Enabling Physical Analytics in Retail Stores Using Smart Glasses

Swati Rallapalli*

Joint work with
Aishwarya Ganesan+, Krishna Chintalapudi+, Venkat Padmanabhan+, Lili Qiu*

*The University of Texas at Austin      +Microsoft Research India

ACM MobiCom 2014
September 8, 2014
Important to capture shopper behavior not only in the **online world** but also in the **physical world**

Even excluding those categories:

**Online sales in U.S. in 2013 only**

11%
Loyalty programs track in-store purchases... but...

In-store browsing?

Alice is interested in a product... but puts it back, does not buy it

Human behavior in the physical world:
✓ Contains signals missed by web-analytics & loyalty programs
✓ Indicates reinforced interest in buying
Understanding the intent of the shoppers in the physical world

Benefits

Alice sees a coupon as she is about to walk away!

Enable contextual recommendations

Shopping list reminders

Guides to product locations
Our approach

Enable a wide coverage and obtain rich user profiles!

Do not tie ourselves closely to any store..
Technology

Localization, Product layouts, User analytics

Incentives

Stores: increased sales
Physical analytics provider: share of profits by partnering with stores
Users: discounts, shopping

Privacy

Survey: Co
Mar 14, 2014 at 3:11pm EST
Shoppers Willing to Tell All
January 17, 2012, 9:00am EST

A new consumer indoor location consumer access overblown.

Teresa Novellino
Upstart Business Journal Entrepreneurs & Enterprises Editor
Email | Twitter

It might surprise retailers, but a new IBM study reveals that consumers are much more willing to give up information about themselves.
Overview: ThirdEye - AutoLayout

Key Idea:

Few Glass users
Glass “sees” products

Many Non-Glass users

**Offline:**

AutoLayout
Localize Glass users

Build product map
AP locations

**Online:**

Localize Non-Glass users
Contextual recommendations

Shopping list reminders

Guides to product locations

User Location

Product Locations

Offline

Crowd-source: No manual effort
Virtual coordinates

Access Point Parameters

Frequently changing Items in stock

Online

Leveraged to localize/analyze glass/non-glass

Offline Phase (AutoLayout)
Problem formulation: unknowns

$sLoc^i_k$: 2D location of $k^{th}$ shopper after $i^{th}$ step

$pLoc_j$: 2D location of $j^{th}$ product in store

$<P_l, aLoc_l, \gamma_l>$: parameters of $l^{th}$ access point

- Path loss constant
- Access point location
- Transmit power

Critical to localize non-glass users!
Problem formulation

**Minimize**

\[
\begin{align*}
\text{Minimize} & \quad \mathbf{w}_1 \cdot \mathbf{r}(s\text{Loc}, a\text{Loc}, P, \gamma) + \mathbf{w}_2 \cdot \mathbf{p}(s\text{Loc}, t) + \mathbf{w}_3 \cdot \mathbf{q}(s\text{Loc}, p\text{Loc}) \\
\text{Leveraged by [EZ]: ties shopper locations with} & \quad \text{AP parameters (AP location, AP Tx power, ..)} \\
\text{Wi-Fi term} & \quad \text{Inertial sensor term} \\
\text{Camera term} & \quad \\
\text{Log Distance Path Loss (LDPL) model:} & \quad \text{eRSS}(s\text{Loc}, a\text{Loc}, ap\text{TxPwr}, \gamma) = ap\text{TxPwr} - 10\gamma \log(\| s\text{Loc} - ap\text{Loc} \|)
\end{align*}
\]

Minimizes error in measured RSS values and those estimated by parameters describing the LDPL model

\[
\sum_l \sum_{k,i} \| m\text{RSS}_{l,k,i} - e\text{RSS}(s\text{Loc}_{k}^{i}, a\text{Loc}_{l}, P_{l}, \gamma_{l}) \|
\]

\(l^{th}\) AP Over all measured RSS values from that AP across all users
Incorporate mobility: inertial sensors

Minimize

\[ \sum_i \sum_k \| s_{Loc}^{i+1} - s_{Loc}^i - \hat{e}_k^i \|^2 \]

- For all shoppers, at all steps:
  - \( x_{i+1} \approx x_i + d \times \cos(\theta) \)
  - \( y_{i+1} \approx y_i + d \times \sin(\theta) \)

Accelerometer: step-count [Zee, UnLoc] \( \rightarrow \) distance

Compass: heading direction

Displacement between inferred locations at consecutive steps close to the estimate of displacement from the inertial sensors

\[ \hat{e}_k^i = [\cos \theta_k^i \sin \theta_k^i]^T \]
Tie in product locations: camera

Minimize

\[ \text{Wi-Fi term} \quad w_1 \cdot r(sLoc, aLoc, P, \gamma) + \text{Inertial sensor term} \quad w_2 \cdot p(sLoc, t) + \text{Camera term} \quad w_3 \cdot q(sLoc, pLoc) \]

Leverage Google Reverse Image search to obtain labels for product images

All shopper locations from where a particular product was seen must be close to each other

\[ \sum_j \sum_{<k,m> \in L_j} \| sLoc_k^m - pLoc_j \|^2 \]

\( j^{th} \) product \( k^{th} \) shopper saw \( j^{th} \) product at \( m^{th} \) step
Optimization

- Origin: \( sLoc^0_0 = (0,0) \)
- Leverage mobility: Locations within a walk are connected via inertial sensor data
- Gradient descent: refine initial estimates
- Run BFS to initialize all shopper & product locations
- Leverage product locations to align different walks

Milk seen by shopper 0 and shopper 2
Example walks around aisles in Target

After BFS: all tracks are in same coordinate system

After optimization: tracks look closer to actual walk
Inferred layout for H-E-B: improves with more shoppers
Overview: ThirdEye - User analytics

In a retail setting
Analyzed 3 hours of shopping videos from 7 shoppers wearing Google Glass from 2 large stores: H-E-B and Target

Time spent: dwell (50.7%) > gaze (23.7%) > walk (17.3%) > reach-out (8.2%)

- Most frequent inter-state transitions: dwell and gaze
- All states: tend to remain in same state for few seconds
Behavior classification algorithm

1. Start

2. Shopper’s head steady? (Inertial sensing)
   - Yes
   - No

3. Shopper’s view is Steady (Video)
   - Yes
   - No

4. Is there a hand near the items (Video)
   - Yes
     - Reach Out
   - No
     - Gaze

5. Shopper’s velocity < threshold (Inertial sensing)
   - Yes
     - Dwell
   - No
     - Walk
Gaze and Reach-out

When shopper is gazing/reaching-out scene in front of him does not change

Leverage vision based technique Optical Flow to detect gaze/reach-out

- Optical flow (of): difference in terms of pixels between consecutive images
- If \( of < of_{gaze} \) detect gaze/reach-out

88% detection rate at 1.8% false detections
Attention identification

A frame may contain as many as 16 items!

Important to identify the part of the frame that user was interested in!

- Estimated X: center of frame with offset for camera position
- Estimated Y: Function of head tilt

Accuracy of product identification: Top 1: 76%, Top 3: 90%
Behavior classification algorithm

Start

Shopper’s head steady? (Inertial sensing)

Yes

Shopper’s view is Steady (Video)

Yes

Is there a hand near the items (Video)

Yes

Reach Out

No

Gaze

No

Dwell

Yes

Shopper’s velocity < threshold (Inertial sensing)

No

Walk
Reach-out detection

Reach-out indicates high degree of interest: important to detect

Hand seen in the frame:
— detect hands to detect reach-out

Train TextonBoost Classifier
Leveraging TextonBoost classifier

Divided hand
Cluster together nearby segments

Spurious hand
Ignore very small segments

Detection success rate: 86% False detection rate: 15%
Behavior classification algorithm

Start

Shopper’s head steady? (Inertial sensing)

Yes

Shopper’s view is Steady (Video)

Yes

Is there a hand near the items (Video)

Yes

Reach Out

No

Gaze

No

Dwell

No

Shopper’s velocity < threshold (Inertial sensing)

Yes

No

Walk
Dwell detection

Accelerometer showing that user is static?

Shopper may not be static, he may take few steps looking at nearby items

Dwell characterized by small net displacement!

Detect dwell based on periods of low net displacement

Suppose K steps in time window $\tau$ and heading at step $i$ is $\theta_i$

— Detect steps using prior work Zee [MobiCom 2011]

Magnitude of net velocity vector

\[
\|\nu\| = \sqrt{\left(\sum_{i=1}^{K} \cos \theta_i\right)^2 + \left(\sum_{i=1}^{K} \sin \theta_i\right)^2}
\]

Net displacement in 5 sec

Dwell if:

\[
\|\nu\| < \|\nu\|_{dwell}
\]

95% detection rate at 10% false alarms
Related work

Research

— Localization [MobiSys12, MobiCom10, MobiCom12, ..]
— Vision [ICCV09, ICCV13, ..]
— Robotics [UbiComp09, ..]
— Human-activity sensing [Sensys07, ..]
— Shopping-behavior [UbiCom08, Pervasive Computing11, ..]

Industry

Get In-Store Notifications
Conclusion

Our contributions

— Fuse Wi-Fi, inertial sensor and video data from smart glasses
— **AutoLayout**: Map the store without any user or store input
— Use these inferences to track glass/non-glass users in online phase
— Characterize *walk, dwell, gaze* and *reaching-out* activities of shoppers
— Attention identification within the captured frame

Future work

— Larger data-set for patterns representative of more diverse population
— In-depth analytics of shoppers
Thank You!

swati@cs.utexas.edu