# Acquiring Shaking-free Route Panorama by Stationary Blurring

*Hongyuan Cai, Jiang Yu Zheng, Hiromi Tanaka\**
Indiana University Purdue University Indianapolis, USA
*Ritsumeikan University, Japan

## ABSTRACT

*This work obtains shaking-free route panoramas from a vehicle borne camera. We detect the shaking and waving profile of the camera and rectify a long route panorama automatically for an ideal parallel-perspective projection. Different from the traditional flow based approach that works on entire frame by point tracking, we create a stationary blurred temporal image called condensed slice during the scene scanning. This image yields long and visible traces from distinct and distant features and horizontal features as the evidences of camera motion, and it diminishes other features irrelevant to the motion as well. We track traces in the condensed slice and rectify the route panorama at the sub-pixel level without shape analysis in video frames. The significance of the approach is in its effectiveness visually confirmable in the results, global rectification as the vehicle moves forward, and efficiency for real-time processing using the minimum data set.*

***Index Terms**— route panorama, motion stabilization, stationary blur, intensity condensing, deshaking*

## 1. INTRODUCTION

To archive urban streets visually for navigation and georeferencing, the *Route Panorama* (RP) has been proposed [1,2,3]. A vehicle borne camera scans scenes with a virtual slit during the vehicle translation so that a 2D image belt can be generated in real time. The route panorama is efficient in real time view generation and compact in data size; it only cuts one pixel line out of a video frame. The route panorama has some shape distortion in aspect ratio due to the employed parallel-perspective projection model, which has been reduced in [3,4]. On the other hand, the dynamic composition of the long image suffers from local jitters and global waves of the vehicle borne camera due to the vehicle shaking and rolling on uneven roads in addition to the required vehicle translation. Such a jerky motion affects the quality of RPs. Video

stabilization [6,8] by matching or tracking 2D points, lines, or regions and finding optical flow has been explored. However, such methods have not produced perfect results for a vehicle moving in the city. The motion blurred features in the video reduces the accuracy of tracking (Fig. 1). Matching errors due to complex occlusions, repetitive patterns, glass transparency and reflection increase the uncertainty of estimation. Moreover, the number of extractable features may be insufficient in open areas.

This paper proposes a unique scheme to select long-lasting features (named *lighthouse feature*) as the reference of camera motion. These features, either at distance or stretching horizontally in the scene, are more reliable than points frequently emerging/disappearing in video frames. To obtain such features, we generate the *stationary blur* [2, 7,13] in a temporal image with the acquisition of a route panorama. Different from the motion blur in dynamic images, the stationary blur appears on slow moving scenes.

In this paper, we accumulate intensities in video frames in the motion directions to generate 1D intensity profiles called *condensed slice* (*CS*) [12]. In the *CS*, both distant and horizontal features appear as long traces that directly reflect the vertical jitters of video frames. The lines irrelevant to the camera shaking are ignored in it. This allows us to deshake the route panorama. We achieve an effective tracking on *CS* and rectify the RP at the sub-pixel level. The results show the advantages of this approach over the shape from motion and the efficiency in using the minimum data set for real time RP deshaking.

In the following, we review the vehicle/camera motion and then the stationary blur in Section 2. Section 3 describes our work in creating large stationary blur in CS and trace tracking. Section 4 introduces the algorithm to detect camera jitters and rectify *RPs*. Experimental results are given in Sec. 5.

## 2. STATIONARY BLURRING

The parallel projection of RP is realized by scanning scenes sideways along a path with a vertical *Plane of Sight*



Figure 2: A section of route panorama, RP(*t,y*), in which a piece of cloud at infinity is projected as a trace (indicated by dotted curve) over two blocks. It reflects the waving of camera as a "lighthouse" feature.

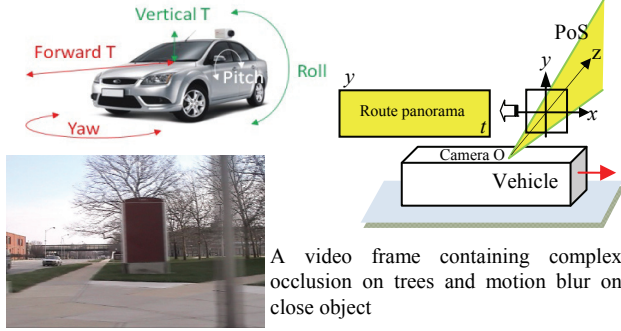A video frame containing complex occlusion on trees and motion blur on close object

Figure 1: Vehicle yaw-roll-pitch corresponds to camera pan-tilt-roll in creating RP. The scanning line is at the image center.

[2] in a fixed angle from vehicle translation $V$. The image plane intersects with the scanning plane at a pixel line, $l$ (e.g., scanning line at the y axis in Fig. 1), on which the pixels are sampled continuously and their values are stored in a long image, $RP(t,y)$. After a camera is mounted on a vehicle, its relation to $V$ is fixed. We compute the Flow of Expansion (FOE) once from the motion vectors caused by vehicle translation; two sampling frames are used as the vehicle moves forward shortly on horizontal ground. Line $l$ can also be located precisely in the frame by referring to a 3D vertical line (e.g., a building rim). Then, the camera starts city journey.

As depicted in Fig.1, the real camera motion on road consists of translation $V(T_x,T_y,0)$, where $T_x=|V|$, and rotation around the $X, Y, Z$ axes described angular velocity $\Omega(R_x,R_y,R_z)$ in radian. For a 3D scene point $P(X,Y,Z)$ in the camera system $O\text{-}XYZ$, its composite motion is $V_c=V+P\times\Omega$. The flow vector $(u,v)$ at image point $p(x,y)$ is

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \dfrac{f}{Z}V + (f + \dfrac{x^2}{f})R_y - \dfrac{xy}{f}R_x - yR_z \\ \dfrac{f}{Z}T_y - (f + \dfrac{y^2}{f})R_x + \dfrac{xy}{f}R_y + xR_z \end{pmatrix} \quad (1)$$

where $f$ is the camera focal length [11]. Rotation $R_y$ for turning is usually small between consecutive frames. Because the wheel base of a vehicle is longer than its wheel span, the vehicle pitch is generally smaller than the vehicle roll, i.e., $R_x>R_z$.

Although the video deshaking based on point flow tracking can reduce abrupt camera movement, the used features may become noisy in the video containing fast moving scenes. Moreover, such short-term stabilization cannot straighten the $RP$ to a long distance that covers large architectures. In this work, we focus on a fewer but more distinct features observable for long time periods and use a data set much smaller than video volume. On the other hand, the stabilization directly on $RP$ has been proposed [5], which has two drawbacks: (a) the employed line tracking is hard to distinguish whether a curve is caused by shaking/waving, or from a slanted line in the 3D space; (b) line tracking can fail easily at repetitive patterns and occluded scenes.
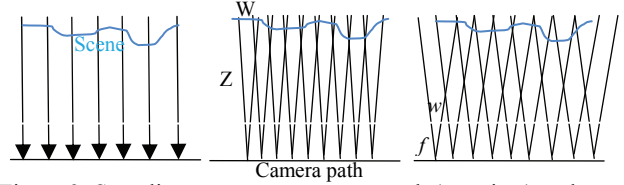


Figure 3: Sampling scenes on a camera path (top view) under an ideal parallel projection and narrow perspective projections respectively in obtaining a route panorama.

As a special effect in the $RP$, the *stationary blur* has been addressed in [2,7,13]. A pixel line in the camera has a certain physical width; the ray through a pixel is a cone realizing a narrow perspective projection (Fig. 3). Intensities are averaged in the cones as they are projected towards a pixel in the RP. Since the camera moves horizontally, consecutive cones overlap partially beyond a certain depth. A part of scenes thus appears in consecutive pixels in the RP, which causes the temporal burring in the RP.

We make positive use of the stationary blur for RP deshaking here. An observation can be made by looking at an extreme example shown in Fig. 2. A cloud pattern ($Z=\infty$) is blurred horizontally for blocks in the RP. Due to the vehicle rolling, the cloud trace is curved in the RP. Generally, distant features (large Z) stay in the video for long time as reliable references of the camera motion. Moreover, their image velocities (or motion parallax) are insensitive to the camera translation $V$ and $T_y$ according to Eq.1. By tracking a distant feature, we can infer the camera tilt and realign the RP for a long distance.

## 3. DIGITALLY INCREASED STATIONARY BLUR

To create a strong stationary blur, we collect incoming light digitally in the areas of $w\times1$ pixel in the image frame. We average $I(x,y)$ in width $w$ along the optical flow direction as

$$B(y) = \frac{1}{w}\sum_{x=-w/2}^{w/2} I(x, y(x)) \quad (2)$$

where $(x,y(x))$ is along the optical flow vector pre-calculated from *FOE*. The spatially collected 1D data from consecutive frames form a *condensed slice* (*CS*)

$$CS(t, y) = \frac{1}{w}\sum_{x=-w/2}^{w/2} I(x, y(x), t) \quad (3)$$

In *CS*, the stationary blur appears along the motion vectors



Figure 4: Spatially condensed slice, $CS(t,y)$, reveals subtle waves of the camera tilt. (top) $CS(t,y)$ ($w=320$ in $640\times480$ frame). (lower) Corresponding $RP(t,y)$ is also displayed.

while the sharpness orthogonal to the motion vector is maintained (Fig. 4). The CS also keeps a small data set to avoid storing fast growing video data during the motion.

***Features in CS:*** Distinct corner points and lines parallel to $V$ in the 3D space leave their weak yet continuous traces in $CS(t,y)$ sufficient to determine the changes in camera tilt. On the contrary, vertical and slanted lines (e.g., roofs on houses) in the space are blurred in $CS(t,y)$. The curves in $CS(t,y)$ thus are mainly caused by shaking. In addition, by stationary blurring, repetitive patterns such as windows and shortly occluded lines are connected as long traces in $CS$ (Fig.4). Moreover, $CS(t,y)$ is less influenced from complex occlusion, specular reflection, transparent scenes, and the camera motion blur in the image that affect stable matching in the flow approach.

***Duration of features appearing in CS(t,y):*** The stationary blurring converts long observable features (distant or horizontally extended) to long traces in $CS$ without computing their depths. Assume the camera sampling rate is $m$ lines per second, the observable duration of a feature in $CS(t,y)$ can be deduced as

$$\Delta t = m\left( L + \frac{2Zw}{f} \right)/V \qquad (5)$$

which is an additive relation of feature length $L$ and depth $Z$ in the 3D space. Thus, features ranging from distant corners (as in Fig. 2) to close lines will stay for long time in $CS$. This allows a global $RP$ rectification over a long distance.

## 4. MOTION DETECTION AND RP RECTIFICATION

In our RP deshaking, horizontal edges in $CS(t,y)$ are extracted at the peaks of its vertically differentiated $CS_y(t,y)$ with a low threshold. As time $t$ increases, a tracking algorithm follows edge points by examining their contrast and color to yield dense traces $y_i(t)$, where $i=1,2,3...$ Meanwhile, it also registers new traces. These traces visualize the camera rotation directly (Fig. 5).

In the $CS$, a stationary-blurred feature provides two components of motion: the horizontal blur indicates the image velocity of features, and the vertical deviation exhibits the camera shaking. Camera roll $R_z$ is small for a long vehicle wheelbase. Besides, a short change of $R_z$ only slightly widen and blur the horizontal traces in $CS(t,y)$. If we focus on the middle/high part of $CS(t,y)$ to exclude road surface (where $Z$ may be small), it is easy to prove from Eq. 1 that features in $CS(t,y)$ have small response to vertical $T_y$,

$$\frac{1}{w} \sum_{x=-w/2}^{w/2} v(x,y) \approx -(f + \frac{y^2}{f})R_x \qquad (6)$$

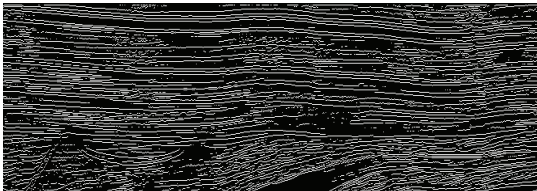Figure 5: Tracking of dense traces (non-crossing) in CS.

Therefore, the waving in CS is related to the rotation $R_x$, i.e.,

$$\frac{\partial y_i(t)}{\partial t} = \frac{CS_t(t,y)}{CS_y(t,y)} = \frac{\frac{\partial}{\partial t}\sum_{x=-w/2}^{w/2} I(x,y,t)}{\frac{\partial}{\partial y}\sum_{x=-w/2}^{w/2} I(x,y,t)} \cong \frac{1}{wI_y}\left( \sum_{x=-w/2}^{w/2} I_t \right) \qquad (7)$$

$$= \frac{1}{w}\sum_{x=-w/2}^{w/2} \frac{I_t}{I_y} = \frac{1}{w}\sum_{x=-w/2}^{w/2} \frac{\partial y(x)}{\partial t} \approx -(f + \frac{y^2}{f})R_x$$

We can thus detect the change of camera tilt from the trace tangent. The rectification of $RP(t,y)$ along a trace is done by shifting pixels at each column from $y(t)$ to $y_1(t)$ as

$$y_1(t) = -\frac{f^2 + y(t)^2}{f}\int_{t_0}^{t} R_x(t)dt + y_1(t_0) \qquad (8)$$

To find shaking locations and values from traces in $CS$, two median filters are applied vertically and horizontally on traces respectively. The first filter obtains a common vertical shift $\hat{y}(t)$ at time $t$ from multiple traces, if the number of traces is over threshold $n_1$. The second horizontal filter has a large span (e.g., 200~300 pixel) along the time axis for removing sharp sparks as noise in the $\hat{y}(t)$ sequence. The shift of vertical columns is at pixel level as shown in Fig. 6 such that both $RP$ and $CS$ do not change their resolution at this stage.

The local jitter removal is not sufficient for rectifying global waves. We further straighten large structures in the $RP$ according to waved traces in the $CS$. As shown in Figure 7, long curves in $CS(t,y)$ are tracked to yield traces $y_i(s_i,e_i)$, $i=1,2,3...$ with end points $s_i$, $e_i$ and length $e_i$-$s_i$, Then, non-overlapping longest traces are marked for every $t$ in the $CS$. For each position $t$, we find the longest trace $tr(t)$ that covers the position, i.e.,

$$\text{length\_tr(t)} = \max_{i=1,2,3...} \{length\_y_i(t)|\, t \in [s_i,e_i]\} \qquad (9)$$

We found that a single reliable trace yields a much better motion parameter than multiple traces do in $RP$ rectification, if it can be tracked for a long time. The rectification of $RP$ is to generate straight scenes while keeping them within the frame for image transmission and visualization.

## 5. EXPERIMENTS

Figure 6: Local $RP$s before and after rectification according to their $CS$.

We obtain $RP(t,y)$ and $CS(t,y)$ without storing entire video. Our vehicle moves at a speed of 15-30 mph depending on average depth of scenes along the street. Figure 7 shows a result of RP rectification where repetitive patterns, depth discontinuity, and large structures are visible. Our method works as long as a horizontal component appears in the image. If there is no single horizontal line or corner in the *CS,* the shaking may not affect the *RP* obviously since jittered lines influence *RP* perception the most. To maintain the quality in deshaking RP, edge traces in $CS(t,y)$ are extracted at the sub-pixel level at the accuracy of 0.33 pixel. To straighten $y_i(s_i,e_i)$ between ends $(s_i,e_i)$, a line $h_i(t)$ calculated at 0.1 pixel level is used as target line. The vertical difference, $\gamma_i(t)=h_i(t)-y_i(t)$, is used to shift *RP* columns at sub-pixel level. The piecewise image shift is converted to $R_x$ for RP rectification.

## 6. CONCLUSION

This paper proposed an efficient method to estimate jerky motion of a vehicle borne camera and rectify the route panorama for a high quality output. We create the stationary blur largely in a compact condensed slice and perform an automatic selection of long observable features. The features are robust to complex phenomena from scene occlusion to specular reflection. The rectification improved the quality of RPs at subpixel level to ensure the true shapes of large scale scenes. This approach is also effective for real time processing on moving vehicles.

### REFERENCES

[1] J. Y. Zheng, S. Tsuji, Panoramic Representation for Route Recognition by a Mobile Robot. Int. Journal Computer Vision, 9(1), 55-76, 1992.

[2] J. Y. Zheng, Digital Route Panorama, IEEE Multimedia, 10(3), 57-68, 2003.

[3] A Roman, G. Garg, M. Levoy, Interactive design of multi-perspective images for visualizing urban landscapes, IEEE Visualization 04, 537-544, 2004.

[4] A. Rav-Acha, G. Engel, S. Peleg, Minimal aspect distortion of long view, Int. Journal Computer Vision, 78(2-3), 187-206, 2008.

[5] J. Y. Zheng, Stablizing Route Panorama, 17th Int. Conf. Pattern Recognition, 4, 348-351, 2004.

[6] P. Preisig, D. Kragic, Robust statistics for 3D object tracking, IEEE, Int. Conf. Robotics and Automation, 2403-2408, 2006.

[7] J. Y. Zheng, M. Shi, Scanning depth of route panorama based on stationary blur, Int. Journal of Computer Vision, 78(2-3), 169-186, 2008.

[8] T. N. Thanh, et al. Robust and real-time rotation estimation of compound omni-directional sensor, IEEE Int. Conf. Robotics and Automation, 2007.

[9] A. Rav-Acha, S. Peleg, Two motion-blurred images are better than one, Pattern Recognition Letter, 26, 311-317, 2005.

[10] J. Ye, G. Fu, P. Poudel, High-accuracy edge detection with blurred edge model, IVC, 23(5), 453-467, 2005.

[11] A. Bruss, B. K. P. Horn, Passive navigation, CVGIP, 1983.

[12] J. Y. Zheng, Y. Bhupalam, H. Tanaka, Understanding vehicle motion via spatial integration, 19th Int. Conf. Pattern Recognition, 1-5, 2008.

[13] J. Y. Zheng, M. Shi, Removing Temporal Stationary Blur in Route Panoramas. 18th Int. Conf. Pattern Recognition, (3) 709-713, 2006.

Figure 7: The process to rectify an RP with distance feature long lasting in the CS. Displayed are, RP (24 second long), CS with tracked traces (red), non-overlapping longest traces (red) in CS, and RP after rectification at sub-pixel level using longest traces.