Technical Report

UCAM-CL-TR-609
ISSN 1476-2986

Number 609

UNIVERSITY OF
CAMBRIDGE
Computer Laboratory

Using camera-phones to interact with context-aware mobile services

Eleanor Toye, Anil Madhavapeddy,
Richard Sharp, David Scott, Alan Blackwell,
Eben Upton

December 2004

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

http://www.cl.cam.ac.uk/

ISSN 1476-2986

Using camera-phones to interact with context-aware mobile services

[♢]Eleanor Toye
`Eleanor.Toye@gmail.com`

[♢]Richard Sharp
`richard.sharp@intel.com`

[♠]Alan Blackwell
`alan.blackwell@cl.cam.ac.uk`

[♠]Anil Madhavapeddy
`anil@recoil.org`

[♠]David Scott
`dave.scott@cl.cam.ac.uk`

[♠]Eben Upton
`eben.upton@cl.cam.ac.uk`

♠ Computer Laboratory
♢ Intel Research

Abstract

We describe an interaction technique for controlling site-specific mobile services using commercially available camera-phones, public information displays and visual tags. We report results from an experimental study validating this technique in terms of pointing speed and accuracy. Our results show that even novices can use camera-phones to "point-and-click" on visual tags quickly and accurately. We have built a publicly available client/server software framework for rapid development of applications that exploit our interaction technique. We describe two prototype applications that were implemented using this framework and present findings from user-experience studies based on these applications.

1 Introduction

Modern mobile phones offer more than just voice communication, providing access to a vast array of data and services over the Internet. However, the practical benefits that users experience when accessing these services are limited by the constraints of small screens and keypads. We have developed an interaction technique that addresses this limitation, exploiting existing public information displays and the integrated cameras that are now becoming commonplace in mobile phones. Hundreds of millions of these *camera-phones* have already been shipped globally. Trends suggest that their rapid rate of adoption is set to continue for years to come [10].

We build on recent work that allows camera-phones to access web pages by photographing machine-readable visual tags representing URLs [2,3,16]. So far, work in this area has focused on defining visual tag formats and developing recognition software that will run on camera-phones. The contribution of our research is an interaction technique

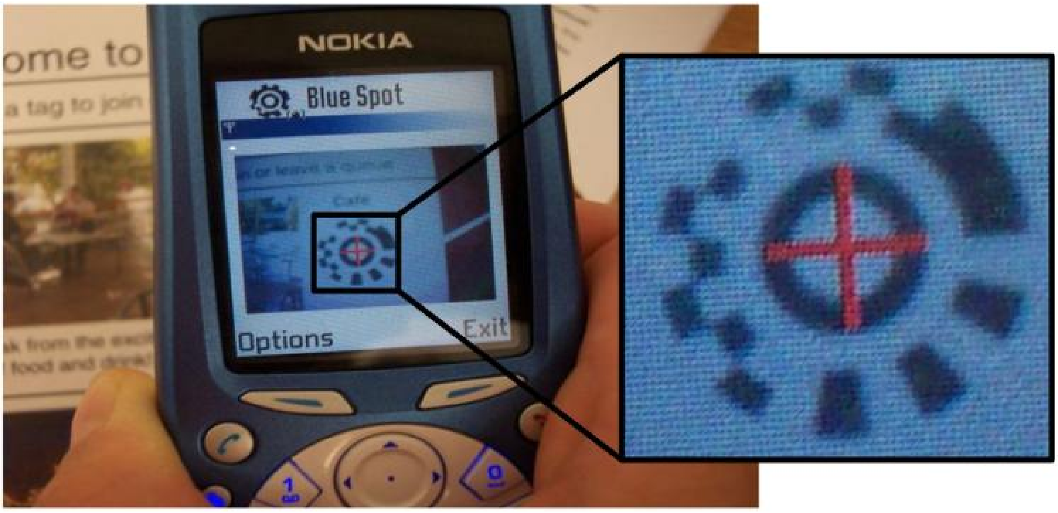


Figure 1: Using the phone-based tag reader.

that enables users to interact with mobile services via their camera-phones and visual tags. This paper specifically considers *site-specific mobile services*: electronic services or applications that augment a specific location (e.g. an interactive map in a shopping mall). The benefits of our approach are:

1. We overcome the limitations of accessing mobile services on phones' small screens and keypads by using visual tags and public information displays.

2. Visual tags can be used to create *Printed Graphical User Interfaces* [9], providing users with some of the benefits of a large electronic display but at a fraction of the cost.

3. Personal information stored on the user's camera-phone is used to drive *context-aware* [20] services that automatically tailor their actions to suit a particular user.

4. Short-range wireless connections to site-specific services offer low-latency, real-time interaction. (Our current implementation is based on Bluetooth.)

We start by describing how a user interacts with site-specific mobile services using camera-phones and tags. We describe the implementation of two prototype applications: (*i*) *Theme Park*: a virtual queueing system; and (*ii*) *Airport*: a personalised flight-information display. User-experience studies are then presented and we discuss participants' impressions, beliefs and opinions arising from their interactions with these applications. We then give a brief account of an experimental evaluation of the camera-phone as a pointing device, presenting results that show even novice users can click on visual tags quickly and accurately. Finally, we describe our implementation of a general software toolkit that enables mobile services based on camera-phones and tags to be built with minimal development effort. This software toolkit, which incorporates a fast and robust camera-phone-based tag reader, is available for download [1] and we present design guidelines for applying it, based on our own experience and the results of our user studies.

2 Overview of Interaction Technique

In this section we describe how a user experiences our interaction technique in practice. When visual tags are seen in the environment (e.g. on a poster or an electronic display) a user can activate our *Mobile Service Explorer* (MSE) application on their camera-phone. The MSE turns the camera-phone's display into a *viewfinder*: a live video feed continually capturing frames from the phone's embedded camera. Whenever a tag becomes visible in the viewfinder, it is highlighted with a red cross-hair and tracked in real-time (see Figure 1). If multiple tags are visible on the screen at the same time, the tag nearest the center of the viewfinder is highlighted. We adopt standard GUI terminology, saying that the highlighted tag *has the focus*. Users can *click* on a visual tag by pressing the select button when the tag has the focus.

A user interacts with a mobile service primarily by aiming and clicking on tags using their camera-phone. Clicking on a tag results either in information being displayed or in an action being performed by the service. The capabilities of the site-specific service determine the ways in which feedback is provided to the user. For example, in the Airport application (described later) a wall-mounted display is available. In this case, clicking on tags causes the screen to be updated and allows users to register for text-message alerts. The Theme Park application (also described later) considers a scenario where active displays are not available; instead, printed tags appear on posters and maps—focussing on tags results in information being displayed directly on users' camera-phone screens.

More sophisticated interfaces can be supported by augmenting basic camera interactions with the camera-phone's keypad and small UI controls displayed on the camera-phone's screen. The MSE allows mobile services to export numerical data-entry fields, list selection fields and multiple choice dialogues (e.g. "Continue? Yes/No") to a user's camera-phone. For example clicking on a tag to join a queue in our Theme Park application results in a numerical data-entry field appearing on the camera-phone's screen prompting the user to enter the number of people who want to join the queue.

Personal information and preferences stored on the user's camera-phone drive *context-aware* mobile services. For example, a mobile service with multi-lingual capability may query the "language preference" property on a user's camera-phone. This exchange of data, which is invisible to the user, enables the service to immediately display information in the user's native tongue. There are a wide range of preferences and items of personal information which one often wants to share with mobile services. The Airport application demonstrates this technique in context. Our user studies explore the extent to which people are prepared to use this technique in practice, addressing privacy concerns.

3 User Experience Studies

Two applications were implemented to collect user experience data: a personalised airport display and a theme park virtual queueing system. Both were designed to obtain observational and interview data regarding participants' impressions, beliefs and opinions relevant to plausible uses of camera-phones and visual tags. The applications were implemented using the general software toolkit presented later in the paper. As such, they also serve as case studies demonstrating its technical capability.

British Airways

Flight Number BA123

Departure Time 20:35

Current Status Boarding

Gate Number C21 (20 minutes walk)

Figure 2: Using the camera-phone in the Airport application

Both Theme Park and Airport applications are "interaction prototypes", designed to give users a realistic impression of how camera-phones and visual tags may be used in the near future but without the full functionality one would expect in deployed applications. We deliberately chose the theme park and airport scenarios because they were situations in which our participants could easily imagine themselves. Therefore, in spite of the fact that the applications were presented in a lab setting, we were able to gather useful data. We consider it important that the applications were implemented on commercially available camera-phones (Nokia 3650s) as this increased the realism of the interaction for the participants.

In the remainder of this section we describe the Airport and Theme Park applications. We then present the procedure and results of our user experience studies.

3.1 Application 1: Personal Airport Display

The Personal Airport Display utilises a large plasma screen, typical of those often found in airport departure lounges. Inspired by McCarthy *et al.*'s work on *proactive displays* [17], our Airport Application allows a user to turn a plasma screen into "their personal display" for the duration of their interaction with the site-specific mobile service.

In its initial default state the plasma screen runs through a slide show containing travelogue-style photos of landscapes and fireworks. A musical soundtrack accompanies the slide show. To the right of the screen is a tag which functions as a rotating volume control. Users can change the volume by focusing on the tag and rotating their camera-phone. At the bottom of the screen is a tag with the word "login" beneath it.

Using the camera-phone to click on the login tag moves the application into its *personalised state*, in which information relevant to the user is displayed. The first time the user interacts with the mobile service, they are prompted to select an airline (by selecting a tag from a menu displayed on the plasma screen) and then to enter their flight number using the numerical keypad on their phone.

On selecting a flight number, the plasma screen displays the current status of that flight (see Figure 2): the boarding and departure times, the gate number (if available), the flight's current status (delayed, gate open etc.) and a map of the airport marking the user's current location. If the gate number is available, the location of the gate is marked on the map and the time taken to walk to the gate is also displayed. The flight status screen also incorporates a visual tag which users can click to toggle *text message updates* on and off. This feature enables users to receive text message updates regarding changes in their personal flight status (e.g. boarding, flight delayed, gate changed etc.). In this way, information arising as a result of interacting with the service is stored on a users' camera-phone for future reference.

When the user clicks on the "login" tag for a second time they do not have to enter their flight number again. Instead, the mobile service recognises them from their previous interaction and immediately displays their personalised flight status information. In a deployed scenario with multiple plasma screens distributed around an airport a user could login to any of the screens and still be presented with their personal flight status information.

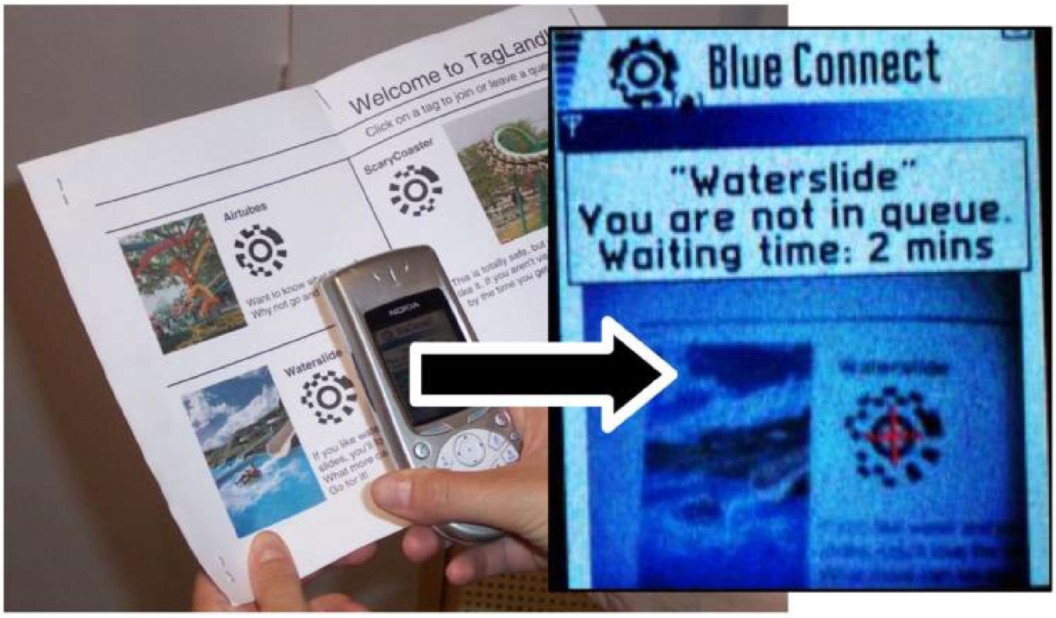
For the purposes of the user study, we implemented two states of the flight status screen. The first state shows the gate as not known and not open. On the second login, the flight status screen is updated to show the flight boarding at Gate C21 with the location of the gate shown on the airport map and a walking distance of 20 minutes to the gate.

The Airport application demonstrates that camera-phones and visual tags offer a number of advantages over simply using (say) a touchscreen to access a mobile service. Firstly, as a result of personal information stored on the camera-phone, the Airport application immediately knows *who* it is dealing with and is able to automatically display their personal flight status as soon as they click "login". Secondly, when the user clicks to turn on text message updates they do not have to enter their mobile phone number manually[1]; instead, the service obtains this information directly from the user's camera-phone.

The application also demonstrates the benefits of our interaction technique over accessing a mobile service using *only* the display and keypad on the phone itself. Rather than relying on menus and forms on a small phone screen, the Airport application is able to present a much richer user interface on its large wall-mounted display, with users able to point-and-click directly on tags within this interface. Furthermore, the "flight status screen" contains much more information than would comfortably fit on the phone's own screen. In particular, the map of the airport is more detailed and significantly larger than could be displayed directly on the camera-phone itself.

---

[1] Entering one's mobile phone number manually is more awkward than readers may expect. Several of our participants reported that they did not know their own phone number off-by-heart.



"Waterslide"
You are not in queue.
Waiting time: 2 mins

Figure 3: Using the camera-phone in the Theme Park Application.

3.2 Application 2: Theme Park Virtual Queueing System

Theme parks often have long queues for rides. Some theme parks have addressed this problem already by devising *virtual queueing systems* in which, rather than waiting in line, people are allocated "tickets" allowing them to get on rides at specified times. Tickets may be pieces of paper or may be virtual capabilities stored on an electronic device. At present, electronic virtual queueing systems deployed in theme parks involve specialised mobile devices or smartcards lent to visitors for the duration of their time in the park. In this section we describe a prototype virtual queueing system that exploits visual tags and users' own camera-phones.

We set up "TagLand Theme Park" in a large student computer laboratory in our building. TagLand consists of six individual posters sited in specific locations across the lab, each representing a different attraction. A large central poster contains a map showing the locations of the six attractions. A *visitor information leaflet* (handed to participants at the start of each study) contains a list of all six attractions, providing some information about what each one entails.

Each attraction has an associated visual tag, which is printed on the poster where the attraction is located, at the relevant position on the central map, and on the visitor information leaflet. For each of these six attractions, holding the camera-phone over the associated tag causes a text overlay to appear at the top of the camera-phone's viewfinder displaying relevant information (see Figure 3). One of the rides is closed for maintenance and holding the camera-phone over its associated tag reveals this information; the other attractions display their current queueing times.

The posters representing attractions and the central map contain text inviting users to "click on a tag to join a queue", and explain that they can leave the queue at any time simply by "clicking the tag again". On clicking a tag, a text-entry field appears on the phone prompting the user to "enter the party size" (the number of people wishing to join the queue). After the party size is entered, using the keypad on the phone, the application records that the participant is now in the queue for a ride.

When a user focuses on a tag corresponding to an attraction for which they have a booking, the resulting text overlay informs them that they are in a virtual queue. The remaining queueing time in minutes and seconds is also displayed, ticking down to zero. Clicking on the tag in this state yields a confirmation dialogue asking the user if they want to leave the queue. Users can select "Yes" or "No" using their camera-phone's keypad.

Shortly before the user reaches the front of a queue, the phone beeps and displays a dialogue notifying them to "Go to ⟨*attraction name*⟩ in 30s". (Clearly in a real deployment more notice would be given.) When the front of a queue is reached a fanfare sound plays and an alert dialogue appears, confirming that they are now able to take the ride.

The Theme Park application demonstrates that visual tags can be used to create *Printed Graphical User Interfaces*, providing users with some of the advantages of a large electronic display, but at a fraction of the cost. While static information (the Theme Park map, posters and visitor information leaflet) is presented on paper, dynamically changing information (e.g. ride queueing times) is overlayed via users' camera-phone screens. Users can interact with paper-based UI controls by clicking on them and entering data via their camera-phone's keypad.

3.3 User Experience Study Participants

18 subjects (9 men and 9 women) aged between 24 and 43 years (mean = 29) served as participants. They were an educated and mostly professional group. None reported difficulties with vision or motor skills. None worked in the field of Computer Science or in the mobile phone industry. Participants were recruited either through flyers posted around the city and University of Cambridge, or by word of mouth and were each rewarded with a book token (value 20 UKP). Sixteen participants were British, one Australian and one Chinese. The Chinese participant had lived in the UK for 10 years and the Australian had lived in the UK for 2 years.

Our rationale for choosing this demographic was that it allowed us to identify any general usability difficulties which affected even healthy, younger adults. We intend to carry out further studies with broader samples of the population in future work.

3.4 User Experience Study Procedure

Participants were asked to explore first the Theme Park application, then the Airport application. Structured interviews were carried out at the end of each session. Initial questions in these interview schedules were entirely open-ended (e.g. "Was there anything about the Theme Park that you found confusing or difficult?"). Next, users were asked to make comparisons between using camera-phones/tags and other familiar technologies. All the comparative questions were worded to avoid drawing attention to specific features of the technology (e.g. speed or convenience). However, two questions in the Airport interview focused specifically on privacy, an issue we particularly wanted to explore. At the end of the study a more general interview was conducted in order to explore their overall impressions of the technology and also participants' attitudes to adoption.

3.4.1 Theme Park Procedure

For the Theme Park study, each participant was given a Nokia 3650 camera-phone and a visitor information leaflet. They were instructed to explore the theme park and use the camera-phone to join and leave queues for the different rides and attractions. They were told that they could do this for as long as they liked and to meet their researcher back at the central map whenever they felt they had finished exploring.

3.4.2 Airport Procedure

For the Airport study, the researcher first described to the participant how camera-phones and circular tags might be used to provide access to personally relevant information in a public space. The participant was given a Nokia 3650 camera-phone and an airline ticket for British Airways flight BA123 from SpotCode Airport to Paris. They were asked to try out using the display, and to talk aloud at the same time about what was happening and what they were doing. They were filmed doing this. Some participants needed to be prompted initially to keep talking about what they were doing, as is standard in "think-aloud" studies.

If the participant used the display to request 'text message updates', they were handed a piece of paper shortly afterwards, representing a text message, telling them that their flight was now boarding. If the participant did not spontaneously logout or login a second time, they were prompted to do so by the researcher. (This was so that the immediate availability of their personal updated flight information on second login was apparent to them.)

3.5 Results and Discussion

3.5.1 Participants' Ability to Focus and Click on Tags

Our data showed that users from the demographic we tested are able to use camera-phones and visual tags without extensive training. All the participants were able to interact successfully with the Theme Park and Airport Applications. However, three (non-critical) usability problems did arise during testing.

Firstly, the rotating volume control which featured in the Airport application was not successful for a variety of reasons. Eleven participants made negative comments about it, citing at least one of four problems: it was generally confusing; it was highlighted in red, which distracted from the login tag; its function was unclear; and it didn't require clicking, but was difficult to rotate. We conclude that rotation controls of this form are not suitable for novice users.

Secondly, 4 (of 18) participants experienced problems clicking on tags in the Airport application. However, nobody had difficulties clicking on tags in the Pointing Device experiment (described later) or in the Theme Park application. It transpired that participants thought they were pressing the select button on their Nokia 3650 camera-phone, when in fact they were pressing one of the cursors[2]. The reason no clicking problems were experienced elsewhere was that, for the other applications, we happened to use a different version of the tag reader which registered a click when *any* key was pressed.

---

[2] The Nokia 3650 has a combined rocker-control that incorporates cursors and the select function.

Thirdly, 2 (of 18) participants commented that they had some trouble with the red cross-hair flickering on the viewfinder, so sometimes they were unsure whether they had successfully focused on a tag. Observation suggested that this was because they were holding the camera-phone just beyond the maximum distance at which the tag could be reliably decoded. We expect that this problem would disappear with further practice.

Despite these issues, however, participants on the whole were able to point-and-click on tags quickly and accurately using our camera-phone-based tag reading software. We justify this statement quantitatively later in the paper by the results of our Pointing Device Experiment.

3.5.2 Participants' perceptions of the technology

A majority of participants saw increased mobile access to information as a positive feature of the system. Talking about their experience of the Airport application, 10 participants made positive comments about the text message notification facility, including:

> *"The text update thing is handy, otherwise you have to sit by the TV screens all the time."*

Describing the Theme Park application, 8 participants made positive comments about having mobile access to current queue lengths and the ability to change their own queue status using tags. For instance:

> *"I liked the ability to get real-time updates by just looking at the tags to find out how far I was from the front of the queue".*

Several people also saw personalisation as a benefit of the technology. In response to the Airport application, 6 participants made positive comments about only receiving personalised information regarding their own flight. For instance:

> *"It's convenient, and it provides information relevant to me so I won't have to look though a long list of flights on the departure screen."*

However, a minority of participants (3 of 18) said they would rather use airport departure boards in their existing form. One participant commented:

> *"If the same information was available on a flight display board and I just had to look up, why would I go through the hassle of using my phone?"*

Overall, participants' responses to the technology were positive. Twelve participants believed the camera-phone-based Airport application to be better than existing flight departure boards (4 thought it was the same, no-one thought it was worse, 2 participants did not answer this question). When asked to compare the Theme Park application to a conventional ticket-based virtual queueing system, 16 participants said the camera-phone system was better. One participant who had recently experienced a ticket-based queueing system at Disneyland last year said of his Disney queueing experience:

> *"... [Disney's queueing system] is better than standing in line but I can't cancel my place on the ride and there'd be queues of people at the ticket machines— and I have to go to the machines, I can't just join queues from wherever I am."*

We see this positive comparison to an existing commercial system as a promising assessment of our research prototype.

3.5.3 Privacy

Many of our participants perceived the personalisation of mobile services to be beneficial. However, personalisation raises important privacy issues. We were interested to ascertain (*i*) how far participants were aware of the privacy implications of personalisation; (*ii*) how concerned they were about sharing personal information with mobile services; and (*iii*) how privacy concerns would affect their behaviour.

When questioned, many participants were unaware that interacting with personalised services (such as the Airport application) would involve giving personal information to service providers. We presented participants with a multiple-choice question, asking whether they noticed that by opting to receive text-message updates in the Airport application they must have implicitly given their mobile phone number to a third party. Seven (of 18) said that they did not notice this at all; 2 said they noticed this issue but had not given it much thought; 6 said they noticed and thought about it, but didn't worry about it; and 3 said that they noticed and worried about this privacy issue while using the application.

This observation subsequently led us to add a confirmation dialogue, displayed on the camera-phone screen, when a service requests personal information (e.g. "The mobile service wants to send you a confirmation email. Do you want to give your email address to this service? Yes/No"). We have not yet evaluated this new feature but we hope it will help users to take control of their personal information when interacting with mobile services.

After we explicitly drew privacy issues to users' attention, almost all said that they would still continue to use personalised services via their camera-phones. Eight participants said that they would use personalised services from known and trusted providers; 8 said that they would use personalised services from any service provider if the provider promised not to spam them with emails or text messages; and 1 participant said that they would be happy to use personalised services whenever convenient. Only 1 participant said that they would not interact with mobile services unless they could tell their phone never to give out personal information.

We are aware that participants did not actually have to give out any of their *own* personal information in our studies; the camera-phone they were using was lent to them at the beginning of the study. Although the above findings are interesting preliminary results, we believe that more detailed studies based on actual deployments of the technology would be required to predict more accurately users' attitudes to sharing personal information with mobile services.

Using public displays to present personal information (such as flight details in the Airport Application) raises another interesting privacy issue—to what extent are people prepared to have personal information presented on large public displays? Even when just displaying flight information on the screen (without, for instance, also displaying the user's name) our participants were still wary. The majority of participants (11 of 18) said that, although they would be happy to use the Airport application on a public display, they would prefer a more private display. (The other 7 participants said that they did not mind their flight information on a public display.) Typical comments included:

> *"Just flight details would be OK but I wouldn't be happy if more personal details were displayed."*

Figure 4: Performing the Pointing Device Task

3.5.4 Attitudes to Adoption

Attitudes to adoption were very positive. All of our participants (not including two who already had Bluetooth camera-phones) said that, when buying a new phone, they would choose one that allowed them to use tags in preference to one that didn't. Nearly half (7 out of 16 participants who did not already own camera-phones) said that they would be prepared to pay more for a mobile phone that enabled them to use tags to interact with mobile services. Of the group tested, all already used mobile phones on a regular basis. It remains an open question whether other social groups, who may not already use mobile phones regularly, would be so open to adopting camera-phones and tags.

4 Pointing Device Experiment

The user experience studies described above were designed to obtain qualitative information regarding participants' perceptions of camera-phones and tags in their everyday lives. In contrast, the pointing device experiment was designed to ascertain from quantitative data how effectively users could click on tags. The same 18 participants (described above) took part in the experiment. The participants performed the pointing device task twice: once before the two user-experience studies and once afterwards. Clicking times and accuracy data were measured. The data allowed us to investigate how users' performance varied as the size of tags changed.

4.1 Experimental Design

In each *trial*, a pair of equal-sized tags appears on a plasma screen, one highlighted with a red box around it (see Figure 4). A participant uses their camera-phone to click first on the highlighted tag, then on the other tag.

The horizontal distance between tags and the diameter of the tags varies across different trials. Every combination of 8 different tag diameters (ranging from 4cm to 11cm, at increments of 0.875cm) and 6 different inter-tag distances (ranging from 18cm to 63cm, at increments of 7.5cm) are presented, resulting in $6 \times 8 = 48$ trials. Each of these 48 trials is performed twice: once for each *clicking order*—that is, whether at the start of a trial the highlighted tag is on the right or the left. Thus, the total trials performed in a single test run $= 48 \times 2 = 96$. Trials with the same clicking order are presented consecutively, meaning that the clicking order only changes once between the two blocks of 48 trials. The order of the trials *within* these 2 blocks of 48 are randomly permuted for each participant.

The 96 trials are divided into sets of 15, each separated by 15 second rest periods, with a short set of 6 trials at the end. The choice of 15 trials per rest period was not determined by the experimental design; we simply found, during pilot testing, that this was the number of trials a participant could perform consecutively before they started to feel tired.

4.2 Procedure

Participants were asked to perform 10 initial practice trials, followed by the 96 clicking trials. For each trial, participants were told to click first on the highlighted tag and then, as quickly as possible, on the other tag. To ensure a standard speed/accuracy tradeoff across participants we instructed them to aim for no more than 1 error in each set of 15 trials (corresponding to a maximum 6% error rate). To avoid ordering effects we ensured that same number of participants started with the left/right clicking order as started with the right/left clicking order.

For each trial, a computer recorded the *clicking time*[3]: the time between the participant clicking on the first tag and clicking on the second tag rounded to the nearest 10ms. (This resolution was sufficient for our purposes.) An error was registered if the participant pressed the select button on their camera-phone when a tag was not in focus, or if the wrong tag was clicked on. At the bottom of the screen a status bar showed the participant the number of trials remaining and the number of trials in the current set of 15 that had so far resulted in error.

Participants were informed that there would be a forced break every 15 trials. They were also told that they could take breaks as often as they liked and for as long as they liked between trials (although in practice no-one actually did this). Every participant performed the pointing device task twice, separated by a break of at least half an hour. This gave us 192 trials per participant.

---

[3] Although recorded by computer, the times were *measured* by the camera-phone itself. This ensured that the latency and jitter of the Bluetooth connection (used by the phone to communicate with the computer) did not affect our data.

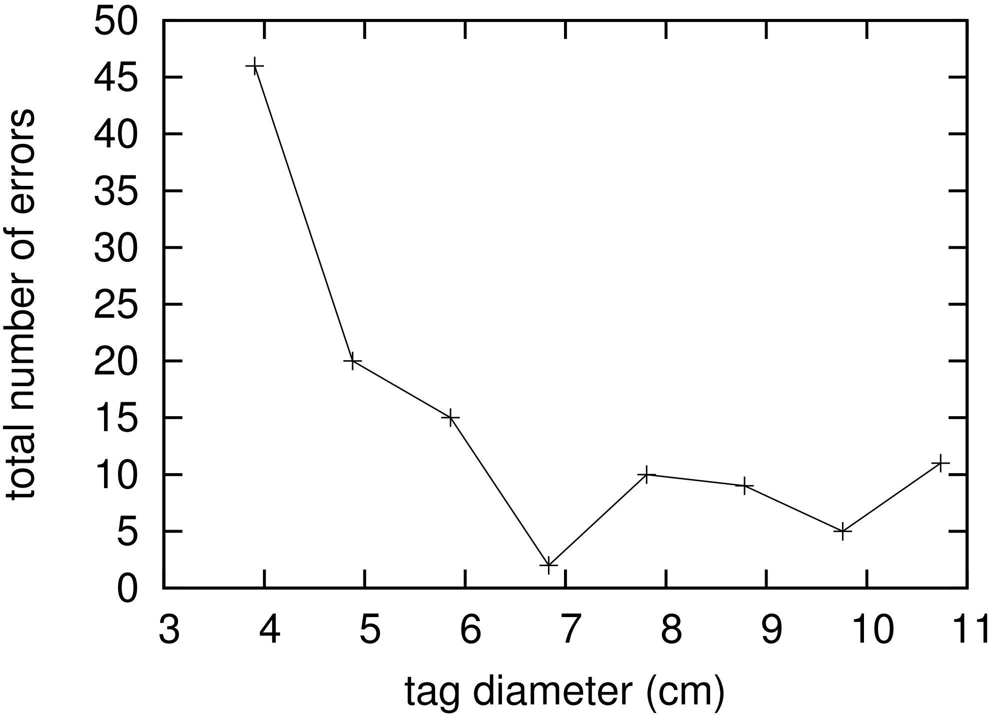


Figure 5: Plot of tag diameter and error rate (summed across all participants for both runs of experiment).

4.3 Results

We found that all participants were able to use our camera-phone-based tag reader to click on tags accurately. Of the 96 trials in a single run of the experiment, participants made an average of 3.2 errors on their first go and 3.3 errors on their second go—well within the specified target range of $< 6\%$. Three participants completed an entire run of the experiment without making a single error; the maximum number of errors made in a single run was 8, an error rate of 8.3%. Figure 5 shows how the total number of trials performed in error (summed across all trials for all participants) varies with tag diameter. The results show that participants made more errors when clicking on smaller tags. The lowest number of errors occurred when the tag diameter was between 6 and 7cm.

In the following reporting of clicking times, trials completed in error have been excluded. Participants clicked fastest (relative to their own performance) on tag diameters between 6cm and 9cm. This result was obtained as follows. First, as is standard practice for reaction time data, we used a log transform to correct for skew. In the log domain, we calculated a mean reaction time for each participant and tag diameter (collapsing across inter-tag distances). These averages were transformed into z-scores, thus expressing each participant's log reaction times in terms of deviations from their own mean. Figure 6 shows a plot of the average of all participants' z-scores for each tag diameter. Positive z-scores indicate that a participant clicked slowly relative to their own average performance; negative z-scores indicate that they clicked quickly relative to their own average performance. This shows the stated result.

Participants completed the vast majority of the trials (80%) in between 1 and 2s, a

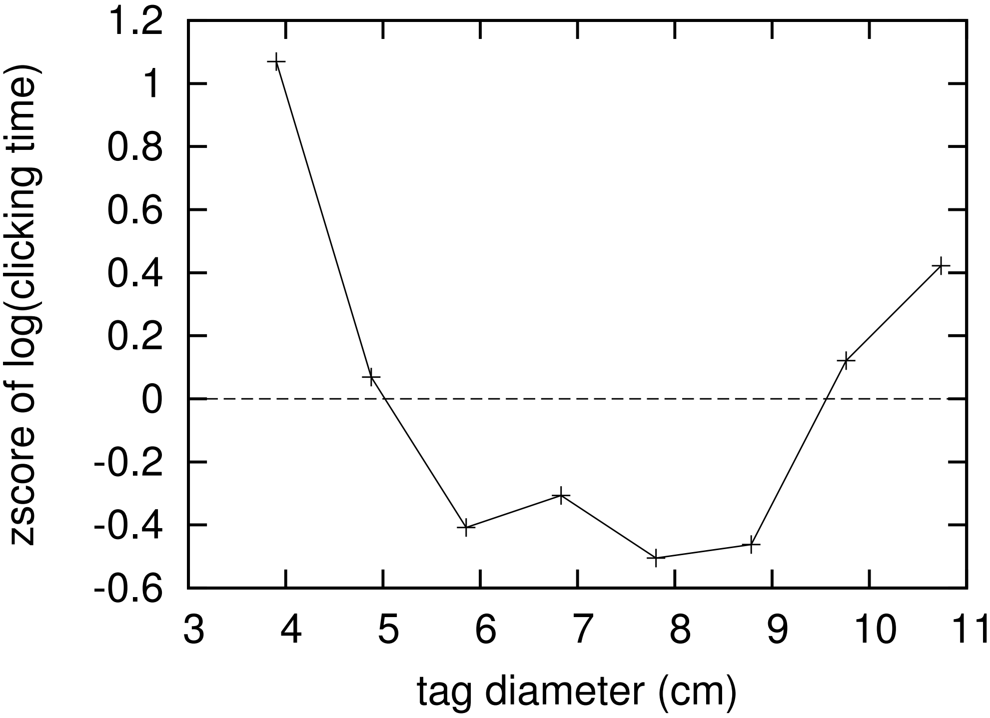


Figure 6: Plot of tag diameter and z-score of log(clicking time) averaged across all participants and both runs of experiment.

timescale which is clearly well within the limits of acceptability for accessing tag-based mobile services. Figure 7 shows mean clicking times and standard deviations for each tag diameter (collapsed across inter-tag distances). As before, we used a log transformation to correct skew when computing means and standard deviations.

Overall, the most successful tag diameter was between 6 and 7cm. Not only did participants click faster on tags of these sizes on average, but more importantly, they made virtually no errors in doing so. Observation suggests that this is because, at the distance users tend to stand from the plasma screen (approximately arm's length), tags of this size comfortably fill the camera-phone's viewfinder. Users therefore do not have to move their camera-phone backwards or forwards in order to frame these tags in the viewfinder.

Although a detailed analysis is beyond the scope of this paper, we found that Fitts' law correlations on the data were weak, taking tag diameter as a measure of Fitts' *target width*. This result is perhaps unsurprising when one considers that the actual target defined by a tag (i.e. the points in 3D space at which the camera-phone can decode it) form a frustum, the parameters of which have as much to do with the camera-phone's optics as the tag diameter itself. Furthermore, the camera-phone can be rotated about a vertical axis to acquire larger tags without having to move it so far. Due to space constraints, a full analysis of the pointing test data will be presented in a forthcoming paper, including detailed Fitts' law analyses, multi-variate analyses and a comparison between novice and expert users of camera-phones and tags.

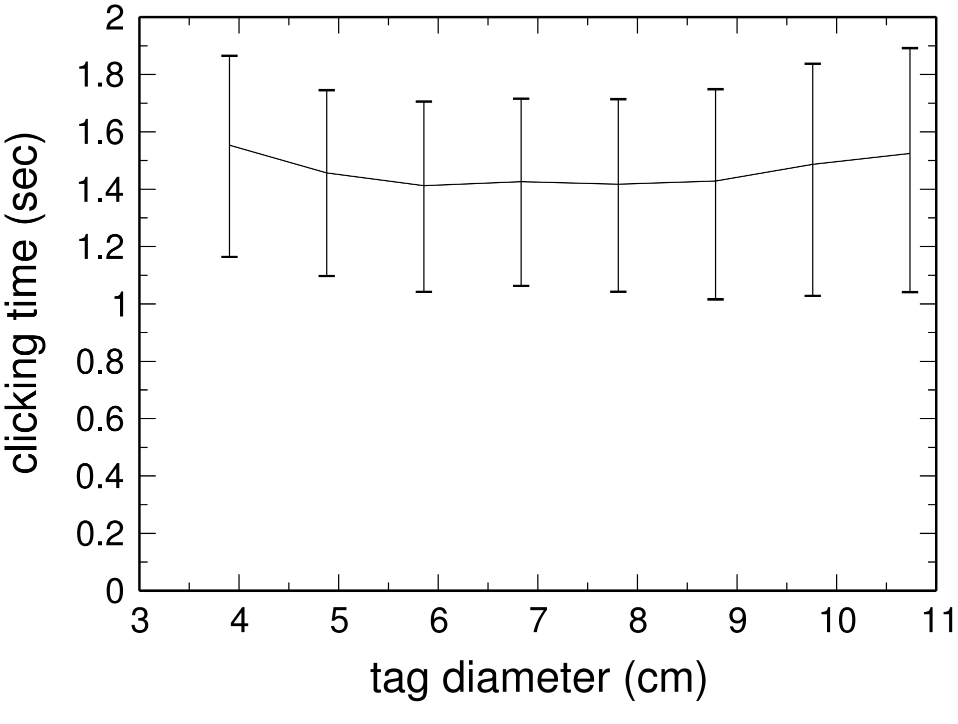


Figure 7: Geometric mean clicking times and standard deviations for different tag diameters (averaged over all participants and both runs of experiment)

5 General Software Toolkit

We have implemented a client/server software toolkit that greatly simplifies the development of mobile services based on camera-phones and visual tags, making it available for download [1]. Our server-side software runs on Linux with the Bluez Bluetooth Stack, FreeBSD and MacOS X.

The client, which we call the Mobile Service Explorer (MSE), runs on users' camera-phones. We have implemented the MSE for *Nokia Series 60* camera-phones (as a native Symbian C++ application). The MSE is the *only* application that users need install on their camera-phones in order to interact with *all* our tag-based mobile services. It is also the reason why our software toolkit dramatically reduces application development time: all camera-phone-based code is taken care of by the MSE[4]; the developer only has to code server-side application logic for the mobile service.

The remainder of this section describes the three main aspects of the MSE in turn: (*i*) the tag reader and the supported visual tag formats; (*ii*) the MSE; and (*iii*) Bluetooth connection-establishment.

5.1 Tag Format and Tag Reader

Based on the authors' previous work on SpotCodes, our circular tags have been designed to work well with the low-resolution fixed-focal-depth cameras found on today's camera-phones. Crucially, the individual bits are large enough to be reliably decoded from distances at which the camera's lens gives a sharp image. In contrast we found that familiar linear (e.g. UPC) barcodes are unsuitable for decoding on camera-phones: when the phone is positioned near to a linear barcode, the image is out of focus and cannot be decoded; at distances large enough to capture sharp images, the resolution is insufficient to make out individual bars.

The MSE recognises circular tags containing 23 sectors and between 1 and 4 data tracks. This allows us to encode 23 data bits (in the 1 track case) through to 92 data bits (in the 4 track case). By supporting tags of different sizes we give application designers the flexibility to trade data capacity against tag real-estate. For more information about the tag format see [21]. Our tag reading software runs at 15 frames per second on the Nokia 3650 with very low latency.

5.2 Mobile Service Explorer (MSE)

The MSE runs on users' camera-phones, performing three main functions: (*i*) tag reading (described above); (*ii*) personalisation with respect to a user-defined privacy policy; and (*iii*) rendering UI controls on the camera-phone's screen and relaying user responses back to mobile services.

The MSE connects to a mobile service over Bluetooth. Once a connection has been established (see below) the MSE acts as a slave, responding to requests issued by the mobile service. Requests take one of four forms: (*i*) give me recent tag sightings and

---

[4] Developing native C++ mobile phone applications is a difficult systems engineering challenge, requiring detailed knowledge of esoteric programming paradigms and obscure build environments.

telemetry information (including co-ordinates in viewfinder and orientation); (*ii*) give me recent keypresses on the phone's keypad; (*iii*) display this UI control on the camera-phone screen and tell me the user's response; and (*iv*) give me this piece of a user's personal information (e.g. mobile phone number, email address).

The MSE stores users' preferences and personal information under a set of pre-defined field names. When a mobile service requires information about who it is dealing with, it sends a request to the MSE. A user-configurable *privacy policy* determines, for each piece of personal information, whether the information should be given out by default or whether to seek user confirmation.

5.3 Bluetooth Connection Establishment

4-track tags can encode both a Bluetooth Device Address (BD_ADDR) as well as some application-specific bits. When one of these tags is decoded, the MSE connects to the specified BD_ADDR. In previous work we have shown that this approach (*i*) leads to order of magnitude connection time improvements over the standard Bluetooth Device Discovery model ($\bar{x} = 1.77$s, $\sigma = 0.47$ using tags; $\bar{x} = 17.12$s, $\sigma = 8.6$ using Device Discovery); and (*ii*) takes the user an order of magnitude fewer keypresses than connecting to a device using the Discovery model (1 keypress to click on a tag; an average of 10 keypresses to perform Device Discovery and select from the resulting list) [21].

6 Related Work

Many researchers have addressed the problems arising from small screens on mobile computing devices, exploring a variety of techniques including panning [12], fish-eye views [8, 19] and novel uses of audio [5]. We see our work as complementary to these approaches. Indeed, applications based on camera-phones and tags could also usefully employ (say) panning and zooming screen displays.

We are not the first to note the potential of the camera as an appealing input medium for interaction. Much of the work in this area involves fixed cameras trained on users [7]. This approach, pioneered by VideoPlace [15], has proved fruitful in areas such as gesture recognition [6] and location-aware computing [11]. In contrast, camera-phones are more suited to the approach often taken by the augmented reality community in which hand-held cameras are pointed at objects in the world. A variety of augmented reality systems based on optical tags have already been developed [4, 13]. These systems influenced our design to a large extent. In particular, the textual overlays that appear when the user focuses on a tag (see the Theme Park Application) were inspired by the NaviCam [18] project.

The CoolTown project [14] proposed the *physical hyperlink*: a way of associating web pages with physical objects. In the spirit of CoolTown, much recent work has focused on using visual tags for camera-phone-based URL resolution [2, 3, 16]. Our use of tags goes further than this: we use tags to initiate connections to site-specific services (analogous to URL resolving for web-based services); but following this, tags continue to be used to interact with the service itself (e.g. clicking on tags to join and leave queues in the Theme Park application). Furthermore, in traditional web-based services, each action has to be initiated by a user clicking on a hyperlink. In contrast, our system enables mobile services to *push* content to users' camera-phones (such as the alert dialogues and text overlays seen in the Theme Park application).

Controlling electronic displays with camera-phones and tags enables standard displays (including CRTs, LCDs, plasma screens and even projected displays) to provide a functionality similar to that of a touchscreen. In this way we enable *existing* display technology to be used to greater effect.

7 Conclusions and Future Work

We have presented interaction techniques that enable users to access mobile services using their camera-phone and visual tags. The benefits to users are (*i*) the use of tags and public information displays helps to overcome the limitations of phones' small screens and keyboards; (*ii*) context-aware mobile services can be personalised automatically using information stored on users' camera-phones; and (*iii*) information arising as a result of interacting with mobile services can be stored on users' camera-phones for future reference. For service providers, camera-phones and tags offer the opportunity to deploy site-specific mobile services cheaply, using visual tags to create Printed Graphical User Interfaces (e.g. the Theme Park application). This approach can provide users with some of the benefits of a large active display, but at a fraction of the cost.

We have evaluated our interaction techniques, presenting both quantitative and qualitative results from our user studies. All participants were able to use our tag-reading software to click on tags quickly and accurately. Users generally saw personalisation of mobile services as a benefit, but had some concerns about the privacy issues this raised. We have addressed these issues by modifying our Mobile Service Explorer to allow users to specify a fine-grained privacy policy, and by providing notification on the camera-phone screen when mobile services request personal information. Users were enthusiastic about having dynamic information pushed to their camera-phones by mobile services (e.g. queueing time updates in the Theme Park, and flight status information in the Airport). Attitudes to the technology as a whole were very positive. All users said that they would preferentially buy a camera-phone that allowed them to use our interaction technique; nearly half said that they would *pay more* for a phone if it allowed them to use our interaction technique.

A general software toolkit, which makes it easy for others to build mobile services that exploit camera-phones and visual tags, has been implemented and released [1] under a BSD licence. At present our software toolkit only supports our own circular tag format and Bluetooth connectivity. In the future we intend to extend the tag reading software to recognise different tag formats (e.g. QR codes [2]) and to experiment with other networking technologies (e.g. GPRS, 3G). We chose to support Bluetooth in our initial implementation for two reasons: (*i*) it has low latency, facilitating real-time interaction; and (*ii*) it is completely free—camera-phone owners do not even need an active SIM-card to access Bluetooth services.

Based on our experiences and the results of our user studies, we advise developers to adhere to the following guidelines when deploying applications based on our software toolkit:

1. Keep data-entry on the phone's keypad to a minimum, relying on tag-based inter-

action instead whenever possible. (In the user-experience interviews, several participants reported that they preferred tag-based interaction to entering data via the camera-phone's keypad.)

2. Be careful to preserve users' privacy when designing applications that present information on public displays (such as the Airport application). If possible, try to partition information so personal information is displayed via dialogues on the user's (more private) camera-phone's screen, whilst more general information appears on the larger public display.

3. Place text or icons that describe the function of a tag very close to the tag itself. Then, when a camera-phone is aimed at the tag, both the tag and its surrounding context will be visible in the camera-phone's viewfinder. This prevents users from having to constantly switch their gaze between two displays; all the information they need to identify a particular tag is on the phone's own display.

More generally we observe that:

1. When information can be partitioned into a static part (e.g. the Theme Park map) and dynamic parts (e.g. Theme Park ride queueing times), then visual tags can be used to form *Printed Graphical User Interfaces* that provide some of the advantages of a large display but more cheaply.

2. Using camera-phones and visual tags to interact with electronic displays in public places offers an alternative to deploying dedicated data-entry devices (e.g. mice or trackerballs). By exploiting users' *own* devices, vandalism and wear-and-tear of shared public data-entry devices can be avoided.

Small embedded cameras are becoming ubiquitous thanks to recent advances in CMOS imaging technology. As cameras increasingly become a regular feature of electronic devices (such as camera-phones and PDAs) [10] preliminary findings suggest that our interaction technique has the capability to deliver real benefits to a wide range of users.

References

[1] Mobile Service Toolkit download. Available from `http://www.highenergymagic.com/`.

[2] "NTT DoCoMo Unveils New mova 506i i-mode Phone Series". Press release. Available from NTT DoCoMo press center.

[3] Semacode website. `http://semacode.org/`.

[4] Ronald T. Azuma. A survey of augmented reality. *Presence*, 6(4):355–385, August 1997.

[5] Stephen A. Brewster. Overcoming the lack of screen space on mobile computers. *Personal and Ubiquitous Computing*, 6(3):188–205, 2002.

[6] K. Dorfmueller-Ulhaas and D. Schmalstieg. Finger tracking for interaction in augmented environments. In *Proc. IEEE International Symposium on Augmented Reality*. IEEE Press, 2001.

[7] Jerry Fails and Dan Olsen. A design tool for camera-based interaction. In *Proc. CHI*, pages 449–456. ACM Press, 2003.

[8] Carl Gutwin and Chris Fedak. Interacting with big interfaces on small screens: a comparison of fisheye, zoom, and panning techniques. In *Proceedings of the 2004 conference on Graphics interface*, pages 145–152. Canadian Human-Computer Communications Society, 2004.

[9] David L. Hecht. Printed embedded data graphical user interfaces. *Computer*, 34(3):47–55, 2001.

[10] IDC Research. Moving pictures 2003: Worldwide camera phone survey, forecast, and analysis, 2003-2007. `http://www.idcresearch.com/`.

[11] Diego Lopez de Ipina, Paulo Mendonca, and Andy Hopper. TRIP: a low-cost vision-based location system for ubiquitous computing. *Personal and Ubiquitous Computing*, 6(3):206–219, 2002.

[12] J. Johnson. A comparison of user interfaces for panning on a touch-controlled display. In *Proc. ACM CHI*, pages 218–225, 1995.

[13] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *International Workshop on Augmented Reality (IWAR)*, pages 85–94, October 1999.

[14] Tim Kindberg, John Barton, Jeff Morgan, Gene Becker, Debbie Caswell, Philippe Debaty, Gita Gopal, Marcos Frid, Venky Krishnan, Howard Morris, John Schettino, Bill Serra, and Mirjana Spasojevic. People, Places, Things: Web Presence for the Real World. In *Proceedings of the third IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'00)*. IEEE Press.

[15] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. VIDEOPLACE—an artificial reality. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, 1985.

[16] Bango Ltd. Bango spots link camera phones straight to mobile content. Press release, available from `http://www.bango.net/`.

[17] Joseph F. McCarthy, David H. Nguyen, Al M. Rashid, and Suzanne Soroczak. Proactive displays: Enhancing awareness and interations in a conference context. Technical Report IRS-TR-04-015, Intel Research Seattle Technical Report.

[18] Jun Rekimoto and Katashi Nagao. The world through the computer: Computer augmented interaction with real world environments. In *ACM Symposium on User Interface Software and Technology*, pages 29–36, 1995.

[19] K. Rodden, N. Milic-Fralyig, Sommerer R, and A. Blackwell. Effective web searching on mobile devices. In *Proc. HCI*, pages 281–296. Springer Verlag, 2003.

[20] Bill Schilit, Norman Adams, and Roy Want. Context-aware computing applications. In *IEEE Workshop on Mobile Computing Systems and Applications*, Santa Cruz, CA, US, 1994.

[21] David Scott, Richard Sharp, Anil Madhavapeddy, and Eben Upton. Using visual tags to bypass bluetooth device discovery. To appear in ACM Mobile Computing and Communications Review (MC2R), January 2005.