Creating Cylindrical Panoramic Mosaic from a Pipeline Video

Chuan Niu[1)] Fan Zhong[1)] Songhua Xu[2)] Chenglei Yang[1)*] Xueying Qin[1)*]
[1)]School of Computer Science and Technology, Shandong University, Jinan, P.R. China
Shandong Provincial Key Laboratory of Software Engineering, Jinan, P.R. China
[2)] Oak Ridge National Laboratory, Oak Ridge, TN 37831, US

Abstract

In geological engineering, stratum structure detection is a fundamental problem in project planning and implementation. One of the most commonly employed detection technologies is to take videos of borehole using a forward moving camera. Following this approach, the problem of stratum structure detection is transformed into the problem of constructing a panoramic image from the taken video sequences, which are typically of low quality. In this paper, we propose a novel method to create a panoramic image of the borehole from the video sequence without camera calibration and tracking. To stitch together pixels of neighboring frame images, our camera model is designed with a focal length changing feature, along with a small rotation freedom in the two-dimensional image space. Essentially, our camera model assumes that target objects lie on a cylindrical wall and the camera moves forward along the central axis of the cylindrical wall. Our method robustly resolves these two degrees-of-freedoms through KLT feature tracking and constructs a panoramic image by stitching strips. Experiment results show that our method could efficiently generate high-quality panoramas for very long video sequences.

1. Introduction

Panoramic mosaicing aims at creating an image of a wide view by aligning and pasting photos or frame images of a video sequence. It is extensively used in many fields such as virtual reality, medical image analysis, geological engineering.

For projects performed in rock and soil masses, their design and implementation must be solidly grounded on engineering geology and hydrogeology conditions, where the problem of stratum structure detection has always been a fundamental issue. By putting a forward camera in a borehole, engineers can obtain videos recording geological structures. Our work aims at creating panoramas based on such videos. The constructed images can be used in detecting faults, cracks, water exits, and coal seams.

One key step in image mosaicing is image alignment, which is critical for the visual quality in the end result. Due to efficiency considerations, feature-based alignment methods are usually employed, which first extract feature points [6], [14] and then match or track these features across frames to obtain transformations between frames, e.g. through SIFT [9], [2] or KLT [10], [26].

Based on different image stitching technologies, existing methods for panorama generation can be mainly classified into three classes. The first class of methods captures images using a camera with pure rotational movement. Such methods first align frames to a pre-selected reference 2D surface and then combine the aligned frames to create a seamless panoramic mosaic. The simplest case is to pan the camera around its optical center, after which a 360 degree panoramic mosaic can be constructed on a cylindrical, squared, or spherical manifold [11], [3], [12], [23]. For more general camera motions, whose transformation among images consists of global affine transformations [5], [8], [24], [25], or planar-projective transformations [18], [7], [21], [22], the aforementioned stitching methods can also be used. The limitations however are, when there is a substantial amount of rotation between images, distortion occurs in the resultant image. Furthermore, such methods do not work for zooming and forward camera motions.

The second class of methods constructs panoramas from sets of images taken from different viewpoints [1], which uses graph cuts to choose a viewpoint for each pixel in the output panorama for minimizing multi-perspective distortions due to motion parallax. This work aims at producing panoramas of long, roughly planar scenes, where camera motions are quite different from our problems.

The last class of methods constructs a panoramic image from a video sequence taken by a camera with motion. Such panorama is called slit-scan panorama or strip panorama, whose generation requires extracting thin strips of pixels from frames of a video sequence as well as aligning them together. Many variations of the methods have been proposed, e.g., the *Pushbroom Camera method* by Gupta and Harltley [15], a general mosaicing method for cameras with free motions [17], [16], [13], and *x-slit* images introduced by Zomet *et al* [19].

For mosaicing, dealing with motion parallax is very problematic, especially for the case when a camera goes through forward motions. B. Rousso *et al* [16] introduced a method by mosaicing curved strips that are perpendicular to optical flows. Their method can deal with cameras of free motions. They further developed a universal mosaicing method using pipe projection to deal with the forward motion of a camera [17]. This latter method assumes that the internal parameters of the camera are known in advance. However, this prerequisite is often hard to satisfy in geological engineering. S. Peleg *et al* [13] further proposed a general method for projecting thin strips onto various manifolds according to the motions of a camera. Their method restricts the distortion of panorama only to motion parallax. In general, strip panorama is suitable for most camera motion models. However, this method is sensitive to registration errors and changing illumination because it employs affine transformations with six parameters

*. Corresponding authors. Emails: {qxy, chl_yang}@sdu.edu.cn

to assign consecutive frame images. In particular, for scenes with large depth variations, strip panoramas will exhibit severe distortions.

In this paper, we propose a novel method to mosaic a low quality video sequence of borehole taken by a forward camera. We simplify the camera motion to a 2D frame image rotation and scaling with only two parameters for improving the robustness and preciseness of image alignment. We then generate panoramic images by selecting strips from images. Our method can create high-quality pipeline panoramas and is adaptable to alignment errors.

The structure of this paper is organized as follows. In section 2, we describe our stitching model and stitching method. section 3 presents our experimental results to validate our algorithm. Then, we conclude our work in section 4.

2. Image Stitching Model

In geological exploration, to get information about engineering geology and hydrogeology conditions, engineers put cameras in the borehole to capture videos. A camera is generally fixed on top of a sliding rail. Camera motions include forward and small rotations around the optical axis (see Fig. 1). Using such videos as input, our purpose is to create pipeline panoramas that show the texture of the borehole.

2.1. A Simplified Camera Model

The quality of a panorama generation result is directly related to the quality of image alignment. However, the more free parameters to be estimated, the more errors will likely occur. Peleg *et al* [13] used an affine model with six-parameters to align images taken by a camera of forward motions. Since a borehole is approximated as a cylinder, and the camera is approximately moved along its central axis as shown in Fig. 1, we further simplify their camera model in our study, which will be explained shortly.

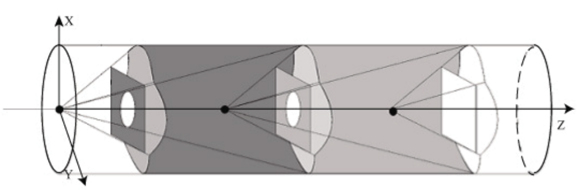


Fig. 1. Camera model: camera moves forward in the borehole.

We assume that the input videos to our problem are captured by a pinhole camera, as is most commonly used in the field practice. According to prior studies [20], [4], if a fisheye lens is used, the distortion can be corrected. In projective geometry, relationships between image point m and a world point M can be expressed by:

$$m \sim \mathbf{K}[\mathbf{R} \quad T]M, \tag{1}$$

where $\mathbf{K}$ is the intrinsic matrix. The extrinsic parameters involved are composed of a rotation matrix $\mathbf{R}$ and a translation vector T. For a forward motion camera model, assuming a world point $M(X, Y, Z)$ has corresponding points $m(x, y)$ and $m'(x', y')$ in two successive frames. Without loss of generality, the camera coordinate system of the first frame is regarded as world coordinate system, and the center of image is considered as the origin point. We therefore have the following relationships:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \sim \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{2}$$

where k represents focal length. From (2), we have:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{k}{Z} \begin{bmatrix} X \\ Y \end{bmatrix} \tag{3}$$

For successive frame images, as mentioned in the above, the translation of a camera is assumed to be along the optical axis direction while the camera rotation is around optical axis. Thus, we assume the translation for m' is $(0, 0, t)$ and the rotation of the camera is β. We then have:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} \sim \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\beta & \sin\beta & 0 & 0 \\ -\sin\beta & \cos\beta & 0 & 0 \\ 0 & 0 & 1 & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4}$$

We can derive the following from (4) by introducing a coefficient k'':

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix} &= \tfrac{k}{(Z+t)} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \\ &= \tfrac{k''}{k} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \cdot \tfrac{k}{Z} \begin{bmatrix} X \\ Y \end{bmatrix} \\ &\quad + (\tfrac{Z}{Z+t} - \tfrac{k''}{k}) \tfrac{k}{Z} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} \\ &= \tfrac{k''}{k} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &\quad + (\tfrac{Z}{Z+t} - \tfrac{k''}{k}) \begin{bmatrix} x'' \\ y'' \end{bmatrix} \end{aligned} \tag{5}$$

where $(x'', y'')^T$ is the corresponding images with only rotational transformation applied onto the original image. Since usually $t \ll Z$, there exists a coefficient k''/k such that:

$$\frac{Z}{Z+t} - \frac{k''}{k} \approx 0. \tag{6}$$

The above equation indicates that the camera model can be approximated by a rotational movement plus a scaling of the first frame image, where k''/k is a fine-tuning parameter for the two frames, and β represents the rotation angle of the frame images. Therefore, the camera model can have two parameters in total, including the coefficient k''/k and the rotation angle β, i.e.:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} \approx \tfrac{k''}{k} \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \tag{7}$$

With these two-free parameters built into our model, the image alignment process can be more robustly carried out, which will be supported by experiments. The accuracy requirement for feature tracking is relatively low and the computational overhead is also much less than that of the models carrying more parameters. In our current implementation, we employ KLT feature tracking [10], [26] to estimate k''/k and β.

2.2. Stitching

Our panoramic space is assumed as a cylinder space sharing the same central axis of the borehole. Our problem is to construct a panoramic image by selecting strips in frame images of the video sequence.

For our input video, the quality is too poor to search for focal length, so it is impossible to precisely establish the relationship between panoramic images and a real scene. With our simplified model, all 3D points with the same depth are on a circle. We select a *standard circle* with radius r_0 in all the frames, and these circles should be a vertical line on the panoramic space. We define the vertical resolution of the panoramic image to be $2\pi r_0$. To get better resolution, r_0 can be very large, as long as it and its projection in next frame do not exceed the boundary of the image. The resolution of video sequence we use is 320×240 pixels, so we set r_0 to be 110 pixels in the experiments. The horizontal resolution is approximated by the one in radial direction around the *standard circle*.

Then we take advantage of the method proposed in [16]. For our camera model, optical flow is radial from the image center after canceling rotation. Assuming a video sequence has N frames, and the panorama is stitched by $N-1$ strips from left to right. Strip i is represented by I_i. The *standard circle* in frame i is represented by C_i.

For C_i, its corresponding projection C_i' in frame $i+1$ can be obtained using (7). The circular broom S_i bounded between C_i' and C_{i+1} is the corresponding projection for strip I_i in frame $i+1$. S_i is warped to become I_i. After warping and stitching, the original optical flow becomes horizontal in the resulting image. The formula (7) has only two unknown numbers, thus the accuracy requirement for feature tracking is low and the computational overhead is very small.

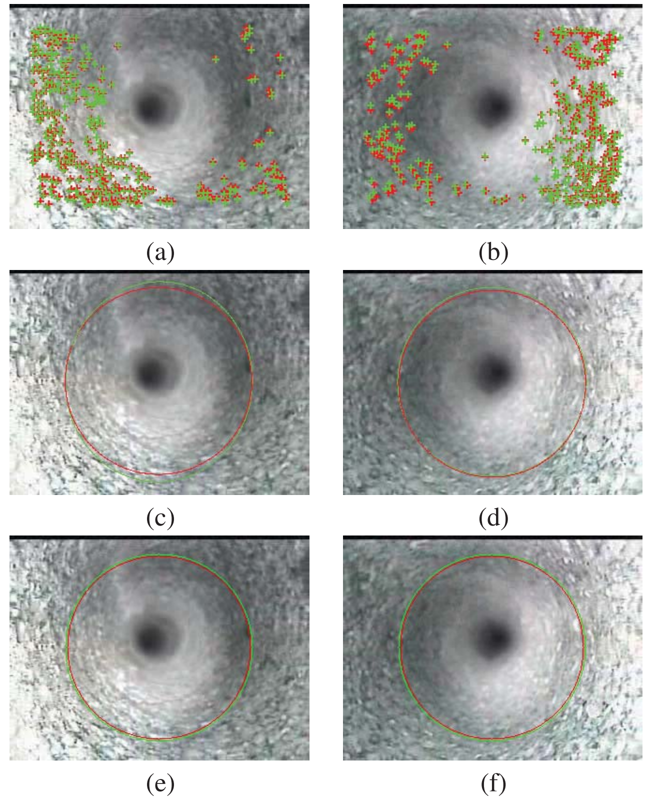
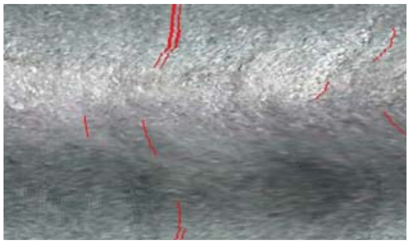


Fig. 2. The ellipses show the corresponding edges of a strip in panoramic space, and the pixels between green curve and red curve in the direction from image center to margin are comparisons between image alignment quality by affine transformation [13] and our proposed method: (a) and (b) show tracked feature points, in which red and green color respectively indicates feature point positions in the last and current frames; (c) and (d) are calculated by affine transformations; (e) and (f) are calculated by our proposed method.

3. Experiment Results

We explore the effectiveness of our algorithm by working with one sample video sequence, taken by a camera mounted on top of a drill, whose focal length is unknown. The video was shot in the midst of a mine construction project operated by a coal mining enterprise. The resolution of video sequence is 320×240 pixels and has 7900 frames in total. The computing platform used in our experiments bears the following configurations: Intel core 2 Duo E6550 @ 2.33GHz CPU; 2GB memory; the Windows XP operating system; ATI Radeon HD graphics card by 2600 Pro (RV630) (256 MB).

We first examine the quality of image alignment. As shown in Figs. 2(a) and (b), feature points are properly tracked, providing quantitative support for calculating transformations between images for matching purpose. The model of affine transformation [13] has 6 parameters, which is very unstable for low quality video sequences. Within this model, the two edges on the panoramic image of a strip are mapped onto two ellipses as shown in Figs. 2(c) and (d). Pixels between the green curves and red curves in the radial direction are warped to horizontal lines in the panoramic space. From these two images we can easily notice that the two ellipses derived for a strip may intersect with each other, leading to strange shapes in the resultant panoramic images. By using the model proposed in this paper, which only carries two parameters, the two ellipses derived will no longer overlap (Figs. 2(e) and (f)). Comparing Fig. 3 (a) and (b), we can easily notice that the affine transformation based method as shown in Fig. 3(a) failed due to the instability of the alignment step, where red colored regions indicate strip areas with crossing situation, and we can notice the distortion, especially in the green box. In contrast, as shown in Fig. 3(b), our method can generate stable result. Fig. 4 demonstrates the same strip sub-images reconstructed from five successive frame images using our new alignment method, which obtain a very decent quality.



(a)

(b)

Fig. 3. Comparison of our method and previous work. (a) is obtained by strip panorama [13], in which the red points are those pixels due to alignment error; (b) is obtained by proposed method.

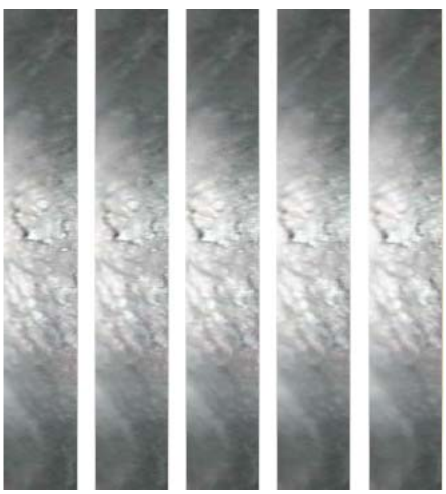


Fig. 4. Sub-images warped from 5 successive images onto the same strip in panorama space.

Furthermore, with our proposed method, we construct panoramic images from this practice video sequence. From Fig. 5, we can recognize the high quality of the panoramic images. In our video, the light source is not stable and the interior surface of the pipe is not Lambertian, but our method still could get nice result. However, some regions are still too bright since all of the pixels at this side are too bright, which cannot erase this effect completely although it can be decreased.

4. Conclusion

In this paper, we proposed a method to create high quality panoramic images from borehole video sequence with very low quality. Our method is highlighted by its image alignment model, involving only two parameters. It thus performs more robustly in image framework alignment without relying on any prior knowledge on camera focal length or distortion parameters.

In the future, many aspects of our work can be improved. Although our method is designed for geological engineering, it is also applicable in other fields, such as creating panorama from endoscope images. We hope to experiment with such data sets in the near future. Second, when camera trajectory is not along the central axis of a cylinder, distortion arises in the resulting panorama. If we can detect the difference between camera trajectory and cylinder center axis, such distortion could be canceled or avoided. In addition, when the camera experiences considerable rotation in its motion that is oriented vertically to the optical axis, our current method is likely to fail. We would like to explore and overcome these issues in our future research.

Acknowledgement

This work is supported by national natural science foundation of China (No. U1035004, No. 61003149, No.60870003), Shandong province outstanding young scientist research award fund (BS2009DX026), and Natural Science Fund for Distinguished Young Scholars of Shandong Province (No. JQ200920). Songhua Xu performed this research as a Eugene P. Wigner Fellow and staff member at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract DE-AC05-00OR22725.

References

[1] A. Agarwala, M. Agrawala, M. Cohen, D. Salesin, and R. Szeliski. Photographing long scenes with multi-viewpoint panoramas. In *ACM Trans. on Graph*, pages 853–861, 2006.
[2] M. Brown and G. Lowe. Automatic panoramic image stitching using invariant features. In *International Journal of Computer Vision*, volume 74, pages 59–73, 2007.
[3] S. E. Chen. Quicktime vr: An image-based approach to virtual environment navigation. In *Proc.ACM SIGGGRAPH 95*, pages 29–38, 1995.
[4] F. Devernay and O. Faugeras. Straight lines have to be straight. In *Machine Vision and Applications*, pages 14–24, 2001.
[5] M. Hansen, P. Anandan, K. Dana, G. van der Wal, and P. Burt. Real-time scene stabilization and mosaic construction. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 54–62, 1994.
[6] C. Harris and M. J. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–152, 1988.
[7] M. Irani, P. Anandan, and S. Hsu. Mosaic based representations of video sequences and their applications. In *Proc. Fifth Int'l Conf Computer Vision*, pages 605–611, 1995.
[8] P. Jaillon and A. Montanvert. Image mosaicking applied to three-dimensional surfaces. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, volume 1, pages 253–257, 1994.
[9] G. Lowe. Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 60, pages 91–110, 2004.
[10] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.
[11] S. Mann and R. Picard. Virtual bellows: Constructing high quality stills from video. In *Proc.IEEE International Conference on Image Processing*, volume 1, pages 363–367, 1994.
[12] L. McMIllan and G. Bishop. Plenoptic modeling: An image-besed rendering system. In *Proc, ACM SIGGGRAPH 95*, pages 39–46, 1995.

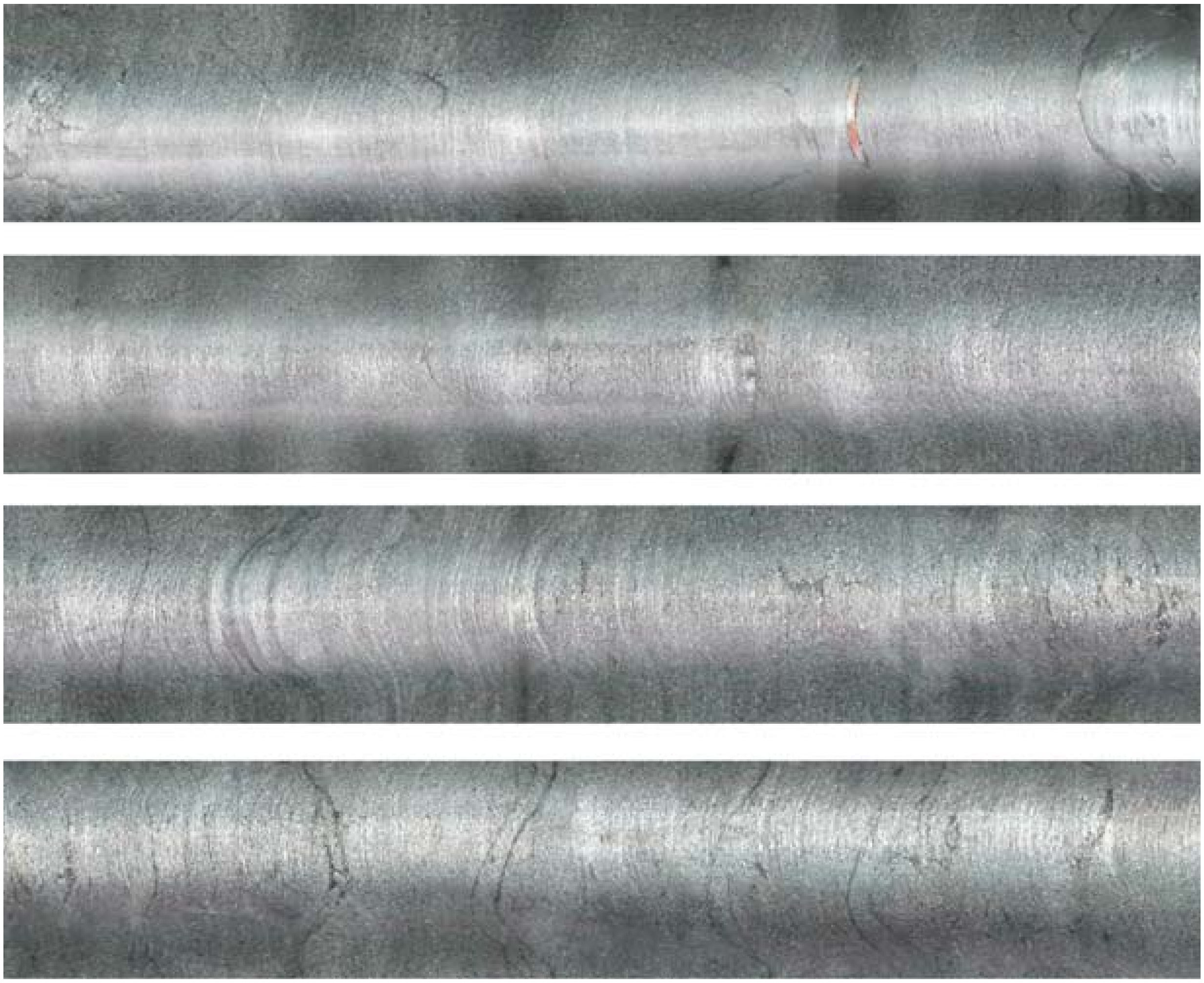


Fig. 5. The complete panoramic image with 15033×691 pixels from a video sequence with 7900 frames, which is divided into 4 parts and aligned from top to bottom due to the space limitation.

[13] S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet. Mosaicing on adaptive manifolds. In *PAMI*, pages 1144–1154, 2000.

[14] E. Rafajlowicz. Susan edge detector reinterpreted, simplified and modified. In *International Workshop on Multidimensional (nD) Systems*, pages 9–74, 27–29, 2007.

[15] I. Richard, Hartley, and R. Gupta. Linear pushbroom cameras. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 963 – 975, 1997.

[16] B. Rousso, S. Peleg, and I. Finci. Mosaicing with generalized strips. In *DARPA Image Understanding Workshop*, pages 255–260, 1997.

[17] B. Rousso, S. Peleg, I. Finci, and A. Rav-Acha. Universal mosaicing using pipe projection. In *ICCV*, pages 945–952, 1998.

[18] H. S. Sawhney, S. Ayer, and M. Gorkani. Model-based 2d and 3d dominant motion estimation for mosaicing and video representation. In *Proc.Fifth Int'l Conf. Computer Vision*, pages 583–590, 1995.

[19] ssaf Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: The crossed-slits projection. In *PAMI*, volume 25, pages 741–754, 2003.

[20] R. Swaminathan and S. Nayar. Non-metric calibration of wide-angle lenses and polycameras. In *CVPR*, volume 2, page 2413, 1999.

[21] R. Szeliski. Image mosaicing for tele-reality applications. In *Technical Report 94/2, Digital Equipment Corporation*, 1994.

[22] R. Szeliski. Image mosaicing for tele-reality applications. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 44–53, 1994.

[23] R. Szeliski. Video mosaics for virtual environments. In *IEEE Computer Graphics and Applications*, volume 16, pages 22–30, 1996.

[24] L. Teodosio and W. Bender. Panoramic overview for navigating real-world scenes. In *ACM Multimedia 93*, pages 359–364, 1993.

[25] L. Teodosio and W. Bender. Salient video stills: Content and context preserved. In *ACM Multimedia 93*, page 3946, 1993.

[26] C. Tomasi and T. Kanade. Detection and tracking of point features. In *Carnegie Mellon University Technical Report CMU-CS-91-132*, 1991.