A Survey of Visual Traffic Surveillance Using Spatio-Temporal Analysis and Mining

Chengcui Zhang, Department of Computer and Information Sciences, The University of Alabama at Birmingham, Birmingham, AL, USA

ABSTRACT

The focus of this survey is on spatio-temporal data mining and database retrieval for visual traffic surveillance systems. In many traffic surveillance applications, such as incident detection, abnormal events detection, vehicle speed estimation, and traffic volume estimation, the data used for reasoning is really in the form of spatio-temporal data (e.g. vehicle trajectories). How to effectively analyze these spatio-temporal data to automatically find its inherent characteristics for different visual traffic surveillance applications has been of great interest. Examples of spatio-temporal patterns extracted from traffic surveillance videos include, but are not limited to, sudden stops, harsh turns, speeding, and collisions. To meet the different needs of various traffic surveillance applications, several application- or event- specific models have been proposed in the literature. This paper provides a survey of different models and data mining algorithms to cover state of the art in spatio-temporal modelling, spatio-temporal data mining, and spatio-temporal retrieval for traffic surveillance video databases. In addition, the database model issues and challenges for traffic surveillance videos are also discussed in this survey.

Keywords: *Content-Based Retrieval of Traffic Events, Intelligent Transportation Systems, Spatial-Temporal Modelling, Traffic Surveillance, Vehicle Classification, Vehicle Tracking*

1. INTRODUCTION

In recent years, the increase of traffic volume has become a significant problem. Consequently, road transportation systems have been subject to considerable increase in congestion and accidents. These include both congestion-related accidents and accident-caused congestions. In Europe accidents cost about 45 billion euros per year, with 45,000 victims annually (Foresti, Micheloni, & Snidaro, 2003). In the United States, in one year alone, in one year alone, motor vehicle crashes cost a total of $100 billion in medical care, rehabilitation, and lost wages (http://www.cnn.com/2010/HEALTH/08/25/motor.vehicle.accident.costs/index.html). On average, each driver in the United States ponies up about $500 a year toward the total costs, according to a recent study by the Centers for Disease Control and Prevention.

DOI: 10.4018/jmdem.2013070103

Early solutions to reducing both traffic accidents and congestion such as adding more lanes are becoming less feasible and the least favorable. Instead of increasing the capacity of existing infrastructure, contemporary solutions of Intelligent Transportation Systems (ITS) try to use roads more efficiently. In recent years, Intelligent Transportation Systems (ITS), which integrate advances in telecommunications, surveillance systems, information systems, automation, and electronics to enhance the efficiency of existing road networks, have been identified as the new paradigm to address the growing mobility problems, and to alleviate congestion and augment the quality of vehicular flow. The demand for better traffic information and thus an increasing reliance on traffic surveillance has resulted in a need for better vehicle detection and traffic analysis tools. The basic components of an intelligent traffic surveillance system include the data acquisition, pre-processing, and interpretation of the traffic conditions (Foresti, et al., 2003). Although loop detectors are the most widely used sensors, they present significant errors in their measurements and incur high costs of installation and maintenance. There are several other sensors using different technologies also being widely adopted, including magnetometer technology, infrared technology, acoustic detection technology, ultrasonic detection technology, CCD (Charged Coupled Devices) camera, Doppler radar technology, etc. as cited in (Weill, Worton, & Garcia-Ortiz, 1998).

To evaluate the quality of surveillance data provided by different sensors, Ozbay and Kachroo (1999) listed four important factors which are reliability, performance under different environmental conditions, data accuracy, and real-time performance. Reliability of a sensor is measured by the frequency of malfunctions or breakdowns. Unreliable sensors could cause high costs in fixation and replacement. Performance under different environmental conditions is another important factor. According to the study presented in (Weill, et al., 1998), inductive loop and magnetometer do not perform well under snow conditions. Infrared sensors do not work well under rain or snow conditions. The performance of CCD cameras is less satisfactory under bad weather conditions (e.g. rain, snow, and fog). It is important to make sure that the selected sensor will perform well under the specific environment conditions of the area where it will be deployed. Data accuracy largely depends on the installation and calibration of the sensor (Weill, et al., 1998). Real-time performance is essential in timely information collecting, analysis, and decision making. High-speed communication technologies can only ensure the timely transmission of available data (e.g. raw data and known knowledge), but cannot help to speed up the process of analyzing and interpreting the data and generating useful insights of traffic conditions. As on-line applications have become a trend and a focus, it is very important that the associated analysis tools for traffic sensors have real-time performance.

In addition to the above four factors, another important factor is the traffic information that could be generated by different sensors. Compared with traditional sensor techniques like loop detectors based on induction or magnetic field sensors which are used in many applications, surveillance cameras, a type of wide area sensors, can provide more specific traffic information and usually monitor several lanes in different directions in parallel. It is expected that more and better traffic information, such as traffic volume, occupancy, and vehicle's speed, which is automatically gathered during the analysis of traffic surveillance videos is emphasized. In addition to these quantitative parameters, more qualitative information such as abnormal events (e.g. traffic accidents, illegal U-turns, etc.) and their confirmation is also desired for the effective execution of traffic and incident management functions. The solutions to extracting this qualitative information based on video analysis have become more feasible recently, due to the advances in computer vision and image processing techniques and the decreasing costs of image processing hardware (http://www.path.berkeley.edu). On the contrary, the analysis of specific traffic behavior in an unknown scene is impossible with loop detectors. A detailed discussion of the pros and cons of each sensor technology as well as the traffic parameters that could be measured by each technology will be presented in Section 2.

While a whole new generation of methodological and algorithmic constructs are being developed to exploit the powerful capabilities afforded by the ITS technologies, concurrent efforts needed to enable practical implementation are lacking in some crucial aspects. One such aspect with sparse focus is the ability to collect, analyze, and store large-scale visual traffic surveillance data for real-time usage. It implies capabilities to:

- Store and catalogue data in an organized manner for easy access;
- Reconstruct traffic situations through offline analysis for addressing traffic safety and control; and
- Automate the process of data indexing and retrieval by obviating the need for human intervention and supervision.

While each of these capabilities significantly enhances operational feasibility, the last capability has critical implications for real-time implementation in terms of substantially reducing computational time for the associated control procedures. One key application domain that addresses these three capabilities is the ability to track vehicular flow from traffic surveillance video sequences both in time and space for easy and unsupervised access.

This paper will focus on the visual traffic surveillance systems for traffic monitoring and present a survey of different methods in visual vehicle segmentation, visual tracking, spatio-temporal modeling and mining, unusual traffic event (e.g. incidents) identification and retrieval, and the traffic surveillance video database model based on the spatio-temporal characteristics of vehicle objects.

2. TRAFFIC SURVEILLANCE AND RELEVANT SENSOR TECHNOLOGIES

Traffic surveillance systems measure traffic flow parameters and send this traffic information to traffic management units. In order to effectively execute traffic and incident management functions, reliable, accurate, and timely traffic information must be obtained. This information includes (but is not limited to) the following:

- Traffic Performance;
- Traffic Congestion;
- Incident Detection and Confirmation;
- Incident Assessment;
- Vehicle Speed and Direction;
- Vehicle Occupancy;
- Vehicle Location.

This information is used to detect and verify traffic incidents and congestion, detect dangerous behaviors (e.g., illegal U-turns, speeding, ghost drivers), initiate incident response, implement traffic control strategies, and monitor network flows. The necessity of such a massive volume of real-time data requires that surveillance technologies be employed in an information-gathering role. The last decade has been an exciting time for traffic surveillance technologies. Since the inception of ITS, many new and effective traffic sensors have been introduced. Traffic information may come from a variety of sensors such as visual (e.g. optical and infrared) and non-visual sensors (radar, laser, ultrasonic range-finders, GPS, loop detectors, microwave sensors, etc.). A review of these technologies with their advantages and disadvantages is presented in sections 2.1 and 2.2, with video surveillance systems being emphasized. Section 2.3 presents sample applications of video technology for traffic management and safety. The issues, challenges, and trends in visual surveillance systems for ITS are presented in Section 2.4.

2.1. Non-Visual Technologies

2.1.1. Inductive Loop Detectors

Among the available sensors, inductive loop detectors are perhaps the most "proven" traffic surveillance technology. Loop detectors are placed in the subsurface of the roadway and when utilized can provide real-time traffic data. This technology is an active detector technology that responds to ferrous mass (cars) (Ozbay & Kachroo, 1999). However, loop detectors present significant errors in their measurements and has been noted that the cost of installation and maintenance of loop detectors can be prohibitively high. For example, in freeze climates, if pavements are in poor condition and the installation is poorly done, the wire can break, causing additional lane closures in order to replace the failed loop (Hallenbeck, 2005). Thus, other technologies must be examined in order to provide a more cost-effective alternative. These alternative technologies provide not only cost-savings but also have the ability to obtain a greater variety of traffic and incident-related data.

2.1.2. Microwave Radar

Microwave radar technology is an active technology that has been used for many years in navigation and traffic control. Microwave radars that are used in the United States in vehicle detection applications transmit energy at 10.525 GHz as regulated by the Federal Communications Commission (FCC). Two types of microwave radars are used in the United States in traffic monitoring applications. The first type of radar measures vehicle speed by emitting electromagnetic energy at a constant frequency. The Doppler principle is used to calculate the velocity of a passing vehicle by evaluating the difference in frequency between the transmitted and received signals. However, the downfall of the constant-frequency microwave radar is its inability to detect a stopped vehicle. Thus this technology cannot be used in a number of important traffic and incident management roles. The second type of microwave radar commonly used in traffic applications emits what is referred to as a "frequency-modulated continuous wave" or FMCW, which varies in frequency with time. Unlike the constant-frequency microwave radar, FMCW emitting radars are capable of detecting the presence of a stopped vehicle as well as vehicle speed. The ability to determine vehicle speed and presence allows the second type of microwave radar (sometimes referred to as a "true presence" radar) to be utilized in a number of applications ranging from signalized intersection control to vehicle counting. Microwave radar systems are not (or less) affected by the weather problems experienced by video surveillance technology (Hallenbeck, 2005). However, they generally produce slightly less accurate volume count information that obtained with loops and/or video detection. Another major disadvantage of the microwave radar technology is the lack of suitable locations for its installation. Because the radar device must be placed beside or above the road, an overhead structure must be constructed. Although the side positions are usually less costly, the above positions produce more accurate information because there are fewer problems with object occlusions (Hallenbeck, 2005). The erection of separate support structures for individual devices is not cost-effective, thereby limiting the placement of devices to pre-existing structures.

2.1.3. Passive Infrared Detectors

Passive infrared devices are sensors that detect the infrared energy emitted by objects that are within the detection range of the device. They operate by measuring changes in energy emissions within the sensor's field-of-view, in that "the change in energy is proportional to the absolute temperature of the vehicle and the emissivity of the vehicle's metal surface (emissivity is equal to the ratio of the energy actually emitted by a material to the energy emitted by a perfect radiator of energy at the same temperature)" (Loral AeroSys, IBI Group, Jackson & Lull, & Edwards and Kelcey, 1995).

The primary advantage of the passive infrared sensor is that it is capable of detecting vehicles at a greater range than sensors that depend on visible wavelengths; however, the accuracy of the detector can be degraded by heavy rain or snow. Passive infrared detectors are also limited by their inability to collect speed data (only vehicle presence data is obtainable).

2.1.4. Active Infrared Detectors

Active infrared detectors operate in much the same manner as microwave radar devices, in that energy from the sensor is reflected off of the vehicle in order to obtain data. These detectors are capable of measuring vehicle presence as well as vehicle speed; however, atmospheric and placement concerns (a suitable structure must be constructed, as in the case of the microwave radar) prevent this technology from being cost-effective.

2.1.5. Ultrasonic Detectors

Ultrasonic detection technology is active and employs reflected sound to detect vehicles. Ultrasonic detectors operate by transmitting sound at the 25-50 KHz range. While more expensive detectors can measure Doppler speed, the most widely used (and low-cost) detectors provide only vehicle presence data. Ultrasonic devices are commonly compact in size and easy to install, however their practical traffic management use is restricted by the limited amount of data obtainable in a cost-effective manner.

2.1.6. Acoustic Detectors

Acoustic detection technology is another type of above-ground sensor technologies. These sensors are passive and employ an array of microphones built into the sensor allowing the device to detect traffic based on spatial processing changes in sound waves received at the sensor. After processing and analyzing the received sound waves, vehicle detection and traffic flow information is then assigned to the appropriate user-specified regions or lane being monitored, forming a picture of the traffic. When acoustic sensors are deployed, their microphone sensitivity is pre-set for normal operating conditions which include typical weather conditions. In addition, the software and operating instructions to control an acoustic sensor require on-site attention to improve and upgrade the capability of the unit, or complete replacement to upgrade the sensor itself.

2.2. Visual Technologies

2.2.1. Video Cameras and Image Processors

Video surveillance cameras like CCD cameras are passive sensors that use the intensity changes in visible light for vehicle detection. Video image processors function by analyzing the consecutive frames supplied by a video camera. While video cameras and processors are typically more expensive than other surveillance mechanisms, and video systems usually involve costs associated with installation and maintenance along with training personnel, their flexibility and additional benefits are incredible. Video cameras can be used to obtain vehicle presence, speed, length, and lane change data for multiple lanes of traffic with a single detector. Video systems can benefit automated enforcement programs which aim to reduce violations and crashes related to speeding, red-light running, and illegal U-turns. In addition, the costs associated with purchase, installation, operations, and maintenance are now lower than that for some traditional methods, which makes video surveillance more attractive than before. For example, the installation of cameras is non-intrusive and typically does not require lane closures, thus, the traffic disruptions are minimized. Additionally, video surveillance allows for rapid incident detection and verification as visual information is readily understandable by human operators. With video surveillance systems, it is also possible to identify incident type, the possible cause, and the type of intervention needed. While this is generally performed by trained personnel, emerging video data mining and information analysis techniques (Chen

& Zhang, 2006) make it more possible than ever to automate this process. For example, the information contained in the video segments immediately preceding an incident can be particularly valuable. Video information analysis can provide a clear understanding of the cause of the accident, and this information can be used not only for safety but also for transportation administration to reduce accidents, enhance traffic planning, and improve traffic emergency response systems. Further, the classification of different vehicle types is also made possible by advanced video mining techniques (Zhang, Chen, & Chen, 2006). Vehicle classification can be used to compute the percentages of vehicle classes which are often counted manually by human operators and often outdated due to the lack of an automated updating system. The use of an automated vehicle classification system can lead to cost-efficient decision about the thickness of road pavements. In addition, it can provide data about vehicle categories that use a particular street.

2.2.2. State of the Art on Visual Traffic Surveillance Systems

As indicated in (Foresti, et al., 2003), the general architecture of a visual-based traffic surveillance system consists of a set of CCD cameras, an background updating and change detection module, an object detection and classification module, an object tracking module, and an event recognition module. According to the study of Zhu et al. (Zhu, Xu, Yang, Shi, & Lin, 2000), a practical visual surveillance system for real traffic surveillance applications must have some key factors. First, it should be easy to install and calibrate, even for non-expert personnel. Second, the system should be able to adapt to varying environmental conditions e.g. different light and illumination conditions and the presence of shadows. Third, it should be able to provide vehicle speed and size information for advanced applications such as traffic intersection control and vehicle type classification. The last but not least, for the wide use of an on-line traffic system, it must enable (near) real-time operations with relatively low costs (Zhu et al., 2000).

Many visual systems that are currently being used are limited to quantitative traffic information such as queue length and road occupancy, to aid congestion detection and control (Chapuis, Potelle, Brame, & Chausse, 1995; Taniguchi, Nakamura, & Furusawa, 1999; Haag & Nagel, 2000). Thus, image processing units within these systems delivers some general traffic parameters without identifying individual vehicles. For example, the approach used in the Image Processing for Automatic Computer Traffic Surveillance (e.g. IMPACTS) system focuses on spatially analyzing image intensities (Roman, 2002). Instead of considering traffic based on individual vehicles, the underlying strategy is to describe how the road space is being utilized at a particular time instance (Roman, 2002). Temporal analysis of the change pattern of these descriptions is then performed in order to determine the abnormal situations in traffic flow. There are three qualitative categories used to describe the use of road space, which are 'no traffic present', 'moving traffic present', or 'stationary traffic'. This approach is computationally inexpensive and can be implemented in real-time. However, its ability in detecting unusual events remains the same as that of traditional sensors.

Some systems provide more detailed information about the behavior of individual vehicles in order to identify unusual events and/or predict incidents. Issues associated with extracting traffic movement and accident information from real-time video sequences are discussed in (Cucchiara, Piccardi, & Mello, 2000; Dailey, Cathey, & Pumrin, 2000; Kamijo, Matsushita, & Ikeuchi, 2000; Huang &Russell, 1998). Two common themes exist in these studies. First, the moving objects (vehicles) are extracted from the video sequences. Next, the behavior of these objects is tracked for immediate decision-making purposes. However, few works have considered the problem of indexing the spatio-temporal data for on-line analysis, storage or off-line pattern querying. For vehicle object detection, three methods for moving object detection within the VSAM (Video Surveillance and Monitoring) test bed were developed (http://www-2.cs.cmu.edu/vsam/research.html#COMPUS). One of them uses temporal differencing to detect moving targets and train the template matching algorithm. These targets are then tracked using template matching. Another approach to moving object detection uses a moving airborne platform (Cohen & Medioni, 1999). A key issue in vehicle detection and tracking is background modeling and subtraction, which has been discussed in the literature (Stauffer & Grimson, 1999; Grimson, Stauffer, Romano, & Lee, 1998; Haritaoglu, Harwood, & Davis, 1998; Chen, Shyu, Peeta, & Zhang, 2003). This usually involves the creation of a background model that is subtracted from the input images to create a difference image. Ideally, the difference image contains only the moving objects (vehicles). These techniques range from modeling the intensity variations of a pixel via a mixture of Gaussian distributions, to simple differencing of successive images. For example, the Autoscope system (Panda, Jacobson, & Bernard, 1996) has been developed for vehicle detection and traffic flow measurement using machine-vision techniques. A three dimensional filter, which includes a temporal dimension, is first used to estimate the background image. The filter is designed to adapt to variations in background intensity due to variations in illumination due to passing clouds and nocturnal illumination (Panda, Jacobson, & Bernard, 1996). Pixels with their intensities different from the estimated background intensities are thresholded and combined with edge images to detect moving objects. Chen et al. (2003) proposed an adaptive background modeling and vehicle detection framework in which the generation of new background images and the segmentation of vehicle objects are performed in an iterative manner. The work of Toyama et al. (Toyama, Krumm, Brumitt, & Meyers, 1999) provides some simple guidelines for the evaluation of various background modeling techniques. There are two key problems in this context: 1) a complex learning model is highly time consuming, and 2) a simple differencing technique cannot guarantee good segmentation performance.

Following the vehicle detection, the vehicle objects are then tracked. Some commercial systems such as Autoscope (Panda, Jacobson, & Bernard, 1996) use feature tracking methods to track vehicles. Visual features include edges, critical points, region textures, etc. These features are tracked between two consecutive frames over a sequence of frames. In Autoscope, there are two methods for tracking vehicles – symbolic tracking and numeric signature tracking. In signature tracking a set of intensity and shape signature features are extracted for each detected object. These features are correlated in the next frame to find the location of the objects in the next frame (Panda, Jacobson, & Bernard, 1996). Signature feature tracking is useful in alleviating the problem of occlusions. In (Smith, Richards, Brandt, & Papanikolopoulos, 1996), Smith et al. proposed an adaptive filtering scheme to track feature windows on the target in spite of the unconstrained motion of the target, possible occlusions of feature windows, and changing target and environmental conditions. This method is relatively fast and robust to varying conditions and there are no explicit target models. In symbolic tracking, objects are independently detected in each frame and tracked between two consecutive frames. The spatial locations and the previous moving patterns (e.g. moving direction) are often used to identify the symbolic correspondence of a set of vehicle objects between two consecutive frames. The tracking mechanism proposed in (Chen, Shyu, Peeta, & Zhang, 2003) also falls into this category, in which a simple yet effective spatial reasoning based on the vehicle's centroid location is performed to build up the correspondence. The temporal trajectory of each vehicle can be used to obtain various spatio-temporal parameters (e.g. velocity, acceleration, and driving direction) and further reveal new knowledge such as usual events like sudden lane changes and accidents. Each of the two tracking methods produces better results under different conditions. There are many other existing video-based approaches (Kanhere, Pundlik, & Birchfield, 2005; Lou, Tan, Hu, Yang, & Maybank, 2005; Song & Nevatia, 2005) that have adopted a single calibrated camera for the tracking task that targets automatic extraction of vehicle parameters (e.g., world position and orientation), which is a fundamental problem in traffic monitoring. However, recovering 3D information from a 2D view is inherently an under-conditioned problem. Thus, they all require some heuristic to serve as the missing constraint, which most often comes from image foreground analysis, and is also most likely to become the major bottleneck of their performances. In (Song & Nevatia, 2005), image foreground blobs are compared to a database of pre-generated 2D projections of vehicle models from a range of viewpoints, so as to gain a constraint on the vehicle orientation. As a result, its performance is dependent on the quality of foreground extraction and noise (shadows and reflections) removal. Lou et al. (2005) iteratively minimizes the residual between foreground edge points and the projection of pre-defined 3D wireframe models to estimate the pose of a vehicle in world coordinates. However, this approach has difficulty handling vehicle-vehicle occlusions, because the merged foreground edges may lead to unexpected local minima. A feature based approach is gaining more attention due to its ability to track under partial occlusions. Its early application in vehicle tracking is explored by Beymer, McLauchlan, Coifman, &Malik 1997) and later improved by Kanhere et al. (2005) to handle the low-angle scenario. Nevertheless, in order to gain any 3D information, they still must provide the missing constraint, although in a more straightforward way: in Beymer et al. (1997), the world coordinates are gained through a homography between the road and the image plane, which is equivalent to assuming that the camera is high enough that all feature points can be treated as lying on the ground (z = 0). In order to remove this restriction, Kanhere et al. (2005) estimate the z coordinate of a feature point by dropping a plumb line down to the boundary of foreground regions. The latter approach, although suitable for more traffic scenarios, relies on image foreground classification, and thus is sensitive to vehicle shadows and occlusions, which may render the height estimate invalid. Moreover, the reconstruction of a feature point has to be delayed until it has been tracked through a block of frames, so that its height estimate can be considered stable. Both of these methods reduce the world coordinates of feature points to 2D-space, which is solvable using one camera.

On the other hand, adding a second camera removes the bottlenecks of these methods and attacks the under-conditioning problem directly, because the additional view provides another two constraints and boosts the problem to an over-conditioned one. This idea formalizes the more recent approaches based on a structure from motion (SfM) framework, where 3D reconstruction from two views is already well studied (Hartley & Zisserman, 2004). Now the major problem becomes finding point correspondences between the two camera views. The corner-based local descriptors (Mikolajczyk & Schmid, 2005) widely used in image matching do not work well here, because many of the corners are blurred away due to vehicle motion, and they cannot handle the case when the two cameras have very different viewpoints. Yang et al. (Yang, Johnstone, & Zhang, 2007) simplify the problem by setting the two cameras very close to each other and using an optical flow based algorithm to find point correspondences across two views, knowing that the target should reside in a small neighborhood. However, this camera setting restricts its usability in different road conditions, and what is worse, the accuracy of 3D triangulation quickly decreases as the two cameras become similarly oriented. Yang et al. (Yang, Johnstone, & Zhang, 2009) further improve their previous work by exploring a new reconstruction scheme, which allows the two cameras to be installed disparately, yet is still able to reconstruct a rich set of vehicle features. In particular, they propose a matching algorithm to correspond feature points between two video streams where two cameras can be installed with a large difference in viewpoint, focal length, and image quality. Then they reduce the full reconstruction of vehicle features to a multiple labeling problem and show that, by adjusting the energy functions, it is possible to solve the problem of reconstruction, grouping of vehicles, and outlier removal at the same time. It is worth noting that while the spatio-temporal tracking of vehicles brings added sophistication to traffic surveillance, it requires more computing power and poses more restrictions in camera positioning in order to handle occlusions.

All described academic/commercial products are able to extract (more or less) traffic parameters. Unusual event recognition is based on these parameters. Thereby, parameters are subject to combined analysis according to an event. Various thresholding methods have been developed based on which certain unusual events can be detected. However, most of them are based on manually selected threshold values, which work for specific events but can be very ad hoc. For example, in (Veeraraghavan, Masoud, & Papanikolopoulos, 2003), a threshold is set on the distance between two vehicles, which is used as the measurement for detecting possible collisions. Similarly, in (Atev, Arumugam, Masoud, & Janardan, 2005), the overlap of two vehicles is regarded as the occurrence of collision. In contrast to thresholding methods, machine learning and data mining techniques, when applied appropriately, can be used to learn independently unusual events. No thresholds are defined. Only video training examples (Dance & Caelli, 1993) showing normal traffic of a particular road intersection or highway are necessary. Once the behavior model of normal traffic has been learnt, it can be used to detect unusual traffic events within the same scene (Roman, 2002). The system can be trained to analyze traffic behavior in an unknown traffic environment by studying the spatio-temporal behavior of vehicles. Various machine learning algorithms are explored for this purpose: 1) Hidden Markov Model, as in Porikli et al. (Porikli & Li, 2004) to estimate traffic congestion without vehicle tracking; in Kamijo et al. (Kamijo, Matsushita, & Ikeuchi, 2000) for traffic monitoring and accident detection at road intersections. 2) Belief Networks. Based on it, Huang et al. (Huang, Koller, Malik, Ogasawara, Rao, Russell, &Weber, 1994) propose a traffic scene analysis algorithm; Buxton and Gong (Buxton & Gong, 1995) use Bayesian belief networks to model dynamic dependencies between parameters involved in visual interpretation. 3) Self-Organizing Map (SOM), as the hierarchical SOM in (Xie, Hu, Tan, & Peng, 2004) and the fuzzy SOM in (Hu, Xiao, Xie, & Tan, 2004). 4) Time Series Neural Networks. In (Chen and Zhang, 2006), a multilayer feed-forward network is designed to process the time series data (e.g. trajectories) extracted from traffic surveillance videos and to identify unusual events like traffic accidents. To sum up, the data mining based event detection systems are highly flexible with less human interactions needed.

More related work in traffic event detection and scene understanding can be found in (Dance & Caelli, 1993; Minsky, 1986; Fernyhough, Cohn, & Hogg, 2000; Jung, Lee, & Ho, 2001; Tang & Gao, 2005; Redden, 2009). Dance and Caelli (1993) present a traffic scene interpretation system, which is based on a cognition model in AI. This model is originally suggested by Marvin Minsky (Minsky, 1986) and is implemented with the object oriented approach in (Dance & Caelli, 1993). Some statistical methods are also applied in this area. Fernyhough et al. (2000) construct qualitative event models in their work. In (Jung, et al., 2001), a video searching scheme is proposed that provides the functions of query by example, by sketch and parameter weighting. The searches in this scheme are based on parameters of moving vehicle trajectories. A nonparametric regression algorithm is examined in (Tang & Gao, 2005) for forecasting traffic flows.

There is yet another class of approaches that detect disturbances in traffic flow by monitoring how the visible road space is being utilized at a particular time instant. This approach is used in the Image Processing for Automatic Computer Traffic Surveillance (e.g.

IMPACTS) system, concentrating on analyzing the spatial distribution/layout of image intensities. It avoids analyzing traffic condition on a vehicle by vehicle basis and detects events by analyzing how these spatial intensity distributions vary over time. A simple categorization schema is to divide the use of road space into three categories: no traffic present, moving traffic present, or stationary traffic. These are essentially qualitative decisions. In (Porikli & Li, 2004), Porikli and Li propose a traffic congestion estimation method that directly extract features (e.g. motion features) in the MPEG compressed domain. These features are combined to classify the current traffic state into one of the five categories: the stopped, heavy, light, open, and empty traffic states. This approach is computationally efficient, robust to illumination changes, and invariant to camera settings.

2.3. Applications of Video Technology for Traffic Management and Safety

"Digital video is an extremely promising and powerful technology that benefits areas ranging from law enforcement to entertainment," said Gail Whipple, vice president, global digital media, IBM. The applications of video technology are primarily for traffic safety, transportation administration, and enforcement purposes:

- **Red-Light Running Enforcement:** Red-light running enforcement through the use of video technology has proven to be an effective deterrent. Usually, when the signal turns red, the detection system becomes active and the camera takes pictures when cars enter the intersection. Photos will be taken of the rear of the car or both the front and the rear ends. The camera will also record the date, time, time elapsed since beginning of red light and the speed of the vehicle (http://safety.fhwa.dot.gov/intersections/rlrcam_tech.htm). Tickets are issued after the photo has been reviewed and verified. There are three camera types that are generally used for red light enforcement: industrial quality 35mm cameras, digital cameras, and video cameras. Recently, the use of video cameras is receiving more attention for red light running enforcement because it can provide traffic parameters such as vehicle speed and predict whether or not a red light running violation will take place. If a red light running violation is predicted, an all-red signal can be created to stop the traffic in order to avoid potential collisions. Thus, it can help to mitigate the potential consequences of the violation;
- **Speed Violations:** Automated enforcement of speed violations operate in a similar way. Typically the camera is connected to a speed measuring device like microwave radar and a computer. The speed measuring device detects speed violations and triggers the camera unit. The photos/videos, with the date, time and speed recorded, are then used to determine the owner of the vehicle and tickets are issued. The implementation of equipment capable of handling speed violations would allow more efficient use of manpower in the area of law enforcement. The biggest challenge would be the legal issues and earning public support (http://ntl.bts.gov/lib/24000/24100/24125/Graettinger-UTCA00463.pdf);
- **Automated Enforcement Technologies in Improving Grade Crossing Safety:** This also involves the installation of video equipment at the grade crossing that monitors the vehicle traffic flow and detects traffic violations. The surveillance camera is activated to record an event only when a violation is detected. The current implementation and use of such systems is mostly "after the fact" safety enhancement, meaning that it does not provide any protection to avoid potential crashes although it can discourage violators. Therefore, public awareness efforts are critical to the success of video enforcement at grade crossings

(Redden, 2009). Legal issues pose another challenge as a judge may not accept videos to convict violators;

- **Adverse Meteorological Conditions Alert Systems:** Along with the use of video cameras at road intersections and highways, it may also be used in the deployment of adverse meteorological conditions alert systems. Adverse meteorological conditions alert systems use video frames and image processing techniques to provide immediate evaluation of the current meteorological conditions and the level of traffic flow. Most approaches are simply based on reference images by making use of a fixed camera placed on the roadway (Bush & Debes, 1998). There are also systems that use onboard cameras. For example, in Pomerleau (1997), Pomerleau adopts an approach to estimate relative visibility distance by measuring a contrast attenuation per meter on the road markings at various distances in front of the vehicle. The detection of road markings is required in this work. Hautiere and Aubert (2003) estimate the real current existing visibility distance under foggy weather. An extension of their work has been proposed in Hautiere, Labayrade, and Aubert (2006) to cover more meteorological conditions;
- **Vehicle Classification:** Vehicle classification can be used to compute the percentages of vehicle classes which are often counted manually by human operators and often outdated due to the lack of an automated updating system. The use of an automated vehicle classification system can lead to cost-efficient decision about the thickness of road pavements. In addition, it can provide data about vehicle categories that use a particular street. With more and more traffic surveillance cameras being installed, computer vision based methods, when coupled with machine learning techniques, can be applied to classify vehicle types. However, there has been relatively little work done in the field of vehicle classification. This is because it is an inherently hard problem. Most of the current work is purely dimension based (such as height and length of a vehicle) or shape based. In Wu, Zhang, and Wang (2009), a parameterized model is proposed to describe vehicles, in which vertices and topological structure are taken as the key features. One requirement of this method is that the image quality has to be sufficiently good to have the topological structures of vehicles exposed. However, this cannot be always satisfied in a real traffic surveillance system. Gupte et al. (2002) propose a system for vehicle detection and classification. The tracked vehicles are classified into two categories: cars and non-cars. The classification is based on dimensions and is implemented at a coarse granularity. Its basic idea is to compute the length and the height of a vehicle, according to which a vehicle is classified as a car or a non-car. Research by Lai and Yang (2000) proposes a virtual loop assignment and direction-based estimation method that is used to identify vehicle types. Each vehicle type is represented by a 1-D signature chart. In their experiment, vehicles are classified into four categories: 7-seat van, fire engine, sedan and motor cycle. With this method, as mentioned in the paper, only a rough estimation of vehicle types based on vehicle length is possible. It cannot distinguish vehicles whose lengths are in approximately the same range, e.g. truck and bus. Another problem of this method is that only those vehicles traversing across virtual loops along the road direction can be detected. Therefore, there is demand for more sophisticated methods that can unveil the real, invariant characteristics of a type of vehicle;
- **Principal Component Analysis (PCA):** PCA is a well-known algorithm in the field of object recognition. It used in the computer vision problem of human face recognition. The similarity between face detection and vehicle detection is that both analyze a 2-D image and try to find out the feature of the image content. In

Zhang, Chen, and Chen (2006), Zhang et al. propose a method based on Principle Component Analysis (PCA) and Support Vector Machines (SVM), which is able to classify a vehicle object extracted from surveillance videos into one of the three vehicle types – "passenger cars", "pickup trucks", "vans and SUVs". Independent Component Analysis (ICA) is another subspace method that has been applied to face recognition. Many works compare between ICA and PCA and show the advantages of ICA (Sahambi & Khorasani, 2003; Fortuna, Schuurman, & Capson, 2002; Sezer, Ercil, & Keskinoz, 2005). In (Sahambi & Khorasani, 2003), the authors applied both methods in analyzing the Coil-20 database. In (Fortuna, Schuurman, & Capson, 2002), the authors demonstrate that ICA outperforms PCA for object recognition under varying illumination. Sezer, Ercil, and Keskinoz (2005) compare the effectiveness of both methods in object recognition. Since the traffic videos are taken during different time periods of the day, it is preferably that the algorithm is robust to varying illumination conditions. Chen and Zhang (2007) propose a method that analyzes vehicle images with ICA and extracts a set of features from each vehicle. These features represent the innate characteristics of the vehicle type and are fed into the classification module - One-Class Support Vector Machine (Schölkopf, et al., 1999). The representative features of vehicles in each vehicle type are used as training data. One classifier is built for each vehicle type which distinguishes that vehicle type from the others. In the testing phase, each set of test vehicles is tested against the classifier of each vehicle type. A test vehicle is then classified into one of the vehicle types according to the highest score it receives from each classifier. To further improve the robustness to color variances, a texture analysis tool is used in this work. The result is an integrated system that can automatically track and categorize vehicles within a traffic surveillance video sequence. The system first tracks and segments vehicles from raw surveillance videos. Then the tracked vehicles and their features are normalized. In the final step vehicles are classified, which can provide more detailed and useful information to traffic administration. Their experimental results show that the ICA-based method outperformed the PC-based method;

- **Spatio-Temporal Modeling of Vehicles and Pedestrians:** Traditional visual surveillance systems are designed to operate with data taken at a time point rather than over a time span. This information alone, typically volume and occupancy, is not sufficient for effective and reliable usual event detection because neither volume nor occupancy is a dynamic measurement. While spatio-temporal applications for Intelligent Transportation Systems have only recently attracted researchers in this field, most of the existing work has concentrated on general-purpose spatio-temporal models for videos (Chen & Özsu, 2002; Chen & Kashyap, 2001; Day, Dagas, Iilo, & Khokhar, 1995). The existing models all have their advantages in modeling certain aspects of the data. However, none of them are "jack-of-all-trades" without any redundancy or loss of any efficiency. Since different spatio-temporal applications may have different emphasis on the properties and queries of the domain-specific data, there is a need for designing domain specific spatio-temporal models. A domain specific spatio-temporal model is especially desired for storing, indexing, and querying traffic surveillance videos. In addition, it is necessary to build an integrated system that captures a comprehensive set of requirements which are needed in building a transportation surveillance video database. Such requirements include the extraction of vehicle objects from surveillance videos, vehicle tracking, vehicle classification, and spatio-temporal modeling that provides efficient data indexing and querying for transportation domain. Xin and Zhang (2006) propose a spatio-temporal multimedia database model for managing transportation surveillance video data. Their objective is to build a spatio-temporal database schema for transportation surveillance videos, in which queries can be answered easily and efficiently. Their spatio-temporal model for transportation surveillance videos combines the strength of two general-purpose spatio-temporal multimedia database models - the Multimedia Augmented Transition Network model (MATN) (Chen & Kashyap 2001) and the Common Appearance Interval (CAI) (Chen & Özsu, 2002) model. While MATN model is good at modeling the replay of the multimedia presentation and the spatial-temporal relations of semantic objects in the video, it is not efficient in modeling or querying the trajectories of moving objects. In the meantime, while Common Appearance Interval (CAI) model can be used to better answer trajectory-based queries, it explicitly stores the spatial relations of pairs of objects in the model, which is considered redundant in transportation video databases. Their proposed model bases its structure on MATNs and adopts the concept of CAI to segment transportation surveillance videos. Since this model is motivated by transportation surveillance applications, it has some domain specific features. It models each traffic light phase in a MATN-like network and models the corresponding video segment using CAIs. In addition, CAIs are further divided into sub-intervals and modeled by the sub-network structure in MATNs.

2.3.1. Content-Based Retrieval of Traffic Surveillance Videos

The abundance of traffic surveillance videos has generated an urgent need for techniques to index and retrieve video clips from surveillance video databases according to video contents, based on automatic video understanding and video mining techniques. Most of the existing research for general-purpose video retrieval focuses on low level visual features (e.g. color, texture, and shape) and simple motion features. However, it is inherently hard to let the machine understand the content of the video data by only reading pixels, frames or signals. There exists a "semantic gap" between the low level features and the high level semantic meaning. Therefore, it is necessary that human provides some guidance to the machine in a semantic-based video retrieval system like a traffic surveillance video retrieval system. The purpose is to automatically learn and retrieve semantic scenes that are related to specific traffic events from the surveillance video database according to the user's query. Semantic retrieval of traffic surveillance videos poses challenges to the following areas: automatic extraction of spatio-temporal features, spatio-temporal mining of the relationship between these features and user-specified query events/scenes, spatio-temporal indexing of video data, and query interface, etc.

In particular, for the retrieval of traffic surveillance videos, not only the contents of the frames (still images) have to be studied but also the relations among them have to be considered. With content based image analysis, the semantic meaning of each object in an image can be extracted. From the perspective of each such object, its moving trajectory in consecutive frames is a kind of spatio-temporal data. The aim of semantic video retrieval is to extract semantic scenes by analyzing the spatio-temporal relations among moving and still objects in the video. The design of the retrieval framework shall first perform the object tracking and segmentation, thus extracting semantic meaning and moving trajectory of each object in the video. In the learning and retrieval phase, the technique of Relevance Feedback (RF) can be incorporated, with which the user provides feedback and the learning algorithm learns from it by penalizing the “irrelevant” scenes and encouraging “relevant” scenes. Such scenes/ events in a traffic surveillance video may be accidents i.e. car crashes. To model an accident scene is obviously different from modeling an “overtaking” event. However, there is no need to build models for every specific event. For instance, a car that bumps into wall and two cars bump into each other could be defined in one event model. Therefore, a proper spatio-temporal database model for traffic surveillance videos should not be restricted to tiny details of specific events as this will lose the generality in event modeling. Meanwhile, such a model should not target at too broad a range of events without actually grasping the characteristics of these events. In addition, the query system should be able to provide a convenient interface for users to specify queries (e.g. query by video example, query by sketches, and query by event types) and to provide feedbacks to the retrieval result. Xin et al. (Xin, Zhang, Chen, & Rubin, 2007) propose a human-centered interactive framework for automatically mining and retrieving semantic events in traffic surveillance videos. The framework first extracts vehicle objects from video frames and records their trajectories by tracking their centroids. Generally categorized semantic events (e.g., accidents) in a particular type of video are then modeled. These object trajectories and event models are fed into the core component of the framework for learning and retrieval. As trajectories are spatio-temporal in nature, the learning component is designed to analyze time series data. The human feedback to the retrieval results provides progressive guidance for improving the retrieval accuracy in the framework. The retrieval results are in the form of video sequences instead of contained trajectories for user convenience. Thus, the trajectories are not directly labeled by the feedback as required by the training algorithm. To solve this problem, a mapping between semantic video retrieval and Multiple Instance Learning (MIL) is established using an adapted neural network for time series prediction.

There are many other applications for which video technology is uniquely well suited. Some example applications are accidents detection, queue length detection, pedestrian identification, etc.

3. ISSUES, CHALLENGES, AND TRENDS

In the previous sections, we have reviewed the state-of-the-art of visual traffic surveillance following a general framework of visual traffic surveillance systems. Although a large amount of work has been done in visual traffic surveillance systems, many issues are still open and deserve further research, especially in the following areas.

3.1. Environmental Adaptation

The primary disadvantage inherent with video technology is operation in changing lighting conditions. As indicated by Zhu et al. (2000), a successful vision based system for real traffic surveillance applications must be able to work under different light conditions, including illumination changes due to changing weather conditions, reduced illumination in the night, vehicle headlights at night, etc. In the work of Kilger (1992), shadows are separated from vehicles by examining an edge map of detected regions (Foresti, et al., 2003). The image analysis algorithms used for night scenes are quite different from that used for daytime scenes. Although video cameras cannot see objects at night, they can see the front/rear light sources. Based on some visual measurements such as the brightness saturation and some spatial constraints such as the relative spatial relationships between a pair of headlights, it is possible to extract the main light sources representing a vehicle from surveillance videos at night or in other low-light conditions. Cucchiara and Piccardi (1999) propose an approach for vehicle detection at nigh time based on morphological analysis of headlight pairs. Other issues include the vehicle detection during the transition from afternoon to night and under difficult weather conditions such as fog and heavy rain. The algorithms that have been developed for working under different lighting conditions can be quite different; therefore, there is a need for developing intelligent video surveillance systems which can automatically select proper sets of image processing algorithms for the current scene condition.

3.2. Occlusion Handling

Occlusion is a major issue in visual traffic surveillance. During occlusions, objects are partially visible, which may cause the 'under segmentation' problem and make the vehicle extraction and tracking algorithms unreliable (Hu, Tang, Wang, & Maybank, 2004). When occlusion occurs between two vehicle objects, some solution is possible through motion tracking and feature matching. For example, Chen et al. (Chen, Shyu, Peeta, & Zhang, 2005) propose a simple, fast, yet effective algorithm to split two partially occluded vehicle objects by *backtrack-chain-update* algorithm together with binary map differencing. This algorithm has been proved effective in handling two-object occlusions without using the global features of objects such as color, edge or texture. However, when multiple objects occlude each other and have similar speeds and driving directions, they tend to be segmented as a single moving object, which makes the vehicle extraction and tracking of individual objects particularly difficult. One partial solution would be to combine the global feature analysis with the rich temporal and motion information contained in video data for multi-object occlusions. For example, if an object changes its illumination or rotates a lot in successive frames, the global feature based methods may not work due to the big change, but the centroid based tracking method is suitable for this situation. On the other hand, if an object makes an agile motion in successive frames, the centroid tracking method may not assure to find the corresponding object but can only identify the possible candidates. However, the analysis of global features will provide some kind of verification in choosing the best one among them. Thus, we can reduce the rate of mismatching, and cut the computation cost caused by global feature matching (for example, the block matching method used in (Kamijo, Matsushita, & Ikeuchi, 2000). Another solution, which is perhaps the most promising and practical method for addressing occlusion, is through the use of multiple cameras as detailed in Section 2. However, using multiple cameras will introduce extra computation cost as well as maintenance cost.

3.3. Other Issues

While the visual data has been explored in depth, the audio data in traffic surveillance videos has long been ignored. In fact, the unusual temporal change in audio feature often indicates traffic incidents or the presence of police cars, fire trucks, or ambulance. Thus, the audio cue can be used to obtain information such as the response time of emergency services. The related research and the actual implementation of such multi-modal traffic surveillance systems have recently been carried out in Japan and Taiwan.

There are many other implementation issues and legal issues as well for traffic video surveillance systems. For example, it has been documented that for transportation officials to efficiently implement video technologies, certain measures must be taken and the issue of future compatibility with existing local systems must be addressed http://ntl.bts.gov/lib/24000/24100/24125/Graettinger-UTCA00463.pdf. The design of local systems should remain as open ended as possible to ensure that they are fully upgradeable. The installation and calibration of video surveillance systems must be easy enough for on-site set-up, reconfiguration, and operation by non-expert personnel. It is also noted that video equipment requires continuous monitoring and ongoing maintenance by the road authority.

Communities must have sufficient computing technology and manpower to operate a video system, as well as an adequate network of wiring to transmit video data. The transmission of videos can be prohibitive because of the size of the data involved, thus partnership with wiring providers is vital.

REFERENCES

Atev, S., Arumugam, H., Masoud, O., & Janardan, R. (2005). A vision-based approach to collision prediction at traffic intersections. *IEEE Transactions on Intelligent Transportation Systems*, *6*(4), 416–423. doi:10.1109/TITS.2005.858786

Beymer, D., McLauchlan, P., Coifman, B., & Malik, J. (1997). A real-time computer vision system for measuring traffic parameters. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 495–501).

Bush, C., & Debes, E. (1998). Wavelet transform for analyzing fog visibility. *IEEE Intelligent Systems*, *13*(6), 66–71. doi:10.1109/5254.736004

Buxton, H., & Gong, S. (1995). Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, *78*, 431–459. doi:10.1016/0004-3702(95)00041-0

Chapuis, R., Potelle, A., Brame, J.-L., & Chausse, F. (1995). Real-time vehicle trajectory supervision on the highway. *The International Journal of Robotics Research*, *14*(6), 531–542. doi:10.1177/027836499501400601

Chen, L., & Özsu, M. T. (2002). Modeling of video objects in a video database. In *Proceedings of IEEE International Conference on Multimedia and Expo,* Lausanne, Switzerland (pp. 217-220).

Chen, S.-C., & Kashyap, R. L. (2001). A spatio-temporal semantic model for multimedia database systems and multimedia information systems. *IEEE Transactions on Knowledge and Data Engineering*, *13*(4), 607–622. doi:10.1109/69.940735

Chen, S.-C., Shyu, M.-L., Peeta, S., & Zhang, C. (2003). Learning-based spatio-temporal vehicle tracking and indexing for transportation multimedia database systems. *IEEE Transactions on Intelligent Transportation Systems*, *4*(3), 154–167. doi:10.1109/TITS.2003.821290

Chen, S.-C., Shyu, M.-L., Peeta, S., & Zhang, C. (2005). Spatio-temporal vehicle tracking using unsupervised learning-based segmentation and object tracking. *IEEE Robotics and Automation Magazine, special issue on Robotic Technologies applied to Intelligent Transportation Systems, 12*(1), 50-58.

Chen, X., & Zhang, C. (2006). A spatio-temporal database model on transportation surveillance videos. In *Proceedings of the 3rd workshop on Spatio-Temporal Database Management. In conjunction with VLDB06*, Seoul, Korea (pp. 17-24).

Chen, X., & Zhang, C. (2006). An interactive semantic video mining and retrieval platform - application in transportation surveillance video for incident detection. In *Proceedings of the 2006 IEEE International Conference on Data Mining (ICDM)*, Hong Kong, China (pp. 731-740).

Chen, X., Zhang, C., Chen, S.-C., & Rubin, S. H. (2009). A human-centered multiple instance learning framework for semantic video retrieval. *IEEE Transactions on Systems, Man, and Cybernetics*. *Part C*, *39*(2), 228–233.

Cohen, I., & Medioni, G. (1999). Detecting and tracking objects in video surveillance. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO (pp. 319-325).

Cucchiara, R., & Piccardi, M. (1999). Vehicle detection under day and night illumination. In *Proceedings of ISCS-IIA99, Special Session on Vehicle Traffic and Surveillance,* Genoa, Italy.

Cucchiara, R., Piccardi, M., & Mello, P. (2000). Image analysis and rule-based reasoning for a traffic monitoring system. In *Proceedings of IEEE International conference on Intelligent Transportation Systems*, Tokyo, Japan (pp. 119-130).

Dailey, D. J., Cathey, F., & Pumrin, S. (2000). An algorithm to estimate mean traffic speed using uncalibrated cameras. *IEEE Transactions on Intelligent Transportation Systems*, *1*(2), 98–107. doi:10.1109/6979.880967

Dance, S., & Caelli, T. (1993). On the symbolic interpretation of traffic scenes. In *Proceedings of the Asian Conference on Computer Vision,* Osaka, Japan (pp. 798-801).

Day, Y. F., Dagtas, S., Iino, M., & Khokhar, A. (1995). Object-oriented conceptual modeling of video data. In *Proceedings of the 11th International Conference on Data Engineering (ICDE)*, Taipei, Taiwan (pp. 401-408).

Fernyhough, J., & Cohn, A. G., & x Hogg, A. G. (2000). Constructing qualitative event models automatically from video input. *Image and Vision Computing*, *18*(2), 81–103. doi:10.1016/S0262-8856(99)00023-2

Foresti, G. L., Micheloni, C., & Snidaro, L. (2003). Advanced visual-based traffic monitoring systems for increasing safety in road transportation. *Advances in Transportation Studies an International Journal*, *A*(1), 27–47.

Fortuna, J., Schuurman, D., & Capson, D. (2002). A comparison of PCA and ICA for object recognition under varying illumination. In *Proceedings of 16th International Conference on Pattern Recognition*, Quebec City, Quebec, Canada (pp. 11-15).

Grimson, W. E. L., Stauffer, C., Romano, R., & Lee, L. (1998). Using adaptive tracking to classify and monitor activities in a site. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA (pp. 22-29).

Gupte, S., Masoud, O., Martin, R. F. K., & Papanikolopoulos, N. P. (2002). Detection and classification of vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *3*(1), 37–47. doi:10.1109/6979.994794

Haag, M., & Nagel, H. H. (2000). Incremental recognition of traffic situations from video image sequences. *Image and Vision Computing*, *18*(2), 137–153. doi:10.1016/S0262-8856(99)00021-9

Hallenbeck, M. (2005). Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation. Retrieved from http://ops.fhwa.dot.gov/congestion_report/appendixa.htm

Haritaoglu, I., Harwood, D., & Davis, L. (1998). W4 - who, where, when, what: A real-time system for detecting and tracking people. In *Proceedings of IEEE 3rd International Conference on Face and Gesture Recognition,* Nara, Japan (pp. 222-227).

Hartley, R. I., & Zisserman, A. (2004). *Multiple view geometry in computer vision*. Cambridge University Press. doi:10.1017/CBO9780511811685

Hautiere, N., & Aubert, D. (2003). Driving assistance: Automatic fog detection and measure of the visibility distance. In *Proceedings of ITS World Congress*, Madrid, Spain.

Hautiere, N., Labayrade, R., & Aubert, D. (2006). Estimation of the visibility distance by stereovision: A generic approach. *IEICE Trans. Inf., & Syst. E (Norwalk, Conn.)*, *89-D*(7), 2084–2091.

Hu, W., Tan, T., Wang, L., & Maybank, S. (2004). A survey on visual surveillance of object motion and behaviors. *IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, *34*(3), 334–352. doi:10.1109/TSMCC.2004.829274

Hu, W., Xiao, X., Xie, D., & Tan, T. (2004). Traffic accident prediction using 3-D model-based vehicle tracking. *IEEE Transactions on Vehicular Technology*, *53*(3), 677–694. doi:10.1109/TVT.2004.825772

Huang, T., Koller, D., Malik, J., Ogasawara, G., Rao, B., Russell, S., & Weber, J. (1994). Automatic symbolic traffic scene analysis using belief networks. In *Proceedings of National Conference on Artificial Intelligence* (pp. 966-972).

Huang, T., & Russell, S. (1998). Object identification: A bayesien analysis with application to traffic surveillance. *Artificial Intelligence*, *103*, 1–17. doi:10.1016/S0004-3702(98)00067-8

Jung, Y.-K., Lee, K.-W., & Ho, Y.-S. (2001). Content-based event retrieval using semantic scene interpretation for automated traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems*, *2*(3), 151–163. doi:10.1109/6979.954548

Kamijo, S., Matsushita, Y., & Ikeuchi, K. (2000). Traffic monitoring and accident detection at intersections. *IEEE Transactions on Intelligent Transportation Systems*, *1*(2), 108–118. doi:10.1109/6979.880968

Kanhere, N. K., Pundlik, S. J., & Birchfield, S. T. (2005). Vehicle segmentation and tracking from a low-angle off-axis camera. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1152–1157).

Kilger, M. (1992) A shadow handler in a video-based real-time traffic monitoring system. In *Proceedings of IEEE International Workshop on Applications of Computer Vision*, Palm Springs, CA (pp. 11-18).

Lai, A. H. S., & Yang, N. H. C. (2000). Vehicle-type identification through automated virtual loop assignment and block-based direction-biased motion estimation. *IEEE Transactions on Intelligent Transportation Systems*, *1*(2), 86–97. doi:10.1109/6979.880965

Loral AeroSys, IBI Group. Jackson & Lull, and Edwards and Kelcey. (1995). *Requirements/Technology: Final project report.* Report I-95 CC3-95-06. I-95 Corridor Coalition.

Lou, J., Tan, T., Hu, W., Yang, H., & Maybank, S. J. (2005). 3-D model-based vehicle tracking. *IEEE Transactions on Image Processing*, *14*(10), 1561–1569. doi:10.1109/TIP.2005.854495 PMID:16238061

Mikolajczyk, K., & Schmid, C. (2005). A performance evaluation of local descriptors. [PAMI]. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, *27*(10), 1615–1630. doi:10.1109/TPAMI.2005.188 PMID:16237996

Minsky, M. (1986). *The society of mind*. Simon and Schuster.

Ozbay, K., & Kachroo, P. (1999). *Incident management in intelligent transportation systems*. Norwood, MA: Artech House Inc.

Panda, D. P., Jacobson, R. D., & Bernard, B. E. (1996). *Image sensing system technology advancement for adaptive traffic management.* ITS World Congress, Orlando, FL.

Pomerleau, D. (1997). Visibility estimation from a moving vehicle using the ralph vision system. In *Proceedings of IEEE Conference on Intelligent Transportation Systems*, Pittsburgh, PA (pp. 906-911).

Porikli, F. M., & Li, X. (2004). Traffic congestion estimation using HMM models without vehicle tracking. In *Proceedings of IEEE Intelligent Vehicles Symposium*, Parma, Italy (pp. 188-193).

Redden, J. W. P. (2009). *Is train horn noise a problem in your town?* Retrieved from http://www.hanson-inc.com/images/Vision/pdfs/Is-Train-Horn-Noise-a-Problem7-09.pdf

Roman, P. (2002). Visual traffic surveillance using real-time. Diploma Thesis. PRIP, TU Wien.

Sahambi, H. S., & Khorasani, K. A. (2003). Neural-network appearance-based 3-D object recognition using independent component analysis. *IEEE Transactions on Neural Networks*, *14*(1), 138–149. doi:10.1109/TNN.2002.806949 PMID:18237997

Schölkopf, B., Platt, J. C., Taylor, J. S., & Smola, A. J. (1999). *Estimating the support of a high-dimensional distribution*. Microsoft Research Corporation Technical Report MSR-TR-99-87, (1999).

Sezer, O. G., Ercil, A., & Keskinoz, M. (2005). Subspace based object recognition using support vector machines. In *Proceedings of European Signal Processing Conference (EUSIPCO)*, Antalya, Turkey.

Smith, C. E., Richards, C. A., Brandt, S. A., & Papanikolopoulos, N. P. (1996). Visual tracking for intelligent vehicle-highway systems. *IEEE Transactions on Vehicular Technology*, *46*(4), 744–759. doi:10.1109/25.543744

Stauffer, C., & Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO (pp. 246-252).

Tang, S., & Gao, H. (2005). Traffic-incident detection algorithm based on nonparametric regression. *IEEE Transactions on Intelligent Transportation Systems*, *6*(1), 38–42. doi:10.1109/TITS.2004.843112

Taniguchi, H., Nakamura, T., & Furusawa, H. (1999). Methods of traffic flow measurement using spatio-temporal image. In *Proceedings of International Conference on Image Processing*, Kobe, Japan (pp. 16-20).

Toyama, K., Krumm, J., Brumitt, B., & Meyers, B. (1999). Wallflower: Principles and practice of background maintenance. In *Proceedings of the 7th International Conference on Computer Vision (ICCV'99)*, Crete (pp. 255-261).

Veeraraghavan, H., Masoud, O., & Papanikolopoulos, N. P. (2003). Computer vision algorithms for intersection monitoring. *IEEE Transactions on Intelligent Transportation Systems*, *4*(2), 78–89. doi:10.1109/TITS.2003.821212

Weill, R., Worton, J., & Garcia-Ortiz, A. (1998). Traffic incident detection: Sensors and algorithms. *Mathematical and Computer Modeling Journal*, *27*(9-11), 257–291. doi:10.1016/S0895-7177(98)00064-8

Wu, W., Zhang, Q., & Wang, M. (2001). A method of vehicle classification using models and neural networks. In *Proceedings of IEEE 53rd Vehicular Technology Conference*, Rhodes, Greece (pp. 3022-3026).

Xie, D., Hu, W., Tan, T., & Peng, J. (2004). Semantic-based traffic video retrieval using activity pattern analysis. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, Singapore (pp. 693-696).

Yang, L., Johnstone, J., & Zhang, C. (2007). A multi-camera approach to vehicle tracking based on features. In *Proceedings of the 9th IEEE International Symposium on Multimedia Workshops*, Taichung, Taiwan (pp. 79-80).

Yang, L., Johnstone, J., & Zhang, C. (2009). Vehicle tracking from disparate views. In *Proceedings of the 2009 IEEE International Conference on Multimedia and Expo*, Cancun, Mexico (pp. 534-537).

Zhang, C., Chen, X., & Chen, W.-B. (2006). A PCA-based vehicle classification framework. In Proceedings of IEEE International Workshop on Multimedia Databases and Data Management. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE 2006)*, Atlanta, GA.

Zhu, Z., Xu, G., Yang, B., Shi, D., & Lin, X. (2000). VISATRAM: A real-time vision system for automatic traffic monitoring. *Image and Vision Computing*, *18*(10), 781–794. doi:10.1016/S0262-8856(99)00046-3

Chengcui Zhang is an Associate Professor of computer and information sciences at the University of Alabama at Birmingham. Her research has been funded by Natural Science Foundation, eBay Inc., IBM, and NIH. She has authored and co-authored more than 120 research papers focusing in the areas of multimedia databases, multimedia data mining, bioinformatics, and geographic information systems. She is currently serving as a program co-chair of the 2013 IEEE International Conference on Information Reuse and Integration (IRI'13). She was a program co-chair for 2011 IEEE International Symposium on Multimedia and the 2012 International Conference on Information Reuse and Integration. She has served on more than 90 program committees of major conferences and workshops in multimedia and database systems. Zhang also serves on the editorial boards for several international journals, including International Journal of Multimedia Data Engineering and Management *and* International Journal of Computer and Informatics.