Expert Systems With Applications 62 (2016) 17–31

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Video-based tracking of vehicles using multiple time-spatial images

Niluthpol Chowdhury Mithun[a,c,1], Tamanna Howlader[b], S M Mahbubur Rahman[c,*]

[a] Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521, USA
[b] Institute of Statistical Research and Training, University of Dhaka, Dhaka 1000, Bangladesh
[c] Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1205, Bangladesh

ARTICLE INFO

Article history:
Received 5 April 2015
Revised 8 June 2016
Accepted 9 June 2016
Available online 11 June 2016

Keywords:
Intelligent traffic management system
Key vehicular frame
Time-spatial images
Vehicle tracking from video

ABSTRACT

An innovative idea of vehicle tracking for video-based intelligent traffic management system is known to bring significant socioeconomic impact. A successful vehicle tracking method is always in demand to monitor different traffic parameters such as the average speed, strange movements, and congestion of vehicles or even to detect accidents automatically on highways or freeways. The challenges of traditional video-based vehicle tracking methods include the initialization of tracking to tackle an unknown number of targets and the reduction of the drift sensitivity of targets from true positions mainly caused by the variations in lighting condition, occlusions and camera position. To address these challenges, this paper presents a novel vehicle tracking method for a traffic management system that introduces the multiple time-spatial images (MTSIs)-based detection in the stochastic filter-based tracking. The MTSI-based tracking employs the concept of multiple numbers of key vehicular frames (KVFs) for each of the vehicular-objects in the traffic. These KVFs provide highly accurate positional information of the vehicles due to the fact that the shape and texture of the vehicles are comparable on the same scale and do not depend on the speed of the traffic. The spatial correspondence of a vehicle in successive KVFs is then incorporated as a low-complexity data association technique to alleviate the common problem of drifting in the stochastic filter-based method and thereby increasing the accuracy in tracking trajectory. Comprehensive experimentations are carried out using two publicly available video databases (EBVT and GRAM-RTM) that have traffics of varying environments to evaluate the vehicle tracking performance of the proposed method as compared to the existing methods. Experimental results demonstrate that the introduction of MTSIs not only automates the initialization of tracking, but also significantly increases the accuracy of the tracking trajectories of the vehicles on roads evaluated both in the presence and absence of ground truths.

1. Introduction

Fatalities and serious injuries caused by the traffic-related accidents are globally recognized as a serious and growing problem due to increasing usage of automobiles (National Highway Traffic Safety Administration, 2011). In order to reduce traffic-related accidents and increase the chances of safe and smooth driving of automobiles on highways or freeways, the development of intelligent traffic management system that automatically tracks the vehicles on roads has been recognized as an active area of research in the past two decades (Sun, Bebis, & Miller, 2006). Further, the automatic tracking of vehicles is required in numerous transportation related surveillance applications including the monitoring of roads to acquire information on well-known traffic parameters such as the average speed as per the categories of vehicles, overstepping on road marks, congestion of vehicles, detection of foreign objects, and inference of suspicious activities on road. Such an expert tracking system embedded in vehicles can be also useful for safety-aware instructions for driver assistance or even for the development of driverless intelligent transportation systems.

1.1. Related works of vehicle tracking

There exists three major approaches of vehicle tracking, viz., wireless distributed sensor network-based tracking (Brooks, Ramanathan, & Sayeed, 2003; Duarte & Hu, 2004), remote-sensing-based tracking, e.g., radar sensing (Gunnarsson, Svensson, Danielsson, & Bengtsson, 2007; Tokoro, Kuroda, Kawakubo, Fujita, & Fujinami, 2003) and lidar sensing (Galceran, Olson, & Eustice, 2015; Premebida, Monteiro, Nunes, & Peixoto, 2007; Weigel, Lindner, & Wanielik, 2009), and video-based tracking (Mandellos, Keramitsoglou, & Kiranoudis, 2011; McCall & Trivedi, 2006; Sivaraman & Trivedi, 2013). Since the performance of sensor network-based

* Corresponding author. Fax: +88028613046.
E-mail addresses: niluthpol.mithun@email.ucr.edu (N.C. Mithun), tamanna@isrt.ac.bd (T. Howlader), mahbubur@eee.buet.ac.bd (S.M.M. Rahman).
[1] Significant part of this work has been done when the author is in BUET.

http://dx.doi.org/10.1016/j.eswa.2016.06.020

Table 1
Strengths and weaknesses of major approaches of vehicle tracking.

Sensor network-based tracking	Remote sensing-based tracking	Video-based tracking
Strength: • Existing radio network such as cellular network can be used *Weaknesses*: • Vehicles need to be sensor mounted • Reliability depends on strength of radio signals • Applicable only when there is availability of network	*Strengths*: • Independent of availability of radio network • Target vehicle or object does not need to be sensor mounted *Weaknesses*: • Insufficient field-of-view due to the use of principle of radar • Highly sensitive to environmental noise	*Strengths*: • Independent of availability of radio network • Sufficient field-of-view of the target objects • Noise robust tracking system can be developed *Weakness*: • Tracking system can be dependent on illumination

tracking methods is highly dependent on the availability of controlling radio frequency signals or sensors, such methods fail to provide instantaneous assistance when the sensor-mounted vehicles are out of range of the network or any surrounding vehicle does not have any sensor. Remote sensing-based tracking methods transmit the radar or laser signals from moving vehicles and estimate the position of surrounding objects using the corresponding received signals (Hoogendoorn, Zuylen, Schreuder, Gorte, & Vosselman, 2003). In many cases, the costly radar or lidar sensors fail to provide sufficient field-of-view required for vehicle tracking and due to high sensitivity to environmental noise such methods very often fail to classify the moving objects (Sivaraman & Trivedi, 2013). In this context, video-based tracking techniques capture video frames using the noise-robust and ultra-fast CCD or CMOS cameras that usually have a wide field-of-view of the vehicle. Table 1 summarizes the strengths and weaknesses of the sensor network, remote sensing, and video-based vehicle tracking approaches. In general, the video-based tracking methods are preferred to the others, since such an approach provides very accurate estimates of the relative positions and classification of surrounding vehicles even in the remote areas (McCall & Trivedi, 2006).

Traditional video-based vehicle tracking techniques follow two approaches - one treats the identification of vehicular objects in a frame and the estimation of correspondences of the objects in successive frames independently (Amer, 2005; Kim, 2008), while the other treats both the issues jointly (Aksel & Acton, 2010; Dellaert & Thorpe, 1998; Isard & Blake, 1998a; Segall, Chen, & Acton, 1999; Stauffer & Grimson, 1999). In the case of simultaneous tracking of multiple vehicles, the second approach is very often preferred to the first to obtain higher accuracy of the estimated trajectories. To identify the vehicles and find their positions in successive frames, the simplest approach can be template matching by assuming the rigid body movements of the objects (Brunelli, 2009). If the pose of the object changes, then the classical Lucas-Kanade affine tracker can be employed (Lucas & Kanade, 1981). In order to obtain a tracking system that is robust to the change of shapes and viewing positions of the vehicles, the corners or points of interest of the deformable vehicular objects are determined and features such as the histograms of oriented gradients (HOGs) (Niknejad, Takeuchi, Mita, & McAllester, 2012; Olmedo, Sastre, Bascon, & Caballero, 2013), the speeded up robust features (SURFs), the scale invariant feature transforms (SIFTs) (Lu, Izumi, Teng, & Wang, 2014; Mantripragada, Trigo, Martins, & Fleury, 2013; Shi & Tomasi, 1994) and the binary robust invariant scalable keypoint (BRISK) features (Hassannejad, Medici, Cardarelli, & Cerri, 2015) are obtained from these points. Due to the fact that the vehicles are identified and their positions are estimated using the HOG, SURF, SIFT or BRISK features generated from the points of interest only, the tracking trajectories estimated from these methods may not provide satisfactory performance for occlusions or noisy environments. Rather, the standard statistical techniques applied to the entire set of pixels of moving objects show a greater success in general for joint recognition of vehicles in the frames and estimation of their positions in the successive frames.

Among statistical methods, moving vehicles in a video traffic are modelled in the state-space frame work, in which the pixel intensities of the vehicular objects are assumed to follow certain random processes. Measurements for the random processes include the position, velocity, acceleration, and the color histogram of vehicular objects in the frames (Lee, Ryoo, Riley, & Aggarwal, 2009). The stereo depths of the moving objects (Ess, Leibe, Schindler, & Gool, 2009; Zhu, Yuan, Zheng, & Ewing, 2012) and the scene geometry or viewing condition (Olmedo et al., 2013) are also used for increasing the accuracy of tracking trajectories. In order to find the solution of the state-space model, the density function of the random processes can be chosen as parametric or non-parametric. The Kalman filter (KF)-based vehicle tracking is the most popular among the parametric approaches (Chiverton, 2012; Shantaiya, Verma, & Mehta, 2015; Sivaraman & Trivedi, 2013), which obtains an analytical solution for tracking by assuming linear dynamics of vehicular movements and a Gaussian distributed intensity of vehicular objects (Shalom, Li, & Kirubarajan, 2001). Due to the variations of traffic, weather or viewing conditions, the intensities of the vehicular objects may follow non-Gaussian statistics and the movements of vehicular objects may follow a non-linear dynamics. In such a case, advanced versions of the KF including the extended Kalman filter (EKF) have been proposed (Li, Wang, Wang, & Li, 2010; Mandellos et al., 2011; Simon & Chia, 2002). In complex traffic environments, the non-parametric approach that uses the Gaussian mixture density function (Stauffer & Grimson, 1999), kernel-based mean-shift filter (Comaniciu, Ramesh, & Meer, 2003), or the particle filter (PF) (Aksel & Acton, 2010; Chan, Huang, Fu, Hsiao, & Lo, 2012; Isard & Blake, 1998a; 1998b; Liu, Li, Wang, & Ni, 2015), has also been adopted to describe the nonlinear and non-Gaussian random processes of the moving objects. To improve the tracking performance, in addition to the pixel intensities of the moving objects, the color cues (Barcellos, Bouvie, Escouto, & Scharcanski, 2015; Lehuger, Lechat, & Perez, 2006; Nummiaro, K-Meier, & Gool, 2002; Yin, Zhang, Sun, & Gu, 2011) or edge features (Kumar & Sivanandam, 2012) have also been used in the traditional PF. Other statistical tracking algorithms include the nearest neighbor (NN), the multiple hypotheses tracking (MHT) (Cox & Hingorani, 1996; Kim, Li, Ciptadi, & Rehg, 2015; Zulkifley & Moran, 2012) and the joint probabilistic data association filter (JPDAF) (Shalom, Fortmann, & Cable, 1990). The NN-based methods are not at all reliable for tracking in the cluttered environment (Raol, 2010). The MHT-based data association relies on the enumeration of hypotheses, the number of which can grow exponentially considering all possibilities (Oh, Russell, & Sastry, 2004). Due to the nature of sequential tracking, in general, the JPDAF-based methods perform better than the MHT-based methods (Shalom, Daum, & Huang, 2009). However, the limitations of the JPDAF-based methods lie in their