Transportation Research Part C 44 (2014) 202–213

Contents lists available at ScienceDirect

Transportation Research Part C

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

Camera calibration and vehicle tracking: Highway traffic video analytics

Yiwen Wan, Yan Huang, Bill Buckles *

Department of Computer Science & Engineering, University of North Texas (UNT), Denton, TX, USA

ARTICLE INFO

Article history:
Received 16 August 2012
Received in revised form 25 February 2014
Accepted 26 February 2014

Keywords:
Autonomous traffic surveillance
Camera calibration
Vehicle tracking
Hungarian algorithm

ABSTRACT

We describe a real-time highway surveillance system (RHSS), which operates autonomously to collect statistics (speed and volume) and generates incident alerts (e.g., stopped vehicles). The system is designed to optimize long-term real-time performance accuracy. It also provides convenient integration to an existing surveillance infrastructure with different levels of service. Innovations include a novel 3-D Hungarian algorithm which is utilized for object tracking and a practical, hands-off mechanism for camera calibration. Speed is estimated based on trajectories after mapping/alignment with respect to dominant paths learned based on an evolutionary dynamics model. The system, RHSS, is intensively evaluated under different scenarios such as rain, low-contrast and high-contrast lightings. Performance is presented in comparison to a current commercial product. The contribution is innovation of new technologies that enable hands-off calibration (i.e., automatic detection of vanishing points) and improved accuracy (i.e., illumination balancing, tracking via a new 3-D Hungarian algorithm, and re-initialization of background detection on-the-fly). Results indicate the capability and applicability of the proposed system in real-time and real-world settings.

1. Introduction

Intelligent traffic systems (ITS) have been expanding with incorporation of multiple technologies into vehicles, roadway, highway, tunnel and bridge surveillance. Such include image processing, pattern recognition, electronics and communication technologies. These have been employed for monitoring traffic conditions, reducing congestion, enhancing mobility, and increasing safety. Vision-based technology is a state-of-the-art approach with advantages of easy maintenance, real-time visualization and high flexibility compared with other existent technologies. This makes it one of the most popular techniques in ITS for traffic control. The most recent widely read description of vision-based highway surveillance is that of Coifman et al. (1998). The last one and a half decades have witnessed improvements in cameras, communications, and video analytics. This paper presents an update to the latter.

Modern commercial vision-based surveillance systems have improved considerably since the *tripwire* systems described by Coifman et al. (1998). Many commercial systems may be tailored to monitor either freeways, arterials, bridges, or tunnels (Econolite's Autoscope Solo Terra, Traficon's VIP/T, Citilog's XCam-I, Kapsch Trafficom AG's VR-2). There are about two dozen products by approximately a dozen vendors that have a freeway option. In contrast to the earlier survey cited (Coifman et al., 1998), today's systems all incorporate vehicle tracking. Incidents reportable by the systems include wrong-way driver

* Corresponding author. Tel.: +1 940 565 4869.
E-mail address: bbuckles@cse.unt.edu (B. Buckles).

http://dx.doi.org/10.1016/j.trc.2014.02.018
0968-090X/

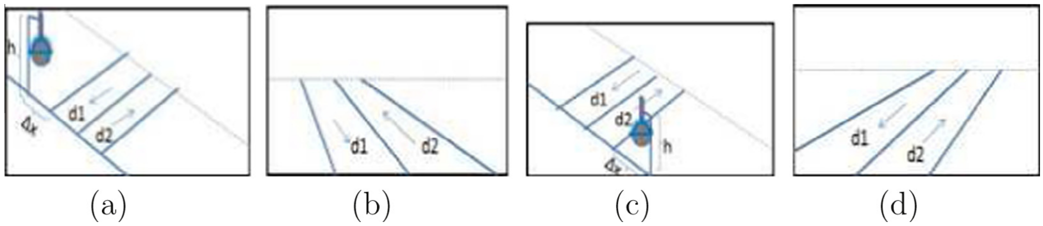

Fig. 1. Typical road structure and camera locations.

(Traficon's VIP/D), stopped vehicle (Econolite's Autoscope Solo Terra), and roadway debris (Iteris's Abacus). Nearly all systems collect data on speed and volume (Iteris's Versicam, Traficon's Traficam Collect-R among others). A few collect data on occupancy (Econolite's RackVision Terra, Iteris's Vantage Edge 2, PEEK's VideoTrak 900) and yet fewer collect data on vehicle class (Iteris's Abacus) or gap time (Traficon's VIP/T).

Many researchers have devoted efforts to investigating different solutions for extracting important traffic information via image and video analysis. Important traffic information includes speed (Mao et al., 2009; Cathey and Dailey, 2005), volume (Pan et al., 2010) and incident detection (Zou et al., 2009). However, for any traffic surveillance system to be capable of such high-level functionality it must be able to detect objects and track them correctly. Additionally, the camera must be calibrated such that the real-world distances can be calculated and utilized for speed estimation.

Normally, for vehicle detection, most methods (Gupte et al., 2002; Beymer et al., 1997; Hsu et al., 2004) assume that the camera is static even if it is a pan-tilt-zoom camera. Given the assumption, foreground vehicles can be detected by image differencing between current frame and estimated background. In order to maintain low computation complexity we have followed this direction. Moreover, we applied a gamma correction based mechanism to suppress sudden illumination changes.

After vehicles are detected, a vehicle tracking mechanism is needed. As mentioned in a tracking survey (Yilmaz et al., 2006), objects are normally represented as points, primitive geometric shapes, object silhouettes and contours, articulated shape models, and skeletal models. Objects in a highway traffic scenario are mostly rigid and a point is an efficient model for representation. The traffic surveillance system proposed in Coifman et al. (1998) detects corner feature points to represent vehicles in order to deal with occlusion. However, corner features may not be easily detected when image quality is low. For instance, a strategy is to reduce transmission bandwidth by reducing resolution or the effects environmental variations. The latter, often consisting of heavy rain or snow, can affect the accuracy and robustness of vehicle analysis. Due to concerns mentioned previously, we utilized the center of the mass to represent objects. Here a 3-D Hungarian tracker is designed to associate points (center of mass) over time. This is effective in dealing with streaming video having low frame rate in order to reduce transmission bandwidth.

Shadow removal is also an issue which proves to be a major source of error in detection and classification. Especially the shadows of large trucks prevent smaller, adjacent vehicles from being detected successfully. Most traffic surveillance systems detect and remove shadows or assume the analyzed sequence includes no shadows. We developed an approach proposed in Xiao et al. (2007) to remove shadows.

Camera calibration is the process of estimating camera parameters so that pixel points in camera coordinates can be mapped into real-world coordinates. In real-time traffic applications the process is more difficult than usual. Perspective effects cause vehicle geometric features such as length, width, and height not to be constant. In other words, different positions at which cameras are installed give different perception angles for each lane. Researchers (Kanhere and Birchfield, 2010) have evaluated a number of both manual and automatic camera calibration techniques and from them we chose the two-vanishing-points-and-known-width (VVW) approach. Practical reasons are: (1) through trajectory analysis one vanishing point can be easily estimated; (2) it requires no manual measurement as would methods starting from camera height or length of a line segment on the pavement; (3) width of a lane is easily accessible because every U.S. state strictly follows highway lane width standards[1] and highway lanes can be easily found through trajectory analysis; and (4) performance-wise, VVW is both accurate and stable. The camera perspective is preferred to be one of the typical views shown in Fig. 1(a) and (c). Basically the height of the camera should be sufficient so that the camera captures a high-angle view such that the vertical occlusion is minimized. Moreover, the horizontal distance from the camera to the road surface, Δx, should be reasonable so that horizontal occlusion is also minimized. In Fig. 2, both poor and reasonable camera perspectives from different viewing angles are illustrated.

The rest of this paper is organized as follows. In the next section, an overview of the entire system is given. Then, initialization (i.e., learning the road structure and camera calibration) is discussed in Section 3. Section 4 describes the operational phase including vehicle tracking, data collection, and event detection. Then, Section 5 reports experimental results over 24 lane hours under various lighting and weather conditions. Finally, the conclusion appears in Section 6.

[1] http://www.fhwa.dot.gov/policyinformation/statistics/2008/hm39.cfm.