

# Understanding Vehicle Motion via Spatial Integration of Intensities

Jiang Yu Zheng<sup>1</sup>, Yayaswy Bhupalam<sup>1</sup>, Hiromi T. Tanaka<sup>2</sup>

<sup>1</sup>Indiana University Purdue University Indianapolis, USA

<sup>2</sup>Ritsumeikan University, Japan

jzheng@cs.iupui.edu hiromi@cv.ci.ritsumei.ac.jp

## Abstract

*On a moving vehicle, speedy motion extraction from video is demanded. Different from the traditional motion estimation methods that track or match 2D features in consecutive motion-blurred images, this work explores a novel approach to find motion without explicit shape analysis. We take spatial integration of intensities in each frame and the obtained consecutive profiles provide distinct motion traces in a visual form. The camera motion can be followed in the integrated images briefly, rather than from complex feature extraction and iterative computation. The resulting motion parameters are used for normalizing length and removing shaking of route panoramas. The results are robust and the method is particularly efficient for real time processing.*

## 1. Introduction

A vehicle borne camera performs on-vehicle tasks such as scanning route scenes [1,15], detecting passing vehicle, and understanding ego motion in 3D scenes. The motion estimation for a freehand camera is based on points [2,4,5,11,14], lines, regions [16], and is not always workable for a vehicle borne camera due to the following reasons: (1) processing video frames with iteration and statistics computation is time consuming [5] and is hard to be implemented on the fly as the vehicle moves; (2) the motion blur [13] in each frame affects the 2D localization and then motion estimation; and (3) the features that provide strong evidence of motion are not always available when the camera moves for a long distance.

To obtain the vehicle motion constantly and robustly, we set a series of linear receptors to collect intensities in the video frame along a certain direction. The temporal 1D profiles from the receptors provide visible traces that directly reflect the vehicle motion. As the results, we are able to detect motion from consecutive image profiles without scene analysis in

2D images. The handled data are much compact compared to the entire video and thus can speed up the motion detection significantly. Also, the spatial integration of intensities serves as a function of selecting long and consistent lines for robust motion detection. The proposed method is verified in the modification of distorted route panoramas.

In the following, we start with the intensity integration in Sec. 2, and apply them to a vehicle-borne camera in Sec. 3. Sec. 4 discusses the motion detection in the spatially integrated images, followed with experiments in Sec. 5 for normalizing and deshaking route panoramas.

## 2. Temporal and Spatial Integration

In an image with a long exposure time as Figure 1, vehicle lights draw bright traces that show their trajectories on a street, while the vehicle shapes are motion-blurred. Such a temporally integrated image [12] is effective for understanding motion without shape analysis. The condition of this approach is the motion consistency and distinct object contrast against the background.



Fig. 1 Temporal integration in the field of view forms an image with traces of moving objects.

In contrast to the temporal integration for object motion, this work explores the spatial integration of images for the camera ego-motion. Figure 2 illustrates a spatial-temporal volume (video volume) representing intensity  $I(t,x,y)$ . An image point  $p(x(t),y(t))$  moving in the Field of View (FOV) generates a trace in the volume. The temporally and spatially integrated images from the video volume are

TM:

$$S(x, y) = \frac{1}{L} \sum_{t=t_0}^{t_0+L} I(t, x, y)$$

SP:  $T(t, x) = \frac{1}{M} \sum_{y=y_0}^{y_0+M} I(t, x, y) \quad R(t, y) = \frac{1}{N} \sum_{x=x_0}^{x_0+N} I(t, x, y) \quad (1)$

where  $L$ ,  $M$ , and  $N$  are the scopes for integration. Different from the temporal integration realized with a slow shutter speed, the spatial integration is obtained digitally or optically with a prism. We refer to  $T(t, x)$  and  $R(t, y)$  as *condensed images* which preserve the motion information to some extent.

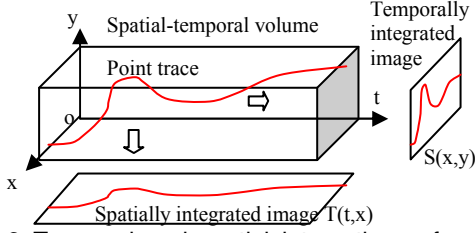


Fig. 2 Temporal and spatial integrations of a spatial-temporal volume for motion sensitive profiles.

The image differential at  $p$  is  $\nabla I = (\partial I(t, x, y) / \partial x, \partial I(t, x, y) / \partial y)$ , and the image velocity  $(u, v)$  is  $(\partial x(t) / \partial t, \partial y(t) / \partial t)$ . On the other hand, the differentials  $\partial T(t, x) / \partial t$  and  $\partial R(t, y) / \partial t$  in condensed images  $T(t, x)$  and  $R(t, y)$  are

$$\frac{\partial T(t, x)}{\partial x} = \frac{1}{M} \sum_y \frac{\partial I}{\partial x} = \frac{1}{M} \sum_y \frac{\partial I}{\partial t} \frac{\partial t}{\partial x} \quad \frac{\partial T(t, x)}{\partial t} = \frac{1}{M} \sum_y \frac{\partial I}{\partial t} \quad (2)$$

$$\frac{\partial R(t, y)}{\partial y} = \frac{1}{N} \sum_x \frac{\partial I}{\partial y} = \frac{1}{N} \sum_x \frac{\partial I}{\partial t} \frac{\partial t}{\partial y} \quad \frac{\partial R(t, y)}{\partial t} = \frac{1}{N} \sum_x \frac{\partial I}{\partial t}$$

according to Eq. 1. If the point has a sufficient contrast to other points horizontally or vertically, the trace becomes visible either in  $T(t, x)$  or  $R(t, y)$ , or even in both. Therefore, tracking the trace in  $T(t, x)$  and  $R(t, y)$  can obtain  $(u, v)$ . Alternatively, computing the gradient orientation in  $T(t, x)$  and  $R(t, y)$  can yield  $(u, v)$  as

$$\frac{\partial x}{\partial t} = \frac{\partial T}{\partial t} / \frac{\partial T}{\partial x} \quad \frac{\partial y}{\partial t} = \frac{\partial R}{\partial t} / \frac{\partial R}{\partial y} \quad (3)$$

In general, feature points visible in the FOV are not as distinctive as lights in a night view. To make shapes distinct in a spatially integrated image for finding motion traces, we align a set of receptors to collect the intensities along the flow direction. The edges along the motion direction enhance the trace contrast, while the edges orthogonal to the direction diverse the contrast (increasing base of intensity). Slanted lines may generate weak traces in  $T(t, x)$  or  $R(t, y)$ . Therefore, the motion that deviate traces can be detected in the condensed images.

### 3. Vehicle-Borne Camera and Images

Different from freehand camera motion, a vehicle borne camera undertakes two dominant motions at any time: the translation velocity along a horizontal path

adjusted by a steering rotation velocity around the vertical axis; both generate the image velocity. In addition, a vehicle is disturbed by bumping due to uneven roads. Figure 3 depicts the vehicle motion and our camera setting, which is characterized in translation velocities  $V(V_x(t), V_y(t), V_z(t))$  with  $V_x$  in heading and  $V_y$  in vertical direction. Because a four wheeled vehicle has no sideslip,  $V_z=0$ . Rotation velocities are  $\Omega(\omega_x(t), \omega_y(t), \omega_z(t))$  in radian around  $X, Y, Z$  axes, respectively.

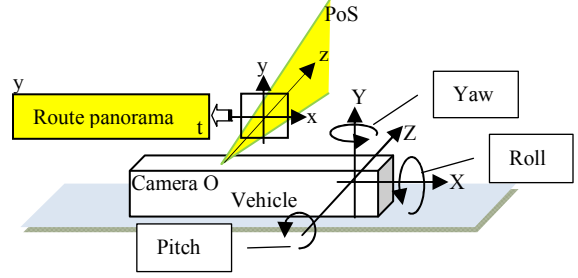


Fig. 3 Pitch-yaw-roll of a vehicle and corresponding pan-tilt-roll of a camera mounted sideways.

The purposes of our motion understanding are for shake detection and aspect ratio improvement for the route panorama [1]. Two types of linear structures are observable on streets: (1) vertical lines on buildings and poles, (2) horizontal lines on architecture, lanes, and sidewalk. Our ideal camera setting aligns the optical axis horizontally and orthogonal to the vehicle heading. The projections of these lines in the image are hence vertical and horizontal as well. If the camera is not set precisely in the ideal direction, a calibration matrix can map the real image to a virtual image plane in the ideal direction.

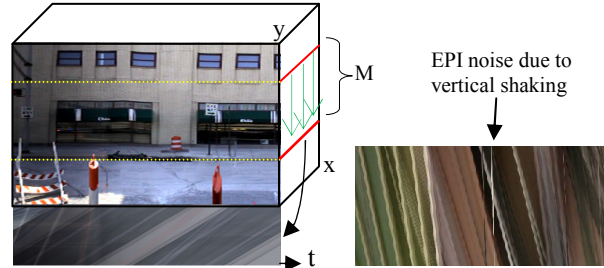


Fig. 4 A spatially integrated image and some failure of EPI due to up and down shaking of camera

As shown in Fig. 4, the vertical accumulation of intensities in each frame obtains a profile and the piled profiles along the time axis form condensed image  $T(t, x)$ . The horizontal movements of vertical lines display as traces in  $T(t, x)$ , while non-vertical lines are blurred in  $T(t, x)$ . The sharpness and length of vertical lines contribute to the contrast of traces. Small vertical jitters of the lines will rarely affect the trace contrasts because of the intensity integration effect.

This condensed image is similar to an Epipolar Plane Image (EPI) [2,9] from a single pixel line. However, they differ at several aspects. (1) The EPI is not defined for general rotation. Curved Epipolar surfaces are undeterminable from motion parameters only. (2) An EPI is susceptible to vertical jitters. The traces in an EPI may be frequently disturbed by edges irrelevant to the motion (Fig. 4). Condensed image  $T(t,x)$  is, however, strong to vertical image motion.

The horizontal intensity integration for  $R(t,y)$  extracts horizontal structural lines along the vehicle path. As the camera experiences translation  $V_x$ , features move horizontally in the frame. Linear structures along the path direction add their intensities to the vertical profile and yield sharp edges in it. The temporally connected profiles form condensed image  $R(t,y)$  as Fig. 5, in which long horizontal lines leave their edges. The detailed camera motion can be figured out visually from the deviated edge traces. We can filter the gradient on the curves to estimate the motion.

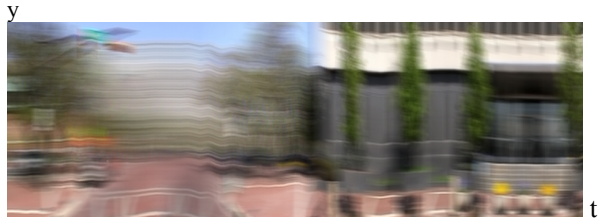


Fig. 5 Integration of intensities within a width of 32 pixels along the motion direction in the image plane. The curves in the condensed image show the camera shaking due to vehicle bumping.



Fig. 6 Horizontally integrated image for motion detection. (a) Route panorama. (b) Condensed image in which traces are continuous on repetitive windows.

Condensed image  $R(t,y)$  is similar to the route panorama (RP) obtained by piling temporal data from a vertical pixel line in the image frame [1]. However,  $R(t,y)$  emphasizes more motion trajectory than shape. Because non-horizontal lines are *stationary blurred* [15], the horizontal curves are more distinct in  $R(t,y)$  than in  $RP(t,y)$ , which is suitable for robust gradient calculation. As shown in Fig. 6, an  $RP(t,y)$  image has many vertical lines that may prevent continuous tracking of horizontal curves [3]. The corresponding

$R(t,y)$ , however, provides longer traces than  $RP(t,y)$  even on repetitive patterns.

#### 4. Vehicle Motion from Condensed Images

The spatial integration of our video corresponds to intensity accumulations horizontally and vertically in the camera frame. For a 3D point,  $\mathbf{P}(X,Y,Z)$ , its relative velocity with respect to the camera centered coordinate system is

$$\mathbf{V}_c(t) = \mathbf{V}(t) + \boldsymbol{\Omega}(t) \times \mathbf{P}(t) \quad (4)$$

Assuming a known camera focal length  $f$ , the image velocity  $(u,v)$  of  $\mathbf{P}$  can be obtained through perspective projection to  $\mathbf{p}(x,y)$  along with differential  $\partial \mathbf{p} / \partial t$ , i.e.,

$$\begin{pmatrix} u(t,x,y) \\ v(t,x,y) \end{pmatrix} = \begin{pmatrix} \frac{f}{Z} V_x(t) + (f + \frac{x^2}{f}) \omega_y(t) - \frac{xy}{f} \omega_x(t) - y \omega_z(t) \\ \frac{f}{Z} V_y(t) - (f + \frac{y^2}{f}) \omega_x(t) + \frac{xy}{f} \omega_y(t) + x \omega_z(t) \end{pmatrix} \quad (5)$$

Besides the dominant motion  $V_x$  and  $\omega_y$ , a vehicle bumping is from abrupt changes in  $V_y$ , vehicle roll (camera tilt)  $\omega_x$  and vehicle pitch (camera roll)  $\omega_z$ . Generally,  $\omega_x > \omega_z$  based on vehicle structures. In Eq. 5, only  $fV_x/Z$  and  $fV_y/Z$  are scene related; other terms are from camera and motion. If a vertical line is through the entire integration scope, the image velocity in  $T$

$$\mu(t,x) = \frac{1}{M} \sum_{y=-M/2}^{M/2} u(t,x,y) \approx \frac{f}{Z} V_x(t) + (f + \frac{x^2}{f}) \omega_y(t) \quad (6)$$

Similarly, we can obtain the image velocity in  $R(t,y)$  as

$$\eta(t,y) = \frac{1}{N} \sum_{x=-N/2}^{N/2} v(t,x,y) \approx \frac{f}{Z} V_y(t) - (f + \frac{y^2}{f}) \omega_x(t) \quad (7)$$

for a long horizontal line. For instant rotation  $\omega_z$  in a few images, slanted lines may only blur edges in the corresponding profiles and then weaken the trace contrast in the condensed images.

The motion extraction is implemented separately in condensed images  $T(t,x)$  and  $R(t,y)$ . In  $T(t,x)$  image, the vehicle translation along a road and the rotation at an intersection can be dealt with separately. During the vehicle translation (with our camera setting), traces are linear in  $T(t,x)$  for a constant  $V_x$ , and their slopes depend on object distances  $Z$  [2]. At intersections, the vehicle steering skews entire  $T(t,x)$  shortly.

We extract image points satisfying  $\|\nabla T\| > \delta_1$  in  $T(t,x)$  and obtain  $\mu(t,x)$  on traces. The traces are curved if the vehicle has acceleration  $\partial V_x / \partial t$ . We straighten all the traces in  $T(t,x)$ . To implement this, we calculate a function  $A(t)$  at every moment to make a distinct curve,  $trace_i$ , linear, i.e.,  $\mu_i(t,x) A(t) = C_i$  where  $C_i$  is a constant. We multiply  $\mu(t,x)$  in the entire profile with  $A(t)$  to normalize  $T(t,x)$  as well as the corresponding  $RP(t,y)$ . This local scaling process will switch to another trace,  $trace_{i+1}$ , after  $trace_i$  ends.

On the other hand, traces in  $R(t,y)$  are mainly

jittered by camera tilt  $\omega_x$ . The influence from  $V_y$  is small due to large  $Z$  to most scenes (except road area close to bottom of  $R(t,y)$ ). We first find edge points in  $R(t,y)$  according to the gradient, i.e.,  $\|\nabla R\| > \delta_2$ , where  $\delta_2$  is a small threshold. The edge points are then followed for traces that exceed a certain length. To guarantee that the motion estimation is from traces rather than slanted object shapes, we track non-vertical edges and calculate gradient  $\eta(t,y)$  only on long traces. Rotation  $\omega_x(t)$  is thus obtained from the common value of  $\eta(t,y)$  at different traces using the least squared error method. Further,  $V_y(t)/Z$  is obtained for points in  $R(t,y)$  based on the common  $\omega_x(t)$  at each instance.

## 5. Experiments

The intensity accumulation for  $T(t,x)$  and  $R(t,y)$  can be done fast at the image input stage or even at hardware level; it achieves a low-cost feature selection function. We have tested long vehicle sequences along many streets at the rate of 60Hz. If the camera is not precisely aligned, the projections of vertical and horizontal lines may have their extensions passing through a vanishing point and a Focus of Expansion (FOE) respectively as shown in Fig. 7. Nevertheless, one can calibrate the camera and map a view to the image of the ideal camera setting in Fig. 3 for the spatial integration. Alternatively, we can accumulate intensities along the pixel lines through the vanishing point and FOE respectively for condensed images.

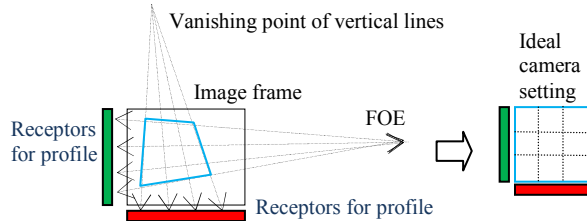


Fig. 7 Field of View of a camera and directions of projected structure lines

To verify the correctness of the extracted data, we use the angle  $\omega_x(t)$  to rectify distorted route panoramas  $RP(y,t)$ . Due to the vehicle shaking, the horizontal structure lines are jittered, and this can be detected in  $R(t,y)$ . Figure 8 and 9 display jitters in  $R(t,y)$  and are rectified using the extracted motion values. The results proved the sensitivity of the algorithm in finding precise motion. Figure 10 is another experiment on a longer road segment.

The experiment on the length normalization of the local RP is shown in Fig. 11 such that the resulting RP corresponds to the one from a constant vehicle speed. The width of a scene is proportional to its real width in the 3D space (Fig. 11a). Moreover, we transform the

RP such that the major trace in  $T(t,x)$  are aligned at 45 degree piecewise (Fig.11b). This yields an aspect ratio of objects in the RP close to the ratio in a perspective image [1,15,16].

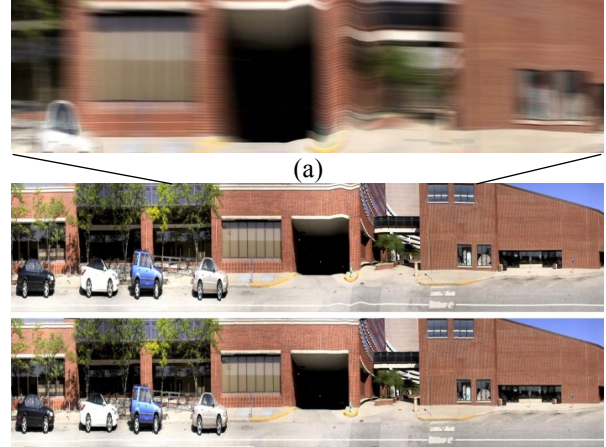


Fig. 8 A rectified route panorama by removing jitters caused by camera bumping. (a) Condensed image with dense traces from building texture. (b) RPs before and after rectification based on the gradient of traces.

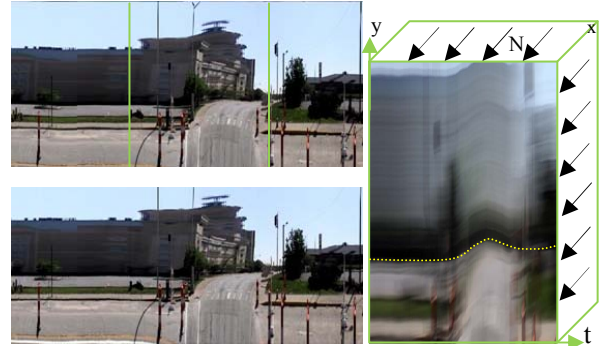


Fig. 9: A condensed image section with visible motion traces. (left) Original  $RP(t,y)$  and rectified  $RP(t,y)$ . (right) Enlarged condensed image  $R(t,y)$  shows curved building traces in a range marked in  $RP(t,y)$ .

## 6. Conclusion

In this paper, we have proposed a visual approach to detect vehicle-borne camera motion without intensive shape analysis in images. We integrate intensities along the flow directions to obtain condensed images where object motion is displayed as traces directly. The process not only simplifies the 2D motion estimation to 1D profile processing, but also selects and emphasizes distinct features for robust motion extraction. It is suitable for sensor development on a fast moving vehicle and archiving motion in a compact form for interactive processing.

## References

- [1] J. Y. Zheng, Digital route panorama, IEEE Multimedia,

- 10(3), 57-68, 2003.
- [2] M. Fishler, R. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, *Comm. ACM*, 24, 381-395, 1981.
  - [3] J. Y. Zheng, Stabilizing Route Panorama, 17th ICPR, 4, 348-351, 2004.
  - [4] Y. S. Yao, R. Chellappa, Selective stabilization of image acquired by unmanned ground vehicles, *IEEE Trans. Robotics and Automation*, 13, 693-708, 1997.
  - [5] P. Preisig, D. Kragic, Robust statistics for 3D object tracking, *IEEE, ICRA*, 2403-2408, 2006.
  - [6] S. Srinivasa, Extracting structure from optical flow using the fast error search technique, *IJCV*, 37(3), 203-230, 2000.
  - [7] J. Weng, Y. Cui, N. Ahuja, Transitory image sequences, asymptotic properties, and estimation of motion and structure, *IEEE PAMI*, 19(5), 451-464, 1997.
  - [8] J. Kim, V. Kolmogorov, R. Zabih, Visual correspondence using energy minimization and mutual information, *ICCV 03*, 1033-1040, 2003.
  - [9] Y. Yagi, K. Imai, K. Tsuji, M. Yachida, Iconic memory-based omnidirectional route panorama navigation. *IEEE PAMI*, 27(1): 78-87, 2005.
  - [10] C. Lee, J. S. Bethel, Geo-registration of airborne hyperspectral image data, *IEEE Trans. Geo. and Remote Sensing*, 66(4), 385-392, 2000.
  - [11] D. Lowe, Distinctive image features from scale-invariant keypoints, *IJCV*, 60(2), 91-110, 2004.
  - [12] M. Irani, B. Rousso, S. Peleg, Detecting and tracking multiple moving objects using temporal integration. *ECCV 1992*: 282-287.
  - [13] M. Ben-Ezra, S.K. Nayar, Motion-based Motion Deblurring, *IEEE PAMI*, 26(6), 689-698, 2004.
  - [14] T. N. Thanh, et al. Robust and real-time rotation estimation of compound omni-directional sensor, *IEEE Int. Conf. Robotics and Automation*, 2007.
  - [15] J. Y. Zheng, M. Shi, Scanning depth of route panorama based on stationary blur, *IJCV*, 78(2-3), 169-186, 2008.
  - [16] A. Rav-Acha, G. Engel, S. Peleg, Minimal aspect distortion mosaic, *IJCV*, 78(2-3), 187-206, 2008.

Fig. 10 (right two images): Rectifying waved route panorama by tracking major curves in the condensed image. Condensed image and RP pairs before and after rectification are displayed in order. Scenes are straightened according to the tracked long curves.

Fig. 11 (right most images): Route panoramas  $RP(t,y)$  and attached condensed images  $T(t,x)$  showing the length changes after local length normalization and aspect ratio adjustment in  $RP(t,y)$  overtime. Pairs of  $RP$  and  $R$  before and after normalization are displayed.

