Advanced Engineering Informatics 38 (2018) 392–403

Contents lists available at ScienceDirect

Advanced Engineering Informatics

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

Full length article

A vision-based statistical methodology for automatically modeling continuous urban traffic flows

Hana Rabbouch[a,c,*], Foued Saâdaoui[b,c,e], Rafaa Mraihi[d]

[a] Université de Tunis, Institut Supérieur de Gestion de Tunis, Cité Bouchoucha, Tunis 2000, Tunisia
[b] Department of Insurance and Risk Management, College of Economics and Administrative Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), P.O Box 5701, Riyadh, Saudi Arabia
[c] University of Monastir, Laboratoire d'Algèbre, Théorie de Nombres et Analyse Non-linéaire, Faculté des Sciences, Monastir 5019, Tunisia
[d] Université de Manouba, Ecole Supérieure de Commerce de Tunis, Campus Universitaire de La Manouba, Tunis 2010, Tunisia
[e] Department of Statistics, Faculty of Sciences, King Abdulaziz University, P.O Box 80203, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Keywords:
Urban traffic flow
Video summarization
Wavelet-based regression
Information retrieval
Long memory models
Simulation

ABSTRACT

We introduce an online video-based virtual sensor allowing to automatically estimate and forecast the number of vehicles passing through a road section over a continuous time interval. The strategy consists, in the first place, in defining a Motion Intensity Index (MII) whose role is to quantify the visual activity in a traffic video. A wavelet-based cause-and-effect statistical model is then used to match the actual number of vehicles to their respective motion scores. This leads to an efficient estimator of the urban traffic flow. The implementation is well optimized in such a way that the local sampling rate is directly proportional to the amount of visual activity in localized sub-shot units of the video. The procedure allows designing an autonomous sensor giving every moment a measure of the flow on a road section and an expectation of its future levels. The device can be very useful for optimizing transportation management, facilitating strategic decision-making, and analyzing networks with the purpose of optimizing transportation equipment efficiency.

1. Introduction

The last decades, because of the rapid expansion of digital video technologies, information indexing and retrieval has become a more and more important topic and several researches have been devoted to the video summarization based on visual features. Many advanced engineering applications have grown in volume and diversity, especially in some relevant fields like surveillance and tracking, recognition systems, and transportation and pedestrian flows modeling [2,6,27,34]. In transportation, the use of video sensors and computer-aided vision techniques for creating databases about the urban traffic have been intensely investigated in the recent years, since traffic videos provide more information about the traffic of vehicles than other classes of sensors (e.g. inductive loops, sonar or microwave detectors) [28]. Today, there are several informatics methods for detecting and counting vehicles in traffic videos, where foreground-background segmentation methods have been the most used. Background subtraction algorithms have been exploited for vehicles detection, with exceptional conditions of lighting and weather [16]. However, these approaches require a priori information about the scene without any moving vehicles and have problems with occlusions. Hsu et al. [9] detected foreground pixels and removed shadows, and then computed the exponential entropy in the detection zone to test whether a vehicle exists or not. Hsieh et al. [8] used Kalman filter for tracking vehicles extracted from background models. They implemented a shadow removal algorithm to extract the size and linearity features of vehicles for the purpose of categorizing them. A double-difference operator with gradient magnitude has been used in Cucchiara et al. [4] to detect vehicles, however, it failed to appropriately handle inter-frame luminance variations. Optical flow techniques have been used to estimate the motion between subsequent frames, see works of Li and Chellappa [12] and Tao et al. [26]. Ma et al. [14] proposed a real-time vehicle behavior analysis, which can be used in traffic jams and under adverse weather conditions. The main advantage of their approach was to make a point tracking system for the vehicles' behavior without a difficult image segmentation procedure. However, the method has been only tested on highways, which are known to be less complex than urban environments.

Thus, many techniques still need to be extended to operate more efficiently in high-frequency traffic conditions with robustly handling

* Corresponding author at: Université de Tunis, Institut Supérieur de Gestion de Tunis, Cité Bouchoucha, Tunis 2000, Tunisia.
E-mail address: hana.rabbouch@gmail.com (H. Rabbouch).

https://doi.org/10.1016/j.aei.2018.08.006
Received 23 December 2017; Received in revised form 5 August 2018; Accepted 15 August 2018

illumination changes, noise, occlusions and shadow effects. Modeling high-frequency traffic flow data offers transportation engineers and researchers the opportunity to better understand the dynamics of certain nonrecurrent events such as the duration of congestion, which frustrate commuters, companies and traffic operators because they cause unexpected delays. Once extracted, traffic databases have also to be efficiently exploited within powerful decision-making models. Extrapolating motion data in the time scale, on the basis of optimal models, allows to plan ahead alternative scenarios, thus leading to an early, efficient and cost-effective intervention for the resolution of the main transportation problems. Vlahogianni et al. [29,30] discussed the performance of the traffic models with data coming from novel technologies, including video-based devices, and how they are fused to improve short-term traffic forecasting. Wang et al. [32] explored the predictability of traffic and congestion by using comprehensive records of global position system devices installed in vehicles. Providing velocity and locations of a large number of taxis in real time, the databases have been exploited in the investigation and quantification of the predictability of segments in main roads in an urban network. Wang and Yao [31] used virtual loops as vehicle detection zones in the image, and then, trained a neuro-fuzzy network to build a data-driven adaptive classifier, which judges whether a vehicle is located in the virtual loop. Recently, Rabbouch et al. [19] proposed a summarization approach allowing to both count and recognize vehicles. The principle of the method is to fixe a straight line of pixels (fictive sensor) transversally to the road in the traffic video so that it imitates the old-fashioned magnetic inductive loop. Then, the motion intensity all along the pixels line, from one frame to another, forms a number of signals containing anarchic information about the road traffic. Unsupervised clustering is then performed to model vector signals, thus providing key information about traffic flows.

Nevertheless, several constraints have impeded the development of an online approach which, at the same time, collects and models data and forecasts future traffic levels. Among most impacting constraints are calculation and storage costs. The strategy we propose in this paper extends the method recently developed in [19], by proposing a more effective alternative allowing to not only to easily handle larger amounts of vehicles, but also to assess expected future flow levels. The strategy begins by extracting meaningful sub-shot units of the traffic video sequence and calculating the motion activity in each of these units using an appropriate Motion Intensity Index (MII). Sampling is adaptively carried out in such a way that sub-shot elements throughout periods of high flow rates, are sampled at a higher rate than in periods of low flow rates. Then, a cause-and-effect statistical model is used to match the number of vehicles to their respective amounts of motion. Regressing on wavelet subspaces play an important role in breaking down intensity signals into smooth components while reducing the noise which has been caused by the video discretization. This leads to an efficient estimator of the urban traffic flow. The accuracy of estimators is improved by splitting days into subintervals where local regressions are performed. Such a remedy allows to overcome vision problems which are related to the underlying weather conditions. Finally, an adequate time series model is assigned and forecasting tasks are carried out.

The proposed approach represents an online data-collecting and forecasting tool whose main advantage, in comparison with state-of-the-art methods, is that it combines online traffic motion analysis to statistical modeling and forecasting. In the literature, counting and modeling were commonly two independent tasks. On one hand, there is a corpus devoted to the main innovations in the field of vehicles counting and recognizing, including pneumatic road tubes, inductive loops, traffic video analysis, etc. On the other hand, statistical modeling and time series forecasting have been based on early collected and processed data bases [1,17,18,30]. Hence, the current strategy is the first and unique approach blending a vision-based data collection with a forecasting model. Another interesting feature of this approach, is that it does not necessitate a complex implementation in both computational and storage standpoints. All these features are of great importance to engineers and managers since the current system has the advantage of being straightforward, objective and efficient. What makes this approach special is also its data sampling, which unlike the few works in this context, rather it is performed on an extremely high frequency basis. This is a very important issue, since it gives engineering analysis more promptness and accuracy, giving also the opportunity to adjust fractionally integrated autoregressive models, thus opening new insights towards understanding many complex problems such as those related to pollution and road injuries and the underlying formation mechanisms.

The paper is structured as follows. Section 2 provides an illustration of the mathematical formulation of the data acquisition and preprocessing stages. The wavelet-based regression method used for assessing the streamflow is presented in Section 3. Section 4 briefly reviews the properties of the forecasting model. The experimental setup and implementation are described in Section 5. Finally, the empirical results and error analyses are reported and discussed along with recommendations for subsequent uses of the new technique.

2. Data acquisition and preprocessing

The strategy consists in assessing the amount of visual activity within each sub-shot unit. In traffic videos, each pixel defines a time-dependent sequence that can then be considered as,

$$\widetilde{p}_{x_1,x_2}(t), \quad t = t_0 + \Delta t, \ \dots, t_0 + T\Delta t, \tag{1}$$

with $(x_1, x_2) \in \Omega \subset \mathbb{N}^2$ and T is the number of sampled frames. In other words, fixing x_1 and x_2 is equivalent to focusing on a single pixel, which from a frame to another yields one discrete signal. It deserves also to be noted that the intensity data are often prone to noise contamination due to the video time-discretization. Thus, we can consider Eq. (2) as

$$\widetilde{p}_{x_1,x_2}(t) = s_{x_1,x_2}(t) + \epsilon_t, \quad t = 1, 2, \dots \tag{2}$$

where $s_{x_1,x_2}(t)$ is a smooth component and $\epsilon_t \sim i.\ i.\ d.\ (0, \sigma^2)$ is an additive independent and identically distributed noise. Below, we also assume that the stepsize $\Delta t = 1$ and $t_0 = 0$. Below, we will explain how we need to process these signals properly to retrieve the smooth component. That is why, the principle of adaptive smoothing will be much discussed below.

Now, for each pixel localized at (x_1, x_2) coordinates, and at time t, we consider an indication variable denoted $\delta^t(x_1, x_2)$. Within our focus-frame, if the intensity of the pixel (x_1, x_2), $\widetilde{p}_{x_1,x_2}(t)$, exceeds some threshold ϵ, then a value 1 is assigned to $\delta^t(x_1, x_2)$, otherwise, 0 is assigned. This is formulated as:

$$\delta^t(x_1, x_2) = \begin{cases} 1 & \text{if } \widetilde{p}_{x_1,x_2}(t) > \epsilon \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Accordingly, we consider a time-dependent Motion Intensity Index (MII), which firstly consists in calculating, at each one of the sampled frames, the sum of binarized pixels. Then, we cumulate calculated scores frame-by-frame. The cumulated score between two points s and t ($s < t$) in the time interval is equivalent to quantify the windowed motion within a sub-shot of length $t-s$. Thus, we define $m: \mathbb{N} \to [0, 1]$ the MII discrete function expressed as follows:

$$m_t = \frac{1}{L} \sum_{s<i\leqslant t} \sum_{x_1,x_2} \delta^i(x_1, x_2) \tag{4}$$

where $\delta^i(x, y)$, $x_1 = 1, 2, \dots, x_2 = 1, 2, \dots$ is the ith binarized frame of the motion sequence within the lath sub-shot unit, and L is a normalizing parameter which is equal to length of the sequence T_s multiplied by the dimension of the analysis window d. Therefore, the resulting MII curve indicates the temporal variations of the visual activity in the original video sequence.