Expert Systems With Applications 91 (2018) 472–479

Contents lists available at ScienceDirect

Expert Systems With Applications

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

PHD filter for vehicle tracking based on a monocular camera

Fernando García[a,*], A. Prioletti[b], P. Cerri[b], A. Broggi[b]

[a] Intelligent Systems Laboratory, Universidad Carlos III de Madrid/ Avda. de la Universidad 30, 28911, Leganés, Madrid, Spain
[b] Vislab from University degli studi di Parma, Italy

A R T I C L E I N F O

Article history:
Received 6 January 2017
Revised 3 August 2017
Accepted 9 September 2017
Available online 13 September 2017

Keywords:
PHD filter
Vehicle detection
Computer vision
Vehicle tracking

A B S T R A C T

Novel advance driver assistance systems, such as emergency braking and adaptive cruise control require the most reliable detection algorithms. Furthermore, in the recent years, the use of computer vision approaches in these type of applications is becoming more frequent. However, when dealing with these technologies, reliability is a very important factor that still requires improvement. On this paper, it is presented a tracking algorithm which aims in improving the accuracy of these applications, based on computer vision and modern Probability Hypothesis Density (PHD) Filter technique. The tracking is performed on the features detected within the bounding box provided by a computer video based vehicle detection algorithm. The features tracked are combined in a last stage, providing accurate monocular camera tracking. Test provided, allowed to identify the best method for feature combination. Furthermore, it was proved that under the proper visibility conditions, the PHD filter design is able to improve current methods such as Unscented Kalman Filter.

1. Introduction

Vehicle detection is one of the main applications among Advance Driver Assistant Systems to know the location of other road users. It is important in order to fully understand the situation of the road. Classical sensing technologies (e.g. radar or laser technologies) have showed trustable results regarding vehicle detection, but their high costs are prohibitive for the general use.

Modern computer vision technologies are able to detect and track vehicles based in with low cost cameras. However, one of the main drawbacks of computer vision approaches is the lack of reliability and accuracy of the detections. The work presented in this paper provides an expert system which focuses on the use of a Probability Hypothesis Density (PHD) filter to track visual features belonging to a detected vehicle using a monocular camera. The use of the PHD filter improves the accuracy of the detection system by means of a modern tracking technique, enhancing the full performance of the detection algorithm, thus providing higher reliability and accuracy to computer vision based algorithms.

The aim of the present work is to provide a step forward in the field of Expert Systems, in two directions. First, to provide a low cost application (monocular camera based) for vehicle tracking in road environments. The second is to enhance the tracking techniques in road safety applications, which plays a key-role on the Expert Systems Applications in the field of road safety, by means of a modern and advance technique, the PHD filter.

2. State of the art

Vision based vehicle detection is a common topic in Expert Systems and Intelligent Transport Systems (ITS) fields. In the first, applications such as autonomous overtaking system (Milanés et al., 2012a) or automatic stopping (Milanés et al., 2012b) are based on advanced vision together with control algorithms. Other classical applications are advanced localization (Gruyer, Belaroussi, & Revilloud, 2016), pedestrian detection (García, García, Ponz, de la Escalera, & Armingol, 2014), driver drowsiness detection (Jo, Lee, Park, Kim, & Kim, 2014), lane departure (Son, Yoo, Kim, & Sohn, 2015), motorcycle detection (Anaya, Ponz, García, & Talavera, 2017), rear obstacle detection (Kim, Choi, Yoo, Yang, & Sohn, 2015), and roundabout moving obstacle detection (Hassannejad, Medici, Cardarelli, & Cerri, 2015) and vehicle detection and tracking (Lee, Mallipeddi, & Lee, 2017). Other advanced control applications use V2V information, such as cooperative control for highways (Pérez, Milanés, Godoy, Villagrá, & Onieva, 2013) and intersection management (Bi, Srinivasan, Lu, Sun, & Zeng, 2014).

In the ITS field, computer vision applications for vehicle detections are also a common topic. Sivaraman and Trivedi (2013a) provide wide state of the art for vision based vehicle detection systems. Regarding monocular detection, work can be divided into those that use appearance features and those that use optical flow based approaches. Appearance features approaches include Histograms of Oriented Gradients (HOG), presented on Dalal and

* Corresponding author.
E-mail addresses: fegarcia@ing.uc3m.es (F. García), prio@vislab.it (A. Prioletti), cerri@vislab.it (P. Cerri), broggi@vislab.it (A. Broggi).

http://dx.doi.org/10.1016/j.eswa.2017.09.018
0957-4174/

Triggs (2005) for pedestrian detection, applications which are used in Cheon, Lee, Yoon, and Park (2012) and Teoh and Bräunl (2012) for vehicle detection. Haar-Like features (Viola & Jones, 2001), are used in Sivaraman and Trivedi (2009) and Castangia, Grisleri, Medici, Prioletti, and Signifredi (2014) for vehicle detection. In Garcia, de la Escalera, Armingol, and Jimenez (2012) HOG, features are used together with a laser scanner for vehicle detection and tracking. Optical flow based approaches take advantage of the motion of the vehicles to identify them, in Arróspide, Salgado, Nieto, and Jaureguizar (2008) the authors combine optical flow and symmetry and in Garcia, Cerri, Broggi, de la Escalera, and Armingol (2012) optical flow is fused with radar for overtaking detection. Other approaches provide vehicle detections based on sensor fusion: In Bertozzi et al. (2008) obstacle detection and classification is performed based on radar and computer vision and Garcia, Martin, de la Escalera, and Armingol (2017) takes advantage of the laser scanner to enhance the vision detection. Modern techniques take advantage of the modern graphic processing units, which makes possible the use of deep learning for vision applications, based on convolutional neural networks, such as the work of Fan, Brown, and Smith (2016) for vehicle detection and Guindel, Martin, and Armingol (2017) for road users identification and pose estimation.

In vehicle detection algorithms, tracking is an important task which enhances vehicle detection, as well as vehicle localization. By means of vehicle tracking, vehicle detection can be performed even in difficult situations such as occlusions or misdetections. Furthermore, vehicle tracking allows to infer the motion of the vehicles being tracked, identifying not only their location, but their speed and direction. In monocular approaches, vehicles are usually tracked in the image plane, ¡ such as in Liu, Wen, Duan, Yuan, and Wang (2007) where template matching is used, unbiased finite memory filter in Choi et al. (2015), appearance and position similarities based on greedy data association in Lee et al. (2017) and a fuzzy cellular automata to enhance the vision based tracking (Darwish, 2017). Modern sensors such as 3D Lidars, allow to provide similar optical flow based tracking algorithm using more robust information such as in Daraei, Vu, and Manduchi (2017). However, the classical configurations are based on bayesian filters such as Kalman Filter (KF) (Haselhoff & Kummert, 2009) and Particle Filters(PF) (Sivaraman & Trivedi, 2013b; Liu, Li, Wang, & Ni, 2015). The authors in Hoffmann (2006) infer the 3D information from monocular camera to perform tracking based on 3D world coordinates and KF. Although KF is a common and widely adopted approach, the non-linearity of the observations leads to estimation errors. These errors are addressed by further approaches specifically designed to overcome these situations: i.e. Extended Kalman Filter, used in Barth and Franke (2009), and in Lim, Lee, Kwon, and Kim (2011) for stereovision detection, and Unscented Kalman Filters (Garcia et al., 2017). In recent years the increasing popularity of PHD filters has led to the development of modern and practical applications, where these estimation tools are used to enhance visual approaches performance. In Meissner, Reuter, and Dietmayer (2013) the authors used PHD filters with Multiple Models for tracking road users (pedestrians and vehicles) at an intersection, based on multiple laser scanners. The authors in Lamard, Chapuis, and Boyer (2013) used Cardinalized PHD to provide multisensor detection for vehicles and pedestrians. Zhang, Chen, Stahle, Buckl, and Knoll (2012) and Zhang, Stahle, Gaschler, Buckl, and Knoll (2012) used PHD together with feature detection to provide visual odometry.

3. General description

The tracking algorithm proposed is performed in three stages: First the vehicle detection algorithm provides visual based detection by means of machine learning methods. Once the vehicle is detected and the visual features are located, a PHD filter applied to those features is used to track the movement of the vehicle. The last stage provides movement estimation based on the combination of the features tracked. This paper focuses on the novel tracking method designed that relates to the second and third stages, therefore information related to the vehicle detection and feature identification will be introduced to add consistency, although as it is not part of the tracking system no further information is provided.

The proposed method was developed within the BRAiVE (BRAin-drIVE) project (Grisleri & Fedriga, 2010). BRAiVE is the prototype vehicle developed by Vislab for Advance Driver Assistance Systems (ADAS) and Autonomous Driving research. It is equipped with 10 cameras, 5 laser scanners, 1 GPS + IMU, 1 e-Stop system covering a 360° view around the vehicle.

4. Vehicle detection and feature identification

The vehicle detection algorithm is based on a soft-cascade approach (Adaboost) and Haar-like features (Viola & Jones, 2001). In order to remove false positives and speed up the processing stage, the detection is performed only in regions that, according to the calibration parameters, respect the typical range of vehicle sizes.

The detection algorithm has been tested on a training set of 75,000 rear images of cars and 110,000 negatives examples, after the bootstrapping process, with a pattern of 32×32 pixels. This technique is later combined with further reclassification techniques and calibration in order to remove false positives.

Once the vehicle is detected and the bounding box surrounding the vehicle delimited, features are searched based on multiple local convolution, key points and descriptors, extracted from two different hash images as described in Geiger, Ziegler, and Stiller (2011). Stable feature locations are obtained by filtering the input images with 5×5 blob and corner masks. Finally, a non-maximum- and non-minimum-suppression was applied.

The vehicle detection assumes that the lower part of the bounding box is located at $z = 0$, therefore, according to Eq. (1), based on the pin-hole model, it is possible to calculate distances to the detected vehicle in meters x and y (3D world coordinates) from the camera coordinate system (2D) in pixels, once one of the axis (z) is fixed.

$$\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = K\left[R|\hat{t}_0\right]\begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = P\begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} \tag{1}$$

where K is the intrinsic calibration parameters matrix, R and $\hat{t}_0$ are the extrinsic calibration parameters matrix (rotation and translation respectively). x_i, y_i and z_i are the world coordinates that represent the distance to the detected vehicle (in meters) and u_i and v_i are the coordinates in pixels.

Camera movement is compensated through a pitch detection algorithm based on visual odometry and line detection. A combination of both approaches is needed due to the drift introduced by features in the visual odometry and lines lack in some scenarios.

4. Feature based tracking

Once the features are selected, the bounding box in the new frame is calculated using optical flow. When the new bounding box