Verifying Patterns from Stopped Traffic at Intersections

Ciera Christopher
Institute for Software Research, International
Carnegie Mellon University
cchristo@cs.cmu.edu

Abstract
This paper examine where a driver stops relative to the white stop line. We use a public web camera to gather and analyze this data. From this, we learn about interesting results that could be used for security purposes. Our initial data shows a pattern of traffic, and we show how statistics can be used to find a pattern statistically significant. The initial data also highlights some data outside of the normal pattern due to a local event near the intersection examined.

Introduction

Web cameras are increasingly used for monitoring traffic on freeways and at major intersections. Many of these cameras are already installed in major cities for the purpose of checking for traffic patterns and vehicle accidents.

The purpose of this paper is to show how we can get interesting data from a seemingly simple observation taken from publicly available web cameras. If these observations are sufficiently simple, it is possible to programmatically analyze the data. Since the web cameras are already set up and publicly available, we can gather large amounts of data for minimal cost.

The observation we will use for this paper is the number of drivers who cross over the white stop line at an intersection. This observation should be simple enough that a program could do it. Due to the way our web camera is setup, we will not be able to do this programmatically. However, it is possible to sample data in small time intervals in order to approximate the entire population of data.

In this paper, we will describe our methods for counting the number of drivers who cross over the white stop line at an intersection. We will gather the data on normal weekdays and show that a clear pattern emerges. Finally, we will use statistics to discover whether the data is statistically significant.

Methods

Our method is to count all the cars that are stopped at an intersection. To do this, we need to answer three questions:

1. Where should we look for cars?
2. Which cars will be counted?
3. What defines a car as being over the line?
4. How shall we count the cars?
5. How shall we determine the validity of the data?

We answer each of these questions in this section.

Location

We must find public web cameras of intersections that give a clear view of the stopping lines and the cars. Preferably, we would like a top down view of a full intersection. We also prefer intersections that are busy enough to have a car in every lane at the change of every traffic cycle. This will provide us with more data. Finally, the web camera of the intersection must provide streaming video. This way it is easy to distinguish between traffic cycles and to differentiate stopped cars from moving cars.

Figure 1: The intersection of Forbes and Bigelow, midday

For this report, I have examined the intersection at Forbes and Bigelow in Pittsburgh, PA. The camera watching this intersection is located at the University of Pittsburgh's Cathedral of Learning [1]. Figure 1 is a picture of the entire intersection. Unfortunately, it is not possible to do a programmatic analysis because the camera can be moved by visitors of the camera's website. However, it is the only camera found of a busy intersection that provides a clear view of the entire intersection and the white stop lines. Figure 1 shows a sample image from the web camera.

While it is not possible to see the traffic lights from this angle, it is easy to follow the traffic flow because there are always cars stopped in one of the directions of traffic. Unfortunately, there are only three directions of traffic because Forbes is a one-way street. However, watching this camera for 15-minute increments usually provides a statistically significant amount of data of over 50 cars stopped.

Counting cars

We will only count first car in a lane that must stop at the intersection. Cars that go through the intersection on a green light are ignored because we cannot tell what they would have done if they were stopped. Likewise, cars beyond the first are ignored. In Figure 2, we count three cars stopped at the light.

Figure 2: Three cars stopped on Forbes

Classifying cars

We need to decide what constitutes a car as being "over the line". We have three groupings for cars: those that obviously behind the line, those that are very close to the line on either side, and those that are obviously over the line. In Figure 2, the closest car is on the line, while the other two are behind the line. Figure 3 shows an example of a car obviously over the white line.[1]

Figure 3: A car over the white line

Should a car move forward after stopping, we will adjust the category if the driver is forced to stop again due to a red light. That is, drivers making a right turn who stop for pedestrians will keep their original designation. Drivers who move forward and stop because the light is still red will receive a new designation.

Additionally, we will count the number of cars in each category that "speed into" the stop. While this is not an exact measurement, it is possible to identify cars that are moving faster than the rest of traffic.

Sampling of Cars

This observation is simple enough that software should be able to take data from a web camera and analyze it programmatically. However, because unknown operators can move our selected web camera, we cannot depend on a programmatic solution. Instead, we will sample data from the population. The population is all the cars that must stop at the intersection at a particular time of day. We will watch the intersection in 10-15 minute intervals throughout the day[2], and we will use this data as a statistical sample.

Validity of the Results

From previous papers [2,3], we found that a pattern emerged on a typical weekday. In this paper, we attempt to verify whether these results are statistically significant or simply a random pattern. To do this, we will gather data over the course of several days. This will provide us with several data points for each time of day. If the pattern is significant, then the averages for the points before and after a given point in time should fall outside of the 95% confidence interval for that time of day.

---

[1] In fact, this car is so far over that other cars must make a wide arc to turn left, and pedestrians cross around the car. This causes severe problems when buses and trucks need to make turns on Pittsburgh's narrow streets.

[2] In practice, this gives a statistically significant number of cars that must stop at the intersection, usually 30-50 depending on the time of day.

For example, consider the data displayed in Table 1.

Table 1: Average proportions of cars behind the line

	mid afternoon	late afternoon	early evening
Behind	. 6669	.6157	.5937

For this data to show a statistically significant pattern, we expect that the proportions for mid afternoon and early eveing are outside of the 95% confidence interval for late evening. Put another way, we need a small enough standard deviation that .6157 + 2(StdDev) is less than .6669, and .6157 - 2(StdDev) is greater than .5937. This will show that there is indeed a pattern, rather than just random noise.

In order to perform this validation, we will need many data points. Due to the time limit on this project, we could not gather enough data to make a small standard deviation. In the next section, we display the current results and discuss how this could be used for future work.

Results

I gathered the data throughout the course of a series of weekdays. I grouped the data based upon the time of day. The averages for these results are in Table 2, and they are displayed graphically in Figure 4.

Also, I noticed that there were no cars that sped up to the light and then did not cross the line. Every car that was moving faster than the regular flow of traffic would cross the line. Since there were so many of these cars, I made a separate listing for them.[3]

Table 2:Averages of cars in each category on a weekday

	early morning	mid morning	late morning	midday	early afternoon	mid afternoon	late afternoon	early evening	mid evening	late evening	night
Behind	21	26.5	20	25	27	26	29	28.5	28	25	13.5
On	8	14.5	9.5	11	8	12.5	12	8.5	13	11	5.5
Over	10	9	8.5	7.5	4	3	5.5	7.5	11	11	7
Over/Moving Fast	2	3	3	2	0	0.5	1	3.5	2	2	2.5
Total	41	53	41	45.5	39	42	47.5	48	54	49	28.5

[3] During observations, I saw four cars run through a red light, once dodging oncoming traffic! I counted these cars as crossing the line and speeding to the light.

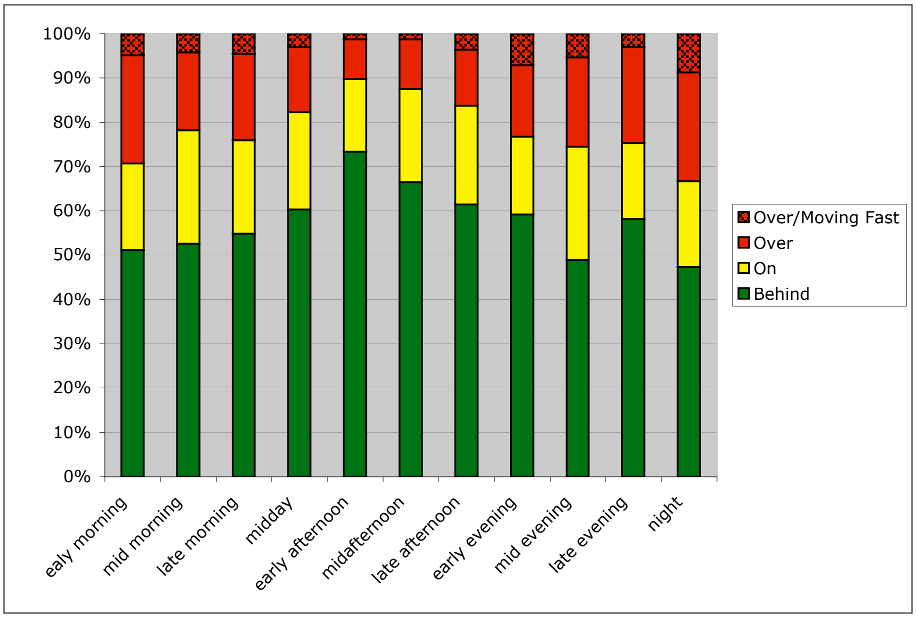

Figure 4: Average proportions of where cars stopped at the intersection

As discussed in [2, 3], we continue to see the pattern where drivers are aggressive in the morning, become less aggressive in the middle of the day at peak traffic for this intersection, and become highly aggressive in the evening.

At this point, we are interested in finding out if this pattern is significant. For example, we see that approximately 50% of drivers stop behind the line in the morning, and almost 70% of drivers stop behind the line in the afternoon. We'd like to verify that these numbers have significance. To do this, we will follow our method outlined in the previous section.

For each category (behind/on/cross/fast), we created a table of the data points gathered. The table for the "behind" category is displayed in Table 3. Data points where we could not gather data we left blank. These points were not included in further calculations.

Table 3: Proportions of cars behind the line

	early morning	mid morning	late morning	midday	early afternoon	mid afternoon	late afternoon	early evening	mid evening	late evening	night
4/6		0.4643	0.4444	0.4681		0.5238	0.5882	0.6364			0.4615
4/10		0.5400	0.5714	0.6364	0.6923		0.6364	0.5577		0.5102	
4/11						0.7143			0.5185		0.4839
4/12	0.5122										
4/20		0.5532		0.7111	0.7750					0.6341	
4/21					0.5769	0.7073	0.6000				
4/24		0.5714	0.6471			0.7222	0.6383	0.5870	0.4500		
Average	0.5122	0.5322	0.5543	0.6052	0.6814	0.6669	0.6157	0.5937	0.4843	0.5722	0.4727
Std. Deviation	0.0000	0.0408	0.0836	0.1016	0.0812	0.0828	0.0220	0.0325	0.0343	0.0620	0.0112

At most, we are averaging only four data points at each time slot[4], so our data isn't very good yet. We can graph the information in Table 3 to get a pictorial view of what is happening. This graph is in Figure 5.[5]

[4] In the early morning case, the standard deviation is completely useless since there was only one time slot. If the gathering could be automated, we would receive much better results.

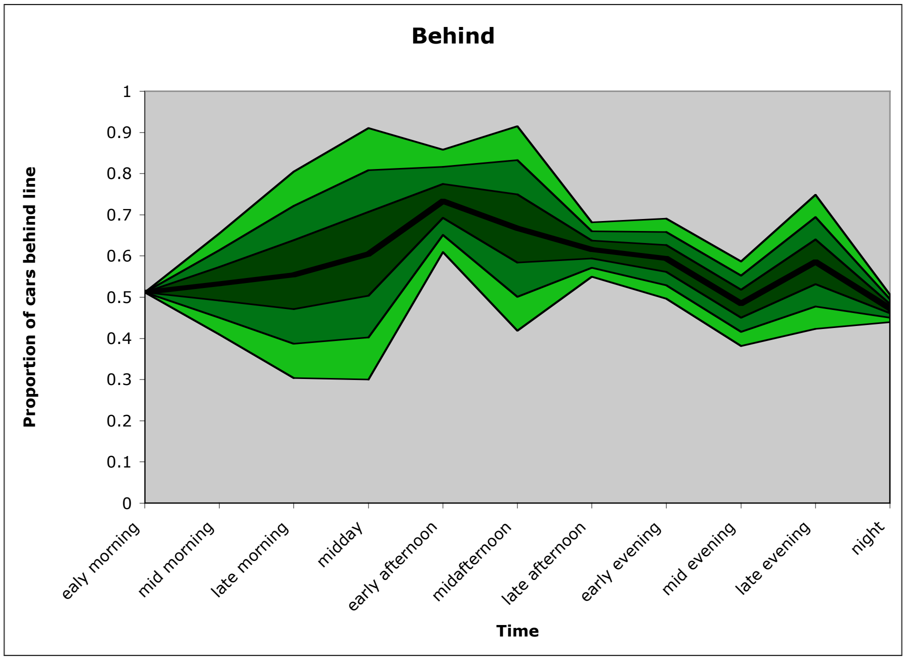

Figure 5: Standard Deviations of cars stopped behind the line

The black line in the middle represents the average. The dark green area is one standard deviation away, the middle green is two standard deviations, and the light green is three standard deviations. For our results, we are concerned with the 95% confidence interval, so we inspect the area two standard deviations away from the average. As we can see, this interval varies widely and spans such a large range that we cannot be sure of the pattern. However, the author believes that this is due to the very few data points.

There is one other interesting point to note about the data. On 4/21, Carnegie Mellon University was having their annual carnival just two blocks from the intersection where the data was gathered. If we calculate the standard deviation without the early afternoon data point, the 95% confidence interval is 0.6510-0.8163. Interestingly enough, the removed data point was outside the confidence interval at 0.5769. While we don't have enough data points to produce a true confidence interval, this does give hope that our technique will highlight days (like CMU carnival) where an unusual event causes drivers to behave differently.

Discussion

In our first paper [1], we considered the following topic:

> *Presume that cars traveling fast do cross the line even when the light is already red. We could count the cars crossing over the line to determine if more cars are speeding up to the light. This could signal an emergency that people are attempting to drive away from.*

The observation of a clear pattern means that we can count cars cross over stop lines as a reliable signal of "normalcy". In [4], observing the number of human faces on a street corner is used to detect an emergency. The concept is that if the current observation strays a statistically significant amount from the normal reading, something may be wrong. Either people are running away from something or flocking to something. Either way, the authorities would like to be notified right away. However, [4] counts human faces, and this is very difficult for a program to do. The software does not take profiles and partially hidden faces into account, and it tends to count objects (such as shoes) that are not faces.

---

[5] The author recognizes that this is not the best way to graph this information. A better graph would have been a floating bar graph to show the disjoint nature of the data. However, Excel would not allow this type of graph, so this was the next best option.

In this paper, we have provided a method that will show whether this pattern is statistically significant. Due to the low amount of data gathered, we cannot still be sure whether the data produces a reliable pattern. However, the method presented will determine this with more data. Additionally, the initial data implies that this will not only produce a pattern, but it will also show anomalous data.

References

[1] Cathedral of Learning Web Camera. University of Pittsburgh. http://www.discover.pitt.edu/tour/cl_cam.html

[2] Christopher, Ciera. Topics of Interest from a Simple Observation in Web Cameras. Data Privacy Course, Carnegie Mellon University. March 2006.

[3] Christopher, Ciera. Patterns from Stopped Traffic at Intersections. Data Privacy Course, Carnegie Mellon University. March 2006.

[4] Sweeney, L and Gross, R. Mining Images in Publicly-Available Cameras for Homeland Security. AAAI Spring Symposium on AI Technologies for Homeland Security. Palo Alto, 2005.