2018

PASSENGER DETECTION THOUGH VIDEO PROCESSING AND SIGNAL SENSORS IN THE BOSTON SUBWAY TO ADDRESS LEFT-BEHIND PASSENGERS

Andronikos Keklikoglou
Umass Amherst

PASSENGER DETECTION THOUGH VIDEO PROCESSING AND SIGNAL SENSORS IN THE BOSTON SUBWAY TO ADDRESS LEFT-BEHIND PASSENGERS

A Project Presented

by

ANDRONIKOS KEKLIKOGLOU

Master of Science in Civil Engineering (or Environmental Engineering)

Department of Civil and Environmental Engineering
University of Massachusetts
Amherst, MA 01003

05/09/2018

Table of Contents

Table of Contents .... **ii**

Table of Figures .... **ii**

Table of Tables .... **iii**

Table of Acronyms .... **iii**

Abstract .... **iv**

1.0 Introduction **1**

2.0 Literature Review **1**

2.1 Crowding **1**
2.2 Technologies for Passenger Counting **3**
2.2.1 Digital Image Processing for Object Detection 4
2.2.2 You Only Look Once 5
2.2.3 Passenger Counting Using Bluetooth and Wi-Fi Detectors 5
2.2.4 Bluetooth and Wi-Fi sensors manufacturers 7

3.0 Methodology **8**

3.1 Crowding Analysis **8**
3.1.1 Passenger Flows 9
3.1.2 Identification of stations and times with overcrowding phenomena 11
3.2 Camera rating **14**
3.3 Threshold selection **14**

4.0 Results **15**

4.1 Passenger Detection Time Series **15**
4.2 Bluetooth and Wi-Fi signal detection **20**

5.0 Conclusions **22**

5.1 Future Work **24**

References **26**

Table of Figures

Figure 1: Passenger Load Colormap Oak Grove to Forest Hills 12
Figure 2: Passenger Load Colormap Forest Hills to Oak Grove 13
Figure 3:Testing Thresholds 14
Figure 4: Number of Passenger on the Platform over time Unsmoothed 15
Figure 5: Number of Passengers on Platform Smoothed 16
Figure 6: North Station 11/15/2017 Smoothed Counts 17
Figure 7: Sullivan Square 11/15/2017 Smoothed Counts 18
Figure 8: North Station 1/31/2018 Smoothed Counts 18
Figure 9: Sullivan Square 1/31/2018 Smoothed Counts 19

Figure 10: Comparison of Manual and YOLO counts in North Station 11/15/2017 19
Figure 11: Comparison of Manual and YOLO counts in North Station 1/31/2018 20
Figure 12: Comparison of Wireless and Manual Counts– North Station 1/31/2018 21
Figure 13: Comparison of Waiting Times – North Station 1/31/2018 22

Table of Tables

Table 1: Orange Line Headways ………………………………………… 10
Table 2: Carriage capacity in Time Periods ……………………………... 10
Table 3: Weekday Time Period Definitions ……………………………... 11
Table 4: Threshold errors from YOLO counts ……………………………... 15
Table 5: Bluetooth and Wi-Fi observations 1/31/2018 …………………….. 21

Table of Acronyms

Acronym	**Expansion**
CNN	Convolutional Neural Network
COCO	Common Object in Context
MAC	Media Access Control
MBTA	Massachusetts Bay Transportation Authority
R-CNN	Region Based Convolutional Neural Network
SVM	Support Vector Machine
TCQSM	Transit Capacity and Quality of Service Manual
YOLO	You Only Look Once

Abstract

Crowding is one the most common problems for public transportation systems worldwide. It has been proven to cause anxiety to commuters and create reliability problems when commuters are not able to board on the first train or bus that arrives. These commuters are referred as left-behind passengers, and their number is directly related to various basic performance measures of public transportation systems that represent the user's experience. Among these measures the most significant are ridership, service quality and, more importantly, travel time. Identifying left behind passengers is a tool to address crowding in stations and respond appropriately, by applying various operational strategies such as decreasing headways.

The methodology proposed in this study has been applied to two stations with high probability of left behind passengers, Sullivan Square and North Station on the MBTA Orange Line in Boston, Massachusetts. Two types of technologies were used to detect passengers being left behind in the platform. The first one was an object detection software, namely You Only Look Once (YOLO), using surveillance cameras. The second type was a Bluetooth and Wi-Fi sensor mounted on the two selected stations. Moreover, manual counts of left behind passengers were collected in the two stations. Both technologies will be individually compared with the manual counts to test accuracy and precision. Finally, the two technologies are compared with the manual counts to determine a best way to detect left behind passengers.

1.0 Introduction

Public transportation is a major part in a commuter's daily routine, especially in large cities. Transit ridership has increased in the past few years, as commuters prefer public transit from passenger cars due to the increased congestion within the urban area as well as the suburban areas. However, one of the most critical components of this choice is the reliability and cost-effectiveness of each mode choice. It is evident that public transportation is more cost-effective than a passenger car, but not as reliable and comfortable with some limited exceptions. Therefore, it is of vast importance for any public transit agency to improve those performance measures, increase ridership, relieve congestion and reduce financial pressure from passenger car users while increasing the agency's revenue. Crowding is a very common problem for public transportation systems and commuters might not be able to board on the first train or bus that arrives. These commuters are referred as left-behind passengers and their number is directly related to various basic performance measures of public transportation systems related to the user's experience, such as ridership measures, service quality and reliability. Ridership measures are focusing on the level of public transportation riders using the services (Grant M., 2011). Addressing the existence of left behind passengers will allow public transportation agencies to identify the locations and magnitude of this issue and act accordingly, in order to improve quality of service and reliability. Most of the studies that have been completed on left behind passenger detection, occurred in transit authorities that had a tap-out system such as the London Underground (Zhu Y., 2017), which gives exact information about travel times, arrivals and departures of passengers. In this project a non-tap-out system is examined and investigated using emerging technologies to address the left-behind passengers.

2.0 Literature Review

2.1 Crowding

Crowding is a major issue in public transit systems all over the world, due to the inconvenience and effects that it causes on operating speed, waiting time, travel time, reliability and route choice (Tirachini A., 2006). Studies on crowding has have shown an increase anxiety, stress and feeling of invasion of privacy (Lundberg, 1976).

In addition, crowding is directly related to a high density of passengers on vehicles, platforms and stations which furtherly establishes the above-mentioned disadvantages.

The Transit Capacity and Quality of Service Manual (TCQSM) (Kittelson & Associates Inc, 2013) determines some guidelines for measuring the quality of service as an effort to track passenger related metrics. It states that crowding affects several aspects of availability and all the elements of comfort and convenience related to the quality of service framework.

The indicators of availability as presented on the TCQSM are frequency, service span and access. The effects of crowding such as vehicles operating more slowly or passengers being left behind at stations and stops, cause an effective reduction in service frequency for users, which represents a limitation on the availability of transit service for users compared to the same uncrowded system. The hours of service in a day that a transit system operates are not typically affected by crowding. However, an exception could occur if crowding on the last run of the night prevents some users from being able to use the system during its hours of operation. Access also affects crowding indirectly in the sense of users being able to physically get to the system which in some cases may delay a passenger's ability to get to an otherwise accessible station (Kittelson & Associates Inc, 2013).

The indicators of comfort and convenience as defined by TCQSM are passenger load, reliability and travel time. Passenger load directly affects crowding and there is a demonstrated relationship between crowding and passengers' perception of time. For instance, more crowded vehicles increase the likelihood of passengers having to stand or squeeze with other passengers, which deteriorates the quality of the passenger experience. Evidence has shown that the users perceive longer waiting and travel times in crowded conditions than in uncrowded conditions (Fan Y., 2015).

Likewise, crowding contributes to diminished reliability with regard to on-time performance and maintaining consistent headways, because boarding and alighting are delayed when there are many people in vehicles and at stations (Carrion C., 2013). Another consequence for reliability is that left-behind passengers essentially experience one or more extra headways of waiting time if they are unable to board on the first vehicle that arrives due to crowding conditions.

Finally, like increasing the perceived travel time, the actual travel times are also increased by crowding due to delays on performance and headways. Therefore, vehicles operate more slowly, especially because of the above-mentioned delays associated with boarding and alighting. Besides, passengers that are left behind experience longer total travel times due to the additional time they must wait to board a vehicle.

It is also critical to state that crowding is included in several demand models as a parameter (Douglas N., 2005) and that it has been shown that waiting and in-vehicle travel time saving are inversely proportional to the number of people in the platform or vehicles. Therefore, an external parameter is produced called crowding externality or crowding cost (M., 1991). All the above mentioned prove the importance and influence of crowding on passengers and agencies decisions. The vast majority of the literature on crowding has focused on passenger discomfort. Limited research has been completed on actions that transit agencies can apply to relieve crowding. Moreover, crowding has demonstrated to affect the demand patterns on bus and rail systems (Tirachini A., 2006). Several studies have investigated the value of crowding from the perspective of the user, in terms of value of time and willingness to pay an extra fee to avoid crowding (Li Z., 2011) (Haywood L. K. M., 2015) (Haywood L. K. M., 2017) (Hörcher D., 2017). Furthermore, various studies aimed to determine the effect that crowding has on passengers' travel decisions and path choice. For instance, research that has been completed in Seoul, South Korea, suggests that crowding affects the path choice in networks that are large and connected enough to offer multiple path choices to users between origin-destination pairs (Kim K.M., 2015).

2.2 Technologies for Passenger Counting

There are a number of technologies that can be used to observe and count pedestrians and pedestrian movements in an area. The two main categories of technologies that are considered are the following:

- Image processing techniques through surveillance videos to directly observe and track pedestrians
- Device detection technologies that register a unique device identifier associated with Bluetooth and Wi-Fi signals, called media access control(MAC) address

Numerous technologies for simple pedestrian counting exist, mostly based on manual counts for a short period of time and applying different models for volume predictions (Schneider R. J., 2008). Those models are often misleading and cannot accurately address whether or not there are left behind passengers.

2.2.1 Digital Image Processing for Object Detection

The detection of objects in surveillance videos is an invaluable tool for passenger counting and has numerous applications. For example, object detection can be used for passenger counting or tracking, crowding recognition, hazardous object recognition and safety evaluation of autonomous technologies that use object detection to avoid conflicts. Computer vision is the duplicate of human vision aiming to electronically perceive, understand (Sonka M., 2014) and store information regarding one or multiple images. There are various techniques of using computer to process an image for detecting objects, by extracting useful information.

Recent methods for detecting objects use feature-based techniques, rather than segmentation of a moving foreground from a static background that was used in the past. Then, the detected features are extracted and subjected to a classification stage, typically using either boosted classifiers or Support Vector Machine (SVM) methods (Viola, 1993) (Cheng D., 2015). SVM is one of the most popular methods used in object detection algorithms and especially passenger counting, because it offers a method to estimate a hyperplane that splits feature vectors extracted from pedestrian and negative samples (Cheng D., 2015), differentiating pedestrians from other unwanted features. Boosting aims to use a sequence of algorithms to convert weak learners to strong learners (Zhi-Hua, 2012). The main idea of the boosted classifiers is weighting weak classifiers and combining them to form a strong hypothesis when training the algorithm to attain an accurate detection. Current methods for object detection take a classifier for an object and evaluate it at several locations and scales in a test image which has been found to be time-consuming and created numerous computational instabilities at large scales (Deng J. B. A.-F., 2010).

The most recent methods such as Region Based Convolutional Neural Network (R-CNN), use another method to decrease the region the classifier runs and include the SVM. Firstly, category-independent regions are proposed to generate potential bounding boxes. Secondly, the classifier runs and extracts fixed-length feature vector for each of the proposed regions. Finally, the bounding boxes are refined by the elimination of duplicate detections and rescoring the boxes based on other objects on the scene using SVMs (Girshick R., 2014).

The technique that will be used in this project is bounding boxes prediction. The bounding box is a rectangular box located around the detections in order to represent their detection (Coniglio C., 2017) (Lézoray O., 2012). Object detection datasets are images with tags used to classify different categories (Deng J. D. W.-J.-F., 2009) (Everingham M., 2010).

2.2.2 You Only Look Once

You Only Look Once (YOLO) software uses a different method than the above-mentioned techniques for object detection. It generates a single regression problem, straight from image pixels to estimate bounding box coordinates and class probabilities (Redmon, 2016). YOLO uses a single convolutional network that simultaneously predicts multiple bounding boxes and class probabilities for these boxes (Redmon J., 2015). The minimum bounding box size is 13x13 tiles. The ability to train YOLO on images has the potential to directly optimize the detection performance (Redmon J., 2015) and increase the bounding box probabilities. Another advantage of YOLO is that, unlike other techniques such as SMVs, it sees the entire image globally instead of sections of the image. This feature enables YOLO to implicitly transform contextual information to the code about classes and their appearance and at the same time makes YOLO more accurate, making less than half the number of errors compared to Fast R-CNN (Redmon J., 2015). YOLO uses COCO which is a large-scale object detection, segmentation, and captioning dataset (COCO Common Objects in Context, 2018). The minimum bounding box restricted size is 13x13 tiles (Redmon, 2016).

Additionally, YOLO can learn and detect generalizable representations of objects, outperforming other detection methods, including R-CNN. It is imperative to state that YOLO could be used in numerous applications and it is less likely to break down when applied to new domains or unexpected inputs due to the ability to generalize (Redmon, 2016) (Redmon J., 2015).

2.2.3 Passenger Counting Using Bluetooth and Wi-Fi Detectors

There are a number of technologies for detecting electronic devices which can be used for passenger counting. The widespread standard is Bluetooth technology, which facilitates radio communications between smart devices. In order to be detected, a Bluetooth device must be set to discoverable, and this is reportedly between 5%–12 % of potential Bluetooth devices (Brennan T.M., 2010). The Bluetooth device detects the unique media access control (MAC) address for each device within range. The Bluetooth detector pings for devices over a period of time repeatedly, a running list of detected devices is collected with time-stamps for the times that they were observed. By considering devices as a proxy for passengers, this data provides observations of the time that a passenger arrives in a station (based on the first observed timestamp) and the time that they leave a station (based on the last observed timestamp), which is critical for detecting whether that passenger was left behind. An additional benefit of tracing MAC addresses is that it does not change for the same device, so if the same MAC address is later observed at another station, it represents a direct observation of a passenger movement from one location to another. Some previous studies have sought to use Bluetooth data to estimate transit wait times (Kurkcu A., 2017) and origin-destination flows (V., 2008) (Dunlap M., 2016).

Bluetooth scanning is based on polling, and not on passive listening. This makes Bluetooth detection slow and leaves the chance for a device to avoid detection by ignoring a polling request. Any smartphone can be configured to be visible or not by other Bluetooth devices. Setting this option as “NOT VISIBLE” will make the smartphone undetectable by any other Bluetooth device or sensor. This relates to the major downside of Bluetooth detection, which is that the sampling rate is very low, as stated above. When collecting data to aggregate over long periods of time, this may not be a big problem because the aggregation of a low sampling rate can still yield a large data set. Hence, estimates of how many passengers were left behind due to crowding could not be reliably made for a specific date and time. For the problem of identifying left behind passengers, it would be useful to have a much richer data set. For this purpose, there has been recent development of sensors that use both Bluetooth and Wi-Fi signals to detect devices.

Some products even use cellular or Wi-Fi signal detection and make use of a communications channel that allows devices to connect to a wireless local area network. This is a common communication for smart phones, tablets, and laptop computers that passengers often carry with them. A prominent manufacturer of combined Bluetooth and Wi-Fi detectors claims that by using Wi-Fi and Bluetooth signals, as many as 95% of smartphone, tablets, hands free devices, and laptops are detected by their MAC address within the detection range, Libelium Meshlium Scanner (Guide, n.d.). Detection of Wi-Fi enabled devices requires that Wi-Fi is on, and this is more likely to be the case in environments where people are used to using free Wi-Fi services. Nevertheless, more and more devices are left to scan for Wi-Fi signals at all times, so the detection rate is likely to be high, and certain to be higher than Bluetooth alone. A few challenges and complications are related to the use of wireless detection devices to identify passengers.

- Scanning devices must be installed – Unlike surveillance systems that have cameras already installed in stations, new devices would have to be acquired. For a long-term solution, these devices would have to be wired into communications channels in order to log records in a database.
- Scanner range – The range of detection systems varies greatly from outdoor to indoor settings. It is not clear what range the devices will have in rail transit stations, especially those that have many concrete columns and walls, which are likely to block signals. Therefore, depending on the architecture more than one devices might need to be installed
- Electronic devices do not map one-to-one with passengers – The essence of the technology is that it detects electronic devices that are enabled with communications, typically included in smart phones, tablets, computers, etc. Many commuters carry multiple devices, so it is likely that some passengers will be double counted. Likewise, some commuters do not carry any device at all or may not be detected at all. There is a risk that data from these sources will oversample relatively wealthier socioeconomic groups and undersample others. This raises some potential concerns for equity and sampling rates which will need to be carefully considered as part of a data collection trial.

2.2.4 Bluetooth and Wi-Fi sensors manufacturers

There are a few manufacturers who produce scanners that detect Bluetooth and Wi-Fi signals, but the manufacturers with applications related to public transportation systems and specifically platforms of subway stations are:

- Libelium1: Products from Libelium include a high-powered scanner, called Meshlium, that is designed to collect maximum number of MAC address signals using a combination of Bluetooth and Wi-Fi signals. Their applications include indoor environments where the scanner is used to count and track pedestrian movements.
- BlueMark Innovations2: BlueMark produces a modular platform to detect, track and locate smartphones based on Wi-Fi and Bluetooth (Classic, Low Energy, iBeacon, Eddystone) technology. They offer a components dashboard to view metrics, such

as unique visitors, admin portal to detect users. They claim to have a 25-meter range in an indoor location with pillars. The platform also offers ports of 3G/4G detection.

- SMATS3: A highly portable product called TrafficTab and mountable product called the TrafficBox provide Bluetooth and Wi-Fi detection capabilities in a portable case that can be easily mounted for temporary data collection. A more permanent product called TrafficXHub connects with a constant power supply for an extended scanning range and long-term data collection.

3.0 Methodology

The software used for this part of an ongoing research project, YOLO, is an open source software designed for real-time object detection in video streams (Redmon, 2016). The main element of its performance is setting the threshold of detection confidence (Redmon J., 2015), which means that the computer vision is able to determine how accurate the outputs will be. The videos that have been used come from the MBTA Orange Line in Boston, Massachusetts and more specifically from the security cameras that constitute the security system of Sullivan Square and North Station platforms.

In order to identify left behind with higher accuracy a crowding analysis was completed. Subsequently, the passenger detection software was used to determine the most representative views with the highest accuracy in object detection in comparison to the manual counts. Finally, the optimal threshold was computed by finding the most precise detections validated with manual counts.

3.1 Crowding Analysis

Crowding analysis is a necessary step in the methodology applied to identify the times and stations where crowding is observed and left behinds have a higher probability of occurring. The data used in this part of the analysis have been extracted from the MBTA Research and Analytics Platform (MBTA, 2018). More specifically, they represent rail flow data from the Winter of 2017, which was the most recent period of available data in the MBTA Research and Analytics Database.

The analysis in the Orange Line of the MBTA rail system has been focused on identifying the stops in which overcrowding phenomena lead to the higher probability of observing left-behind passengers during the day. An additional task has been the identification of the exact 15-minute time periods when such phenomena are expected.

Finally, the scope of this part, is to define the location and time that commuters are most likely to be left behind while using the MBTA commuter rail system in order to focus where and when to collect data to validate left-behind detections.

3.1.1 Passenger Flows

Cumulative counts of the numbers of passengers boarding and passengers alighting have been created with respect to stations along the direction of train travel. For a 15-minute time period, $B(n)$ is the count of all of the passengers that are assumed to board trains in the direction of interest at stations preceding and including station n. Similarly, the cumulative number of passengers alighting, $A(n)$, is the count of all passengers that are assumed to have exited trains traveling in the direction of interest at stations preceding and including station n.

It should always be true that $A(n) \leq B(n)$, because passengers can only alight a train after boarding it. The difference between the cumulative boardings, $B(n)$, and alightings, $A(n)$, provides an estimation of the passenger flow, $Q(n)$, between adjacent stations during each 15-minute time period.

$$Q(n) = B(n) - A(n) \tag{1}$$

This calculation is approximate, because cumulative counts are calculated for a single 15-minute time period, and real trains take more than 15 minutes to traverse the length of a line. Moreover, to calculate the crowding on trains, the passenger flow per time period should be converted to a passenger occupancy, $O(n)$ (passengers/train), which is calculated by multiplying the passenger flow by the scheduled headway, h (minutes), of trains.

$$O(n) = Q(n)\frac{h}{15} \tag{2}$$

In this equation, the headway is divided by 15 minutes to account for the fact that the passenger flow is per 15-minute time period. This measure is an approximation of the number of passengers onboard each train that is based on the assumption that real headways are uniform.

According to the official MBTA website, the Table 1 shows the scheduled headways for the Orange Line.

Table 1: Orange Line Headways

Orange Line	**First Trip**	**AM Peak**	**Midday**	**PM Peak**	**Evening**	**Late Night**	**Last Trip**
Oak Grove	5:16AM	6min	9min	6min	10min	10min	12:30AM
Forest Hills	5:16AM	6min	9min	6min	10min	10min	12:28AM

A major assumption is that there are no delays in the arrival and departure of the trains, thus the schedule is being strictly respected and the headways remain as reported by MBTA (MBTA, 2018).

The 2017 MBTA Service Delivery Policy (Service Delivery Policy, 2017) has been used in the following steps of this research. From Table 2, the capacity of each train on the Orange Line has been extracted. Trains on each line are 6 cars long, so each of these vehicle capacities is multiplied by 6 to obtain the capacity of a train.

The maximum vehicle load, according to the Service Delivery Policy, is 245% of seating capacity in the peak hours and 143% of the seating capacity in other hours. This standard was presented more explicitly in the 2010 Service Delivery Policy (Service Delivery Policy, 2010) and not explicitly emphasized in the 2017 revision due to the challenges associated with measuring occupancy and crowding on trains.

The comparison between the passenger load expressed as a percentage of seating capacity and the passenger load is theoretically an important indicator of left-behind passengers. Another way of identifying potential left-behind phenomena has been the comparison among the number of passengers boarding, the number of passengers alighting as well as, the passenger load per line and per 15-minute increments.

The Table 2 states that for Orange Line (both directions) the vehicle load during peak hours, as they are presented in Table 2, is 86 passengers which equals to approximately 148% of the seating capacity which is 58. In addition, during non-peak hours, the vehicle load, 50 passengers, equals to approximately 86% of seating capacity.

Table 2: Carriage capacity in Time Periods (Service Delivery Policy, 2017)

Line	Number of Seats	Total Number of Passengers			
		Early AM/ AM Peak 6:00 AM – 8:59 AM	*Midday Base* 9:00 AM – 1:29 PM	*Midday School/ PM Peak* 1:30 PM – 6:29 PM	*Evenings & Weekends* 6:30 PM – 9:59 PM
Orange Line	88	86	50	86	50

The time period as defined by the Service Delivery Policy (Service Delivery Policy, 2017) are shown in Table 2 above. The combination of tables 1 and 2 enables us to validate findings and determine the critical point between normal conditions and overcrowding.

3.1.2 Identification of stations and times with overcrowding phenomena

The stations with the maximum passenger load per 15 minute-time-period (both directions) have been identified in this part of the analysis. Two samples of the results of this procedure are given below, aiming at visualizing where and when trains are most crowded. This is a critical information in order to prioritize stations for further investigation.

As indicated by the tables presented above, if we only take account the maximum passenger load per train, the stops that could be more efficient to research due to overcrowding are the North Station and the Sullivan Square.

In order to observe the data more efficiently and reach to more firm conclusions, the following color map diagrams have been created, Figure 1 and 2. The color indicates the magnitude of the passenger loads, with blue being the lowest and red the highest. The capacity is 516 passengers/train and it is illustrated with orange. Any shade of red shows overcrowding and the darker the shape the higher the load. The Figures 1 & 2 show the passenger loads across all stations of the Orange Line in both directions throughout the course of a day.

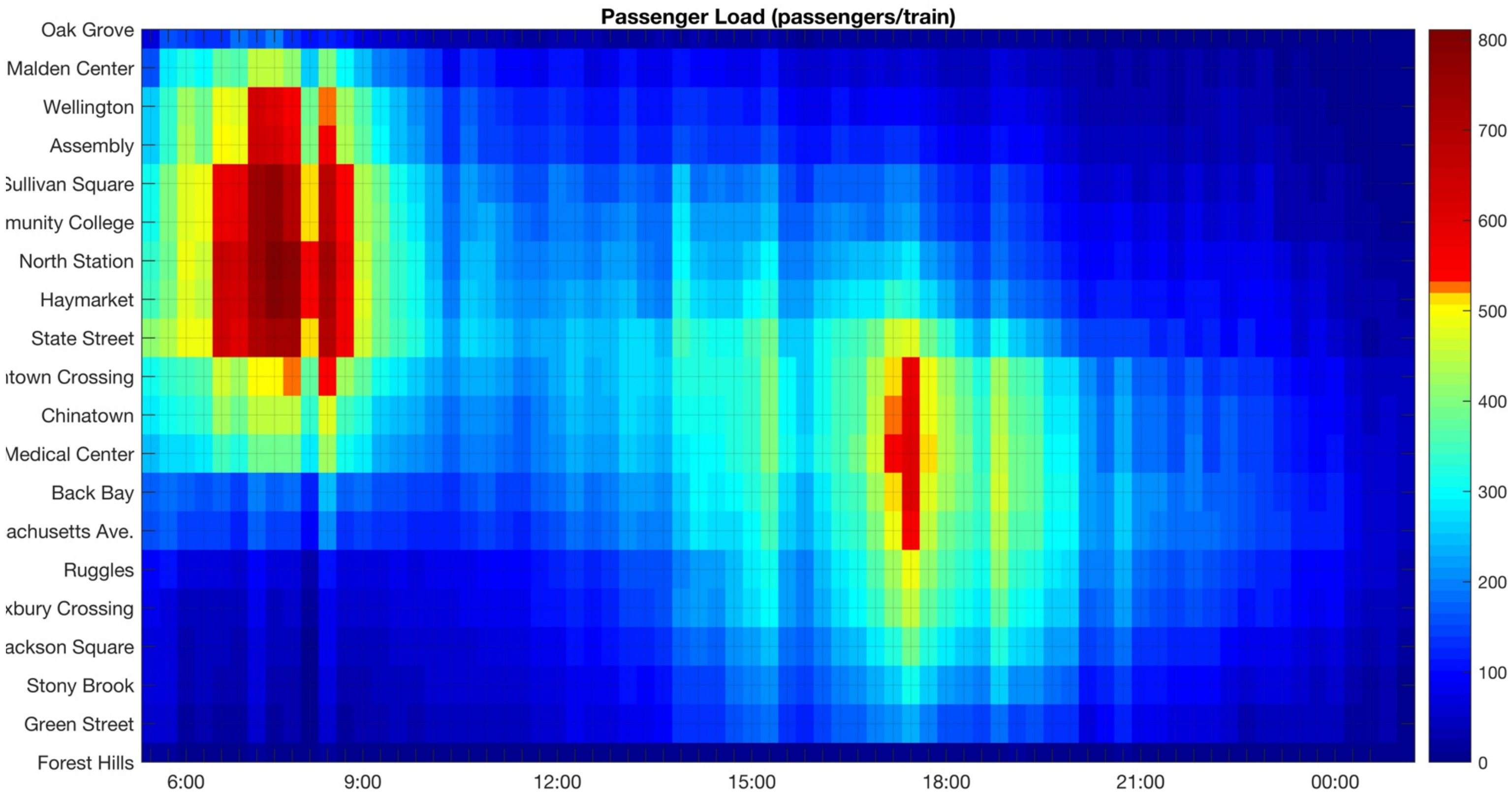

Figure 1: Passenger Load Colormap Oak Grove to Forest Hills

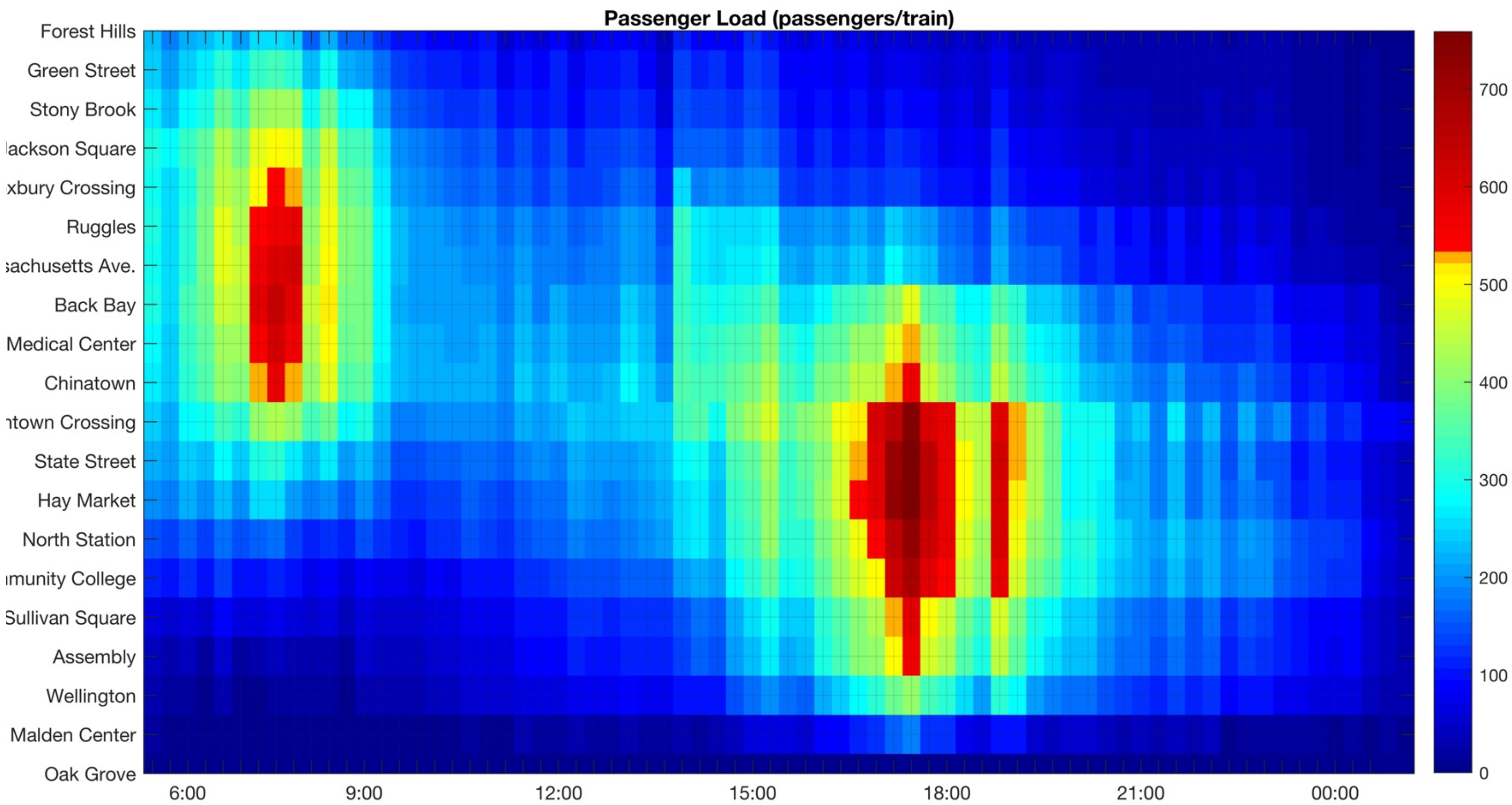

Figure 2: Passenger Load Colormap Forest Hills to Oak Grove

3.2 Camera rating

Camera rating has been implemented to identify the cameras that would give us the best views as input for our video processing analysis, with the purpose of detecting the maximum number of passengers.

Using the results of the crowding analysis, regarding the stations with maximum possibility of having left-behind passengers, we proceeded with the evaluation of the related camera videos. The sample of camera views from the MBTA transit line has been reviewed and the best camera views have been chosen according to number of passengers they detect through YOLO. In order to perform the evaluation of the views, we compared our counts of passengers with YOLO counts.

3.3 Threshold selection

In order to identify the optimal threshold, all the available camera screenshots from all the stations were analyzed. Each screenshot was run separately for threshold values ranging from 6% to 25% so as to determine the optimal threshold value in relation to the human eye count from each screenshot. According to the results of the image processing, we determined that the optimal threshold was 7%. This 7% threshold was chosen due to the smallest mean squared error difference between YOLO and human eye counts. Figure 3 below shows the progress of the threshold testing procedure using different threshold values.

Figure 3:Testing Thresholds

In Table 4, a sample of the images that were analyzed in YOLO are shown above, as well as the detected passengers for each different threshold value tested.

Table 3: Threshold errors from YOLO counts

Threshold	Mean Error	Mean Squared Error
25%	6.82352941	62
20%	5.82352941	43.70588235
15%	4.58823529	28.47058824
10%	3.05882353	14.11764706
9%	2.23529412	8
8%	1.35294118	4.647058824
7%	0.11764706	1.176470588
6%	-1.5294118	6.823529412

4.0 Results and Discussion

4.1 Passenger Detection Time Series

Applying this threshold to the analysis enables us to derive the maximum accuracy of counts of passengers being on the platform over time. Using YOLO a text file including location information for each detection, the frame of the video when it was detected and its nature (e.g. handbag, train, person) was produced. Using the output file, a time series of the people detected at each second was computed. Moreover, by adding the times the doors on the train close on the time series and looking at a shot time interval after that time, we can determine the number of passengers that may be left behind on the platform at this point in time. In Figure 4 below, the passenger counts time series illustrate the number of passengers detected by YOLO over time. The green and the red vertical lines represent the times that the doors were opening and closing, accordingly. The times of opening and closing doors were collected manually.

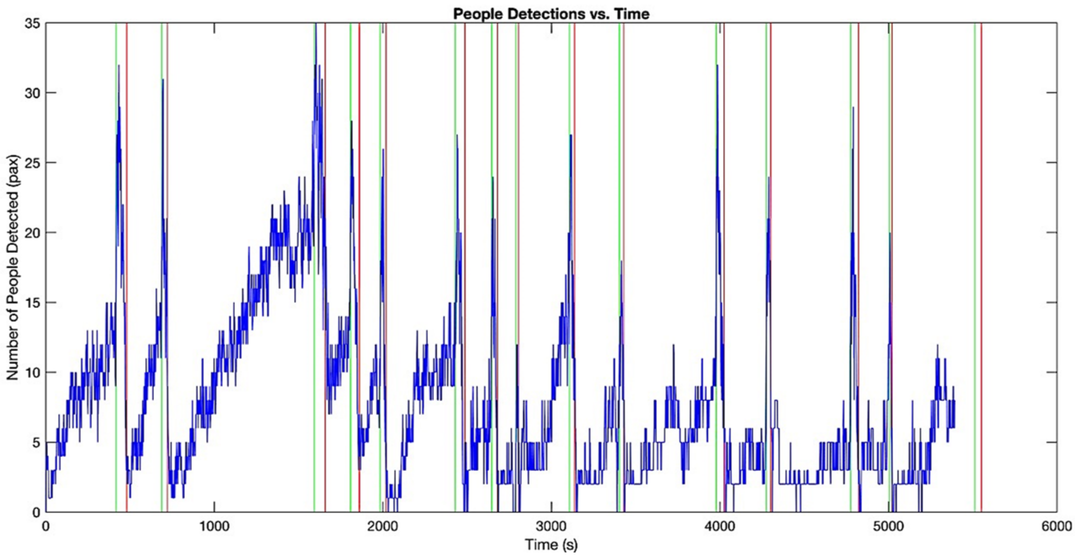

Figure 4: Number of Passenger on the Platform over time Unsmoothed

However, Figure 4 has a lot of noise and does not illustrate clearly the peaks and troughs and for this purpose a moving average of the time series was created shown in Figure 5 below. The moving average averaged ten numbers before and after each second's detection. Now, it is clearly visible where the trains left and how many passengers were detected after the trains left.

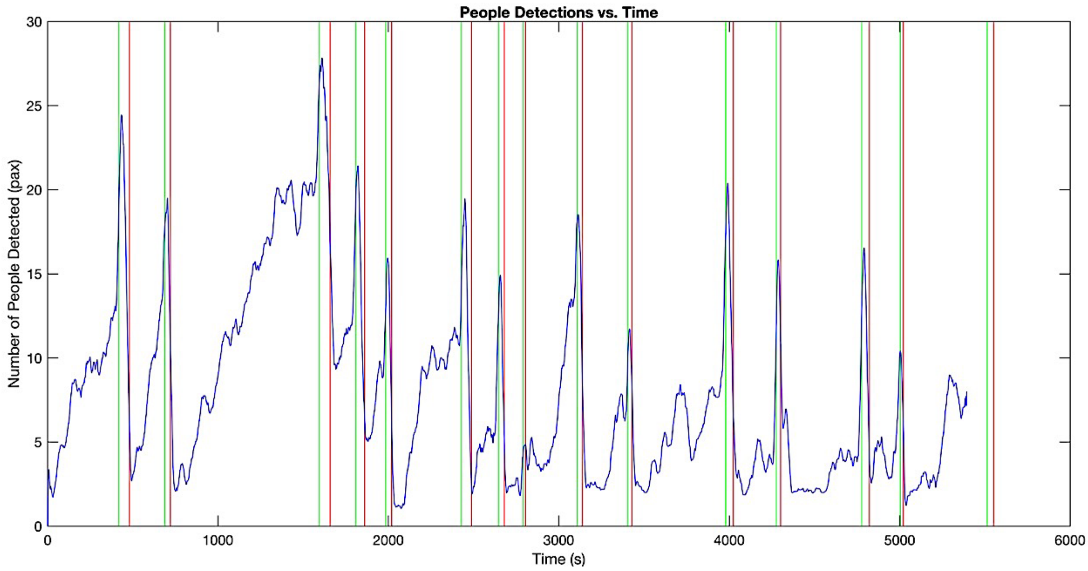

Figure 5: Number of Passengers on Platform Smoothed

In the figure above, the green lines represent the times that the train doors open and the red lines the times that the train doors close. The concept in this case is that the number of passengers on the platform right after the doors close may be a good indicator of the number of passengers being left-behind or the occurrence of left behind passengers.

Figures 6 and 7 below illustrate the passenger detections in both North Station and Sullivan Square in November. The first graph in Figure 6 shows the detections at North Station from 3:30 PM to 5:00 PM. The second graph in Figure 6 illustrates the detections at North Station from 5:00 PM to 6:30 PM.

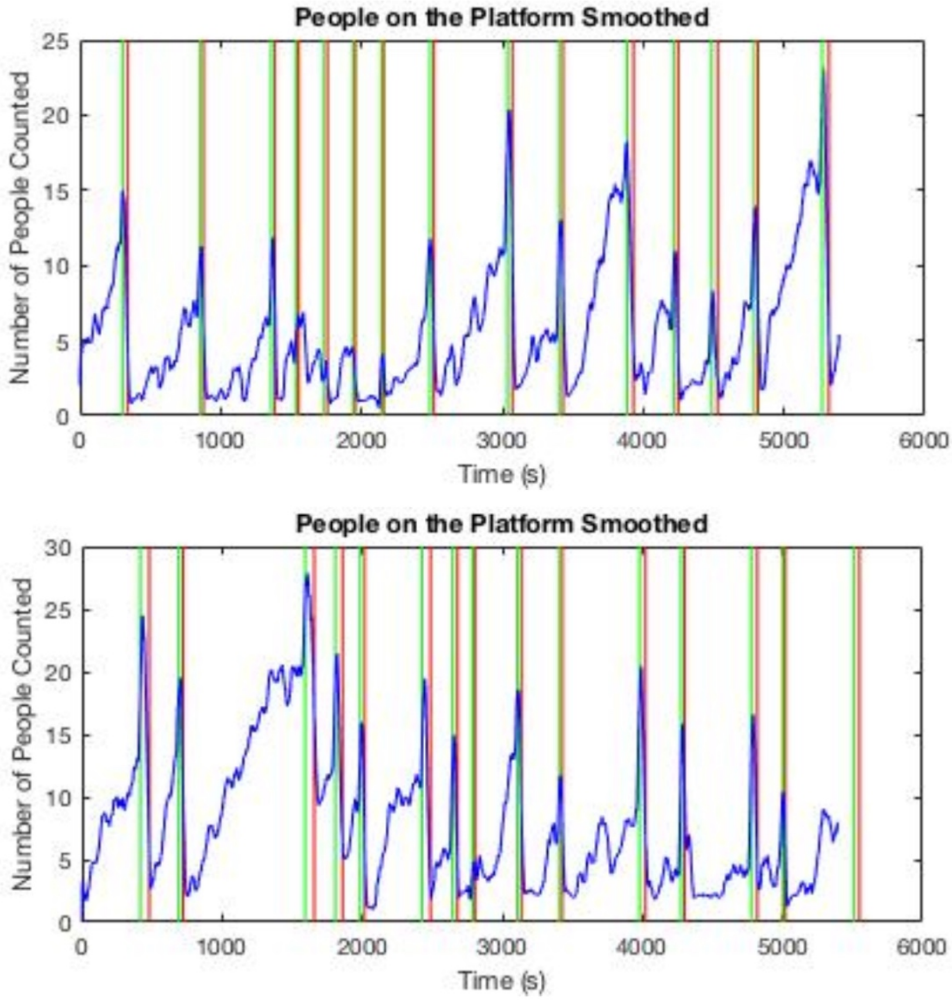

Figure 6: North Station 11/15/2017 Smoothed Counts

The first graph in Figure 7 shows the detections at Sullivan Square from 6:30 AM to 7:45 AM. The second graph in Figure 7 also illustrates the detections at Sullivan Square from 7:45 AM to 9:30 AM.

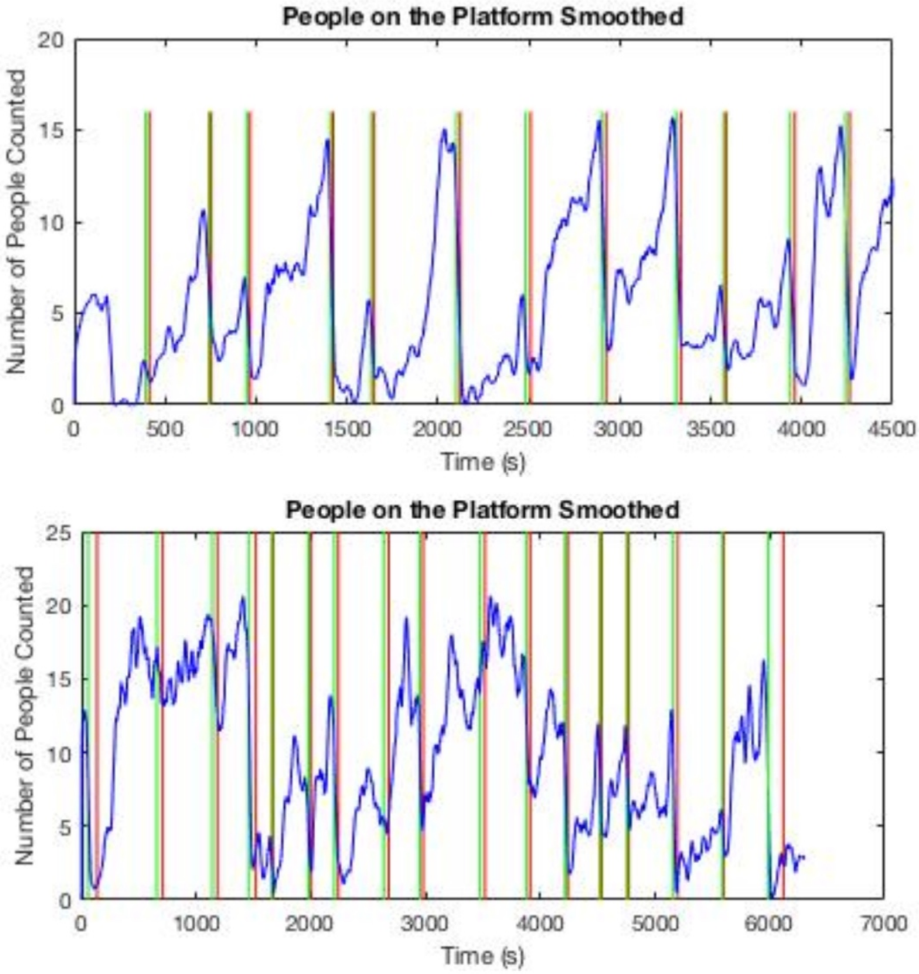

Figure 7: Sullivan Square 11/15/2017 Smoothed Counts

Figures 8 and 9 below illustrate the passenger detections in both North Station and Sullivan Square in January. More specifically, Figure 8 shows the detections at North Station from 3:30 PM to 6:30 PM.

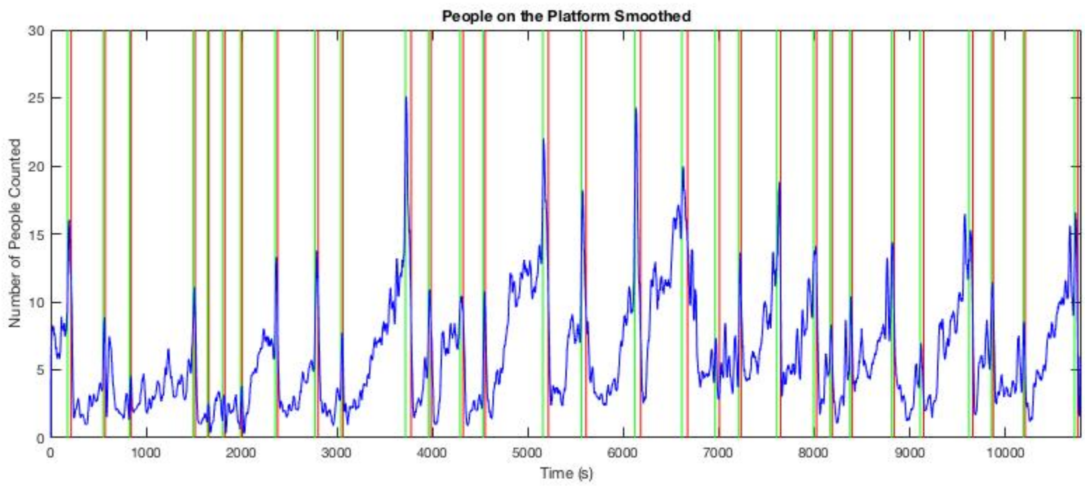

Figure 8: North Station 1/31/2018 Smoothed Counts

Figure 9 shows the detections at Sullivan Square from 6:30 AM to 9:30 AM.

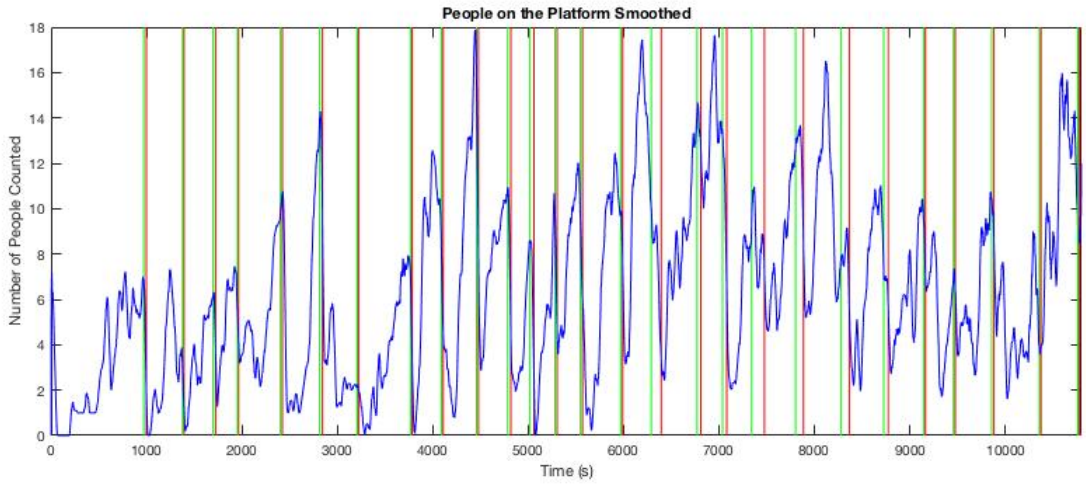

Figure 9: Sullivan Square 1/31/2018 Smoothed Counts

Apart from the aforementioned time series, a similar time series was produced using the surveillance videos of the stairs and escalators. The data collection was a manual procedure counting the passengers entering and exiting the platform and logging the time using an open source software. The difference of passengers entering and exiting the platform estimates precisely the number of passengers in the platform at each second, assuming zero passengers when the counting began. In Figure 10, the combination of smoothed YOLO counts and number of passengers on the platform are shown, as well as the scaled YOLO counts to observe their fit to the actual counts. The scaled counts fit the actual counts with an R^2=0.74 using a regression analysis.

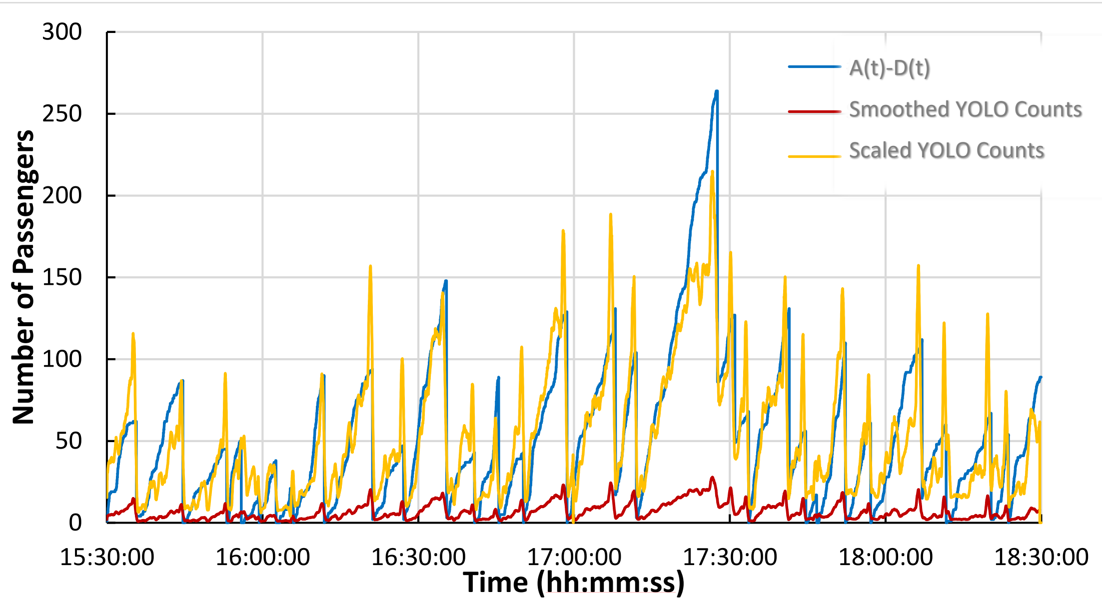

Figure 10: Comparison of Manual and YOLO counts in North Station 11/15/2017

The same graph was produced for the January 31, 2018 collected data and the new YOLO detections from the surveillance cameras. The fit was improved to R^2=0.84, 10% better than November's data, the graph is shown on Figure 11 below.

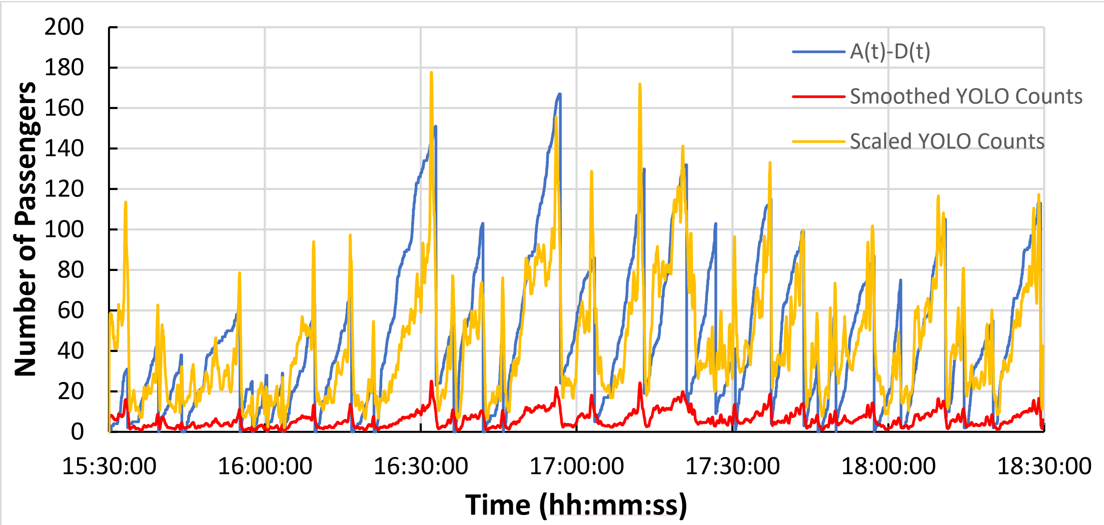

Figure 11: Comparison of Manual and YOLO counts in North Station 1/31/2018

4.2 Bluetooth and Wi-Fi signal detection

Another technology widely used for passenger detection is identifying smartphone and other electronic devices' signals. Through logging individual addresses for each electronic device over time a similar pattern is expected to be observed as the one using object detection through surveillance cameras. For this purpose, four boxes in total, two were mounted in Sullivan Square and two North Station, each including one Bluetooth and one Wi-Fi sensor. The boxes collected data the night before the manual data collection when they were mounted and throughout the next day, when the manual data and surveillance videos were collected too. The batteries of the boxes collected data for about 20 hours before they run out. The total number of observations was almost 1,5 million with 87% of those being Wi-Fi signals. However, the period of interest was the three hours of manual data collection from 6:30 am to 9:30 am in Sullivan Square and from 3:30 pm to 6:30 pm in North Station. Therefore, the data corresponding to those two periods were 375,000 observations. The unique MAC addresses of those observations were 28,800. Those observations were far more than the observed ones, from the manual data collection. There might are many reasons for this result, including that there are devices there for very long periods of times, such as routers, modems etc. and devices that are not associated with the platform that was studied. Hence, two thresholds were set in order to eliminate unwanted electronic devices. Firstly, devices that were detected for less than five seconds and more than three peak hour headways were eliminated from the data and secondly, devices that did not have their last observation within 120 seconds of a departing train were deleted. Table 5 below shows the number of observations that satisfied both constraints.

Table 4: Bluetooth and Wi-Fi observations 1/31/2018

	Sullivan Square (AM Peak)		North Station (PM Peak)	
Data	Box 3	Box 4	Box 1	Box 2
Total Number of Observations (~20 hours)	187,732	439,294	306,156	553,673
Observations During Peak (3 hrs)	55,628	115,719	79,239	128,425
Unique MAC Addresses in Peak	16,396		12,431	
*Filtered MAC Addresses	3,963		8,406	

*Duration from first to last observation ∈ (5, 960) seconds & last observation within 120 sec of a departing train

Moreover, in order to compare the detections with the manual counts, a time series using the number of detecting at each second of the filtered MAC addresses was produced. Figure 12 below illustrates the comparison between manual and combined Bluetooth and Wi-Fi counts.

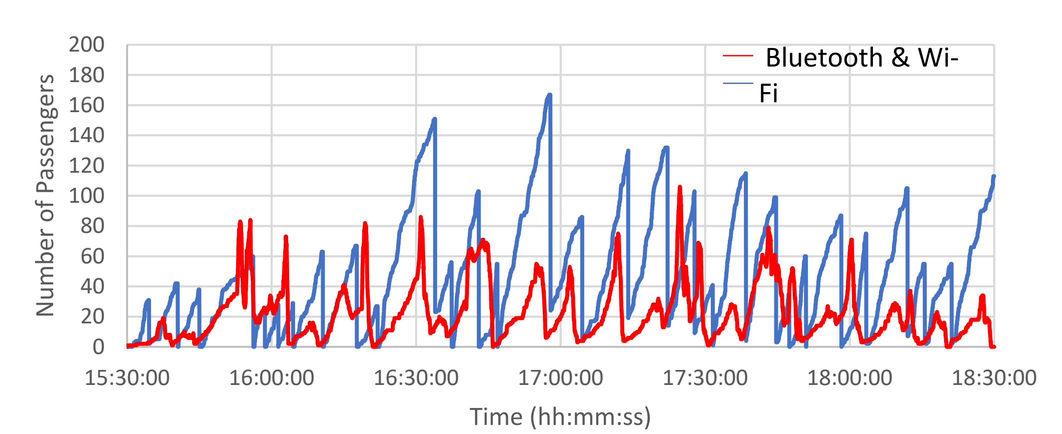

Figure 12: Comparison of Wireless and Manual Counts– North Station 1/31/2018

The Bluetooth and Wi-Fi counts shows a similar pattern with the manual counts but there is no statistically significant re-scaling factor and therefore this source is not indicative for left behind passengers. However, using the Wi-Fi and Bluetooth data the cumulative percentile against time was calculated in comparison with the waiting times of the manual counts observed. Figure 13 illustrates the abovementioned metrics that may be indicators of the conditions on the platform, in terms of the reliability standard by MBTA Service Delivery Policy. Specifically, the MBTA Service and Delivery Policy states that 90% of the commuters should be served within one headway, which, in peak hours, is 6 minutes.

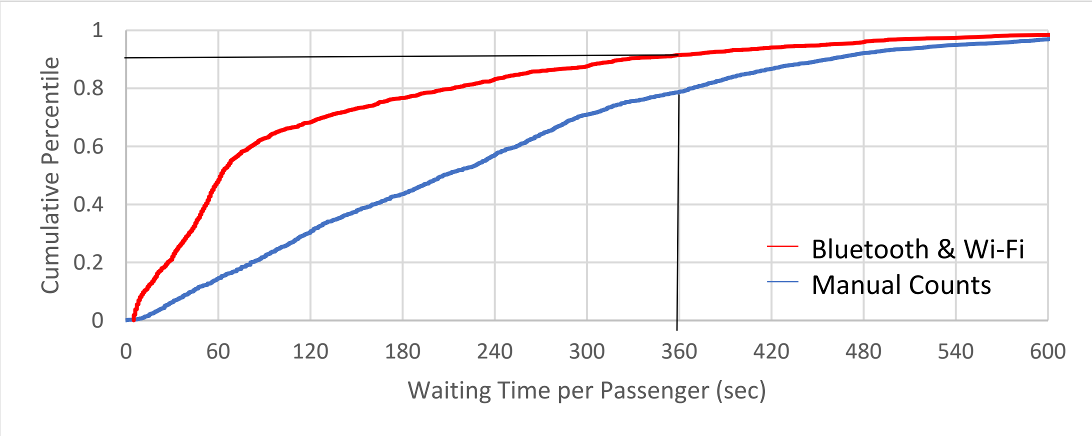

Figure 13: Comparison of Waiting Times – North Station 1/31/2018

Figure 13 shows that Bluetooth and Wi-Fi counts overestimate the percentile by almost 10% but they might be a good indicator of the reliability of the system. Additionally, the Bluetooth and Wi-Fi counts can be used to detect devices in multiple stations and inferring origin and destination pairs since a tap-out system does not exist in the Boston Subway yet. The sensors are not accurate enough to detect devices and left behind passengers because the signal ping is not constant and there is no way of controlling it. Therefore, the exact time of arrival and departure of each unique MAC detection is not accurate and might be totally misleading.

5.0 Conclusions

The scope of this project is to measure passengers being left behind due to crowding on the MBTA Orange Line in Boston using emerging technologies. In order to achieve that, a crowding analysis was completed to determine the stations with the highest probability of detecting left behinds.

Manual counts of left behind passengers were collected in two different dates on the two most crowded stations, Sullivan Square and North Station. Likewise, the surveillance videos of these stations were provided by MBTA, in order to extract passenger detections using YOLO, an object detection software, to compare with manual counts. Furthermore, four Bluetooth and Wi-Fi sensors were installed in the two stations. Again, the data collected from the sensors are going to be compared to the manual counts. Finally, the two methods using technologies to detect left behinds are going to be compared to each other.

To sum up crowding leading to left-behind passengers is a regular occurrence on MBTA heavy rail, even on uneventful days. The demand is very near capacity, therefore even some small fluctuations of headways lead to overcrowding condition and induce left behind passengers. Moreover, accounting for left-behind passengers reduces the reliability measure according to the Service Delivery Policy for the passengers waiting less than one headway in the peak hours. It was observed that headway and dwell time are strong determinants of left-behinds at Sullivan Square and North Station, due to occurrence of left behind passengers when those two parameters are increased. Introducing logistic regression models would validate this observation and would prove the significance of headway and dwell times in relation to the occurrence of left behind passengers.

In terms of the video processing analysis, it was based on a simple off-the-shelf algorithm, that could be improved for increased accuracy and precision of detections. Additionally, a dataset specifically made for detecting pedestrian could be implemented instead of COCO to further optimize the detections accuracy.

The direct observations from video feeds are associated with errors. For example, the range of detection which included a small part of the platform and individual bodies were difficult to distinguish in crowded conditions. Finally, many camera angles are blocked or obscured by objects that increased the possibility of undercounting passengers on the platform and lead to misleading results. However, the scaled counts were statistically significant and represented real conditions with high accuracy.

Likewise, the predictive models built on video detection observations provide good predictions of metrics of interest namely, the number of people left behind per rush period, the occurrences of trains leaving people behind and the distribution of waiting times experienced by passengers.

Conversely, the estimates of Left-Behinds or platform counts were very noisy and not as reliable as video counts from the Bluetooth and Wi-Fi sensors. The main reason for this result was the relatively low and highly variable sampling rate leading to inaccurate station entry and boarding estimates. In future research, the sensors are advised to be used in multiple station for the estimation of Origin and Destination pairs, as exact time of arrival and departure cannot be acquired.

Similarly, it was extremely difficult to determine which devices were associated with the train in the platform studied in crowded situations and when many trains from different lines arrive in a quick succession and short time interval. Even when the detections were matched to the corresponding train departures the measured duration from wireless sensors will always be less than actual wait time in the station causing the cumulative waiting time estimates to get biased. Finally, the statistical models' performance was not improved by wireless sensors' data.

5.1 Next Steps

There is much room for improvement in terms of detection accuracy using both the surveillance cameras and the sensors. In order to improve object detection through video processing several aspects can be addressed accordingly. There are other faster and more accurate video detection algorithms that can be used for detection which are expected to further reduce false negatives. However, such algorithms are not free of charge.

Adding passenger tracking algorithms to link observations in consecutive frames can significantly reduce false positive detections and would also allow the tracking of passenger movements that would identify demand patterns across the platform. There is an opportunity to train algorithm to detect only the heads of passengers rather than their whole bodies which will possibly increase the number of bounding boxes that can fit in each frame and therefore, increase the accuracy of the detections.

Expanding the analysis to compare a broader number of stations and settings, such as stations with more obscured views, less consistent demand patterns, etc. would be able to examine the methodology and systemwide implementation possibilities.

The is a need to experiment further with the capabilities of the Bluetooth and Wi-Fi sensors by distributing the devices and mounting for extended periods of time in order to evaluate origin-destination patterns or origin-destination travel times.

Ultimately, a costly suggestion for MBTA is to add new data sources such as Automated Passenger Counters (APC) on trains or a tap-out system that will provide very useful and precise measures of crowding that are relevant to passengers being left behind and exact origin-destination travel times

References

Bell S., Z. C. (2015). *Insideoutside net: Detecting objects in context with skip pooling and recurrent neural networks.* arXiv preprint arXiv:1512.04143.

Brennan T.M., E. J. (2010). *Influence of vertical sensor placement on data collection efficiency from Bluetooth MAC address collection devices.* Journal of Transportation Engineering.

Carrion C., L. D. (2013). *Uncovering the influence of commuters' perception on the reliability ratio.* Washington D.C.

Cheng D., W. J. (2015). *Training mixture of weighted SVM for object detection using EM algorithm.* Neurocomputing.

Coniglio C., M. C. (2017). *People silhouette extraction from people detection bounding boxes in images.* Pattern Recognition Letters.

Deng J., B. A.-F. (2010). *What does classifying more than 10,000 image categories tell us?* ECCV.

Deng J., D. W.-J.-F. (2009). *Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition.* IEEE.

Douglas N., K. G. (2005). *Estimating the cost to passengers of station crowding.* 28th Australasian Transport Research Forum (ATRF).

Dunlap M., L. Z. (2016). *Estimation of origin and destination information from Bluetooth and Wi-Fi sensing for transit.* Transportation Research Record.

Everingham M., V. G. (2010). *The pascal visual object classes (voc) challenge.* International journal of computer vision.

Fan Y., G. A. (2015). *Perception of Waiting Time at Transit Stops and Stations.* University of Minnesota. Transportation Research Part A: Policy and Practice.

Girshick R., D. J. (2014). *Rich feature hierarchies for accurate object detection and semantic segmentation.* In Computer Vision and Pattern Recognition (CVPR).

Grant M., P. T. (2011). *State DOT Public Transportation Performance Measures: State of the Practice and Future Needs.* NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM.

Guide, M. T. (n.d.). *http://www.libelium.com/development/meshlium/documentation/meshlium-technicalguide/.* Retrieved from Libelium.

Hörcher D., G. D. (2017). *Crowding cost estimation with large scale smart card and vehicle location data.* Transportation Research Part B.

Haywood L., K. M. (2015). *The distribution of crowding costs in public transport: New evidence from Paris.* Transportation Research Part A.

Haywood L., K. M. (2017). *Crowding in public transport: Who cares and why?* Transportation Research Part A.

Kim K.M., H. S.-P.-J. (2015). *Does crowding affect the path choice of metro passengers.* Transportation Research Part A.

Kittelson & Associates Inc, A. (2013). *Transit Service and Quality of Service Manual.* Parsons Brinckerhoff, KFH Group Inc., Texas A&M Transportation Institute. Washington D.C.: Transportation Research Board of the National Academies.

Kurkcu A., O. K. (2017). *Estimating pedestrian densities, wait times, and flows using wi-fi and Bluetooth sensors.* Washington D.C.: Transportation Research Board 96th Annual Meeting.

Lézoray O., G. L. (2012). *Image processing and analysis with graphs: theory and practice.* Digital Imaging and Computer Vision. CRC Press / Taylor and Francis.

Li Z., H. D. (2011). *Crowding and public transport: A review of willingness to pay evidence and its relevance to project appraisal.* Transport Policy.

Lundberg, U. (1976). *Urban commuting: crowdedness and catecholamine excretion.* Journal of Human Stress.

M., K. (1991). *Discomfort externalities and marginal cost transit fares.* Journal of Urban Economics.

MBTA. (2018, 2 20). *MBTA Research and Analytics*. Retrieved from https://mbta01.mghpcc.org/MbtaResearch/home.xhtml

P., V. (1993). *Feature-Based Recognition of Objects.* MIT AI Lab. AAAI Technical Report FS-93-04.

Redmon J., D. S. (2015). You only look once: Unified, real-time object detection.

Redmon, J. (2016). *Darknet: Open source neural networks in c*. Retrieved 1 29, 2017, from http://pjreddie.com/darknet/

Schneider R. J., A. L. (n.d.). *Use of Automated Counters to Extrapolate Weekly Volumes from Short Manual Counts.* Transportation Research Record 2140.

Service Delivery Policy. (2010). *Massachusetts Bay Transportation Authority*.

Service Delivery Policy. (2017). *Massachusetts Bay Transportation Authority*, 9-45.

Siala M., K. N. (2009). *People detection in complex scene using a cascade of boosted classifiers based on haar-like-features.*

Sonka M., H. V. (2014). *Image Processing, Analysis, and Machine Vision.* Cengage Learning.

Tirachini A., H. D. (2006). *Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand.* Sydney: Transportation Research Part A: Policy and Practice.

V., K. (2008). *Using Bluetooth to capture passenger trips on public transport buses.*

Z., Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms.* Chapman and Hall/CRC Press.