MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

3-phase Vehicle Tracking using Image Alignment and Haar Transform

Lap-Chi, Cheung
Dept. of Computer Science & Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
lccheung@cse.cuhk.edu.hk

Yiu-Sang, Moon
Dept. of Computer Science & Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
ysmoon@cse.cuhk.edu.hk

Abstract

The large number of rear end collisions due to driver inattention has been identified as a major automotive safety issue. In this paper, we describe a 3-phase vehicle tracking methodology with a single moving camera mounted on the driver's automobile as input for use in detecting rear vehicles on highways and city streets. In the first phase, vehicle detection, a rear vehicle is detected by symmetrical measurement and Haar transform. A template is created simultaneously which is used for object tracking employing image alignment techniques in the second phase. The template is continuously monitored to handle abrupt changes in surrounding conditions and is updated when necessary in the third phase. We also present a method to compute the relative distances between the rear vehicles and the driver's car. Initial experiments demonstrate a successful rate over 90%.

1. Introduction

We have developed a real-time rear vehicle detection system, which can give an advance warning to a driver when the relative distance between a rear vehicle and the driver's car is too small, for the purpose of prevention of collision. On the highway, many *active* obstacle detection methods such as laser and radar cannot be possibly employed due to low spatial resolution data and interference among vehicles [1]. Therefore, it is worthwhile to investigate completely passive, vision based approaches for vehicle detection.

In phase 1 of our system, vehicles in an image are first located by applying the symmetrical function to the image. Then, features like vehicle roof, vehicle bottom and vehicle sides are identified for each vehicle using techniques like horizontal edge response, horizontal pixel intensity response and over-complete Haar transform so that we are able to enclose the vehicle with a bounding polygon. The polygon is known as the template. Instead of simply detecting vehicles in each frame, a methodology is used to track the changing polygon whereas in traditional vehicle detection techniques an image region is simply treated as some moving "stuff" [2] and hence cannot be used to track the changes in the polygons associated with each vehicle according to a viewing camera. Moreover, information about these changes also give us he clue, for computing the relative distance between the vehicle and the camera, and thus, are able to generate appropriate driving advice or advance warning. In phase 3, we check for significant changes for the vehicles templates appearing in consecutive images. This task is necessary due to the potential variability in the appearances of the vehicles over time. Such variability arises from two principle sources: variation in vehicle pose or deformations and variation in illumination.

The rest of the paper is organized as follows. Section 2 elaborates the vehicle detection method and template generation processes in details. The vehicle template tracking using inverse compositional image alignment is described in Section 3. Then, the principle of vehicle template update is introduced in Section 4. In Section 5, the technique of finding the relative distance from the target vehicle is introduced. Finally, experiments and conclusion are discussed in Section 6 and Section 7 respectively.

2. Vehicle Detection

Vehicle detection is the first phase of the 3-phase vehicle tracking methodology. Our first task is to detect and locate the candidate vehicle in an image and then pass the segmented part of the image as a template to the vehicle template tracking phase. Vehicle detection, also interpreted as the vehicle template segmentation, aims to segment the vehicle part from the whole image. A review of on-road vehicle detection using optical sensors is given by Sun, Bebis and Miller [1] in which vehicle detection is classified into three categories: stereo-vision, motion, and knowledge based. In general, stereo-vision based methods are accurate if the stereo parameters have been estimated accurately, which is really hard to guarantee for the on-road scenario. Optical flow calculation of vehicles is the representative method in the motion-based category. However, the lack of vehicle texture and the existence of vibration cause poor performance for this kind of method. Knowledge-based methods employ a-priori knowledge of such vehicle features like color, shadow, corners, horizontal/vertical edges, texture, and vehicle lights for their detection. Among these features, locating the roofs of vehicles by horizontal edge detection and locating the bottoms using shadow information as sign patterns of the vehicles are popular techniques since these features, which are often quite uniquely distinguished from the background objects, can often be identified easily. However, it is not practical to search the roof and bottom of vehicles from the entire image. We propose using symmetrical measurement to reduce the searching area of the roof, bottom and sides of vehicles. Locating the vehicle center – Symmetrical Measurement

By minimizing the searching area of image, the computation time of finding the interesting objects could be decreased. In our proposal, the vehicle is first located by applying a symmetrical function to the image. To find a reliable measure of the degree of symmetry, we begin with an elementary theory of one dimensional function [3], namely, a function $f(x)$ can be divided in two parts: a purely symmetric function $f_s(x)$ and a purely asymmetric function $f_{as}(x)$. that is

$$f(x) = f_s(x) + f_{as}(x) \quad \begin{cases} \text{with } f_s(-x) = f_s(x) \\ \text{and } f_{as}(-x) = -f_{as}(x) \end{cases}$$

Now, a measure of the degree of symmetry $S\{f(x)\}$ is introduced. Apparently, we can compare the contribution of the symmetrical part of the function with the entire contribution of the function, as follows:

$$S\{f(x)\} = \frac{\|f_s(x)\|^2}{\|f(x)\|^2} = \frac{\|f_s(x)\|^2}{\|f_s(x)\|^2 + \|f_{as}(x)\|^2} \qquad \begin{aligned} &S\{f(x)\} \in [0,1] \\ &S\{f(x)\} = 1 \text{ if } f(x) \text{ is purely symmetric} \\ &S\{f(x)\} = 0 \text{ if } f(x) \text{ is purely antisymmetric} \end{aligned}$$

As the camera and target vehicles move, the illumination variation may cause the target vehicles to appear unlike a symmetrical object in grayscale intensity. To reduce this kind of influence, gradient images are input for symmetrical measurement. In order to highlight the measurements surrounding the vehicles, the symmetrical measurement around vehicles, are further enhanced by using the gradient images with horizontal edges only as input for vehicle detection. A result is shown in Figure 1 in which the center of the vehicle is marked by a dotted line.

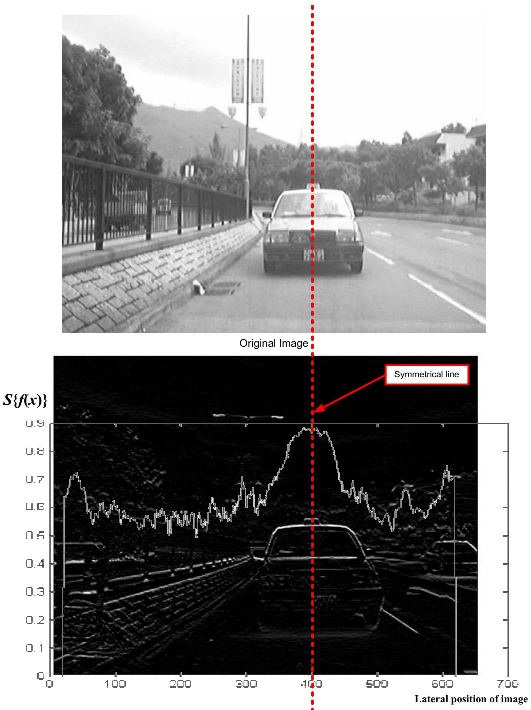

Figure 1. Symmetrical measurement using gradient image with horizontal edges as input

2.1. Locating vehicle roof and bottom

After the region of interest has been located in the image by the symmetrical measurement, the vehicle roof and the vehicle bottom can be located in a relatively small searching window instead of searching in the whole image. The vehicle roof is located at the first horizontal position searching from the top to the bottom with high horizontal edge response. On the other hand, the vehicle bottom is located at the first horizontal position searching from the bottom to the top with low horizontal pixel intensity response.

2.2. Locating the vehicle sides – Over-complete Haar Transform

The remaining task of vehicle segmentation is to locate the vehicles' sides. The moving and variant-illumination conditions make the ordinary pixel difference method unsuitable to represent the sides of vehicles. As a result, we use over-complete Haar wavelet transform, proposed by Papageorgiou [4], for vehicle learning in which there is a large set of features responding to local intensity differences. In [4], a complete vehicle images is transformed by three oriented Haar wavelets – vertical, horizontal and diagonal for classifier training.

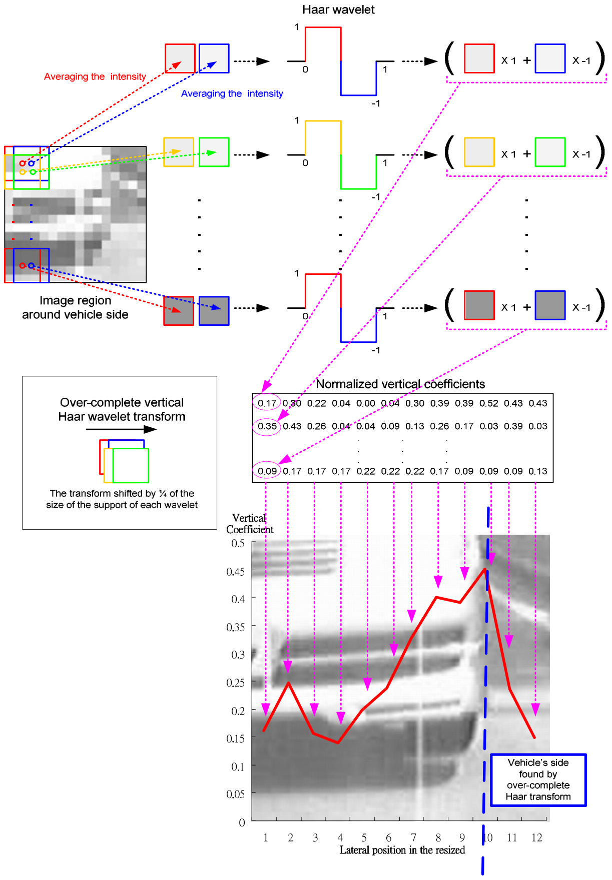

Figure 2. The over-complete Haar transform

In our work, only the vertical over-complete Haar wavelet transform, which gives high response to the sides of vehicles, is used since there is no training process required and the wavelet transforms are computed around the estimated locations of the vehicles' sides instead of the complete image of the vehicle. Differing from the original Haar wavelet transform, [4] also suggested shifting the transform by 1/4 of the size of the support of each wavelet, yielding an over-complete (quadruple density) dictionary of wavelet features. The framework of the over-complete Haar transform is shown in Figure 2. The vertical coefficients are detailed. We propose using the resulting high dimensional vertical feature table to locate the sides of vehicles.

By summarizing and averaging the coefficients of each column, the lateral position with the highest vertical coefficient is located as the vehicle's side. The high-resolution vertical coefficients of the over-complete Haar transform along the lateral direction enable the vehicle sides to be located precisely. By the above techniques, the vehicle is detected in the image and the bounding region, called vehicle template, is passed to the vehicle template tracking phase.

3. Vehicle Template Tracking – Image Alignment

We propose to use image alignment as the technique of object tracking. Image alignment consists of moving deforming (aligning) a template to minimize the differences among the templates (*T*) in consecutive images (*I*). The applications of image alignment include tracking, layered motion estimation, mosaic construction and medical image registration. We will follow the unifying framework proposed and described in the Lucas-Kanade algorithm [5].

The newly aligned (tracking) position and the aligned (tracking) bounding region are usually called the warped position (*W*) and the warped image respectively. The inverse compositional image alignment equation is used to describe the difference between the template and the warped image

$$\sum_x \left[T(W(x;\Delta p)) - I(W(x;p))\right]^2$$

where p are the parameters describing the warped position). The image tracking process then:

1) Aligns the warped position and update the parameters of the warped position; perform a first order Taylor expression of the image alignment equation and minimize its value by a least-squares method.
2) Minimize the image alignment equation by iterating and updating the warped parameters *p*.

For each iteration, the warped parameters are updated by:

$$\Delta p = H^{-1}\sum_x \left[\nabla T\frac{\partial W}{\partial p}\right]^T \left[I(W(x;p)) - T(x)\right]$$

The steps involved for the inverse compositional image alignment are shown as followings:

Pre-compute:
(1) Evaluate the gradient ∇T of the template T(x)
(2) Evaluate the Jacobian $\frac{\partial W}{\partial p}$ at (x;0)
(3) Compute the steepest descent images $\nabla T\frac{\partial W}{\partial p}$
(4) Compute the Hessian matrix $H = \sum_x \left[\nabla T\frac{\partial W}{\partial p}\right]^T \left[\nabla T\frac{\partial W}{\partial p}\right]$

Iterate:
(5) Warp *I* with **W**(**x**; **p**) to compute **I**(**W**(**x**; **p**))
(6) Compute the error image **I**(**W**(**x**; **p**)) – **T**(**x**)
(7) Compute the steepest descent parameter $\sum_x \left[\nabla T\frac{\partial W}{\partial p}\right]^T \left[I(W(x;p)) - T(x)\right]$
(8) Compute Δp
(9) Update the warp $\mathbf{W}(\mathbf{x};\mathbf{p}) \leftarrow \mathbf{W}(\mathbf{x};\mathbf{p}) \circ \mathbf{W}(\mathbf{x};\Delta\mathbf{p})^{-1}$
until $\|\Delta p\| \le \varepsilon$ or after a constant number of iterations. Typically, it is set to 10. (The corresponding steps are shown in Figure 3)

In this section, we have shown that how vehicle tracking is achieved by image alignment technique in an image sequence. By tracking the rear vehicles, with the vehicle height, an advance warning to a driver can be given to a driver when the relative distance computed by the newly proposed formula introduced in Section 5 is below a certain threshold.

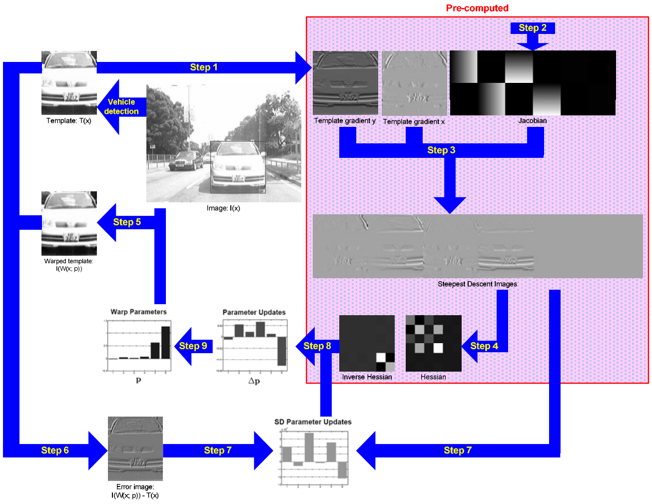

Figure 3. The framework of the inverse compositional algorithm

4. Vehicle Template Update

Variation in vehicle pose or deformations and variation in illumination can cause the tracking algorithm to fail. They can lead to the necessity of updating the template in order to let the tracking process remain effective. Updating the template for tracking is simple by “Making the new template fitting the vehicle in the current image”. The procedure is similar to vehicle detection but the searching is limited to the image area around the currently warped image. The update is composed of updating the new positions of vehicle symmetrical line, roof and bottom.

5. Finding the relative distance from the target vehicle

The distance formula was first introduced by Mobileye [6] company using the law of perspective with the height of vehicles obtained in tracking:

$$Z = \frac{fH}{y}$$

f: the focal length of camera in pixel
H: the height of the camera from the ground
y: the height of vehicles in pixel
Z: camera-to-vehicle distance

This formula has the assumption that the roof of the vehicle is at the same horizontal level of the camera. However, different types of vehicles can invalidate this assumption. Therefore, we propose a new version of formula:

$$Z = \frac{fH}{y}\left(1 + \frac{y_k}{y_h}\right)$$

y_k: the number of pixel from the horizontal line to the roof of the vehicle
y_h: the number of pixel from the horizontal line to the bottom of the vehicle

The formula is based on the fact that the ratio of the part from the horizontal line to the top of the vehicle to the part from the horizontal line to the bottom of the vehicle to remain invariant in the 2D image and the 3D world. The comparison between the camera-to-vehicle distance computed by the formula proposed by Mobileye and the new formula proposed is shown in Figure 4. The camera-to-vehicle distance computed by the Mobileye formula is shorter than the actual distance, leading to possibility of untimely warning.

Distance computed by the newly proposed formula	Distance computed by Mobileye formula
	

Figure 4. The camera-to-vehicle distance computed by the newly proposed formula

6. Experiment

We have conducted an experiment to analyze the reliability of the proposed 3-phase vehicle tracking methodology. There are 4 image sequences of the road traffic taken by a camera mounted on the rear side of our car. All of the sequences were taken during daytime on highways or city streets with durations ranging from 4 seconds to 50 seconds. The camera setting is shown in Table 1.

Table 1. Setting of camera mounted on our car

Camera Model:	Logitech webcam 4000
Frame rate:	15 frames / second
Frame size:	320 x 240 pixels

By observing the number of frame when tracking became discontinuous, the reliability of our methodology can be analyzed. Figure 5 shows some of the tracking results with camera-to-vehicle distance which is indicated by the numbers on the top of images and there are multiple vehicles following and overtaking our car. Results show that the 3-phase methodology is able to track the vehicles over 90% of frames. Failures occur when vehicles are running on winding roads and when the appearance of the target vehicles deforms frequently due to their unstable movement.

7. Conclusion

In this paper, we have described a 3-phase vehicle tracking methodology for detecting vehicles in image sequences captured by a single camera mounted on driver's car with distance information in order to prevent automobile collisions. Our methodology differs from most previous methods which simply detect vehicles in each frame without correlation between frames. The 3-phase structure provides valuable information to accommodate changes in the viewpoint and position of the vehicles and allow thus the distances from the target vehicles to our vehicle be computed using a newly proposed distance formula.

Basically, our 3-phase vehicle tracking methodology is able to track the rear vehicles. Further enhancement can be made to reduce the failure rate, especially when the vehicles are traveling some winding roads. Ultimately, we hope that our system can help to alert a driver for vehicles in his blind spot when he attempts to change lane while driving.

8. Acknowledgments

The research described in this paper was partially supported by the Hong Kong Research Grants Council CERG Project 415205.

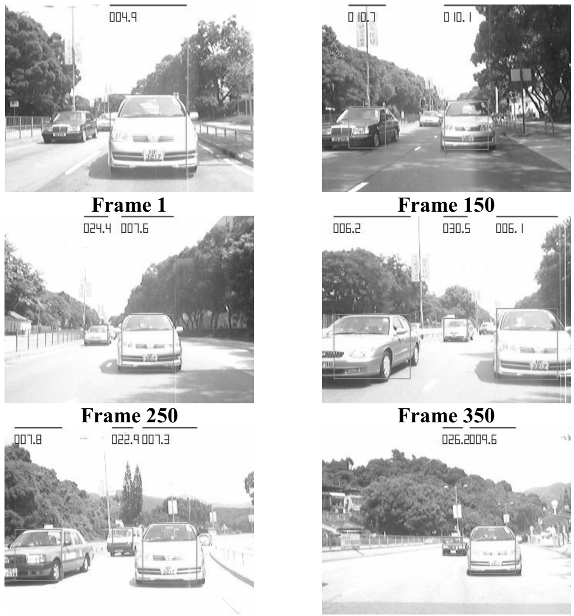

Figure 5. A sequence of vehicle tracking images with distance information

References

[1] Z. Sun, G. Bebis and R. Miller: "On-Road Vehicle Detection Using Optical Sensors: A Review", *IEEE Intelligent Transportation Systems Conference*, Oct 2004.

[2] E.H. Adelson and J.R. Bergen: "The plenoptic function and the elements of early vision", In Computational Models of Visual Processing, M. Landy and J.A. Movshon, MIT Press: Boston, MA, pp. 1-20.

[3] N. Kiryati and Y. Gofman: "Detecting Symmetry in Grey Level Images: The Global Optimization Approach", International Journal of Computer Vision, Vol. 29 No.1, pp 29-45, 1998.

[4] C. P. Papageorgiou, T. Poggio: "A Trainable Object Detection System: Car Detection in Static Images", MIT AI Memo No. 180. October, 1999.

[5] S. Baker and I. Matthews. Lucas-Kanade 20 Years On: A unifying Framework. *International Journal of Computer Vision*, Vol. 56, No. 3, March, 2004, pp. 221 – 255.

[6] G.P. Stein, O. Mano and A. Shashua: "Vision-based ACC with a Single Camera: Bounds on Range and Range Rate Accuracy", IEEE Intelligent Vehicles Symposium, Jun. 2003, Columbus, OH.