Moving Vehicle Segmentation from Plane Constraint

Dong-Joong Kang[1], Jong-Eun Ha[2], Jin-Young Kim[1], Min-Sung Kim[3], Tae-Jung Lho[1]

[1] Dept. of Mechatronics Eng., Tongmyong University of Information Technology, Busan, Korea
(Tel: +82-51-610-8361, {djkang,kjy97,tjlho}@tit.ac.kr)

[2] Dept. of Multimedia Eng., Tongmyong University of Information Technology, Busan, Korea
(Tel: +82-51-610-8416, jeha@tit.ac.kr)

[3] Dept. of Information & Communication Eng., Tongmyong University of Information Technology, Busan, Korea
(Tel: +82-51-610-8378, minsung@tit.ac.kr)

Abstract: We present a method to detect on-road vehicle using geometric invariant of feature points on side planes of the vehicle. The vehicles are assumed into *a set of planes* and the invariant from motion information of features on the plane segments the plane from the theory that a geometric invariant value defined by five points on a plane is preserved under a projective transform. Harris corners as a salient image point are used to give motion information with the normalized correlation centered at these points. We define a probabilistic criterion to test the similarity of invariant values between sequential frames. Experimental results using images of real road scenes are presented.

Keywords: Geometric invariant, plane constraint, motion, vehicle detection, active safety vehicle

1. INTRODUCTION

Intelligent on-road vehicles, guided by computer vision systems, are a main issue in developing experimental or commercial vehicles in numerous places in the world [1-8].

Reliable vehicle detection in images acquired by a moving vehicle is an important problem for the applications such as active safety vehicles (ASV) equipped with driver assistance system to avoid collision and dangerous accidents. Several factors including changing environmental condition affect on-road vehicle detection and the appearance changes of foregoing vehicles with scale, location, orientation, and pose transition makes the problem very challenging.

Foregoing vehicles are come into several views with different speeds and may always vary in shape, size, and color. Vehicle appearance depends on relative pose between observer and foregoing vehicles and occlusion by nearby objects affects the detection performance. In case of real implementation of intelligent road vehicle, real-time processing is another important issue.

We consider the problem of rear-view detection of foregoing vehicles from gray-scale images. Several previous researches assume two main steps to detect road vehicles [1]. The first step of any vehicle detection system is hypothesizing the locations in images where vehicles are present. Then, verification is applied to test the hypotheses. Both steps are equally important and challenging. Well known approaches to generate the locations of vehicles in images include using motion information, symmetry, shadows, and vertical/horizontal edges [5-8].

The purpose of this paper is to provide a method for the hypothetical candidates of on-road vehicle detection. Once the hypothetical regions including vehicles are extracted first, then several methods could be applied to verify the initial detection.

Detecting moving objects from images acquired by a static camera can be usually performed by simple image difference based methods. However, when the camera undergoes an arbitrary motion through a scene, the task is much more difficult since the scene is no longer static in the image. Simple image differencing techniques no longer apply. Road vehicle detection problem belongs to the second category because the observer camera is mounted on a moving vehicle. For general segmentation, optical flow from all image points or corresponding information from prominent image features can be used.

In this paper, we present a method based on the geometric invariant and motion information. Based on the sparsely obtained motion field or corresponding data, the method selects an initial segmentation cluster by using geometric invariant technique that can be described by a plane constraint. Moving vehicle segmentation is based on the fact that a geometric invariant value of point-set defined on a plane of the vehicle is preserved after motion of the plane [9][11-12]. Harris corner is a good image feature to provide motion information with the normalized correlation [13]. The probabilistic criterions to test the similarity of invariant values between frames and to merge neighboring regions are introduced without a need of magic factors for the threshold. The proposed method is more exact in initial segmentation than simple methods clustering similar velocity vectors because a side or rear part of a vehicle could be separately extracted under the *strong plane constraint*.

Among the vehicles surrounding my host vehicle, close-by front and rear, and overtaking side vehicles are more dangerous for collision and threaten car driver [1]. Methods detecting vehicles in these regions might be better to employ motion information because there are large intensity changes and detailed image features such as edges and corners by close view. Detecting vehicles in the far distance region is relatively easier since the full view of a vehicle is available and appearance is more stable.

2. INITIAL SEGMENTATION OF MOVING VEHICLES

2.1 Projective invariants

Projective invariants are quantities which do not change under projective transformations. Detailed contents of the uses of invariants are given in Mundy and Zisserman [9]. There are two convenient invariants that can be defined for groups of five points. Four points (no three collinear) form a projective basis for the plane and the invariants correspond to the two degrees of freedom of the projective position of the fifth point with respect to the first four - there exist positions that invariants do not change their values in some directions.

The two invariants may conveniently be written as the ratios of determinants of matrices of the form $\mathbf{M}_{ijk}$, which denotes area of a triangle consisting of three image points. Then the two invariants are given by [9][11] :

$$\mathbf{I}_1 = \frac{|\mathbf{M}_{124}||\mathbf{M}_{135}|}{|\mathbf{M}_{134}||\mathbf{M}_{125}|} \tag{1}$$

$$\mathbf{I}_2 = \frac{|\mathbf{M}_{241}||\mathbf{M}_{235}|}{|\mathbf{M}_{234}||\mathbf{M}_{215}|} \tag{2}$$

where $\mathbf{M}_{ijk} = (\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ and $\mathbf{x}_i$ is position (x_i, y_i) of an image point. These two quantities may be seen to be preserved under a projective transformation if $\mathbf{x}'$ is substituted for $\mathbf{x}$,

$$\frac{|\mathbf{M}'_{124}||\mathbf{M}'_{135}|}{|\mathbf{M}'_{134}||\mathbf{M}'_{125}|} = \frac{|\lambda_1\mathbf{Px}_1, \lambda_2\mathbf{Px}_2, \lambda_4\mathbf{Px}_4||\lambda_1\mathbf{Px}_1, \lambda_3\mathbf{Px}_3, \lambda_5\mathbf{Px}_5|}{|\lambda_1\mathbf{Px}_1, \lambda_3\mathbf{Px}_3, \lambda_4\mathbf{Px}_4||\lambda_1\mathbf{Px}_1, \lambda_2\mathbf{Px}_2, \lambda_5\mathbf{Px}_5|} \tag{3}$$

which gives,

$$\frac{|\mathbf{M}'_{124}||\mathbf{M}'_{135}|}{|\mathbf{M}'_{134}||\mathbf{M}'_{125}|} = \frac{\lambda_1^2\lambda_2\lambda_3\lambda_4\lambda_5|\mathbf{P}|^2|\mathbf{M}_{124}||\mathbf{M}_{135}|}{\lambda_1^2\lambda_2\lambda_3\lambda_4\lambda_5|\mathbf{P}|^2|\mathbf{M}_{134}||\mathbf{M}_{125}|}. \tag{4}$$

where $\mathbf{P}$ is the projectivity matrix and λ_i is scaling factor.

2.2 Extract points groups from motion data

Point features corresponding to high curvature points are extracted from image before motion. A salient image feature should be consistently extracted for different views of object and there should be enough information in the neighborhood of the feature points so that corresponding points can be automatically matched. We use the Harris corner detector [13]:

$$\mathbf{R}(x, y) = \det[\mathbf{C}] - k \cdot trace^2[\mathbf{C}] \tag{5}$$

where $\mathbf{C}$ is

$$\mathbf{C} = w \cdot \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix}. \tag{6}$$

The notation g denotes gray scale image and w is the Gaussian smoothing operator, k is a parameter usually set to 0.04. And g_x and g_y indicates the x and y directional derivative for the grey image, respectively. Corners are defined as local maxima of the corner response function $\mathbf{R}$.

Given the high curvature points, we can use a correlation window of small size $(2r+1)\times(2c+1)$ centered at these points. We define a rectangle search area of size $(2d_x+1)\times(2d_y+1)$ around the points in the next image, and perform a correlation operation on a given window between a point in the first image and pixels within a search area in the second image. This operation provides the matched vectors.

We select 5 points from n correspondences to define invariant values. There are $N={}_nC_5$ independent ways of choosing five points from n. Therefore, it is impractical to test all possible combinations of five points in an image. Instead of, groups of five points are selected as four nearest neighbors outside a small circular neighborhood of the fifth point and inside of a larger circle.

All selected 5-point sets define invariant values. A set before motion defines a model invariant value and the corresponding set after motion defines the corresponding invariant value. If this pair exists on the same plane or on an object moving under weakly perspective projection assuming far distance from the observer, the difference of two invariant values will be small. This constraint makes the initial segmentation possible. Pairs of points giving a similar invariant value between after and before motion are considered as independently moving objects. We introduce a threshold to test the similarity of two invariant values.

$$|\mathbf{I}'_i - \mathbf{I}_i| < Thres_i \tag{7}$$

For selecting thresholds, we could use a measurement uncertainty associated with the estimated position of corner features [10-12].

The invariant is a function of five points:

$$\mathbf{I} = \mathbf{I}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5). \tag{8}$$

Let $\mathbf{x}_i$ be the true and $\tilde{\mathbf{x}}_i$ be the noisy observation of (x_i, y_i), then we have

$$\tilde{x}_i = x_i + \xi_i \tag{9a}$$

$$\tilde{y}_i = y_i + \eta_i \tag{9b}$$

where the noise terms ξ_i and η_i denote independently distributed noise terms having mean 0 and variance σ_i^2. From these noisy measurements, we define the noisy invariant,

$$\tilde{\mathbf{I}}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \tilde{\mathbf{x}}_3, \tilde{\mathbf{x}}_4, \tilde{\mathbf{x}}_5) \tag{10}$$

To determine the expected value and variance of $\tilde{\mathbf{I}}$, we expand $\mathbf{I}$ as a Taylor series at $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5)$:

$$\begin{aligned} \mathbf{I} &\approx \tilde{\mathbf{I}} + \sum_{i=1}^{5}\left[(\tilde{x}_i - x_i)\frac{\partial\tilde{\mathbf{I}}}{\partial\tilde{x}_i} + (\tilde{y}_i - y_i)\frac{\partial\tilde{\mathbf{I}}}{\partial\tilde{y}_i}\right] \\ &= \tilde{\mathbf{I}} + \sum_{i=1}^{5}\left[\xi_i\frac{\partial\tilde{\mathbf{I}}}{\partial\tilde{x}_i} + \eta_i\frac{\partial\tilde{\mathbf{I}}}{\partial\tilde{y}_i}\right] \end{aligned} \tag{11}$$

Then, the variance becomes

$$E\left[(\tilde{\mathbf{I}} - \mathbf{I})^2\right] = \sigma_0^2\sum_{i=1}^{5}\left[\left(\frac{\partial\tilde{\mathbf{I}}}{\partial\tilde{x}_i}\right)^2 + \left(\frac{\partial\tilde{\mathbf{I}}}{\partial\tilde{y}_i}\right)^2\right]. \tag{12}$$

Hence, for a given invariant $\mathbf{I}$, we can determine a threshold:

$$\Delta\mathbf{I} = 3 \cdot \sqrt{E[(\tilde{\mathbf{I}} - \mathbf{I})^2]}. \tag{13}$$

The partial derivative $\partial\mathbf{I}_1 / \partial x_1$, for example, is given by

$$\frac{\partial\mathbf{I}_1}{\partial x_1} = \mathbf{I}_1\left(\frac{y_2 - y_4}{|\mathbf{M}_{124}|} + \frac{y_3 - y_5}{|\mathbf{M}_{135}|} + \frac{y_3 - y_4}{|\mathbf{M}_{134}|} + \frac{y_2 - y_5}{|\mathbf{M}_{125}|}\right) \tag{14}$$

If there are several point groups smaller than the threshold values in the band region, a candidate with the minimum difference is selected among the combination candidates.

4. MERGING AND EXPANDING SEGMENTED REGIONS

Section 2 provides the small regions that include five points defined under the plane constraint. We can extract a maximum boundary region (MBR) for each point-set. Because initial clusters by the invariant test could be concentrated on a vehicle side that is approximately planar, many MBR are overlapped at similar positions and neighboring MBR having a similar mean motion should be merged together. We use the student t-test to combine the near MBR [14]. A criterion merging two MBR having mean $\bar{t}(i)$ and $\bar{t}(j)$ follows as:

$$d_x(i,j) = \frac{\bar{t}_x(i) - \bar{t}_x(j)}{\sqrt{\frac{\sigma_x^2(i)}{n_i} + \frac{\sigma_x^2(j)}{n_j}}} \quad (15)$$

where σ_x is the variance of $\bar{t}_x$. $d_y(i,j)$ is defined analogously. Threshold values for $d_x(i,j)$ and $d_y(i,j)$ are determined from the t-distribution table. MBR formed by two or more regions become the motion candidates of independently moving objects and their motion parameters are updated by using the inside points of merged regions. After the step merging overlapped regions, each MBR expands to include neighboring motion vectors. Smith [5] proposed a simple method to compare a motion vector u_i with mean vector $\bar{t}$ that is estimated by the current MBR:

$$D_x = \frac{|u_i - \bar{t}_x|}{\frac{|u_i| + |\bar{t}_x|}{2} + \sigma_x} \quad (16)$$

where σ_x is the variance of the flow of a candidate region. The distance D_y is given analogously. Two steps of eq. (15) and (16) repeat until there is no merging region. Finally, the remained MBR are tested to verify whether the regions include or not a vehicle by using the ratio of horizontal and vertical length of each MBR.

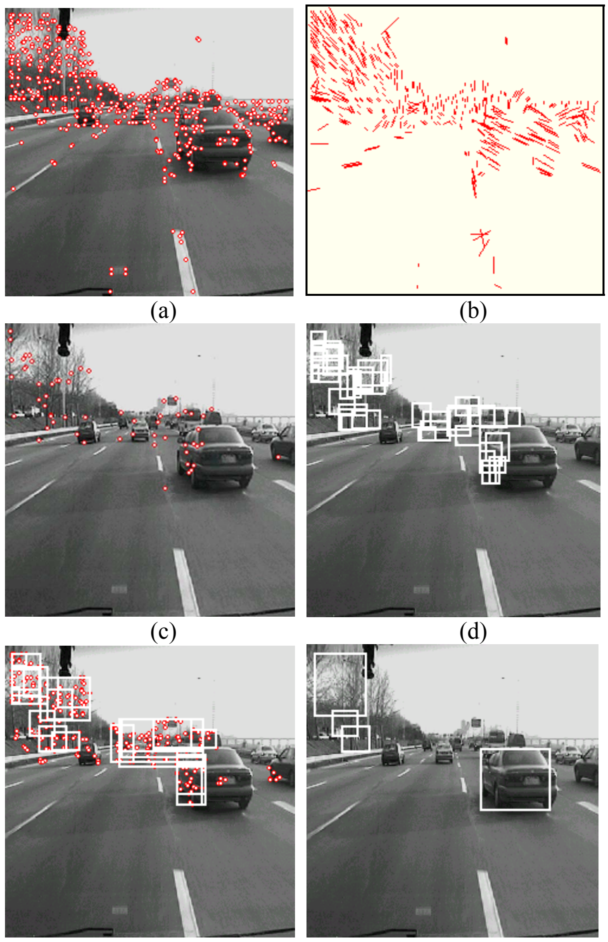
4. EXPERIMENTS

Experiments show that foregoing close-by vehicles are well detected. Fig. 1 shows a detection example for a real road scene. Corners on the rear part of a vehicle are extracted by Harris corner algorithm as shown in Fig. 1(a). We set $r_1 = 5$ and $r_2 = 30$ pixels at the circular band of Fig. 1 for calculation of the geometric invariant values.

Fig. 1(b) shows motion vectors from the normalized correlation employed to Harris corners of Fig. 1(a) between two sequential frames. The correlation is performed with small size window of 9x9 pixels for search windows of 11x11 pixels in next frame. Points giving a small invariant difference between frames are showed as the points in Fig. 1(c), and Fig. 1(d) shows each MBR from five point sets. Position variance σ_o in eq. (12) sets to 0.2. As the center points of detected five point sets are appeared on side of vehicle, the side of a vehicle is recognized as an approximate planar object with the invariant values preserved during moving of the vehicle.

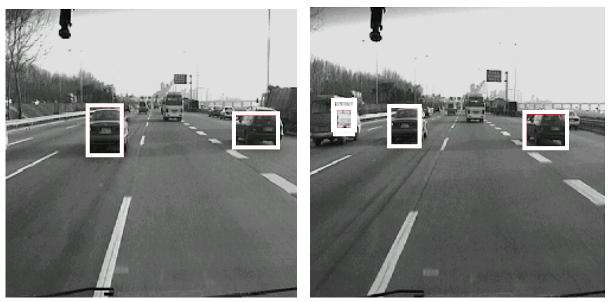
These points become seeds for the following segmentation steps. During the merging and expanding steps of Section 3, each MBR expands and then is merged with neighboring regions having similar motion. Fig. 1(e) and 1(f) present the MBR combined by the region merging test and the final result after a few repeat of Section 3 steps, respectively. Threshold values for D_x and D_y are set to 0.15 [5]. After merging and expanding process of Section 3, final MBR was tested to verify the vehicle from the ratio of horizontal and vertical length of each MBR.

Fig. 2 shows the results for a few more frames. Overtaking vehicles are detected by starting from side part of the vehicles. The computing time of the method is about 0.5 sec for 256x256 pixels image.



(a) (b) (c) (d) (e) (f)

Fig. 1. Invariant based segmentation for a real road scene. (a) Harris corner points; (b) Motion information with the normalized correlation; (c) The points with small changes of invariant values; (d) Initial MBR; (e) MBR combined by the region merging test; (f) Final result



(a) (b)

Fig. 2. Experiments for a few following frames. A new appearing vehicle is detected by starting from side part of the vehicle

5. CONCLUSIONS

We present an invariant and motion based methods for vision-based vehicle detection. The segmentation method uses the geometric invariant with the correspondence information obtained from corner points of sequential frames. The proposed technique assumes the 3-D moving vehicle into a set of planar surfaces that come from front and rear, and side surface parts of a vehicle. The method is more exact in initial segmentation than a similar velocity field clustering method because a side or rear part of a vehicle could be independently extracted under strong plane constraint.

For two consecutive images, a corner extraction algorithm detects prominent corner features in the first image. Correspondence information of detected corner points is then found by the normalized correlation method for a next sequential image. Based on the sparsely obtained motion data, the segmentation algorithm selects points one a plane that maintain consistent geometric invariants between frames. These points set form initial clusters to segment the plane and the clusters merge the neighboring points having a similar motion. The process repeats and the regions expand according to the probabilistic criterions. Finally, the merged clusters become candidates of independently moving vehicles. Through the experiments of real road scenes, the proposed method presents a strict segmentation of fast moving and overtaking vehicles is possible.

REFERENCES

[1] Sun, Z., Bebis G., Miller R.: On-road vehicle detection using optical sensors: A review, *IEEE Intelligent Transportation Systems Conference*, Washington D.C., pp.585-590, 2004

[2] A. Polk and R. Jain, I.Masaki: Vision-based Vehicle Guidance, *Springer-Verlag*, 1992

[3] Kang, D.J., Choi, J.W., Kweon, I.S.: Finding and Tracking Road Lanes using Line Snakes, *IEEE Intelligent Vehicles Symposium*, 1996

[4] Kang, D.J., Jeong M.H.: Road lane segmentation using dynamic programming for active safety vehicles, *Pattern Recognition Letters*, Vol.24 pp.3177-3185, 2003

[5] Smith, S.M.: ASSET-2 Real-Time Motion Segmentation and Object Tracking, *Int. Conf. on Computer Vision*, pp.237-244, 1995

[6] Kuehnle, A.: Symmetry-based recognition for vehicle rears, *Pattern Recognition Letters*, Vol.12, pp.249-258, 1991

[7] Handmann, U., Kalinke, T., Tzomakas, C., Werner, M., Seelen, W.: An image processing system for driver assistance, *Image and Vision Computing*, Vol. 18, pp.367-376, 2000

[8] Betke, M., Haritaglu, E., Davis, L.: Multiple vehicle detection and tracking in hard real time, *IEEE Intelligent Vehicles Symposium*, pp.351-356, 1996

[9] Joseph, L., Mundy, Zisserman, A.: Geometric Invariance in Computer Vision, *MIT Press*, 1992

[10] Kang, D.J., Ha, J.E., Kweon, I.S.: Fast Object Recognition using Dynamic Programming from Combination of Salient Line Groups. *Pattern Recognition*, Vol.36, pp.79-90, 2003

[11] Roh, K.S., Kweon, I.S.: 2-D Object Recognition Using Invariant Contour Descriptor and Projective Refinement, *Pattern Recognition*, Vol. 31, pp.441-455, 1988

[12] Sinclair, D., Blake A.: Quantitative Planar Region Detection, *Int. Journal of Computer Vision*, Vol.18, pp.77-91, 1996

[13] Harris, C., Stephens, M.: A combined corner and edge detector, *The Fourth Alvey Vision Conference*, pp.147-151, 1998

[14] Press, W., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P.: Numerical Recipes in C, *Cambridge University Express*, 1992