Designing a Vision-based Mobile Interface for In-store Shopping

Yan Xu[†] [‡] [*], Mirjana Spasojevic[‡], Jiang Gao[‡], Matthias Jacob[‡]

[†] Georgia Institute of Technology
85 fifth street NW
Atlanta, GA 30308, USA
yan.xu@cc.gatech.edu

[‡] Nokia Research Center, Palo Alto laboratory
955 Page Mill Road
Palo Alto, CA 94304-1003, USA
{yan.xu, mirjana.spasojevic, jiang.gao, matthias.jacob}@Nokia.com

ABSTRACT

Due to the situated nature of mobile applications, designing them requires more emphasis on users' cognitive load and interaction style. Considering that users can only devote limited and fragmented attention to mobile interface when moving between locations, what interaction-styles and services are appropriate to a specific user scenario? To explore this issue for in-store shopping, we designed a vision-based mobile interface for supporting shopper's communicational and organizational requirements on-the-go. With this interface, the physical objects can be automatically recognized by the camera phone in real time, so that shoppers can easily access related internet services. In this paper, we present an ethnographic study from which the design rationale is generated, and a formative evaluation to understand how mobile visual interface can be used in the field. The issues uncovered and lessons learned are applicable to our design improvement. Moreover, we use this to motivate the discussion on vision-based mobile interfaces in general, including embodied interaction and alternate interfaces.

Categories and Subject Descriptors

H5.2 [**Information interfaces and presentation**] User Interfaces—*Graphical user interfaces*.

General Terms

Design, Human Factors.

Keywords

Vision-based object recognition; Mobile computing; Filed study; Formative evaluation; Diary study

1. INTRODUCTION

Mobile applications are penetrating into our everyday life, changing human behaviors [7], and generating significant social and economic impacts [28]. Among these efforts, mobile applications for shopping come at the intersection of ubiquitous computing and electronic commerce, and are gaining research attention from both communities. While much of the prior work is to support transactions on mobile phones and information consumption on-the-go [19], the experiential aspect of shopping remain largely unsupported. However, should the consumerism culture be implicitly accepted as the value for designing mobile shopping applications? What's the design space for mobile shopping, in contrast to online shopping with desktop computers and in-store shopping? How can mobile device based interaction get embedded in the shopping experience? In this paper, we target to explore these questions.

To reexamine the design space for mobile shopping applications, we start from a fundamental cognition-level difference between mobile and desktop applications: the visual attention resource. As found by prior research and our own ethnographic study, the visual attention that users spend on mobile interface is often limited and fragmented due to the its situated nature [13, 16]. In the case of shopping scenarios, shoppers must frequently multitask in the environment of well-designed stores, moving themselves, navigating the space and objects, and easily getting interrupted. Consequently, mobile applications for in-store shopping need to be able to gain *just enough* attention, but not so much as to interfere with the shopping experience itself. Can contemporary mobile applications fulfill this requirement? With this question in mind, we conducted an ethnographic (a diary study and two in-depth interviews) to understand how mobile phones have been used in people's shopping experience, and generate design principles.

With the guidelines generated from the situated nature of mobile shopping application, we designed and built a prototype called Point&Find (see Figure 1). In this prototype, a vision-based mobile interface is introduced and the non-transactional functions of shopping are supported. Recent advances in mobile computing and computer vision raises promise for rapid object identification [6, 12], which provides the opportunity of directly connecting real-world objects with related online services. This kind of interface, sometimes referred to as mobile augmented reality, can be applied to many domains due to the absence of required infrastructure [8]. In the Point&Find prototype, this interface connects physical objects with the non-transactional

NordiCHI 2008: Using Bridges, 18-22 October, Lund, Sweden

[*] This work is done during the first author's internship in Nokia Research Center, Palo Alto Lab.

(communicational and organizational) services, which is different from much of the prior work's focus. This design choice is based on our understanding that shopping is not a "pin-point" activity (the activity that happens only in-the-moment.) Instead, it is a long term activity interwoven into people's daily life. This is also backed up by O'Hara and Perry's research, which found that experiential aspects of shopping "might be a more important part (than the in-the-moment transactional aspect) for mobile shopping services and applications" [24].

Figure 1: Point&Find, a vision-based mobile interface. The object is recognized in real-time, candidate results are ranked and listed below the view finder window.

Although the idea of connecting physical objects with digital information by detected context is not new [4, 5, 10], understanding how a vision-based mobile interface can be embedded in people's shopping activity has not been researched. We conducted a formative evaluation in a real-world store, from which we observed users' *active* interaction with the recognition results and the *emerging* patterns of attention switching between the virtual and real world. We also received preliminary yet promising feedback about the non-transactional functions designed in the interface. As a formative study, we are not taking a hypothesis-based approach, instead, open-ended and qualitative approaches contributes to the major findings. The goal of the field study is to shape the dimensions and issues that may be involved in mobile visual interfaces, and motivate the discussion about designing mobile applications that can be used in the field.

In the following sections, we present our process of designing, developing and evaluating a vision-based recognition prototype *Point&Find*. Our work extends the body of research on mobile interfaces and empirical studies of shopping experience with mobile applications in the following ways:

- We designed and implemented a mobile interface based on findings from ethnographic study and literatures;
- We conducted a formative evaluation in the field, and found how the mobile interface is embedded in shopping activities;
- We reported design issues and opportunities for similar mobile interfaces that use vision-based technology.

2. RELATED WORK

In their exploration of mobile HCI issues, Kristoffersen and Ljungberg reported that for the working context, mobile devices require so much attention that users need to "make place" to make it work [16]. Later, Pascoe et al. identified four specific characteristics of people using mobile technology at work: dynamic user configuration, limited attention capacity, high-speed interaction, and context dependency. In particular, they generated two design principles: the "Minimal Attention User Interface" and context awareness [27]. These two earlier works analyzed the issues for mobile HCI from the perspective of cognitive resource conflicts, shifting much of the research focus from the ever diminishing size of the hardware to ways for minimizing the attention resources required for mobile applications. Afterwards, Oulasvirta et al. recorded, coded and analyzed the attention duration and switch times for mobile web searching when users were moving between locations (streets, café, commute, escalator) [26]. The results show that the average duration of attention on mobile displays is 4-8 seconds. This research identified the significant competition for users' attention, which needs to be addressed when designing for the field. "Shorten interaction units" and "automate or eliminate tasks" [25] are the two of the design guidelines emerging from that research. While our study is motivated by Oulasvirta's findings, we take a more qualitative, situated approach. Our findings also show different attention distribution patterns for using mobile HCI for in-store shopping.

To assist shoppers' decision-making and reduce the cognitive effort for manipulating mobile devices on-the-go, various automatic recognition interfaces have been proposed. For example, in the early prototype of "Shopper's eye" [10], shopping suggestions were generated and transferred to shopper's mobile device basing on the match between the pre-recorded shopping list and real-time location. Later on, the prototype of Pocket BargainFinder brought forward the idea of bridging the gap between online and in-store shopping by providing web-based price comparison for bar-code detected products [4]. More shopping assistant applications were designed and developed in later research, such as smart shopping assistant [29], which used Radio Frequency Identification (RFID) sensors to transparently recognize and record users' actions in order to anticipate users' shopping plans. In these systems, the mobile device is used as a detector and collector of context information (location, object of interest, shopping plans), which triggers the access to related content. Our research prototype also shares this idea of triggering information access with real world objects or clues. However, our focus is on non-transactional and experiential aspects of shopping rather than to support in-the-moment decision making.

Much literature from sociology and cultural studies researched the experiential aspect of shopping both online and in physical stores. For example, Miller's ethnographic study of shopping in North London discusses how shopping indicated and shaped social relationships within families [20]. Wolifinbarger pointed how experiential aspects of shopping online emerge, and that shoppers had a substantially increased sense of freedom and control compared to offline shopping [35]. Most closely related to our work is O'Hara and Perry's study of user centered opportunities for supporting consumer behavior through ubiquitous computing [24]. Through a picture diary study and interviews, they examined reasons for buyers' deferring their "purchase impulses." Their data suggests that non-transactional/experiential values, as well as the strategic ways people organize and think about shopping, are at least as important as in-the-moment transactions.

As mentioned above, our research was motivated by recent advances in vision-based object recognition on camera phones. A substantial number of commercial applications are starting to use

the cameras embedded in mobile phones by computer vision technology. One successful example is the QR code, a matrix code created by the Japanese firm Denso-Wave in 1994. By 2005, about 30 million people in Japan were carrying QR code reading software, tucked inside their cell phones [11]. With a snapshot of a product's QR code, the information is decoded, and directs the phone's web browser to coupons, games or further product details. In another recently launched service, SnapTell [1], a user sends a picture of a product's textual name as it appears on its packaging, via MMS or email to the SnapTell servers, and then receives price and product details. Different from the QR code, this interaction doesn't require attaching specially printed symbols to the product, and the end users do not need to install any software to their mobile phones. But the disadvantage is that the feedback is not real time, which makes it less compelling in the mobile context. These applications have shown the potential of visual recognition based interfaces; specifically, automation of object identification, unobtrusive use of the device, and the pervasiveness of camera phones are the advantages of this interface.

Although many researchers have pointed out the importance of improving usability for mobile commerce [33], we found a lack of research on evaluation. Ngai and Gunasekaran's reviewed 149 papers on mobile commerce after 2000 [22], but few of the papers were about evaluating mobile commerce services in real world environments. Thus for our research, we have applied the evaluation methodologies from other related areas of mobile computing. Brush et al. conducted a 5-week long field study to understand how AURA, a bar-code based PDA application for shopping, is used in the field. They provided several design guidelines, including: ease of use of the interface, creating critical mass, and multiple data sources [5]. Another important piece of related work is a field study of a vision based interface for automated species identification [34]. Although this work was not done for the domain of mobile shopping, we share the similar goal of finding how visual recognition-based interface is used in the field. Their finding on how users inspect and compare the recognition results, how users interact with the algorithm to improve accuracy, and emerging system usage patterns resonate closely with the findings of our study.

3. BACKGROUND RESEARCH FOR DESIGN

In order to develop design guidelines for our prototype mobile interfaces, we first ran an ethnographic study of how mobile devices are being used in shopping scenarios.

3.1 Understanding How Mobile Phones are Being Used for Shopping

We conducted a one-month long diary study, including two rounds of one-on-one in-depth interviews with 12 participants (9 female), ranging in age from 21 to 42. We recruited participants from Craigslist.org and through the social networks of our co-workers. All of the participants were from California, USA. None of the participants were employed by the IT industry. Participants were screened to include those that shopped at least twice a week, including both online and in real world store, and excluding shopping for everyday items such as groceries. Six participants had internet access on their mobile phone, and three had unlimited data plans; all but one of the participants used camera phones.

The goal of the diary study was to capture specific instances of shopping activities, including full context and motivation, and to understand any barriers to the use of mobile phones in those situations. Before the study, we conducted an interview to understand general attitudes, patterns and preferences related to shopping. During the diary study, participants used an on-line diary tool for answering questions about their shopping activity, and they are encouraged to use their mobile phones to capture "snippets of information" (text, images, audio) as contextual information [3]. Over the course of the study we collected 128 entries, of which 58 had pictures. This data was used as the basis for subsequent in-depth interviews, which were focused on the reflections of, and motivations for, the recorded shopping experiences. Below we report some of our findings that are closely related to mobile applications for shopping:

Categories of mobile phone usage

The study revealed four categories of roles that mobile phones play in shopping (Table 1). Communication and organization were the two most frequently mentioned categories. Informational and transactional uses were tried by early adopters who had internet access on their phones, but with less frequency.

Table 1. Categories of usage of mobile phone for shopping

Communication	• Taking and keeping pictures on the phone for later discussion, including gift ideas, opinion-seeking, or for fun; • Taking and sending pictures to friends or families to get opinions or for fun; • Calling families or friends to find out what to buy; • Chatting with friends when waiting in line;
Organizational	• Remembering product requirements (size, other's preference, gift idea) for purchase; • Recording the location and phone number of the store; • Reminding of the timing for purchase (usually birthday);
Informative	• Monitoring bidding on eBay (for 2 intensive eBay users); • Searching for prices in the store; • Receiving local ads and promotional info by text message (for 1 participant when she lived in Norway)
Transactional	• Purchasing by mobile phone (for small-ticket items).

As found in our study, the pictures taken by the camera phones were frequently used as a common ground of communication – a finding backed up by previous research on the usage of camera phones [15]. One participant also talked about her choice between two channels for sending pictures: MMS and email. She preferred email because she believed that it was more "polite" in the sense that email was less of an interruption and less obligatory to view or provide a response.

We also found that users were facing the increased memory demands for in-store shopping. "Online research" before visiting a store, was a frequently mentioned activity, especially for electronic products and large-ticket items. As mentioned by our participants, online research was done in order to "know the price range", "get a coupon or discount", or "find local stores". While this kind of research increased users' feelings of "confidence" and "control" when walking into a store, it also placed increased memory loads. But none of the participants reported to record this kind of additional information with a mobile device. Instead, we found some cases in which the mobile device is used to record the size, preference, date of birthday and gift idea for their families and close friends. We suspect this is due to the difficulty of transferring information from desktop computers to mobile devices. Since only two of our participants had full QWERTY keyboards on their phones, the task of manually transferring data from the computer to the phone would often be impractical.

Concerns about using mobile internet in shopping

We asked participants whether they would use the internet on their mobile phones to make, or assist with making, purchases (via the phone's built-in web browser). We found participants expressed several issues with doing so. First, there were concerns about the time required to find the desired content – especially given the limited time frame when moving in and between stores. In particular, participants commented on the difficulty of performing searching/navigating operations, as well as the depth of the web page hierarchy. Even basic actions entailed a significant burden, such as starting the mobile web browser and switching between pages. Additional impediments included the use of the keypad for entering data, the small size of the display, and service costs. These factors led people to find alternative resources, such as real world clues or staff, to be more desirable.

Below is an example from a 34 year old female participant who was buying a gift for her boyfriend during her lunch hour. She was trying to find out the meaning of some symbols written on the product. She went back to her office to investigate the product, and she later returned to the store and purchased the product after work. We inquired why she didn't try to find the information in the store by using her phone, especially since she had an unlimited data plan:

"…(I would use the internet,) if I had a larger screen (on the phone). But still I have to find xxx.com (the website of the store) first of all, and then, I have to find if it was under a man's or woman's section; I've never been to their website before; is it under accessories or jewelries? Do they even have a jewelry category? Being that I was in a rush on lunch hour, on a little phone, I don't think it will work… If the retailers have some capability that you can Bluetooth in, or something in the air, you can easily connect to, taking out the trouble steps of logging into the Internet, finding the website and trying to find which page you need, then I will probably use it…. But I won't pay for it (the service), me, use; they (the stores), pay… (Laugh)"

From this description, we found that one user may have multiple concerns in a single use case, which caused our participant to avoid using internet on her mobile phone in the store, even though she had the motivation of an information query at that time. Moreover, we note that many of the difficulties the user describes come from her envisioning herself using the mobile phone in the same manner as desktop computers, which suggests to us that a different interaction style would be necessary on a mobile device.

3.2 Design Principles

Stemming from our diary study and related research we formulated these design principles:

- Mobile shopping applications should not demand a high, continuous level of attention [31];
- Shopping is not a discrete, short-lived "pin-point" activity [24], hence the portability of the mobile phone platform, and its connection to other resources, need to support the shopping activity over a longer term and across different locations;
- Mobile shopping applications need to simplify access to product information as compared to their desktop counterparts; that is, the traditional searching and browsing model of desktop internet usage does not fit well in field contexts;
- Mobile application needs to provide highly relevant content: relevance can be achieved by using context information, such as location, object of interest, time, and personal preference.
- Mobile applications need to focus more on supporting organizational and communicational functions, as these are the common usage of mobile devices in shopping currently.
- The online and in-store shopping experiences need to be bridged. We must minimize the amount of information the user must remember when switching between the two contexts.

In summary, mobile applications need to deal with the physical and social context in which they are embedded. Reducing the input and output effort of using the system and minimizing the cognitive load of shopping tasks (e.g. memory load) are two of the fundamental guidelines for designing mobile interface.

4. SYSTEM AND INTERFACE

Our prototype system consists of three parts: the visual matching engine, which runs as a background program and keeps updating the image matching results; object-recognition based interface, the main interface that shows the recognition results with the viewfinder; and the internet services, including the non-transactional functions for the selected product.

4.1 Visual Matching Engine

We implemented an intelligent image search engine on smartphones running the Symbian S60 operating system. The visual matching engine [12] is configured to perform image matching between a query image and a set of images within a prerecorded database. The results, a set of matching images, are generated by comparing multiple corresponding image features. The visual matching system is designed to run efficiently on a mobile phone at the rate of three to five frames per second, allowing image matching to be performed on a live video feed from a phone camera – thus providing continuous feedback to users. We compare the images using a combination of histograms of pre-defined image features which are known to be among the best for Content-based Image Retrieval (CBIR) [30].

Figure 2 shows performance of the current image matching engine, when applied to poster images, and street/office scene

pictures. The test was done using datasets collected from both outdoor and indoor environments.

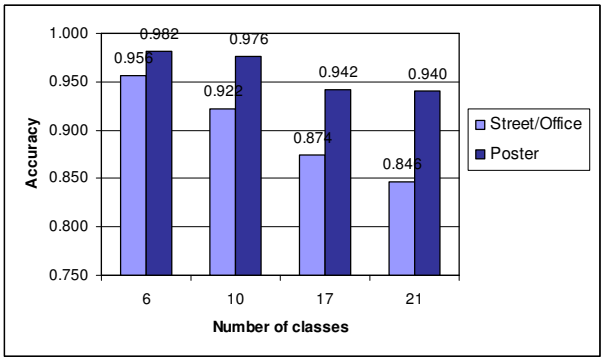

Figure 2. The relationship between number of classes and accuracy

Image matching accuracy depends on the number of classes the query image needs to match with. Class means the unique object that can to be recognized by the system. The accuracy also depends on different types of visual objects, and how the images were taken. For example, the poster images normally get higher recognition accuracy, due to their two dimensional structure and well textured regions. For more general street scene and office pictures, however, the problems of large variation of view angles, three dimensional structures and blurred images poses much greater challenges. These will serve to explain the different image classification performances between posters and outdoor environments.

Our current algorithm has the limitation on scalability. However, there is much research about using contextual information (such as location) to enable the scalability and assist vision-based algorithms [8, 12]. We agree that the large-scale usage of this application is an interesting research topic. But in this paper, we focus on the lower level of interaction that happens between an individual user and the mobile interface.

4.2 Object Recognition-based Interface

The components of the mobile mixed interface are shown in Figure 3. The phone's camera viewfinder is on the top of the screen and the search results are on the bottom. When the user focuses the viewfinder on an object whose image is already in the database, the frame around the viewfinder turns blue. The deeper the blue color is, the higher the confidence level is. This feedback is visible to the user within 1 or 2 seconds after she has the object in focus. In addition, real-time feedback of updated results (model name and price) is listed in the results section, and ranked by the confidence level. A blue selection marker indicates the currently selected label, and when the user presses the navigation button on it, a webpage with associated content is displayed. This interface distinguishes itself by the real-time feedback, compared to other object detection techniques, such as QR code [11] and SnapTell [1]. More importantly, with the video based interface, users can choose to drill down to the information/service of the specific object they want, or stay on the level of browsing different products. And users can always jump back to the main recognition-based interface anytime by clicking one button.

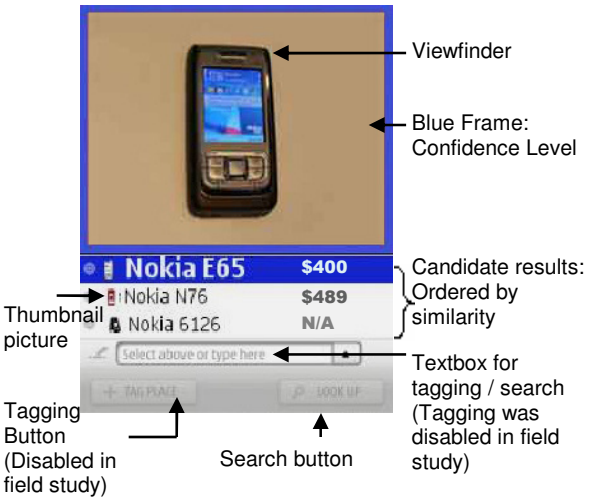

Figure 3. The screen shot from the mobile application

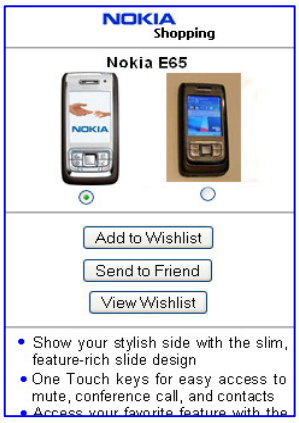

Figure 4. The main web page for a product. The left image is a stock photo; the right one is the image just taken by the camera phone. Three services are listed below the images as buttons.

4.3 Non-transactional Functions

After selecting a label, the user is directed to an associated service HTML page in the mobile web browser (Figure 4). Following our design implications, we selected only three out of over 20 potential services generated from our ethnographic study, trying to cover the important aspects of shopping:

- Organizational – *"wish list"*, for bookmarking the object when the shopping impulse was generated and for later research; this function considers the gap between "intention" and "action", and taking shopping as a long-term activity related to other parts of everyday life;
- Communicational – *"send to friend"*, for communicating with friends or families by email. Users can choose to send the image they took with their phone in the store or the stock picture of the product.
- Informative – *"specifications and link to a more detailed information page"*.

These simple yet representative functions are designed for supporting non-transactional values by mobile computing. We followed the "simple and shallow" design principle for mobile applications [18], and used these functions to explore the space of designing mobile interfaces for non-transactional aspects of in-store shopping.

5. FORMATIVE EVALUATION IN THE FIELD

We conduct a formative evaluation in the field to understand how the vision-based mobile interface is embedded into the activity of shopping, and whether the experience of shopping is augmented. We believe that some of the discovered issues and learned lessons will carry over to other vision-based mobile interfaces. Specifically, we target to answer these questions from the study:

- How is vision-based mobile interface incorporated in the activity of shopping?
- How do the users cope with errors and failures of visual recognition results?
- Where do shoppers direct their attention while using the application during a shopping experience?
- What is user's feedback on the organizational and communicational functions?

5.1 Choice of Field Study

We chose an in-situ field study instead of a laboratory experiment because it can find more usability issues, especially on cognitive load and interaction style [23]. In this study, participants were recruited from shoppers who walked into the store. There were no pre-defined task sets for participants to perform. We considered two reasons for this choice: *First*, the store environment is hard to be simulated or recreated in laboratories. And the real environment is very important, as indicated in Brush et al.'s user study for a bar-code reader based mobile system. They found that the field study highlighted several places that their "seemingly reasonable design assumption" did not match with real usage [5]. *Second*, we planned to recruit "real" shoppers who walked into the store with specific motivations. At the same time, they were under realistic time constraints for store visiting.

As a consequence, we faced the similar difficulties as other in-situ studies for mobile applications, such as data collection processes, instrumentation and control of extraneous influences [21]. Before the study, we did a few pilot runs, and found that the average usage time for this application was about several minutes, which was reasonable because the average visit time to this small store was around 5 minutes according to the staff. However, the small amount of usage time makes the logged quantitative data have little statistical power. Also, we found that it was hard to require more than half an hour of our participants, which led to a reduction to the duration of post shopping survey and interviews.

5.2 Description of the Environment

We conduct the field study in Nokia Experience store (see Figure 5.a) in the Stonestown shopping mall in San Francisco, USA. The purpose of this store was to encourage people to walk in, experience the products and ask questions. This store didn't actually sell any of the products on display, although they can be purchased elsewhere in the same mall. In total, 26 products were displayed in the store, including mobile phones and their accessories (see Fig.5.b). While logistical concerns played into our decision to select this store, it actually turns out that mobile phones are a good representative of the general consumer electronic product category, for which consumers typically conduct extensive research before committing to a purchase.

Figure 5. (a). the in-store environment, with one customer playing with one of the products with one hand; (b). the products are displayed in a row on the table; (c). details of a product display, the red dotted rectangle marks the label for the model's name.

5.3 Participants and the Procedure

The participants were recruited from the customers who walked into the store and agreed to try our application prototype after being informed about the purpose. In total, 17 participants were recruited. Their age ranged from 13 to 52, with the average of 27.4. 5 participants were female; 7 were Nokia phone users; 14 used Internet everyday for more than 2 hours. The purposes for their visit to the store varied, including: "research for a purchase", "having fun", "window shopping", "keeping updated", "randomly walking in", "asking help from staff".

After the shoppers agreed to join the study, we handed them the Nokia N95 mobile phone with the application to them, explained how the application worked and showed how to point the camera phone to the product. Then we asked them to try it out themselves, and answered their questions during the trial. Most participants understood how to use the system immediately, while some of them had questions about scrolling through a web page by using navigation keys. After that, participants were asked to keep the device and use the application during their visit and spend as much time as they wanted in the store. One researcher observed them. A survey and a semi-structured interview were conducted afterwards and lasted about 15-30 minutes. At the end of the study, participants were thanked and given a $10 Starbucks card.

5.4 Results and Discussion

From the observations, survey results, and interviews, we found that the general feedback to the application was positive. Figure 6 showed the subjective evaluation of the ease of use, enjoyment and shopping experience change in a 1-5 likert scale.

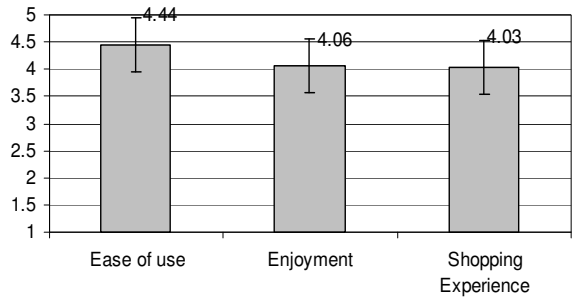

Figure 6. Ease of use (1-difficult, 5-easy); Enjoyment (1- boring, 5-enjoyable); Shopping experience (1-disturbing, 5-augmenting)

Three of the participants held neutral opinion about whether their experience was augmented or interfered. And one of the participants thought his shopping experience was interfered. Their major comment was: the function can partially be replaced by alternative resources, such as staff or in-store clue, and not willing to rely on a "high-tech" application. But some other participants prefer the independent access to information/service as an advantage for reducing unwanted interaction with staff to avoid the "sales pressure" (even though in this particular case there was none).

Beyond the general feedback, below we discuss the answers to the outlined research questions, as supported mostly by the qualitative data from the user study.

5.4.1 Dealing with Vision-based Mobile Interface

- **Automated object recognition is easy to learn and use**

As noted before, the participants reported that this vision-based mobile interface was easy to learn and easy to use (see Fig. 6). All the participants got the idea of "pointing" and then "clicking" real-world objects almost immediately. After knowing how the interface could work, some of the participants were curious about which objects can be recognized by the prototype, and tried it over many objects in the store, some of which are not included in our image database, such as plants, posters and even a friend's face. This kind of behavior, spontaneously browsing real-world object by camera phones, highlighted the benefits of automated object recognition, as well as the real time update of the recognition result. It significantly reduced the effort to access related services. Nine of the participants mentioned that "immediate information accessibility" was what they like about the prototype.

As found in prior study, vision-based recognition interface may introduce usability issues that the users do not know what objects are recognizable by the systems [8]. Although our interface tries to create "point and click" experience, only a few objects can be interacted with. In the user study, it did not become an issue. Users quickly learned that all the candidate search results shown on the interface are the products in the store. We attribute this to the match between objects of interest for shoppers and the pre-stored object images from the system.

- **Users actively interact with recognition results**

We observed that participants actively interacted with recognition results very often. Figure 7 showed examples of how the participants were holding the camera phone in their hand(s). As described in the previous section, the interface listed three candidate results, ranked by the confidence level. When the correct result does not appear on the top, participants adjust their camera phone position by changing the viewing angle or zooming in and out until they got the right result listed as the top choice. Only two participants used the navigation key to scroll down and select the second or third options. Above observation implied the potential of leveraging user interaction to assist object recognition. It also showed that motion-based interaction, rather than keypad based interaction, fits better with vision-based mobile interface.

Another common behavior we observed was that, all of the participants checked the physical clue (the label placed in front of the product, see Fig.5.c) before they clicked the label for a product. This confirmation behavior caused users' attention to switch back and forth between the mobile display and the physical environment. It indicated the existence of a "gate" between physical world to digital world, and real-world clue is an important input into the process of interaction.

(a) **(b)** **(c)**

Figure 7: (a) and (b). two participants were holding the device (within read frame) and product in each hand; (c). a female participant was holding the phone by two hands and try to reduce the jittering of labels by holding the phone more steadily.

- **Lessons: coping with the error**

In Abowd and Mynatt's ubiquitous computing review paper [2], they pointed out that eliminating errors for recognition-based system may not be possible, while on the other hand, recognition accuracy is not necessarily the determinant factor for user satisfaction. As shown on the accuracy chart in previous section (see Figure 2), our algorithm is not perfect either. To assist the algorithm, we designed a blue frame (see Figure 3) around the viewfinder to indicate the confidence level of recognition result. However, none of the participants noticed and actually used this feature. The failure of this design might be caused by the nuance of the clue.

The most common critique we received about current prototype was the "flipping label" problem. What caused this problem? Our interface updated recognition results three to five times per second. And for each update, the lighting and slight motion may cause the ranking of the candidate results change. The sensitivity to subtle changes in environment is a common issue for many other

computer vision-based interfaces. We found the severity of this issue in the field study, in which the environmental lighting change may happen more frequently than laboratories.

This problem shows that, it is not always true that the higher recognition speed is, the better the user interface will be, especially for vision-based interfaces, in which jittery has been a problem that can hardly be controlled. One solution is to reduce the speed for updating object recognition results, so that users have enough time for interaction. We believe that a tuning process is required to trade-off the update speed and the stability of user interface.

5.4.2 Distributed Attention between the Physical and Virtual

- **User's attention is redistributed after introducing the mobile interface**

To understand how user's attention is redistributed after introducing the artifact of our prototype, we first tried to understand normally how shoppers behave in a store, by talking to the staff and observing people. Shoppers usually started from browsing the products. Once they got interested in a particular model, they would start touching and playing with it. They might ask the store staff questions for knowing the price and feature of the product. Some of them came to the store with certain products in mind, and they directly approached them or found them by asking the staff, and then initiated the exploration as mentioned above. During this process, the products occupied most of the visual and tactile attention of the user.

After we introduced the mobile interface, it redistributes users' attention resource. The mobile device usually occupied one hand of the user, except p14 (see Fig. 7.c), who tried to hold the device steady with two hands, because she thought that the "flipping problem" could be overcome by doing so. One of the users put away the device in his pocket temporally after using the mobile interface, and freed two hands for touching and playing with the products. While our prototype took away much tactile attention from the product, it also changed visual attention distribution. How deep was the user immersed in the interface? How the visual attentions switch between the virtual and the real? The findings are as follows:

- **User's different immersion level to the interface**

Participants showed different immersion levels on the mobile interface. Five of the participants browsed over several products by our prototype, without drilling down to the product page. Their attention spent on the digital content of each product was no more than a few seconds. For them, the camera phone screen was used as a *lens* to view a video of the real world objects with a little digital information (price and name) attached. Thirteen participants clicked the label of product for more than once and reading the specifications. One participant did both browsing and clicking in his trial. The difference of the immersion level is correlated with the different intentions that people come to the store with (binary logistic regression analysis, $R=0.48;\ p=0.05$): Among 9 of the participants who claimed to be researching for purchase, 8 were spending several minutes on using the functions for specific products.

- **Emerging patterns of attention switch**

Three different patterns of attention switch were emerged from our study, in terms of the navigation between the physical and the digital (see Fig. 8):

Browsers (five participants) – they used the camera phone to browse over several products and only read the label text, never going deeper into content; they preferred touching and feeling the products.

Frequent switchers (seven participants) – they switched their attention between the real object and digital services one by one; they spent time on both the services we provided and trying the products. After they finished exploring one product, they would move to the next target.

Immersed researchers (five participants) – they separated the shopping experience into two phases: "research" and "play"; they spend a few minutes on the using the interface on multiple products; and then touch and play with them; or vise versa.

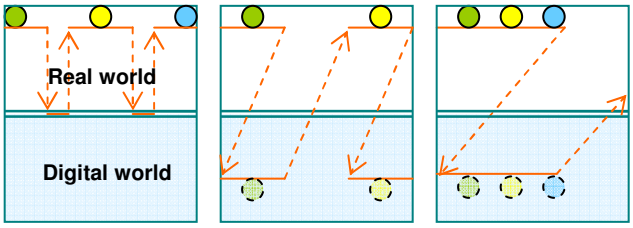

Figure 8: Three models for shoppers' attention switching between physical and digital spaces. (left).Browsers; (center).Frequent switchers; (right),Immersed researchers

Note: colorful balls with solid outline represent products; those with dotted outline are the digital information/services about the product; orange solid lines show duration of attention; orange dotted lines represent the switch of attention.

The diversity of attention switching patterns showed the nature of in-situ actions, which had no fixed task flow and highly depended on the context in which the actions happened. Moreover, these patterns demonstrated how users settled with the newly introduced artifact and embedded it into their experience, while they still kept the part of touching and feeling the physical objects. Compared to the findings from Oulasvirta's experimental study about attention for mobile HCI [26], the duration of the attention on our mobile HCI was longer, and frequency of attention switching between interface and environment was lower. We analyze that there were two reasons for the difference: *the physical environment*, and *user intentions*. The store environment required less navigation on the environment than streets, escalators or buses. Stores were usually designed to be pleasant and friendly. Moreover, in Oulasvirta's study, the tasks on the mobile phone and in the physical world was unrelated to the participants; while in our case, the mobile task was triggered by the interest toward physical objects. It showed the importance of overlapping goals between physical and digital tasks, which is the space that mobile computing can play an important role in.

5.4.3 Feedback for non-transactional functions

From the field study, we received preliminary but promising feedback about the non-transactional functions, specifically on organizational and communication aspects.

- **Feedback for "add to wishlist" function**

Eight of our participants mentioned the wish list function as their favorable feature of the prototype. They liked this function for multiple reasons. Below are some examples:

"The wish list is useful for me. I usually take time to make sure I am getting the right thing; I could take as long as a couple of weeks to do that…." (p6, a 25-year-old male)

"…I am obsessive of getting the right thing… last year, I spent 7 months on buying my laptop… I like mobile phone stores; usually I will get interested in 5 or 6 models. But I can't remember all the model names, so usually what I do is to pick up two and try to remember them…" (*p10,* a 14-year-old girl);

"…you can write it (the name of the product) down of course, but a picture is more useful to help you remember the moment and place…" (*p9,* a 29-year-old male)*;*

"…I cannot afford the phones now, but I can save it for later…" (*p5,* a 27-year-old male).

Participants liked the function because it matched with their established shopping behaviors. The time span from the shopping impulse to purchase could be more than a few days. The function of "add to wishlist" recorded and contextualized the impulses when it happened, without adding more memory loads.

- **Feedback for "send to friend" function**

Five of our participants mentioned that they liked "send to friend" function of the prototype. Below are some examples of people's feedback:

"…It's good to take a picture and ask opinions right there when you get interested, but with this prototype you can send more information around, not only pictures, but also links to specs…" (p12, a 34-year-old male)

"…I would like to let my wife know what I want. She might buy me this later (laugh)…" (p12)

"… I've being doing this (sending pictures to friends) all the time… we send pictures for anything fun…" (p4, a 13-year-old boy)

Participants' answers indicated that this function is a good fit with their shopping related social activities, such as discussions around an inspirational item, gift ideas and opinion seeking behavior [24].

Also we received many suggestions about how to improve the usability of "send to friend" function, including: avoid the requirement for remembering and typing email address by importing it from contact list, show feedbacks after the picture is correctly sent, and include other types of media in the email.

6. DISCUSSION AND FUTURE WORK

In this paper, we presented our work designing and evaluating a vision-based mobile interface for shopping in physical stores. From the formative evaluation in the field, we not only evaluated the usability and usefulness of the prototype, but also explored the opportunities and issues with vision-based mobile interface that can be applicable to other domains. We found the importance of taking error and failure into consideration for designing error prone systems, especially for vision-based interface. We also find that it cannot be taken for granted that "the faster object recognition speed is, the better interaction experience will be". Instead, the stability of the result and the update speed need to be balanced. Given the fact that many vision-based algorithms are sensitive to environmental changes, the stability of recognition result is hard to achieve. One possible solution is to leverage users' active interaction with the recognition result. Similar to White et al.'s work [34], in which users assist object segmentation to increase the recognition accuracy, we also find in our study that users are naturally interacting with the recognition result by adjusting angles and positions.

In the field study, we also find the emerging patterns of how users switch their attention between the physical and digital world. The variety of the patterns implies that mobile interface design needs to allow the flexibility of interaction process. Some of the traditional design methods, such as task analysis, may not fit very well with designing such applications due to its rigidity. Moreover, connecting the mobile tasks with real world intentions need to be considered in the design. This finding is applicable for other interfaces in which interplay between the physical and digital world is highly required.

We begin an exploration into the experiential and non-transactional aspects of shopping, and we receive promising feedback. We see this work as the first step to explore the opportunities for designing mobile shopping applications by focusing on the experiential aspects. For example, we are building an integrated web portal, through which the online shopping and mobile shopping experience can be connected. This portal is also designed to support richer social functions, such as sharing and commenting on friends' activities.

Although the field study yielded interesting findings related to our research questions, we did not conduct a quantitative comparison between mobile augmented reality interface and other interfaces, such as barcode reader-based and RFID based ones. Future work of controlled experiments can answer the questions related to the advantages, disadvantages, applicable conditions of each of the interfaces.

Another promising topic is to research is how the knowledge of products can be tagged and shared on social networks. In current prototype, the image database was populated by developers; but in the future, we plan to develop an infrastructure for end users to share and tag location-related images. The scalability issue of the algorithm (as mentioned in section 4.1) can hopefully be solved by introducing contextual information and improving object recognition algorithms. By using mobile devices, a critical mass can contribute to this system by taking images of products in stores, tagging and sharing them with others. Also, this application can be migrated to other domains, such as tourism and pervasive games.

Throughout this field study, we mostly ignored the other stakeholders (e.g. store owners), who might have concerns about people taking and sending pictures of the products in the store, and accessing the pricing information. New business models that can benefit both stores and consumers need to be explored [17] to fit the shopping ecosystem.

7. ACKNOWLEDGEMENTS

We gratefully acknowledge Markku Pulkkinen, Philipp Schloter, Kari Pulli for their contributions to the system. And we would like to thank Joel Brandt for providing us the tool of "txt 4l8r" for diary study. We also appreciate the valuable feedback from Craig Tashman.

8. REFERENCES

[1] SanpTell.http://www.snaptell.com/.

[2] Abowd, G.D. and Mynatt, E.D. 2000. Charting Past, Present, and Future Research in Ubiquitous Computing. ACM Transactions on Computer-Human Interaction 7, 1, 29-58.

[3] Brandt, J., Weiss, N. and Klemmer, S.R. 2007. txt 4 l8r: lowering the burden for diary studies under mobile conditions. Ext. abstracts CHI'07, ACM Press (2007).

[4] Brody, A.B. and Gottsman, E.J. 1999. Pocket BargainFinder: A Handheld Device for Augmented Commerce. in Proc. HUC'99 (1999).

[5] Brush, A.J.B., Turner, T.C., Smith, M.A. and Gupta, N. 2005. Scanning Objects in the Wild: Assessing an Object Triggered Information System. in Proc. Ubicomp, Springer 2005, 305-322.

[6] Chen, W.-C., Xiong, Y., Gao, J., Gelfand, N., Grzeszczuk, R. 2007. Efficient Extraction of Robust Image Features on Mobile Devices. Proc. ISMAR 2007.

[7] Consolvo, S., Paulos, E. and Smith, I. 2007. Mobile Persuasion for Everyday Behavior Change. Mobile Persuasion, Stanford Captology Media.

[8] Cuellar G., E.D., Spasojevic M. 2008. Photos for information: a field study of cameraphone computer vision interactions in tourism. CHI'08 extended abstract.

[9] Dunne, A.J. and Raby, F. Design Noir. 2001. The Secret Life of Electronic Objects. Birkhäuser, 2001.

[10] Fano, A.E. 1998. Shopper's Eye: Using Location-based Filtering for a Shopping Agent in the Physical World. Proc. Agents 98, 416-421.

[11] Fowler, G.A. 2005. QR codes: In Japan, Billboards Take Code-Crazy Ads to New Heights. Wall Street Journal 10.

[12] Gao J., S.M., Jacob M., Setlur V., Reponen E., Pulkkinen M., Schloter P., Pulli K. 2007. Intelligent Visual Matching for Providing Context-Aware Information to Mobile Users. UbiComp 2007 Adjunct Proceedings, 68-71.

[13] González, V.M. and Mark, G. 2004. " Constant, constant, multi-tasking craziness": managing multiple working spheres. Proceedings of SIGCHI conference on Human factors in computing systems 2004, 113-120.

[14] Kallio, T. and Kaikkonen, A. 2005. Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. Journal of Usability Studies 1, 4-6.

[15] Kindberg, T., Spasojevic, M., Fleck, R. and Sellen, A. 2005. The ubiquitous camera: an in-depth study of camera phone use. Pervasive Computing, IEEE 4, 2, 42-50.

[16] Kristoffersen, S. and Ljungberg, F., Making place to make IT work: empirical explorations of HCI for mobile CSCW. Proceedings of the international ACM SIGGROUP conference on Supporting group work (1999).

[17] Lee, K.J. and Seo, Y.H. 2006. A pervasive comparison shopping business model for integrating offline and online marketplace. in Proc.of the 8th ICEC 2006, 289-294.

[18] Lee, Y.E. and Benbasat, I. 2003. Interface design for mobile commerce. Communications of the ACM 46, 12, 48-52.

[19] May, P. 2001. Mobile Commerce: Opportunities, Applications, and Technologies of Wireless Business. Cambridge University Press.

[20] Miller, D. 1998. A Theory of Shopping. Cornell University Press.

[21] Newcomb, E., Pashley, T. and Stasko, J. 2003. Mobile computing in the retail arena. in Proc. CHI 2003, ACM Press, 337-344.

[22] Ngai, E.W.T. and Gunasekaran, A. 2005. A review for mobile commerce research and applications. Decision Support Systems 2005.

[23] Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J. and Stenild, S. 2006. It's worth the hassle!: the added value of evaluating the usability of mobile systems in the field. Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles 2006, 272-280.

[24] O'Hara, K. and Perry, M. 2003. User centred opportunities for supporting consumer behaviour through handheld and ubiquitous computing. System Sciences, Proceedings of the 36th Annual Hawaii International Conference on 9.

[25] Oulasvirta, A. 2005. The fragmentation of attention in mobile interaction, and what to do with it, interactions, v. 12 n. 6. November+ December, 16-18.

[26] Oulasvirta, A., Tamminen, S., Roto, V. and Kuorelahti, J. 2005. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 05, ACM press, 919-928.

[27] Pascoe, J., Ryan, N. and Morse, D. 2000. Using while moving: HCI issues in fieldwork environments. ACM Transactions on Computer-Human Interaction (TOCHI) 7, 3, 417-437.

[28] Rheingold, H. 2003. Smart Mobs: The Next Social Revolution. Basic Books.

[29] Schneider, M. 2004. Towards a Transparent Proactive User Interface for a Shopping Assistant. Workshop on Multi-User and Ubiquitous User Interfaces 04.

[30] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A. and Jain, R. 2000. Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22, 12, 1349-1380.

[31] Tarasewich, P. 2003. Designing mobile commerce applications. Communications of the ACM 46, 12, 57-60.

[32] Underhill, P. 1999. Why we buy: The Science of Shopping. Simon & Schuster New York.

[33] Venkatesh, V., Ramesh, V. and Massey, A.P. 2003. Understanding usability in mobile commerce. Communications of the ACM 46, 12, 53-56.

[34] White, S.M., Marino, D. and Feiner, S. 2007. Designing a mobile user interface for automated species identification. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems CHI 07.

[35] Wolfinbarger, M. and Gilly, M.C. 2001. Shopping online for freedom, control, and fun. California Management Review.