Tracking Bodies in Washington University's Red Square

Allison M Naaktgeboren
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15213
pinkrobot@cmu.edu

Abstract

Determining the number of people in a public space without violating privacy would be beneficial and yield numerous applications. Despite interesting prior work, the field has not been fully explored. For the case of Washington University's Red Square, images from a public webcam were analyzed to determine the number of people present in each frame using an algorithm that does not generate or use any personal identifiable information. Several methods were tested, and the best one was found to have an accuracy of .997, a 1-specificity of .001, and a sensitivity of .459. The algorithm has several limitations, and required a fair amount of manual inspection, which could be improved upon in further work.

Introduction

Although counting the number of people crossing a public space may seem like a trivial and superfluous task, it may have important implications. Knowing the average number of people present allows for the determination of deviations form the norm, which could be a signal of suspicious activity. Unusually fewer people in a public space may imply that a greater number of people are sick, an important concern for those monitoring for bioagents and bioweapons. People moving in unusually large numbers may also imply that something untoward is happening within a population, something that a large number of people find threatening. Unlike close surveillance, only watching

bodies from a distance would hide personal details, and hence privacy, while still providing the benefits associated with such monitoring, e.g. information on group behavior, size of the population at large, etc. Automated people counting in a public space would be helpful in all these cases.

Background

There has been recent interest in monitoring public spaces for characteristic behaviors of human populations. Webcams alone have resulted in many studies of diverse nature and purposes. Homeland Security and Public Health are interested in determining if an abnormal proportion of the population is ill in order to detect early outbreaks of bioweapons[2]. Those interested in data mining for business organizations have used a webcam to measure the use of a phone booth in a major metropolitan area[3]. Others have used webcam to study the behavior of people in waiting rooms, including the duration of the stay in the room[1]. Another researcher monitored consumption behaviors of a particular population through a webcam[4].

Methods

The operational definition of a body is any 11 by 20 pixel rectangle normal to the x-axis of the image, which contains an area that is more than sixty percent filled after the image has been processed. Any region that is not a body is defined as void space, even if the region is partially filled after the image has been processed.

The dimensions for the definition of a body were determined experimentally by randomly choosing ten randomly selected bodies from ten randomly selected images and

averaging the dimensions. The amount of area filled within that rectangle, the sixty percent, was chosen arbitrarily.

The final method of analyzing images consisted of five steps, harvesting, converting, processing, filtering, and body searching. The images were extracted every five minutes from the public webcam at http://www.washington.edu/cambots/ via a small script. However, images harvested from the page were gif files, difficult to analyze. Furthermore, more than fifty percent of the image is of the sky and buildings, regions of noninterest. A second script was used to convert and crop the images. The original images were 640 x 480 pixels in size, the region of interest, where humans can be found is 446 x 113 pixels.

The processing step was the most varied and was reworked several times during the development of the algorithm. The final form of processing marked regions more than two standard deviations above the mean of the image as filled, green color, and all regions below as non interest, not filled, colored black in the processed images.

The first processing algorithm searched only for dark regions. However, people also wear light colored clothing, and this search would miss these people. The second processing algorithm measured the statistical variance in the image. The third iteration averaged values over cursor squares, replaced that value in the top left pixel, and looked for high variation over the image. Cursor squares of ten by ten, five by five, and two by two, were tested. After that, thresholding each image against the image which occurred immediately before it, was tested. The next version simply looked for any pixel whose value was more than one standard deviation above the mean. The final version looked for any pixel whose value was greater than two standard deviations above the mean.

The filtering step occurred to remove regions of static noise, noise that occurred in all images. Static noise resulted from stairs on one building, the corner of another, two trees, and a white strip in the foreground. This filtering was performed manually on all images analyzed. Although automatic filtering was attempted, it produced problematic and inconsistent results, while manual filtration provided more even and consistent filtering at the expense of speed.

After filtering, the body searching phase commenced. Starting in the upper left hand corner, the image was systemically scanned for filled regions, regions of green. When green was detected, it was then checked to see if it filled more than 60 percent of the dimensions defined as a body. If it did, the region met the operational definition of a body, and was declared a body. Although automation of the body searching was attempted, the analysis was performed manually in all cases.

Since all the images contain much more nonbody space, manual determination of true negatives was difficult and prone to error. For example, two separate attempts on the sample image produced counts that differed by more than one hundred. Given the large number of true negatives, and the difficulty of counting them by hand, a mathematical method was used to determine the number of true negatives. For each image, the number of true positives, false positives, and false negatives was determined manually. The area of all these was determined and subtracted from the total area of the image, and the regions left have to be true negatives, and the number of bodies that would fit inside that area were calculated and reported as the number of true negative.

Results

The results of the first iterations of the processing algorithms were disappointing. The statistical variance algorithm produced the image pictured in figure 4, which contained no useful output for locating bodies in the image. The averaging square method, for square of ten, five, and two, produced the blank image depicted in figure 5, which also contained no useful information for body detection. Next, thresholding an image against the image taken immediately before, where immediately is defined as five minutes, was attempted, however this produced the results in figure 6, where were not deemed viable. The threshold algorithm, set at one standard deviation produced potentially viable results, however they were subject to a great deal of noise, as exemplified in figure 7. The best results of the set were obtained when the threshold was set to two standard deviations, as pictured in figure 8.

Regions that are often false negatives typically contain enough filled region that a human observer could guess that a body was present, but the algorithm would not see it because the filled region was left the required sixty percent.

Statistical analysis was performed to determine quantatively the utility of the algorithm. A contingency table was constructed based on a five percent random sampling. The table is pictured in figure 1. The accuracy of the final method was determined to be .997, a relatively high value. The 1-Specificity of the algorithm, or the rate of false positives, was calculated to be .001, an excellent value. The sensitivity of the algorithm was not as strong as its accuracy or 1-specificity; it was .459, indicating a propensity toward false negatives.

	Positive	Negative
True	108	6637

False	67	127

Fig.1 Contingency Table from a 5% random sample

Fig. 2 Original Image

Fig. 3 Image reduced to relevant area

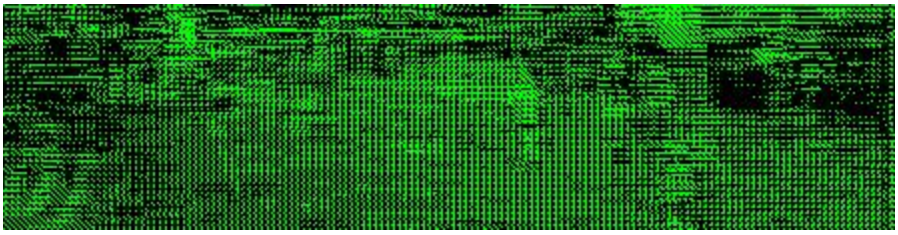

Fig. 4 Image from after application of the variance algorithm

Fig. 5 Image resulting from average square algorithm (same results for squares of 10, 5, and 2)

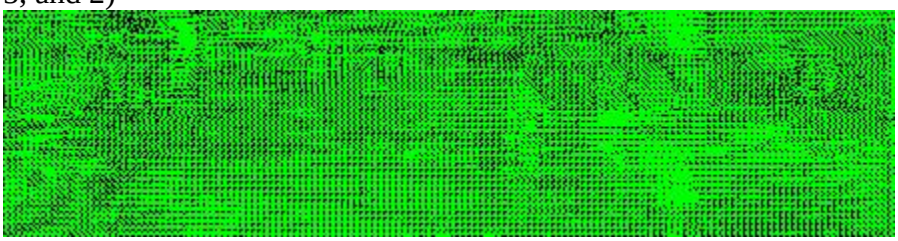

Fig. 6 Image resulting from thresholding against previous image

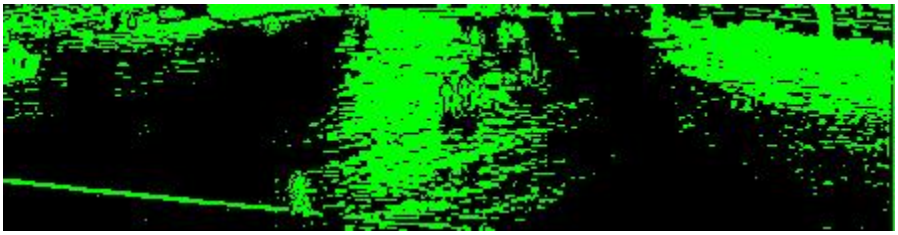

Fig. 7 Image after application of thresholding on one standard deviation

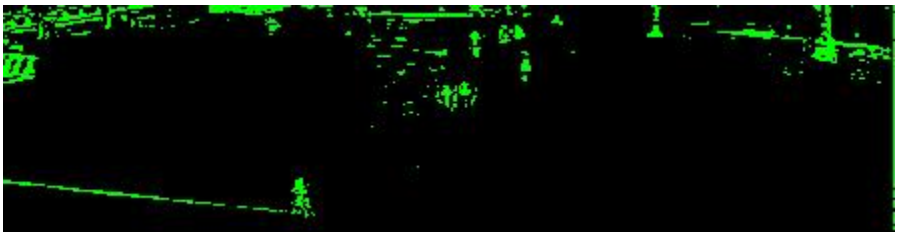

Fig. 8 Image after application of thresholding on two standard deviations, prior to filtering

Discussion

The final choice of algorithm is subject to several flaws and limitations. The images gathered after dusk or before dawn are too dark for either a human or a computer to discern to presence of people. Rain and other forms of precipitation can obscure the view of the square below. Puddles, even small ones, formed from excess water produce false positives. Precipitation also can land on the lens, as the camera is not protected from the elements. Droplets on the lens leave the images blurry for hours or sometimes days after the precipitation has ceased.

In addition to nature, other factors limit the applicability of the algorithm. Two or more people walking close together are often interpreted as one body instead of the correct two. People in the background are often not registered as bodies; most of the false negatives came from this case. Filtering out the stairs and building regions also causes anyone present on those structures to be filtered out as well, the price of filtering those regions. If a situation occurs when there are more bodies than ground visible in camera, if there is a rally for example, the algorithm will fail.

The positioning of the camera is also limiting. If the camera is not stationary, this algorithm will fail, and produce results of no significance or use. The requirement that the camera be stationary limits the general application of this algorithm. This also extends to streaming webcams, if this algorithm were run in real-time on streaming video, the movement detected would cause the algorithm to generate meaningless results. This could be circumvented by extracted stills from the camera, but that carries the potential for missing people not caught on those stills but still seen by the camera.

Although the size of the body has chosen from experimental data, the region that had to be filled to count as a body was chosen arbitrarily. Research into the optimal value of coverage would be helpful for future iterations of body finding. A body is always defined as perpendicular to the x axis, so would not be applicable to cameras in such situations where people might be diagonal, such as a waiting room, or if the camera is tilted.

The algorithm for locating bodies has room for a large degree of improvement. Much of the more important analysis is done manually. It would be vastly easier to analyze a large data set if the filtering of the static environmental variables, the stairs, the

trees, the stripe on the ground, were automatic. Searching the processed image looking for bodies was also done manually, and was prone to human error. Automation of this would make large or complex applications feasible. An excellent extension would be tracking groups of bodies. It would also be advantageous to know when it is raining, so the algorithm could be turned off, or a special one applied to deal with rain.

Conclusion

Determining the number of bodies present in a public space would yield data of interest to those studying patterns of behavior in a population, for example the department of Homeland Security. Different algorithms were tested before a final form was chosen for further analysis. The final algorithm for filtering and searching for bodies has good preliminary results, accuracy of .997, 1-specificity of .001, yet still has far to go before it can be used in large scale applications. Its sensitivity was also low, calculated as .459, and the algorithm is not effective at all during night hours, during precipitation or for periods of time following precipitation.

References

[1] V. Bedford, Use of Publicly Available Webcams in Naturalistic Observation Studies, Data Privacy and Technology, Carnegie Mellon University
[2] L. Sweeny, Surveillance Using Webcams, Data Privacy Laboratory, Carnegie Mellon University
[3] I. Fette, Analysis of Call Length and Placement Times, Data Privacy and Technology, Carnegie Mellon University
[4] E. Daimler, A Study of Food Consumption by Carnegie Mellon's Computer Science Department, Data Privacy and Technology, Carnegie Mellon University