Modeling Driver and Pedestrian Behavior from Normal Driving and Crash Events in One Year of Virginia Traffic Camera Data

Max Bareiss

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctorate of Philosophy
In
Biomedical Engineering

Zachary Doerzaph, co-chair

Hampton C. Gabler, co-chair

Tom Dingus

Douglas Gabauer

Miguel Perez

John M. Scanlon

January 30, 2023
Blacksburg, VA

Keywords: Traffic Camera, Deep Learning, CNNs, Driver Behavior, Pedestrian Behavior, Computer Vision

Modeling Driver and Pedestrian Behavior from Normal Driving and Crash Events in One Year of Virginia Traffic Camera Data

Max Bareiss

ABSTRACT

Traffic cameras are those cameras operated with the purpose of observing traffic, often streaming video in real-time to traffic management centers. These camera video streams allow transportation authorities to respond to traffic events and maintain situational awareness. However, traffic cameras also have the potential to directly capture crashes and conflicts, providing enough information to perform reconstruction and gain insights regarding causation and remediation. Beyond crash events, traffic camera video also offers an opportunity to study normal driving. Normal driver behavior is important for traffic planners, vehicle designers, and in the form of numerical driver models is vital information for the development of automated vehicles. Traffic cameras installed by state departments of transportation have already been placed in locations relevant to their interests. A wide range of driver behavior can be studied from these locations by observing vehicles at all times and under all weather conditions.

Current systems to analyze traffic camera video focus on detecting when traffic events occur, with very little information about the specifics of those events. Prior studies into traffic event detection or reconstruction used 1-7 cameras placed by the researchers and collected dozens of hours of video. Crashes and other interesting events are rare and cannot be sufficiently characterized by camera installations of that size.

The objective of this dissertation was to explore the utility of traffic camera data for transportation research by modeling and characterizing crash and non-crash behavior in pedestrians and drivers using a captured dataset of traffic camera video from the Commonwealth

of Virginia, named the VT-CAST (Virginia Traffic Cameras for Advanced Safety Technologies) 2020 dataset.

A total of 6,779,726 hours of traffic camera video was captured from live internet streams from December 17, 2019 at 4:00PM to 11:59PM on December 31, 2020. Video was analyzed by a custom R-CNN convolutional neural network keypoint detector to identify the locations of vehicles on the ground. The OpenPifPaf model was used to identify the locations of pedestrians on the ground. The location, pan, tilt, zoom, and altitude of each traffic camera was reconstructed to develop a mapping between the locations of vehicles and pedestrians on-screen and their physical location on the surface of the Earth. These physical detections were tracked across time to determine the trajectories on the surface of the Earth for each visible vehicle and pedestrian in a random sample of the captured video.

Traffic camera video offers a unique opportunity to study crashes in-depth which are not police reported. Crashes in the traffic camera video were identified, analyzed, and compared to nationally representative datasets. Potential crashes were identified during the study interval by inspecting Virginia 511 traffic alerts for events which occurred near traffic cameras and impacted the flow of traffic. The video from these cameras was manually reviewed to determine whether a crash was visible. Pedestrian crashes, which did not significantly impact traffic, were identified from police accident reports (PARs) as a separate analysis. A total of 292 crashes were identified from traffic alerts, and six pedestrian crashes were identified from PARs. Road departure and rear-end crashes occurred in similar proportions to national databases, but intersection crashes were underrepresented and severe and rollover cases were overrepresented. Among these crashes, 32% of single-vehicle crashes and 50% of multi-vehicle crashes did not appear in the Virginia crash

database. This finding shows promise for traffic cameras as a future data source for crash reconstruction, indicating traffic cameras are a capable tool to study unreported crashes.

The safe operation of autonomous vehicles requires perception systems which make accurate short-term predictions of driver and pedestrian behavior. While road user behavior can be observed by the autonomous vehicles themselves, traffic camera video offers another potential information source for algorithm development. As a fixed roadside data source, these cameras capture a very large number of traffic interactions at a single location. This allows for detailed analyses of important roadway configurations across a wide range of drivers. To evaluate the efficacy of this approach, a total of 58 intersections in the VT-CAST 2020 dataset were sampled for driver trajectories at intersection entry, yielding 58,180 intersection entry trajectories. K-means clustering was used to group these trajectories into a family of 45 trajectory clusters. Likely as a function of signal phase, distinct groups of accelerating, constant speed, and decelerating trajectories were present. Accelerating and decelerating trajectories each occurred more frequently than constant speed trajectories. The results indicate that roadside data may be useful for understanding broad trends in typical intersection approaches for application to automated vehicle systems or other investigations; however, data utility would be enhanced with detailed signal phase information.

A similar analysis was conducted of the interactions between drivers and pedestrians. A total of 35 crosswalks were identified in the VT-CAST 2020 dataset with sufficient trajectory information, yielding 1,488 trajectories of drivers interacting with pedestrians. K-means clustering was used to group these trajectories into a family of 16 trajectory clusters. Distinct groups of accelerating, constant speed, and decelerating trajectories were present, including trajectory clusters which described vehicles slowing down around pedestrians. Constant speed trajectories

occurred the most often, followed by accelerating trajectories and decelerating trajectories. As with the prior investigation, this finding suggests that roadside data may be used in the development of driver-pedestrian interaction models for automated vehicles and other use cases involving a combination of pedestrians and vehicles.

Overall, this dissertation demonstrates the utility of standard traffic camera data for use in traffic safety research. As evidence, there are already three current studies (beyond this dissertation) using the video data and trajectories from the VT-CAST 2020 dataset. Potential future studies include analyzing the mobile phone use of pedestrians, analyzing mid-block pedestrian crossings, automatically performing roadway safety assessments, considering the behavior of drivers following congested driving, evaluating the effectiveness of work zone hazard countermeasures, and understanding roadway encroachments.

Modeling Driver and Pedestrian Behavior from Normal Driving and Crash Events in One Year of Virginia Traffic Camera Data

Max Bareiss

GENERAL AUDIENCE ABSTRACT

Traffic cameras are those cameras operated with the purpose of observing traffic, often streaming video in real-time to traffic management centers. These video streams allow transportation authorities to maintain situational awareness and respond to traffic events. However, traffic cameras also have the potential to directly capture crashes, providing enough information to perform reconstruction and gain insights regarding causation and remediation. Beyond crash events, traffic camera video also offers an opportunity to study normal driving, which is vital information for the operation of automated vehicles. Traffic cameras installed by state departments of transportation have already been placed in thousands of locations around the country capturing traffic scenes relevant to their interests. A wide range of driver and pedestrian behavior can be studied from these locations by observing vehicles at all times and under all weather conditions.

Current systems to analyze traffic camera video focus on detecting when traffic events occur, with very little information about the specifics of those events. Previous studies into traffic event detection or reconstruction used 1-7 cameras placed by the researchers and collected dozens of hours of video. Crashes and other interesting events are rare and cannot be sufficiently characterized by camera installations of that size.

The objective of this dissertation was to explore the utility of traffic camera data for transportation research by modeling and characterizing crash and non-crash behavior in pedestrians and drivers using a dataset of statewide traffic camera video captured from the Commonwealth of Virginia.

A total of 6,779,726 hours of traffic camera video from live internet streams was captured from December 17, 2019 at 4:00PM to 11:59PM on December 31, 2020. This captured video was processed by a trajectory analysis system which determined the path on the ground for each visible vehicle and pedestrian in a random sample of the captured video. Additionally, 298 crashes visible in the traffic camera video were analyzed, comparing them to nationally representative crash datasets. With anticipated uses in traffic modeling and automated vehicle development, two additional potential use cases of the dataset were explored: cases where a driver enters an intersection, and cases where a driver interacts with a pedestrian.

ACKNOWLEDGEMENTS

First, I would like to acknowledge the influence of my co-chair Dr. Hampton C. Gabler (deceased: January 11, 2021). His impact on this work, my personal character, and career trajectory, years after his passing, remains strong.

My time as a PhD student was in part funded by the Toyota Collaborative Safety Research Center and Toyota Motor Corporation. I give special thanks to Katsuhiko Iwazaki, Takashi Hasegawa, and Rini Sherony of Toyota for sharing their technical insights and expertise throughout the project. I acknowledge the Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within this dissertation. This research was done using services provided by the OSG Consortium (*1*, *2*), which is supported by the National Science Foundation awards #2030508 and #1836650.

I would like to thank Calvin Winkowski, Ken Strickler, Eric Wonderley, and Neal Feierabend at Virginia Tech Transportation Institute for both providing computational resources and support and dealing with my strange computing problems. I would also like to thank Brady Siegel, Edward Shangin, Kyle Brown, Nikhil Pradeep, Stephanie Bandy, Teddy Bird, and others I cannot remember for their help as undergraduate interns over the years.

A detailed treatise on traffic video analysis is not the natural result of birth; there are many people who guided me to this outcome. Special thanks to my parents; Stephen Pendergrast and Mike Bloodworth at Pope John Robotics; Chris Land at Sussex County Technical School; Dr. Eric Constans and Dr. Krishan Bhatia at Rowan University; and my grad school friends and labmates including Luke Riexinger, Sammy Haus, Whitney Tatem, Grace Wusk, Keegan Yates, Lexi Basantis, Morgan Dean, Shannon King, Nick Johnson, and many, many others.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CDS	Crashworthiness Data System
CISS	Crash Investigation Sampling System
CRSS	Crash Report Sampling System
FARS	Fatality Analysis Reporting System
NASS	National Automotive Safety System
NHTSA	National Highway Traffic Safety Administration
NMVCCS	National Motor Vehicle Crash Causation Survey
PAR	Police Accident Report
VDOT	Virginia Department of Transportation
VTIS	Virginia Traffic Information System
VT-CAST	Virginia Traffic Cameras for Advanced Safety Technologies

1. INTRODUCTION

There are several sources of crash data compiled in the United States at the state and national level. For example, national crash databases, collected by the National Highway Traffic Safety Administration (NHTSA) include the Crash Investigation Sampling System (CISS) (*3*), Crash Report Sampling System (CRSS) (*4*), Fatality Analysis Reporting System (FARS) (*5*), and older studies such as the National Motor Vehicle Crash Causation Survey (NMVCCS) (*6*). Nationally representative conclusions can be drawn from these databases, due to the sampling schemes of CISS, CRSS, and NMVCCS and the census nature of FARS. However, these databases each have their drawbacks when considering certain crash analyses. While CISS includes detailed injury information, scene diagrams, and event data recorder information for most cases, the limited annual sample size and short duration of the study limit the statistical analyses which can be performed. CRSS samples a larger number of cases, but relies heavily on police reports, which are not filled out by trained crash investigators and do not contain reliable injury information. Because FARS is a census with no sampling scheme, it is not subject to the same statistical limitations as CRSS and CISS. However, because it is derived from police reports the information available is not as detailed as CISS. The data in the NMVCCS study was in many cases collected directly from the crash scene while the vehicles were located in their final rest positions, allowing for full crash reconstructions. However, NMVCCS does not contain detailed injury information and was conducted from 2005 to 2007, before recent changes such as the broad adoption of smart phones and the introduction of modern active safety systems; limiting the generalization of recent and future analyses.

State databases of police reports, in contrast to federal datasets, cannot be used directly to make traffic safety claims applicable to the entire United States. However, state datasets often include all police reported crashes, allowing for higher statistical power in analyses. These databases, by only including police reports, are limited in the number of data elements and crashes cannot be reconstructed. Additionally, CISS, CRSS, and FARS require a significant amount of processing at NHTSA and there is a considerable delay from the time of the crash to its presence in the public dataset; at the time of writing this dissertation in early 2023, the most recent crash data available was from December 2021. In contrast, police report data from the Commonwealth of Virginia was available from October 2022. A tabular comparison of these databases is given in Table 1. To proactively study traffic safety, there exists an opportunity for a recently collected, if not continuously ongoing, large database with detailed information for crash reconstruction.

Table 1. Tabular comparison of modern crash datasets. † High-level FARS summary statistics are released quarterly.

	Cases Per Year (within study population)	Study Population	Crash Reconstruction Potential	Data Availability
CISS	~5,000	Police Reported Towaway	**Possible**	12-14 month reporting lag
CRSS	**~50,000**	Police Reported	Not Possible	12-14 month reporting lag
FARS	**All cases included**	Fatality	Not Possible	12-14 month reporting lag†
NMVCCS	~5,000	EMS Activation	**Possible**	Stopped collection in 2007
VA State Police Reports	**All cases included**	Police Reported	Not Possible	**3 months reporting lag**
This Dissertation	**59% of cases**	Crashes visible near traffic cameras	**Possible**	**1 hour reporting lag**

Traffic cameras are those cameras operated with the purpose of observing traffic, which allow transportation authorities to respond to traffic events and maintain situational awareness. Traffic cameras have the potential to directly observe crashes, providing enough information to perform detailed reconstruction. Correlating these crashes with police reports would allow for the integration of occupant, vehicle, and environment information to analyze cases for injury risk modeling, active safety benefits estimation, developing driver models for applications such as autonomous vehicle development, and other study designs.

Beyond crash events, traffic camera video also offers an opportunity to study normal driving. Traffic cameras installed by state departments of transportation have already been placed

in locations relevant to their interests. Traffic cameras in most states are located at regular intervals on interstate highways, and are often present at large intersections on non-access-controlled roads. A wide range of driver behavior can be studied from these locations by observing the large number of vehicles which pass through the observed roadway at all times and under all weather conditions.

The objective of this dissertation was to capture streaming traffic camera video from the Commonwealth of Virginia and to evaluate efficacy for computing the trajectories of vehicles and pedestrians. Furthermore, this dissertation leveraged these trajectories to examine crash and non-crash behavior in pedestrians and drivers to evaluate the ability of these data for safety and driver modeling analyses. The name of the dataset of traffic camera video is the Virginia Traffic Cameras for Advanced Safety Technologies 2020 dataset, or VT-CAST 2020.

Prior Work

Historical approaches to use camera data in traffic observation and automated driving focused on lane detection and vehicle detection and have been detailed by Kastrinaki, Zervakis, and Kalaitzakis (*7*). Lane detection focused on automated methods of determining the locations of lanes within the camera view for both infrastructure-mounted cameras and vehicle-mounted mobile cameras. Methods for lane detection considered three primary approaches: activity detection, feature detection, and model fitting. Activity detection relies on segmenting areas of the image with frequent and common movement, as these areas will likely be lanes from a fixed-view camera. Feature detection methods use contrast and brightness information to identify lane lines. Model fitting methods deform a template to minimize error between the template and the camera image. These techniques have the advantage that they operate in camera space without

transforming the image to a top-down view and are suitable for lane line detection in an unlimited number of locations. However, these methods cannot determine the true identity of a lane; i.e. they can identify a lane's location and shape, but cannot connect this information to pre-existing highway maps. Vehicle detection as described by Kastrinaki, Zervakis, and Kalaitzakis has been completely supplanted by deep learning methods which were developed in the following years and used in this work.

Other research groups have performed studies using stationary cameras to observe road vehicles. In Wu et. al. (*8*), the authors describe a computer vision based system for tracking vehicles on a circular course 260m in diameter for the purpose of studying driver behavior in traffic. The standard error of position measurement was 11cm and the standard error of velocity measurement was 9cm/s. Because of the controlled track, fixed camera, and fixed fleet of vehicles, a clustering algorithm using manually generated vehicle templates was effective at detecting and tracking vehicles over time. While demonstrating the potential of camera-based vehicle detection methods, the simplicity of their experimental design shows the limits of template matching approaches and the need for advanced methods, such as the deep learning methods in this work.

In Liu, Weinert, and Amin (*9*), the authors collected traffic camera video from seven cameras over 53 days between November 6, 2017 and January 31, 2018. Cameras recorded frames every three minutes, and a total of 259,830 frames were collected. The Google Cloud Vision service was used to annotate these images with keywords, and trends in these keywords over time were used to analyze traffic and weather events. This study demonstrates how informative analysis of video streaming from pre-existing cameras can be; however the analysis method is simple and

opaque, relying entirely on a proprietary service for analysis. This motivates the transparent and advanced vehicle detection method in this work.

In Kanhere et al. (*10*), the authors describe a system for detecting motorcycles from stationary cameras for the purpose of traffic volume estimation. A total of 1.5 hours of video was collected from two cameras. Traffic count accuracy was at most 6% for motorcycles and 3.6% for all vehicle types, demonstrating a clear need for additional improvement.

In Tiezzi et al. (*11*), the authors describe a system for detecting stationary vehicles located along a roadway, stationary vehicles at the moment they leave, vehicles driving the wrong way, and crashes from traffic cameras located along highways in Italy. A total of 10.3 hours of video was collected from 130 cameras. The F1 score of detecting an event in a one-minute video clip ranged from 0.82 to 0.98, suggesting that event detection directly from traffic camera video can be effective. However, this method simply detects the events without any additional information, like waiting time for stationary vehicles, speed and behavior of wrong-way vehicles, or crash severity information for crash vehicles. While promising for detecting real traffic events, it only considers a limited data collection interval.

In Wu et al. (*12*), the authors describe a system for identifying vehicle-pedestrian near-crashes from a roadside LIDAR sensor. Two LIDAR sites collected two hours of data. Trajectories were extracted for each vehicle and pedestrian. A total of nine events were identified as either crash-relevant conflicts or near-crashes (using the same definitions as the SHRP 2 NDS (*13*)) or events where the driver broke Nevada law. The study shows a promising application of vehicle-pedestrian event detection in trajectory data which could be performed using the VT-CAST 2020 dataset.

In the CICAS-V project (*14*, *15*), the authors instrumented six stop sign controlled intersections with radar and camera units. A total of 7,318 hours of data were collected, and models were developed to predict the likelihood of drivers failing to stop at an intersection for use in Intersection Collision Avoidance Systems (ICAS). While effective, the study of pre-existing traffic cameras has the potential to collect similar information with a greater diversity of collection sites.

In Xing et al. (*16*), the authors observed a single toll plaza from an overhead UAV. A total of 1,103 vehicles were tracked by in the study, and statistical modeling was used to analyze pairs of vehicles using each vehicle's speed, initial and target lanes, and other factors to predict TTC as a proxy for crash risk. Vehicles were manually labeled and tracked by a template matching tracker.

In Ismail et al. (*17*), the authors describe a system for automatically capturing pedestrian-vehicle conflicts from video data. Twenty hours of video were captured from one location. Approximately 7,000 vehicles and 2,100 pedestrians were observed at a potential conflict point. Behavioral metrics such as Time to Collision (TTC) and Post-Encroachment Time (PET) were computed for each potential collision pair. A projection was developed between the camera view and an orthographic view of the road surface to determine pedestrian and driver position using matching points in both views. Features on pedestrians and vehicles were identified using a custom feature detector and tracked using the standard KLT tracker (*18*, *19*). The system was able to predict approximately 90% of conflicts and 72% of important events as compared to a human evaluator using the FHWA observer's guide. However, almost 55% of normal non-conflict events were identified as important events, suggesting the system performance as a conflict/important event detector is poor. This study is strong but limited by the number of sites where data collection

occurred. By considering streaming video from pre-existing traffic cameras, a more complete range of infrastructure and driving situations could be considered.

Beyond stationary cameras, researchers have considered vehicle mounted cameras to study the behavior of the ego vehicle (the vehicle to which the camera is mounted) and other surrounding vehicles. The largest of these studies are naturalistic driving studies (NDS), where participants receive data collection equipment, including exterior-facing and interior-facing cameras, in their personal vehicles which records their normal driving for the duration of the study. The largest naturalistic driving study conducted at the time of this dissertation was the Second Strategic Highway Research Program (SHRP 2) NDS (*13*, *20*). The SHRP 2 NDS collected approximately 1.4 million hours of data from 3,362 participant vehicles, containing considerably more data than the previously mentioned static camera studies. SHRP 2 data collection concluded in 2013, and therefore did not capture any significant effects of automated vehicle technologies entering the US market. Additionally, telematics providers for large trucks offer datasets of crash events, such as those from SmartDrive and Lytx. Teoh (*21*) considered a dataset of 38,463,692 hours of driving which contained 2,058 crashes in class 8 trucks equipped with SmartDrive telematics and concluded that Automatic Emergency Braking and Forward Collision Warning reduce the occurrence of front-to-rear crashes by 41% and 44%, respectively. Soccolich and Hickman (*22*) found 201,810 safety-critical events from approximately 10,000 drivers using the Lytx driver monitoring system over a 90-day period. These datasets unfortunately are not publicly available.

A summary of these prior studies in comparison with this dissertation is given in Table Table 2. VT-CAST considers a large selection of cameras while allowing similar analyses to

previous static camera studies and SHRP 2. No study to date has been identified which uses pre-existing traffic camera installations to generate traffic safety outcomes.

Table 2. Tabular comparison of camera-based vehicle studies.

	Study Year	Study Cameras / Location	Pre-existing Camera	Study Hours	Vehicle Information	Pedestrian Information
Wu et al.	2018	1 Fixed	No	Unknown	**Trajectories**	None
Lui, Weinert, and Amin	2019	7 Fixed	No	4.8	None	None
Kanhere et al.	2010	2 Fixed	No	1.5	Counts	None
Tiezzi et al.	2018	**130** Fixed	**Yes**	10.3	Events	None
Wu et al.	2020	2 LIDAR Fixed	No	2	**Trajectories**	**Trajectories**
Doerzaph	2007	6 Camera + Radar Fixed	No	7,318	**Trajectories**	None
Xing et al.	2020	1 Fixed	No	Unknown	**Trajectories**	None
Ismail et al.	2009	1 Fixed	No	20	**Trajectories**	**Trajectories**
SHRP 2 NDS	2013	**3,362** Mobile	No	**~1,370,000**	**Trajectories**	**Trajectories**
Teoh	2021	Unknown Mobile	No	**38,463,692**	**Trajectories**	None
This Dissertation	2020	**1,263** Fixed	**Yes**	**6,779,726**	**Trajectories**	**Trajectories**

Traffic Camera Installations in the United States

There are no federal standards describing the required capabilities of traffic camera systems, and not all states provide the same level of service. To characterize variations in traffic camera installations across the United States, in 2021 the motorist information portal for each state was identified and the available traffic camera images and video were evaluated, when present. A summary of state traffic camera capabilities is given in Figure 1. Fifteen states offered traffic camera video for public consumption, and a further 27 states and territories offered still images of traffic. Nine states offered no traffic camera video for public consumption. These data demonstrate the growing opportunity for recording traffic cameras, as many states already have pre-existing cameras, and future states may expand their systems with the proliferation of intelligent transportation systems.

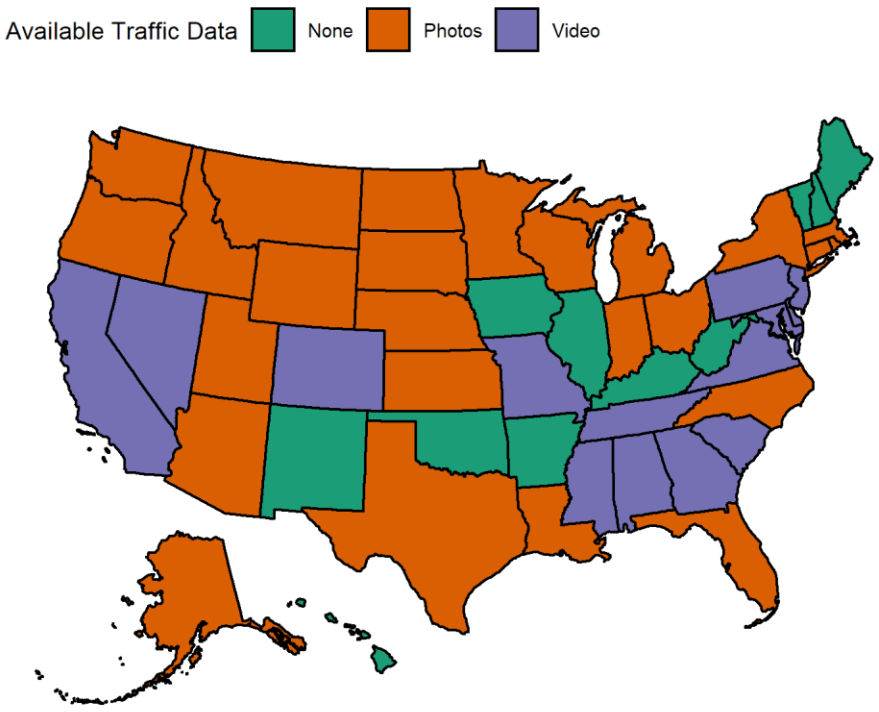

Figure 1. State traffic camera capabilities. Oklahoma offered very slow video (two frames per second).

2. METHOD TO CAPTURE TRAFFIC CAMERA VIDEO IN VIRGINIA

Methods

The source of video in this work was the Virginia Traffic Information System (VTIS) (*23*). A total of 1,263 individual cameras streaming data into the internet were available at the time of data recording. Approval from the Virginia Department of Transportation (VDOT) was obtained for capturing this database. All cameras in VTIS were included in the study. The underlying codec for all video streams was H.264. Streams were delivered using the RTMP protocol. Streams could also be delivered using RTSP or HLS, but RTMP was found empirically to be the most stable delivery format and required the least computational resources. Not all video streams had the same frame rate, resolution, and bit rate, but most were approximately 15 frames per second, 320x240

Figure 2. An image from the camera located at the intersection of Kings Highway and Richmond Highway in Alexandria, Virginia. This image was captured at approximately 2:44:34PM on March 22, 2020. A pedestrian is visible crossing the center concrete divider on Richmond Highway, highlighted in the orange circle.

(QVGA) resolution, and 200 kb/s. An example frame is shown in Figure 2. Video from all Virginia cameras was recorded from 4:00PM on December 17, 2019 until 11:59PM on December 31, 2020. One year of data collection was chosen to capture any seasonal variation in driver behavior which may occur.

The initial motivation for this endeavor was basic curiosity of the naisent research method, so the first attempt at the recording system was a low-cost off-the-shelf video recording solution. The FFmpeg (*24*) command line tool was used to record each stream of video individually, but this process was overly resource intensive for the available hardware, as the overhead for each process was too high. To rectify this issue, a custom tool was written using Python to record video from multiple streams per process. This was more performant than using discrete FFmpeg processes, but the number of concurrent streams which could be recorded was limited by the Python interpreter. After taking these steps, the Python application was re-written in C++ to efficiently use the FFmpeg C API. The C++ application was able to record approximately 200 streams concurrently per recording host. An analysis of the traffic camera URLs showed that seven servers were distributing video data from VDOT. To balance the load from each recording host to each VDOT server, the traffic cameras served from each VDOT server were evenly distributed across the six recording hosts. Video stream processing was performed using the libav suite of open source libraries (*25*). Due to the network data load associated with beginning an RTMP stream, video recording was “stagger started” so the number of streams opening per recording host was limited to two per second. In addition, at startup, streams were opened from each VDOT server in round-robin order to further limit the stream opening rate seen by a single VDOT server. The

total data recording rate of the system under nominal conditions was approximately 203Mbits/s or 2.2TB per day.

Video data was stored using the MPEG-4 (MP4) container. Midway through the data collection interval, video was stored using the Matroska (MKV) container, as this format was more resilient to failures in the capture process. Video recording stopped three minutes before the end of each hour to ensure each file was closed properly and all data was written to disk. After the implementation of the MKV container format, 90 seconds were allotted to this closeout process to record more video, as the close-out process was less resource intensive for MKV than MP4. Then, at the beginning of the next hour, video streams were re-opened and recording would begin again. Video files were roughly one hour in length and roughly 80MB. This was sufficiently large to bypass “small file” limitations in most filesystems but small enough to facilitate manual video scrubbing and limit the amount of lost video if corruption or failure occurs. Video recording was restarted if a problem with the camera stream was identified. To avoid triggering rate limiting or cause excessive load, a cooldown period was required to limit the frequency at which a given stream would be restarted. A cap of 30 restarts per hour was applied to each stream to limit the retry behavior for unrecordable streams.

These system design choices were made for performance or resource usage reasons. System performance was important for the video capture application because it had to run in real time, directly capturing video from the VTIS. Because the system was operating on a normal Linux kernel, normal network buffer sizes were available, which would quickly fill due to the volume of data. Because of this, only the smallest lapses in realtime performance could be absorbed by network buffers, and any performance variation due to disk access, system scheduler contention,

kernel time, or other system processes would result in lost video data. Common optimization techniques such as caching were unsuitable for the streaming processing in the recording hosts, as new, unique data was constantly processed.

At the beginning of the data collection project, the video capture system was deployed to the Virginia Tech Advanced Research Computing (VT ARC) on-campus cloud. Six virtual machines (VMs) were provisioned as the recording hosts. The ARC Gluster file system was accessible to each node through an NFS-GFS bridge. Data was stored on the Gluster file system over the course of one recording day. A total of 20TB of storage space was available on the Gluster file system, allowing several days of recording to be held temporarily. After this space was exahusted, the video from the oldest stored day of recording was manually packaged into an uncompressed tar (*26*) archive. GZIP compression was attempted, but compression could not be completed in real time, and no storage space gains were achieved because of the highly compressed nature of H.264 video. Each tar archive was moved onto the VT ARC tape backup system for long-term storage.

The data collection scheme had to be modified in April 2020 because of resource limitations at VT ARC. In the updated data collection scheme, tar archives of each day of video were split into 100GB video chunks and copied to Google Drive. At the time of this project Google offered unlimited storage to Google for Education customers, such as Virginia Tech. At the time of this project Google Drive maintained a 750GB per account upload limit. Because roughly 2.2TB of camera data was generated each day, four Google accounts were used, and the video chunks were uploaded to each account in round-robin order. The data was uploaded to each Google account concurrently using several instances of the rclone (*27*) software application. One recording

day (~2.2TB) could be uploaded to Google Drive in roughly 4.5 hours, preserving the real time nature of the process. Upload to Google Drive occurred at the same time new video was recorded.

The virtual machines used at first for data collection were considered “ephemeral” and their contents could be re-built at any time using the Ansible system configuration management tool. Rebuilding a capture VM took approximately 30 minutes from provisioning through VT ARC to the start of live data capture. After the switch to Google Drive for data storage, a change was made to package the data capture application using the Singularity HPC application framework. Data capture was then performed on normal ARC nodes using the SLURM scheduler. This reduced the support burden on ARC personnel to manage the virtual machines and application performance increased, as the capture system was running on an unvirtualized kernel. A network proxy was used, as ARC compute nodes did not have direct access to the Internet in the cluster used.

Video could not be recorded from each camera throughout the entire study period. Cameras were not functional at all times, and some cameras were added to and removed from the VTIS during the course of the dissertation. Cameras added to the VTIS during the dissertation were enrolled in data collection once their addition was noticed. In addition, VT ARC storage was not functional at all times and recording could not take place when the storage was not available.

In addition to the video data, crash data was also captured from the Virginia Traffic Information System. Traffic data in GeoJSON format (*28*) used to generate the publicly visible crash map layer was polled every ten seconds for the duration of video recording using a program written in Python.

Results

A total of 6,779,726 hours of video were captured from 4:00PM on December 17, 2019 to 11:59PM on December 31, 2020. If all cameras were captured at all times, a total of 11,528,664 hours of video would have been captured, resulting in a capture completeness of 58.8%. Video was lost as a result of issues relating to computational limitations, camera stream availability, storage downtime, time spent switching between capture systems, and the general performance of the capture application. With refined methods based on the work described herein, greater uptime could be achieved for future data collection efforts.

A map of camera locations is given in Figure 3. The route assigned by VDOT to each camera is given in Table 3. Most cameras (62%) were located on Interstate highways. A further 19% were located on US routes, and 11% were located on Virginia routes. The remaining 8% were located on small local roads maintained by VDOT.

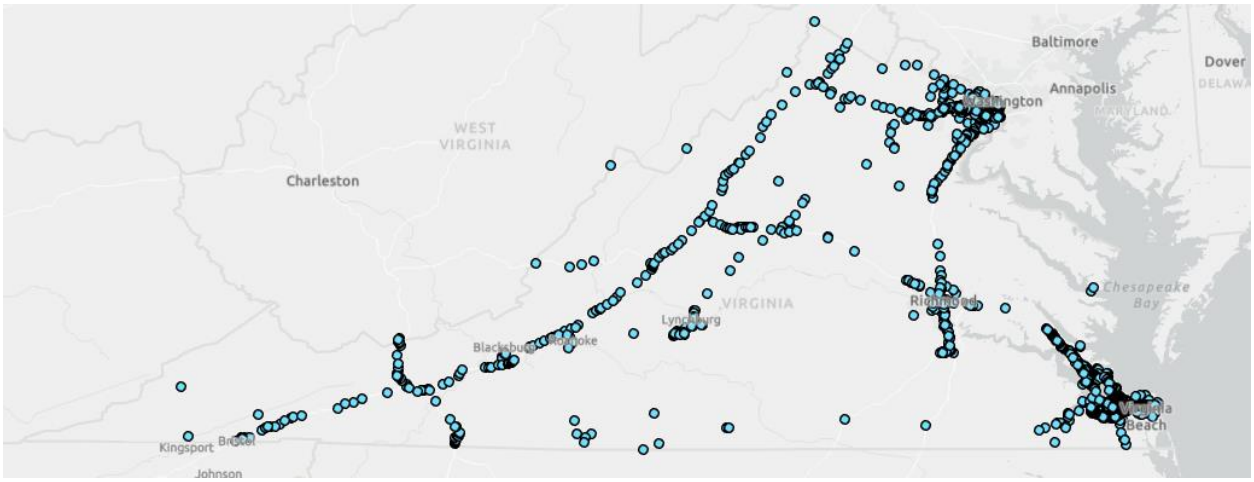

Figure 3. VT-CAST 2020 camera locations in the Commonwealth of Virginia.

Table 3. Camera Locations in the VT-CAST 2020 dataset.

	Number of Cameras	**%**
All Cameras		
All Cameras	1,263	100%
Primary Route		
Interstate	780	62%
US Route	234	19%
Virginia Route	143	11%
Local Route	106	8%

It is important to note the heterogeneity of the cameras; cameras were installed and decommissioned as needed, using the cameras available at the time of purchase; the whole system therefore represents a wide variety of camera abilities, technologies, and system parameters. The bitrate of the video streams is given in Figure 4. The bitrate was within 10% of 200 kb/s for 69.8% of cameras, and a greater proportion of cameras had lower bitrates than higher bitrates. The video resolutions and frame rates observed are detailed in Table 4. The vast majority of video resolutions (93%) were 320 pixels in width and 240 pixels in height. Most videos (77%) were within 10% of 15 frames per second (fps) nominally. A total of 18% of the videos were variable frame rates under 15 fps nominally, however, the true frame rate for many of these video clips was often 15 fps.

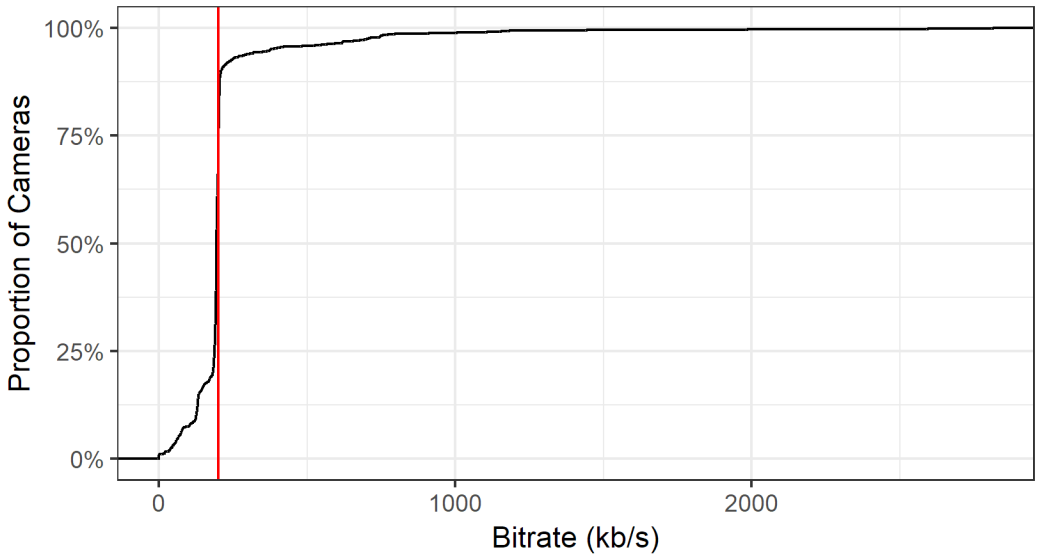

Figure 4. The bitrate of camera streams in the VT-CAST 2020 dataset.

Table 4. Camera resolutions and frame rates present in the VT-CAST 2020 dataset.

	Number of Cameras	**%**
All Cameras		
All Cameras	1,263	100%
Dimensions (WxH)		
320x240	1,169	93%
240x180	39	3%
Unknown	37	3%
352x240	7	1%
352x288	6	0%
720x480	2	0%
320x180	2	0%
704x480	1	0%
Nominal Frame Rate		
Less Than 15 fps	228	18%
15 fps	967	77%
25 fps	3	0%
30 fps	25	2%
Unknown	40	3%

The distribution of the concurrent number of cameras recorded is given in Figure 5. For more than half of the study interval, at least 800 cameras were recorded concurrently. At no time were all 1,263 of the cameras recorded simultaneously. This was because some cameras listed by VDOT as publicly available were added during the study, and thus not available at the start, while some cameras were removed during the study, so at no point were all cameras available. The presence of a time-related bias was investigated in Figure 6. No particular day of week or hour of day showed a significant deficit or surplus in hours recorded.

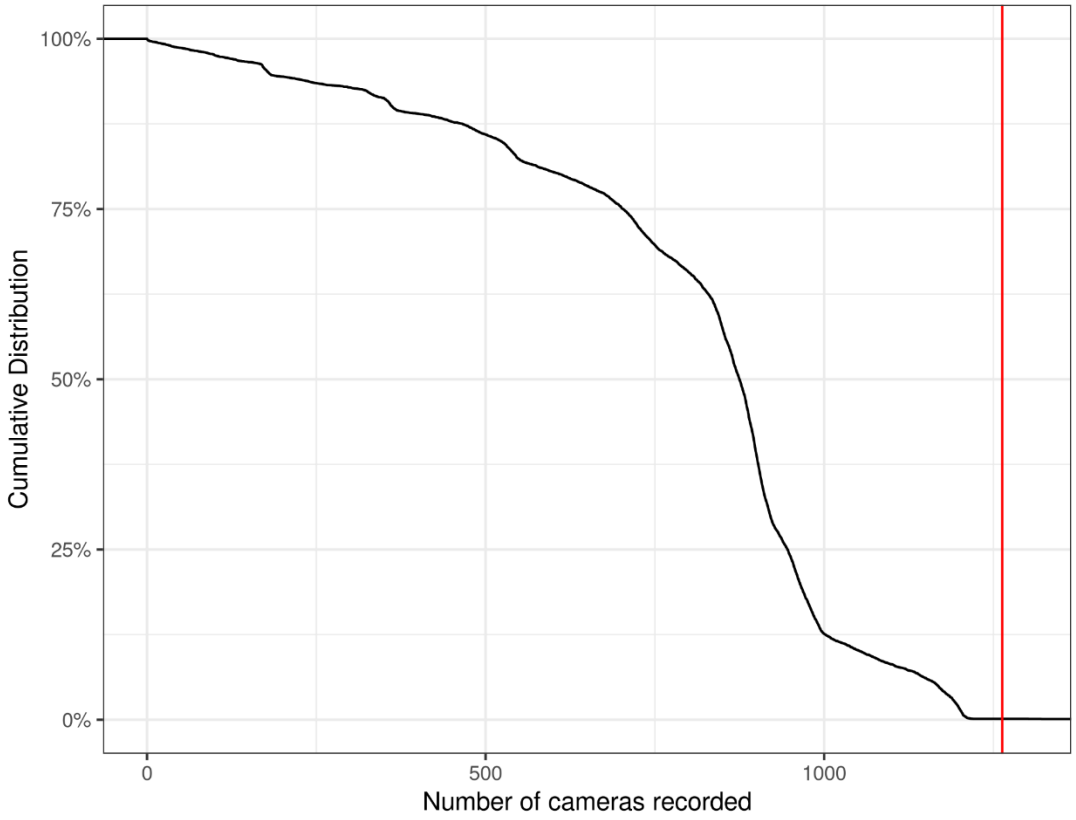

Figure 5. Cumulative distribution of the number of traffic cameras recorded concurrently. The red line indicates 1,263 cameras, the total number of cameras presented by VDOT.

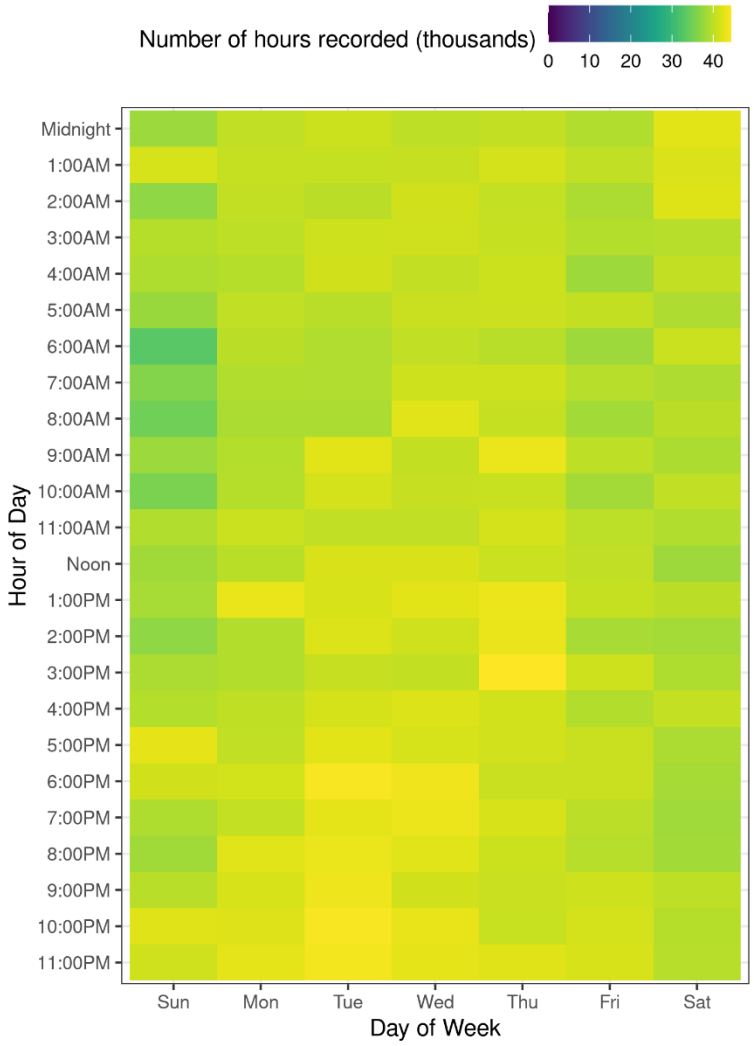

Figure 6. Hours captured by day of week and hour of day.

Discussion

The traffic cameras in this study were installed by VDOT and are only present on roads they maintain. This means that small and low-traffic roads which are maintained by counties and municipalities are not captured in camera video. A small number of county and municipal roads cross state roads or appear in the background of some camera views, but these roads are not covered deliberately or in any systematic way. However, some roads were maintained by VDOT

but were considered for census purposes to be local roads and had traffic cameras. VDOT does not provide any information or justification to describe why the camera locations were chosen. The distribution of camera locations in VT-CAST 2020 is repeated below in Table 5, along with the distribution of road miles in Virginia (*29*), and those road miles weighted by their last measured average annualized daily traffic (AADT) (*30*). Local roads represented the vast majority of roadway miles in Virginia (84%) but a small proportion of traffic camera locations (8%). Interstate cameras were also overrepresented among miles installed in Virginia, as there was 2.5 times *less* interstate mileage than US route and VA route mileage, but 2.1 times *more* interstate cameras than US route and VA route cameras. However, when weighting for AADT we can consider the miles traversed in Virginia, in contrast to the miles installed. A greater proportion of cameras were installed on interstates (62%) than the proportion of AADT-miles traveled on interstates (41.5%). US Routes were underrepresented among cameras installed (19%) as compared to AADT-miles traveled (58%), but Virginia Routes were significantly overrepresented, consisting of 11% of traffic cameras but only 0.7% of AADT-miles. AADT was not measured everywhere and information was only available for state roads, so a comparison could not be made to the AADT of county and local roads. The proportion of roads without AADT information was not known.

Table 5. Comparison of Virginia traffic camera locations to the roadway inventory in Virginia.

	Number of Cameras	%	Virginia Miles	%	Virginia AADT-Miles	%
All Cameras						
All Cameras	1,263	100%	30,427	100%	176,439,368	100%
Primary Route						
Interstate	780	62%	1,366	4.5%	73,171,334	41.5%
US Route	234	19%	3,365	11.1%	101,902,855	57.8%
Virginia Route	143	11%			1,365,179	0.7%
Local Routes	106	8%	25,696	84.4%	-	-

A significant amount of available traffic camera video was not recorded (41.2%, or 4,748,938 hours). The failures in capture were mostly random, or linked to software or hardware failure in the VDOT camera and public video distribution system. A random subset will be taken of the dataset for computational reasons in future tasks. If the recording failures are interpreted as a random sub-setting procedure, they should not significantly limit the scientific value of the dataset. A future study could reduce the amount of unrecorded video by applying the following remedies:

- Overprovision the video recording servers so that capture failures on one server do not cause total loss of video for all the cameras handled by that server;
- Use diverse storage servers or automatically upload to Google Drive to eliminate the requirement for large institutional storage systems which do not have uptime guarantees;
- Coordinate with VDOT to understand camera failures and potentially integrate directly with camera systems, bypassing the public system intended for the motoring public; and/or

- Automate the capture software provisioning so that manual restarts are unnecessary. This must be coupled with automated reporting and analytics to ensure that a bad system state is identified and rectified quickly.

Conclusion

This chapter described the recording of the VT-CAST 2020 dataset. The VT-CAST 2020 dataset contains approximately 6.22 million hours of video from 1,263 traffic cameras in the Commonwealth of Virginia, collected from December 17, 2019 to December 31, 2020. This dataset represents an incredible resource to study normal driver behavior in a variety of situations. The extent of data collection means that rare events such as near-crash and crash behavior were captured as well. Using Virginia AADT-miles on state-maintained roads as a baseline, cameras were overrepresented on interstates, underrepresented on US Routes, and overrepresented on Virginia Routes. It should be noted that in general, the location of the cameras was likely motivated by VDOT's specific needs for monitoring traffic coupled with a variety of practical factors (for example, availability of power and network communications) rather than the purpose of capturing a statewide representative video sample. Analysis and intersection of these data has been performed with this limitation in mind.

3. ANALYSIS OF TRAFFIC CAMERA VIDEO TO DETERMINE VEHICLE TRAJECTORIES

Introduction

The objective of this chapter was to process the traffic camera video captured using the method described in the previous chapter to determine the time series trajectory of each visible vehicle and pedestrian in the dataset. This process was decomposed into a series of self-contained software tasks which were independently tested. The four components of the video analysis pipeline were:

1. The vehicle and pedestrian detection component, which determined the locations of vehicle and pedestrian keypoints in each frame of video,
2. The coordinate transformation component, which determined the real-world location of vehicle keypoints, using camera tracking information,
3. The tracking component, which determined the most likely time-series trajectory of each vehicle given a series of position measurements, and
4. The camera tracking information component, which determined the most likely extrinsic and intrinsic camera parameters for each camera over time.

Methods

Pipeline

The full pipeline of the traffic camera video analysis procedure is given in Figure 7. Traffic camera video was only directly processed by two components: the vehicle and pedestrian detection component and the camera tracking component. The vehicle and pedestrian detection and camera

tracking information processed by these software components was used by the coordinate transformation component to observe the locations of vehicles on the ground. Then, the tracking component determined the mostly likely set of real vehicle trajectories explained by observations across multiple frames.

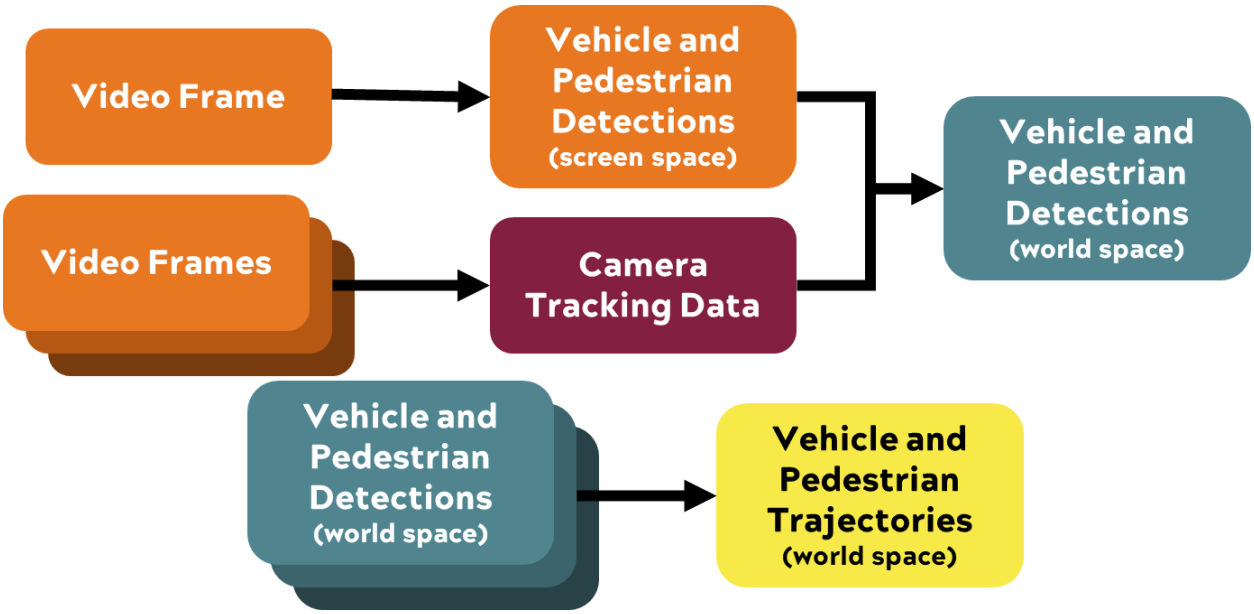

Figure 7. Pipeline for analysis of traffic camera video data to find vehicle trajectories.

Vehicle and Pedestrian Detection Component

The purpose of the vehicle and pedestrian detection component was to identify the keypoint locations of each vehicle and pedestrian visible in each video frame. The keypoints for each vehicle were defined to be the four locations along the perimeter of the vehicle where the lateral extreme of each tire (point farthest from the vehicle centerline) touched the ground. For vehicles with more than two axles, the tires on the front-most and rear-most axles were used. The entire truck-trailer unit was considered together for articulating vehicles, as the centroid of the front truck axle and rear trailer axle provide a more meaningful representation of the location of the truck than

considering the truck or trailer separately. These keypoints were chosen as they are directly on the ground and easily identifiable. Points on the ground were necessary as the keypoint locations were combined with a terrain model in a later step to determine their physical location on the roadway. Generally speaking, the 3D location of points in space cannot be directly measured from a single camera view. Prior work has considered 2D bounding boxes, which are the smallest rectangle drawn on the image that contains the vehicle. While convenient to draw and relevant for computer vision research, the sides and corners of the bounding boxes are not related to any physical location and cannot be used for accurate trajectory reconstruction of vehicles. Research systems have been developed which directly predict 3D bounding boxes from 2D views (*31*–*33*). Because it is not actually possible to determine 3D information from a single viewpoint 2D image, these methods estimate depth from the images using visual cues. These systems are highly sensitive to the intrinsic parameters of the camera(s) used in training and do not generalize well to other cameras, which would be unsuitable for use with the traffic cameras in Virginia. The Virginia traffic cameras vary considerably and can vary intrinsic parameters (such as zoom level and distortion) while operating. The keypoints for each pedestrian were defined to be the center of the left and right ankles. These points were chosen for their proximity to the ground and inclusion in previously developed human pose estimation models. While one could have considered detecting the centroid of the vehicles and pedestrians directly, the centroid does not identify a physical point on a vehicle or pedestrian and would have been impossible to verify a correctly labeled point by inspection.

Vehicle detection was performed using a custom neural network keypoint detector based on the Detectron2 (*34*) R-CNN (*35*) architecture. No suitable pre-existing vehicle detector existed at the time of model development; this represents a novel approach to vehicle detection, motivated

by the reasons stated above. The network was trained on a manually labeled dataset of 1,006 images. Images were labeled by an undergraduate researcher and manually verified. For most viewpoints, the body of the vehicle obscures one or more keypoints. The positions of these obscured keypoints were manually estimated in the manually labeled dataset, and the neural network was trained to generate similar estimates. In cases where two keypoints were occluded, this estimation served to also approximate track width and/or wheelbase. Researchers conducting the manual labeling process were instructed to label each point within three pixels of accuracy, knowing it was not possible to evaluate this concretely for obstructed labeling points.

Person detection was accomplished using OpenPifPaf (*36*) human pose estimation model. Human pose estimation is a well explored area, especially in the years since the task was included in the 2016 COCO (*37*) challenge. Runtime performance, suitability for traffic camera imagery, and detection performance were considered when selecting this model.

Sources of uncertainty in vehicle and pedestrian keypoint estimation include the wide range of scales present in traffic camera video and the detection confidence of the neural network. By estimating mean and standard deviation instead of a simple value, uncertainty information was made available to downstream processes. Uncertainty information was computed by assuming detection error was modeled by gaussian noise, where 95% of detected keypoints were within three pixels of the true location, as this was the stated error goal in the manual keypoint labeling procedure. This uncertainty, transformed into a coordinate system on the ground, was characterized using a cumulative distribution to evaluate the choice of tolerable keypoint labeling error on vehicle trajectories.

Camera Tracking Component

The traffic cameras in this study were primarily pan-tilt-zoom cameras and used by VDOT to maintain situational awareness and examine particular roadside features. The purpose of the camera tracking component was to identify the orientation (pan and tilt), focal length (zoom), physical location, and altitude of each camera. The coarse physical location and altitude for each camera was assumed to be fixed for the data collection interval (pan, tilt, and zoom were not). The VTIS contained information about the coarse latitude and longitude for each camera. Some indicated camera positions were significantly incorrect (hundreds of meters in positional error), so a manual review was conducted to correct each camera location by checking satellite imagery. A 3D model of the environment around each camera was constructed by combining orthoimagery from the Virginia Base Mapping Program (*38*, *39*) and elevation data from the USGS 3D elevation program (*40*). A region 0.002 degrees (7.2 seconds of arc, or approximately 720 feet) around each camera was considered, as this contained enough descriptive landscape features to locate the camera without processing an unreasonable amount of data.

Features in the terrain model were manually aligned with locations from each camera video in its "reference" state. Automatic alignment was tested using ORB (*41*), SIFT (*42*), and BRISK (*43*) features but was unsuccessful. An example of this alignment is given in Figure 8. Any features were used where a specific point could be clearly identified on the ground in both satellite imagery and camera video. Features used included but were not limited to transitions between pavement types, guardrail end terminals, painted road lines, the bases of signs, and manhole covers. Most VDOT traffic cameras remain in one location, and automatically return to that location after some time if the operator ceases to move the controls. A LM-BFGS minimizer (*44*) was used to

determine the most likely altitude, pan angle, tilt angle, and zoom ratio given the camera-terrain point correspondences using the cost function in Equations 1-4, where z was the altitude of the traffic camera in meters on the WGS84 ellipsoid, θ was the pan angle of the camera in radians with East as the reference angle, ϕ was the tilt angle in radians with horizontal as the reference angle, $\hat{f}$ was the zoom ratio of the camera (unitless), with 240 pixels as the reference zoom level, u_k was the easting value of point k, v_k was the northing value of point k, $Q_{u,k}$ was the easting value of the projection of point k onto the surface of the earth, $Q_{v,k}$ was the northing value of the projection of point k onto the surface of the earth, s was the dimensionless scaling factor of the projective mapping, U was the width of the camera frame in pixels, V was the height of the camera frame in pixels, R was the rotation matrix of the camera, T was the transformation matrix of the camera, and P_k was a row vector representing the location on the surface of the earth of point k. With this information, the projective mapping between each pixel in the reference camera view and the 3D terrain location could be determined. This second step mapped the features captured in the reference state to the real-world physical locations in the orthoimagery.

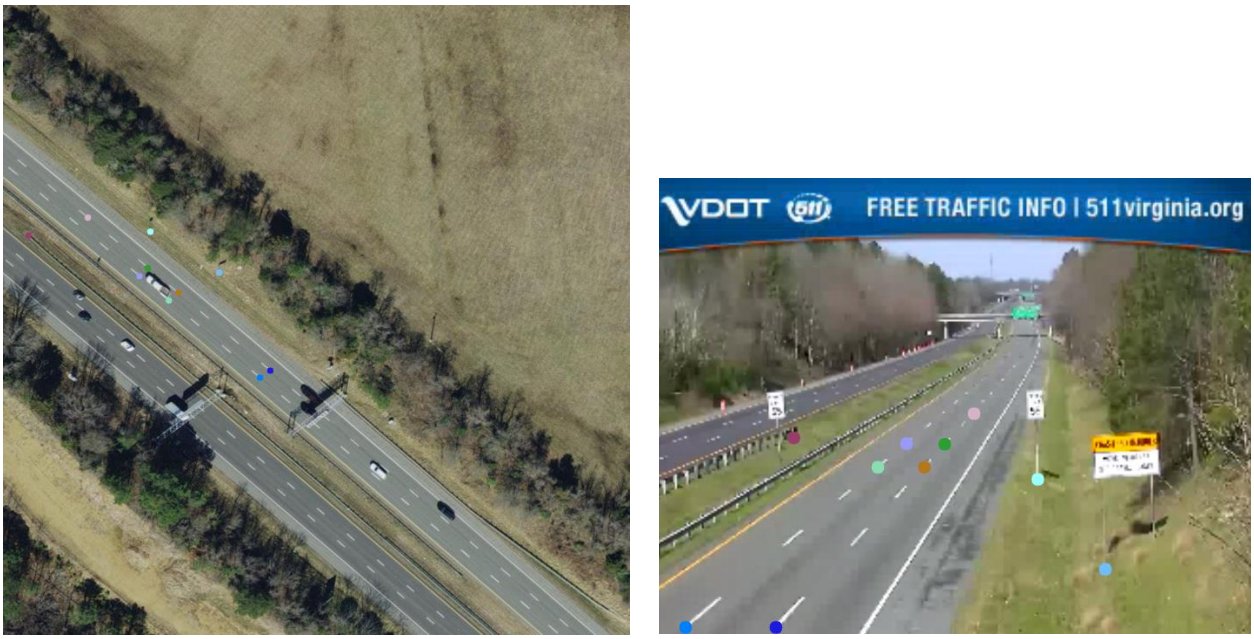

Figure 8. Satellite and camera imagery near mile marker 194.3 on I-64 Westbound in Henrico County, VA. Corresponding points are shown with matching colors.

$$z, \theta, \phi, \hat{f} = \underset{z,\theta,\phi,\hat{f}}{\operatorname{argmin}} \left\| \begin{matrix} u_0 - Q_{u,0} \\ v_0 - Q_{v,0} \\ u_1 - Q_{u,1} \\ v_1 - Q_{v,1} \\ \vdots \\ u_k - Q_{u,k} \\ v_k - Q_{v,k} \end{matrix} \right\| \tag{1}$$

$$\begin{bmatrix} sQ_{u,k} \\ sQ_{v,k} \\ s \end{bmatrix} = \begin{bmatrix} 240 * \hat{f} & 0 & \frac{U}{2} \\ 0 & 240 * \hat{f} & \frac{V}{2} \\ 0 & 0 & 1 \end{bmatrix} [R \quad T] \begin{bmatrix} P_k R \\ 1 \end{bmatrix} \tag{2}$$

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\left(-\phi + \frac{\pi}{2}\right) & -\sin\left(-\phi + \frac{\pi}{2}\right) \\ 0 & \sin\left(-\phi + \frac{\pi}{2}\right) & \cos\left(-\phi + \frac{\pi}{2}\right) \end{bmatrix} \begin{bmatrix} \cos(-\theta + \pi) & -\sin(-\theta + \pi) & 0 \\ \sin(-\theta + \pi) & \cos(-\theta + \pi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

$$T = \begin{bmatrix} 0 \\ 0 \\ -z \end{bmatrix} \quad (4)$$

Because some of the cameras can pan, tilt, and zoom, it is necessary to know how each camera moved to accurately map all of the video to real-world physical locations. This information was used to determine the location of each camera at all times through the dataset. For each frame of video in the VT-CAST 2020 dataset, ORB features were identified. These features were then matched with those found in the reference image. These features did not necessarily appear in the same location as the reference image, because a VDOT operator could have moved the camera, or the camera could have moved due to thermal expansion of an underlying bridge or wind. The pan, tilt, and zoom values which best explain the difference in position between the reference image and video frame were determined using a numerical minimizer. This measures the pan, tilt, and zoom of the camera in each frame. To determine the maximum likelihood values of these dynamic parameters, a Rauch-Tung-Striebel (RTS) smoother with a 2nd order dynamics model was used. A 2nd order model was appropriate as the pan, tilt, and zoom were all achieved using physical mechanisms with electronic velocity control that was smoothly moved by a manual operator.

Coordinate Transformation Component

The vehicle detection component detected the pixel locations of each vehicle in each video frame, and the camera tracking component determined the relationship between the camera view and the real world. The objective of the coordinate transformation component was to determine the real-world position of each vehicle in each frame. This was accomplished by casting a ray from the camera location through the location of each keypoint (Figure 9). The intersection of that ray

with the terrain model surrounding the camera represented the physical location on the surface of the Earth where that keypoint was present.

The ray casting procedure was performed using the Mitsuba (*45*, *46*) research renderer. The digital terrain model and camera were combined in a Mitsuba scene, and a view from the camera was rendered using the Mitsuba Arbitrary Output Variable integrator to determine the X, Y, and Z map position for each point in the camera view. This served as a lookup table for each camera view, where the value of the lookup table at the position of a keypoint was its position on the Earth.

After keypoints were mapped onto the terrain model, the centroids of the vehicles and pedestrians were computed. As the coordinate transformation is a nonlinear process, the centroid of the keypoints in the video frame transformed onto the terrain model was not equivalent to the centroid of the transformed keypoints, the latter more closely representing the centroid of the real pedestrian or vehicle.

Figure 9. A pixel location shown in the camera view at the left is projected through the ray shown at the right until it intersects the terrain at the location shown. The right image is a screenshot of Google Earth in the area around the traffic camera. The origin of the ray is the location of the traffic camera.

Casting a ray from the camera to each vehicle and pedestrian location represents a non-linear transformation, so the uncertainty information encoded by the vehicle and pedestrian detectors cannot be used directly. The unscented transform (*47*) was used to determine the covariance of the keypoint measurements on the surface of the Earth.

Tracking Component

The purpose of the tracking component was to combine all observations of all vehicle and pedestrian keypoints to determine the most likely set of real vehicles and pedestrians which existed at the time of video capture and determine their maximum likelihood trajectories.

First, vehicle detections were modified to remove poor quality corners. A common failure mode of the vehicle detector was to report one significantly incorrect corner prediction and three correct ones. Given that the four corners of a vehicle trace out a rectangle, it was possible to remove the incorrect corner and recompute its likely location based on the locations of the other three corners.

Subsequently, invalid vehicle and pedestrian detections were filtered out. Vehicle and pedestrian detections below a confidence threshold were not considered. Vehicle detections were not considered for trajectory generation if they were less than 2 meters in width or wheelbase, more than 4 meters in width, or more than 27 meters in wheelbase; a common failure mode of the vehicle detector was to detect vehicles of near-zero or wheelbase with a confidence above the threshold. When vehicles left the area of the lookup table generated by the coordinate transformation component, the effect was to report a very large vehicle, so these were filtered out.

Vehicles have fixed parameters, wheelbase and track width, which we assume did not change while the vehicle was visible. Vehicles also had dynamic parameters, position and heading,

which changed over time but were constrained by vehicle dynamics. For each vehicle, an RTS smoother Kalman filter was used to determine the maximum likelihood values of the fixed and dynamic parameters. A 0^{th} order model for the fixed parameters and a point-mass model for the dynamic parameters was used. The Munkres assignment ("Hungarian") algorithm (*48*) was used to associate vehicle keypoint detections across frames. Covariance information was used from the coordinate transformation component to optimally consider information from the camera view due to the properties of the Kalman filter. Because traffic cameras generally view the ground at a glancing angle, a tighter state covariance estimate was possible *across* the field of view than *into* the field of view, thus measurements were weighted more heavily *across* the field of view and the dynamics model was weighted more heavily *into* the field of view.

Pedestrians were tracked in a similar manner to vehicles. An RTS smoother Kalman filter was used to determine the maximum likelihood value of the position on the ground of the center of mass of the pedestrian. A 2^{nd} order point-mass model was used for the position of the pedestrian. While stance width could be computed analogously to vehicle track width by comparing the locations of the left and right ankle positions, this value would be poorly measured because pedestrian feet frequently lift off the ground, violating the ray-casting assumption made in the coordinate transformation component.

These vehicle and pedestrian trajectories were the final processed outcome of the analysis pipeline. They encode rich information about driver and pedestrian behavior, are time correlated with the behavior of surrounding vehicles and pedestrians, and include detailed information about measurement uncertainty.

Results

A total of 35,080 vehicle miles traveled and 139 pedestrian miles traveled were reconstructed from approximately 36,487 and 49,267 hours of video, respectively. This is a random subset of 3.5% of the video from cameras with full projective mappings between camera image locations and physical locations on the ground. Example detections and trajectories are given in Figure 10. Computation occurred using systems at the Virginia Tech Transportation Institute and across the United States using resources in OSPool (*1*, *2*). The entire traffic camera dataset was not processed due to limited available computing resources. A common failure mode of the vehicle and pedestrian detectors was to permute the keypoint locations; detecting the front-left tire in the location of the rear-right tire, for example. Trajectories were generated from the centroid of the vehicle, so this failure did not significantly impact the number of available trajectories. However, this failure served to limit the value of the heading, wheelbase, and track width information.

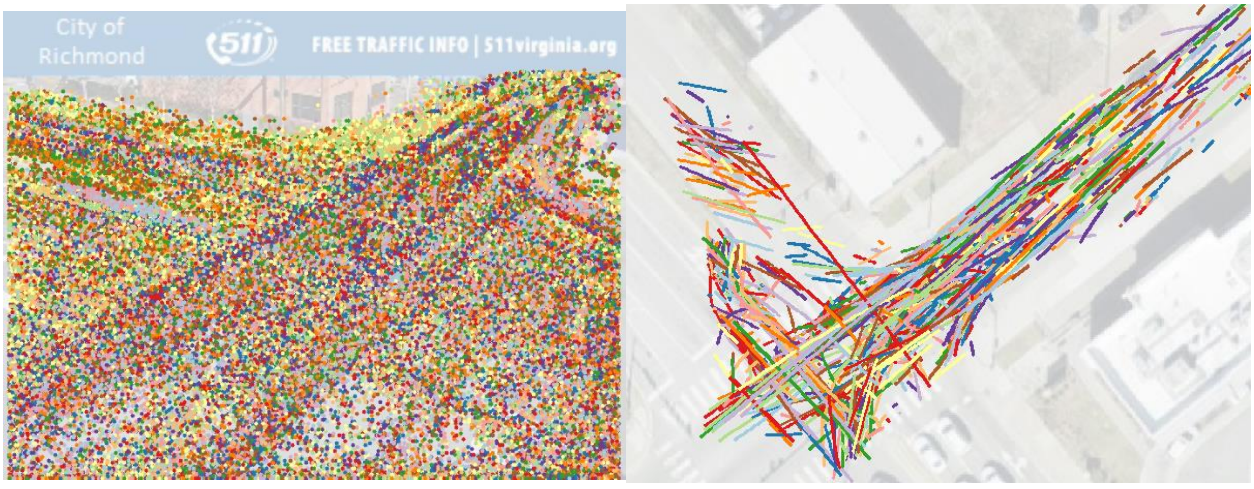

Figure 10.Vehicle keypoints (left) and trajectories (right) from 9:00AM to 10:00AM on January 1, 2020 at the intersection between Commerce Rd and Hull St (US-360) in Richmond, VA. Only trajectories longer than one second are shown.

The manual alignment process for the camera tracking component could not be completed for all camera views due to the absence of distinguishing features along some very straight or lightly curved road sections. In practice, many road sections only had lane lines as distinguishing features, which repeat regularly along the surface and cannot be used as a position reference. Additionally, alignments between the camera image and the digital terrain model could not be made in cases where road construction was occurring, as the terrain model and traffic camera would not be describing the same visual information, or when bridges were present in the video, as bridges and other manmade structures were not present in the terrain model. An example of a failure in manual alignment is given in Figure 11. A total of 154 camera views had complete manual reconstructions. In addition, the numeric minimizer which used the manual reconstruction information to determine the camera position, altitude, and intrinsic parameters could not converge for all cameras. A total of 78 camera views (6.2% of all cameras) had a full projective mapping between camera image positions and physical locations on the Earth. Trajectory information was only available from these camera views, which we will refer to as trajectory cameras. The locations of these cameras are given in Table 6, alongside the camera locations in the full dataset. Interstate highways has the smallest proportion of camera views with reconstructions, as these roads had the most failures in the manual alignment process due to the lack of identifying landmarks on the roadway and the frequent presence of bridges which could not be used for keypoints. Many US route and local route cameras were placed at intersections, so the manual reconstruction process was more successful. A greater proportion of these cameras (13.2% and 24.5%) were trajectory cameras.

Table 6. Quantity and proportion of trajectory cameras by route type.

	Number of Cameras (Total)	Number of Trajectory Cameras	Proportion of Cameras with Trajectories
All Cameras			
All Cameras	1,263	78	6.2%
Primary Route			
Interstate	780	16	2.1%
US Route	234	31	13.2%
Virginia Route	143	5	3.5%
Local Route	106	26	24.5%

Figure 11. The relatively featureless area around the camera labeled “MMBT1130” made it unsuitable for reconstruction.

The distribution of trajectory time spans is given in Figure 12. The median trajectory had a duration of eight frames, or 533 milliseconds at 15 fps. Vehicles were typically visible on screen for several seconds, so this indicates a significant amount of vehicle re-identification, a failure mode of the tracking component where the relationships between vehicle detections across frames were "forgotten" and previously labeled vehicles were given new identifiers. Some vehicles were identified for very long periods; these were typically parked vehicles.

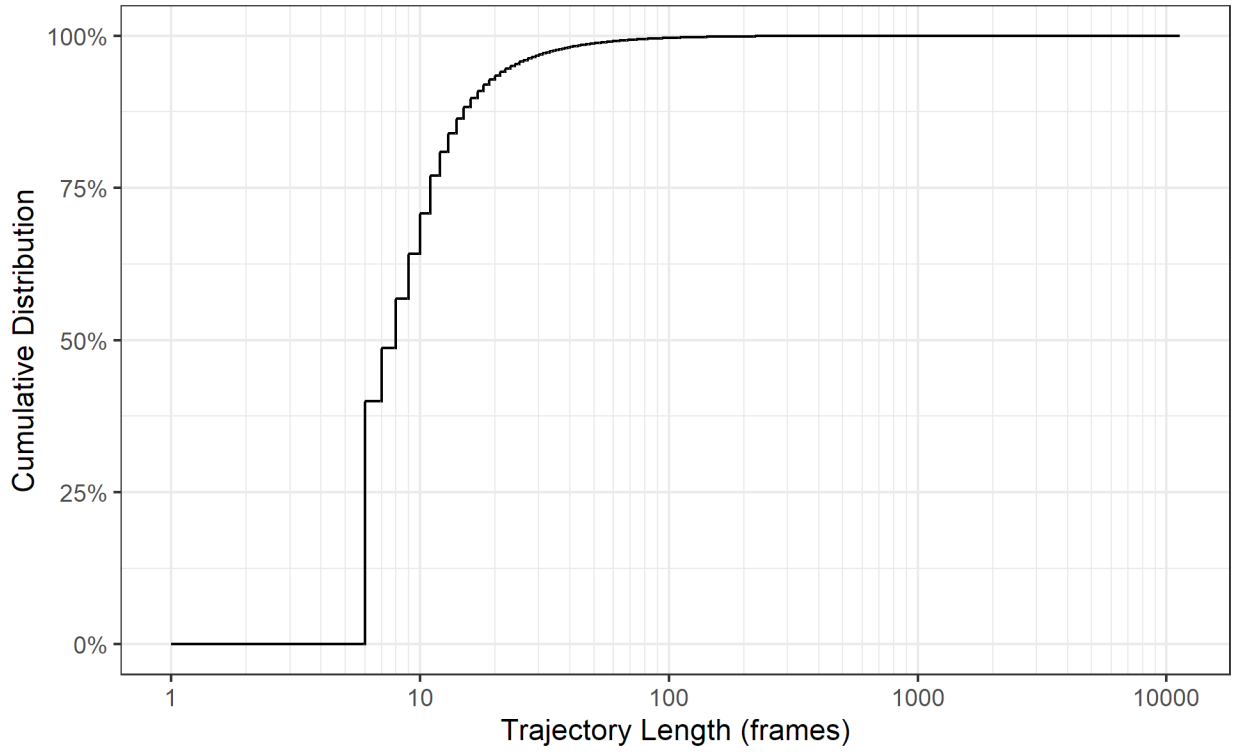

Figure 12. Distribution of trajectory lengths detected in VT-CAST 2020.

The distribution of uncertainties in vehicle position detection are given in Figure 13 and pedestrian position detection in Figure 14. The median vehicle centroid detection had a standard error of 37 cm and the median pedestrian centroid detection had a standard error of 26 cm.

Uncertainty was characterized by the square root of the larger (first) eigenvalue of the covariance matrix, which corresponds to the major radius of the confidence interval ellipse (*49*). Confidence intervals tended to be highly eccentric ellipses, as the ability to resolve a vehicle or pedestrian's position was high *across* the field of view of the camera but low *along* it. The pedestrian error was thought to be lower because pedestrians are smaller than vehicles, so the median detected pedestrian was closer to the camera than the median detected vehicle, and their position could be measured more accurately. The distribution of velocity errors is given in Figure 15. The median velocity error was 0.50 m/s, and error was roughly uniformly distributed from 0.37 m/s to 0.62 m/s.

For comparison, the cumulative distribution of Global Position System (GPS) Standard Positioning Service (SPS) error is given by a red line, computed by comparing the GPS SPS position to the surveyed position of markers at several airports within the United States (*50*). The median position standard deviation for vehicles or pedestrians was smaller than the median GPS error of 79 cm. While error and standard deviation are not directly comparable, this suggests that the position measurements made using traffic cameras are similar or better accuracy than those made by GPS.

However, when compared to GPS Wide Area Augmentation System (WAAS), shown in green, the performance of traffic camera position measurements was comparable or poorer than those made by GPS WAAS. Error was measured (*51*) in a similar manner to GPS SPS error. WAAS is a system to broadcast GPS correction data to the entire United States. This system is intended for aircraft, so the presence and availability of WAAS corrections may suffer in urban

environments at ground level. Other augmentation systems exist for GPS which offer comparable gains in accuracy.

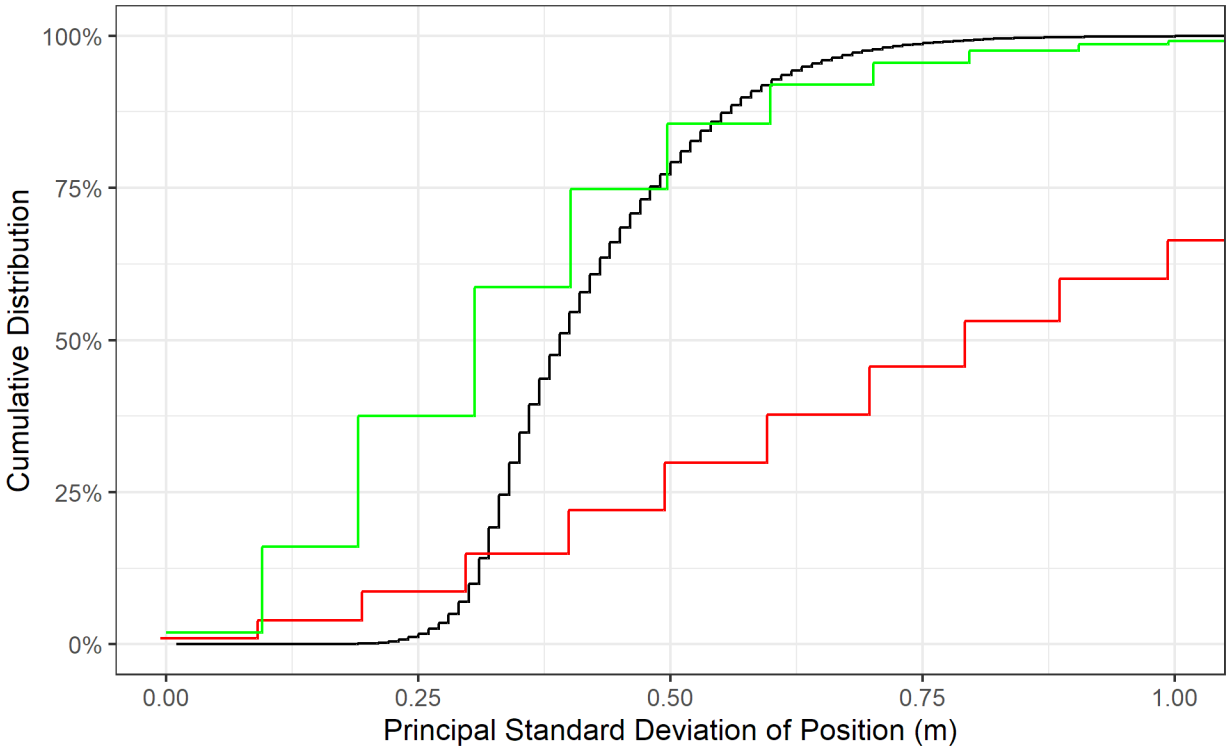

Figure 13. Cumulative distribution of the principal eigenvalue of the covariance matrix of vehicle centroid locations. The red line gives the error distribution for GPS SPS. The green line givens the error distribution for GPS WAAS.

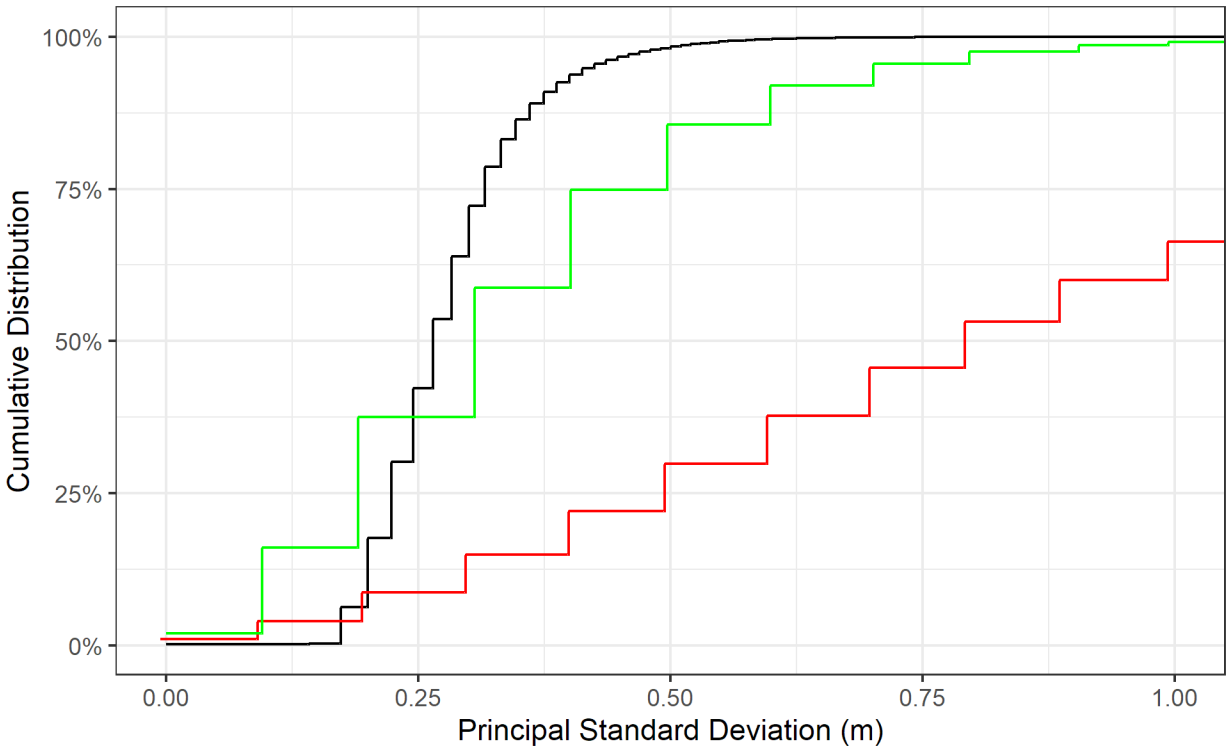

Figure 14. Cumulative distribution of the principal eigenvalue of the covariance matrix of pedestrian centroid locations. The red line gives the error distribution for GPS SPS. The green line givens the error distribution for GPS WAAS.

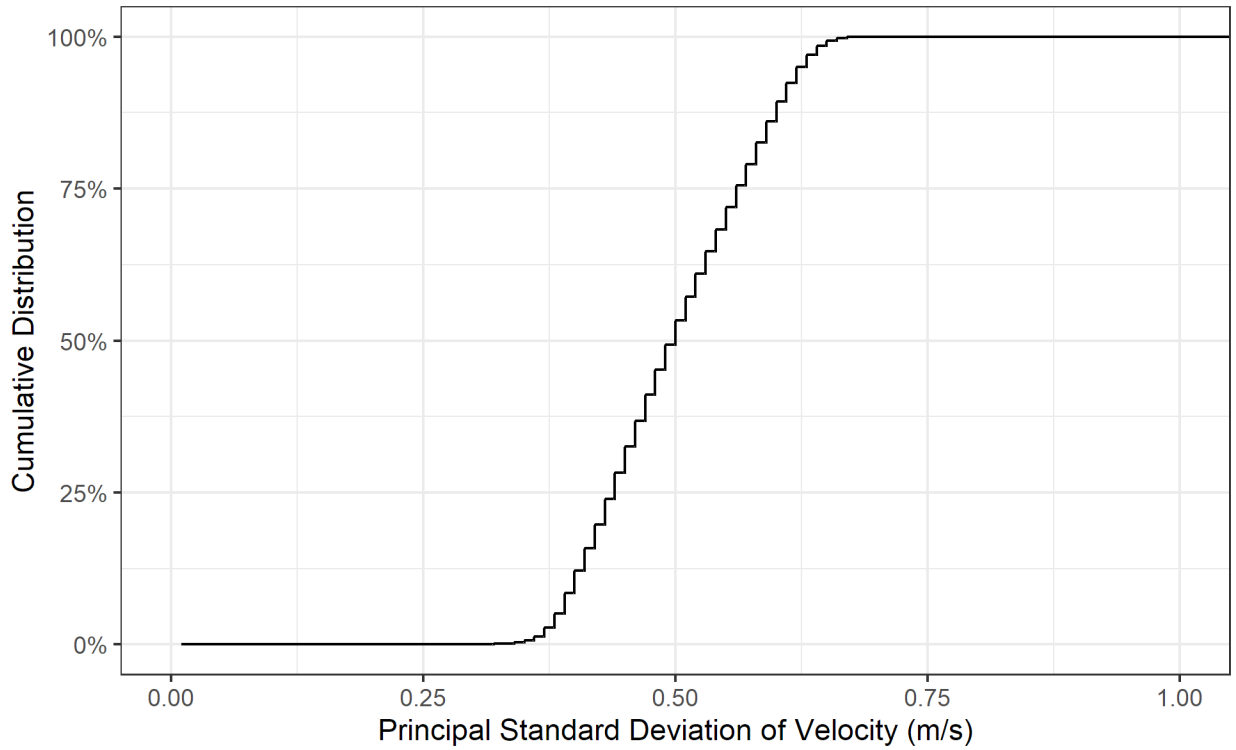

Figure 15. Cumulative distribution of the principal eigenvalue of the covariance matrix of vehicle centroid velocities.

Future Work

A future study could improve upon the VT-CAST 2020 analysis pipeline in the following areas:

- Aggressively optimize software performance of the video processing components to make the best use of scarce computing resources.
- Consider new sources of information for the terrain model which would contain accurate geometry from man-made structures, such as USGS Point Clouds (*52*) or Google Earth 3D Data. These data sources would be challenging to use, as USGS Point Clouds are unprocessed sensor scans, and Google Earth 3D Data is not licensed for use in research projects.

- Consider new methods for determining the extrinsic and intrinsic parameters of the traffic cameras, possibly coupled with the new data sources mentioned above.
- Increase the performance of the tracking component, enabling downstream tasks to use a larger proportion of the detected trajectories and minimize the significant impact of vehicles which were assigned multiple identifiers.

4. ANALYSIS OF POLICE REPORTED AND UNREPORTED CRASHES IN VIRGINIA TRAFFIC CAMERA VIDEO

Introduction

Traffic cameras, operated by states for the good of the public, have the potential to directly observe crashes, providing an untapped source of information to perform reconstructions. Correlating these crashes with police reports would allow for the integration of occupant, vehicle, and environment information to analyze cases for injury risk modeling, active safety benefits estimation, and other study designs. Such capabilities may allow for proactive safety evaluations that provide an ability to identify and remediate safety deficiencies faster than traditional crash databases.

Analysis of traffic camera video can uncover information about vehicle behavior comparable to that which is available from event data recorders (EDRs). Modern crash datasets such as CISS contain information from the EDR when retrievable. However, in crashes where EDRs are available from multiple vehicles, little information is present to synchronize the time series data between EDRs and the moment of impact, limiting the accuracy of crash reconstructions. Because traffic cameras record position information for all vehicles simultaneously, information such as acceleration, travel speed, and braking magnitude can be compared between vehicles and pedestrians in multivehicle crashes and related to a common moment of impact. EDRs record a fixed, limited amount of pre-crash information at a low sample rate, required by federal law to be at least 1 Hz, typically 2 Hz, and 10 Hz in some vehicles. Because the traffic camera video was recorded continuously, the amount of pre-crash information

was only limited by the time each vehicle was visible, and most traffic cameras recorded at 15 Hz, providing enhanced information for highly dynamic events. Not all vehicles are equipped with EDRs; in 2022 63% of registered vehicles had an EDR accessible by the Bosch CDR tool (*53*). The Bosch CDR tool reads most vehicle EDRs and is used for collecting EDR information in CISS. In traffic camera video, the vehicle make/model/model year does not affect the availability of trajectory information. Finally, EDRs only permanently save information if an impact delta-v greater than 5 km/h is recorded or the airbags are deployed. EDR data is not routinely collected for minor crashes, and if collected, the desired information may not be saved in the EDR memory. Because traffic camera trajectory information is available in crashes of all severities, traffic camera data offers a method to determine trajectories in low severity and near-crash cases. Additional context, such as traffic density, lighting levels, weather, the behavior of non-crash vehicles, and other information is available which is not present in EDRs or traditional crash databases. Because of these factors, a large number of crash attributes which could typically only be obtained by recovering the vehicle EDR are available to study in crashes observed by traffic cameras.

Dash cams, commercial telematics systems, and naturalistic driving studies, such as SmartDrive, Lytx, and the SHRP 2 NDS, have the ability to capture much of the same information as EDRs in crashes by recording forward-facing video and vehicle electronic data. The primary limitation of these systems is the imbalance of data; the front of the subject vehicle is well captured, but video from other areas is often limited or nonexistent. Among crash vehicles in police-reported cases, 47% did not have an initial contact point in the front of the vehicle (defined to be clock points 11, 12, and 1) (*54*), and thus critical information may not be recorded. Traffic cameras, with

their overhead perspective, have the ability to capture more context and a greater proportion of pre-crash information.

National and state crash databases only contain police reported crashes. By directly observing the roadway, traffic cameras do not rely on police reports and have the ability to capture detailed information about non-police reported crashes. Previous studies have investigated unreported crashes. The MDAC study (*55*) surveyed a nationally representative sample of households in 2010 for those which had a crash in the preceding year. They estimated approximately 30 percent of crashes were unreported. Blincoe et al. (*56*) expanded upon this analysis to conclude that in total 7,546,213 crashes (53%) were unreported to police in 2019. In these studies, information about the crash mode was unknown. Understanding crash mode is critical to evaluating the future crash safety problem, as new and proposed crash prevention systems are designed to activate for specific crash modes. By using traffic cameras, we can characterize the types of unreported crashes found in Virginia. National estimates of total crashes can then be estimated by crash type.

The objective of this chapter was to identify non-police-reported crashes in traffic camera video and compare them to police reported crashes. With the analyses described, this dissertation establishes traffic camera data as a viable method for substantively increasing the data available for a variety of future work.

Methods

The source of video in this work was the VTIS (*23*) described in Chapter 2. Videos were stored in hour-long segments for each camera. Some video files were missing or shorter than one

hour in cases where the cameras or video capture system were not functioning; however, the overall sample captured remains sufficiently large to support statistically robust analyses.

Crash Identification

The source of data for crash identification was the crash map layer in the Virginia Traffic Information System, described in Chapter 2. When an event with the text "crash" or "accident" in the description was found in the crash map layer for the first time, the location of this event as indicated by VDOT was compared to the locations of the traffic cameras in the dataset. If an event was found to have occurred within 800 feet of a traffic camera, the entire hour-long segment for that traffic camera was marked as a candidate video segment. A candidate crash distance of 800 feet was chosen as vehicle behavior was found to be visible no further than 800 feet from most cameras.

Each hour-long segment was then manually reviewed for crashes. A crash was included in this study if the type of crash and moment of crash were clearly discernable. A crash was defined to be any event where the body of a vehicle (excluding the tires) contacted the environment, another vehicle, a pedestrian, or an animal. Crashes which were occluded by vehicles or infrastructure were not included, and crashes which occurred too far away from the camera to determine the crash type were not considered. Loss of control events which did not result in road departure were not considered crashes. While VDOT did not have clear publicly-stated criteria for what was considered a crash at the time of data collection, it appeared to include any roadway event which affected the flow of traffic, so not all events identified by VDOT were included in the dataset for this chapter.

A separate analysis was performed to identify pedestrian impacts. Because pedestrian impacts typically do not affect the flow of traffic, no VTIS events were pedestrian impacts at intersections. In order to study pedestrian impacts, PARs from the Commonwealth of Virginia during the study period were queried. PARs were selected where a pedestrian impact occurred within 800 feet of a traffic camera. These candidate PARs were manually reviewed to determine if a pedestrian impact was visible.

Crash Data

For each crash, the crash type for the first event was coded using the same categories as the ACCTYPE variable in the NASS/CDS (*57*). Additionally, the number of ambulances which responded to the crash was coded. This was used as a proxy for severity that was directly observable in the video. The presence of ambulances could not be identified for video clips where these vehicles arrived after the video clip ended, which occurred in 17% of cases. Finally, the presence of rollover was coded for any vehicles in each case. When possible, a Virginia PAR was matched with each identified crash using its time and location and KABCO crash severity was extracted. Cases where no PAR could be matched were considered unreported crashes.

Reported and Unreported Crash Estimation

The total number of crashes was estimated in the United States by using the reporting rate and the formula given below, where CR_K was the total number of crashes of type k, $CRSS_k$was the number of crashes in CISS of type k, and PAR_k was the proportion of VT-CAST 2020 crashes of type k which had police reports. CRSS did not use the ACCTYPE system in NASS/NDS, so the ACCTYPE value for each crash was re-encoded to the equivalent manner of collision (MANCOL) value in CRSS.

$$CR_k = \frac{CRSS_k}{PAR_k}$$

Results

A total of 13,969 candidate video segments were identified from the Virginia Traffic Information System and manually reviewed for crashes. During September and October 2020, the event reporting system was deactivated without the authors' knowledge, and a new web page presenting similar information was identified in November 2020. This created a gap in time where no crashes were identified.

Despite the relatively infrequent placement and narrow field of view of traffic cameras, some cameras were "in the right place at the right time" and a total of 292 crashes were identified within the traffic camera dataset using the VTIS method. Only 2.1% of candidate video segments contained crashes. For the other cases, the crash did not occur within the camera view or the event identified by VDOT was not considered a crash for the purposes of this chapter. An example frame showing a crash involving heavy vehicles is depicted in Figure 16. A total of 134 pedestrian impact PARs were found within 800 feet of a traffic camera. After manual review, six pedestrian impacts were visible.

The crash types identified in the dataset are described in Table 7. The most common crash types identified were road departure, rear-end, angle/sideswipe, and intersection. Forward impact crashes, fixed object, and pedestrian crashes were less common. In 2019, single-vehicle crashes represented 28% of police reported crashes nationally in CRSS 2019 (*58*) compared to the 38% of road departure crashes identified in this dataset. Rear-end crashes represented 36% of crashes in both CRSS 2019 and this dataset. This suggests that this dataset could be representative of road

departure and rear-end crashes as they occur nationally. However, intersection and intersection related crashes represented 9% of the cases in this dataset and 45% of the cases in CRSS 2019; likely as a consequence of camera location.

Figure 16. A crash between two tractor-trailers which occurred at the intersection between US 58 and US 501 in South Boston, VA at 10:31AM local time. The tractor on the left was turning across the intersection and had the right of way while the tractor on the right was traveling left and ran the light.

The proportion of crash types which had police reports is also detailed in Table 7. Approximately 41% of crashes identified through traffic cameras were police reported. The police reporting rate was 32% among single-vehicle crashes and 50% among multi-vehicle crashes, and strikingly the rate among intersection crashes was 100%. Blincoe et al. (*56*) estimated that 40.3% of property damage only crashes and 68.1% of injury crashes were reported to police.

In CRSS 2019, 6.5% of crash occupants were transported from the scene of the crash to a medical facility. However, in this dataset, ambulances were present in 31.8% of cases, although whether any individual used the ambulances for medical transport was unknown. This suggests that there were many cases where ambulances did not transport crash participants, or that the crashes identified in this dataset could be more severe than the national sample of police reported

crashes identified in CRSS. Rollover occurred in 6.9% of cases in this dataset. In CRSS 2019, 3.2% of cases involved a rollover.

Table 7. Crash types identified in traffic camera video, and the proportion of cases in each crash type which were reported to police. The percentages of police reported crashes are computed for each crash type and do not add up to 100%. † Pedestrian crashes were found directly from police reports.

Type	Count	%	Police Reported	Police Reported %
Road Departure	111	37.5%	36	32.4%
Rear-End	104	35.1%	39	37.5%
Angle or Sideswipe	29	9.8%	14	48.3%
Intersection	25	8.4%	25	100.0%
Forward Impact	10	3.3%	3	30.0%
Pedestrian	6	2.0%	6	100.0% †
Other/Unknown	6	2.0%	2	33.3%
Struck Fixed Object	6	2.0%	2	33.3%
Total	**298**	**100.0%**	**127**	

Among crashes with police reports, the distribution of KABCO severity is given in Table 8. Property damage only crashes (those with no apparent injury) occurred most frequently (63% of cases). Minor injury cases were less common (24.2% of cases), followed by serious injury cases (6.7% of cases) and possible injury cases (4.2% of cases). The dataset included two cases with fatal injuries as determined by police officers. The proportion of KABCO scores represented in the VT-CAST 2020 dataset was comparable to the proportion of KABCO scores across all Virginia PARs during the period which traffic camera video was collected. This suggests that an unbiased distribution of Virginia crash severities was sampled.

Table 8. Injury severity in crashes with police reports identified by the VTIS method, compared to the proportion of KABCO scores in all Virginia crashes during the study interval.

Severity	Count	%	All VA PARs %
O – No Apparent Injury	76	63.3%	67.1%
C – Possible Injury	5	4.2%	8.0%
B – Suspected Minor Injury	29	24.2%	19.1%
A – Suspected Serious Injury	8	6.7%	5.2%
K – Fatal Injury (Killed)	2	1.7%	0.7%
	120	**100.0%**	**100.0%**

To relate the crash outcomes in this dataset to national databases, the ratio between CRSS 2019 manners of collision and manners of collision in this dataset was computed (Table 11). Assuming the reweighting factor to be 1.00 in cases where it could not be computed, the total number of reported and unreported estimated crashes in 2019 is given in Table 12. Approximately 14,928,545 crashes occurred in 2019, considering the reweighting factors computed from traffic camera crashes. This compares favorably to the estimate in Blincoe et al. (*56*) for 2019 of 14,193,727 total crashes. However, this estimate contains information about the crash mode: While front-to-rear crashes and single-vehicle crashes still represent the most common manners of collision, the third most common manner of collision among police reported crashes was angle crashes, while the third most common manner of collision estimated among all crashes was sideswipe – same direction crashes.

Table 9. Comparison of CRSS 2019 Manner of Collision and VT-CAST 2020 Manner of Collision in cases identified by the VTIS method.

Manner of Collision (MANCOL_IM)	CRSS 2019	VT-CAST 2020	VT-CAST 2020 % Police Reported	Reweighting Factor CR_k
Single-Vehicle	28.3%	43	35.2%	2.84
Front-to-Rear	32.5%	42	36.8%	2.71
Front-to-Front	2.7%	0	0.0%	-
Angle	22.1%	25	100.0%	1.00

Sideswipe – Same Direction	11.4%	14	48.3%	2.07
Sideswipe – Opposite Direction	1.6%	0	0.0%	-
Rear-to-Side	0.9%	0	0.0%	-
Rear-to-Rear	0.1%	0	0.0%	-
Other	0.6%	2	33.3%	3.00
Total	**100.0%**	**292**	-	-

Table 10. Total estimated crashes in 2019 by manner of collision using the reweighting factors computed from VT-CAST 2020.

Manner of Collision (MANCOL_IM)	**CRSS 2019 Police Reported**	**Total Estimated Crashes**
Front-to-Rear	2,194,966	5,957,765
Single-Vehicle	1,909,264	5,416,982
Sideswipe – Same Direction	770,185	1,595,383
Angle	1,492,532	1,492,532
Front-to-Front	179,532	179,532
Other	38,495	115,485
Sideswipe – Opposite Direction	108,291	108,291
Rear-to-Side	58,964	58,964
Rear-to-Rear	3,865	3,865
Total	**6,755,841**	**14,928,545**

Discussion

This chapter showed that rare events such as crashes could be analyzed in video form by capturing massive amounts of video and filtering. A large number of crashes were observed and studied for this analysis which were not reported to police, particularly single-vehicle crashes. It is important to consider the distribution of camera locations given earlier in Table 5 from Chapter 2; if a certain crash mode was disproportionately unreported to police on local roads, for example, the reweighting factor CR_k would be underestimated. This might occur if lower severity crashes occurred on local roads.

Automotive stakeholders such as manufacturers, safety advocates, ratings agencies, NHTSA, state DOTs, and others prioritize their actions based on the relative occurrence and severity of crashes. Not all actions apply to every crash mode, so an understanding of the relative occurrence of crash modes is essential. This work suggests that sideswipe crashes occur more frequently than angle crashes, in opposition to nationally representative samples of police reports, because angle crashes are well covered by police reports and sideswipe crashes are not. As such, these data may provide a better estimate of overall crash prevalence and distribution for a specified location than traditional PAR-based databases. Stakeholders should consider this difference when evaluating their priorities and applying the traffic camera research methods.

Future work could expand the scope of data collection. The fifteen states which offer traffic video (Figure 1, Chapter 1) had 15,302 traffic fatalities in 2020, 39% of all traffic fatalities that year (*59*). A more nationally representative understanding of crashes could come from considering these camera installations across the United States, not just in Virginia. Because many of the cameras are placed at regular intervals along highways, they are not necessarily placed to capture areas which have a disproportionate number of crashes. By investing in camera installations, state transportation agencies have the ability to capture a greater number of crashes and perform analyses beyond that which is possible from police reports.

Conclusion

The objective of this chapter was to directly observe crashes found in Virginia Traffic camera video. A total of 298 events were successfully identified from traffic camera video. Road departure and rear-end crashes occurred in similar proportions to national databases, but

intersection crashes were underrepresented and severe and rollover cases were overrepresented. This shows promise as a future data source for crash reconstruction, particularly for specific geolocations which are equipped with traffic cameras. Among these crashes, 32% of single-vehicle crashes and 50% of multi-vehicle crashes did not appear in the Virginia crash database. An estimated 14.9 million police reported and unreported crashes occurred in the United States in 2019, concurring with previous estimates but uncovering the underlying distribution of crash types. The distribution of all crashes was similar to that of police reported crashes, with the distinction that sideswipe crashes occur more often than expected from police reports, outranking angle crashes in incidence.

5. PROOF OF CONCEPT QUANTITATIVE CHARACTERIZATION OF DRIVER BEHAVIOR IN INTERSECTION TRAVERSALS

Introduction

Automated vehicles, in general, operate model predictive control (MPC) schemes, where information from the environment is gathered to inform short-term predictions of their own behavior and that of surrounding traffic actors. The optimal control signals (throttle position, indicator lamp status, brake pressure, steering torque, etc.) to achieve short-term navigation goals, obey the law, and avoid collisions are chosen based on these predictions. Computing accurate short-term predictions of traffic actors is therefore an essential component of these systems.

Road user behavior can be directly observed by driving the instrumented vehicles (often the automated vehicle with or without automation active), collecting data, and synthesizing models from the sensor data captured. Public road user data directly observed by driving automated vehicles has been released by Waymo (*60*) and other developers. Such methods are useful; however, another way to collect this road user behavior is to study traffic camera video. Traffic cameras are installed in a variety of locations, allowing the analysis of highway driving, surface street driving, pedestrian walking, and cycling. The large volume of traffic camera video available allows for significant statistical power and the ability to validate road user behavior in rare situations at a much lower cost than acquiring a comparable volume of information from vehicle-based collection. Clusters of driver behavior can be determined by computational methods in a data-driven way, considering the full range of behavior analyzed in the VT-CAST 2020 dataset and capturing the natural variation and multitude of influences on driver behavior.

The objective of this chapter was to investigate the efficacy of traffic camera data for understanding driver behavior. As a proof of concept, this analysis developed a family of curves to describe the behavior of drivers traversing intersections with the aim of creating a process and reference dataset for developers of automated vehicle control systems.

Methods

Trajectories

The source dataset for this task was the trajectories determined from Chapter 3. Only camera views with trajectory information were considered; these were further filtered to only consider those camera views which covered intersections. Segments of vehicle trajectories were selected where drivers entered an intersection and at least two seconds of trajectory information was available both before and after entering the intersection.

The intersection boundary was manually annotated using the JOSM software package (*61*). The intersection boundary when entering the intersection was defined to be the location of the stop bar for the purposes of this chapter. In the absence of the stop bar, asphalt seams, road wear, and the positions of vehicles in reference satellite imagery were used to estimate the typical location where stopped drivers would stop. Dashed lines at the edge of a traffic circle were used as the intersection boundary in those cases.

Model Development

To develop a family of curves describing driver behavior, the k-means statistical clustering technique was used as implemented in scikit-learn (*62*). The dataset for the clustering algorithm was the vehicle velocity sampled at 15 Hz for both two seconds before and after intersection entry,

along with the moment of intersection entry. Each timestep represented a distinct model feature, so 61 model features were present. Two seconds was chosen to balance the extent of the vehicle trajectory with the availability of contiguous trajectories. The number of long-duration trajectories was limited by the performance of the vehicle tracking component in Chapter 3. The number of clusters was selected so that the RMS error of individual trajectories to their assigned clusters, on average, was equal to the median velocity measurement error determined in Chapter 3. Increasing the cluster count further would serve primarily to capture measurement noise.

Results

After considering camera views of intersections with trajectory information, 59 intersections were included in the analysis for this chapter. All intersections were signalized, except for one traffic circle. An exemplar image from the traffic circle camera, located on US-58 in the Virginia Beach area, is available in Figure 17. The number of lanes of traffic present at each intersection are given in Figure 18. Most intersections had between five and eight lanes.

Figure 17. An exemplar frame from the only traffic circle included in the dataset.

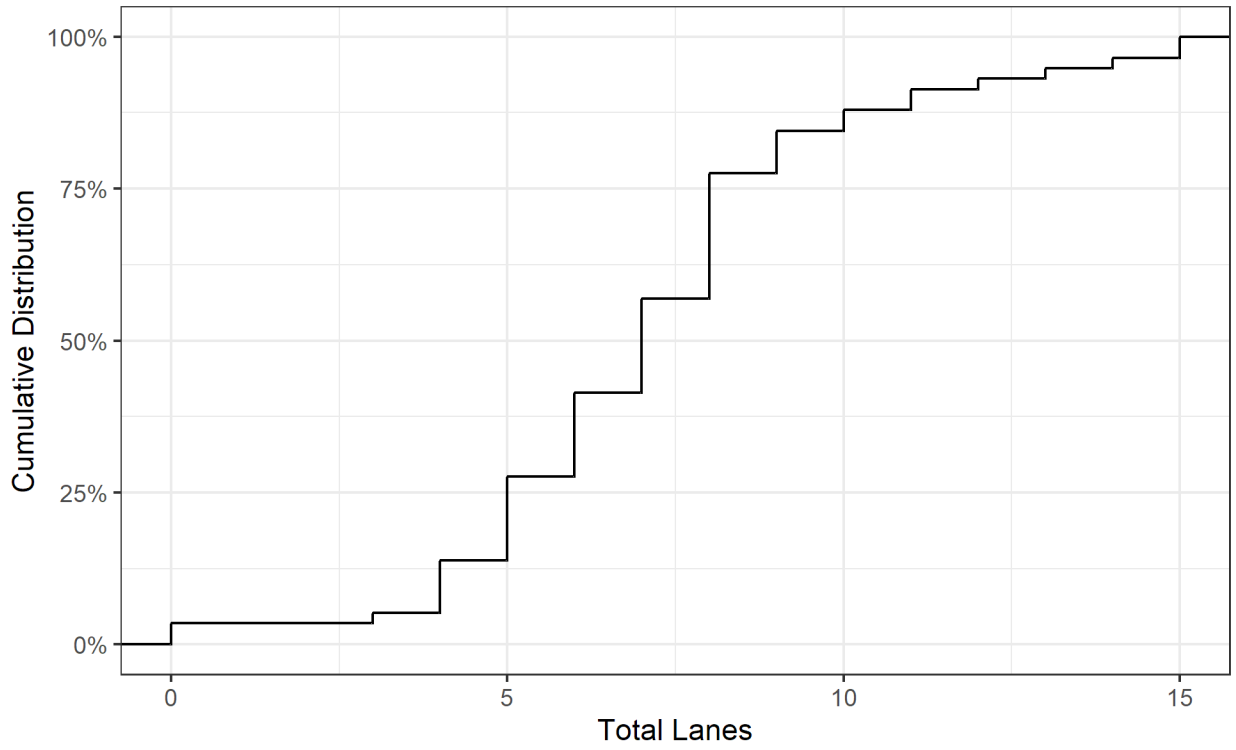

Figure 18. Distribution of the number of lanes at each intersection considered for analysis in this chapter.

A total of 58,180 eligible vehicle trajectories were found from the selected camera views. Approximately 36,990 hours of video were considered. To model driver behavior, 45 clusters were selected. The family of trajectories (centroids of the k-means clusters) is given in Figure 19. Each trajectory in the source data was associated with a cluster. The largest cluster represented 6,099 trajectories, 10.5% of all trajectories considered. The smallest cluster represented 154 trajectories, 0.3% of all trajectories considered.

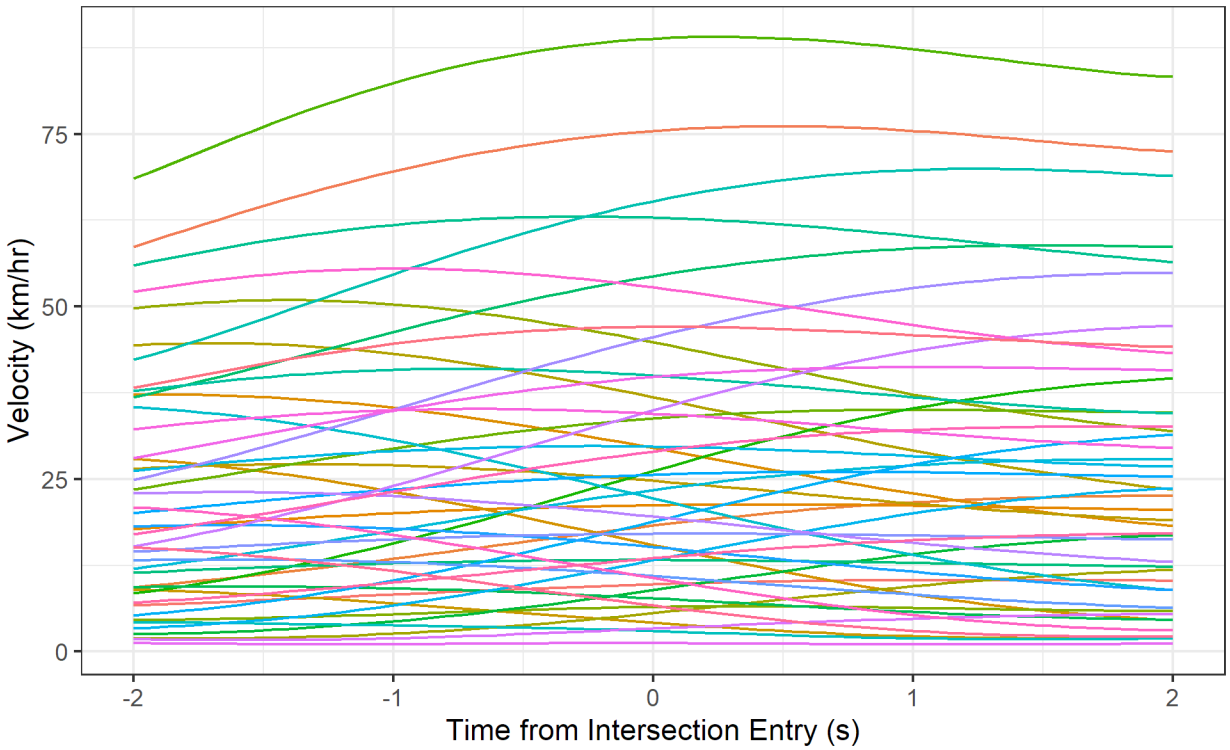

Figure 19. Centroids of the algorithmically generated clusters found in the velocity of vehicles entering intersections.

Three gross behaviors were demonstrated: vehicles which decelerated, vehicles which traveled at approximately the same speed, and vehicles which accelerated. These categories are presented separately in Figure 20. Trajectories with an average acceleration between -0.1 m/s^2 and 0.1 m/s^2 were considered to have constant speed. Vehicles which decelerated at the time of intersection entry were generally traveling lower speeds than those accelerating into the intersection, suggesting those vehicles started slowing earlier than two seconds before intersection entry. The relative frequencies of these three trajectory modes are given in Table 11. Accelerating and decelerating trajectories occurred in similar proportion to each other, and more often than constant speed trajectories.

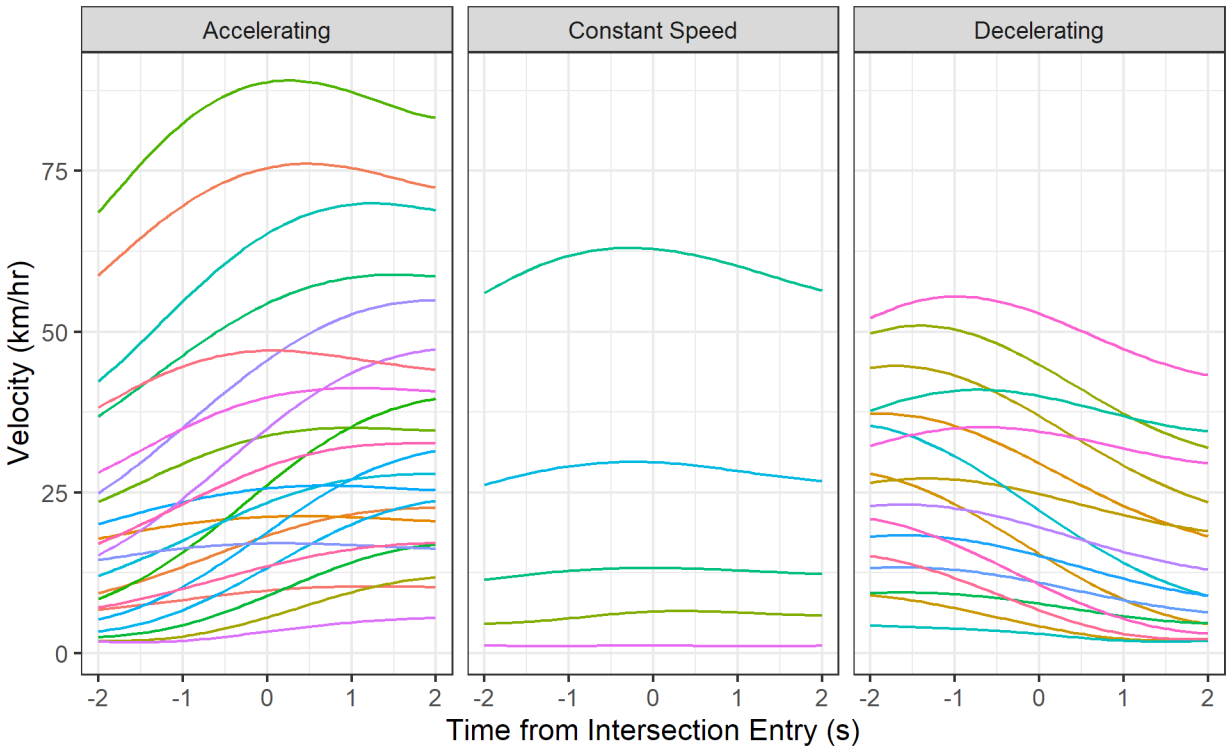

Figure 20. Centroids of the algorithmically generated clusters found in the velocity of vehicles entering intersections.

Table 11. Relative frequency of trajectory types.

	Number of Trajectories	**%**
Accelerating	24,507	42.1%
Constant Speed	11,762	20.2%
Decelerating	21,911	37.7%
Total	58,180	100.0%

Discussion

Arguably the most significant factor in driver behavior around signalized intersections is the phase of the light experienced during the intersection approach. Signal phase information was not available directly in this study. However, for some camera views, the traffic signals themselves were visible. An example of this is given in Figure 21. In this example, the signal was not visible

for the intersection approach that was visible; the *north*bound signal was visible, but the *south*bound road was visible. There were no camera views where an intersection approach was visible alongside the signal controlling that approach. A future study could algorithmically observe the traffic camera video to determine which traffic signal is lit and further analyze driver behavior in the limited population of traffic camera views with signal phase information, considering that signal information is only indirect, like in Figure 21. A significantly better approach would be to directly receive signal phase and timing data streams from the signal controllers. Such information was not readily available at the time of data collection; however, it may be feasible to obtain such data in the future as Virginia is migrating to a statewide signal system. Substantially more meaningful analyses of the vehicle approach trajectories could be performed if the exact signal phase and timing synchronized with each trajectory was treated as a covariant.

Figure 21. A camera image from the intersection between Main Street and Chain Bridge Road in Fairfax, VA.

The method of aligning driver trajectories used in this chapter was most useful at characterizing driver behavior when vehicles stop at the intersection boundary. Drivers frequently

stop before, near, or beyond this location, and future work could consider alternate trajectory alignment points.

Previous work by Noble et al. and Scanlon has considered the trajectories of vehicles at intersections in the 100-car NDS (*63*, *64*). This chapter considered a greater number of trajectories (58,180 vs. 41,479), however the Noble et al. and Scanlon studies considered a significant number of intersections with stop signs. No such intersections were present in the dataset of cameras considered for this study. This study improves upon the work by Noble et al. and Scanlon by considering a larger dataset and a family of driver trajectories, each of which closely aligns with the behavior of a distinct group of driver trajectories, in contrast to developing a single 'master' trajectory with high statistical variance.

Three gross behaviors were characterized in this analysis: accelerating, constant speed, and decelerating trajectories. To physically interpret the causation of these behaviors and more finely divide them, future work could consider signal phase, ultimate destination of the vehicle, traffic volume, intersection size, and other environmental factors.

Expansion of the trajectory dataset developed in Chapter 3 could enable an updated version of the study presented in this chapter. Many vehicles were re-assigned identifiers as they traveled, instead of receiving one stable identifier. Reducing vehicle re-identification would increase the proportion of trajectory data amenable to this analysis. A future version of this study could also consider the effects of lead vehicles on the behavior of drivers; one might expect drivers to behave differently between car following situations and free driving. The number of k-means clusters selected was limited by the detection accuracy of the camera system. A future study with better vehicle detection algorithms or higher resolution video could potentially include more clusters and

model more subtle driver behaviors. Future work could additionally consider the following factors in intersection driver behavior:

- time of day (day/dawn/dusk/night);
- weather;
- the action taken by the driver (traveling through, turning left, turning right);
- and the size of the intersection and lane of travel of each vehicle.

Conclusion

The objective of this chapter was to investigate the efficacy of traffic camera data for understanding driver behavior. As a proof of concept, this analysis developed a family of curves to describe the behavior of drivers traversing intersections with the aim of creating a process and reference dataset for developers of automated vehicle control systems. A total of 58,180 trajectories of intersection entry were grouped into 45 trajectory clusters. Distinct subgroups of accelerating, constant speed, and decelerating trajectories were present. Accelerating and decelerating trajectories each occurred more frequently than constant speed trajectories. This study improves upon the work by Noble et al. and Scanlon by considering a larger dataset and a family of driver trajectories, each of which closely aligns with the behavior of a distinct group of driver trajectories, in contrast to developing a single ‘master’ trajectory with high statistical variance. Future work should consider signal phase and timing, environmental factors which would explain the causes of driver behavior, time of day, weather, and other factors. Higher quality trajectories developed in future work could enable more fine-grained clusters.

5. PROOF OF CONCEPT: QUANTITIVE MODEL OF DRIVER-PEDESTRIAN INTERACTION AT CROSSWALKS

Introduction

Future automated vehicle systems will need an understanding of pedestrian behavior to perform normal driving and prevent collisions. Traditionally, information on pedestrian behavior has come from observational studies in which pedestrians were observed at a fixed number of locations for a fixed period of time. Bosina and Weidmann (*65*) present a detailed review of pedestrian observational studies. Capturing the paths and intentions of pedestrians is not only important for road safety work, but is also central to planning walkways, movement in public spaces, and evacuation scenarios. Kerridge et al. (*66*) discuss a method for automatically collecting pedestrian trajectories using several linear thermal imaging sensors. Trajectory information was available at 30 Hz for up to 10 people, and the update rate was lower when more pedestrians were visible. No accuracy information was given.

Previous observational studies (*65–67*) have relied on positioning cameras or observers near pedestrian crossing locations and observing behavior. Because the researchers in these studies position the cameras themselves, the selection of data collection sites can be driven by convenience or other operational concerns. In this work we propose to overcome this limitation by observing pedestrian behavior using pre-existing traffic cameras in the Commonwealth of Virginia. These traffic cameras are located across the commonwealth and no physical setup is required for their use. This allows the locations to be easily selected and changed as part of the sample design, and a larger number of cameras were available for analysis than for any one study.

The objective of this chapter was to demonstrate a methodology to determine a family of curves describing the behavior of drivers in the presence of pedestrians using traffic camera video at crosswalk crossings.

Methods

Pedestrian and Vehicle Trajectories

In a similar method as the previous chapter, the source dataset for this task was the trajectories determined from Chapter 3. Only camera views with trajectory information were considered; these were further filtered to only consider those camera views which covered crosswalks. All valid vehicle and pedestrian trajectories were considered where drivers could react to the pedestrian. We defined this to be cases where, at any point during the trajectory, a pedestrian was contained within a sightline arc 180 degrees wide and 40 meters in radius pointed along the centerline and centered at the front of the vehicle. A radius of 25 meters was used at night, to match pedestrian detection range estimates by Haus and Bhagavathula (*68, 69*). Valid trajectories were those where at least one second of trajectory data was available. Trajectories with fewer data points were often mis-detections or could not be fully smoothed by the Kalman filter. One second was chosen to balance the extent of vehicle trajectory with simultaneously available vehicle and pedestrian trajectories. For those trajectories where vehicles and pedestrians were on a collision course, Time to Collision (TTC) was computed. TTC was defined as the ratio between the distance to the shared collision point and the velocity of the vehicle.

Model Development

To develop a family of curves describing driver behavior in the presence of pedestrians, the k-means statistical clustering algorithm was used as implemented in scikit-learn (*62*). Both vehicle clusters and pedestrian clusters were identified. The dataset for the clustering algorithm was the vehicle or pedestrian velocity sampled at 15 Hz for 0.5s before and after the pedestrian proximity point, along with the pedestrian proximity point itself. The pedestrian proximity point was defined to be the point in the trajectory where the driver was closest to a pedestrian. Each timestep represented a distinct model feature, so 15 model features were present. The trajectory duration was chosen to balance the number of trajectories available with the amount of context available in each trajectory. Vehicle velocities were lower when vehicles were interacting with pedestrians, necessitating a different cluster quantity cutoff from the one used in Chapter 5. The number of clusters was selected so that the RMS error between individual trajectories and their assigned cluster was less than 5% of the mean vehicle velocity in each trajectory. This served to reduce the acceptable error for trajectories at low speeds and fully capture their behavior. Also, the difference between each velocity point and the mean velocity was considered instead of the true velocity at each point; this normalized the large range of velocity values considered.

Results

After considering camera views of crosswalks with trajectory information, 35 camera views were included in the analysis for this chapter. A total of 1,488 driver-pedestrian encounters were found from approximately 13,238 hours of traffic camera video. Using the method described in the previous paragraph, 16 vehicle trajectory clusters were selected. The family of trajectories

(centroids of the k-means clusters) is given in Figure 22. Each trajectory in the source data was associated with a cluster. The largest cluster represented 595 trajectories, 40.0% of all trajectories considered. The smallest cluster represented 9 trajectories, 0.6% of all trajectories considered.

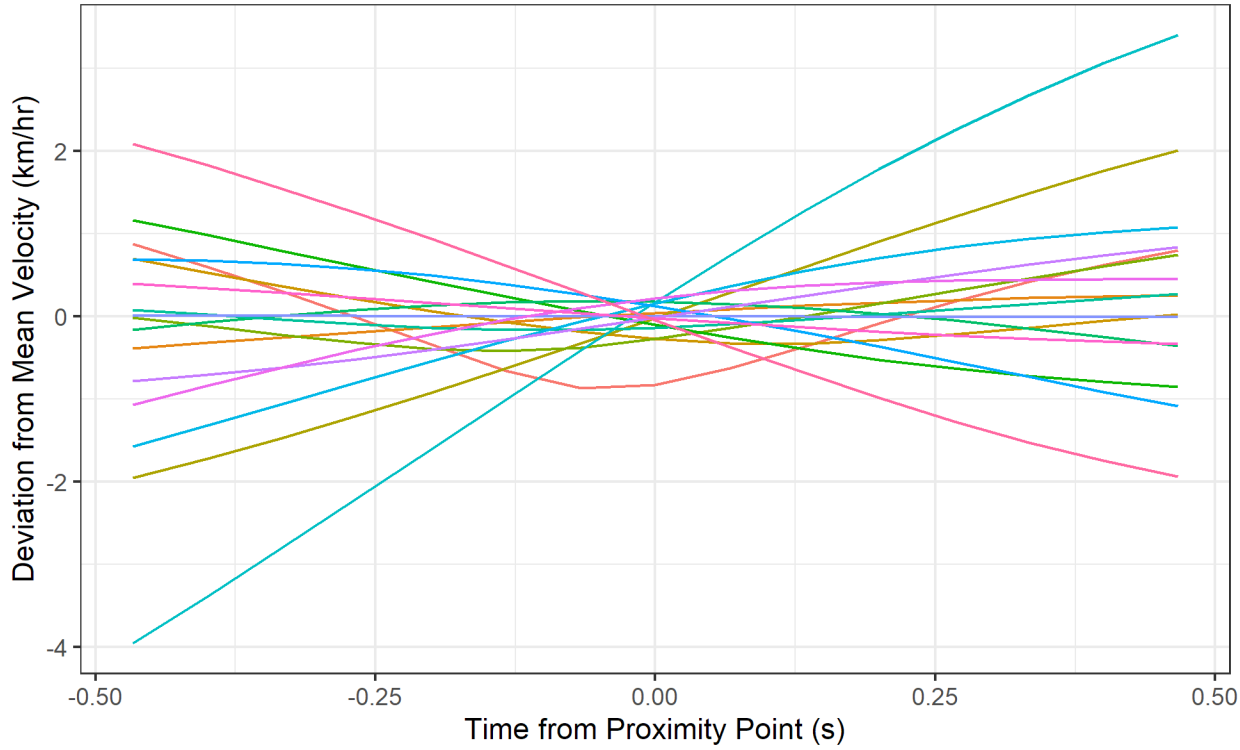

Figure 22. Centroids of the algorithmically generated clusters found in the velocity of vehicles interacting with pedestrians near crosswalks.

In a similar manner to the analysis in Chapter 5, three gross behaviors were shown: vehicles which accelerated, vehicles which traveled at approximately the same speed, and vehicles which decelerated. These categories are presented separately in Figure 23. Trajectories with an average acceleration between -0.1 m/s^2 and 0.1 m/s^2 were considered to have constant speed. Interestingly, one centroid with a mean acceleration near zero shows a significant slowing around the pedestrian of 2 km/hr; the deceleration prior to and acceleration after encountering the pedestrian was 1.1 m/s^2. This centroid represented 9 trajectories, or 0.6% of all trajectories considered. Other

trajectories also showed a slowing around the pedestrian. The relative frequencies of these three trajectory modes are given in Table 12. Constant speed trajectories occurred the most often, followed by accelerating trajectories and decelerating trajectories.

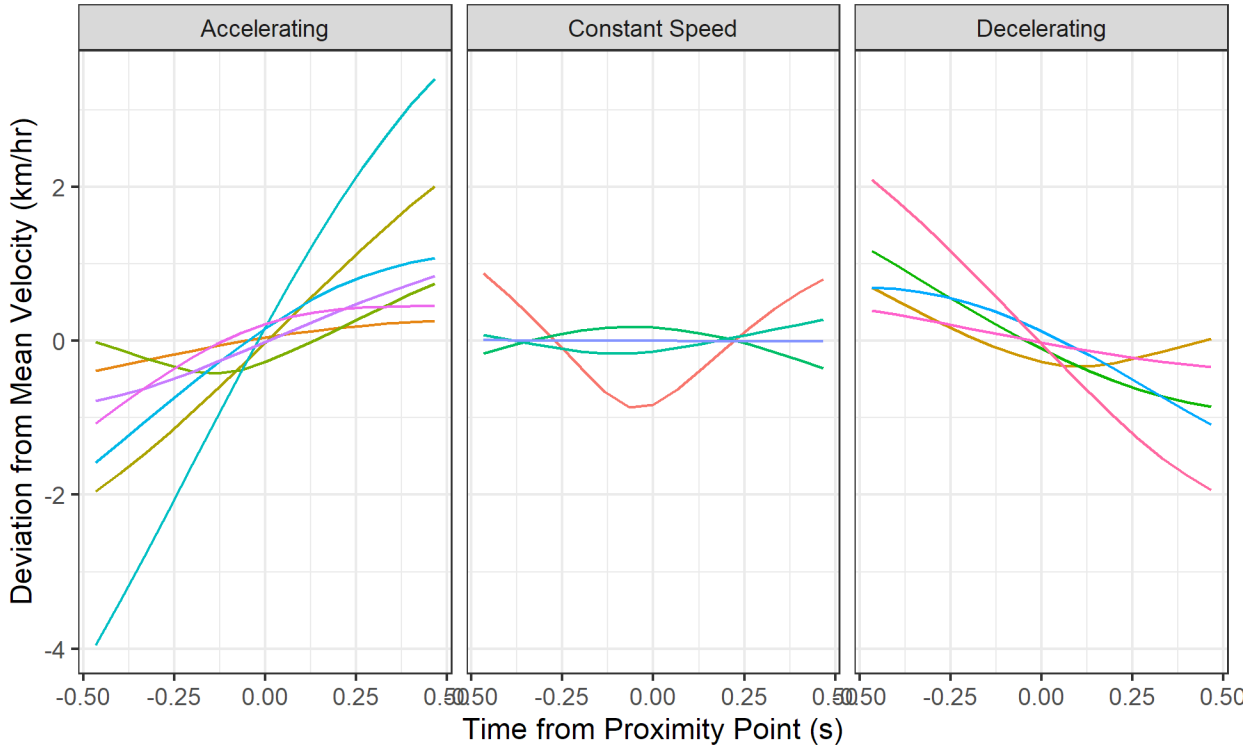

Figure 23. Categories of the centroids of the algorithmically generated clusters found in the velocity of vehicles interacting with pedestrians near crosswalks.

Table 12. Relative frequency of trajectory types.

	Number of Trajectories	**%**
Accelerating	466	31.3%
Constant Speed	717	48.2%
Decelerating	305	20.5%
Total	1488	100.0%

A total of eight pedestrian trajectory clusters were identified. The distribution of pedestrian instantaneous velocities is given in Figure 24. A cutoff of 1.7 km/hr was used to differentiate

motionless pedestrians from those who are moving. This corresponds to the 1st percentile pedestrian speed across all non-stairs datasets considered by Bosina and Weidmann (*65*). The median velocity across all time points was 1.0 km/hr. Approximately 56% of pedestrian time was spent motionless. Among moving pedestrians, the median velocity was 3.4 km/hr. The median reference speed of a North American pedestrian given by Bosina and Weidmann was 5.0 km/hr, faster than that found in this dataset. We propose two possible explanations of this: 1) because these instantaneous velocities include the time some pedestrians spend accelerating and decelerating before and after their motion across the roadway, respectively, a greater proportion of slower motions are included, and/or 2) as these pedestrians are interacting with vehicles, they may not achieve their free-flow velocity. The family of pedestrian trajectories is given in Figure 25. The largest cluster represented 895 pedestrian trajectories, or 60.1% of all trajectories considered. The smallest cluster represented 10 pedestrian trajectories, 0.7% of all trajectories considered. Most trajectories (81%) were approximately constant speed (less than 0.1 m/s^2 of acceleration/deceleration). Approximately 15% of trajectories were accelerating faster than 0.1 m/s^2, and the remainder (4%) of trajectories were decelerating. The trajectory labeled X5 depicts a pedestrian slowing prior to the proximity point; this occurred in 0.6% of cases.

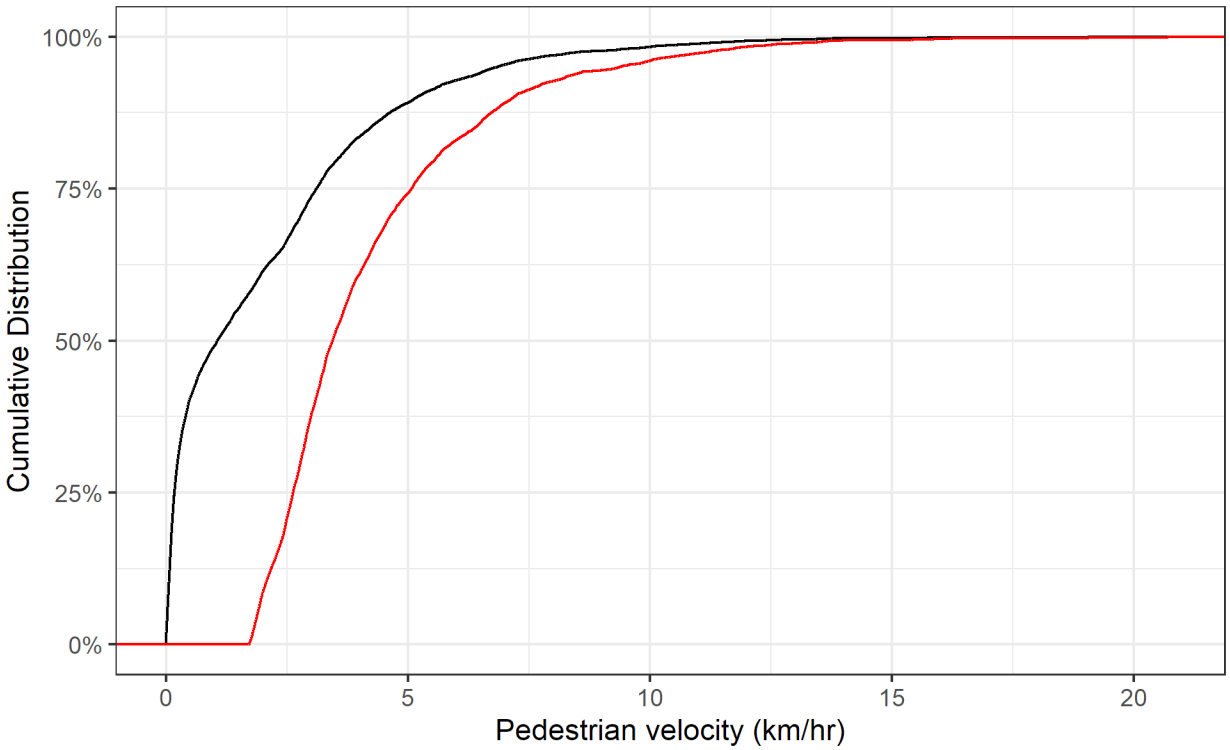

Figure 24. Distribution of instantaneous velocities of pedestrians observed interacting with vehicles near crosswalks. The black line indicates all velocities; the red line indicates all moving velocities (those greater than 1.7 km/hr).

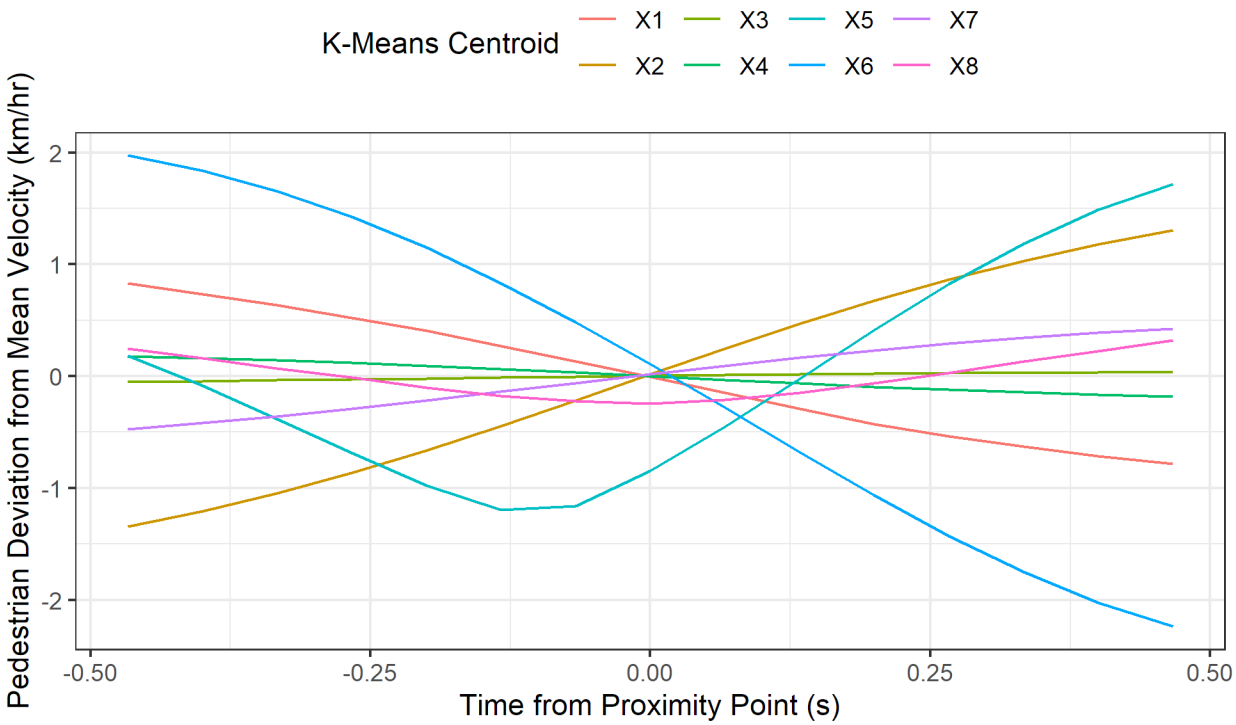

Figure 25. Centroids of the algorithmically generated clusters found in the velocity of pedestrians interacting with vehicles near crosswalks.

For those vehicle-pedestrian pairs on a collision course (1,316 pairs), the distribution of TTC values is given in Figure 26. TTC values were uniformly distributed from approximately TTC=1s to TTC=15s, and infrequent below that threshold.

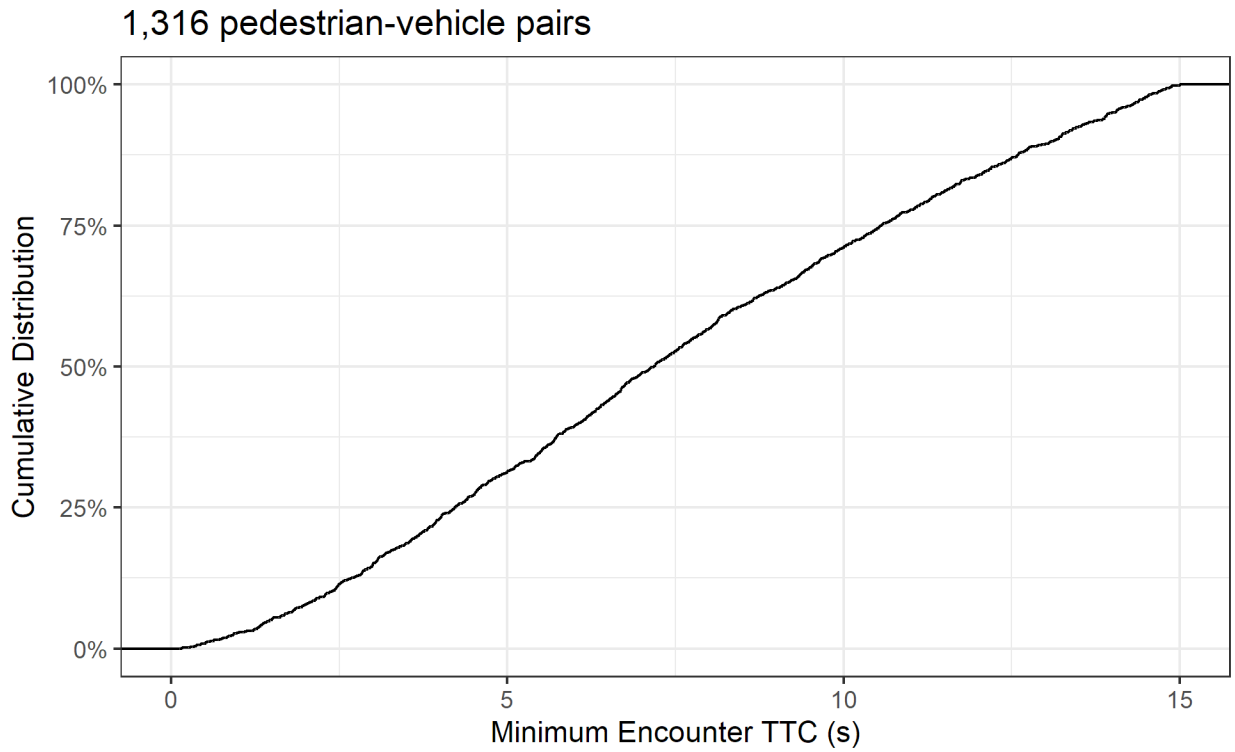

Figure 26. Distribution of minimum TTC values for vehicle-pedestrian pairs which were on a collision course.

Discussion

Similar work was performed by Ismail et al. in 2009 (*17*). This chapter considered a similar number of pedestrians (2,218) as their study (2,100). However, this chapter considered a larger number of sites (35 vs. 1).

Arguably the most significant factor in driver behavior around intersections and crosswalks is the phase of the light and pedestrian walk sign. Signal phase information was not available directly in this study. However, for some camera views, the traffic signals or pedestrian signs themselves were visible. An example of this is given in Figure 27. While the pedestrian signals are visible in this case, the eggcrate pattern and small size of the signal makes the walk/stop status unidentifiable from the traffic camera image. There were no cameras where the pedestrian

walk/stop status was clearly visible. A future study would have to consider an alternate source for signal phase information or use higher resolution video to clearly identify pedestrian signal status. As discussion in the previous chapter, obtaining streams of signal and crosswalk phase information is increasingly feasible as the deployment of advanced signal controllers and intelligent transportation system networks proliferates.

Figure 27. Camera view from the intersection between Carlin Springs Road and the Route 50 exit/entrance ramps.

Expansion of the trajectory dataset developed in Chapter 3 could enable an updated version of the study presented in this chapter. Many vehicles and pedestrians were re-assigned identifiers as they traveled, instead of receiving one stable identifier. Reducing vehicle and pedestrian re-identification would increase the proportion of trajectory data amenable to this analysis and would allow expanding the time horizon considered for each encounter. A future version of this study could also consider the effects of lead vehicles on the behavior of drivers; one might expect drivers to behave differently between car following situations versus directly interacting with a pedestrian. The number of k-means clusters selected was limited by the detection accuracy of the camera system. A future study with better vehicle detection algorithms, pedestrian detection algorithms,

or higher resolution video could potentially include more clusters and model more subtle driver behaviors. Analogously to the previous chapter, future work could additionally consider the following factors in driver-pedestrian behavior:

- time of day (day/dawn/dusk/night), beyond the inclusion or exclusion of cases;
- weather;
- the action taken by the driver (traveling through, turning left, turning right);
- and the action of the pedestrian (walking across the vehicle path, walking along/opposite the vehicle path).

Conclusion

The objective of this chapter was to demonstrate a methodology to determine a family of curves describing the behavior of drivers in the presence of pedestrians using traffic camera video at crosswalk crossings. From 35 crosswalks in the VT-CAST 2020 dataset, all vehicle and pedestrian trajectories were considered where drivers could react to the pedestrian, yielding 1,488 trajectories of driver-pedestrian interaction. K-means clustering was used to group vehicle and pedestrian trajectories into 16 and eight trajectory clusters, respectively. Distinct subgroups of accelerating, constant speed, and decelerating trajectories were present, including trajectory clusters which described vehicles slowing down around pedestrians. Constant speed trajectories occurred the most often, followed by accelerating trajectories and decelerating trajectories. This analysis demonstrates the efficacy of analyzing pedestrian and vehicle interactions at crosswalks using traffic cameras. This potentially powerful research method will enable the study of a very large number of interactions at specific intersection locations at a very low cost relative to

instrumented vehicles or other methods. With the addition of signal phase and timing data, the method will provide detailed information for a variety of applications with a particular opportunity for developing driver/pedestrian models which may be used in automated vehicle or other technology development efforts (e.g., advanced signal systems).

6. CONCLUSION

There exists a need for a database of crashes and normal driving collected recently with detailed information to enable crash and driving reconstruction. In this work, 6,779,726 hours of video from publicly available internet streams were recorded from 1,263 traffic cameras in the Commonwealth of Virginia. The trajectories of vehicles and pedestrians in the video were automatically determined using computer vision techniques. A total of 35,080 vehicle miles traveled and 139 pedestrian miles traveled were reconstructed. Within the dataset, 298 crashes were observed. Road departure and rear-end crashes occurred in similar proportions to those in national databases, but intersection crashes were underrepresented while severe and rollover crashes were overrepresented. Among these crashes, 32% of single-vehicle crashes and 50% of multi-vehicle crashes did not appear in the Virginia crash database. An estimated 14.9 million police reported and unreported crashes occurred in the United States in 2019, concurring with previous estimates but uncovering the underlying distribution of crash types. The distribution of all crashes was similar to that of police reported crashes, with the distinction that sideswipe crashes occur more often than expected from police reports, outranking angle crashes in incidence.

Furthermore, two proof of concept analyses were performed to explore and demonstrate the potential of these data as a large-scale robust source for future safety analyses. First, a total of 58,180 intersection entries were examined and grouped into a family of 45 trajectory clusters. The relative occurrences of these clusters were considered, capturing rare driver behaviors which occurred in 0.3% of traversals. A similar analysis was conducted for driver-pedestrian interactions. A total of 1,488 driver-pedestrian interactions were examined and grouped into a families of 16 vehicle trajectory clusters and eight pedestrian trajectory clusters. The relative occurrences of these

clusters were considered, capturing rare driver-pedestrian interactions which occurred in 0.6% of cases.

The primary benefit of a dataset is to enable a wide range of research, including study designs the dataset authors could not have considered. As evidence, there are already three studies currently in progress (beyond this dissertation) which use the VT-CAST 2020 dataset:

- "Modeling driver behavior and I-ADAS in intersection traversals", which uses the vehicle trajectory information to study traffic conflicts in a manner similar to Chapter 5.
- "Real time risk prediction at signalized intersection using graph neural network", which uses the traffic camera video with alternative algorithm to the one presented in Chapter 3 to ultimately develop a near real-time risk score for a traffic scene (*70*).
- "Methods to encourage slow-moving trucks to travel only in designated lanes", which aims to understand the prevalence of commercial motor vehicles using the left lane in situations violating Code of Virginia § 46.2-803.1, and the impact this may have on other motorists who are driving behind these vehicles by determining the level of impedance.

Limitations of Work

A significant amount of available traffic camera video was not recorded (41.2% of potential recordable hours); failures in capture were mostly random and linked to software and/or hardware failures in the VDOT camera and public video distribution system. This means that not all possible events, such as crashes, were contained in the dataset. Significant improvements in performance are possible based on the lessons learned within this work; particularly if the research team closely

collaborates with the road operator to maximize camera data availability (little direct collaboration was exercised in this case, rather the publicly available data streams were captured via the internet).

The vehicle tracking and coordinate transformation algorithms assumed that the roadway was located on the USGS digital terrain model. This means that vehicles traveling over bridges or in tunnels were not tracked. Additionally, the manual alignment process for the camera tracking component could not be completed for cameras views of road construction or camera views with no distinguishing features.

The median trajectory had a duration of eight frames, or 0.53 seconds at 15 fps. This means that long-term vehicle tracking could only be performed on a small subset of the data, and a naïve count of traffic would significantly overestimate the number of cars and pedestrians at each location, as most vehicles were assigned multiple identifiers. Despite these limitations which reduce the amount of available data, the source dataset of video was so large that the studies in this dissertation could still be completed. Future work which uses the trajectory data must consider these issues.

Due to limitations on available computational resources, not all valid video clips have trajectory information for vehicles or pedestrians. Future studies can complete more computation, and future computing systems (both CPUs and GPUs) may have the ability to complete the trajectory analyses in a shorter period of time, enabling studies designs with higher statistical power that consider more rare behaviors.

The trajectory analyses in this work would be substantially more meaningful with precise signal phase and timing information, for both vehicles and pedestrians. This information is valuable for understanding driver and pedestrian behavior around intersections. With the increased

prevalence and integration of statewide intelligent transportation systems (ITS), especially in Virginia, future studies should consider capturing all relevant ITS data, such as signal phase and timing, pedestrian walk sign phase and timing, and highway variable sign text. Generally, tighter integration with systems provided by the roadway operator could allow for higher quality video, more camera views, better uptime, more data, and other benefits.

Potential Study Designs for Future Work

At the onset of this work, little prior research had leveraged traffic cameras for safety research. However, traffic camera networks are increasingly common across the United States, with many road operators developing regional and statewide systems which capture hundreds to thousands of cameras within a single platform. In some instances, these platforms will also contain various additional data, such as signal phase and timing, roadway weather, and other potential data content of interest. While this study verified the efficacy of the method, there are a number of potential studies which could be conducted using the VT-CAST 2020 dataset and comparable future datasets:

- One potential area of future study is to consider the effect of mobile phone use for pedestrians. Hatfield and Murphy (*71*) previously performed an observational study in which researchers directly observed 270 female and 276 male pedestrian crossings. Female pedestrians engaged in more dangerous behaviors while using a cell phone, such as crossing the street more slowly and observing oncoming traffic less frequently. Male pedestrians crossed the street more slowly at unsignalized crossings when using a mobile phone. A similar analysis could be performed using the dataset in this study. While mobile

phone use was not automatically annotated in the VT-CAST 2020 dataset, some traffic cameras could be selected with pedestrians large enough in the field of view to determine mobile phone use by manually reviewing the video. The advantages of this study design compared to the Hatfield and Murphy study would be ease of selecting a representative sample and potential sample size. A random selection of traffic camera data would provide representative distributions of light level, time of day, weather, road surface condition, and other factors without the logistical hurdle of organizing numerous data collection events. Since the traffic camera video is pre-recorded, times where no pedestrians are visible could be skipped, allowing researchers to consider more pedestrians for their time and conduct a larger study.

- The STRIDE 2012-016S project (*72*) aimed to understand vehicle-pedestrian interaction at midblock crossings. Data was collected from 27 sites across North Carolina, Florida, and Alabama. Sites were chosen across urban and university locations. Sites with more than three lanes, raised crosswalks, or infrequent pedestrian crossings were not considered. Posted speed limits varied from 15 to 40 mph. Researchers manually coded pedestrian-vehicle interactions on a written form. A radar speed gun was used to measure the speed of the potential conflict vehicle. Video recording of the conflict event was later reviewed to measure the duration of pedestrian crossing and review any details which may have been missed at the moment of the event. A separate driving study was also conducted in this project, where researchers installed cameras and datalogging equipment in participant vehicles. The participants were then instructed to drive around one of two routes on a university campus. Similar data elements as the pedestrian study were coded from the

video. A total of 975 vehicle-pedestrian interactions were considered, in which 53.3% of interactions resulted in pedestrian yielding and the remainder resulted in vehicle yielding. State flow diagrams and binary logit models were developed to model the decision process of the driver and pedestrian and the likelihood of accepting a vehicle gap for crossing, respectively. Then, traffic microsimulation models were developed for VISSIM. A similar analysis framework could be applied to the VT-CAST dataset. The VT-CAST dataset would allow for much larger statistical power due to the larger number of cameras. Many of the VT-CAST 2020 cameras would meet the requirements of the STRIDE 2012-016S project, and 25 out of the 30 explanatory variables considered in this study could be determined from the VT-CAST data without any additional manual annotation. With additional manual annotation, it may be possible to determine the other five variables: whether the oncoming vehicle was a heavy vehicle, approximate age of the pedestrian, pedestrian distraction, presence of business attire on the pedestrian, and pedestrian gender.

- Another potential application of the dataset is to automatically perform roadway safety assessment, using rating scales such as the US Road Assessment Program. Previous methods have shown the ability to perform this analysis automatically from panoramic images (*73, 74*).
- Drivers have been found to be more aggressive immediately after congested driving on freely-moving roads (*75*). In that study, drivers in a simulator spent ten minutes in heavy traffic followed by ten minutes in light traffic, or the reverse. Drivers in light traffic following an experience in heavy traffic were found to travel at higher speed, accelerate faster, and change lanes more often. This study, however, was limited in scope due to the

number of participants which could be recruited. Additionally, the relationship between simulator behavior and real-life driving could be questioned due to the absence of real-life stressors caused by travel delays which were not present in the study design. A follow-up study could be conducted using the VT-CAST 2020 dataset. Periods of heavy, congested traffic could be identified from traffic cameras located on interstates. Camera views would then be included in the study design by selecting those which are immediately downstream of the congestion, either on interstates, highways, or secondary roads. Heading rate, velocity, lateral and longitudinal accelerations, and lane position would be compared between the post-congestion driving and normal driving at other times. This would improve upon the Li et al. study by considering a much larger number of drivers, seasonal variation, urban vs. rural traffic behavior, and other real-world factors which would not be present in a simulator.

- In Thapa et al. (*76*), a method was described to evaluate the effectiveness of work zone signage, including variable message signs. The SHRP 2 NDS was used to identify 299 time series traces across 25 work zones. The authors found that the ends of lanes, speed limit signs, and variable message signs reduced the speed of drivers as they passed the point at which they were legible. The video capture method in this dissertation could be used to perform a similar study while the work zone is active and report effectiveness to state DOTs. This would allow states to take action to adjust or supplement signage which does not meet standards of effectiveness.
- The Hutchinson and Kennedy (*77*) and Cooper (*78*) studies represent some of the best available measurements of encroachment events, those where drivers leave their lane of

travel and depart the road. These studies were conducted in 1966 and 1980, respectively. Vehicles and roadside design have evolved significantly since this time, so there exists the need for an updated dataset of encroachment events. Encroachments could be detected at each of the 1,263 camera locations in the VT-CAST 2020 dataset, significantly more than the five sites in the Hutchinson and Kennedy study or the 59 sites in the Cooper study. Departure angle, re-entry angle, departure speed, and other departure relevant factors could be automatically determined from vehicle trajectories. Other studies (*79*) have considered encroachments in the SHRP 2 NDS. A complementary study performed using the VT-CAST 2020 dataset would allow a more detailed analysis of the roadside hardware present for each encroachment, as a fixed, smaller amount of roadside hardware would be considered.

Final Thoughts

In summary, a large dataset of publicly available streaming traffic camera video from the Commonwealth of Virginia was recorded using a system limited by the software and hardware available. From this, the trajectories of vehicles and pedestrian in the video were automatically determined using computer vision techniques, limited by the software and hardware available. The video was also manually reviewed to find crashes, limited by the manual review time available. The proof-of-concept analyses of intersection and driver-pedestrian behavior were limited by the trajectories available (software and hardware limits) and the resolution of the video. Technological progress is rapid; this alone bodes well for future attempts at similar projects.

7. PUBLICATION SUMMARY AND PLAN

The following studies were presented as a part of my PhD program:

- "Method to Capture Naturalistic Pedestrian Behavior in Virginia Traffic Camera Video", presented at the 2020 AAAM Annual Conference.
- "Properties of Virginia Crashes Directly Observed in Traffic Camera Video", presented at the 2021 AAAM Annual Conference.

The following publications and presentations are proposed:

- "Analysis of Virginia Crash Kinematics Directly Observed in Traffic Camera Video"
- "Numerical Models of Driver Behavior at Intersections in Virginia"
- "Numerical Models of Pedestrian-Driver Interactions in Virginia".
- "System for Real-Time Analysis of Vehicle Trajectories, Enabling Real-Time Crash Detection", presentation to VDOT.

REFERENCES

1. Pordes, R., D. Petravick, B. Kramer, D. Olson, M. Livny, A. Roy, P. Avery, K. Blackburn, T. Wenaus, F. Würthwein, I. Foster, R. Gardner, M. Wilde, A. Blatecky, J. McGee, and R. Quick. The Open Science Grid. *Journal of Physics: Conference Series*, Vol. 78, 2007, p. 012057. https://doi.org/10.1088/1742-6596/78/1/012057.
2. Sfiligoi, I., D. C. Bradley, B. Holzman, P. Mhashilkar, S. Padhi, and F. Wurthwein. The Pilot Way to Grid Resources Using GlideinWMS. No. 2, 2009, pp. 428–432.
3. NHTSA. *Overview of the 2021 Crash Investigation Sampling System*. Publication DOT HS 813 397. 2022.
4. NHTSA. *CRSS Analytical Users Manual 2016-2020*. Publication DOT HS 813 236. 2022.
5. NHTSA. *Fatality Analysis Reporting System Analytical User's Manual, 1975-2020*. Publication DOT HS 813 254. 2022.
6. NHTSA. *The National Motor Vehicle Crash Causation Survey*. Publication DOT HS 811 057. 2008.

7. Kastrinaki, V., M. Zervakis, and K. Kalaitzakis. A Survey of Video Processing Techniques for Traffic Applications. *Image and Vision Computing*, Vol. 21, No. 4, 2003, pp. 359–381. https://doi.org/10.1016/S0262-8856(03)00004-0.
8. Wu, F., R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, R. Haulcy, B. Piccoli, B. Seibold, J. Sprinkle, and D. B. Work. Tracking Vehicle Trajectories and Fuel Rates in Phantom Traffic Jams: Methodology and Data. *Transportation Research Part C: Emerging Technologies*, Vol. 99, 2019, pp. 82–109. https://doi.org/10.1016/j.trc.2018.12.012.
9. Liu, J., A. Weinert, and S. Amin. Semantic Analysis of Traffic Camera Data: Topic Signal Extraction and Anomalous Event Detection. *arXiv:1905.07332 [cs]*, 2019.
10. Kanhere, N. K., S. T. Birchfield, W. A. Sarasua, and S. Khoeini. Traffic Monitoring of Motorcycles during Special Events Using Video Detection. *Transportation Research Record*, Vol. 2160, No. 1, 2010, pp. 69–76. https://doi.org/10.3141/2160-08.
11. Tiezzi, M., S. Melacci, M. Maggini, and A. Frosini. Video Surveillance of Highway Traffic Events by Deep Learning Architectures. Cham, 2018.
12. Wu, J., H. Xu, Y. Zhang, and R. Sun. An Improved Vehicle-Pedestrian near-Crash Identification Method with a Roadside LiDAR Sensor. *Journal of Safety Research*, 2020. https://doi.org/10.1016/j.jsr.2020.03.006.
13. Hankey, J., J. A. McClaffety, and M. A. Perez. *Description of the SHRP 2 Naturalistic Database and the Crash, Near-Crash, and Baseline Data Sets*. 2016.
14. Doerzaph, Z. R. *Development of a Threat Assessment Algorithm for Intersection Collision Avoidance Systems*. Virginia Polytechnic Institute and State University, 2007.
15. Doerzaph, Z. R., V. L. Neale, J. R. Bowman, D. C. Viita, and M. A. Maile. *Cooperative Intersection Collision Avoidance System Limited to Stop Sign and Traffic Signal Violations Subtask 3.2 Interim Report*. Publication FHWA-JPO-10-068. 2008.
16. Xing, L., J. He, M. Abdel-Aty, Y. Wu, and J. Yuan. Time-Varying Analysis of Traffic Conflicts at the Upstream Approach of Toll Plaza. *Accident Analysis & Prevention*, Vol. 141, 2020, p. 105539. https://doi.org/10.1016/j.aap.2020.105539.
17. Ismail, K., T. Sayed, N. Saunier, and C. Lim. Automated Analysis of Pedestrian–Vehicle Conflicts Using Video Data. *Transportation Research Record*, Vol. 2140, No. 1, 2009, pp. 44–54. https://doi.org/10.3141/2140-05.
18. Lucas, B. D., and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. 1981.
19. Tomasi, C., and T. Kanade. *Detection and Tracking of Point Features*. Publication CMU-CS-91-132. 1991, p. 22.
20. Transportation Research Board, E. National Academies of Sciences, and Medicine. *Naturalistic Driving Study: Technical Coordination and Quality Control*. The National Academies Press, Washington, DC, 2014.
21. Teoh, E. R. Effectiveness of Front Crash Prevention Systems in Reducing Large Truck Real-World Crash Rates. *Traffic Injury Prevention*, Vol. 22, No. 4, 2021, pp. 284–289. https://doi.org/10.1080/15389588.2021.1893700.
22. Soccolich, S. A., and J. S. Hickman. Potential Reduction in Large Truck and Bus Traffic Fatalities and Injuries Using Lytx's DriveCam Program. 2014.

23. Virginia Department of Transportation. Virginia 511 Web. https://www.511virginia.org/. Accessed Jul. 25, 2019.
24. Bellard, F. and FFmpeg Authors. FFmpeg. , 2020.
25. Documentation. https://ffmpeg.org/documentation.html. Accessed Apr. 14, 2020.
26. IEEE/ISO/IEC International Standard - Information Technology Portable Operating System Interface (POSIX(TM)) Base Specifications, Issue 7. *ISO/IEC/IEEE 9945:2009(E)*, 2009, pp. 1–3880. https://doi.org/10.1109/IEEESTD.2009.5393893.
27. Nick Craig-Wood. Rclone. *Rclone syncs your files to cloud storage*. https://rclone.org/authors/. Accessed Dec. 15, 2022.
28. Butler, H., M. Daly, A. Doyle, S. Gillies, S. Hagen, and T. Schaub. *The GeoJSON Format*. Publication RFC7946. 2016.
29. Virginia GIS Clearinghouse. https://vgin.vdem.virginia.gov/pages/clearinghouse. Accessed Jan. 17, 2023.
30. Traffic Data - Info | Virginia Department of Transportation. https://www.virginiadot.org/info/ct-TrafficCounts.asp. Accessed Jan. 17, 2023.
31. Wu, H., C. Wen, W. Li, X. Li, R. Yang, and C. Wang. Transformation-Equivariant 3D Object Detection for Autonomous Driving. http://arxiv.org/abs/2211.11962. Accessed Jan. 18, 2023.
32. Wu, X., L. Peng, H. Yang, L. Xie, C. Huang, C. Deng, H. Liu, and D. Cai. Sparse Fuse Dense: Towards High Quality 3D Detection with Depth Completion. Presented at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022.
33. Yang, H., Z. Liu, X. Wu, W. Wang, W. Qian, X. He, and D. Cai. Graph R-CNN: Towards Accurate 3D Object Detection with Semantic-Decorated Local Graph. http://arxiv.org/abs/2208.03624. Accessed Jan. 18, 2023.
34. Wu, Y., A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick. Detectron2. , 2019.
35. Ren, S., K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv:1506.01497 [cs]*, 2015.
36. Kreiss, S., L. Bertoni, and A. Alahi. PifPaf: Composite Fields for Human Pose Estimation. 2019.
37. Lin, T.-Y., M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár. Microsoft COCO: Common Objects in Context. *arXiv:1405.0312 [cs]*, 2014.
38. *Geographic Information Network Division Established; Powers and Duties; Division Coordinator*. Code of Virginia, Virginia Code § 44-146.18:6, 2020.
39. Orthoimagery. https://vgin.vdem.virginia.gov/pages/orthoimagery. Accessed Aug. 17, 2022.
40. U.S. Geological Survey. 3D Elevation Program. https://www.usgs.gov/3d-elevation-program. Accessed Aug. 19, 2022.
41. Rublee, E., V. Rabaud, K. Konolige, and G. Bradski. ORB: An Efficient Alternative to SIFT or SURF. 2011.
42. Lowe, D. G. Object Recognition from Local Scale-Invariant Features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, No. 2, 1999, pp. 1150–1157 vol.2.

43. Leutenegger, S., M. Chli, and R. Y. Siegwart. BRISK: Binary Robust Invariant Scalable Keypoints. Presented at the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 2011.
44. Liu, D. C., and J. Nocedal. On the Limited Memory BFGS Method for Large Scale Optimization. *Mathematical Programming*, Vol. 45, No. 1–3, 1989, pp. 503–528. https://doi.org/10.1007/BF01589116.
45. Jakob, W. Dr.Jit: A Just-In-Time Compiler for Differentiable Rendering | RGL. https://rgl.epfl.ch/publications/Jakob2022DrJit. Accessed Jan. 16, 2023.
46. Jakob, W., S. Speierer, N. Roussel, M. Nimier-David, D. Vicini, T. Zeltner, B. Nicolet, M. Crespo, V. Leroy, and Z. Zhang. Mitsuba 3 Renderer. , 2022.
47. Uhlmann, J. *Dynamic Map Building and Localization : New Theoretical Foundations.* Ph.D. University of Oxford, 1995.
48. Munkres, J. Algorithms for the Assignment and Transportation Problems. *Journal of the Society for Industrial and Applied Mathematics*, Vol. 5, No. 1, 1957, pp. 32–38. https://doi.org/10.1137/0105003.
49. How to Draw Ellipse of Covariance Matrix. https://cookierobotics.com/007/. Accessed Jan. 18, 2023.
50. Federal Aviation Administration. *Global Positioning System Standard Positioning Service Performance Analysis Report*. Publication 112. 2021.
51. Federal Aviation Administration. *WAAS Performance Analysis Report*. Publication 83. 2023.
52. Lidar Point Cloud - USGS National Map 3DEP Downloadable Data Collection | USGS Science Data Catalog. https://data.usgs.gov/datacatalog/data/USGS:b7e353d2-325f-4fc6-8d95-01254705638a. Accessed Jan. 18, 2023.
53. Ruth, R. State of EDR in the US CDR Update. Jun 14, 2022.
54. NHTSA. *Crash Report Sampling System: 2020 CRSS Databook*. Publication DOT HS 813 253. 2022.
55. M. Davis and Company, Inc. *National Telephone Survey of Reported and Unreported Motor Vehicle Crashes*. Publication DOT HS 812 183. 2015.
56. Blincoe, L., T. R. Miller, J.-S. Wang, D. Swedler, T. Coughlin, B. Lawrence, F. Guo, S. Klauer, and T. Dingus. *The Economic and Societal Impact of Motor Vehicle Crashes, 2019*. Publication DOT HS 813 403. 2022.
57. NHTSA. National Automotive Sampling System - Crashworthiness Data System: 2014 Coding and Editing Manual. 2015, p. 1337.
58. Crash Report Sampling System 2019 CRSS Databook. 2020, p. 564.
59. NHTSA. *2020 Data: State Traffic Data Fact Sheet*. Publication DOT HS 813 368. 2022.
60. Sun, P., H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, S. Zhao, S. Cheng, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. http://arxiv.org/abs/1912.04838. Accessed Jan. 20, 2023.
61. Immanuel Scholz and Dirk Stöcker. Java OpenStreetMap Editor (JOSM). .
62. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher,

M. Perrot, and E. Duchesnay. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research*, Vol. 12, 2011, pp. 2825–2830.
63. Noble, A. M., K. D. Kusano, J. M. Scanlon, Z. R. Doerzaph, and H. C. Gabler. Driver Approach and Traversal Trajectories for Signalized Intersections Using Naturalistic Data. Presented at the Transportation Research Board 95th Annual MeetingTransportation Research Board, 2016.
64. Scanlon, J. *Evaluating the Potential of an Intersection Driver Assistance System to Prevent U.S. Intersection Crashes*. Virginia Polytechnic Institute and State University, Blacksburg, VA, 2017.
65. Bosina, E., and U. Weidmann. Estimating Pedestrian Speed Using Aggregated Literature Data. *Physica A: Statistical Mechanics and its Applications*, Vol. 468, 2017, pp. 1–29. https://doi.org/10.1016/j.physa.2016.09.044.
66. Kerridge, J., S. Keller, T. Chamberlain, and N. Sumpter. Collecting Pedestrian Trajectory Data In Real-Time. In *Pedestrian and Evacuation Dynamics 2005* (N. Waldau, P. Gattermann, H. Knoflacher, and M. Schreckenberg, eds.), Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 27–39.
67. Knoblauch, R. L., M. T. Pietrucha, and M. Nitzburg. Field Studies of Pedestrian Walking Speed and Start-Up Time: *Transportation Research Record*, 1996. https://doi.org/10.1177/0361198196153800104.
68. Haus, S. H. Effectiveness of Automatic Emergency Braking for Protection of Pedestrians and Bicyclists in the U.S. 2021, p. 214.
69. Bhagavathula, R., R. B. Gibbons, and C. J. Edwards. Effect of Static and Moving Objects on Driver Eye Movements and Detection Distances. 2012.
70. Real Time Risk Prediction at Signalized Intersection Using Graph Neural Network | Safe-D: Safety through Disruption. .
71. Hatfield, J., and S. Murphy. The Effects of Mobile Phone Use on Pedestrian Crossing Behaviour at Signalised and Unsignalised Intersections. *Accident Analysis & Prevention*, Vol. 39, No. 1, 2007, pp. 197–205. https://doi.org/10.1016/j.aap.2006.07.001.
72. Schroeder, B., N. Rouphail, K. Salamati, E. Hunter, B. Phillips, L. Elefteriadou, T. Chase, Y. Zheng, V. P. Sisiopiku, and S. Mamidipalli. *Empirically-Based Performance Assessment & Simulation of Pedestrian Behavior at Unsignalized Crossings.* Southeastern Transportation Research, Innovation, Development and Education …, 2014.
73. Song, W. *Image-Based Roadway Assessment Using Convolutional Neural Networks*. University of Kentucky, 2019.
74. Song, W., S. Workman, A. Hadzic, X. Zhang, E. Green, M. Chen, R. Souleyrette, and N. Jacobs. FARSA: Fully Automated Roadway Safety Assessment. *arXiv:1901.06013 [cs]*, 2019.
75. Li, G., W. Lai, X. Sui, X. Li, X. Qu, T. Zhang, and Y. Li. Influence of Traffic Congestion on Driver Behavior in Post-Congestion Driving. *Accident Analysis & Prevention*, Vol. 141, 2020, p. 105508. https://doi.org/10.1016/j.aap.2020.105508.
76. Thapa, R., S. Hallmark, O. Smadi, and A. Goswamy. Assessing Driving Behavior Upstream of Work Zones by Detecting Response Points in Speed Profile: A Naturalistic Driving Study.

Traffic Injury Prevention, Vol. 0, No. 0, 2019, pp. 1–6. https://doi.org/10.1080/15389588.2019.1663348.
77. Hutchinson, J. W., and T. W. Kennedy. *Medians of Divided Highways-Frequency and Nature of Vehicle Encroachments*. University of Illinois at Urbana Champaign, College of Engineering …, 1966.
78. Cooper, P. Analysis of Roadside Encroachments - Single Vehicle Run-off Accident Data Analysis for Five Provinces. 1980.
79. Bareiss, M., D. J. Gabauer, L. E. Riexinger, and H. C. Gabler. Properties of Encroachments and the Associated Roadway and Roadside Environment in the Second Strategic Highway Research Program. *Transportation Research Record*, 2022, p. 03611981221141896. https://doi.org/10.1177/03611981221141896.