Demo: Tracking User Browsing on a Demo Floor

Aishwarya Ganesan
Microsoft Research India
t-aigane@microsoft.com

Swati Rallapalli
The University of Texas at Austin
swati@cs.utexas.edu

Krishna Kant Chintalapudi
Microsoft Research India
krchinta@microsoft.com

Venkata N. Padmanabhan
Microsoft Research India
padmanab@microsoft.com

Lili Qiu
The University of Texas at Austin
lili@cs.utexas.edu

Categories and Subject Descriptors

C.2.m [**Computer Systems Organization**]: COMPUTER - COMMUNICATION NETWORKS - Miscellaneous

Keywords

Physical Analytics, User Browsing, Wearables, Smart Glasses

1. TECHNICAL OVERVIEW

Our demo tracks physical browsing by users in indoor spaces. Analogous to online browsing, where users choose to go to certain webpages, dwell on a subset of pages of interest to them, and click on links of interest while ignoring others, we can draw parallels in the physical setting, where a user might *walk* purposefully to a section of interest, *dwell* there for a while, and *gaze* at specific items that they wish to know more about.

We have a developed a system called ThirdEye to track physical browsing by users in indoor spaces, such as retail stores. ThirdEye classifies the user's browsing into different modes, such as walking, dwelling, and gazing, analogous to the various modes of online (web) browsing. It further identifies which object and what type of object the user is gazing.

The design of ThirdEye centers on leveraging smart glasses worn by users as they browse, together with the smartphone carried by them. The front-facing camera on the smart glasses is in a position to capture the scene being viewed by a user. However, turning on the camera imposes a very significant drain on the battery, so ThirdEye uses the less expensive inertial sensors to the extent possible and triggers the camera only when the user is likely to be gazing. Once it is extablished that the user is gazing, images captured by the camera are passed to a cloud service to identify the objects in the scene. Below we briefly describe our approach. For more details on ThirdEye, including evaluation in actual retail stores, please see our paper [6] from Mobicom 2014.

1.1 Behavior Classification

ThirdEye automatically classifies shopper behavior into walking, dwelling, and gazing — using sensor data obtained from the shopper's smart glasses: inertial data obtained from the accelerometer, 3-axis compass, gyroscope sensors, and the video feed obtained from the camera sensor. ThirdEye also makes use of data from the inertial sensors on shoppers' smartphones, where available, as discussed later in this section.

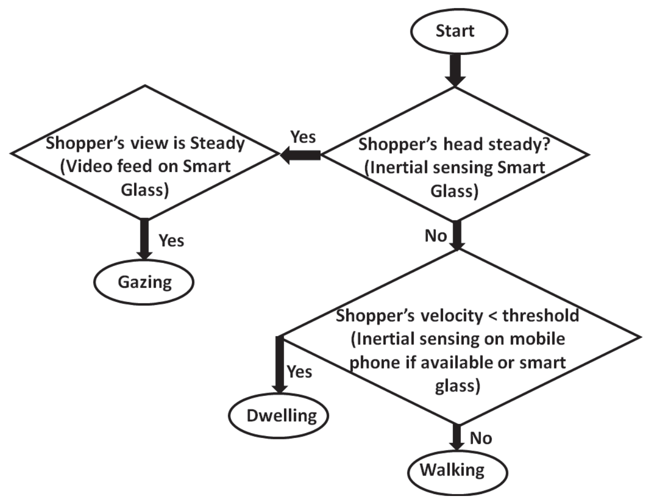

Figure 1: Overview of Behavior Classification Algorithm

Turning on the video on smart glasses causes a high power drain on the device, while inertial sensing results in much less power drain. Consequently, in our behavior classification scheme, we try to minimize the duration for which the video is turned on. Figure 1 depicts the high-level flow of our behavior classification scheme.

Gaze detection: The first step in our scheme is to detect whether or not the person is gazing. Since shoppers typically hold their head steady while gazing and are also not walking, this can be detected by measuring the standard deviation in inertial sensor measurements on the smart glass. However, just relying on inertial sensors leads to a large number of false positives. Consequently, we use a preliminary detection based on a low standard deviation in the inertial sensor measurements on the smart glass as a trigger to turn on the video and analyze the scene to confirm that the person is indeed gazing.

Dwelling and Walking: If there is a high standard deviation in the inertial sensors, this could mean that the person is either dwelling or walking. We use the step counter in Zee [5] to detect steps from the accelerometer. The accuracy of the step counter is higher when the accelerometer readings from the shopper's smartphone are used compared to when measurements from the smart glasses are used.

MobiCom'14, September 7–11, 2014, Maui, HI, USA.
ACM 978-1-4503-2783-1/14/09 ...$15.00.
http://dx.doi.org/10.1145/2639108.2641754.

This is because of spurious step detections caused by head movement. The step count is combined with compass data from the smart glasses to estimate the average velocity of the shopper. If the magnitude of average velocity is less than a certain threshold, the behavior is classified as dwelling, otherwise it is classified as walking.

1.2 Attention Identification

Once a user is identified as gazing, ThirdEye further determines which object in the field of view the user is gazing at and what is the object.

In general, the number of items within the field of view of the Glass could be large. In some cases, the number of items seen can be as large as 16. It is likely, however, that the shopper is only focused on one or perhaps a small number out of these items in his/her field of view. So our goal is to identify the shopper's focus within the field of view.

Our evaluation shows that a user's focus often does not lie at the center of the field of view of his/her smart glasses. So we develop the following scheme based on users' head-tilt.

Dependence of Attention on Head-Tilt: Our analysis revealed an interesting fact – *the center of attention in fact depends on the tilt of the head of the person.* In hindsight, this makes sense, since when a shopper looks up or down towards a product, he/she accomplishes this by partly turning head up/down and partly by rolling their eyeballs. Consequently, the shopper's focus within their field of view would tend to shift in the same direction as their head is tilting.

Probability Distributions for center of attention: Since it is in principle impossible to determine exactly which item the shopper is focused on, we resort to assigning a probability value $P(x, y|\delta)$ to an item in his field of view at the x and y pixel position, given the head-tilt, δ, as obtained from the compass:

$$P(x, y|\delta) = \Phi_X(x - x_o)\Phi_Y(y - m\delta - c) \qquad (1)$$

The distributions $\Phi_X(z)$ and $\Phi_Y(z)$ are obtained using histograms from actual data. Figure 2 depicts the function $P(x, y|\delta)$ as a function of $x - x_o$, and $y - m\delta - c$.

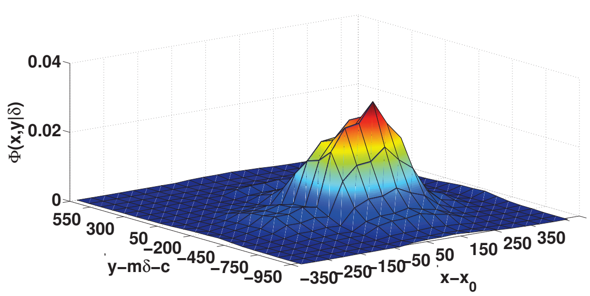

Figure 2: $\Phi(x, y|\delta)$

1.3 Object Recognition

ThirdEye captures images and invokes a cloud service to perform either reverse image search or image matching to identify the objects contained within the image. In our current implementation, we use Google's service [2], although ThirdEye is not tied to it and could also invoke alternative services, such as Bing Vision [1] and Nokia Point & Find [3].

Reverse image search involves object recognition, which in turn involves extracting image features (*e.g.*, using an algorithm such as SIFT [4]), comparing these with precomputed features from a large corpus of existing images to find a match, and retrieving meta-data associated with the matching image(s) in the corpus (*e.g.*, product name). The scale and complexity of this computation necessitates running it on the cloud (*e.g.*, the corpus of product images could run into the tens of millions). In turn, this calls for being selective in how frequently reverse image search requests are sent.

Therefore, only when gazing is detected, does ThirdEye capture images and invoke the cloud service. During each period of gazing, ThirdEye captures all the images during the gazing period.

2. PROPOSED DEMO

The proposed demo includes (i) behavior classification to identify if a user is dwelling, walking, and gazing, (ii) attention identification to detect which object the user is gazing at, and (iii) using reverse image search or image matching to recognize the type of the object. We have a few smart glasses, which can be given to visitors to wear. As a user walks around a demo floor wearing smart glasses, ThirdEye will track their mode of browsing: walking around the floor, dwelling in a particular area, or gazing. This output will be displayed on a monitor at the demo booth, so that it can be viewed by the visitors to the booth. We will also show the energy-saving techniques employed by ThirdEye in action, reporting when the camera is turned on or off.

3. EQUIPMENT NEEDS AND LOGISTICS

We will bring the laptops, smartphones, and smart glasses needed for the demo. We will need the conference to provide a large display to be placed in the demo booth to show the results of tracking visitors.

4. REFERENCES

[1] Bing vision. http://en.wikipedia.org/wiki/Bing_Vision.

[2] Google search by image. http://www.google.com/insidesearch/features/images/searchbyimage.html.

[3] Nokia point and find. https://betalabs.nokia.com/trials/nokia-point-and-find.

[4] D. Lowe. Object Recognition from Local Scale-Invariant Features. In *Proc. of ICCV*, 1999.

[5] A. Rai, K. Chintalapudi, V. N. Padmanabhan, and R. Sen. Zee: Zero-Effort Crowdsourcing for Indoor Localization. In *Proc. of MobiCom*, 2012.

[6] S. Rallapalli, A. Ganesan, K. Chintalapudi, V. N. Padmanabhan, and L. Qiu. Enabling Physical Analytics in Retail Stores Using Smart Glasses. In *Proc. of MobiCom*, 2014.