J. Vis. Commun. Image R. 46 (2017) 250–259

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads ☆

Deng-Yuan Huang [a], Chao-Ho Chen [b,*], Tsong-Yi Chen [b], Wu-Chih Hu [c], Kai-Wei Feng [b]

[a] Department of Electrical Engineering, Da-Yeh University, 168 University Rd., Dacun, Changhua 515, Taiwan, ROC
[b] Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, 415 Chien Kung Rd., Kaohsiung 807, Taiwan, ROC
[c] Department of Computer Science and Information Engineering, National Penghu University of Science and Technology, 300 Liu-Ho Rd., Makung, Penghu 880, Taiwan, ROC

ARTICLE INFO

Article history:
Received 30 December 2016
Accepted 22 April 2017
Available online 26 April 2017

Keywords:
Vehicle detection
Inter-vehicle distance estimation
Background subtraction
Histogram of oriented gradient (HOG)
Support vector machine (SVM)

ABSTRACT

This paper presents a driver assistance system for vehicle detection and inter-vehicle distance estimation using a single-lens video camera on urban/suburb roads. The task of vehicle detection on urban/suburb roads is more challenging due to their high scene complexity. In this work, the still area of frame inside the host vehicle is first removed using temporal differencing, followed by detecting vanishing point. Segmentation of road regions is then conducted using vanishing point and road's edge lines. Shadow regions at the bottoms of vehicles verified using the HOG feature and an SVM classifier are utilized to detect vehicle positions. The distances between the host and its front vehicles are estimated based on the locations of detected vehicles and vanishing point. Experimental results show varied performance of vehicle detection with different scenes of urban/suburb roads and the detection rate can achieve up to 94.08%, indicating the feasibility of the proposed method.

1. Introduction

Traffic safety has become an increasing concern in urban and suburban areas due to the increasing amount of vehicles. Rear-end collisions account for the majority of traffic accidents on highways and urban/suburb roads. Hence, developing an automotive driver assistance system (ADAS) to reduce such accidents is urgent and important. Intelligent ADAS based on a video camera recorder has gradually become standard equipment. Common smart functions include lane departure warning, pedestrian detection, and forward collision warning (FCW). Most vision-based FCW systems have a higher vehicle detection rate on highways, compared to that on urban/suburban roads, due to the former's simpler road structures and backgrounds. Vehicle detection on urban/suburb roads is more challenging because of their variable backgrounds and scene complexity, which inevitably affect the performance of moving object detection [1–3].

In general, FCW systems can be broadly divided into three categories, namely active, passive and hybrid, depending on whether they use active sensors, such as Radar, Lidar, or GPS, to sense environmental conditions. Although estimation of inter-vehicle distance using active sensors is more accurate than that of using the vision-based approaches, it is usually high-cost. In recent years, the vision-based vehicle detection without the aid of active sensors has increased in popularity due to the limitations of active sensors in detecting multiple vehicles [4]. As reported in many works [5–8], the accuracy of vision-based vehicle detection is greatly affected by road structures and scene complexity, and in general, a higher detection rate is achieved in highway scenes due to their relatively simple road structures and backgrounds. However, few investigations have been carried out using scenes of urban/suburb roads. This motivates that the present work is aimed at efficient vehicle detection and inter-vehicle distance estimation for such scenes.

The rest of this paper is organized as follows. Related work is briefly reviewed in Section 2 and the proposed method for vehicle detection and inter-vehicle distance estimation is described in Section 3. Section 4 provides the experimental results for demonstrating the performance of the proposed method and finally the concluding remarks are given in Section 5.

2. Related works

Vision-based monocular vehicle detection approaches can be broadly divided into two categories: appearance-based and

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail addresses: kevin@mail.dyu.edu.tw (D.-Y. Huang), thouho@cc.kuas.edu.tw (C.-H. Chen), chentso@cc.kuas.edu.tw (T.-Y. Chen), wchu@npu.edu.tw (W.-C. Hu), 1100305149@cc.kuas.edu.tw (K.-W. Feng).

http://dx.doi.org/10.1016/j.jvcir.2017.04.006
1047-3203/

motion-based methods [4]. In general, appearance-based methods are more popular than motion-based methods for monocular vehicle detection because they recognize vehicles by relying only on features extracted directly from images, whereas motion-based methods depend on the context from a sequence of images for vehicle detection. For appearance-based methods, various appearance features, including Haar-like features [7–10], histogram of oriented gradient (HOG) features [11–14], a combination of Haar-like and HOG features [15], and Gabor features [16,17], are extracted from images and used for vehicle detection. Moreover, a dimension reduction of the feature space can be applied using a combination of principal component analysis and independent component analysis for vehicle detection in static images [18].

Most methods reported in [19–26] for front vehicle detection and distance estimation are aimed at a single vehicle and involve vanishing point detection, road detection, and vehicle segmentation. Recently, several vision-based vehicle segmentation methods are focused on structured roads (e.g., highways, expressways, freeways) [5–8,11,12,27,28] due to their relatively simple scenes. For detecting the vanishing point on unstructured roads, an adaptive soft voting scheme based on a variable-sized voting region using confidence-weighted Gabor filters for computing the dominant texture orientation at each pixel was proposed [19]. By combining both time and spatial information using the efficient random walker algorithm, a fully automatic road segmentation algorithm was proposed based on a vanishing-point-constrained edge detection technique developed for detecting road boundaries [24].

To detect a vehicle for real-time collision warning systems in daytime and nighttime, Sobel edge detection and the Hough transform techniques have been used in the lane marking detection stage to extract lane marking information. Two features of vehicle shadows and horizontal edges are then extracted to detect the locations of vehicles, where these two features can be obtained by Otsu's method and horizontal edge detection method, respectively. Moreover, the distance between the host vehicle and a vehicle in front of it can be estimated using exponential functions [6]. However, this method is only suitable for the structured roads (e.g., highways, expressways, freeways) and might not work well on unstructured roads (e.g., urban/suburb roads) because features, such as lanes, vehicle shadows, and horizontal edges, are not explicit.

For using a hand-held video camera in a moving vehicle, the shadow feature and the vanishing point were employed to detect vehicles [5]. In the hypothesis generation (HG) stage, the shadow feature of a vehicle is used to generate potential vehicle locations, and in the hypothesis verification (HV) stage, these locations are verified using the constraint of the vanishing point. Otsu's method is used to segment shadow regions; unfortunately, this approach is sensitive to illumination variation, making shadow segmentation difficult in complex scenes. A similar difficulty was encountered in another method [6].

A monocular vision-based vehicle detection and inter-vehicle distance estimation method was proposed in [8] for a driving assistance system. In this method, a Haar-like feature descriptor is used to extract the shadow feature in the HG stage, which is then verified using directional edge features in the HV stage to reduce false positives. Furthermore, inter-vehicle distance is estimated using improved versions of existing position-based and width-based algorithms. In another method [7], the same features were also used and the vehicle/non-vehicle classification is implemented using an Adaboost cascade classifier.

Recently, the HOG descriptors have been exploited to detect potential vehicle locations and then they can be verified using a support vector machine (SVM) classifier as vehicle or non-vehicle. Based on such HOG & SVM approach, an efficient lane and vehicle detection with integrated synergies (ELVIS) for both lane and vehicle detection is developed [12], where the HOG features are extracted from the mask of a sliding window and verified using an SVM classifier. However, the high computational cost due to the use of multi-scale sliding windows to scan the whole image prohibits real-time operation. HOG features and an SVM classifier have also been used in another method [11].

Most existing methods for vehicle detection and distance estimation are suitable for only structured roads or can extract only one vehicle on urban/suburb roads. The present study thus proposes monocular vision-based vehicle detection and distance estimation methods for roads with varied scenes (e.g., urban/suburb roads).

3. Proposed method

This paper is an extension of our previous work [29]. Fig. 1 shows a flowchart of the proposed method for vehicle detection and inter-vehicle distance estimation, which is composed of several subsystems, including image preprocessing, information collection, vanishing point detection, segmentation of road regions, front vehicle detection and estimation of inter-vehicle distance. Input images are first converted into gray-scale ones because color information is not used in this work. Then, down-sampling and removal of the still area of frame are introduced for increasing computational efficiency. The information collection about detection of the vanishing point involves the Canny edge detection, the removal of horizontal and vertical lines, the Hough transform, and identification of left and right lines. During this stage, vanishing point candidates are found and then verified using the DB-SCAN (density-based spatial clustering of applications with noise) algorithm in order to obtain the final vanishing point. Based on the vanishing point and both lines of road's edges, road regions can be completely segmented. Afterward, front vehicles detection within the segmented road regions is conducted using the shadow feature, followed by the verification of the hypothesis generation (HG) using the HOG feature and SVM classifier. Finally, the inter-vehicle distance is estimated using the information of the vanishing point and the ratio of the true distance to the image pixels. The details of these processes are described in the following subsections.

In Section 3.1, the image preprocessing is described, and information collection and detection of the vanishing point are introduced in s 3.2 and 3.3, respectively. Then, the segmentation of road regions is explained in Section 3.4. Finally, the front vehicle detection and estimation of inter-vehicle distance are given in s 3.5 and 3.6, respectively.

3.1. Image preprocessing

Vehicle video camera recorders (also called dashcam) with full HD (i.e., 1920 × 1280 pixels) format are commonly used in ADAS. HD pictures have rich information but suffer from high computational complexity for computer vision applications. In the proposed system, full color images are first transformed into grayscale images, followed by downsizing through the bilinear interpolation approach. The sizes of still area of frame taken inside the vehicle will vary with different dashcam positions on the windshield, as shown in Fig. 2(a). Such still area can be removed by temporal differencing of $k+1$ consecutive frames, as shown in Fig. 2(b). The number of different pixels in $k+1$ consecutive frames can be summed as:

$$N(y) = \sum_{i=0}^{k-1}\sum_{x=0}^{W-1} F_i(x,y) \oplus F_{i+1}(x,y), \quad 0 \leqslant y < H, \quad 0 \leqslant i < k+1 \quad (1)$$