Creating and sharing multi-media packages using large situated public displays and mobile phones

Andrew Maunder
Dept. of Computer Science
University of Cape Town
Cape Town, South Africa
amaunder@cs.uct.ac.za

Gary Marsden
Dept. of Computer Science
University of Cape Town
Cape Town, South Africa
amaunder@cs.uct.ac.za

Richard Harper
Microsoft Research
7 JJ Thomson Ave., Cambridge,
CB2 0DF, UK
r.harper@microsoft.com

ABSTRACT

This paper will describe a novel interaction technique that allows mobile phone users to create and share contextualised media packages between their personal, BlueTooth enabled camera phones, and situated public displays. Unlike other solutions to this problem, the one presented in this paper does not require any specialist software or hardware on the user's handset. We believe this technique has the potential to revolutionise how people donate and retrieve digital media files without incurring any direct cost.

Author Keywords

Public displays, camera phone, BlueTooth, human computer interaction, multi-media, content sharing.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Situated public displays are a common feature in public spaces where dynamic, contextualized information needs to be conveyed to members of the public. At present, however, most of these screens are passive, displaying a series of static images. More recently, various research groups have evaluated user interaction techniques with large displays in an effort to make these installations more useful to those consuming the information they display. One particularly interesting form of interaction with large public displays is that of *'personal device based'* interactions, wherein users interact with public displays via their cellular handsets [1]. Although Ballagas et al [1] highlight two important benefits of this approach, (namely the growing ubiquity of such devices and the user's familiarity with their own device), we believe there are many more interesting applications beyond this:

- The display system can act as a limited information source in areas and contexts where access to mobile networks is poor (within a museum, for example).
- It can also be used as a free information source in contexts where users may be interested in information, but unwilling to pay for it (maps in an airport or shopping mall)
- Any contextualised display information that is transferred onto a user's personal device can be taken away and reviewed and reflected upon at a later time.
- Finally, as two of this paper's authors are based in the developing world they are acutely aware of the problems facing many users within Africa whereby they can afford a handset, but cannot afford the airtime to make use of it. Thus mobile network services may not be a feasible approach and alternative, cost effective ways of content provision must be investigated. Situated displays, accessible via unregulated, personal area networks, would allow users to access relevant information (e.g. health information and job opportunities) in public places without incurring any network charges.

It is hardly surprising therefore that, given the benefits on offer, many such systems have been developed and deployed in public spaces. It is our belief, however, that none of these existing systems have provided a holistic, practical solution that may be deployed in a realistic public setting. Our goal for the rest of this paper is to examine the limitations of existing systems and propose a new interaction technique that works around them.

Tarpits and Design Goals

Most recent interactions between situated displays and cellular handsets have been based on the BlueTooth protocol. This is an obvious choice as it is currently the most common personal area network protocol supported by cellular handsets[2]. Taking BlueTooth as a given, what are the problems one needs to work around in order to design intuitive and rapid interactions?

The main problem lies in pairing devices. Core to the BlueTooth protocol is the notion that some sort of identification must take place before devices can select and share information. This takes the form of a pairing process, by which the systems (in our case the public display and cellular handset) must exchange a PIN code. For public displays where many people may wish to swap information simultaneously, this scheme becomes unworkable. No solution should therefore require pairing of devices.

If pairing is to be overcome, then one solution is to use BlueCasting, whereby a server sends information to every suitably configured BlueTooth device which comes within range. However, this approach is irritating to users who neither initiate the communication, nor have choice over what they receive. In an example, Filter UK, a mobile marketing, application and content agency, utilised BlueCasting to promote the release of an album in 2005 [11]. The BlueCasting system attempted to broadcast media clips to devices in range of large display screens within a public space (in this case a train station). Over 87,000 devices were *'discovered'* but only 15% of those device owners accepted the transfer. This statistic highlights the scope of the Bluecasting problem, 74 000 people received unwanted messages. In addition, the type of media clip received was determined by the server's promotional policies, not the device owner's choice on the matter. Any successful system must therefore ensure that the user initiates the interaction and is able to choose what they wish to receive.

In order to meet the criteria in the last sentence, many researchers have resorted to installing software in the client device. This software can either remotely control a display cursor [10,12] or read and interpret a tag (RFID or visually via the camera [7]) to indicate the media required. Whilst this does solve the BlueCasting problem, the solution relies on software that must be pre-installed on the handset. This means that the user has to download the software before using the system and that the provider of the situated display has software that is compatible with every handset that people might be using when interacting with the system. Clearly, installing client software is not a viable route. What options then remain open?

SNAP AND GRAB

After working through many prototypes and synthesizing the best aspects of the systems that had gone before, we eventually developed an interaction technique which we call 'Snap and Grab' – see Figure 1. In essence, it allows users of camera-phones to select media for download by taking a photograph of the portion of the screen that relates to the media they want. The photograph is then sent, via BlueTooth, to the public display. On arrival the image is processed and the corresponding media is sent back to the user's handset as a Bluetooth object (this is possible as the image sent by the user to the display has the device's unique Bluetooth address embedded within). In order to explain the technique more fully, we will compare it with Hermes[2], one of the most widely researched and familiar systems in this field.

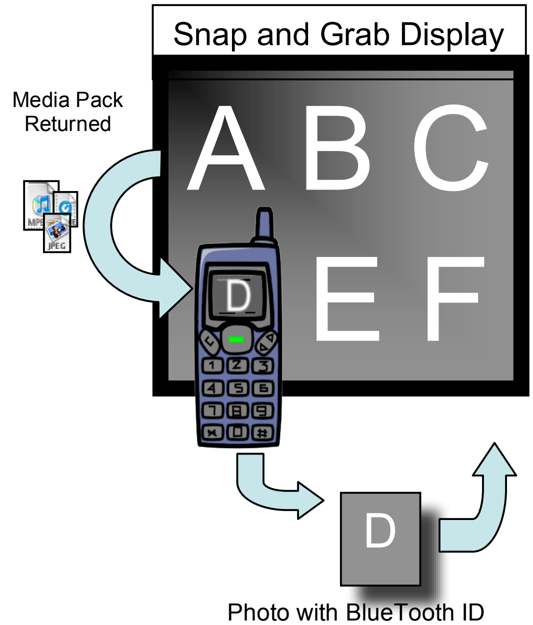

Figure 1 – User is interested in item 'D' so takes a photo of it and sends it to the display. The display sends back an MMS over BlueTooth containing the media related to 'D'.

Comparative Systems

Introduced by Cheverst et. al. [2], the Hermes photo display [2] was deployed in a semi-public office environment and allowed users to contribute photos to the display via their mobile phone. In a second activity, users were also able to select photos for download from the display onto their mobile phone. A typical Snap and Grab configuration consists of a much larger display than that used in the Hermes project and is deployed in a public area. In a similar way, Snap and Grab users are again able to contribute or select and download various multi-media objects (not just photos) via their mobile phones.

As with Hermes, the Snap and Grab design relies on the BlueTooth personal area networking (PAN) protocol for the transfer of content between the user's mobile device and the display system. Furthermore, neither system requires softwre to be installed on the client device.

Selecting display content for download

The Hermes display utilises a display panel with touch-screen input capabilities. A user is able to quickly select the photo they wish to download to their mobile phone. A drawback to this approach is that the system has no way of determining the phone's BlueTooth address. The Hermes designers solved this problem by prompting the user with a list of nearby BlueTooth device names and requiring them to select their device from this list.

The Snap and Grab design approaches display object selection through a camera-phone based interaction. In recent years significant research has been conducted on camera-phones as input devices [1,5,9]. Of particular interest are camera-phone interaction techniques that include visually tagged display objects where the visual tags encode machine-readable information, such as the ShotCode system[3]. Snap and Grab visual tags are embedded within display images and when processed, are used to determine the X-Y co-ordinates of the received image with respect to the display. The image co-ordinates are used to determine which display object the user was pointing at (selecting). This approach is a simplification of the work presented by Rohs et. al. [4]. Snap and Grab explored the use of standard digital photographs as visual tags as opposed to Rohs' use of visual codes. As mentioned by Rohs [4] and Shi [12], visual codes are visually unappealing and we wanted to create a solution that was visually appealing.

If a user submits a photo containing a visual tag, a server-side feature matching algorithm determines if a current display image matches the image submitted by the user. Feature matching algorithms such as those proposed by Lowe [9] have proved to be successful for this purpose.

A display object is selected in the following way:

1. The user activates the camera application on his/her phone and utilises the view-finder to ensure that a single display object is captured as part of a photo.
2. The photo is then sent via the native BlueTooth OBEX push service of the phone to the Snap and Grab display. In a typical scenario the user would execute a *'send via BlueTooth'* type command. The *'Snap and Grab'* device is selected from a list of active BlueTooth devices within range.

Receiving content from the display

Should the user's image contain a tagged display image and a positive match is made, the content associated with that display image is sent to the user. Once again the OBEX push service is used. BlueTooth based systems, such as Snap and Grab, would need to know the address of the client device before it is able to send any data objects. A client device address can be determined by:

1. Prompting users with a list of BlueTooth device names and allowing them to select their device. The Hermes [2] display utilises this technique with the aid of a touch screen display. Although practical, the approach was deemed unsuitable due to the high cost of a large touch screen displays, the delay associated with BlueTooth device scans and the interruption of other users that may be viewing the display.
2. Another approach would be to embed the BlueTooth device addresses within exchanged objects. This proved to be ideal. The Snap and Grab server simply extracts the BlueTooth device address from the image (Bluetooth object) uploaded by the user. The approach is beneficial in that it bypasses the need for device pairing and discovery. The Snap and Grab system goes further by not even requiring the client's device to be set to *'discoverable'*. This effectively hides the device during scans and thereby prevents unwanted data objects being received.

Both Scott [8] and Rohs et. al. [4] have utilised an embedded approach but in both cases the visual code was processed by software on the client device and the BlueTooth address of the target device extracted. This achieves the goal of avoiding any device scans (device discovery), making the interaction more efficient. As mentioned, Snap and Grab devices have no specific client software and thus no prior knowledge of the visual codes used by the display system, ruling out this approach. Instead, the unprocessed image is returned, containing a visual tag along with the device's *'return-address',* thereby providing the Snap and Grab server with all the information required for content delivery.

Media Packages

Media packages refer to a set of related multi-media objects that the system logically groups together via a BlueTooth address (for user generated packages) or title (for editorialised packages).

In a typical scenario, a media package related to a particular theme, such as London, or an advertisement for holiday accommodation, may be posted on the Snap and Grab display. The content may include various multi-media objects that are designed for mobile device platforms. Examples include Web pages (HTML file), mobile format video (GP3, MP4 etc.), audio file (MP3), contact details (VCard), calendar entries (VCal), image files (jpeg) and, more recently, text files (PDF and txt). The Snap and Grab system harnesses the multi-media capabilities of modern mobile devices for the purpose of reviewing a downloaded media package or creating/ authoring content that will be submitted to a media-package for others to download.

Submitting content to the display

An important goal of the Snap and Grab system design was to allow editorial and user generated content submission. This was achieved through the creation of two content submission interfaces. Firstly, the editorial interface was aimed at content providers that wish to author media-packages for whatever purpose. The content provider would purchase screen collateral and a time slot for when the download key (visual tag) for their media package would be displayed. The second submission interface would allow user generated content (UGC) to be submitted, via BlueTooth OBEX services, into a Snap and Grab media package. The Snap and Grab system is able to generate a download key from any image submitted as UGC. Subsequent media sent by that user, such as videos, VCards etc, is added to the media package and the associated download descriptor. When the system displays the UGC download key, other users may 'Snap and Grab' the media package left by the original user.

Remote content management

The Snap and Grab design utilises a Web-based architecture, thereby ensuring that the content on each display can be managed via a remote Web server. A content management service enforces various remote and local system policies. These policies cover issues such as user content moderation or identifying local content with broader relevance. In addition, these policies govern screen collateral assignment, visual tag sizing and time-slot management.

Possible Usage Scenario

To illustrate usage, consider the following scenario:

> Imagine you are a musician and wish to advertise the next show you are playing in. Snap and Grab presents the artist with two ways of advertising this event. Firstly, should his manager/agent wish to purchase some screen collateral, a professionally designed and maintained media-package may be presented to the public across multiple Snap and Grab displays. Audio samples, video, photos, album details, show dates, calendar entries (VCal) may all be made available. Secondly, the artist may approach a single Snap and Grab deployment and upload media that was authored on his mobile device onto the display system. Again live or recorded audio clips, videos, photos, calendar entries (VCal) or even his contact details for ad-hoc bookings (VCard) may be provided. Members of the public are able to download artist media packages, reflect on the content and hopefully support the artist in the future.

EVALUATION

To date we have only conducted informal ethnographies on a sample system deployed in a semi-public area within the building where the system was developed. As a measure of the technique's success, we have yet to complete a live demonstration as members of the audience 'get' the design instantly and start snapping and grabbing before we can give a full account. These reactions have now prompted us to go into full scale trials both in the developed and developing world.

CONCLUSION

In this paper we presented a new interaction technique that we believe overcomes the problems inherent in existing interactions between public displays and cellular handsets. The system is truly walk-up-and-use, requiring no client software or complex interaction with the display – if the user knows how to send and receive files via BlueTooth, we believe they will be able to use this system.

ACKNOWLEDGMENTS

This work was supported through a visiting researcher grant and internship from Microsoft Resarch in Cambridge.

REFERENCES

1. Rafael Ballagas, Michael Rohs, Jennifer G. Sheridan, and Jan Borchers. *BYOD: Bring Your Own Device*. In Proceedings of the Workshop on Ubiquitous Display Environments, Ubicomp 2004.
2. Cheverst K., Dix A., Fitton D., Kray C., Rouncefield M., Sas C., Saslis-Lagoudakis and Sheridan J., Exploring BlueTooth based Mobile Phone Interaction with the Hermes Photo Display. *In Proceedings of the 7th international conference on Human computer interaction with mobile devices and services,* ACM press, 2005, 47-54.
3. www.shotcode.com Last visited 28/2/07
4. Ballagas R., Rohs M. and Sheridan J. Sweep and Point & Shoot: Phonecam-Based Interactions for Large Public Displays. *In the extended abstracts of Human factors in computing systems (CHI2005)*, ACM press, 2005, 1200-1203.
5. Rohs, M., Sheridan, J. G. and Ballagas, R. Direct Manipulation Techniques for Large Displays using Camera Phones. *Proceedings of 2nd International Symposium on Ubiquitous Computing Systems* (USC2004), November 8-9, Tokyo, Japan.
6. Toye E., Sharp R., Madhavapeddy A., Scott D., Upton E. and Blackwell A. *Interacting with mobile services: an evaluation of camera-phones and visual tags.* Personal and Ubiquitous Computing, 11(2), 97-106. 2007.
7. Rohs, M. Real-world interaction with camera phones, *Proc. UCS.* IPSJ Press (2004).
8. Scott D., Sharp R., Madhavapeddy A. and Upton E. Using visual tags to bypass BlueTooth device discovery. Mobile Computing and Communication Review, 9(1), 41- 53.
9. David G. Lowe, *Object recognition from local scale-invariant features***,** International Conference on Computer Vision, Corfu, Greece, 1150-1157, 1999.
10. Feldbusch F., Paar A., Odendahl M. and Ivanov I. The BTRC Bluetooth remote control system. Personal and Ubiquitous Computing, Springer-Verlag, 7, 102-112. 2003.
11. BlueCasting a hit for Coldplay fans. Available at: www.mobileburn.com Last visited 28/2/07.
12. Jiang H., Ofek E., Moraveji N. and Shi Y. *Direct Pointer: Direct manipulation for large-display interaction using handheld cameras*. Proc. CHI 2006, Montreal, 1107-1110, 2006.