A Real-Time Vision System for Automatic Traffic Monitoring Based on 2D Spatio-Temporal Images

Zhigang Zhu, Bo Yang, Guangyou Xu and Dingji Shi
Department of Computer Science and Technology
Tsinghua University, Beijing 100084, China
zzg@vision.dcs.tsinghua.edu.cn

Abstract

In this paper we present a novel approach using 2D spatio-temporal images for automatic traffic monitoring. A TV camera is mounted above the highway to monitor the traffic through two slice windows for each traffic lane. One slice window is along the lane and the other perpendicular to the lane axis. Two types of 2D spatio-temporal(ST) images are used in our system: the panoramic view image (PVI) and the epipolar plane image (EPI). Our real-time vision system for automatic traffic monitoring, VISATRAM, an inexpensive system with a PC486 and an image frame grabber has been tested with real road images. The system can not only count the vehicles and estimate their speeds, but also classify the passing vehicles using 3D measurements (length, width and height). The VISATRAM works robustly under various light conditions including shadows in the day and vehicle lights at night, and automatically copes with the gradual and abrupt changes of the environment.

1. Introduction

Automatic traffic monitoring plays an important role in the truly Intelligent Vehicle/Highway System(IVHS). Vision-based approach is promising since it requires no pavement adjustments and has more potential advantages such as larger detection areas and more flexibility. However traffic flow raises interesting but difficult problems for image processing. The various light conditions places a strong need on the robust algorithms, which require a great amount of computational power to meet the real-time operations of the traffic monitoring system. Much research effort has been made in this area [1, 2, 5, 6, 7], but most of the current commercial traffic monitoring image systems are cost expensive (e.g. [1], [3]).

In this paper we present a novel approach using 2D spatio-temporal images for automatic traffic monitoring. A TV camera is mounted above the highway to monitor the traffic through two slice windows for each traffic lane. One slice window is along the lane and the other perpendicular to the lane axis. Two types of 2D spatio-temporal(ST) images are used in our system: the panoramic view image (PVI) and the epipolar plane image (EPI). The vehicle counting, speed estimation, and vehicle classification are solved through analyzing these two 2D ST images. Our real-time vision system for automatic traffic monitoring, VISATRAM, an inexpensive system with a PC486 and an image frame grabber has been tested with real road images. Our VISATRAM works robustly under various light conditions including shadows in the day and vehicle lights at night, and automatically copes with the gradual and abrupt changes of the environment.

Nakanishi and Ishii[4] also presented a method for extracting images of laterally moving vehicles from image sequences based on spatio-temporal image analysis. The problem of occlusion and background updating under typical daylight variations were addressed. They detected the locus of the vehicle using Hough transform and classified the vehicle type based on silhouette analysis. Their experiments were carried out in a Sparc Station 1 but the real-time operations and nighttime operations were not mentioned. Our work is mostly related to their work, but the camera setting and algorithms are quite different from theirs. Our approach has the following new features and advantages:

(1) *Real-time and robust performance*: VISATRAM is a real-time visual system working robustly under various light conditions including shadows and vehicle lighting. It can automatically cope with the slow and sudden illumination changes. The system can automatically recover from false sensing or abrupt environment changing.

(2) *Enhanced functions.* Our system can not only count the vehicles and estimate the speed, but also classify the passing vehicles using 3D measurements (length, width and height). Moreover, robust speed and height estimation are obtained from the loci of the vehicle's front and rear edges instead of using the vehicle's locations at two different instants [1].

(3) *Low cost* . More traffic parameters are obtained by using 2D spatio-temporal image techniques in inexpensive image processing system (PC486 + frame grabber). Only a few scan lines are processed in each frame, and ST images are more generic and simple than frame images in this special application. Narrow spatial viewing windows are

0-8186-7620-5/96 $5.00

compensated by dense temporal sequences, vehicles which are partially viewed in a single frame can be reconstructed by using ST images.

(4) Compact representation. PVIs are a compressed and panoramic representation of the traffic flow and they can be saved on the hard disk. ST images are also suitable for performance analysis of the traffic monitoring system when the ground truth information is lacked.

2. Spatio-temporal geometry

2.1. Camera setting and ST image geometry

In the ideal system setup, the camera is mounted over the center of the highway, although other camera mountings are possible. The pan/zoom/tilt settings have to be fixed to retain detection configurations. Basically, we assume that vehicles move away from the camera with constant speeds along the straight lanes within the range of camera's viewing zone (Fig. 1). This setting is suitable for detecting and tracking the vehicle since the vehicle enters into the field of view in the high resolution end of the image (see Fig. 2). This setting also reduces the negative effect of the vehicle's front light during nighttime, since the light is not directly reflected to the camera. The system is designed to work in various conditions including heavy shadows, dim light, and nighttime conditions. Auto iris is permitted and sometime advantageous for the background updating.

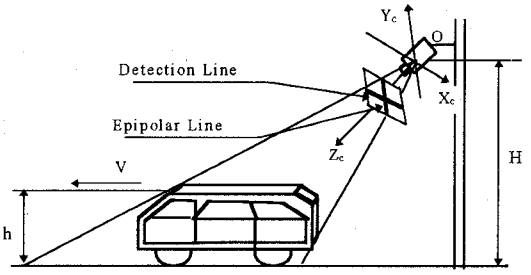
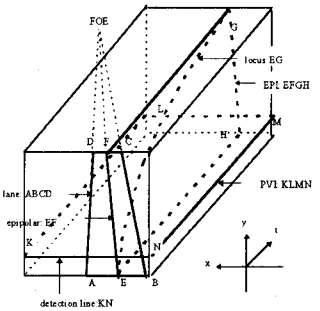


Fig. 1. The camera setting



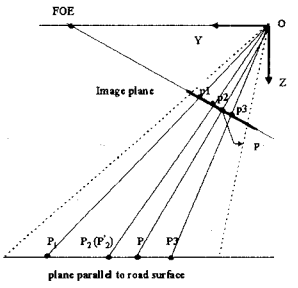


Fig. 2. ST geometry **Fig. 3. Image rectification**

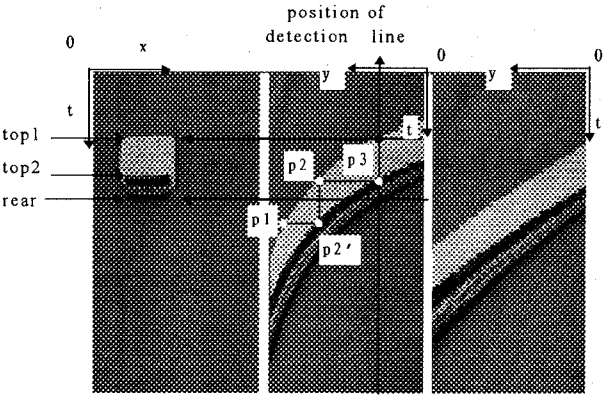
Inside the 3D ST image cube xyt, PVI and EPI are two kinds of representative 2D intersecting planes that reveal most of the traffic flow information (Fig. 2, Fig. 4). The PVI is formed by piling the horizontal detection lines from the consecutive frames. It shows the presence, the width W and duration time T of the vehicle across the detection line, while the EPI tells the speed V, length L and height h of the vehicle. The size and class of the vehicle can be easily obtained by integrating measures from the above two sources. Based on these measures, other traffic parameters such as volume, occupancy, headway, etc., can also be recovered easily.

2.2. ST image rectification

The camera is calibrated in order to find the relationship between the world coordinates and the image coordinates. First FOE is estimated by using the lane boundaries. Then a given image point of a vehicle will moves along the epipolar line that passes through both this point and the FOE (Fig. 2). However the locus of the point in the EPI, which is formed by piling up the epipolar lines, could not be a straight line if the optical axis is not perpendicular to the road surface, which is almost the case in the practical setting (Fig. 4(2)). Therefore, we re-project the sensor image to a rectified image plane parallel to the road surface. The rectification is only made along the selected epipolar lines inside the lanes and the cross ratio invariance of projective geometry is applied (Fig. 3):

$$\frac{\overline{PP_1}/\overline{PP_2}}{\overline{pp_1}/\overline{pp_2}} = \frac{\overline{P_3P_1}/\overline{P_3P_2}}{\overline{p_3p_1}/\overline{p_3p_2}} = \lambda(\text{constant}) \qquad (1)$$

Without measuring any actual points in the world coordinates, we use the fact that the real size of the moving vehicle is not changed (i.e. $P_3P_2 = P_2P_1$, where P_3 and P_2 are the y coordinates , at a time instant, of two points on top of the vehicle with the same height, while P_1 is the new coodinate of point P_2 at the instant when P_3 (P'_2) reaches P_2. Their corresponding points in the EPI , p_1, p_2, p'_2 and p_3 , are shown in Fig. 4(2)). In this way the original EPI with curved locus is transformed to a rectified EPI with straight locus, whose slope is proportional to the speed of the vehicle (Fig. 4(3)).



(1)PVI (2)Original EPI (3) Rectified EPI

Fig. 4. 2D ST images

2.3. 2D ST image models

After the image rectification, the image coordinates, (x,t) in the PVI and (y,t) in the EPI, of the world coordinates

(X,Y,Z) are measured in the rectified "camera" and image plane under the pinhole camera model

$$x(t) = f\frac{X(t)\sin\theta}{H-h},\ y(t) = f\frac{Y(t)}{H-h} \quad (2)$$

where H is the height of the camera from the road , h is the height of the vehicle's point (i.e. Z=H-h), f is the equivalent focal length, and θ is angle between road surface and the plane passing through the optical center and the detection line. The coordinates x(t), y(t), X(t) and Y(t) are all the functions of time t. H, f and θ can be easily decided by a simple calibration procedure using the PVI and rectified EPI of a vehicle with known size(length, width and height).

3. Vehicle metric measurements

3.1. Speed and size estimation

The EPI approach is superior to the two-frame approach in simplicity and robustness. There is no correspondence problem. Speed can be estimated using more than two points in the locus. All we need to do is to extract the (straight) locus and calculate the slope.

When the vehicle moves away from the camera, the point on the rear is roughly considered as ground point (i.e. h=0) and the point on the front is considered on the roof of the vehicle. The speed V and the height h of the vehicle can be estimated using the locus slope of the rear, v_g, and that of the front, v_h:

$$V = \frac{H}{f}v_g,\quad h = H(1-\frac{v_g}{v_h}) \quad (3)$$

Sometime the front and the rear of a vehicle can not be presented in the same view if the vehicle is too large. But there is no problem in the ST image approach. The length L of the vehicle can be calculated as the production the speed V and the duration time T. Considering the projective distortion, the length and width of the vehicle can be simply estimated by the following equations when a simplified cuboid vehicle model is used:

$$L = V\cdot T - h\cdot ctg\theta \quad (4a)$$

$$W = \frac{1}{f\cdot\sin\theta}\cdot(x_R\cdot(H-h)-x_L\cdot H) \quad (4b)$$

where x_L and x_R are left and right boundaries of the vehicle in the PVI. Based on these measurements, other traffic parameters for each lane, such as class (classified by the size, e.g. car, truck, trailer...), volume (number of vehicle detected during the time intervals), occupancy (lane occupancy measured in percent of time), headway (average time interval between vehicles), mean speed (average vehicle speed in the lane), etc., can be obtained without difficulties.

3.2. Performance analysis

The system monitors the road at the speed of 50 fields per second (M = 50 fields/second). If the length of the vehicle is L_{veh} and the speed is V (km/hr) , and the length and width of the effective field of view (FOV) are L_{FOV} and W_{FOV} respectively, then we have

$$N_{PVI} = M\cdot n\cdot L_{veh}/V \quad (5)$$

and

$$N_{EPI} = M\cdot L_{FOV}/V \quad (6)$$

where N_{PVI} and N_{EPI} are the numbers of vehicle's presence at the detection lines and epipolar lines respectively, n is the number of scan lines extracted inside the detection slice, and it varies according to the mean speed of the vehicles during a certain time interval. The precision of the measured length and width of a vehicle can be estimated as

$$\Delta L = \frac{L_{FOV}}{N_y}\cdot\Delta y,\ \Delta W = \frac{W_{FOV}}{Nx}\cdot\Delta x \quad (7)$$

where Ny and Nx are the effective field image resolutions in y and x directions (e.g. Ny=256 pixels, Nx = 256 pixels) respectively, and Δ y and Δ x are the localizing precision in the image. Suppose $L_{veh} = 4m$, $V = 144Km/hr$, $L_{FOV} = 13.2m$, $W_{FOV} = 16m$, $n = 3$, $\Delta y = \Delta x = 2$ pixels, then we have $N_{PVI} = 18$, $N_{EPI} = 20$, ΔL=0.103m, Δ W=0.125m. It shows that the vehicle can be reliably detected and measured, and the system is capable of detecting vehicles from very small to very huge.

The range of vehicle speed that can be estimated is from 0 to 160 km/hr. The speed precision can be estimated as follows. From equation (3) we have

$$V = \frac{H}{f}\cdot\frac{d}{\tau}\cdot s \quad (8)$$

where d is the pixel interval (mm/pixel) along y axis, τ is the temporal sample rate($1/50\ s$), s is the slope of the image locus of a ground point measured in pixels. So the absolute error and the relative error of the speed can be computed as

$$\Delta V = \frac{H}{f}\cdot\frac{d}{\tau}\cdot\Delta s,\quad \frac{\Delta V}{V} = \frac{\Delta s}{s} \quad (9)$$

If only two points on the locus are used to calculate the slope, then we have

$$s_2 = \frac{N_y}{N_{EPI}},\ \Delta s_2 = \frac{N_{EPI}\cdot dN_y + N_y\cdot dN_{EPI}}{N_{EPI}^2} \quad (10)$$

where dN_y and dN_{EPI} are the localization errors for N_y and N_{EPI} respectively. The accuracy can be improved if the slope is estimated by fitting the locus using least square mean method. The procedure can be considered approximately as

calculating the average of $\frac{N_{EPI}}{2}$ slope values s_2 that come from $\frac{N_{EPI}}{2}$ pairs of points. If the errors Δs_2 are independent Gaussian noises, then the error of s can be reduced to

$$\Delta s = \frac{\frac{N_{EPI}}{2} dN_y + \frac{N_y}{2} dN_{EPI}}{\left(\frac{N_{EPI}}{2}\right)^2} / \left(\frac{N_{EPI}}{2}\right) = \frac{4}{N_{EPI}} \Delta s_2 \qquad (11)$$

Suppose H=10 m, f = 5 mm, N_y = 256, d=6.6/256 mm/pixel, L_{FOV}=13.2m, dN_y=2, dN_{EPI} = 2, then the speed precision can be estimated using equations (6) and (8) to (11). Table 1 gives the theoretical results under different vehicle speeds.

From eqn. (3) the absolute error for the height estimation can be calculated as

$$\Delta h = H \frac{s \cdot \Delta s_v + s_v \cdot \Delta s}{s_v^2} \qquad (12)$$

where s_v is the slope of the image locus of point on the vehicle's roof measured in pixels. The absolute errors for a two-meter-high vehicle in different speeds are also listed in Table 1.

Table 1. performance analysis

V (km/hr)	0	10	50	80	120	160
s	0	1.077	5.387	8.620	12.929	17.239
s_v	0	1.347	6.733	10.775	16.160	20.994
N_{EPI}		238	48	30	20	15
Δs_2		1.74e-2	2.64e-1	6.36e-1	1.38	2.41
Δs	0	2.92e-4	2.20e-2	8.48e-2	2.76e-1	6.42e-1
ΔV (km/hr)	0	2.71e-3	0.204	0.787	2.56	5.96
$\Delta V / V$	*	0.03%	0.41%	0.98%	2.13%	3.72%
Δh (m)	0	0.004	0.059	0.142	0.312	0.556

4. Real-time implementation

4.1. System overview

A VIsual System for Automatic Traffic Monitoring, VISATRAM, has been realized in a PC486-based image system. The frame grabber is the OFG-100 of Itex Technology, Co. and the resolutions of the frame image are 512x768 pixels. The VISATRAM can deal with two or three lanes simultaneously in the current implementation. More lanes can be considered if a faster computer is used. A single detection slice window covers all the lanes, therefore a single PVI is formed. On the other hand one EPI is formed for each lane. In practice the PVI and EPI are processed while they are being formed, so the system is operated at field rate while the captured images are active. Fig. 5 gives the system diagram. One example of the real-time operation is show in Fig. 6 where two lanes are processed.

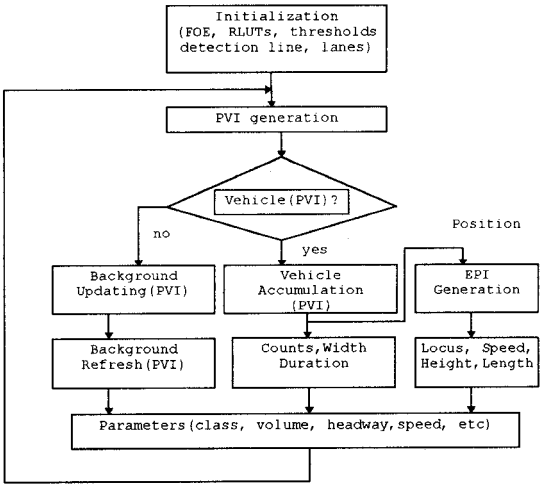


Fig. 5. System diagram

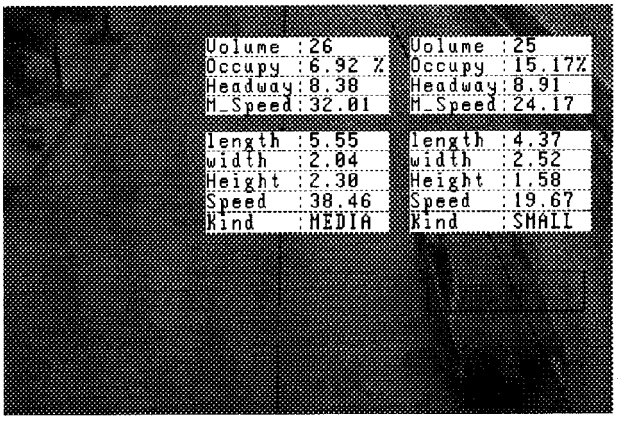


Fig. 6. An example of the real-time operation

4.2. Vehicle extraction

The extraction of the vehicle is performed by fusing multiple cues including intensity, gradients and models of vehicles and the environment. The vehicle is separated from the road surface, shadows and the vehicle lights in the following ways:

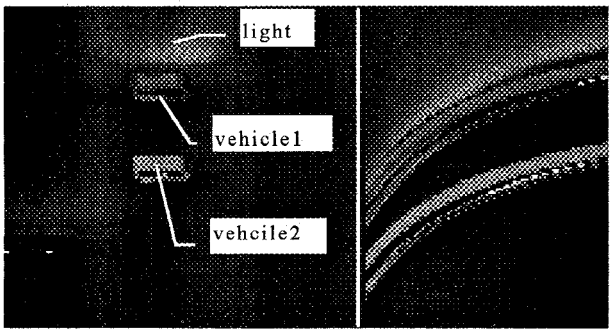
1) Background subtracting: There are intensity differences between vehicle and the road surface. In the daytime those portions whose intensity is higher than that of the road are directly classified as belonging to vehicle. Those portions with lower intensity need further analysis. The intensity of shadows is always lower than that of the road. In nighttime operation the intensity in the area corresponding to the vehicle's light projection is higher than that of the road surface, so further investigation is also needed.

2) ST differentiating This procedure is based on the fact that there are rich intensity changes inside the vehicle, especially along the longitude direction and around the boundary of the vehicle; while the intensities of shadow areas

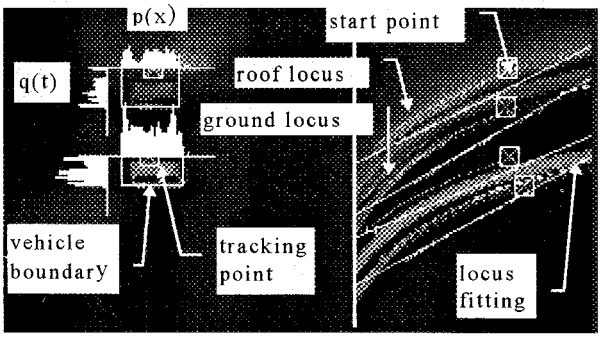
of a vehicle are almost constant, and the area where the vehicle's light projects has no distinctive edges.

3) Models of vehicles, shadows and lights: The symmetry and the length of the vehicle as well as the minimum headway between the vehicle are used as the model to group the different portions into a vehicle. The direction and the size of the shadow are estimated using the knowledge of the sun's position in different time of a day in a given place. The knowledge that the front light of the vehicle always projects in front of the vehicle can also be used in the segmentation of the vehicle.

4) Merging multiple lanes: To handle the situation when a vehicle is crossing over lanes or the shadow of a vehicle projects onto other lanes, the detection line covers all the related lanes and several lanes are considered together.



(1). Original PVI and EPI



(2). Results

Fig. 7. Vehicle separation and loci extraction

Considering the above principles, we design the basic algorithm of vehicle detection and separation. A single detection line f(x,t) is processed at each time t. The background is initialized as b(x) and is updated from time to time.

Step 1. *Background subtracting and ST differentiating*:

$$I(x,t) = f(x,t) - b(x) \tag{13}$$

$$d(x,t) = \left|\frac{\partial I(x,t)}{\partial x}\right| + \left|\frac{\partial I(x,t)}{\partial t}\right| \tag{14}$$

Step 2. *Vehicle detection*. For a given time t, the point whose gradient d(x,t) is greater than a given threshold Td is regarded as a possible vehicle's point:

$$e(x,t) = \begin{cases} 1 & d(x,t) > T_d \\ 0 & d(x,t) \le T_d \end{cases} \tag{15}$$

Edge is accumulated along the detection line

$$q(t) = \sum_x e(x,t) \tag{16}$$

and if q(t) is greater than a certain threshold Nw (the minimum vehicle's width measured in pixel number), a possible vehicle line is labeled.

Step 3. *Grouping and separation*. In the 2D PVI e(x,t), those labeled lines, which satisfy the criterion of minimum headway between the vehicles, duration time and width of a vehicle, are grouped into a possible vehicle. Thus the duration time (T) of the vehicle across the detection line is obtained.

Step 4. *Vehicle's width*. During the time period, we project the e(x,t) in the time axis as

$$p(x) = \sum_{t \in T} e(x,t) \tag{17}$$

The left and right boundaries of the vehicle in image are searched from both sides as x_L and x_R if p(xi) (i=L,R) are greater than N_L, the minimum vehicle's length measured in number of pixels. The width is calculated by eqn. (4b).

Fig. 7 shows an example of vehicle detection and separation during nighttime operations.

4.3. Background updating

To make the system adaptive to varying light conditions, the background should be updated from time to time. In our approach only a fraction of image (several scan lines for the PVI) needs to be updated. The background intensity of the PVI is initialized and then updated whenever the detection lines are definitely not covered by vehicles, shadows and vehicle's light. The background is updated in the following three cases:

(1). The background changes gradually. For the current time t, if q(t) < Nw, then the background is modified as

$$b_{new}(x) = \alpha b_{old}(x) + \beta I(x,t) \tag{18}$$

where α = q(t)]/Nw, β =[Nw - q(t)]/Nw.

(2). The background changes abruptly, for example, when a piece of cloud is passed over the road in the day, or the light is turned on in the evening. If vehicle's duration T is greater than the maximum duration for the largest vehicle, then an abrupt background change is assumed. In this case an alarm is given to show that the system enters a recovery period and the background is refreshed as:

$$b_{new}(x) = \frac{1}{T_s}\sum_{T_s} I(x,t) \tag{19}$$

where Ts is a certain time interval during the recovery period. Vehicles may pass the detection line while the background is re-initialized. So the background should be updated using eqn.(18) and the alarm is on until the detection turns to the normal condition.

During the recovery period and the previous period T, the passing vehicles may be missed by the VISTRAM. So the PVI section during those periods is saved and then reprocessed afterward using the refreshed background information.

(3). The background is changed frequently. This situation may occur when the energy-saving traffic lamp is used at night. The intensities of the background are changed periodically while the light flashes at a certain frequency, so the short headway and vehicle duration are detected frequently. In this case the PVI is smoothed using a 5x5 or large Gaussian operator according to the flashing frequency.

The gradient threshold Td is also changed according to the changes of the background, and the differences between vehicles and the road surface in different time of the day.

4.4. Loci extraction

For real-time implementation, we used re-projection LUTs to map the original epipolar lines to the rectified ones. When the vehicles move bottom-up in the image and the detection line is set at the bottom of the image, the lateral position of the epipolar line is selected adaptive to the position of the vehicle inside the lane. So we have several (e.g., 4) LUTs for each lane.

While the EPI is formed, the locus of the front and the rear of the vehicle is tracked. Theoretically, it is advantageous to rectify the EPI before we track the locus since the locus is straight after the rectification. However in practice, the rectification of the original EPI will degrade the resolution of the image. So we only reproject the locus points while track them in the original EPI, and straight line constraints can also be applied to guide the tracking. The tracking of the loci of vehicle's front and rear is relatively easy because they are the border edges of the loci's pattern of a vehicle. The energy function for the front or the rear loci tracking of a gradient image g(y,t) of the EPI is

$$E = \alpha\left|y - y^*\right| + \beta\left|g - g^*\right| + \gamma g + \kappa d \qquad (20)$$

where $\alpha, \beta, \gamma, \kappa$ are the normalized and weight coefficients for the measurements of locus's straightness, gradient similarity, gradient magnitude and "border-ness" for the point (y,t). In eqn. (20) y^* is the estimated value for y using the straight locus constraint in the rectified EPI, g^* is the average gradient value of the tracked points, d is the distance between the point (y,t) and the most outlying edge-like point. Initially the locus point is determined using the detecting result in the PVI (see Fig. 7). The locus is tracked by using a heuristic search method with the cost function expressed in eqn. (20). The values of $\alpha, \beta, \gamma, \kappa$ are changed during the tracking and in different light conditions. For example, at the beginning of a tracking, we set α=0, β = small value. The locus straightness and gradient similarity have larger weights when a certain number of tracking points has been obtained. In the nighttime operation the weight κ is set to a relatively small value to reduce the negative effect of the projection of the vehicle's light.

When enough points are obtained for both loci, Two straight lines are fitted to the rectified loci. The speed and height can be obtained using eqn.(3), and the length is estimated using eqn.(4a). Fig. 7 shown the loci tracking and fitting results of front and rear edges.

5 Conclusion

We have developed a visual traffic monitoring system VISATRAM based on 2D spatio-temporal image analysis. Experiment results of real images are encouraging as the common difficult situations such as daylight variation, shadows, nighttime operations are all tested. The system is also very cost-effective as the high performance is achieved on a simple 486-class PC.

Acknowledgments The authors would like to thank the reviewers for their insightful comments and suggestions. Thanks are also given to Dr. Chih-ho Yu for his helpful assistance on presentation of this material.

References

[1] Ferrier,N J, Roew, S M, Blake,A(1994), Realtime traffic monitoring, Proc. IEEE Workshop on Application of Computer Vision, 81-88

[2] Kilger, M.(1992) A shadow handler in a video-based real-time traffic monitoring system, IEEE Workshop on Application of Computer Vision, 11-18

[3] Michalopoulos, P (1991), Vehicle detection video through image processing: the AUTOSCOPE system, IEEE Trans. on Vehicular Techno., 40(1), 21.

[4] Nakanishi, T, Ishii,K(1992), Automatic vehicle image extraction based on spatio-temporal image analysis, Proc 11th ICPR, 500-504.

[5] Sal D'Agostino(1991) Commercial machine vision system for traffic monitoring and control, SPIE vol. 1615, 180-186

[6] Takatoo, M., et al(1989) Traffic flow measuring system using image processing, SPIE vol. 1197,172-180

[7] Zielke,.T. et al (1993) Intensity and edge-based symmetry detection with an application to car following, CVGIP: Image Understanding, vol. 58, no 2, 177-190