Jim Kerchenfaut
15-394
25 April 2006

Automated Detection and Correlation of Waiting Room Occupancies

Abstract:

A previous report (Bedford) briefly looked at trends in the occupancy of a single waiting room by observing the number of people in the room on a web-cam. In this study, we seek to expand on that analysis by again examining how busy the waiting room is at different times of day and asking how occupancy in different offices correlates to each other. Are they generally busy at the same times? Previous results (Bedford) show that there is possibly some correlation between the time of day and the number of people in the room. We examine data from an entire day at 6 waiting room locations to determine what conclusions, if any, can be drawn about correlation between different offices. We also examine some naive methods of estimating how busy the waiting room is in order to determine whether a computer can easily determine the number of people in the room.

Methods:

The Alaska Department of Motor Vehicles (DMV) has publicly available web-cams at each of 6 office locations: Eagle River, Fairbanks, Palmer, Soldotna, and 2 in Anchorage (Benson Boulevard and Fairbanks Street which are hereafter refered to as Benson and Anchorage respectively). See Figure 1. The Anchorage and Eagle River offices are open from 10am to 6pm PST Tuesday through Saturday. All of the other offices are open 8:30am to 4:30pm PST Monday through Friday. There is no reason given for the cameras to be there and no reason to belive that the people know they are being photographed. The web-cams theoretically record an image every 5 minutes. While examining the data we noticed that for some reason there were a couple instances where after five minutes the same image was still displayed. This could be due to any number of network problems or trouble with the camera. We also noticed that around the time of closing the cameras froze on the last recored image until the office opened again. Sometimes this image was taken while there were still customers in the office. There were also some other issues. The quality of the cameras is not very high and in the images from Palmer there is noticeable pixelation.This makes differentiating people more difficult. In some of the offices, particularly Soldotna, you can also see through the windows to outside of the office. This can cause errors because of peopl walking outside or changes in the outside scenery.

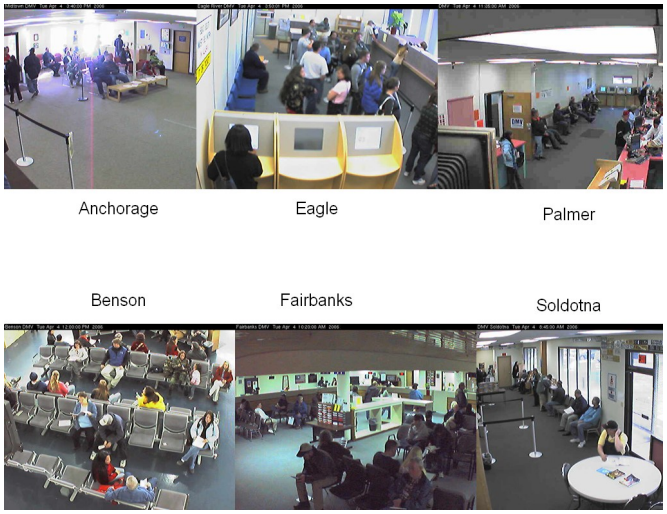

Figure 1: Typical images of all 6 waiting rooms

Using a Perl script, an image was automatically collected every 5 minutes from each of these cameras on Tuesday April 4th from 11:45 AM till 5 minutes past midnight EST. This should comprise the entire day of operations for all of the offices. There were some oddities in this dataset. For instance, in Anchorage there was an entire column of chairs that disappeared between the image the camera froze on the day before and when it started taking pictures that morning.

In order to do comparisons we used three different methods to determine how busy the waiting room was. First of all, we counted the number of people in each picture by hand. We encountered several problems using this method. It was often difficult to tell exactly how many people were in a picture. People were often partially obscured and conditions such as glare and poor resolution made differentiating people difficult. The criteria we used were to count a

person if they were in the photograph beyond a reasonable doubt and they were physically inside the waiting room. We did not count people who were visible through open doors or through the windows. We did count employees who were visible behind counters.

The second and third methods compared the photograph with a reference image by taking the sum of the absolute values of the differences of each pixel in the image. The images that we used as references for each office were an average of all of the images from that office and an image of the office when it was empty. The second method compared angainst an empty image, which was in all but one case chosen from the data gathered. See Figure 3. In the other case a suitable image was not found in the data so an empty image of the room that had been gathered previously was used. The average images for the third method were generated using Java software provided by Latanya Sweeney. See Figure 2. You can see that the average images are for the most part empty rooms with some differences such as the chairs being blury and the lighting being more diffuse.

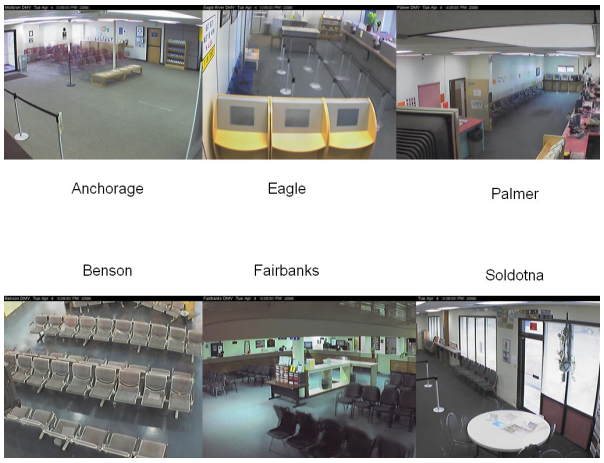

Figure 2: Average Images

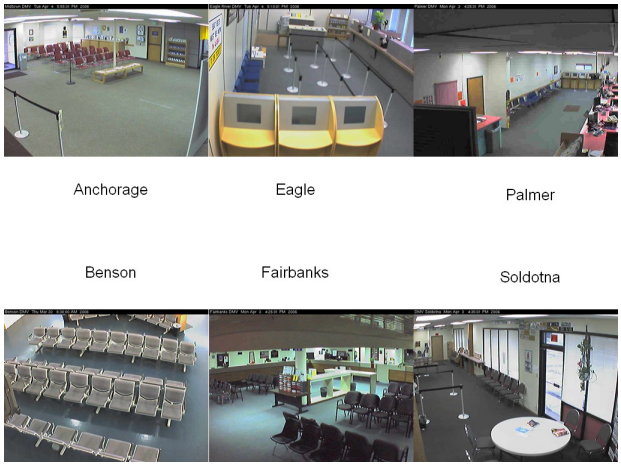

Figure 3: Empty Images

Results:

Comparing the hand counts to the difference images we notice that despite some significant issues some of the results map fairly well. See Figures 4 through 6. In particular, the Eagle River and Benson difference counts follow almost exactly the same curve as their hand counts. In the Palmer set, however the features found in the hand count curve, while they are somewhat subtle, are almost entirely lost in the difference curves.

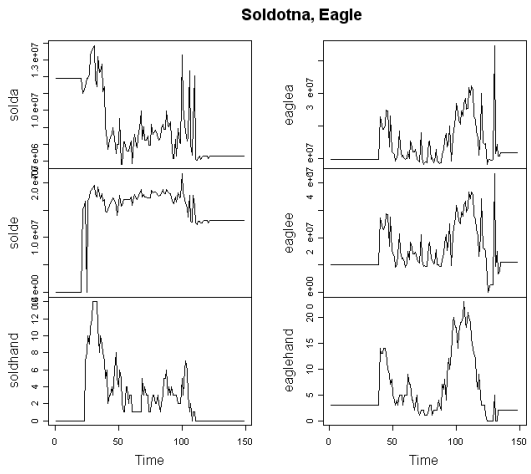

Figure 4: All counts for Soldotna and Eagle River

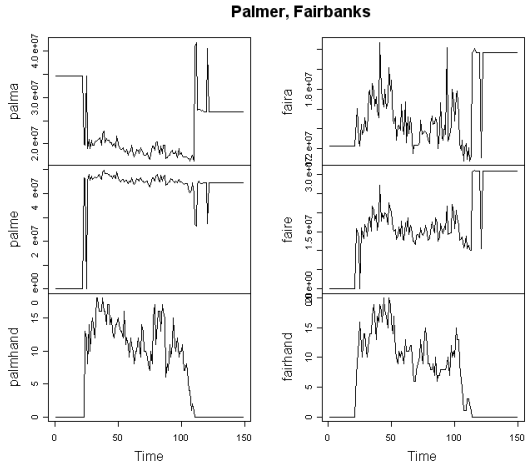

Figure 5: All counts for Palmer and Fairbanks

The bigest problem that we notice in the difference counts is due to the fact that the e value at the end of the day and the value at the beginning of the day do not match up and are in some cases not even close. This appears to mainly be due to the effects of different lighting in the morning and the evening. This prevents us from obtaining useful results in most statistical tests for correlation between the hand counts and the difference counts. The difference against the average image seems to perform better than the difference against the empty image in this respect. This problem,

however, makes it extremely difficult to equate zero people to any specific value of the difference score which in turn prevents us from having a baseline to use for converting the difference count into a count of people.

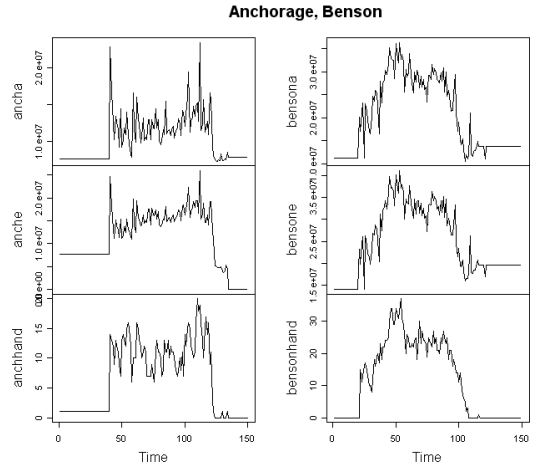

Figure 6: All counts for Anchorage and Benson

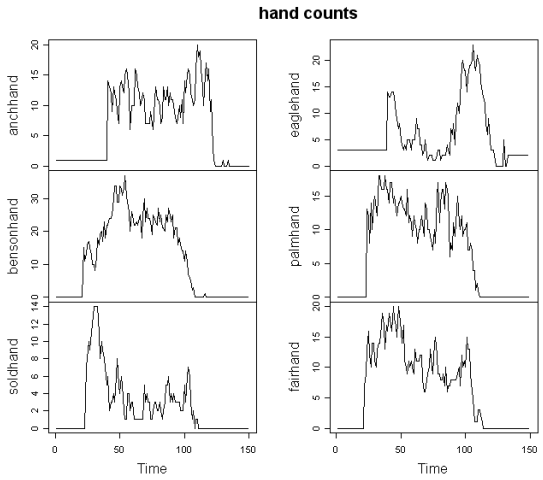

Figure 7: Hand counts for all offices

Looking at the hand counts we notice that the curves for Palmer and Fairbanks line up very well. See Figure 7. The correlation between other locations seems small at best. We do, however notice a similar trend in all but the Benson image. All of the others appear to be bimodal with peaks soon after opening and just before closing.

Discussion:

Overall the issues with using the difference counts probably eliminate any chance of reliably using them to quantitatively estimate the numbe of people in the room. They do, however, appear to in many cases follow the same trends as the actual number of people in the room as counted by hand. The factor which most throws them off this metric is simply variations in natural light coming from windows, but there are many waiting rooms without windows, including 3 in this study. Therefore, this naive measurement might be useful for determining things such as if a large crowd suddenly appeared or what times of day are the busiest on average even if it is not useful for directly determining the number of people in the room.

It appears from these results that the peak times for most DMV offices are right after each office opens and right before each office closes. This implies that it may be best to go in the middle of the day when there are fewer customers waiting. It does not, however, appear that there is any way to reliable determine the number of people in one office by simply looking at another office. It was interesting to note that the number of people in the Palmer and Fairbanks offices, which are seperated by over 300 miles lined up very closely while the two offices in Anchorage exibited very different patterns of counts.

All in all, this is good news for privacy enthusiasts. We have shown that to provide any better than rough estimates for the number of people in the room non-naive methods must be used. We have also shown that marked differences exist between traffic patterns at offices and therefore a person looking at a webcam of one office cannot necessarily make inferences cannot necessarily make assumptions about another office without more information.

Works Cited:

Bedford, Virginia “Use of Publicly Available Webcams in Naturalistic Observation Studies”. http://privacy.cs.cmu.edu/courses/dp1/refs/surveillance/samples/Bedford.pdf (3-30-06)