

# Employing a Fish-Eye for Scene Tunnel Scanning

Jiang Yu Zheng<sup>1</sup>, Shigang Li<sup>2</sup>

<sup>1</sup> Dept. of Computer Science, Indiana University Purdue University Indianapolis,  
723 W. Michigan St. Indianapolis, IN46202, USA

[jzheng@cs.iupui.edu](mailto:jzheng@cs.iupui.edu)

<http://www.cs.iupui.edu/~jzheng>

<sup>2</sup> Dept. Computer and Information Science, Iwate University,  
4-3-5 Ueda, Morioka, Iwate 020-8551, Japan

[li@cis.iwate-u.ac.jp](mailto:li@cis.iwate-u.ac.jp)

<http://cv.lk.cis.iwate-u.ac.jp>

**Abstract.** This work employs a fish-eye to scan cityscapes along a street and register scenes in a compact scene tunnel image. A fish-eye has complete field of view along a route. We mount the fish-eye camera on a vehicle and estimate its pose initially with respect to the vehicle by referring to 3D structure lines of such as roads and buildings on a street. Sampling curves are then allocated in the image frame for dynamic scanning route scenes as the vehicle moves forward. The accurate alignment of the curves ensures less distortion of shapes in the scene tunnel image. We also analyze the scanned results and evaluate alignments of the sampling curves to improve the scanning. The resulting scene tunnel is a continuous archive of the entire route in a city, which can be used for environment visualization and assessment, Internet based virtual navigation, city information indexing, etc.

## 1 Introduction

To register cityscapes for a visual map, the *route panorama* has been proposed to archive street scenes [1][2], which is a continuous image different from local panoramic views at static positions. A video camera moves along a smooth path and a long image of the path is generated with the slit scanning mechanism [3][4][5][6][10][17]. The slit scanning differs from most of the image patch stitching [12] or video mosaicing, since only a pixel line is collected at an instance when the camera undertakes translation. The slit scanning requires no inter-frame matching and morphing in creating the image so that it is suitable for transiting cameras viewing scenes even with complex occlusions, and real time route scene archiving. In order to capture both sides of a street, multiple cameras have been stacked to scan a *scene tunnel* that contains complete heights, three distinct aspects, and a long image length [6]. The compact data size and complete coverage of scenes benefit scene visualization, texture mapping on urban models, image transmission for navigation, virtual tour, etc.

In this paper, we explore the use of a fish-eye camera to achieve the scene tunnel acquisition. A fish-eye camera can capture half space scenes in an image, and thus is efficient to scan entire scene around the camera trajectory. It avoids many issues so

far such as calibrating and synchronizing multiple cameras, integrating scenes at different heights and sides, and normalizing photometric properties in the connected images for generating the scene tunnel. In this paper, we explore the following issues.

- (1) How to mount a fish-eye camera properly on a vehicle for route scenes?
- (2) How to align sampling pixels (lines/curves) in the image frame to scan a scene tunnel so that the shapes of typical structures on the street can be preserved?
- (3) How to calibrate the camera external parameter with respect to the street structure for localizing the sampling curves, if the camera is not set ideally on the vehicle?
- (4) How the shapes are distorted due to an imprecise setting of a sampling curve, and how it can be improved?

Many works on the fish-eye camera calibration have been reported for 360 degree local panoramic view acquisition [7][8][9][13][14][15]. We will use these results for the calculation of the camera pose in an outdoor environment, in order to locate sampling curves and implement scene tunnel scanning for long distances.

In the following, we introduce the plane of scanning to acquire a scene tunnel, and the setting of a fish-eye camera in section 2. The selection of the sampling curves and their initial calibration in the image frame are discussed in section 3. The scanned results are analyzed and the refinement is given in section 4.

## 2. Acquiring Scene Tunnel along a Camera Path

There is an expectation to project long route scenes to a single image, which can be simply retrieved with maps in GIS and many other applications. To capture complete views, a spherical retina is located on a vehicle  $O-X'Y'Z'$  moving along a smooth path on a horizontal roadway. The ideal vehicle motion has a translation  $V$  in the heading direction and a rotation  $R_y$  around the vertical axis, realized by a four-wheeled vehicle with a good suspension system.

### 2.1 Scene Tunnel Scanning under Ideal Camera Setting

To capture non-redundant scenes for a compact representation, a *plane of scanning* ( $PoS$ ) is set from the retina center to scan scenes along its trajectory [5]. As the vehicle moves forward, the temporal data on the  $PoS$  are projected towards the retina center and are imaged at a pixel curve  $C(\theta)$  that is the intersection of the  $PoS$  and the retina surface. Here  $\theta$  is azimuth angle of a line of sight in the  $PoS$  (correspond to latitude on the sphere). If we copy the temporal data from the sampling curve  $C(\theta)$  and list them along the time axis, a scene tunnel  $T(t, \theta)$  is obtained. The reason to select a 3D plane for scanning is that many structure lines such as building rims and road edges exist in urban environments. A plane can scan them instantaneously to obtain straight shapes in the resulting scene tunnel.

A  $PoS$  is desired to be set vertically in the 3D space when the vehicle is moving on a horizontal road; all architectures projected in the *scene tunnel* thus keep vertical for visualization and texture mapping [1][2]. The *scene tunnel* is the generalization of the *route panorama* that extends a pixel line to a sampling ring for the complete heights

and sides [6]. To achieve this goal, multiple cameras are directed to different heights, which increases the system complexity.

Denote the angle between the vertical *PoS* and heading direction  $V$  by  $\alpha$ . If a *PoS* is non-orthogonal to  $V$  ( $\alpha \neq \pi/2$ ), the scanned scene tunnel includes not only building fronts but also partial side façades on the street. If three *PoS* ( $\alpha < \pi/2$ ,  $\alpha = \pi/2$ ,  $\alpha > \pi/2$ ) are set for fore-side, side, and rear-side scenes along a street, scene tunnels can cover all the visible route scenes.

Because the scene tunnel image is composed of consecutive 1D views, it has special object shape distortion under the parallel-central projection [1][2]. We focus on three types of lines in the analysis. In a street space, there are vertical lines, horizontal lines orthogonal to the roadway, and lines along the roadway on the street structures (Fig. 1). These lines, denoted as line sets  $L$ ,  $M$ , and  $N$ , are orthogonal to each other, and the remaining lines can be expressed as the linear combinations of them. The projections of  $L$ ,  $M$ , and  $N$  sets in the image are denoted by  $l$ ,  $m$ , and  $n$ , respectively. Considering the camera motion along  $N$ , *PoS* are set parallel to  $L$  and  $M$  for scanning architectures and ground.

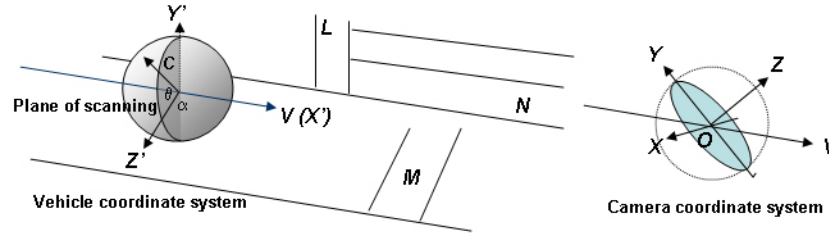


Fig. 1 Scanning a scene tunnel by a spherical retina along a path with typical structure lines.

## 2.2 Scene Scanning with a Fish-eye Camera

A fish-eye camera  $O$ - $XYZ$  maps half-space scenes onto a circular image  $I(x,y)$ . According to the equidistance projection model of the fish-eye lens [8], point  $P(X,Y,Z)$  in the space is projected to  $p(x,y)$  in the image and is represented in a polar coordinates system  $p(\varphi, \gamma)$  as

$$\rho = (x^2 + y^2)^{1/2} \quad \tan \varphi = y/x \quad \gamma = \rho/f \quad (1)$$

where  $\gamma$  is the angle between the line of sight  $OP$  and the optical axis  $OZ$ , and  $f$  is the camera focal length. The line of sight can also be represented by a vector in the  $O$ - $XYZ$  as

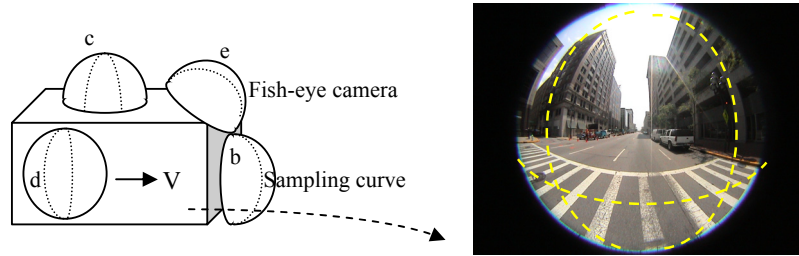
$$(X,Y,Z) = v(\varphi, \gamma) = (\sin \gamma \cos \varphi, \sin \gamma \sin \varphi, \cos \gamma) \quad (2)$$

We use a fish-eye lens (Olympus FCON-w2) in front of a video camera to capture image  $I'(x',y')$ . By calibrating the internal parameters of the lens [15] such as optical center  $o(x_0, y_0)$ , focal length  $f$ , and radial distortion, the coordinates of a point in  $I'(x',y')$  can be converted to that in  $I(x,y)$  that obeys the equidistance model.

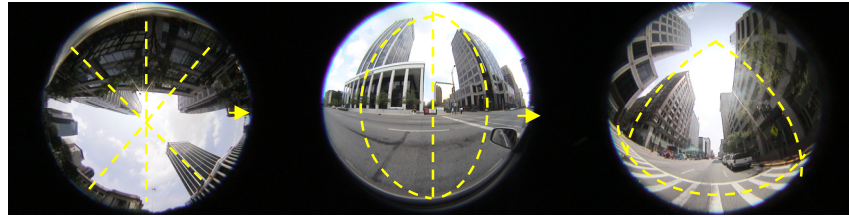
There are several ways to set the fish-eye camera for street scanning as depicted in Fig. 2. The camera can be mounted at front, side, top of the vehicle to face forward, side, up, or fore-up directions, respectively. The sampling curves allocated must scan optical flows in the fish-eye image, in order to generate the scene tunnel image.

- (1) The forward setting obtains both building fronts and the ground. Two curves aligned with  $l$  and another curve on  $m$  can be set for scanning. The drawback is incapable of taking rear-side views. Moreover, angle  $\alpha$  cannot be close to  $\pi/2$ . Otherwise, the image perimeter is sampled and the image quality is low. The same defect appears at the top. A large FOV is assigned to the ground that is not considered as important as the side scenes.
- (2) The camera set at the vehicle top can scan fore-side, side and rear-side scene tunnels with three sampling curves. Scenes lower than the camera position are not included in the scene tunnel. It wastes a large FOV on sky area if most scenes along a street are low.
- (3) Placing a fish-eye camera sideways covers entire side scenes and the camera is easy to mount. Three curves can be set for fore-side, side, and rear-side scene tunnels. However, two fish-eye cameras are needed for both-side scenes.
- (4) Directing the camera forward-up to include full building heights and partial road surface. It is an ideal setting we employ for our street scanning.

For all the camera settings, the angular resolutions assigned to a high-rise are the same; directing the camera upward does not increase the building size in the image. The sampling curves are set according to the distinct  $PoS$  to scan  $L$  and  $M$  lines in  $O-X'Y'Z'$ . As the vehicle traverses a street, a video is taken and the 1D data on the sampling curves is copied to the scene tunnel at a fixed rate (30~60Hz).



(a) Locating fish-eye cameras on a vehicle. (b) Forward: curves scanning  $l_{left}$ ,  $l_{right}$ ,  $m$



(c) Upward:  $l_{rear}$ ,  $l_{side}$ ,  $l_{fore}$  (d) Sideways:  $l_{rear}$ ,  $l_{side}$ ,  $l_{fore}$  (e) Forward-up:  $l_{left}$ ,  $l_{right}$ ,  $m$ .

**Fig.2** Different settings of a fish-eye camera and their images with designed sampling curves.

### 3. Calibration of Sampling Curves for Scene Scanning

After the fish-eye camera is mounted, there is no guarantee that the camera axis is exactly aligned with the  $V$  direction. We have to calibrate the orientation of the camera system  $O-XYZ$  with respect to the vehicle system  $O-X'Y'Z'$ . This ensures a correct allocation of sampling curves for the designed  $PoS$ . We park the vehicle on a straight road with high buildings around. The road surface is measured to be horizontal and the mounted camera takes a sample image with many straight lines. Different from the fish-eye lens calibration focusing on the internal parameters, we use parallel structure lines in the real scene to locate their vanishing points, which characterize the camera pose. Then the sampling curves are computed accordingly for vertical  $PoS$ . Although the located  $PoS$  may change instantly in the route scanning afterwards due to vehicle bumping on uneven roads, the major road sections will yield reasonably good results.

#### 3.1 Calibrating Camera Pose Using Structure Lines in Sample Images

Let us examine the shapes of structure lines  $L$ ,  $M$ , and  $N$  in the fish-eye images (Fig. 3). In general, the projection of a line in the fish-eye image is a curve. A line in the space determines its *plane of sight*  $\Gamma$  with  $O$ . The plane  $\Gamma$  intersects the spherical retina at a *great circle*  $C$ . A 3D line segment  $P_1P_2$  determines the normal of  $\Gamma$  from  $OP_1 \times OP_2$ . Further, parallel lines in the space determine planes of sight  $\Gamma_j, j=1,2,3,\dots$ . These planes intersect each other at a vector  $\mathbf{Q}$  through  $O$ , which is parallel to the line set. Its projection to the spherical retina is the penetrating point at  $\mathbf{q}$  as Fig. 3(c) depicted. A  $\Gamma_j$  is projected to a great circle  $C_j$  on the retina. Since  $\mathbf{Q}$  is the intersection of all  $\Gamma_j$ ,  $\mathbf{q}$  is the crossing point of all  $C_j$ . Therefore,  $\mathbf{q}$  is the *vanishing point* of the line set. Detecting the position of the vanishing point in the image tells the camera orientation with respect to the line set. Fig. 3 shows a forward and a slanted setting of a fish-eye camera and their images. We can find structure lines  $L$ ,  $M$ , and  $N$  projected as curves  $l$ ,  $m$ , and  $n$  in Fig. 4. These lines can be characterized by three pairs of vanishing points  $\mathbf{q}_L$ ,  $\mathbf{q}_M$ , and  $\mathbf{q}_N$ , respectively. Some points are not in the image (on the other half of the spherical retina).

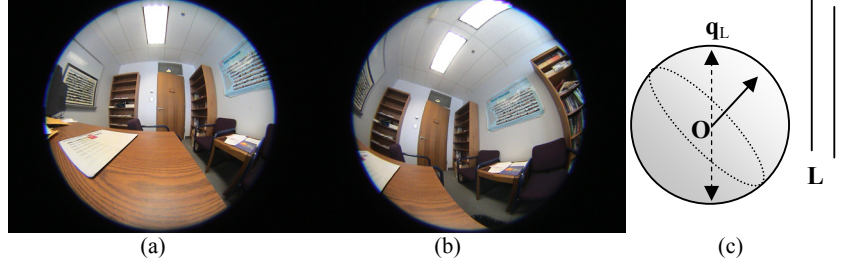
We first extract line segments on buildings in the sample image  $I'(x',y')$  by edge detector and tracking algorithm, and convert the coordinates to that in the non-distorted image  $I(x,y)$ . Denoting the  $i$ th point on  $j$ th curve by  $\mathbf{x}_{ij}=(x_{ij},y_{ij})$ ,  $i,j=1,2,3,\dots$  in  $I(x,y)$ , we calculate the line of sight through  $\mathbf{x}_{ij}$  is  $(X_{ij},Y_{ij},Z_{ij})$  according to (2). Manually selecting an extracted line  $l_j$ , all its points are filled into a total least squared error method for the normal  $\mathbf{n}_j=(a_j, b_j, c_j)$  of  $\Gamma_j$  in the  $O-XYZ$ , where  $a_j^2+b_j^2+c_j^2=1$ . It minimizes

$$\sum_i [(X_{ij}, Y_{ij}, Z_{ij}) \bullet (a_j, b_j, c_j)]^2 \rightarrow \min \quad (3)$$

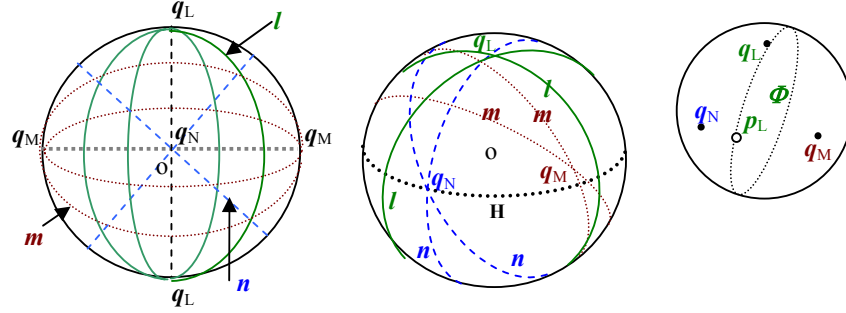
Denote the coordinates of vanishing point by  $\mathbf{Q}(X_L, Y_L, Z_L)$  in  $O-XYZ$ , it can be obtained through the second least squared error estimation from multiple normal  $\mathbf{n}_j$ , by

$$\sum_j [(X_L, Y_L, Z_L) \bullet (a_j, b_j, c_j)]^2 \rightarrow \min \quad (4)$$

By tracking edge points on  $n$  and  $m$  curves, we calculate the vanishing points  $q_N$  and  $q_M$  in the same way as for  $q_L$ .



**Fig. 3** Different fish-eye camera poses and their captured images. (a) An ideal forward setting of fish-eye camera. (b) An inclining camera setting by changing camera yaw, pitch and roll. (c) Vanishing point on spherical retina surface.



**Fig. 4** Projected structure lines and the corresponding vanishing points in the fish-eye image. (a) Ideal camera setting in Fig. 3(a). (b) Projected curves  $l$ ,  $m$ ,  $n$  converge to their vanishing points in Fig. 3(b). (c) Vanishing point and pin point for setting PoS.

### 3.2 Initial Setting of Sampling Curves in Fish-eye Image Frame

Because the fish-eye camera  $O-XYZ$  captures a half space, at least one quadrant of the  $O-X'Y'Z'$  system is covered. Among six vanishing points of three orthogonal line sets  $L$ ,  $M$ , and  $N$  in  $O-X'Y'Z'$ , at least one set  $q_L$ ,  $q_M$ , and  $q_N$  can be found within the fish-eye image. After detecting their positions from tracked curve sets  $l$ ,  $m$ , and  $n$ , we generate sampling curves for scene tunnel scanning.

For the calculated vanishing point  $q_L$ , we select a pin point  $p_L$  in the image (Fig. 4(c)), considering the orientation of  $PoS(\alpha)$ . The lines of sight  $v(q_L)$  and  $v(p_L)$  through the two points can form a  $PoS$  parallel to line set  $L$ , with the normal

$$n_L = v(q_L) \times v(p_L) \quad (5)$$

A line of sight  $\nu(\varphi, \gamma)$  on such a *PoS* must satisfy

$$\mathbf{n}_i \bullet \nu(\varphi, \gamma) = 0 \quad (6)$$

Given  $\varphi \in [0, 2\pi]$  with a fine interval, we can obtain corresponding  $\gamma$  for the sampling curve. It is further converted to  $(x, y)$  by using (2) and then to  $(x', y')$  in the fish-eye image frame. The distinct pixels on the curve are accumulated as a continuous  $C$  for sampling. Figure 5(a) displays a scene tunnel  $T(t, \theta)$  copied from  $C_{Lleft}$  and  $C_{Lright}$  curves aligned with  $l$  lines in the image of  $640 \times 480$  pixels.

For line set  $M$ , the process to locate a curve for scanning the ground is similar. The resulting ground part thus has all the crossing pedestrians appearing straight in the scene tunnel from a linear vehicle path. Lines  $N$  appear straight and parallel to the  $t$  axis in the scene tunnel, and are disturbed if the vehicle deviates from a straight path.

## 4. Result Analysis and Refinement

### 4.1 Scene Tunnel Affected by Feature Distance

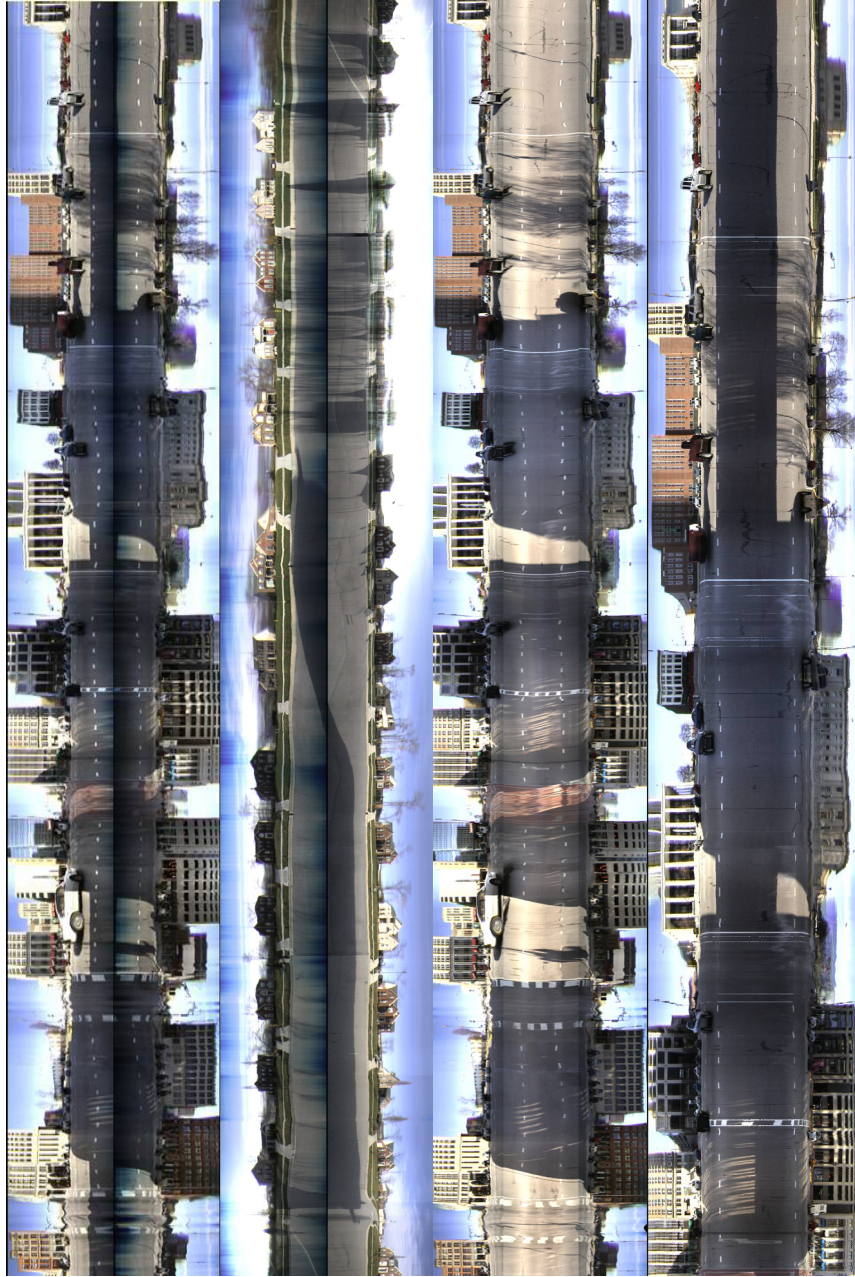
How important the alignment of sampling curves is and what happens if they are not precisely located? An immediate result one can imaging is that buildings are all slanted. However, some more serious defect is caused from an imperfect alignment. Fig. 6 shows a section of a scene tunnel where vertical lines at distance are more slanted than close ones. Because the *scene tunnel* at two sides contains scenes with different depth, a 2D skew transform is unable to rectify the image. This additional effect happens because the physical width of the sampling curve (one pixel) is not infinitely thin. The fish-eye video camera has a low-resolution image in over a half space. A pixel has a large angular coverage in the space. Through a pixel, the *Point Spread Function* (PSF) gets wide as the distance increases.

When the *PSF* moves across features at speed  $V$  (Fig. 7), a distant feature has longer response duration than a close feature because a wider *PSF* covers on it. This causes the distant edge blurred in the scene tunnel, which is named *stationary blur* in [11]. The degree of blur is more significant as the depth increases. Fig. 5(b) shows a scene tunnel from two side sampling curves. Stationary blur is obvious on the houses and trees because of their far distances from the roadsides. The close parts are sharp; it even has no motion blur on the vehicle shadow on the road surface. If the sampling line is non-parallel to the vertical edges as shown in Fig. 7, the responses at different heights are more sheared for a large *PSF* than a small *PSF*, even if the *PSF* regions are slanted in the same degree. This difference makes the edges incline differently.



**Fig. 6** Slanted vertical rims in the scene tunnel if the sampling line is not precisely aligned with the projection of a vertical line. (a) Declined edges due to imperfect alignment of the pixel line. (b) A precisely set sampling line generates a good shape of vertical lines.

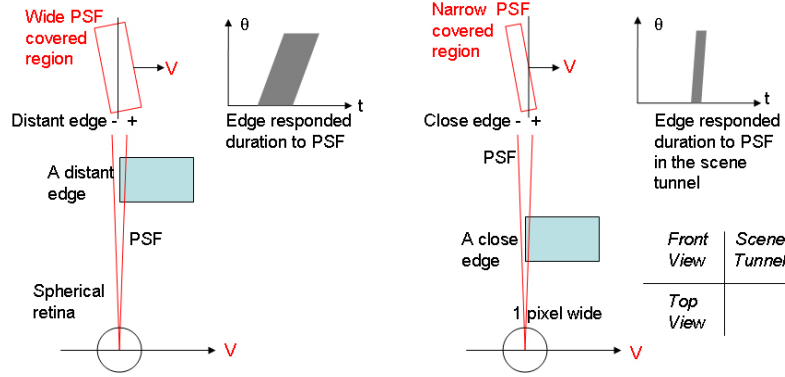




(a) Two sides only. (b) Stationary blurred. (c) Sampling curve inaccurate. (d) Refined tunnel

**Fig. 5** Fish-eye camera scanned scene tunnels  $T(\theta, t)$ ,  $\theta \in [0, 800]$  pixels along urban and suburban streets.





**Fig. 7** Different sizes of the Point Spread Function of a sampling line at different distances, and the scanned edges in the scene tunnel. Edges at different depths have different slanting angles.

Based on above analysis, close features are less influenced from slanted sampling curves than distance features. If a street has close building fronts, the error caused from the calibration of the sampling lines is not significant. For distance features, we can observe large, coarse edges slanted in the background. If a sampling line is located imperfectly, a common slanting direction and bending can be observed in the resulting scene tunnel (Fig. 5(c)).

#### 4.2 Scanning Refinement of Scene Tunnel

From a limited number of sample images in a long video sequence, it is difficult to obtain an accurate camera pose for scanning all the routes. Based on the scanned result of a scene tunnel, we can refine the position and shape of the sampling curves for the second scanning of the recorded video.

Because the setting of sampling curves has position error, we have to assess the calibration based on the scanned scene tunnel. On the time axis of the scene tunnel, we locate long periods without shaking, verified by consistent orientations (or slanting directions) of the projected building rims. From these periods, we pick up some time instances when a high building (or a high rim) is imaged. The video frames at such moments are loaded again for the refinement of the sampling curves. A curve segment is adjusted precisely on a projection of the long rim in the image. The recorded video is scanned again with the updated sampling curve. This refinement guarantees that most parts of a route are scanned in good shape (except the shaking road segments). If no structure line can be found in the entire scanned scene tunnel, no sample image will be retaken from the video sequence for further refinement. The deformed shapes in the remaining nature scenes are not critical to the visualization. Fig. 5(c) shows such an initial scanning with bended structure lines on the ground and slightly on buildings. Some shaking road segments are at the street crossings. Such defects are reduced after the refinement in Fig. 5(d).

## 5. Conclusion

This work explores the using of a fish-eye camera, localization and calibration of sampling curves for scanning scene tunnels from a moving vehicle. It avoids many system issues with multiple cameras in the scene scanning. We located multiple sampling curves on the projection of structure lines in order to generate the scene tunnel with preserved shapes. We calculated the camera pose from the structure lines, and scanned long scene tunnels successfully. We further analyzed the influences on shapes from an inaccurate setting of the sampling line. The future work will be the scene tunnel scanning with a high-definition TV camcorder, in order to obtain high-resolution scene tunnel with less stationary blurs. The scene tunnel is then a digest of street views for VR visualization, information indexing, and environment assessment.

## References

1. Zheng, J. Y., Route Panorama, *IEEE Multimedia*, 10(3), 57-68, 2003.
2. Zheng, J. Y., Shi, M., Mapping cityscapes into cyberspace for visualization, *J. Computer Animation and Virtual Worlds*, 16(2), 97-107, 2005.
3. Zheng, J. Y., Tsuji, S., Panoramic Representation for route recognition by a mobile robot, *IJCV*, 9(1), 55-76, 1992.
4. Gupta, R., Hartley, R., Linear pushbroom cameras, *IEEE PAMI*, 19(9), 963-975, 1997.
5. Zheng, J. Y., Tsuji, S., Generating dynamic projection images for scene representation and recognition, *Computer Vision and Image Understanding*, 72(3), 237-256, 1998.
6. Zheng, J. Y., Zhou, Y., Shi, M., Scene tunnels for seamless virtual tour, 12<sup>th</sup> ACM Multimedia, 448-451, 2004.
7. Xiong, Y., Turkowski, K., Creating image-based VR using a self-calibrating fisheye lens, *IEEE CVPR*, 237-241, 1997.
8. Kannala, J., Brandt, S., A General Camera Calibration Method for fish-eye lens, *ICPR2004*, Vol. 1, 692-695, 2004.
9. Li, S., Nakano, M., Chiba, N., Acquisition of spherical image by fish-eye conversion lens, *IEEE VR04* 235-236, 2004.
10. Peleg, S., Rousso, B., Rav-Acha, A., Zomet, A., Mosaicing on adaptive manifolds, *IEEE PAMI*, 22(10), 1144-1154, 2000.
11. Shi, M., Zheng, J. Y., A slit scanning depth of route panorama based on stationary blur, *IEEE CVPR05*, Vol. 1, 1047-1054, 2005.
12. Zhu, Z., Hanson, A., Riseman, E. M., Generalized parallel-perspective Stereo Mosaics from Airborne Video. *IEEE PAMI* 26(2): 226-237, 2004
13. Bakstein, H., Bajdla, T., Panoramic mosaicing with 180 field of view lens, *Omnidirectional vision workshop*, 60-68, 2002.
14. Swaminathan, R., Nayar, S. K., Nonmetric calibration of wide-angle lenses and polycameras, *IEEE PAMI* 22(10), 2000.
15. Li, S., Estimating head orientation based on sky-ground representation, *IEEE/RSJ Int. Conf. Intelligent Robots and Systems* 05.
16. Uyttendaele, M. *et al.* Image-based interactive exploration of real-world environments. *IEEE Computer Graphics and Applications*, 24(3), 2004.
17. Li, S., Hayashi, A., Robot navigation in outdoor environments by using GPS information and panoramic views, *IEEE/RSJ Conf. Intelligent Robots and Systems*, 570-575, 1998.