In this project, we have designed and developed an effective end-to-end system based on advanced Artificial Intelligence (AI), machine learning, and computer vision to automatically monitor, detect, track, and count pedestrians and bicyclists. The main objective of this project is to improve the safety of pedestrians and bicyclists, by applying self-sensed and AI-powered systems to monitor and control the flow of pedestrians/bicyclists. The developed system includes algorithms for detecting the pedestrians and bicyclists, as well as algorithms for tracking and counting the pedestrians. We evaluated the developed system on real videos captured by actual traffic cameras in the city of Los Angeles. Despite the low quality of some of the videos, the results demonstrated high accuracy and effectiveness of the developed system in automatically detecting and counting pedestrians and bicyclists.
DISCLAIMER STATEMENT

This document is disseminated in the interest of information exchange. The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California or the Federal Highway Administration. This publication does not constitute a standard, specification or regulation. This report does not constitute an endorsement by the Department of any product described herein.

For individuals with sensory disabilities, this document is available in alternate formats. For information, call (916) 654-8899, TTY 711, or write to California Department of Transportation, Division of Research, Innovation and System Information, MS-83, P.O. Box 942873, Sacramento, CA 94273-0001.
Automatic Traffic Monitoring and Management for Pedestrian and Cyclist Safety Using Deep Learning and Artificial Intelligence

Mohammad Pourhomayoun, PhD
MINETA TRANSPORTATION INSTITUTE
LEAD UNIVERSITY OF
Mineta Consortium for Transportation Mobility

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation’s transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the four-university Mineta Consortium for Transportation Mobility, a Tier 1 University Transportation Center funded by the U.S. Department of Transportation’s Office of the Assistant Secretary for Research and Technology (OST-R), the California Department of Transportation (Caltrans), and by private grants and donations.

MTI’s transportation policy work is centered on three primary responsibilities:

**Research**
MTI works to provide policy-oriented research for all levels of government and the private sector to foster the development of optimum surface transportation systems. Research areas include: bicycle and pedestrian issues; financing public and private sector transportation improvements; intermodal connectivity and integration; safety and security of transportation systems; sustainability of transportation systems; transportation/land use/environment; and transportation planning and policy development. Certified Research Associates conduct the research. Certification requires an advanced degree, generally a Ph.D., a record of academic publications, and professional references. Research projects culminate in a peer-reviewed publication, available on TransWeb, the MTI website (http://transweb.sjsu.edu).

**Education**
The Institute supports education programs for students seeking a career in the development and operation of surface transportation systems. MTI, through San José State University, offers an AACSB-accredited Master of Science in Transportation Management and graduate certificates in Transportation Management, Transportation Security, and High-Speed Rail Management that serve to prepare the nation’s transportation managers for the 21st century. With the active assistance of the California Department of Transportation (Caltrans), MTI delivers its classes over a state-of-the-art videoconference network throughout the state of California and via webcasting beyond, allowing working transportation professionals to pursue an advanced degree regardless of their location. To meet the needs of employers seeking a diverse workforce, MTI’s education program promotes enrollment to under-represented groups.

**Information and Technology Transfer**
MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and journals and works to integrate the research findings into the graduate education program. MTI’s extensive collection of transportation-related publications is integrated into San José State University’s world-class Martin Luther King, Jr. Library.

**Disclaimer**
The contents of this report reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation’s University Transportation Centers Program. This report does not necessarily reflect the official views or policies of the U.S. government, State of California, or the Mineta Transportation Institute, who assume no liability for the contents or use thereof. This report does not constitute a standard specification, design standard, or regulation.
AUTOMATIC TRAFFIC MONITORING AND MANAGEMENT FOR PEDESTRIAN AND CYCLIST SAFETY USING DEEP LEARNING AND ARTIFICIAL INTELLIGENCE

Mohammad Pourhomayoun, PhD

September 2020
ACKNOWLEDGMENTS

The authors would like to thank the California Department of Transportation, and the Mineta Transportation Institute, San Jose State University, for their support of this project. We especially appreciate the tremendous support, guidance, and collaboration of Dr. Hilary Nixon, the Deputy Executive Director of Mineta Transportation Institute. We also deeply appreciate the work of Jerry Kwong of the Division of Research and Innovation in Caltrans for his great help and support.

The author thanks Editing Press, for editorial services, as well as MTI staff, including Executive Director Karen Philbrick, PhD; Graphic Designer Alverina Eka Weinardy; and Executive Administrative Assistant Jill Carter.
# TABLE OF CONTENTS

I. Introduction ................................................. 1  
II. System Architecture ........................................ 4  
III. Data Preprocessing ......................................... 5  
IV. Machine Learning and Deep Learning for Object Detection ........................................ 8  
V. Trajectory Prediction for Traffic Tracking and Counting ........................................ 11  
VI. Results and Conclusion ..................................... 13  

Bibliography .................................................. 16  
About the Author .............................................. 18  
Peer Review .................................................... 19  

Mineta Transportation Institute
# LIST OF FIGURES

1. Examples of Extreme Visibility Conditions in Traffic Video Images 2

2. End-to-End System Architecture 4

3. Background Subtraction: (a) Original Video Frame; (b) Estimated Background; and (c) Moving Objects After Background Subtraction 6

4. Background Change in Only 13 Minutes 7

5. (a) General Structure of a ConvNet and (b) AlexNet 9

6. Pedestrian Detection Using Machine Learning Algorithms: (a) Using Hog Features and SVM Classifier, and (b) Using YOLO. 10

7. Location Prediction and Trajectory Estimation 12

8. System Results on Real-Time Traffic Video Streams: (a) Bicyclist Tracking and Counting, (b) Pedestrian Tracking and Counting 14
LIST OF TABLES

1. Pedestrian Counting Results for Actual Videos Captured from a Traffic Camera over 12 Hours (A View of the Camera is Shown in Figure 8-B) 15
I. INTRODUCTION

According to the US Department of Transportation and the Insurance Institute for Highway Safety (IIHS) Highway Loss Data Institute (HLDI), the number of traffic fatalities in the state of California was 3,623 in 2016, which is more than 9.2 deaths per 100,000 population (USDOT 2016). The city of Los Angeles alone has one of the highest rates of traffic death among large U.S. cities. Currently, the city of Los Angeles has a strong understanding of vehicular travel at key intersections and corridors. Caltrans and local departments of transportation can optimize their traffic signal system to improve vehicular travel times using the Automated Traffic Surveillance and Control (ATSAC) System for vehicles and Regional Integration of Intelligent Transportation Systems (RIITS). However, understanding the movement of people, bicycles, and their interaction with vehicles is critical to avoiding traffic accidents and improving safety. Currently, there is no efficient automated system for monitoring the movement of pedestrians and bicyclists across the state of California and in major urban areas. Such a system could also provide valuable information about the traffic flow once implemented and calibrated.

With the advancement of technology, automated traffic monitoring has been gaining attention over the past decades. In particular, several methods have been proposed for pedestrian detection in the past couple of years (Dalal and Triggs 2005, Dollar et al. 2011, Benenson et al. 2014). These methods have used different techniques including image and video processing algorithms, as well as machine learning techniques to detect human targets (pedestrian) through images or video. Most of these contributions have used standard datasets including images or videos to evaluate the performance of the algorithm (Benenson et al. 2014). Two of the well-known pedestrian datasets include Daimler pedestrian benchmark dataset (Enweiler 2009), and Penn-Fudan database for pedestrian detection (Wang et al. 2007).

Dollar et al. (Dollar et al. 2011) and Beneson et al. (Benenson et al. 2014) performed extensive evaluations of the state of practice. They put together the most popular and well-annotated pedestrian detection datasets, and evaluated the performance of the most promising pedestrian detectors across several datasets. They have shown that despite significant progress in the past few years, there is much room for improvement. In particularly, the pedestrian detection results are disappointing at low resolution images and/or videos as well as for occluded pedestrians in images (Dollar et al. 2011).

Implementing pedestrian/bicyclist detection algorithms in practice, on real-life video streams captured by regular traffic cameras, can be different from ideal settings. There are challenges and difficulties associated with the process when applying the algorithm to a populated urban area. Some of these challenges include:

- poor quality of video streams due to low resolution, light conditions, disturbed and disoriented lens, or weather conditions;
- dealing with stretched, convex, or squeezed images collected by existing traffic cameras and wide-angle lenses;
- undesired angles, locations, and the direction of cameras;
• vibration of cameras due to wind or the passing of heavy vehicles;

• light distortion at night time;

• inconsistent lighting during the daytime and shadow effects; and

• moving or stationary objects that may block the view of the target.

Figure 1 illustrates some examples of these difficult situations from traffic cameras. In Figure-1a, the pedestrian is partially visible and the view is blocked by the wall. Figure-1b shows an example of poor and inconsistent lighting condition in which it is even difficult to detect the pedestrian with the human eye. Furthermore, the image is stretched and convex, as the gutter alignment is completely distorted. Figure-1c shows a condition where the pedestrian is hardly visible because of shadowing and low lighting.

Figure 1. Examples of Extreme Visibility Conditions in Traffic Video Images

In this project, the AI and Data Science Research Lab at California State University Los Angeles has designed and developed an effective system to monitor, track, and count pedestrians and bicyclists based on computer vision and machine learning algorithms. The developed system includes algorithms for detecting the pedestrians and bicyclists, as well
as algorithms for tracking and counting the pedestrians. It is important to notice that in this project, we are not planning to install new cameras or add new components or sensors to the traffic operation infrastructure. The system can use the videos captured by existing traffic cameras operated by Caltrans.

The rest of this report is organized as follows: Chapter 2 discusses the system architecture and the pipeline for detection, tracking, and counting. Chapter 3 discusses the preprocessing algorithms that we have developed to improve the quality of the input data, and consequently improve the outcome accuracy. Chapter 4 discusses the machine learning and deep learning algorithms that we have developed for pedestrian detection. Chapter 5 presents the developed algorithms for trajectory prediction, tracking and counting algorithms, and finally, Chapter 6 discusses the results and conclusion.
II. SYSTEM ARCHITECTURE

In this project, we developed an end-to-end system including a series of image/video processing algorithms, computer vision algorithms, Machine Learning algorithms, and optimal state estimator algorithms that receive videos and monitor, recognize, track, and count pedestrians and cyclists in the video.

Figure 2 shows the high-level system architecture.

![Figure 2. End-to-End System Architecture](image)

The next chapters describe the 3 main parts of the system: (1) raw video processing; (2) feature engineering and Machine Learning for object detection; and (3) trajectory prediction for traffic tracking and counting.
III. DATA PREPROCESSING

The first step in our end-to-end traffic vision system is the raw video preprocessing, which includes a series of algorithms for quality enhancement, and brightness and contrast adjustment.

The next step in background estimation and subtraction (we can also call it foreground detection or moving-object detection). In this framework, any moving object is considered as foreground, and any stationary object in a period of time (i.e., an object with fixed location in a number of sequential frames) is considered as background. We have tried several effective algorithms for background estimation/subtraction including frame differencing, mean filter, running Gaussian average, and mixture of Gaussian modeling (MOG) (Wang et al. 2018, Piccardi 2004, Bouwmans, El Baf, and Vachon 2008). It turned out that a mixture of Gaussian modeling (MOG) and mean filtering achieved the best results for background subtraction. Figure 3 shows the results of background subtraction (i.e., moving object detection) based on mean filtering.

It is important to notice that the background continuously changes during the daytime because the sunlight direction and intensity changes. Figure 4 shows the estimated background of a video captured by a traffic camera at 7:06AM and another time at 7:19AM. As we see, the background has significantly changed in only 13 minutes. Thus, we need to continuously estimate and update the background in order to always have the best background subtraction performance.

We have to notice that background removal can improve the performance and accuracy of the object recognition algorithm (i.e., the next step, which is the machine learning algorithm), and can also reduce the computational load of the object recognition algorithm by reducing the size of the area of interest. This can be even more important when we want to use computationally expensive machine learning algorithms, such as the Convolutional Neural Networks (ConvNet also known as CNN) (Krizhevsky, Sutskever, and Hinton 2012).
Figure 3. Background Subtraction: (a) Original Video Frame; (b) Estimated Background; and (c) Moving Objects After Background Subtraction
Figure 4. Background Change in Only 13 Minutes
IV. MACHINE LEARNING AND DEEP LEARNING FOR OBJECT DETECTION

After data preprocessing, the next step is to extract and select the best set of computer vision features that can be used in machine learning algorithms for object detection. Depending on the type of machine learning algorithm, this step may include feature extraction, feature selection, and/or dimensionality reduction. We have tried many different types of features and machine learning algorithms for object recognition, which we will briefly discuss below, and more details are available in (Wang et al. 2018, Pourhomayoun et al. 2019).

One of the most effective and popular features are the Histogram of Oriented Gradient (HOG) (Dalal and Triggs 2005), which along with SVM classifier can form an effective method for pedestrian recognition (Dalal and Triggs 2005). HOG is a feature descriptor that counts occurrences of gradient orientation in localized portions of an image (Dalal and Triggs 2005). It has been proven to be one of the most effective hand-made features that can be used for object recognition.

We have also tried deep learning methods, particularly the Convolutional Neural Networks (ConvNet), including the R-CNN (Region-based Convolutional Network) and YOLO (You Only Look Once) algorithms (Ren et al. 2015, Redmon et al. 2016, Zarchan and Musoff 2000). A big advantage of ConvNet methods compared to other classic machine learning algorithms is that there is no need to generate and use hand-made features for ConvNet. The algorithm automatically learns to generate the best set of convolutional features that can best represent the image. However, ConvNet is computationally expensive and sometimes makes it more difficult to run in real-time on high-frame-rate videos. In addition, when the training dataset is not large enough, it is usually difficult to train a deep neural network. In this case, Transfer Learning methods that take advantage of a pre-trained neural network model on another dataset can be very helpful to ease and expedite the training process (Wang et al. 2018, Pourhomayoun et al. 2019).

Figure 5-(a) shows the high-level general structure of a ConvNet, and Figure 5-(b) shows a specific design of ConvNet named AlexNet (Krizhevsky, Sutskever, and Hinton 2012).

Figure 6-(a) shows our pedestrian detection results using HOG features and SVM classifier. Figure 6-(b) shows our results using YOLO algorithm.
Figure 5. (a) General Structure of a ConvNet and (b) AlexNet
Figure 6. Pedestrian Detection Using Machine Learning Algorithms: (a) Using Hog Features and SVM Classifier, and (b) Using YOLO.
V. TRAJECTORY PREDICTION FOR TRAFFIC TRACKING AND COUNTING

After detecting a target object (e.g., a pedestrian or bicyclist) in several sequential frames, we use *Optimal State Estimator* to estimate the *Trajectory* of each target object. Since several objects may exist in each frame at a time (e.g., several pedestrians walking together in the same direction or different directions), it is essential to estimate the trajectory of each object individually.

To this end, we use the Kalman Filter (Zarchan and Musoff 2000) as an optimal state estimator to predict the next location of the object and estimate the trajectory of the object over time. This allows us to track each object individually during the video. For example, suppose that we want to track a pedestrian. We use the Kalman filter to predict the next location of the pedestrian in the next frame based on its previous locations and walking pace. Then, after receiving the next frame, we compare our prediction with the actual pedestrian detected in the next frame. This comparison tells us if this pedestrian was the same person as in the previous frame or, instead, a different person. If the predicted location and actual location match, we consider this pedestrian as being identical to the previous one, and then continue completing the trajectory of this pedestrian (see Figure 7).

Using this approach, we can build a trajectory map including individual trajectories for all pedestrians in the video, and then track each pedestrian from the first frame they enter until the last frame when they move out.

Every time we detect a pedestrian whose location does not match any of the previously predicted locations (i.e., the pedestrian is not located on any of the existing estimated trajectories), we consider that person as a new pedestrian and, consequently, add to the pedestrian counter. This will allow us to track and count each pedestrian everywhere in the video, and avoid double counting them in sequential frames.
Figure 7. Location Prediction and Trajectory Estimation
VI. RESULTS AND CONCLUSION

The city of Los Angeles has one of the highest rates of traffic deaths among large U.S. cities. Fortunately, the city has launched the Vision Zero initiative as a strategy and commitment to reduce traffic fatalities. The main goal of LA’s Vision Zero is to eliminate all traffic fatalities by 2025. Since the most vulnerable components of traffic accidents are pedestrians and bicyclists, it is essential to develop intelligent transportation systems, and human-centered traffic approaches to protect pedestrians and bicyclists. This project introduced an effective end-to-end system based on computer vision and machine learning to detect, monitor, track, and count pedestrians and bicyclists. This approach particularly enables us to recognize and monitor busy intersections that are prone to traffic accidents, and allows us to control and manage traffic in those intersections to protect our pedestrians and bicyclists.

Despite many practical challenges (as mentioned in Chapter 1), the developed system works very well with the existing regular traffic videos. This system may help increase safety and traffic flow through better traffic management and planning.

We evaluated our developed system on 12 hours of real videos captured from actual traffic cameras. Figure 8 shows some of the results for pedestrian and bicyclist detection, tracking, and counting. Table 1 shows the pedestrian counting results on the video streams captured from an actual traffic camera for 12 hours (a view of the camera is shown in Figure 8-b). The first column of Table 1 shows the hour number; the second column shows the number of pedestrians counted automatically by the developed system; the third column shows the actual number of pedestrians counted by a human expert as the ground truth; and the last column is the hourly percent error.

We used the following equation to calculate the Percent Error:

\[ \text{Percent Error} = \frac{|A - B|}{B} \times 10 \]

where \( A \) is the number of pedestrians counted automatically by the developed system, and \( B \) is the correct number of pedestrians counted by a human expert.
Table 1 presents the pedestrian count results from a traffic camera over 12 hours. The first column shows the hour number; the second column shows the number of pedestrians counted by the developed system; the third column shows the actual number of pedestrians counted by a human expert; and the last column is the percent hourly error.
### Table 1. Pedestrian Counting Results for Actual Videos Captured from a Traffic Camera over 12 Hours (A View of the Camera is Shown in Figure 8-B)

<table>
<thead>
<tr>
<th>Hour No</th>
<th>Automated Counted by Developed System</th>
<th>Ground Truth Counted by Human</th>
<th>Hourly Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>89</td>
<td>86</td>
<td>% 3.5</td>
</tr>
<tr>
<td>2</td>
<td>94</td>
<td>90</td>
<td>% 4.4</td>
</tr>
<tr>
<td>3</td>
<td>101</td>
<td>107</td>
<td>% 5.6</td>
</tr>
<tr>
<td>4</td>
<td>148</td>
<td>139</td>
<td>% 6.5</td>
</tr>
<tr>
<td>5</td>
<td>120</td>
<td>110</td>
<td>% 9.1</td>
</tr>
<tr>
<td>6</td>
<td>153</td>
<td>160</td>
<td>% 4.4</td>
</tr>
<tr>
<td>7</td>
<td>217</td>
<td>210</td>
<td>% 3.3</td>
</tr>
<tr>
<td>8</td>
<td>242</td>
<td>234</td>
<td>% 3.4</td>
</tr>
<tr>
<td>9</td>
<td>222</td>
<td>229</td>
<td>% 3.1</td>
</tr>
<tr>
<td>10</td>
<td>260</td>
<td>261</td>
<td>% 0.4</td>
</tr>
<tr>
<td>11</td>
<td>331</td>
<td>324</td>
<td>% 2.2</td>
</tr>
<tr>
<td>12</td>
<td>291</td>
<td>280</td>
<td>% 3.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2268</strong></td>
<td><strong>2230</strong></td>
<td><strong>% 4.1</strong></td>
</tr>
</tbody>
</table>

Average of Hourly Errors = % 4.1
BIBLIOGRAPHY

Benenson, Rodrigo, Mohamed Omran, Jan Hosang, and Bernt Schiele. 2014. “Ten years of pedestrian detection, what have we learned?” European Conference on Computer Vision.


ABOUT THE AUTHOR

Dr. Mohammad Pourhomayoun is an Assistant Professor of Computer Science at California State University Los Angeles (CSULA). He is the founder and director of Artificial Intelligence & Data Science Research Lab at CSULA. Dr. Pourhomayoun’s research interests focus on Data Science, Artificial Intelligence (AI), and Machine Learning for social good.
PEER REVIEW

San José State University, of the California State University system, and the Mineta Transportation Institute (MTI) Board of Trustees have agreed upon a peer review process required for all research published by MTI. The purpose of the review process is to ensure that the results presented are based upon a professionally acceptable research protocol.
MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Founder, Honorable Norman Mineta*
Secretary (ret.), US Department of Transportation

Chair,
Abbas Mohaddes
President & COO, Econolite Group Inc.

Vice Chair,
Will Kempton
Executive Director, Sacramento Transportation Authority

Executive Director,
Karen Philbrick, PhD*
Mineta Transportation Institute, San José State University

Winsome Bowen
Chief Regional Transportation Strategy, Facebook

David Castagnetti
Co-Founder, Mehlan Castagnetti, Rosen & Thomas

Maria Cino
Vice President, America & U.S. Government Relations Hewlett-Packard Enterprise

Grace Crunican**
Owner, Crunican LLC

Donna DeMartino
Managing Director, Los Angeles-San Diego-San Luis Obispo Rail Corridor Agency

Nuria Fernandez**
General Manager & CEO, Santa Clara Valley Transportation Authority (VTA)

John Flaherty
Senior Fellow, Silicon Valley Leadership Form

William Flynn*
President & CEO, Amtrak

Rose Guilbault
Board Member, Peninsula Corridor Joint Powers Board

Ian Jefferies*
President & CEO, Association of American Railroads

Diane Woodend Jones
Principal & Chair of Board, Lea + Elliott, Inc.

David S. Kim*
Secretary, California State Transportation Agency (CALSTA)

Nuria McMillan
Executive Director, Metropolitan Transportation Commission (MTC)

Bradley Mims
President & CEO, Conference of Minority Transportation Officials (COMTO)

Jeff Morales
Managing Principal, InfraStrategies, LLC

Dan Moshavi, PhD*
Dean, Lucas College and Graduate School of Business, San José State University

Toks Omishakin*
Director, California Department of Transportation (Caltrans)

Takayoshi Oshima
Chairman & CEO, Allied Telesis, Inc.

Paul Skoutelas*
President & CEO, American Public Transportation Association (APTA)

Beverley Swaim-Staley
President, Union Station Redevelopment Corporation

Jim Tymon*
Executive Director, American Association of State Highway and Transportation Officials (AASHTO)

Larry Willis*
President, Transportation Trades Dept., AFL-CIO

* = Ex-Officio
** = Past Chair, Board of Trustees

Directors

Karen Philbrick, PhD
Executive Director

Hilary Nixon, PhD
Deputy Executive Director

Asha Weinstein Agrawal, PhD
Education Director, National Transportation Finance Center Director

Brian Michael Jenkins
National Transportation Security Center Director