Active Traffic Monitoring Through Large Scale Processing of Aerial Camera Array Networks

Final Report

Authored by

Wayne Sarasua, Ph.D., P.E.
sarasua@clemson.edu
Clemson University

Xi Zhao
xiz@whut.edu.cn
Wuhan University of Technology

William J. Davis, Ph.D., P.E.
jeff.davis@citadel.edu
The Citadel

Contact information
Wayne Sarasua, Ph.D., P.E.
310-B Lowry Hall, Clemson University
Clemson, South Carolina 29634
Phone: (864) 656-3318; *E-mail:* sarasua@clemson.edu

December 2019

Center for Connected Multimodal Mobility (C^2M^2)

200 Lowry Hall, Clemson University
Clemson, SC 29634

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by the Center for Connected Multimodal Mobility (C^2M^2) (Tier 1 University Transportation Center) Grant, which is headquartered at Clemson University, Clemson, South Carolina, USA, from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

Non-exclusive rights are retained by the U.S. DOT.

ACKNOWLEDGMENT

The authors would like to thank C^2M^2 for funding this project and the University of Tennessee Southeastern Transportation Center (STC) for funding prior research related to this work. The authors would also like to thank Persistent Surveillance Systems *for providing the persistent monitoring dataset*.

Technical Report Documentation Page

1. Report No.	**2. Government Accession No.**	**3. Recipient's Catalog No.**

4. Title and Subtitle Active Traffic Monitoring Through Large Scale Processing of Aerial Camera Array Networks	**5. Report Date** March 2020
	6. Performing Organization Code

7. Author(s)
Wayne Sarasua, Ph.D.; ORCID: 0000-0002-3910-446X;
Xi Zhao; ORCID: 0000-0001-9105-5872;
William Davis, Ph.D.; ORCID: 0000-0002-3812-8654
	8. Performing Organization Report No.

9. Performing Organization Name and Address Clemson University 310-B Lowry Hall, Clemson, SC 29634	**10. Work Unit No.**
	11. Contract or Grant No. 69A3551747117
12. Sponsoring Agency Name and Address Center for Connected Multimodal Mobility (C^2M^2) Clemson University 200 Lowry Hall, Clemson, SC 29634	**13. Type of Report and Period Covered** Final Report (March 2018 – November 2019)
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This research focuses on automated processing of video from a persistent surveillance camera array to extract traffic data from a 25 square mile area. In previous research, we developed an automated traffic surveillance system capable of processing aerial camera array imagery to extract valid and useful traffic data for diverse applications. In this research, we continue to improve the system's capability by adding a novel multiple hypothesis tracking capabilities to improve vehicle tracking in congested traffic and adding a location identification algorithm to map vehicles throughout a network. Our evaluation has shown that the proposed system is capable of collecting speed, density, and volume data with an acceptable level of accuracy for many applications. The mapping of vehicles for a sample area was also successful. With further research, improved video preprocessing, enhanced resolution, and a higher frame rate, the accuracy of tracking vehicles can be improved significantly which will eventually allow the envisioned system to be able to accurately map the location of all vehicles throughout a camera array image sequence. A digital real-time "traffic map" created by the envisioned system will provide a robust data set where data mining methods could be applied to enhance traffic management and provide data for a variety of traffic studies. A connected vehicle camera array application can open up plenty of possibilities in real-time traffic surveillance where erratic drivers can be identified automatically and warnings or even shut down commands can be sent to the erratic vehicles. The active sensing capability of such a system can potentially prevent some incidents from occurring thereby increasing safety and reducing incident induced traffic congestion.

17. Keywords Traffic Surveillance, Aerial Camera Arrays, Vehicle Tracking, Vehicle Mapping, Computer Vision, Deep Learning	**18. Distribution Statement** This report or any part of this report is restricted to publish until prior permission from the authors.

19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	27	NA

Table of Contents

List of Tables

List of Figures

EXECUTIVE SUMMARY

Traditional data collection sensors serve as the primary source for monitoring traffic and collecting data at fixed points strategically located throughout a traffic network. Collected data commonly includes traffic volume, time-mean speed, vehicle classification, and occupancy. Conventional sensors are unable to track vehicles through a network even though attainable results would serve to support widescale traffic applications through the collection of microscopic mobility parameters for individual vehicles. Vehicle tracking offers the potential of generating enumerated values of stopped delay, running speeds, acceleration and deceleration, and other useful driver behavior characteristics. With the recent technological advent of aerial digital camera arrays, wide view high-resolution video from aircraft operating overhead can be used as a valuable data source for vehicle tracking within a pre-established roadway network through the use of appropriately robust post processing algorithms. Traditional manual inspection affords the possibility of tracking any vehicle of interest for targeted safety and security purposes, however, applications of this method have been limited due to cumbersome, time-consuming and resource intensive workforce requirements.

Automated processing of video from a camera array performed for the purpose of extracting network traffic data has not been extensively explored due in part to the novelty of this innovative approach and perceived systematic complexities. Technological challenges include video stabilization, image registration and rectification, object recognition, and low-frame-rate tracking. Previous work conducted by the research team focused on the development of an automated traffic surveillance system capable of processing aerial camera array imagery for the purpose of extracting valid and useful traffic data for a diverse list of applications specifically benefitting traffic data monitoring and traffic safety. This research initiative serves as a next step in continuing in improving the system's algorithms by adding a novel tracking algorithm to further advance vehicle tracking capabilities and adapting a location identification algorithm to map vehicles traveling within an established pre-defined transportation network.

Our previous approach that combined individual vehicle feature-based tracking with vehicle detection based on deep learning provided promising results for collecting speed, density, and volume along uninterrupted flow road segments located throughout a 25 mi^2 coverage area. Improvements to the tracking system to improve data collection in congested traffic are based on an innovative analytical approach that combines multiple hypothesis tracking (MHT) with a kinematics and appearance model (KAM) to improve vehicle tracking capabilities. Research findings indicate the use of a combined MHT-KAM processing approach offers the capability of producing promising analytical results for individual vehicle tracking, even under commonly occurring problematic conditions of dense traffic flows.

Research evaluations indicate the proposed system provides the capability of collecting speed, density, and volume data within an acceptable level of accuracy for a variety of practical transportation network performance applications. Additionally, mapping of individual vehicle paths was determined achievable when an appropriate number of control points are used to support the processing algorithms. With further

research, improved video preprocessing, enhanced resolution, and higher frame rates, the accuracy of vehicle tracking can be further improved, ultimately allowing the envisioned system to accurately map all vehicle locations and paths throughout a camera array image sequence. From an increase in mapping accuracy of vehicle location within a spatio-temporal context, additional traffic parameters can be extracted microscopically, including intersection turning movements, traffic signal phasing, and timing, vehicle queues, identification of erratic drivers, trip origin-destination, and route decision making. Furthermore, the resulting high-functioning analytical system could facilitate traffic management through traffic effective data mining routines. Traffic data mining techniques offer the potential to identify reliable recurring patterns extracted from large traffic data sets with less complexity than current approaches. The ability to accurately predict traffic patterns and associated vehicle travel path parameters can facilitate and support the creation of data and information to more efficiently manage and operate traffic networks.

Digital real-time “traffic maps” created from widescale implementation of proposed vehicle tracking algorithms can provide a robust data set where data mining techniques could be applied to enhance traffic management, operations, and provide a rich source of data for a variety of insightful traffic studies. Data produced from network-wide vehicle tracking systems can potentially facilitate or even automate the creation and calibration of microsimulation and activity-based travel demand forecasting models. Areawide and regional mobility models commonly require many months, or even years, to develop and calibrate, from which the consequential results are used to plan and design the transportation network of the future.

Lastly, a connected vehicle camera array application can provide and support an array of possible beneficial safety applications in real-time traffic surveillance where erratic drivers can be identified automatically, and warnings or even shut down commands could be sent to the erractic vehicles. The active sensing capability of such a revolutionary vehicle tracking system can potentially prevent some traffic incidents from occurring thereby increasing safety and proactively mitigating incident induced traffic congestion.

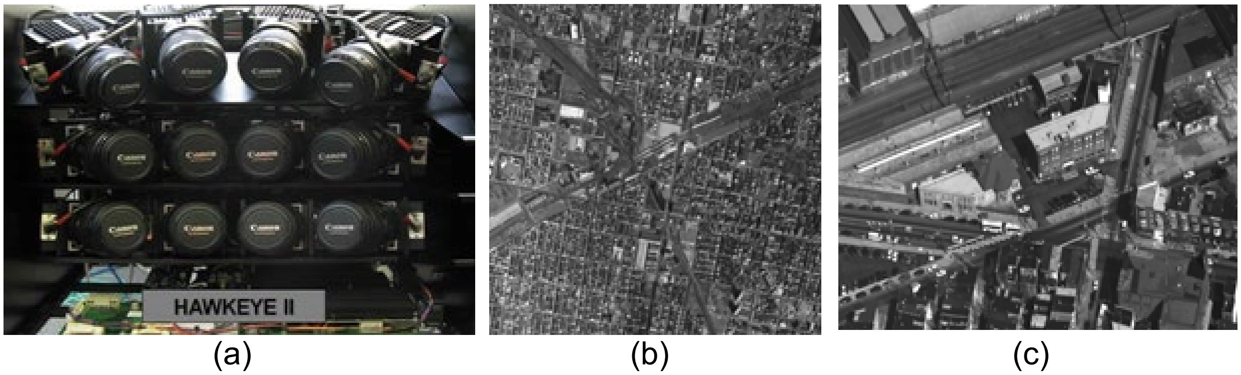
CHAPTER 1

Introduction

1.1 Description of Problem

Conventional data collection sensors (microwave radar, infrared devices, piezos, road tube sensors, and inductive loop detectors) are frequently used to monitor traffic and collect data at fixed points throughout a roadway network. Data commonly include time-mean speed, vehicle classification, traffic volume, and vehicle occupancy. Unfortunately, conventional sensors are unable to track vehicles traveling through a network, even though results from this technological advancement would offer broad traffic applications through determining microscopic parameters for individual vehicles. Vehicle tracking offers the potential of calculating values for the stopped delay, running speeds, acceleration and deceleration, and other useful driver behavior characteristics. Many existing ITS (Intelligent Transportation Systems) network applications using digital cameras mounted along roadside locations were developed to collect traffic image data for various applications including collision detection (Saunier and Sayed 2007) or driving behavior (Tsai et al. 2011). With the relatively recent advent of aerial digital camera arrays, the possibility to record high-resolution video for a wide field of view from aircraft overhead has become an achievable reality. Fig. 1 (a) provides an illustration of a camera array, which is configurable and adaptable for a variety of aircraft. Coverage of this aerial system is largely dependent on altitude, nevertheless, video captured from a typical device configuration, illustrated in Fig. 1 (b) and (c), covers approximately 25 square miles (5-miles wide by 5-miles long).

Algorithm-based automated computer processing programs of video from readily available commercial camera arrays used for the purpose of extracting traffic data have not been extensively explored due in part to the novelty of this innovative approach and perceived systematic technological challenges including video stabilization, image registration, image rectification, object recognition, and low-frame-rate tracking. Previous work conducted by the research team focused on the development of an automated traffic surveillance system capable of processing aerial camera array imagery to extract reliable, repeatable, and beneficial traffic data for diverse applications including traffic monitoring operations and safety. This follow-on research initiative serves as a logical next step in continuing to improve the system’s capability by expanding to include a novel multiple hypotheses tracking to improve vehicle tracking and adding a location identification algorithm to better map vehicles traveling throughout a per-determined roadway network. Precise mapping of vehicles combined with vehicle-to-system (active cameras) communication will allow one-to-one correspondence allowing vehicles to be tracked throughout an active camera array image. Vehicle-to-system communication can benefit the traveling public, optimize network operation, support traffic management, and potentially reduce persistently occurring incidents by allowing authorities to shut down roadways under extreme conditions when necessary to resolve unsafe conditions resulting from confirmed miscreant erratic drivers.



(a) (b) (c)

Fig. 1. (a) Aerial camera array; (b) High-resolution image; (c) Zoomed view (images from Persistent Surveillance Systems, Dayton, OH, www.pss-1.com)

CHAPTER 2

Literature Review and Background

2.1 Previous Advances in Digital Image Processing of Aerial Video of Traffic

A variety of algorithms have been programmed and adapted to detect vehicles automatically from aerial images, however, the majority of testing conducted was based upon images from a single camera, or captured from low altitudes within a limited small field of view. For example, University of Arizona researchers used a computer vision-based approach to collect traffic parameters from one low-resolution camera (720 × 480) mounted to a helicopter flying at an altitude of under 305 m (1,000 feet), providing a field of view of less than 244 m (800 feet) (Angel et al. 2002). Subsequent research led to the creation of an innovative software program dubbed “Tracking and Registration of Airborne Video Image Sequences” (TRAVIS) that can extract vehicle positions from airborne imagery for use in the analysis of microscopic traffic behavior. Data input for the TRAVIS software program was comprised of a sequence of images captured from airborne video. TRAVIS registers image sequence to an initial common reference frame, detects vehicles contained within the images, and tracks vehicles through the image sequence using a blob tracking algorithm. The output from TRAVIS links a sequence of pixel coordinates for vehicles as the basis for creating tracks extending through an image sequence (Hickman and Mirchandani 2006). Follow-up research was conducted by Du and Hickman (2012) to improve vehicle detection, reduce the probability of false detection, and optimize computation runtime by masking areas located outside roadway limits. Researchers also improved tracking algorithms to better accommodate vehicles with limited contrast relative to roadway pavement surface coloring. The dataset created for this investigative research work was collected from an individual camera mounted on a helicopter, providing a 0.4-m (1.3-ft) pixel size.

Previous work on vehicle detection and vehicle tracking can be categorized into three distinct categories: a) feature detectors, b) background subtraction/modeling, and c) machine learning. A variety of feature detectors, such as KLT and SIFT, have been widely applied to either track vehicles or model the digital appearance of vehicles for detection and tracking purposes. Moon et al. (2002) created a vehicle detector algorithm by combining four elongated edge operators. Detection performance was considerably impacted by camera angles and illuminations. Kim and Malik (2003) presented a model-based 3D vehicle detection routine using description-based online feature algorithms; concluding that this analytical approach outperformed Zhao and Nevatia’s algorithm (2003). Hinz (2005) modeled vehicles at a local level through the use of 3D wireframe representations on a global-scale by grouping vehicles within queues. Beneficial results from the use of this approach include processing did not rely on external information and data were not limited to constrained environments. Palaniappan et al. (2010) presented an interactive tracking system based on feature detection using appearance modeling and motion prediction. Cao et al. (2012) proposed a framework for vehicle detection and robust to partial occlusion vehicle tracking based on KLT features. Pelapur et al. (2012) presented a feature tracking system that used a pre-established adaptive set of feature descriptors with posterior fusion modeling.

Background subtraction and object modeling are frequently used requiring either a stabilized camera or with a system capable of providing accurate image registration. Reinartz et al. (2006) applied background subtraction techniques to identify vehicles and image patch correlation to associate vehicle matching between frames. This approach experienced reliability issues when mistakenly detecting pedestrians as vehicles and errantly grouping vehicles operating in close proximity. Xiao et al. (2010) proposed a probabilistic relation graphical function to combine a vehicle behavior model with a road network as the basis for vehicle detection and tracking within the context of a wide area video. Shi et al. (2013) proposed the use of a maximum consistency context model for background subtraction-based multiple object tracking by leveraging discriminative power and robustness embedded into scenario processing. Prokaj and Medioni (2014) demonstrated the application of a multiple object tracking approach using two trackers in parallel; one based on detection by background subtraction, and one based on the use of a template-based regression tracker. Saleemi and Shah (2014) presented a framework capable of tracking thousands of vehicles based on input from low frame rate aerial videos using background modeling. Chen and Medioni (2015) applied two methods of adapting the background model to produce more accurate background subtraction, specifically addressing the presence of parallax (e.g. interference from adjacent multi-story buildings). For this approach, the first method used a dense 3D model of the landscape and the second method was predicated upon an epipolar flow constraint.

Machine learning constitutes a widely relied upon approach for vehicle detection, commonly employed in association with either feature detectors, and/or background subtraction modeling. Zhao and Nevatia (2003) introduced a passenger vehicle detection system by modeling passenger vehicles as 3-D objects through the use of Bayesian network parameters. Nguyen et al. (2007) established an automatic car detection framework trained for three types of features through the use of AdaBoost (Adaptive Boosting), a machine learning meta-algorithm commonly used in conjunction with other types of learning algorithms to improve processing reliability and performance. Tuermer et al. (2010) used a preprocessing algorithm to constrain the limits of search space to develop a reliable detector using Real AdaBoost with HoG (Histograms of Oriented Gradients) features. Cheng et al. (2012) introduced a pixel-wise classification approach in which the Dynamic Bayesian Network (DBN) variables were constructed for use in determining pixel classification.

Most vehicle tracking approaches are based on vehicle detections that use visual information from digital images to initialize the tracker or support the vehicle tracking algorithm process to match correspondences between adjacent frames. However, other analytical process approaches for vehicle tracking and vehicle detection treat data input constructs as mutually dependent. Kalal et al. (2012) proposed the use of a TLD (Tracking-Learning-Detection) framework as the basis for establishing a vehicle tracking algorithm that generates training data for improving the detector by initializing and re-initializing the tracker simultaneously.

None of the vehicle detection and vehicle tracking approaches cited in the literature have been adapted to process datasets produced from aerial camera arrays in a similar manner as proposed in this research initiative. Currently, the application of these methods is generally considered not yet feasible for adaptation to provide near real-time processing due to inherent characteristic limitations of expansive data collected by aerial camera arrays that include pixel size, image illumination, and related consequential challenges of data processing reliant upon orthorectification and mosaicking.

2.2 Multiple Hypotheses Tracking

An algorithm identified by Reid (1979) dubbed Multiple Hypothesis Tracking (MHT) has been widely adopted as a preferred method for data association in conducting multiple target tracking (MTT) tasks. Practical implementations of MHT are challenging due to an underlying high degree of computational complexity associated with this approach. However, through enhanced implementation methods, Cox and Hingorani (1996), conducted work that effectively upgraded computational hardware capabilities allowing practical real-time implementations to become an achievable reality (Blackman et al. 2001). Applications of MHT have been studied by researchers in both the radar detection and the computer vision professional communities, however, technological advances have not yet been included in the mainstream of the latter (Kim et al. 2015). In addition to MHT, computer vision and image processing researchers have focused on exploring other data association methods reliant upon the use of visual information for multiple object tracking (MOT) task routines (Zhang et al. 2008; Shitrit et al. 2011; Luo et al. 2014; Leal-Taixé et al. 2017). However, a noteworthy supposition concludes none of these MHT algorithms have been used to consider vehicle tracking within the application context of wide-area camera array images, as proposed in this research initiative.

Conventional MHT approaches commonly employ a singular Kalman filter for maintaining and updating projected vehicle tracks by estimating point-by-point track states typically including kinematic measurements such as position, velocity, and acceleration. A further expansion of this technique that involves an interacting multiple model (IMM) using multiple Kalman filters has received wide-scale acceptance based on high performance in tracking maneuvered targets (Blackman 2004). IMM deploys differing filter models in parallel which are specifically selected to harmonize with different types of maneuvers. The combined state estimated values and covariance are computed by either switching across available outputs generated from different Kalman filters or correspondingly, from weighted compositions using them (Genovese 2001).

Effective surveillance applications generally necessitate the capability of tracking multiple targets, as a result, MTT is widely considered one of the most important tasks for surveillance system processing. Similarly, most analytical frameworks incorporate a) sensor modules including radar, infrared and sonar, report measurements, or detections, b) MHT modules, c) MHT modules implement sensor data associations, and d) output tracking results for MTT. These types of high-functioning frameworks have been applied across numerous application fields including track confirmation, agile beam radar, missile defense systems, and ground target tracking which is largely considered the most challenging technical application (Blackman 2004).

Security restrictions, governmental confidentiality regulations, and industry redaction policies have constrained or prevented publication of much of the previous analytical work and advancements related to vehicle tracking (Blackman 2004). Nevertheless, Arambel et al. (2004) presented a synopsis report proposing an automated video-based ground targeting system for unmanned aerial vehicles (UAVs). This system used background subtraction and site modeling to extract features and measurement values to an MHT module for tracking multiple ground targets simultaneously. Spraul et al. (2017) proposed a similar analytical approach for persistent vehicle tracking within the context of wide-area imaging. The framework incorporated a median background subtraction approach for reliable vehicle detection and applied an adaptive MHT methodology that integrated several extensions for optimally effective vehicle tracking.

This research initiative uses foundational detection and tracking knowledge to specifically explore how a novel and innovative MHT approach can improve vehicle tracking system accuracy beyond the previously established algorithm approaches for the extraction of data from aerial camera arrays. This next step improvement focuses on the need to a) further advance vehicle tracking capabilities, b) adopt a location identification algorithm to map vehicles traveling within a defined transportation network, and c) to facilitate active camera communication with tracked vehicles within established limits of a pre-defined transportation network.

2.3 Persistent Monitoring Using Aerial Camera Arrays

Airborne video imaging using camera arrays comprises an emerging and newly evolving technological field enabling persistent coverage of large geographical regions dependent upon platform altitude and aerial camera array configuration (Palaniappan et al. 2010). An aerial camera array combined with computer vision techniques allows the creation of a virtual view of the region being monitored. Surveillance systems commonly deploy a circular flight path while cruising at a constant altitude. As an aircraft maneuvers overhead, the camera array is adjusted simultaneously to maintain a constant orientation of the camera array fixed upon a pre-established point visible on the ground. The continuous coverage of geographic region for an array can remain constant for a number of hours contingent upon aircraft range and flight time.

Numerous potential network applications of this system are possible within the public sector, however, most current common applications are deployed by law enforcement and event security contractors. Under these applications, vehicles or individuals of interest are manually tracked, and connected information of concern or relevance is relayed to law enforcement personnel located on the ground. Automated tracking of vehicles using data captured from aerial camera arrays has not yet been accomplished except in previously conducted work by the researchers (Zhao et al. 2017). Palaniappan et al. (2010) conclude that while potential applications of automated processing of aerial camera array data provide promising analytical advancements, a number of persistent challenges remain as consequential impediments to broader commercial adoption. Noteworthy obstacles to system implementation include a) need for improved camera calibration, b) better estimation of platform dynamics, accounting for lighting variability, and c) seamless image mosaicking. Additionally, existing approaches previously discussed in this literature review are not capable of processing aerial camera array data as a result of requirements for higher resolution imagery needed to analyze connecting adjacent pixels surrounding vehicle locations within static images; whereas resolution of aerial imagery, when collected at high altitude, is too poor (distorted, pixelated, grainy) to be able to distinguish vehicles from other visible objects contained within the static images. Other approaches have successfully used either background subtraction or frame differencing, which cannot be applied to aerial camera array videos unless adjacent image frames are absolutely stabilized.

2.4 Mapping of Vehicles

Limited research is readily available for citation from published technical literature on the mapping of vehicles from airborne imagery. Maturana et al. (2017) described a semantic mapping system applied in conjunction with a deep learning 2D semantic segmentation algorithm that uses an occupancy grid mapping to identify and determine locational metrics of vehicles based upon imagery recorded from aerial drones. Sengupta et al. (2012) presented a processing system using digital images and depth to create 3-D segmentation for street-level imagery. Other researchers (Kunda et al. 2014; Savinov et al. 2017) used monocular imagery for semantic segmentation and 3-D reconstruction. For the application methodology of this research initiative, the use of a planar occupancy triangulated network approach is proposed, which relies upon extensive use of control points to calculate vehicle location metrics extracted from a camera array video. The research team believes this approach provides an optimally efficient method in contrast to comparable 3-D reconstruction methodologies.

CHAPTER 3

Algorithm Development, Methodology and Results

3.1 Persistent monitoring dataset

The approach and methodology adopted for this research focus on implementing algorithms, testing image processing, and applying computer vision techniques using a persistent monitoring dataset. The aerial image mosaic dataset used as the basis for conducting this research was provided by Persistent Surveillance Systems (PSS). The PSS data was collected using a HawkEye II camera array system. Findings from the research and accompanying data analysis identified numerous technical challenges with processing the video sequences. These challenges need needs to be addressed and resolved for automated processing algorithms to commercially produce repeatable and reliable results in the future. An overview of technical challenges in applying this research methodology is summarized as follows:

1. Sub-images from different cameras in the array were not exactly aligned after the images were digitally seamed, as shown in Fig. 2 (b).
2. Images were not completely stabilized, as shown in Fig. 2 (a) and (b). It should be noted the image in Fig. 2 (b) was shifted to the left from Fig. 2 (a).
3. Illumination of sub-images from different cameras in the array was not consistent, as illustrated in Fig. 2 (d).
4. Illumination in a single sub-image created from an individual camera in the array was not consistent, as illustrated in Fig. 2 (d).
5. Video images were preprocessed by the monitoring system. The preprocessing software was proprietary and researchers do not have access nor insight to specific details of how image data was analytically preprocessed.
6. Image resolution was low, about 0.5 m by 0.5 m per pixel, resulting in few detailed features being available to detect vehicles using a static image. As illustrated in Fig. 2 (c), difficulties occurred in distinguishing vehicles (right two) and other objects (left two).
7. Images comprised very large data files, 16384×16384 pixels for 8.05 km by 8.05 km (5-miles by 5-miles). Some image processing and computer vision hardware cannot support immense data requiring extremely high-level resolution.
8. The frame rate was only 1 Hertz. On a freeway, a vehicle traveling at 97 km/h (60 mph) travels 26.8m in a second (53.6 pixels), making associated linkages between corresponding adjacent image frames challenging.
9. The amount of data was tremendously large. A single compressed frame was 40-50 MB and a single uncompressed frame was 1 GB in size, as a result, data processing algorithms require extremely high computational and associated memory efficiency to achieve near real-time execution processing.

By overcoming these technical challenges, automated persistent traffic monitoring can be achieved, and diverse traffic data can be extracted with adequate algorithms for vehicle detection and vehicle tracking.

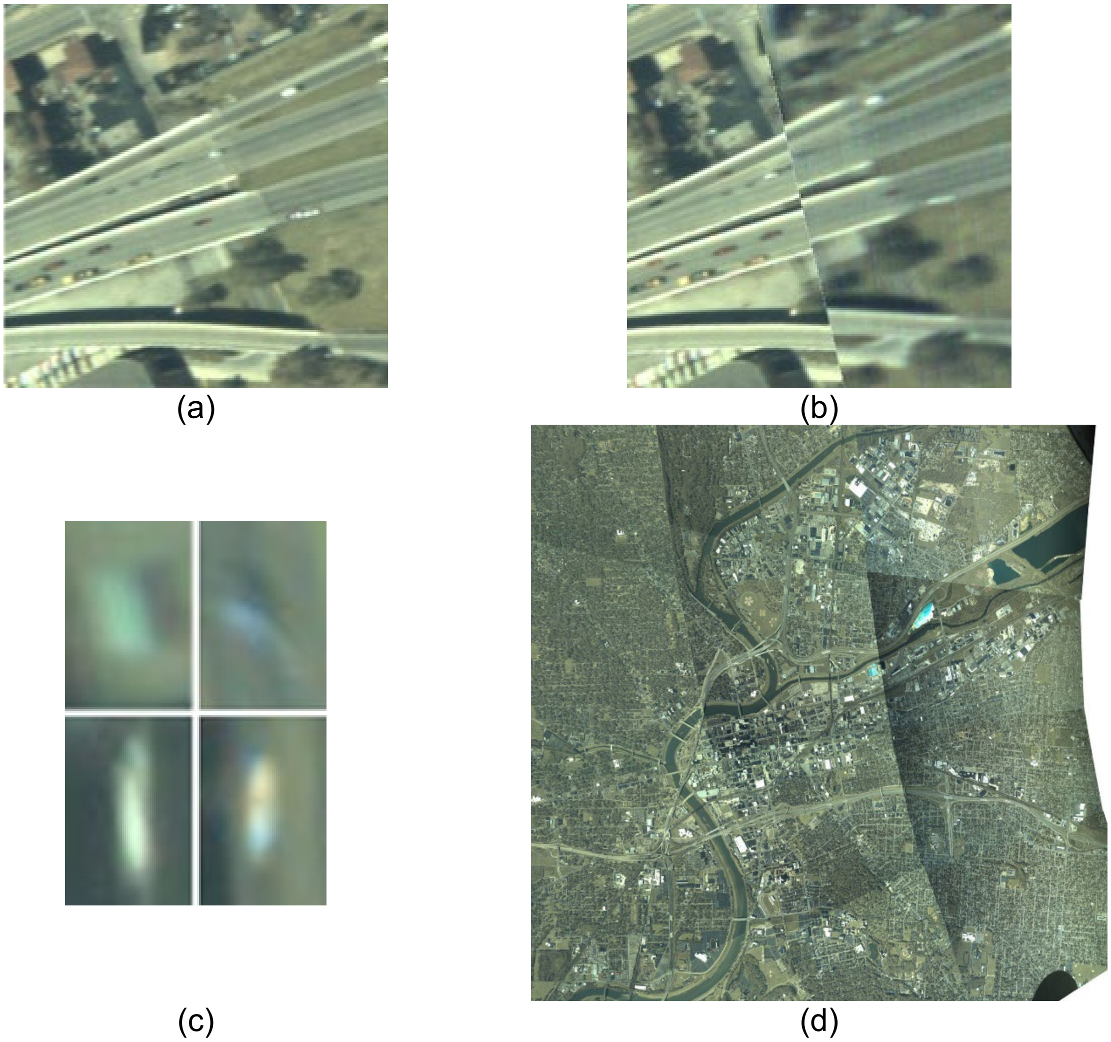


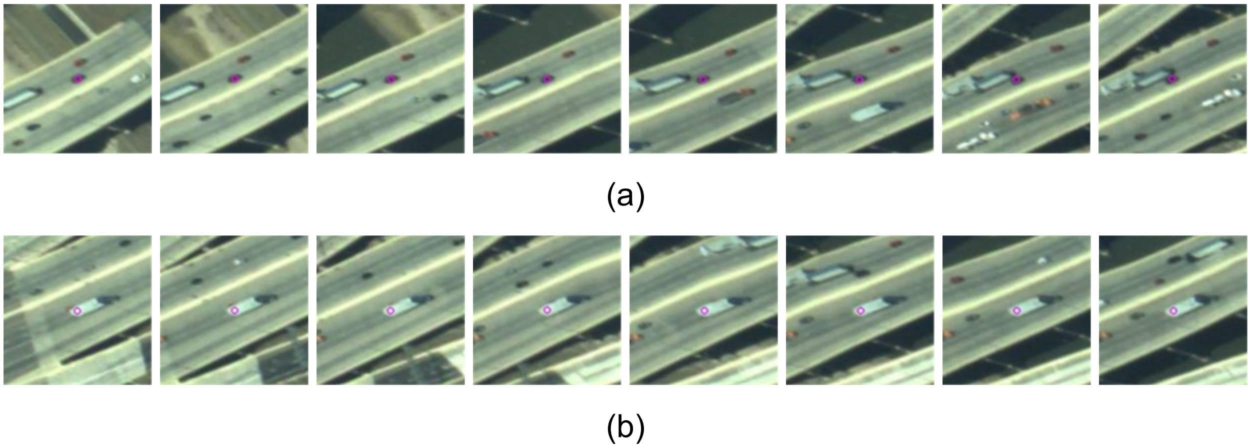
Fig. 2. (a) A roadway view from an image frame; (b) The same roadway view from an adjacent frame, showing undesirable digital seam; (c) Zoomed portions of a frame illustrating vehicles (right two) and non vehicular objects (left two); (d) A mosaic frame comprising the input data.

3.2 Overview of Previous work

Previous work conducted by the research team experimented with multiple approaches to illustrate the feasibility for establishing a reliable basis for producing a high-functioning traffic surveillance system to collect traffic data based on aerial camera arrays (Zhao et al. (2017). In the following sections, a summary of two promising approaches will be presented including 1) a feature-based vehicle tracking framework, and 2) a vehicle tracking framework based on deep learning. The testing was conducted for a selected number of sample vehicles; however, both approaches are scalable to include all of the vehicles visible within an entire image mosaic.

3.2.1 A feature-based vehicle tracking framework

The vehicle tracking framework developed in the previous research was dependent upon feature detection and matching. The method constitutes a heuristic-based approach that clusters features representing a tracked object predicated upon the requirement the object's appearance does not morph dramatically between consecutive frames. Thus, a representative sample object, or vehicle, emerges as identifiable across adjacent image frames by explicitly matching specific corresponding features of the object through the use of multiple methods (detection, matching, and filtering). Sample feature detection and matching vehicle tracking illustrated in Fig. 3 (a) and (b) are based on a SURF (Speeded Up Robust Features) detector and descriptor. A simplified constant acceleration model was applied to estimate speed, acceleration, and orientation for use in determining a vehicle's predicted location and assisting the targeted feature search and matching range routines. As demonstrated in Fig. 3 (a), a black vehicle (identified with a magenta circle) is tracked within the traffic flow as the vehicle crosses a bridge. In Fig. 3 (b), a white vehicle is tracked using a similarly successfully image-based routine. Results indicate the consistent application of SURF-based tracking methods were capable of successfully tracking most vehicles when accompanying images exhibited a stable identifiable visual appearance across numerous recurring roadway network conditions.

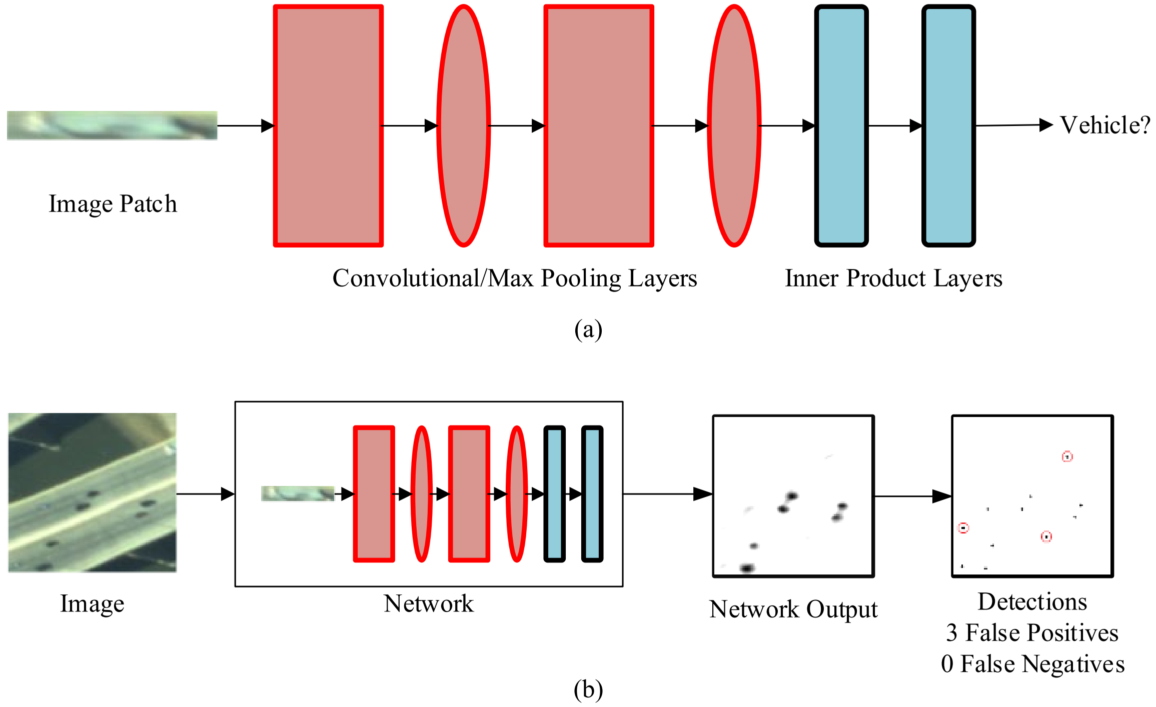

(a)

(b)

Fig. 3. (a) Tracking of a black vehicle through analysis of multiple adjacent image frames); (b) Similarly tracking of a white vehicle.

3.2.2 Deep Learning-Based Vehicle Detector

Deep learning-based vehicle detector algorithms were applied through the use of a customized deep learning convolutional neural network (CNN) performed in affiliation with the Caffe Library, developed at the University of California at Berkeley (Jia et al 2014). The Caffe Library provides a customized network application that includes an embedded ability to learn and operate on a GPU (Graphics Processing Unit), speeding up the data process running time. CNN is designed to detect whether an image patch includes the presence of a vehicle, or not. Fig. 4 (a) provides an illustration of the detection operation within the network from a single image patch. The image patch was processed through a two-stage procedure: 1) convolutional/max-pooling layers (illustrated in red), and 2) two inner product layers (illustrated in blue). Fig. 4 (b) presents a schematic depiction of the vehicle detector analysis procedure for an entire image by splitting the image into numerous separate image patches, each of which was processed through the detector. Patches were pulled from the image centered about each pixel and tested using the detector. This produced a score for each pixel indicating the likelihood that the pixel contains a vehicle. In Fig. 4 (b), this process was graphically illustrated through the use of image-grayscale shading where the level of brightness provides an indication of the likelihood of vehicle presence. Non-maximal suppression was used to identify the peak in this grayscale image shading and associated peaks are designated as vehicles. For the example input image represented in Fig. 4 (b), all eight vehicles were detected, three of which were false positives. Removal of false positives can be systematically accomplished through the implementation of the vehicle tracking steps. More specific details of the CNN deep learning detector were described and delineated by Zhao et al. (2017).



(a)

(b)

Fig. 4. (a) Diagram of the CNN deep learning detector operating on an image patch; (b) Diagram of the detector being used.

Deep learning-based vehicle detector network algorithms provide a sophisticated functional method for effectively matching vehicle detections between digital image frames. Application of the Caffe library supports and facilitates training of Siamese Networks so that two image patches can be simultaneously processed through an identical set of layers and digitally compared. This algorithm capability allows training of network processing routines to become more adept at distinguishing between different vehicles. The resulting output from this network can be reflected as a number, quantitively indicating how likely the two image patches are actually representing the same vehicle.

For the purpose of carefully testing this vehicle detector configuration, a simple tracking system was implemented that included acceptable robust functionality necessary to produce reliable and repeatable results. Fig. 5 provides an illustration of two vehicles being tracked across image seams and differing lighting illumination variations. It is important to note that each image provided in Fig. 5 represents an individual image frame. A closeup view of a vehicle of interest is provided in the top left-hand corner, and a closeup view of a second vehicle of interest is shown in the bottom right-hand corner. In viewing these images, it should be noted vehicles advance nearly 100-feet between frames when traveling freeway speeds.

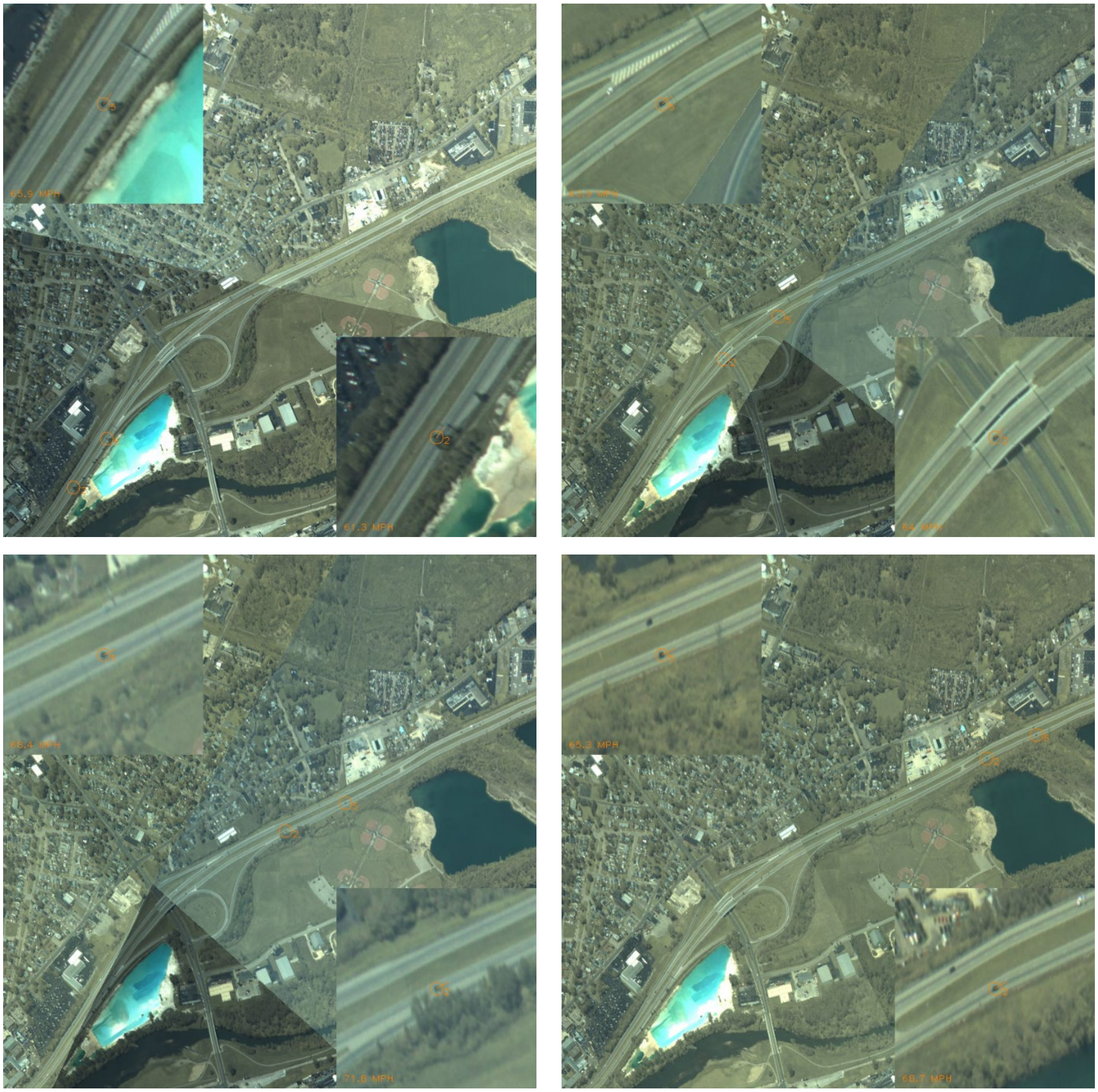


Fig. 5. The deep learning detector tracking two vehicles across the seams and lighting changes.

3.2.3 Data Extraction and Testing

Testing results indicate that the use of the deep learning approach produced promising results inferring that this method provides the potential capability of achieving reliable and repeatable performance. Due to data quality limitations, vehicle detection for individual frames determined independently of other adjacent frames was not achievable even with the benefit of extensive Caffe Library training. Thus, a combined method including a feature tracking approach with deep learning was determined to produce the most reliable vehicle detection and vehicle tracking capabilities. This cooperative approach was used to extract traffic parameters including speed, density, and volume, as identified through the deep learning vehicle detector algorithm to locate potential vehicles, and feature-based tracking algorithm to match the vehicles between frames.

The combined system proved reliable in collecting common traffic parameters measured by traffic monitoring devices, such as speed and volume. An additional associated traffic parameter that can be directly determined through the use of a camera array video was the meaningful variable of vehicle density. Direct detection was achievable as a result of optimally collecting data along an entire pre-established roadway segment, rather than conventionally back-calculating based extraction of spot data at discrete locations. Traditional methods used to extract speed, density, and flow data are discussed in Zhao, et al. (2017). Automatically and manually collected data for eight uninterrupted flow roadway segments are summarized in Table 1 as extracted from a single frame. Roadway segments were determined using the same imagery from which the previously described network algorithms were trained. A constraining shortcoming stemmed from the inclusion of only 23 (0.6%) of 3,828 test segment vehicles for use in process training routines. Unfortunately, with a limited dataset and randomly selected training samples, adverse overlaps between training and testing were inevitable. Nevertheless, the presence of small overlaps should not detract from a successful demonstration of the overall reliability and functionality of combining deep learning detection with a feature-based tracking algorithm. Values allowing ground truth comparisons of 1) density were obtained by manually counting vehicles present within each frame, and 2) accompanying speeds were obtained by averaging the Euclidean distance of the movement across frames for randomly selected vehicles, and 3) corresponding volumes were determined by multiplying associated values for density and speed. For algorithm validation, manually labeled ground truth data was directly compared to automatically extracted network measurements. From these comparisons, algorithm determined density data was identified as providing the highest accuracy while algorithm determined speed data was also deemed to be accurate based on the preceding discussion, both values of which can produce accurate estimates of LOS and volume.

TABLE 1 Traffic Data Measurements

Road Segments	No. of Lanes	Length (mi)	Count (veh)		Speed (mph)		Density (v/mi/ln)		Volume (v/hr/ln)		LOS		Accuracy	
			A[1]	M[2]	A	M	A	M	A	M	A	M	Density	Volume
OH-4 (WB)	2	3.05	30	36	66.9	69.8	4.9	5.9	329	412	A	A	83.33%	79.82%
OH-4 (EB)	2	3.04	33	36	62.8	66.8	5.4	5.9	341	396	A	A	91.67%	86.19%
I75 (SB)	3	1.68	96	104	63.1	56.2	19.0	20.6	1202	1160	C	C	92.31%	96.42%
I75 (NB)	3	1.68	60	76	58.8	68.2	11.9	15.1	700	1029	B	B	78.95%	68.06%
US-35 (EB) [1]	4	0.83	23	30	57.3	50.0	6.9	9.0	397	452	A	A	76.67%	87.88%
US-35 (WB) [1]	4	0.83	11	12	56.0	65.2	3.3	3.6	186	236	A	A	91.67%	78.94%
US-35 (WB) [2]	3	0.88	54	54	61.5	65.7	20.5	20.5	1258	1345	C	C	100.00%	93.55%
US-35 (EB) [2]	3	0.88	61	67	58.8	62.9	23.1	25.4	1359	1596	C	C	91.04%	85.15%
OH-4 (WB) *	2	3.05	34.66	36.98	65.6	NA	5.7	6.1	373	NA	A	A	93.73%	NA
OH-4 (EB) *	2	3.04	32.12	32.72	62.1	NA	5.3	5.4	327	NA	A	A	98.17%	NA

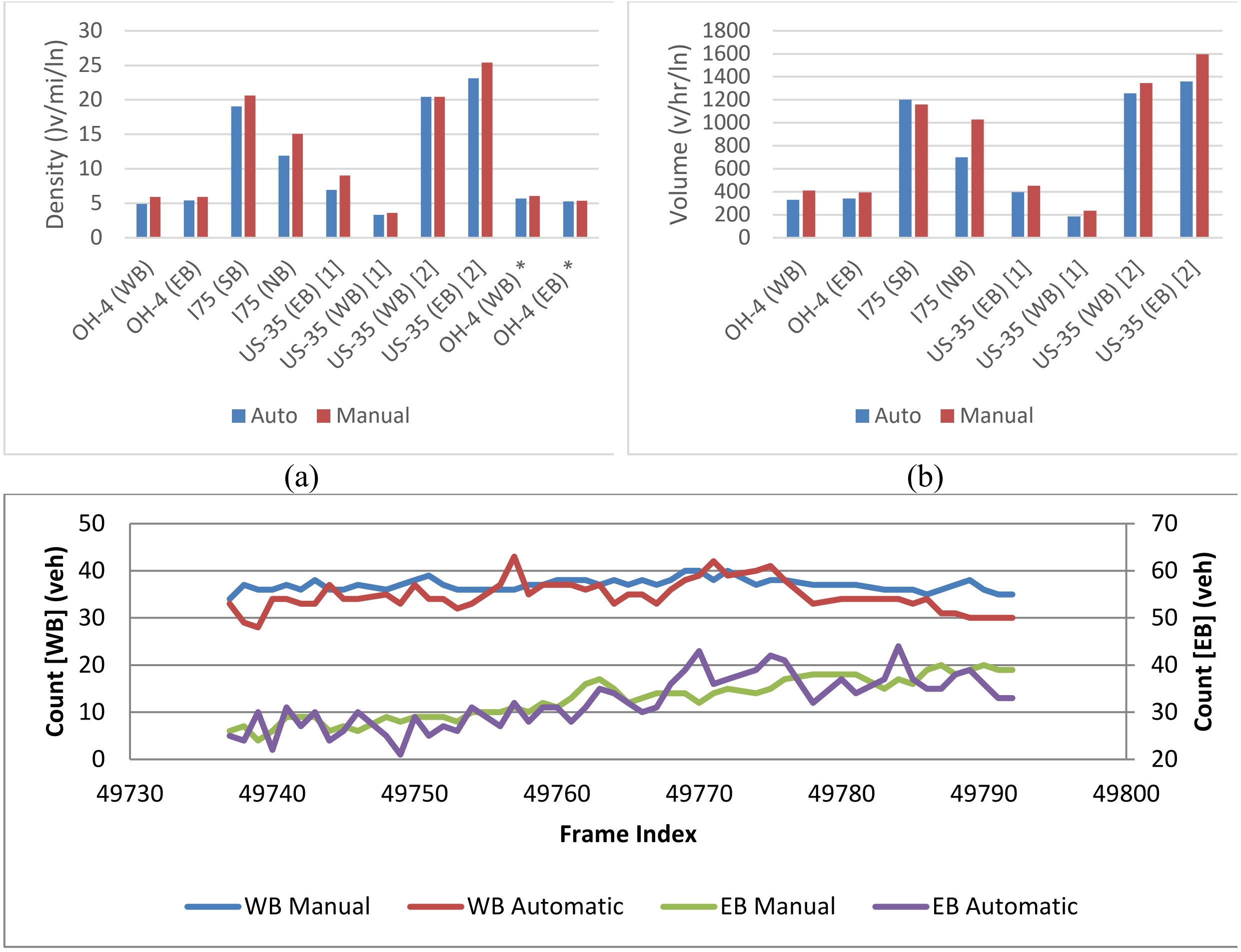
[1] Automatic measurement

[2] Manually measured ground truth

* Data based on 50 frames.

Automatically produced algorithm density data was deemed to provide precise values. Comparable ground truth densities were determined through painstakingly manual hand-counted vehicles along with segments from each frame recorded during an entire one-minute video duration and calculating an overall average. Understandably using a single frame, average errors for estimating density were determined to be 11.8%, representative of all eight segments. However, when density measurements were averaged across 50 frames, an accuracy as high as 98.2% (1.8% average error) was achieved, as summarized in the results presented in Fig. 6 (a). Results indicate that the proposed algorithm approach produced reliable network density data for roadway segments as determined from a sequence of camera array images. Reliable density values can be used to calculate LOS, which when based upon reliable density estimates, produce highly accurate and reliable performance indicator data for all of the network segments evaluated.

Algorithm produced estimates for vehicle and segment speeds were determined to be particularly sensitive to the instability of digital imagery data. Comparable values for ground truth speeds were calculated by randomly sampling a subset of vehicles from each network segment and manually tracking selected vehicles across associated images and using values as the basis for determining average speeds. Estimates using instantaneous speed determined based on the correspondence of two adjacent frames oftentimes does not produce useful and reliable results due to shifting and rotation of the second frame. However, the calculation of average vehicle speeds across multiple frames can produce accurate and useful values when the application of appropriate filter techniques is incorporated.



(a)

(b)

(c)

Fig. 6. Comparison of extracted algorithm values to manually collected ground truth values for (a) density, (b) volume, and (c) counts using 50 frames on OH-4.

Algorithm produced estimates for network volumes along pre-established roadway segments were deemed to provide accurate values. Even though instantaneous speed estimates were not reliable, volume data calculated using density were able to produce reasonable and reliable results as shown in Fig. 6 (b). The average accuracy is 84.5% (15.5% average error) across all eight segments. The accuracy of volume data can be greatly improved based on the application of appropriate filtering techniques to average speed across multiple frames.

Algorithm produced data and affiliated ground-truthing indicate that the vehicle tracking method derived from the systematic analysis of multiple frames were successful in producing reliable results. The performance of the proposed approach provided a reliable method for counting vehicles, supporting the determination of precise average density values over time as summarized in the results presented in Fig. 6 (c). Similarly, other algorithm measurements produced more precise results when averaged across multiple frames. Unfortunately, performing manual extraction as a baseline ground truthing is not practical for each vehicle within every pair of frames occurring across the entire image sequence as a means for calculating precise speed ground truth values to validate this approach.

3.3 Improvements to Tracking

The combination of a deep learning CNN approach with a feature-based tracking framework created an appropriately sophisticated algorithm that was able to account for instability, mosaic seams, and inconsistent image illumination sequences using data collected at a one-hertz frame rate with problematic mosaic seams, low image quality, and poorly preprocessed images. Beyond these plausible limitations, the performance of proposed vehicle tracking algorithms remained dependent on an additional subset of critical processing influences including occlusions, false detections, and data noise. Furthermore, the performance of the vehicle tracking approach and reliability of results was adversely impacted under the conditions of dense traffic flows.

This research developed, tested and evaluated results from a proposed novel and innovative processing algorithm approach combining multiple hypothesis tracking (MHT) with a kinematics and appearance model (KAM) to improve vehicle tracking reliability. Multiple hypothesis tracking (MHT) was intended to optimally function as a downstream component within the pipeline of vehicle detection and tracking as illustrated in the flowchart present in Fig. 7). This pipeline applied framework was configured to similarly reflect comparable procedures used in many existing radar detection systems. MHT modules use inputs from sensor or detector modules as source data and identify vehicle tracks using the highest probability or calculated scoring values. Many existing radar systems have already proven the reliability of MHT to produce data associations for multiple target tracking (MTT) tasks (Arambel et al. 2004). However, a noteworthy critical difference for most radar systems is the fact that the applications are designed for detecting and tracking fewer targeted vehicles with much wider spacing between vehicle detections than continuously processing traffic surveillance systems. Tracking vehicles under frequently occurring saturated traffic flow conditions comprises a long-standing "closely-spaced targets" processing problem. In order to improve application performance for more universal use in traffic surveillance systems, an algorithm combining MHT with a kinematics and appearance model (KAM) was determined as a promising solution to improve vehicle tracking reliability.

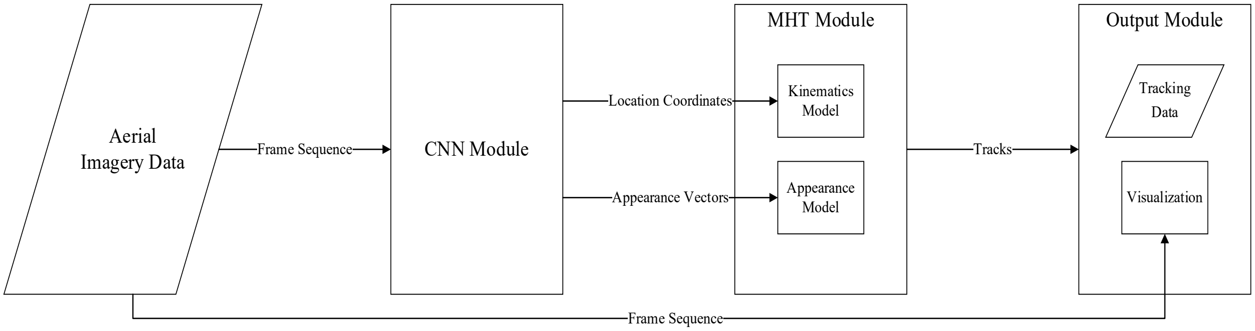


Fig. 7 The framework of vehicle detection and tracking system

Specific details describing the development and potential implementation benefits of MHT-KAM are further discussed by Zhao et al. (2019). In comparison with current MHT approaches, the proposed MHT-KAM method was developed to take full advantage of robust descriptive digital information available from aerial imagery in combination with advanced KAM routines to track vehicles building upon deep learning vehicle detection approaches previously identified by the research team. Findings indicated that an MHT-KAM approach can achieve promising performance for tracking each individual vehicle within traffic flows from the use of either a kinematic model or KAM. Previous experimental results revealed sensitivity from critical factors including frame rate, noise, traffic configurations, target density, and appearance weights on processing performance (Zhao et al 2019). Additionally, the use of the scale-agonistic property for MHT was introduced to systematically evaluate this as an optimal MHT approach for which experimental results confirmed this resolution. The findings provide workable solutions for the long-standing “closely-spaced targets” processing problem commonly resulting from saturated traffic flow conditions and offer solutions to achieve satisfactory performance using wide-area aerial imagery datasets derived from limited quality and frame rates, achieving very low detection noise, applying high appearance weights in KAM, and using large Mahalanobis distance for gating (Mahalanobis, 1936).

3.4 Requirements for Calculation of Vehicle Location Metrics

Persistent surveillance implies a rectified video covering a constant area of interest is available to readily serve as a basis for algorithm processing. Unfortunately, numerous challenges prevent widescale availability of video sequenced imaging necessary to provide an appropriate level of complexity required as a technical basis when trying to calculate coordinates for individual vehicles. The following sections provide a discussion of steps for resolving the technical aspects needed to implement an analytically rigorous vehicle mapping approach for the purposes of this research.

3.4.1 Mapping

Accurate estimation of the location (longitude and latitude) of detected vehicles identified in captured aerial images are predicated upon establishing one-to-one relationships between the geometry of the environment including the roadway and adjacent planimetric features, and image pixels viewable within the field of view for each individual camera included within the network. Projection from real-world coordinates to image coordinates requires consideration of a uniform transferable coordinate system that will allow consistent measurement of all locations of interest identifiable with the environment and adjusted to adequately reflect camera calibration. For the purpose of simplifying mapping practice, it is reasonable to approximate the target moving space (offset to the road on which vehicles move) as a planar surface. The approximation approach proves to be less feasible for large regions or hilly terrains.

3.4.2 Registration

Until improved projection rectification can be effectively and uniformly performed for proprietary video images, an undesirable and cumbersome manual registration process will be necessary to prepare video images for coordinate extraction. Visible control points are required as the basis for estimating the real-world locations of vehicles captured within the images and mosaic conglomerations. Recognizable landmarks are commonly identified as control points in practice to establish a representative coordinate grid. Alternatively, a number of randomly sampled control points theoretically produce a more reliable basis than manually selected control points for conditions when the target moving surface is not relatively “flat”. As camera array video images are actually commonly comprised of 12 videos knitted together, registration requires at least two stationary planimetric points be available for each camera included in the array. Furthermore, additional control points are desirable to better compensate for hilly terrain, lens distortion, and rectification anomalies.

3.4.3 Coordinate extraction

Once the initial registration is complete, the construction of a coordinate occupancy interpolation surface is accomplished by triangulating control points and lifting identified vertices by a magnitude into a dimension orthogonal to the image. Based on the interpolating surface, longitudes and latitudes of targeted vehicles can be estimated. Coordinates of targeted vehicles are determined by traversing the triangulated network to identify the specific triangle limits that enclose the targeted vehicle of interest, after which linear interpolation allows the determination of vehicle coordinates. A primary advantage of linear interpolation is that the interpolation process is performed independently from the triangulation process. Thus, control points can be added or removed without triggering a complete surface re-computation. Fig 8 illustrates automatically identified coordinates for all detected vehicles identified within a zoomed-in area of a camera array image.

3.5.4 Prediction and noise

Since no vehicle detection algorithm can achieve one hundred percent accuracy, false negative and false positive errors can systematically deteriorate the reliability of vehicle mapping. Thus, a Kalman filter was identified as an effective approach for establishing the movement state of a targeted vehicle and predict its movement in the following time steps for the purpose of avoiding or alleviating the detrimental influence of errors and noise interference.

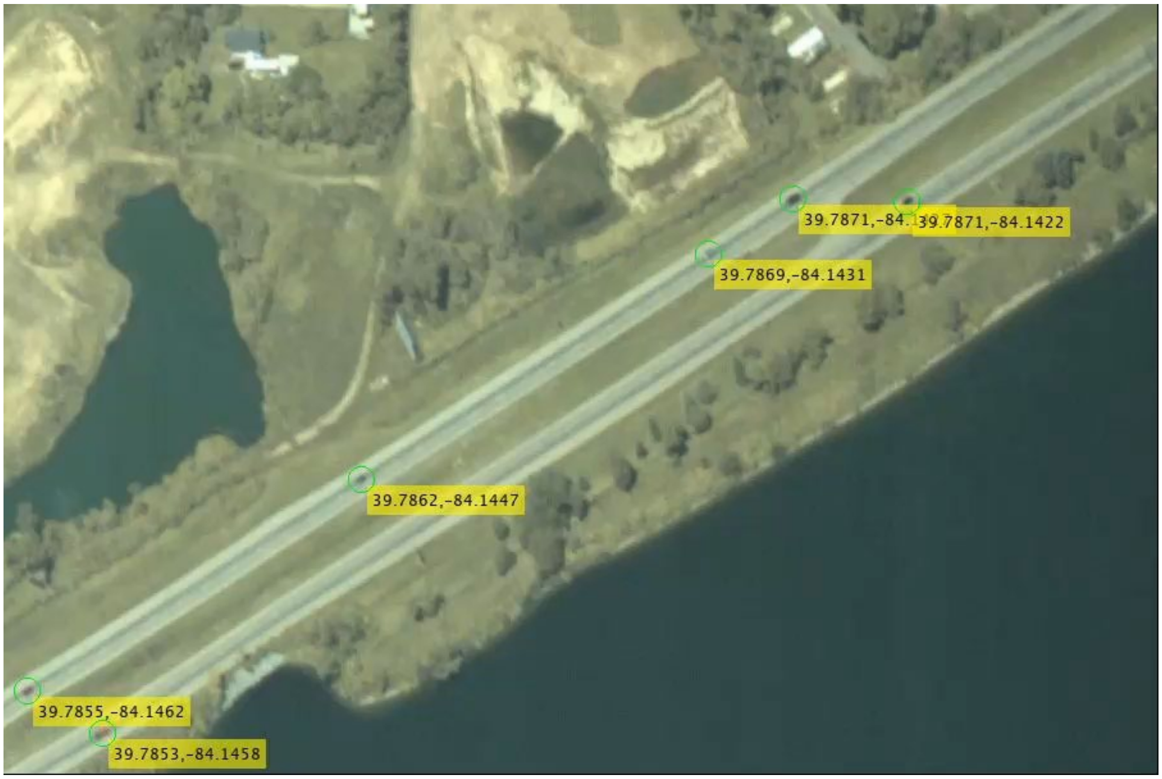


Fig. 8. Automatic identification of coordinates of detected vehicles

CHAPTER 4

Conclusions and Future Research

4.1 Conclusions

The purpose of this research initiative was to continue work on a novel and innovative automated traffic surveillance system based on algorithm processing of aerial imagery from camera arrays that build upon and advance previous tracking methods. Proposed tracking system algorithms combine vehicle feature-based tracking with vehicle detection based on a deep learning approach. This method produced promising results and supports reliable and reputable performance for automated procedures to determine speed, density, and volume along uninterrupted flow segments throughout a 25 mi^2 network coverage area. Improvements to the tracking system based on MHT-KAM developed as part of this research contributed measurably in creating a combined approach producing improved results. Research findings indicate that the proposed MHT-KAM approach can achieve promising performance for tracking individual vehicles even under conditions of dense traffic.

Testing and research evaluations validate the supposition that the proposed system is capable of determining speed, density, and volume values within an acceptable level of accuracy for many network applications. The mapping of vehicle paths across multiple image frames was also successful when vehicles were well separated. The results are promising given an adequately available number of ground-level control points. With likely forthcoming technological advances resulting in improved video preprocessing, enhanced resolution, and a higher frame rate, accuracy of tracking vehicles can be further improved which eventually will allow the envisioned processing system the enhanced capability of accurately mapping the location of all vehicle tracks occurring throughout a camera array image sequence. By increasing the accuracy of the mapped locations of vehicles in a spatio-temporal manner, additional desirable traffic parameters can be extracted microscopically including intersection turning movements, traffic signal phasing and timing, vehicle queues, identification of erratic drivers, trip origin-destination, and route decision making. Furthermore, the resulting high-functioning analytical system could facilitate traffic management through effective traffic data mining. Traffic data mining techniques offer the potential to identify reliable recurring patterns extracted from large traffic data sets with less complexity than current approaches. The ability to accurately predict traffic patterns and associated vehicle travel path parameters can facilitate and support the creation of data and information to more efficiently manage and operate traffic networks. A digital real-time “traffic map” created by the envisioned system will provide a robust data set where data mining methods could be applied to enhance traffic management and provide data for a variety of traffic studies. The data can potentially facilitate or even automate the creation and calibration of microsimulation models and activity-based travel demand forecasting models. Areawide and regional mobility models commonly require many months, or even years, to develop and calibrate, from which the consequential results are used to plan and design the transportation network of the future.

4.2 Future Research

To further improve the accuracy of the MHT-KAM model, we will consider stitching images from multiple unmanned aerial vehicles with higher resolution. We will also explore other deep learning models and compare their performance with the MHT-KAM model to further prove the efficacy of our model. In addition, a connected vehicle camera array application offers a wealth of possibilities in real-time traffic surveillance. As previously identified, one system capability could potentially benefit the traveling public as follows: through video processing, an erratic/drunk driver that is weaving through traffic can be automatically identified. This vehicle's coordinates are mapped in real-time based on our vehicle positioning algorithm. Meanwhile, instrumented vehicles could broadcast their GPS coordinates via a 5G network. Through coordinate matching, a camera array to vehicle communication would be established. Confirmation would be accomplished by comparing vehicle parameters measured through the video with those broadcasted from the matched vehicle. Once confirmation was authenticated, a verbal message for the driver to pull over would be issued. If the driver does not comply, a command message could be broadcast to automatically shut-off the erractic vehicle. This provides a realistic scenario where the active sensing capability of the proposed system could potentially prevent an incident from occurring thereby increasing safety and proactively mitigating incident induced traffic congestion.

REFERENCES

Angel, A., Hickman, M., Chandnani, D., and Mirchandani, P. (2002). "Application of aerial video for traffic flow monitoring and management." *Proc., 7th Int. Conf. Appl. Adv. Technol. Transp.,* ASCE, Reston, VA, 346-353.

Arambel, P. O., Silver, J., Krant, J., Antone, M., and Strat, T. (2004). "Multiple-hypothesis tracking of multiple ground targets from aerial video with dynamic sensor control." *Proc., SPIE Int. Soc. Opt. Eng., SPIE,* Bellingham, WA, 23-32.

Blackman, S. S. (2004). "Multiple hypothesis tracking for multiple target tracking." *IEEE Aerosp. Electron. Syst. Mag.,* 19(1), 5-18.

Blackman, S. S., Dempster, R. J., and Reed, R. W. (2001). "Demonstration of multiple-hypothesis tracking (MHT) practical real-time implementation feasibility." *Proc., SPIE Int. Soc. Opt. Eng., SPIE,* Bellingham, WA, 470-475.

Cao, X., Lan, J., Yan, P., and Li, X. (2012). "Vehicle detection and tracking in airborne videos by multi-motion layer analysis." *Mach. Vision. Appl.,* 23(5), 921-935.

Chen, B., and Medioni, G. (2015). "3-D mediated detection and tracking in wide area aerial surveillance." *Proc., 2015 IEEE Winter Conf. Appl. Comput. Vis. (WACV),* IEEE Computer Society, Los Alamitos, CA, 396-403.

Cheng, H., Weng, C., and Chen, Y. (2012). "Vehicle detection in aerial surveillance using dynamic Bayesian networks." *IEEE Trans. Image Process,* 21(4), 2152-2159.

Cox, I. J., and Hingorani, S. L. (1996). "An efficient implementation of Reid's multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking." *IEEE Trans. Pattern Anal. Mach. Intell.,* 18(2), 138-150.

Du, X., and Hickman, M. (2012). "Estimation of a road mask to improve vehicle detection and tracking in airborne imagery." *Transp. Res. Rec.,* (2291), 93-101.

Genovese, A. F. (2001). "The interacting multiple model algorithm for accurate state estimation of maneuvering targets." *Johns Hopkins APL Tech. Dig.,* 22(4), 614-623.

Hickman, M. D., and Mirchandani, P. B. (2006). "Uses of airborne imagery for microscopic traffic analysis." *Proc., 9th Int. Conf. Appl. Adv. Technol. Transp.,* ASCE, Reston, VA, 238-243.

Highway Capacity Manual. (2010). Transportation Research Board, Washington, DC.

Hinz, S. (2004). " Detection of vehicles and vehicle queues for road monitoring using high resolution aerial images." *Proc., 9th World Multi-Conf. Syst. Cybern. Informatics (WMSCI),* International Social Science Council, Paris, France, 405-410.

Jia, et al. (2014) “Caffe: Convolutional architecture for fast feature embedding.” Proceedings of the ACM International Conference on Multimedia, ACM, 675–678.

Kalal, Z., Mikolajczyk, K., and Matas, J. (2012). "Tracking-learning-detection." *IEEE Trans. Pattern Anal. Mach. Intell.,* 34(7), 1409-1422.

Kim, C., Li, F., Ciptadi, A., and Rehg, J. M. (2015). "Multiple hypothesis tracking revisited." *Proc., IEEE Int. Conf. Comput. Vis., ICCV,* IEEE Computer Society, Los Alamitos, CA, 4696-4704.

Kim, Z., and Malik, J. (2003). "Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking." *Proc., 9th IEEE Int. Conf. Computer Vision (ICCV),* IEEE Computer Society, Los Alamitos, CA, 524-531.

Kundu, A., Li, Y., Dellaert, F. Li, F., and Rehg, J. (2014). “Joint Semantic Segmentation and 3D Reconstruction from Monocular Video,” in ECCV, 1–16

Leal-Taixé, L., Milan, A., Schindler, K., Cremers, D., Reid, I., and Roth, S. (2017). "Tracking the Trackers: An Analysis of the State of the Art in Multiple Object Tracking."

Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., and Kim, T.-K. (2014). "Multiple Object Tracking: A Literature Review."

Mahalanobis, P. C. (1936). "On the generalized distance in statistics." *Proc., Natn. Inst. Sci. India,* 2, 49–55.

Maturana, Daniel & Arora, Sankalp & Scherer, Sebastian. (2017). “Looking forward: A semantic mapping system for scouting with micro-aerial vehicles.” 6691-6698. 10.1109/IROS.2017.8206585.

Moon, H., Chellappa, R., and Rosenfeld, A. (2002). "Performance analysis of a simple vehicle detection algorithm." *Image Vis. Comput.,* 20(1), 1-13.

Nguyen, T. T., Grabner, H., Bischof, H., and Gruber, B. (2007). "On-line boosting for car detection from aerial images." *Proc., 2007 IEEE Int. Conf. Res. Innov. Vis. Future (RIVF),* IEEE Computer Society, Los Alamitos, CA, 87-95.

Palaniappan, K., Bunyak, F., Kumar, P., Ersoy, I., Jaeger, S., Ganguli, K., Haridas, A., Fraser, J., Rao, R. M., and Seetharaman, G. (2010). "Efficient feature extraction and likelihood fusion for vehicle tracking in low frame rate airborne video." *Proc., 13th Int. Conf. Inf. Fusion,* IEEE Computer Society, Los Alamitos, CA, 1-8.

Palaniappan, K., Rao, R. M., and Seetharaman, G. (2011). "Wide-area persistent airborne video: architecture and challenges." *Distributed Video Sensor Networks,* Springer, 349-371.

Pelapur, R., Candemir, S., Bunyak, F., Poostchi, M., Seetharaman, G., and Palaniappan, K. (2012). "Persistent target tracking using likelihood fusion in wide-area and full motion video sequences." *Proc., 15th Int. Conf. Inf. Fusion,* IEEE Computer Society, Los Alamitos, CA, 2420-2427.

Prokaj, J., and Medioni, G. (2014). "Persistent tracking for wide area aerial surveillance." *Proc., 27th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* IEEE Computer Society, Los Alamitos, CA 1186-1193.

Reid, D. B., (1979). "An algorithm for tracking multiple targets." *IEEE Trans. Autom. Control,* 24(6), 843-854.

Reinartz, P., Lachaise, M., Schmeer, E., Krauss, T., and Runge, H. (2006). "Traffic monitoring with serial images from airborne cameras." *ISPRS J. Photogramm. Remote Sens.,* 61(3), 149-158.

Saleemi, I., and Shah, M. (2013). "Multiframe many–many point correspondence for vehicle tracking in high density wide area aerial videos." *Int. J. Comput. Vis.,* 104(2), 198-219.

Saunier, N., and Sayed, T. (2007). "Automated analysis of road safety with video data." *Transp. Res. Rec.,* (2019), 57-64.

Savinov, N., Hane, C., Ladicky, L., and Pollefeys, M. (2016). "Semantic 3d reconstruction with continuous regularization and ray potentials using a visibility consistency constraint," in CVPR, 5460–5469.

Sengupta, S., Sturgess, P., Ladicky, L., andTorr, P. (2012). "Automatic dense visual semantic mapping from street-level imagery," IROS, 857–862.

Shi, X., Li, P., Ling, H., Hu, W., and Blasch, E. (2013). "Using maximum consistency context for multiple target association in wide area traffic scenes." *Proc., 38th IEEE Int. Conf. Acoust. Speech Signal Process (ICASSP),* IEEE, Piscataway, NJ, 2188-2192.

Shitrit, H. B., Berclaz, J., Fleuret, F., and Fua, P. (2001) "Tracking multiple people under global appearance constraints." *Proc., IEEE Int. Conf. Comput. Vis., ICCV,* IEEE Computer Society, Los Alamitos, CA, 137-144.

Spraul, R., Hartung, C., and Schuchert, T. (2017). "Persistent multiple hypothesis tracking for wide area motion imagery." 2017 IEEE International Conference on Image Processing (ICIP), IEEE, 1142–1142.

Tsai, Y., Wang, C., and Wu, Y. (2011). "A vision-based approach to study driver behavior in work zone areas." *Proc., 3rd Int. Conf. Road Safety Simulation,* TRB, Washington, DC, 14-16.

Tuermer, S., Leitloff, J., Reinartz, P., and Stilla, U. (2010). "Automatic vehicle detection in aerial image sequences of urban areas using 3D HoG features." *Proc., ISPRS Technical Commission III Symposium Photogramm. Comput. Vis. Image Analysis (PCV),* International Society for Photogrammetry and Remote Sensing, Saint-Mande, France, 50-54.

Xiao, J., Cheng, H., Sawhney, H., and Han, F. (2010). "Vehicle detection and tracking in wide field-of-view aerial video." *Proc., 23rd IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),* IEEE Computer Society, Los Alamitos, CA, 679-684.

Zhang, L., Li, Y., and Nevatia, R. (2008). "Global data association for multi-object tracking using network flows." *Proc., IEEE Conf. Comput. Vis. Pattern. Recognit., CVPR,* IEEE Computer Society, Los Alamitos, CA, 1-8.

Zhao, T., and Nevatia, R. (2003). "Car detection in low resolution aerial images." *Image Vis. Comput.,* 21(8), 693-703.

Zhao, X., Dawson, D., Sarasua, W. (2019). “Multiple Hypothesis Tracking with Kinematics and Appearance Models on Traffic Flow for Wide Area Traffic Surveillance,” *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, New York, Vol. 33, Issue 3.

Zhao, X., Dawson, D., Sarasua, W., Birchfield, S. (2017). “Automated Traffic Surveillance System with Aerial Camera Arrays Imagery: Macroscopic Data Collection with Vehicle Tracking,” *Journal of Computing in Civil Engineering*, American Society of Civil Engineers, New York, Vol. 31, Issue 3, May 2017.