Real-Time Big Data Analytics for Traffic Monitoring and Management for Pedestrian and Cyclist Safety

Mohammad Pourhomayoun, Haiyan Wang, Mohammad Vahedi, Mehran Mazari
Computer Science Department
California State University Los Angeles
Email: mpourho@calstatela.edu, hwang2@calstatela.edu, mvahedi@calstatela.edu, mmazari2@calstatela.edu

Hunter Owens
Data Science Federation
City of Los Angeles
Los Angeles, USA
Email: hunter.owens@lacity.org

Janna Smith
Department of Transportation
City of Los Angeles, Los Angeles, USA
Email: janna.smith@lacity.org

William Chernicoff
Toyota Mobility Foundation
Washington DC, USA
Email: william.chernicoff@toyota.com

Abstract— **In this study, we design and develop an end-to-end system based on data analytics and deep learning methods to monitor, count, and manage traffic, particularly, pedestrians and bicyclists in real-time. The main objective of this research is to improve the safety of pedestrians and bicyclists, by applying self-sensed and intelligent systems to control and monitor the flow of pedestrians/bicyclists particularly at intersections. This paper proposes an effective end-to-end system for traffic vision, detection, and counting on real-time traffic videos. The developed system is evaluated on 12 hours of real video streams captured from actual traffic cameras in the city of Los Angeles. According to the results, the developed system can count the pedestrians with less than 2% error.**

Keywords - Machine Learning; Deep Learning; Computer Vision; Object Detection.

I. INTRODUCTION

By 2050, 66% of the world's population is projected to be urban [1][2]. As urban populations rise, it is essential for city designers and planners to focus more on designing smart cities and addressing the main challenges, such as traffic issues, and the impacts of increased vehicle use. According to the U.S. Department of Transportation (USDOT), the number of traffic fatalities has increased by nearly 6% in 2016 [3]. The number of traffic fatalities only in the state of California was 3,623 in 2016, which is more than 9.2 deaths per 100,000 population.

Understanding the movement of people, bicycles, and their interaction with vehicles is critical to avoid traffic accidents and improve safety. We know that the most vulnerable components of the traffic collisions are pedestrians and bicyclists. Thus, it is essential to develop intelligent transportation systems, and human-centered traffic approaches to protect our pedestrians and cyclists and ensure that they can travel safely, efficiently, and comfortably.

With the advancement of technology, automated traffic monitoring has been gaining attraction over the past couple of years. In particular, several methods have been proposed for pedestrian detection in the past couple of years [4] – [6]. These methods have used different techniques including image/video processing, as well as machine learning techniques to detect human targets (pedestrian). Most of the previous contributions have used standard datasets including images/videos captured in ideal situations to evaluate the performance of the algorithm [5]. However, when we want to do it in practice, in real-time on video streams from traffic cameras in the scale of a large city like Los Angeles, it will be very different from lab settings, and we need to deal with challenges of Big Data Analytics.

Dollar et al. [4] and Beneson et al. [5] performed an extensive evaluation of the state of the practice. They put together the most popular pedestrian 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, the performance still has much room for improvement. Particularly, the pedestrian detection results are disappointing at low resolutions videos and for occluded pedestrians in the image [4].

The goal of this study is to design and develop an end-to-end system based on computer vision and machine learning to monitor, detect, track, count, and manage traffic, particularly, pedestrians and bicyclists. In this paper, we will evaluate our system on 12 hours of real video streams captured from actual traffic cameras in the city of Los Angeles. According to the results, the developed system can count the pedestrians with less than 2% error.

The rest of the paper is organized as follows: Section II describes the system architecture, methods, and the details of the proposed framework and components. Section III provides the evaluation results on actual data including 12 hours of real video streams captured from actual traffic cameras in the city of Los Angeles. Finally, Section IV includes the conclusion.

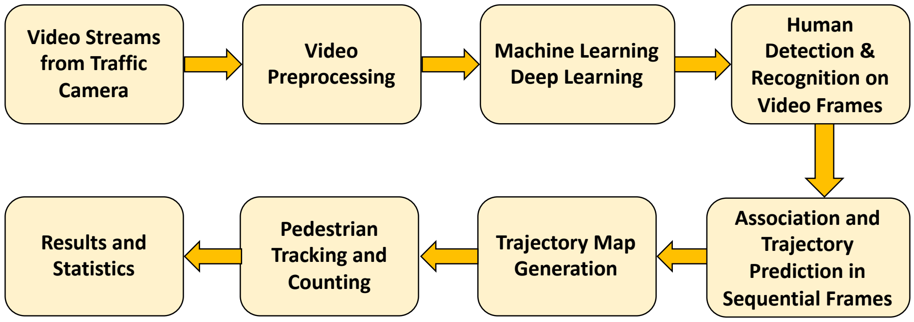


Figure 1. End-to-end system architecture.

II. SYSTEM ARCHITECTURE AND METHOD

In this study, we have developed an end-to-end system including a series of image/video processing, computer vision algorithms, Machine Learning and Deep Learning, and optimal state estimator algorithms that receive video streams in real-time, and detect, recognize, track, and count pedestrians and cyclists in the video.

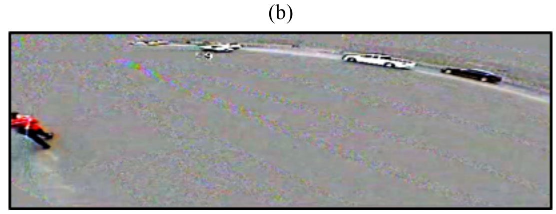
Figure 1 shows the high-level system architecture. The first step in the proposed traffic vision system is raw video preprocessing, which includes a series of algorithms for quality enhancement, and brightness/contrast adjustment. In the case of wide-angle lenses that may make the image convex, we can also use correction algorithms to convert the video back to natural view.

An important step in video preprocessing is background estimation and subtraction. In this concept, any moving object is considered as foreground, and any stationary object (i.e., an object with fixed location in a number of sequential frames) is considered as background. Although most of machine learning algorithms can still perform object recognition without a background removal step, but most of the time, it can improve the performance and accuracy of object recognition algorithm and also reduce the computational load of the object recognition algorithm by reducing the size of the area of interest.

In this study, we tried several effective algorithms for background estimation/subtraction including mean filter, frame differencing, running Gaussian average, and Mixture of Gaussian modeling (MOG) [6][7]. It turned out that MOG, and also mean filtering achieved the best results for background subtraction. Figure 2 shows the results of background subtraction (i.e., moving object detection) based on mean filtering. We have to note that the background continuously changes because the light direction and intensity changes. Thus, it is essential to continuously estimate and update the background to always have the best background subtraction performance.

(a)

(b)



(c)

Figure 2. Background subtraction: (a) Original video frame, (b) Estimated background, (c) Moving objects after background subtraction.

After video 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.

Before recent advancement in deep learning, Histogram of Oriented Gradient (HOG) has been one of the most popular hand-made features for object recognition [8]. HOG features along with Support Vector Machine (SVM) classifier can form an effective method for pedestrian recognition [8]. HOG is a feature descriptor that counts occurrences of gradient orientation in localized portions of an image [8].

In this study, we have also tried various deep learning methods, particularly the Convolutional Neural Networks (ConvNet), R-CNN (Region-based Convolutional Network), and YOLO (You Only Look Once) algorithms [9]-[12]. A big

 ISBN: 978-1-61208-700-9

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 difficult to run in real-time on high-frame-rate videos. In addition, when the training dataset is not large enough, it is usually hard to train an accurate deep neural network. In this case, Transfer Learning methods that take advantage of a pre-trained neural network model on other dataset can be very helpful to expedite the training stage [14].

Figure 3-(a) shows our pedestrian detection results on an actual traffic video using HOG features and SVM classifier. Figure 3-(b) shows our results using YOLO algorithm.

(a)

(b)

Figure 3. Pedestrian detection using machine learning algorithms. (a) using HOG features and SVM classifier, (b) using YOLO.

After detecting/recognizing the object of interest (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 same direction or different directions), it is essential to estimate the trajectory of each object individually.

We use Kalman Filter [13] as an optimal state estimator to predict the next location of the object and estimate the trajectory of the object over time. In this approach, in addition to the location of each bounding box, we extract and use a set of object features to represent each object uniquely. This allows us to recognize, distinguish, and track each object (i.e., each pedestrian or bicyclist) individually during the video, especially in difficult situations when several objects pass or overlap each other.

Suppose that we want to track a pedestrian. We use Kalman filter to predict the next location of each pedestrian in the next frame based on the previous locations and walking pace (extracted from previous frames). Then, after receiving the next frame, we compare our prediction with the new pedestrian detected in the next frame. The association is performed by comparing the bounding box location as well as other object features. This comparison tells us if this pedestrian was the same person in the previous frame, or it is a new one. If the predicted location and actual location match, we consider this pedestrian as previous one, and continue completing the trajectory of this pedestrian (see Figure 4). 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 he enters until the last frame when he moves out.

Figure 4. Location prediction and Trajectory estimation.

In this approach, when we detect a pedestrian whose location does not match to any of the previously predicted locations (it does not locate on any of the existing estimated trajectories), we consider that person as a new pedestrian and consequently, increment the pedestrian counter. This will allow us to track and count each pedestrian everywhere in the scene, and avoid double counting them in sequential frames.

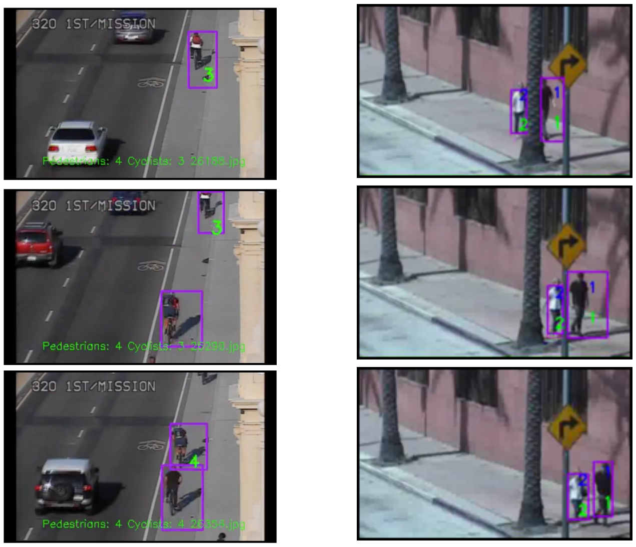
III. RESULTS ON ACTUAL DATA

We evaluated our developed system on 12 hours of real video streams captured from actual traffic cameras in the city of Los Angeles. Figure 5 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 in the city of Los Angeles for 12 hours (a view of the camera is shown in Figure 5-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. The last row in Table 1 shows the Overall Percent Error of 1.7% for counting over 12 hours. We used the following equation to calculate the Percent Error:

$$\text{Percent Error} = \frac{|\,A - B\,|}{B} * 100$$

where A is the number of pedestrian counted automatically by the developed system, and B is the correct number of pedestrians counted by a human expert.



(a) (b)

Figure 5. System results on real-time traffic video streams: (a) Bicyclist tracking and counting, (b) Pedestrian tracking and counting.

IV. CONCLUSION

This paper introduced an effective end-to-end system based on computer vision and machine learning to detect, recognize, monitor, track, and count pedestrians and bicyclists in real-time. 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.

The California State University Los Angeles in partnership with the Los Angeles Department of Transportation (LADOT), the City of Los Angeles, and Toyota Mobility Foundation has developed this effective and scalable system to detect, monitor, track, and count pedestrians and bicyclists in real-time. This system is potentially scalable to the 56,000 miles of streets in Los Angeles. Despite many practical challenges, the developed system works very well with the existing regular traffic cameras and therefore, there is no need to install any special or new cameras for this purpose.

ACKNOWLEDGMENT

The authors would like to thank Toyota Mobility Foundation for supporting this research. The authors would like to thank LADOT, City of LA, and ITA Data Science Federation for valuable help and support.

TABLE I. PEDESTRIAN COUNTING RESULTS FOR REAL VIDEO STREAMS CAPTURED FROM TRAFFIC CAMERAS IN LOS ANGELES FOR 12 HOURS (A VIEW OF THE CAMERA IS SHOWN IN FIGURE 5-B).

Hour No	Automated Counted by Developed System	Ground Truth Counted by Human	Hourly Error
1	89	86	3.5%
2	94	90	4.4%
3	101	107	5.6%
4	148	139	6.5%
5	120	110	9.1%
6	153	160	4.4%
7	217	210	3.3%
8	242	234	3.4%
9	222	229	3.1%
10	260	261	0.4%
11	331	324	2.2%
12	291	280	3.9%
Total	**2268**	**2230**	
Average of Hourly Errors = 4.1%			
Overall Percent Error in 12 hours = 1.7%			

REFERENCES

[1] World Urbanization Prospects, UN-Department of Economic and Social Affairs, 2018.
[2] Unicef, www.unicef.org/sowc2012/urbanmap, 2012.
[3] USDOT, https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data, 2016.
[4] P. Dollar, C. Wojek, B. Schiele, P. Perona, "Pedestrian Detection: An Evaluation of the State of the Art," in IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 34, p.p 743 – 761, 2012.
[5] R. Benenson, et al. “Ten Years of Pedestrian Detection, What Have We Learned?” ECCV, Springer, pp 613-627, 2015.
[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," IEEE CVPR'05, Jun 2005.
[7] M. Piccardi, “Background subtraction techniques: a review”, IEEE Int. Conf. on Systems, Man and Cybernetics, 2004.
[8] T. Bouwman, F. El Baf, B. Vachon, "Background Modeling using Mixture of Gaussians for Foreground Detection – A Survey". Recent Patents on Comp. Science, pp. 219-237, 2008.
[9] R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision (ICCV), 2015.
[10] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Analysis & Machine Intelligence, 2017.
[11] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You Only Look Once:Unified, Real-Time Object Detection," Computer Vision & Pattern Recog., 2016.
[12] A. Krizhevsky, I. Sutskever, G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS 2012.
[13] P. Zarchan and H. Musoff, “Fundamentals of Kalman Filtering: A Practical Approach”, ISBN 978-1-56347-455-2, 2000.
[14] H. Wang, et al., “An End-to-End Traffic Vision and Counting System Using Computer Vision and Machine Learning: The Challenges in Real-Time Processing”, SIGNAL2018, 2018.