Analysis of Customer Flow in a Coffee Shop

Manuel E. Sepúlveda
School of Computer Science
Carnegie Mellon University
msepulve@andrew.cmu.edu

ABSTRACT

As the cost of new technology decreases over time, many more people can afford high quality electronic devices. Webcams are being used in many ways: to have a video conference over the Internet, to build security systems, and to supervise employees' performance. Webcams can gather a large amount of data at relatively low cost. There are many publicly available webcams that transmit their video over the Internet. We used one webcam located at a coffee shop to estimate the times of the day that the store is busiest. This paper introduces an algorithm to perform this analysis automatically and presents the results from the manual and the automated observation. We concluded that the preferred period of the day to get coffee products is approximately from 10:00 am to 2:00 pm.

INTRODUCTION

The Coffee Gallery is a coffee shop in Haleiwa, HI that produces its products on site. On the store's web page, we can find four webcams that produce live streaming video from different spots of the coffee shop [1]. The cameras transmit video during the entire store business hours, which are from 7:00 am to 8:00 pm (Hawaiian time). For this paper, the camera located at the back of the counter will be used. This paper hypothesizes what time periods of the day the store has the most customers.

This hypothesis gives very important information to the store administrators. For example, they can specify the employees' break time to be at the time the store has fewer clients. Also, they can decide what times of the day are better to have special offers that increase their number of customers.

METHODS

The resolution of the cameras located at this coffee shop is 640x480 pixels. The observation was performed using a naturalistic observation approach [2] on April 6th, 2006 from 6:54 am to 2:32 pm and from 5:09 pm to 8:00 pm (Hawaiian time).

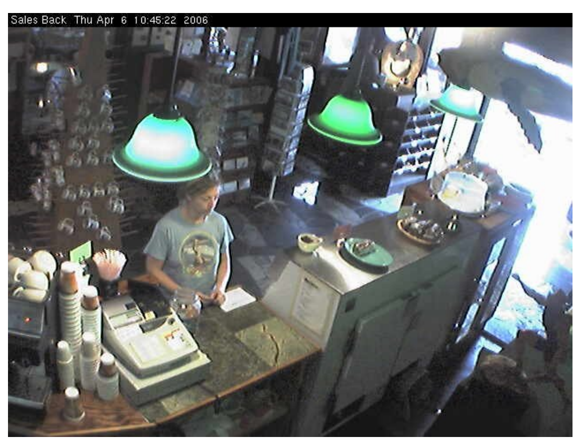

Fig. 1. Sample image from the back of the counter at the Coffee Gallery

For the first three hours of our observation, a two minute interval between each image was used. After that, we changed the capture rate to one image per minute. We ended up with 536 images.

A customer, in this paper, is defined as a person standing in front of the counter that pays the employee for a product. A region, showed in Figure 2, is defined as the area of the image that is processed by the algorithm. A background image is defined as an image where no customers appear on the region previously defined.

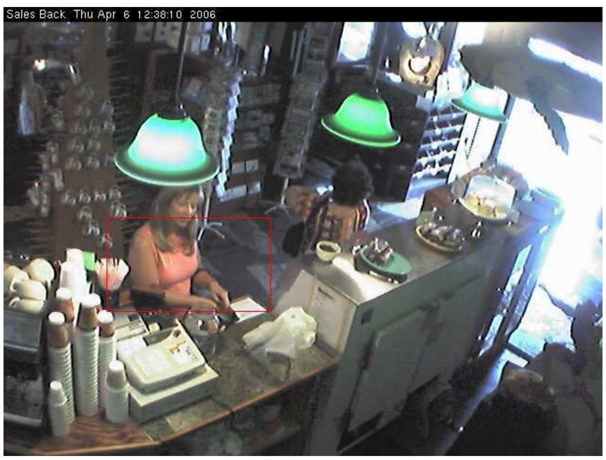

Fig. 2 The area defined by the red rectangle is the region analyzed by the algorithm

The proposed algorithm to perform this observation automatically is the following:

Let

- n be the number of sample images
- S_i be sample image i, such that $0 < i <= n$
- B be the average background image
- R be the analyzed region of the image
- $R(S_i) - R(B)$ be the subtraction of the value of every pixel in $R(S_i)$ minus the value of every pixel in R(B)

For each image S_i:

If $(R(S_i) - R(B))^2$ is greater than some threshold T_1

A customer is on the image

If $(R(S_i) - R(S_{i-1}))^2$ is greater than some threshold T_2

The customer on S_i is different from the one on S_{i-1}

If not

The customer is a new one

If not

No customer is on the image

This algorithm is linear in the number of sample images.

With this algorithm, we can determine basically two things: whether a customer appears on an image or not, and the number of different customers that the store has over a period of time.

This algorithm was implemented in Java and tested on the whole set of sample images.

RESULTS

From the manual analysis performed over the retrieved images, we can conclude that the busiest hours of this store are from 10:00 am to 2:00 pm approximately, and the average number of customers per hour is 20.1. Figure 3 shows the results from the manual and automated analysis.

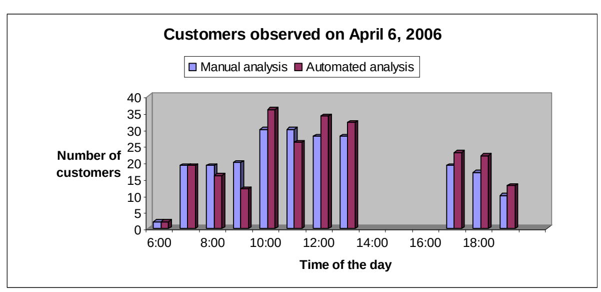

Fig. 3 Results from both analysis on the number of customers per hour observed on April 6, 2006

The application performed very well deciding if a customer appeared on a given image. The accuracy of the implemented version was 97.94%. Figure 4 shows the contingency table of this claim.

		Algorithm	
		NOT CUSTOMER	*CUSTOMER*
Truth	*NOT CUSTOMER*	167	10
	CUSTOMER	1	358

Fig. 4 Contingency table for concluding that a customer appears on an image

On the other hand, the algorithm deciding if a customer in a given image is the same as the one on the previous image didn't perform as well as the previous one. From the contingency table on figure 5, we can see that the implementation retrieved many false positives; that is, the algorithm was fooled and concluded that the costumer was a new one when it really wasn't. Likewise, the algorithm produced several false negatives. The accuracy of the implementation of this task was 75.13%.

		Algorithm	
		NOT NEW CUSTOMER	*NEW CUSTOMER*
Truth	*NOT NEW CUSTOMER*	85	51
	NEW CUSTOMER	38	184

Fig. 5 Contingency table for concluding that the customer on an image is different from the one on the previous image.

DISCUSSION

Many false positives in the first task were obtained because the sunlight hit the store in such a way that changed significantly the color of the region processed by the algorithm. On the other hand, the second task produced false positives every time a costumer appeared on the observed region in a given image k, but in image k+1 the customer wasn't on the region and in image k+2 the customer got back to the observed region; the algorithm concluded that this was a new customer. Likewise, if a costumer was accompanied by someone else, and the escort did not appear on the region in the next image, the algorithm also produced a false positive. False negatives were produced when different customers wore clothing of similar color.

Following work should look for a way to decrease the false positives and false negatives produced on the second task. A possible solution for this is finding a better threshold for deciding if there is a new customer on a given image.

REFERENCES

[1] Coffee Gallery webcams. *Barista Cam.* [online] http://www.roastmaster.com/cameras.htm [March 30, 2006]

[2] "Naturalistic observation". *Wikipedia, the free encyclopedia*. http://en.wikipedia.org/wiki/Naturalistic_observation