) derailments, 124 collisions (0.50%) and 198 (0.80%) others*. Therefore, the detection to the track irregularity is extremely important, which has a significant influence on the safety of train operation. The description and analysis to the irregularity are important for

rolling stock, track design, track maintenance, and quality assessment.

Nowadays, the patrol train has been widely utilized for ensuring the safety of railway. The cameras installed in the various spots of a moving train can be used for different railway inspection tasks such as the track profile measurement, rail track defect detection, rail bolts detection and the inspection of overhead contact system [1]–[4]. One of which is to install a camera in front of the patrol train to capture the whole video along the railway for environment surveillance, including missing communication units and bolts on the track, broken fences, unpredictable objects falling into the rail area or hanging on wires on the top of rails. Moreover, the whole track condition can be perceived visually by browsing the total video to observe and analyze the train-swaying arising from the track irregularity. Traditionally, this work can be performed manually by trained persons who monitor the video. Human inspection is time-consuming, low-efficient and labor-intensive to examine the whole large scale video captured by patrol train repeatedly, and difficult to recognize the unstable video segment arising from the train-swaying.

However, the video captured by the patrol train is of enormous size due to the frequent inspection tasks and long route. It's very challenging to archive the large-scale video data daily and achieve the information retrieval manually. Therefore, the key issue is how to generate a compressed visual representation from large scale video without information loss for fast data access, which is more feasible to implement the automatic decision algorithm than monitoring the video.

Zheng proposed an effective method named route panorama [5]–[7] to scan the temporal scenes in the side direction of the path and connect them to the spatial image. Different from a local panorama at a static viewpoint [8], [9], the route panorama is constructed from partial views at consecutive viewpoints along a path. The route panorama is compressed for temporal sampling, and has covered the entire scenes for the variable viewpoint. Wang [10], [11] introduced a railway scene rendering system for virtual maneuver and inspection based on four route panoramas. The compact data format with sufficient visual information can be transmitted and released in cyberspace for fast access, and automatic retrieval.

Manuscript received November 24, 2016.

Manuscript revised April 14, 2017.

Manuscript publicized June 14, 2017.

[†]The authors are with Infrastructure Inspection Research Institute, China Academy of Railway Sciences, Beijing, 100081, China.

^{††}The author is with the Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing, 100044, China.

a) E-mail: 15600699615@163.com

b) E-mail: yphuang@bjtu.edu.cn

DOI: 10.1587/transinf.2016PCP0020

*<http://safetydata.fra.dot.gov>

In this paper, we propose a simple and effective method to detect the train-swaying quickly and automatically based on previous work [10], [11]. We first generate the long rail track panorama (RTP) by stitching the stripes cut from the video frames, analyze the relationship between the unevenness on the RTP and train-swaying, and extract track profile from the panorama. The unevenness of the track profile 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, i.e. track irregularity. Such a location needs to be examined repeatedly to ensure the safety of the rail. Therefore, the condition of rail track can be assessed intuitively after we obtained the RTP from train-borne video.

2. Perceiving Principle and Track Irregularity

As shown in Fig. 1, The train-swaying often arises from track irregularity, such as different elevation between the two rails will lead to the train rolling around the Z axis; the ascending or descending track plane will result in the train tilting around the Y axis; and the sudden side bend will cause the train panning around the X axis. Note that the camera we mounted has a rigid connection with train, so that the train-swaying is bound to cause camera-shaking which give rise to the shaky video scene. Therefore, if we can extract the shaky information from the video, the information relative to the train-swaying and track condition can be inversely deduced in succession, which is a typical inverse problem. In this paper, we generate the RTP to reveal the camera-shaking and further perceive the rail condition.

Track irregularity is the deviation including track geometry, rail size and spatial position. Generally speaking, uneven or non-straight rail, the deviation of rail's width and height relative to the central line position, non-smooth curved track, deviation from the central line of curved track, superelevation and abnormal gauge value etc. are known as track irregularities.

According to the direction of the excitation to train bodywork, there are five types of irregularities mainly (See Fig. 2). Figure 2(a) shows the rail track fluctuation up and down, which introduces a tilting angle β ; Fig. 2(b) shows the rail track bend left and right, which introduces a panning angle φ ; Fig. 2(c) shows the rail track lean clockwise or anticlockwise, which introduces a rolling angle γ ; Fig. 2(d) shows the abnormal gauge, which introduces a synthetical angle by φ and β ; Fig. 2(e) shows the twist of track, which introduces a synthetical angle by β and γ .

Track irregularity usually appear in the form of transversal wave. As shown in Fig. 2, the L is the wavelength, Δx and Δy are the wave crest, and G is the standard gauge. Therefore, the three angles φ , β and γ can be denoted as below,

$$\begin{aligned}\varphi &= \arctan\left(\frac{2\Delta y}{L}\right); \\ \beta &= \arctan\left(\frac{2\Delta x}{L}\right); \\ \gamma &= \arctan\left(\frac{\Delta x}{G}\right).\end{aligned}\quad (1)$$

According the criteria of track irregularity management

Fig. 1 Train-swaying arising from track irregularity

Fig. 2 The types of track irregularity

and maintenance, the value of wave crest Δx and Δy should be limited under 10mm while the wavelength L is 10m for ensuring railway safety when train run at the speed of $90\text{--}120\text{ km/h}$ [12], [13], i.e. a good detection system should be sensitive to the track irregularity greater than 10mm . As the precision criteria for proposed method should meet the requirements of perceiving the train-swaying accurately when $\varphi > 0.115^\circ$, $\beta > 0.115^\circ$, $\gamma > 0.401^\circ$ (Calculated by Eq. (1)).

3. RTP Acquisition

A forward camera in the train captures video as the train moves on the track. The discrete video frames contain overlapped scenes with non-uniformed resolutions at different depths on the rail. Highly repeated rail patterns between consecutive frames make the monitoring process redundant on the examined portion of track, which can be eliminated by RTP acquisition.

As shown in Fig. 3, we scan the rail scene with a stripe in order to obtain an accumulative image. The resulting image should neither miss small scenes nor repeat scenes that cause redundancy. The stripe is set in the bottom of video frame which covers the rail that the train moves on. The rail scene scanned by the stripes is at the same depth so that they will have a stable resolution in the obtained panoramic image. If the train moves on a smooth track in good condition, the stitching result from all the stripes would display a panorama with straight rail track, otherwise the distorted rail track would be observed in the panorama when the train goes through the irregular railway division.

As shown in Fig. 4, $O - XYZ$ is the World Coordinates System where the Z axis is in the train moving direction

Fig. 3 RTP acquisition by stitching the stripes sampling in each frame

Fig. 4 The relationship between the image velocity v and the train speed V to realize the just-sampling requirement on the rail track.

translating at speed V , and the X axis is perpendicular to the ground. The camera coordinates system is $o - xyz$. We capture a video segment as the train moves on the tracks, and will determine the image velocity, v , at the bottom strip over the rail track for the just-sampling stripe. The just-sampling means the perfect connection of scenes between two successive frames, yielding no information loss as well as pixels overlap. Without loss of generality, we consider the camera face front has no obliquity that means the two coordinate systems coincide.

The sampling stripe is constructed by the outer sampling line l_1 and inner sampling line l_2 . The location of l_1 is fixed in every frame, and l_2 can be determined when image velocity v is given. Two image lines in the world coordinates system $O - XYZ$ are denoted as

$$\begin{aligned} l_1 &: A_1(X_1, Y_{a1}, Z_{a1}) B_1(X_{b1}, Y_{b1}, Z_{b1}); \\ l_2 &: A_2(X_2, Y_{a2}, Z_{a2}) B_2(X_2, Y_{b2}, Z_{b2}). \end{aligned} \quad (2)$$

The plane of sight through l_1 is thus determined from its normal $N_1 = OA_1 \times OB_1 = (X_1, Y_{a1}, Z_{a1}) \times (X_1, Y_{b1}, Z_{b1})$ 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 $X = -H$, where H is the height of the camera from ground.

$$\begin{cases} [X \ Y \ Z] \begin{bmatrix} X_1 Z_{b1} - Z_{a1} X_1 \\ Z_{a1} Y_{b1} - X_{a1} Y_{b1} \\ Y_{a1} X_1 - X_1 Y_{b1} \end{bmatrix} = 0; \\ X = -H. \end{cases} \quad (3)$$

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

$$Z = -\frac{f}{X_1} \cdot H \quad (4)$$

Where f is the calibrated focal length.

In the same way as (3), the inner sampling line l_2 has the image distance v from the outer line (i.e., $X_2 = X_1 + v$) 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 ,

$$\begin{cases} [X \ Y \ Z] \begin{bmatrix} (X_1 + v) Z_{b1} - Z_{a1} (X_1 + v) \\ Z_{a1} Y_{b1} - Y_{a1} Z_{b1} \\ X_{a1} (X_1 + v) - (X_1 + v) X_{b1} \end{bmatrix} = 0; \\ X = -H. \end{cases} \quad (5)$$

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 of lines L_1 and L_2 on the Z axis, and is computed as,

$$D_{L_1 L_2} = \left(-\frac{f}{X_1 + v} + \frac{f}{X_1} \right) H = V/R \quad (6)$$

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

$$v = \frac{V \cdot X_1^2}{V \cdot X_1 - fHR} \quad (7)$$

The just-sampling stripe would be obtained when the image velocity v is determined, the pixels on the sampling patch are mapped onto the real plane (see Fig. 3). The sampled strips from each frame are aligned by homography transformation to obtain the regular rectangle, and then stitching all the stripes from total video frames to generate the RTP.

4. The Unevenness on the RTP and Train-Swaying

Without loss of generality, we assume the width of just sampling stripe is one pixel for simplified analysis in this section, and only consider the rail track close to the camera. In fact, it's not appropriate to analyze another track far away from the camera, which will be discussed in more detailed in Fig. 10.

As shown in Fig. 5 (a), the train run on the smooth rail track along the vector $T(0, 0, V/R)$. A_i and B_i mark two fixed pixel points in the projection area of rails, where i is the current frame number. A'_i and B'_i are the object points in the physical space corresponding to the pixel points A_i and B_i on the rail surface. $I(A_i)$ and $I(B_i)$ denote the gray value of the pixel points A_i and B_i , which are determined by the

(a) Camera sampling on the straight rail track

(b) The static pixel points A_i and B_i **Fig. 5** Even RTP generated from smooth sampling

brightness value $E(A'_i)$ and $E(B'_i)$ of the points A'_i and B'_i , i.e.

$$\begin{cases} I(A_i) = T(E(A'_i)) \\ I(B_i) = T(E(B'_i)) \end{cases} \quad (8)$$

Where $T(\cdot)$ represent the photoelectric conversion.

The physical points on the rail surface have the higher brightness and the value changes very little along the Z axis. If the train retains the smooth running, A_i and B_i will keep sampling the rails, and the pixel value of them will not change (see Fig. 5 (b)). Therefore, we call A_i and B_i as static pixel points (SPP), accordingly A'_i and B'_i are named static rail sampling points (SRSP).

We notice that the static pixel value will always keep static sampling on the rails other than the SRSP stray away from the Z axis. The coordinates of the SRSP at the Y axis and Z axis will both change when camera suffers the shaking introduced by train-swaying. However, the coordinate change along Z axis won't affect the pixel value of SPP due to the constant brightness. So, only the coordinate change of SRSP at the Y axis works on the pixel variation of SPP.

The points A_i and B_i in the image coordinates system $o-xyz$ are denoted as

$$\begin{cases} A_i = (a, -d, f+c) \\ B_i = (b, -d, f+c) \end{cases} \quad (9)$$

Where c is the distance the train has run; a, b, d are given values.

The points A'_i and B'_i in the world coordinates system $O-XYZ$ are denoted as

$$\begin{cases} A'_i = (-H, Y_a^i, Z_a^i) \\ B'_i = (-H, Y_b^i, Z_b^i) \end{cases} \quad (10)$$

For the straight rail track shown in Fig. 5 (a), the coordinates of SRSP can be solved according to the principle of the pinhole camera,

$$\begin{cases} A'_i = (-H, \frac{a \cdot H}{d}, \frac{(f+c) \cdot H}{d}) \\ B'_i = (-H, \frac{b \cdot H}{d}, \frac{(f+c) \cdot H}{d}) \end{cases} \quad (11)$$

(a) Panning Camera sampling on the rail track from side to side

(b) The offset of the static rail sampling points towards panning angle

(c) The unevenness introduced by camera panning

Fig. 6 The relationship between unevenness and panning camera

i.e. the position of two rails r_1 and r_2 can be denoted as below,

$$r_1 : \begin{cases} X = -H \\ Y = \frac{a \cdot H}{d} \end{cases} ; \quad r_2 : \begin{cases} X = -H \\ Y = \frac{b \cdot H}{d} \end{cases} \quad (12)$$

Based on the foregoing analysis, we can estimate the unevenness on the RTP using the following equation

$$D_{r_1}^i = Y_a^i - \frac{a \cdot H}{d}; \quad D_{r_2}^i = Y_b^i - \frac{b \cdot H}{d} \quad (13)$$

Where $D_{r_1}^i$ and $D_{r_2}^i$ indicate the offset of the SRSP stray away from the rails r_1 and r_2 due to the train-swaying, which will bring about the unevenness at the i -th column of the RTP.

As shown in Fig. 6 (a), if the camera is shaking around x axis, i.e. the camera with panning angle φ is sampling the rail track from side to side. The offset of the SRSP $D_{r_1}^i$ and $D_{r_2}^i$ can be obtained by the coordinates system transformation and principle of the pinhole camera,

$$\begin{cases} D_{r_1}^i = \left(\cos \varphi - \frac{f}{a} \sin \varphi - 1 \right) \cdot \frac{aH}{d} \\ D_{r_2}^i = \left(\cos \varphi - \frac{f}{b} \sin \varphi - 1 \right) \cdot \frac{bH}{d} \end{cases} \quad (14)$$

(a) Tilting camera sampling on the rail track up and down

(b) The offset of the static rail sampling points towards tilting angle

(c) The unevenness introduced by camera tilting

Fig. 7 The relationship between unevenness and tilting camera

Now that a, b, H, d, f are known, we can obtain the offset of the static rail sampling points towards panning angle. As shown in Fig. 6(b), A'_i and B'_i move at the same velocity towards the positive (negative) direction of Y axis together when the camera is panned clockwise (anticlockwise), which will lead to the pixels of rails bending to the same direction, so the rails on the RTP will bend to one side as shown in Fig. 6(c).

As shown in Fig. 7(a), if the camera is shaking around y axis, i.e. the camera with tilting angle β is sampling the rail track up and down. The offset of the $SRSP D_{r_1}^i$ and $D_{r_2}^i$ can be expressed as,

$$\begin{cases} D_{r_1}^i = \left(\frac{1}{\cos \beta - \frac{f}{d} \sin \beta} - 1 \right) \cdot \frac{aH}{d} \\ D_{r_2}^i = \left(\frac{1}{\cos \beta - \frac{f}{d} \sin \beta} - 1 \right) \cdot \frac{bH}{d} \end{cases} \quad (15)$$

We can obtain the offset of the static rail sampling points towards tilting angle. As shown in Fig. 7(b), A'_i moves towards the positive (negative) direction of Y axis but B'_i towards the negative (positive) direction when camera is tilted clockwise (anticlockwise). However, both of them keep the same velocity. Note that there is an asymptotic line in Fig. 7(b) when the camera faces the upper side.

(a) Rolling camera sampling on the rail track left and right

(b) The offset of the static rail sampling points towards rolling angle

(c) The unevenness introduced by camera rolling

Fig. 8 The relationship between unevenness and rolling camera

This is because the sight plane through the line $A_i B_i$ has been parallel with the ground plane, so that no intersection exists between them. The width between two rails will expand or shrink on the RTP , as shown in Fig. 7(c).

As shown in Fig. 8(a), if the camera is shaking around the z axis, i.e. the camera with rolling angle γ is sampling the rail track left and right. The offset of the $SRSP D_{r_1}^i$ and $D_{r_2}^i$ can be expressed as

$$\begin{cases} D_{r_1}^i = \left(\frac{1 - \frac{d}{a} \tan \gamma}{1 + \frac{d}{a} \tan \gamma} - 1 \right) \cdot \frac{aH}{d} \\ D_{r_2}^i = \left(\frac{1 - \frac{d}{b} \tan \gamma}{1 + \frac{d}{b} \tan \gamma} - 1 \right) \cdot \frac{bH}{d} \end{cases} \quad (16)$$

As shown in Fig. 8(b), A'_i and B'_i move at the different velocity towards the positive (negative) direction of Y axis in together when the camera is rolled clockwise (anticlockwise), which will lead to the pixels of rails bend to the same direction in varying degrees, so the rails on the RTP will bend to one side with different curvature as shown in Fig. 8(c).

5. Experimental Results

5.1 Experiment Data

The video we used in the experiment is captured by a calibrated camera ($f = 15.46\text{mm}$) installed in the head of the train, as shown in Fig. 9 (a). Such videos may include various illumination conditions and weathers. A frame of the forward motion video is captured at a relatively constant speed of 120km/h and ordinary frame rate (25fps) on a rail path (See Fig. 9 (b)). For recording the train vibration clearly, a metallic mount is fixed to make a rigid connection without any damping between the camera and train. The image resolution of captured video is 1024×768 , and the pixel size of each frame is $10\text{ }\mu\text{m}$. The length in the real world corresponding to a pixel in the video is inversely proportional to the image's depth, since the video is captured in the forward motion way. However, the generated panorama has transformed the non-uniform resolution of forward motion video into an extended image with definite resolution since we extracted the stripes at almost the same location from each video frame and then stitched them into a panorama (See Fig. 3). The rail tracks are standard gauge ($1,435\text{mm}$) and the pixel width of rail in the panorama is about 135, so the length in the real world corresponding to a pixel in the panorama, i.e. spatial resolution can be determined by $\Delta L = 1435/135 = 10.63\text{mm}$. Therefore, the panorama can show the unevenness when the deviation value of $SRSP$ $D_{r_1}^i$ and $D_{r_2}^i$ is greater than 10.63mm . From the Eqs. (13), (14) and (15), we obtain that the minimum angles that our method can perceive are $\varphi = 0.039^\circ$, $\beta = 0.021^\circ$ and $\gamma = 0.138^\circ$, which meet the precision requirements discussed in Sect. 2.

In most cases, the train runs with an almost constant speed, and the width of the just-sampling region is fixed for most frames. Therefore, we only recalculate the just-sampling region when the train speed takes an obvious change, which eliminates the redundant computing task and greatly improves the efficiency of panorama acquiring.

Partial result of *RTP* acquisition is displayed in Fig. 10. The width of the just-sampling stripe is determined by Eq. (7), i.e. $v = 12$. The unevenness of the track shown is due to the fact that the train as well as the camera experiences a pitch shaking at a non-smooth connection spot of the rail tracks. Such a location needs to be examined further to ensure the safety of the rail.

We extract upper track profile including two rails marked r_1 and r_2 from the panorama, because the edge of the under track far away from the camera is not sharp enough due to the serious influence from homography transformation and amplifying noise. As discussed in the next, we perform the automatic detection algorithm on the profile image and count the total number of the unevenness easily.

Fig. 9 The case of the straight and smooth rail: (a) the camera installed in the head of the train; (b) the video frame captured in the forward direction.

Fig. 10 Partial display of the *RTP* acquired from forward motion video

5.2 The Unevenness Detection on the *RTP*

The rail pixels in the generated *RTP* have almost the maximum gray value, so it's very fast and efficient to extract the rail area by image binarization for further analysis. The key is how to choose an appropriate threshold ζ . The illumination variation is inevitable to our task due to the train runs in a long distance, and the weather, surrounding etc. frequently change. So, it's infeasible to perform the global binarization on the entire *RTP*. An alternative is to cut the long panorama into many sub-images, and choose multiple local thresholds ζ_n for better result. We divide the 100000 columns of *RTP* into one hundred of parts, where every part is a 1000 columns of sub-image, which is denoted as

$$RTP = \sum_{n=1}^{100} I_n \quad (17)$$

The local thresholds ζ_n can be chosen according to the mean of gray values of I_n , denoted as

$$\zeta_n = \frac{k}{MN} \sum_{i=1}^M \sum_{j=1}^N I_n(i, j) \quad (18)$$

Where $M = 230$, $N = 1000$ are the size of the sub-image, and k is an empirical parameter for threshold adjustment. We set the value $I_n(i, j) = 0$ if $I_n(i, j) < \zeta_n$, else set $I_n(i, j) = 255$. The binary image contains several disconnected or close isolated points, which can be eliminated easily by morphological operation. The result of rail extraction is shown in Fig. 11 (b), and the rail track profile lines with only one pixel are obtained after calculating the gradient of the binary rail image, which indicate the pixel locations in the image of the rails r_1 and r_2 , denoted as $p_j, q_j (j = 1, 2, \dots, 1000)$.

The vertical displacements of the rails r_1 and r_2 are calculated by

Fig. 11 The case of the straight and smooth rail: (a) the upper panorama cut from the RTP; (b) rail extraction by image binarization and morphological operation; (c) rail track profile line by calculating the gradient of the image in (b)

$$d_j^1 = p_j - h_1; \quad d_j^2 = q_j - h_2 \quad (19)$$

where h_1 and h_2 are the regular positions of the rail r_1 and r_2 , which can be estimated by the statistics of straight lines in Fig. 11 (c).

The vertical displacements d_j^1 and d_j^2 are essentially incurred by the offsets $D_{r_1}^i$ and $D_{r_2}^i$ of the static rail sampling points as we described in Sect. 4, and the following three experimental results also maintain a high level of consistent with the previous theoretical analysis in Sect. 4.

Figure 12 (a) shows the case of rotating around the y axis, i.e. rail track fluctuation introduced by the elevation of both rail tracks changing suddenly. Figure 12 (d) shows the corresponding result of the vertical displacement of r_1 and r_2 . We can see that the bend of the two rail tracks are in the opposite directions.

Figure 13 (a) shows the case of rotating around the x axis, i.e. rail track bend introduced by the sudden side rail bend or train turning. Figure 13 (d) shows the corresponding result of the vertical displacement of r_1 and r_2 . The two conditions can't be distinguished by the instantaneous vertical displacements of the two pixel points on the rails for two offsets are in the same direction. However, train turning is enduring bend while sudden side bend is shorter-rate, which can be separated by observing the duration of the continuous vertical displacements of r_1 and r_2 .

Figure 14 (a) shows the case of rotating around the z axis, i.e. rail track lean introduced by different elevation between the two rails. Figure 14 (d) shows the corresponding result of the vertical displacement of r_1 and r_2 . We can see

Fig. 12 The case of rotation around y axis: (a) the upper panorama cut from the RTP; (b) rail extraction by image binarization and morphological operation; (c) rail track profile line by calculating the gradient of the image in (b); (d) vertical displacement of rail r_1 and r_2

that the pixels of rails r_1 and r_2 bend to the same direction with different displacement, the rail r_2 bend to one side with larger curvature and displacement. In addition, we also observed the case of rotating around the x axis. The two types of rotation didn't take place at the same time, and the clean unevenness information can be distinguishable. This is quite different from the case of train turning, some unevenness can't be distinguishable from mixed information.

When we obtain the vertical displacement of rails (See Figs. 12 (d), 13(d), 14(d)), we conclude the corresponding relation between the vertical pixel displacement d_j^1 , d_j^2 on the RTP and rail condition. The decision rules for perceiving the irregularity of single rail according to the vertical displacement, while ε is used to further classify the irregularity according to the difference between the two rails' vertical displacements. The two empirical parameters ξ and ε must be chosen appropriately; if not, the improper ξ will introduce more false positive or false negative results, while the improper ε will confuse the type of rail track irregularity between rail track bend and lean.

The ground truth we used in the experiment is obtained by inertial laser sensor installed in the bodywork structure and posterior manual review, which includes 13 rail track bend, 10 rail track fluctuation and 19 rail lean. We evaluate the proposed method by defining recall ratio and precision

Fig. 13 The case of rotation around x axis: (a) the upper panorama cut from the RTP; (b) rail extraction by image binarization and morphological operation; (c) rail track profile line by calculating the gradient of the image in (b); (d) vertical displacement of rail r_1 and r_2

Table 1 The vertical pixel displacement and rail condition (ξ and ε are two empirical parameters)

d_j^1 and d_j^2	Train-swaying type	Rail condition
$ d_j^1 > \xi; d_j^2 > \xi; d_j^1 \cdot d_j^2 > 0$ and $ d_j^1 - d_j^2 < \varepsilon$	sway around x axis	rail track bend
$ d_j^1 > \xi; d_j^2 > \xi; d_j^1 \cdot d_j^2 < 0$	sway around y axis	rail track fluctuation
$ d_j^1 > \xi; d_j^2 > \xi; d_j^1 \cdot d_j^2 > 0$ and $ d_j^1 - d_j^2 > \varepsilon$	sway around z axis	rail track lean

ratio as below,

$$\begin{aligned} \text{Precision} &= TP / (TP + FP); \\ \text{Recall} &= TP / (TP + FN). \end{aligned} \quad (20)$$

Where TP is the number of true positive results; FP is the number of false positive results and FN is the number of false negative results. Table 2 shows the variation of TP , FP and FN with the value of ξ in the range from 1 to 10, while Table 3 show the variation of the number of rail track bend (NRB) and rail Track lean (NRL) with the value of ε in the same range when $\xi = 5$.

5.3 Evaluation

The computational environment we use in the experiment is that Intel(R) Core(TM) i5-2500@3.30GHz Processor, 8 GB Memories, and Windows 7 OS. In addition, the algorithm is

Fig. 14 The case of rotation around z axis: (a) the upper panorama cut from the RTP; (b) rail extraction by image binarization and morphological close operation; (c) rail track profile line by calculating the gradient of the image in (b); (d) vertical displacement of rail r_1 and r_2

Table 2 The variation of FP and FN with the value of ξ

ξ	1	2	3	4	5	6	7	8	9	10
TP	41	40	39	39	39	37	36	36	35	34
FP	21	16	14	10	8	8	7	7	7	7
FN	1	2	3	3	3	5	6	6	7	8

Table 3 The variation of NRB and NRL with the value of ε ($\xi = 5$)

ε	1	2	3	4	5	6	7	8	9	10
NRB	7	9	14	16	19	21	22	25	25	26
NRL	27	25	20	18	15	13	12	9	9	8

Table 4 Compression ratio between video and RTP

	Duration(min)	video Size(MB)	RTP Size(MB)	Ratio
Seg.1	51	937	29	32:1
Seg.2	68	1268	42	30:1
Seg.3	57	1053	34	31:1
Seg.4	53	984	31	32:1

implemented by C++ based on Visual Studio 2012 platform.

As shown in Table 4, four segments of video are used for performance evaluation. Compared with the original video with *avi* format, the generated panoramas have smaller size with *jpg* format.

The time costs of RTP acquisition from close to 1 hour long video is about 90 seconds and extra 50 seconds to detect the unevenness on the RTP automatically, which is more

Table 5 Time cost comparison between automatic and manual inspection

	Duration(min)	Automatic cost(min)	Manual cost (min)
Seg.1	51	2.2	70
Seg.2	68	2.5	91
Seg.3	57	2.3	81
Seg.4	53	2.3	75

Fig. 15 The precision and recall curve **Fig. 16** The numerical distribution of rail bend and lean

efficient than manual inspection with eyes by four skilled workers as shown in Table 5.

According to the data from Table 2 and Eq. (20), the Precision and Recall curve can be given directly. As shown in Fig. 15, we have observed that the precision ratio improves with the increase of ξ due to the decrease of the false positive results, while the recall ratio declines due to the increase of the false negative results. From the point of view of ensuring the safety of railway operation, the recall ratio should be perfect while accepting the decreased precision ratio to a certain extent. We assign ξ the value 5, and obtain 92.86% recall and 82.98% precision in this experiment.

After ξ is determined, ε should be chosen to separate the type of detected faults. In Fig. 16, $TNRL$ and $TNRB$ are the true number of rail lean and bend, respectively. An appropriate ε should separate the NRL and NRB as close as possible to the $TNRL$ and $TNRB$. From the Fig. 16, we can observe that the NRL and NRB are closest to the $TNRL$ and $TNRB$ when $\varepsilon = 3$.

The recall of our method is up to 93%, which is a desirable result. However, the false positive is a little high due to the stochastic occasion such as sudden acceleration and deceleration, crosswind or other unpredictable factor. The

Fig. 17 The comparison between proposed method and manual monitoring

error can be reduced by comparing the multiple results from the same railroad, assuming that the unevenness with similar shape will always appear at the irregular location.

The detection result of our method is good at repeatability by compared with manual monitoring by four skilled workers, as shown in Fig. 17. The result of manual monitoring is dependent on workers' experience and conscientiousness, which is subjective and unstable.

6. Conclusion

In this paper, we propose an new detective method to perceive the condition of rail track. A fast and concise method is introduced to generate a desirable panorama for indexing large-scale video, only using simple geometric computation with some known priors (calibrated focal length, train speed and position of the camera form the ground). In our method, a rectangular stripe is located to scan the rail track, and the width is adjusted according to the known train speed. The RTP can reflect the real rail scene, the irregular track gives rise to the unevenness of the track profile in the panorama 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.

The rail track panorama indexes the entire large volume of video. It's more timesaving and stable to examine the panorama automatically than the entire video manually, where only 2 minutes are required to find the irregular spots on the RTP generated from the video close to 1 hour, and obtain a desirable recall ratio up to 93%.

The future work will focus on the more complicated railway division such as switches or crossing road, and analyzing the unevenness track in the panorama systematically for formulating the relationship between the RTP distortion and rail track irregularity.

Acknowledgments

The authors wish to acknowledge the support and motivation provided by the Major Scientific Research Projects of China Railway Corporation under Grant 2015T003-A, J2016G003, Beijing Natural Science Foundation granted No.4152042 and National Basic Research Program of China (973 Program) granted No.2013CB329400.

References

- [1] C. Alippi, E. Casagrande, F. Scotti, and V. Piuri, "Composite Real-Time Image Processing for Track Profile Measurement," *IEEE Trans. Instrumentation and Measurement*, vol.49, no.3, pp.559–564, 2000.
- [2] L. Jie, L. Siwei, L. Qingyong, Z. Hanqing, and R. Shengwei, "Real-time rail head surface defect detection: A geometrical approach," *IEEE Int. Symp. Industrial Electronics*, pp.769–774, 2009.
- [3] Q.Y. Li and S.W. Ren, "A Visual Detection System for Rail Surface Defects," *IEEE Trans. System, Man, and Cybernetics*, vol.42, no.6, pp.1531–1542, 2012.
- [4] A. Rubaai, "A neural-net-based device for monitoring Amtrak railroad track system," *IEEE Trans. Industry Applications*, vol.39, no.2, pp.374–381, 2003.
- [5] J.Y. Zheng, "Digital Route Panoramas," *IEEE Multimedia Mag.*, vol.10, no.3, pp.57–67, 2003.
- [6] J.Y. Zheng, Y. Zhou, and P. Mili, "Scanning scene tunnel for city traversing," *IEEE Trans. Vis. Comput. Graph.*, vol.12, no.2, pp.155–167, 2006.
- [7] J.Y. Zheng and M. Shi, "Scanning depth of route panorama based on stationary blur," *International Journal of Computer Vision*, vol.78, no.2-3, pp.169–186, 2008.
- [8] D. Kim and J. Hahn, "Projective Texture Mapping with Full Panorama," *Computer Graphics Forum*, vol.21, no.3, pp.421–430, 2002.
- [9] X.X. Wu, B.L. Guo, and J. Wang, "Cylindrical Panoramic Image Automatic Mosaic Algorithm Based on Phase Correlation," *Acta Optica Sinica*, vol.29, no.7, pp.1824–1829, 2009.
- [10] S.C. Wang, J.Y. Zheng, S. Luo, X. Luo, Y. Huang, and D. Gao, "Route panorama acquisition and rendering for high-speed railway monitoring," *IEEE International Conference on Multimedia and Expo*, pp.1–6, 2013.
- [11] S.C. Wang, S.W. Luo, Y.P. Huang, J.Y. Zheng, P. Dai, and Q. Han, "Rendering Railway scenes in Cyberspace Based on Route Panoramas," *IEEE Conference on Cyberworlds*, pp.12–19, 2013.
- [12] X.C. Tong, "Characterization Methodologies of Thermal Management Materials," *Advanced Materials for Thermal Management of Electronic Packaging*, Springer New York, pp.59–129, 2011.
- [13] J. Nielsen, E. Berggren et al, "Overview of Methods for Measurement of Track Irregularities Important for Ground-Borne Vibration," *RIVAS Deliverable 2.5*, July 2013.

Peng Dai received the BS, MS and PhD degrees from the Department of Control Science and Engineering of the Harbin Institute of Technology. He worked as an associate researcher at Infrastructure Inspector Center of China Academy of Railway Science. His research interests include statistical machine learning, visual detection and its application to Highspeed railway.

Shengchun Wang was born in 1985. He received his B.S and PHD degree from the School of Computer and Information Technology, Beijing Jiaotong University in 2008 and 2015, respectively. He worked as an assistant researcher at Infrastructure Inspector Center of China Academy of Railway Science. His research interests include scene representation for railway environment, high resolution reconstruction and object detection.

Yaping Huang was born in 1974. She received her B.S., M.S. and Ph.D. degree from Beijing Jiaotong University in 1995, 1998 and 2004, respectively. Since 2012, she has been a professor in the institute of computer and information technology at Beijing Jiaotong University. Her research interests include computer vision, pattern recognition, and machine learning.

Hao Wang was born in 1984. He received his B.S and MS degree from school of Electronic Information Engineering of Beijing University of Aeronautics and Astronautics. He worked as an associate researcher at Infrastructure Inspector Center of China Academy of Railway Science. His research interests include structured light measurement, rail geometry measurement and optical engineering.

Xin-Yu Du received the Ph.D. degree in biomedical engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2011. He is currently a Researcher with the Infrastructure Inspection Institute, China Academy of Railway Sciences, Beijing, China. His research interests focus on the image processing.

Qiang Han received the BS and MS degrees from the School of Science of the Beijing Jiaotong University. He worked as an assistant research at Infrastructure Inspector Center of China Academy of Railway Science. He works in the field of laser and photoelectric detection.