

# Physical-Virtual Linkage with Contextual Bookmarks

<sup>1</sup>Niels Henze, <sup>2</sup>Enrico Rukzio, <sup>3</sup>Andreas Lorenz,

<sup>4</sup>Xavier Righetti, <sup>5</sup>Susanne Boll

<sup>1</sup>OFFIS - Institute for Information Technology, Oldenburg, Germany

<sup>2</sup>Computing Department, Lancaster University, UK

<sup>3</sup>Fraunhofer-Institute for Applied Information Technology FIT, Germany

<sup>4</sup>VRlab, EPFL, Switzerland

<sup>5</sup>Media Informatics and Multimedia Systems, University of Oldenburg, Germany

henze@offis.de, rukzio@comp.lancs.ac.uk, xavier.righetti@epfl.ch,

andreas.lorenz@fit.fraunhofer.de, susanne.boll@uni-oldenburg.de

## ABSTRACT

In our everyday life we often see objects or persons and are aware that there are related digital services such as an online ticket service when seeing a poster advertising a concert. Currently it is a rather time consuming activity to find the related information. Using our Contextual Bookmark system a user can define a snapshot with her mobile phone consisting of a picture, time stamp and location. Such a bookmark can then be stored on the mobile phone, exchanged with friends and in particular be used to access related videos, web pages, and other services. This helps the user to bridge the gap between the virtual and the real world in order to use related services. Combining content and context analysis objects are recognized without any visual markers or electronic tags. We would like to demonstrate our system based on a nomadic usage scenario in which a person defines a Contextual Bookmark of a movie trailer, buys the corresponding movie, plays the movie on a TV, and exchanges the bookmark with a friend.

## Categories and Subject Descriptors

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities, H.5.2 [Interfaces and Presentation]: User Interfaces - Interaction styles; D.2.2 [Software Engineering]: Design Tools and Techniques - User interfaces.

## General Terms

Design, Human Factors.

## Keywords

Mobile interaction, physical-virtual linkage, physical interaction, contextual bookmark, content analysis.

## 1. INTRODUCTION

Mobile phones are the first truly pervasive interaction devices that

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*MobileHCI 2008*, September 2–5, 2008, Amsterdam, the Netherlands.

Copyright is held by the authors/owners.

are so far mainly used for calling, SMS or web browsing. Beside this we saw in the last years a huge interest in industry and academia in using these devices for interactions with people, places and things in the real world [2]. The primary motivation for this is to access mobile services related to such objects. Some Asian countries and in particular Japan play a pioneering role in this area when looking on location based services (e.g. i-area [3]), marker based interactions (e.g. QR codes [4]) and NFC based services (e.g. NTT DoCoMo i-mode Felica [5]). These services were already introduced a few years ago and are now widely used. One important disadvantage of these approaches is that the objects have to be tagged in some way, which either disturbs the object's visual appearance or requires the installation of electronic tags. In addition, the user needs certain knowledge about how to interact with these tags (e.g. visual marker or NFC/RFID tag).

This demonstration paper presents a system in which a user can define Contextual Bookmarks using her camera phone to access information related to a certain object or situation. More specifically a Contextual Bookmark consist of a picture of an object, a related time stamp and location information. These bookmarks are stored on the mobile phone, can be exchanged with friends and can in particular be used to access related mobile services. An example for the latter is for instance accessing a webpage of a concert ticket shop on the mobile phone after creating a Contextual Bookmark of a poster advertising a certain concert.

The Contextual Bookmarks system helps users to overcome the gap between the physical and the virtual world, which currently hinders users to access already existing services that are related to objects we see in our everyday life. The advantages of our approach when compared with others are that the corresponding objects do not have to be tagged with an NFC/RFID tag or a visual marker. Unlike other approaches our system combines different information such as a picture, time, and location in order to narrow the search-space for the vision based object recognition. A further advantage is that the definition of a Contextual Bookmark requires no specific knowledge about a certain interaction technique besides taking a picture with the camera phone. In addition to that allows the Contextual Bookmarks concept the interaction with a variety of objects and persons. The current implementation supports interaction with services related to movie trailers and advertisement posters. We currently work on the recognition of persons and will use this to open a related webpage on the mobile phone.

We envisage the presentation of an implementation of such a Contextual Bookmark system and its usage based on nomadic usage scenario. Here a person takes a Contextual Bookmark of a trailer shown by a public display, uses the bookmark to buy the movie related to the trailer, watches the movie on a remote TV, and exchange the bookmark with a friend. The demonstration we plan to show at Mobile HCI 2008 will be related to this scenario and will show all the necessary interactions.

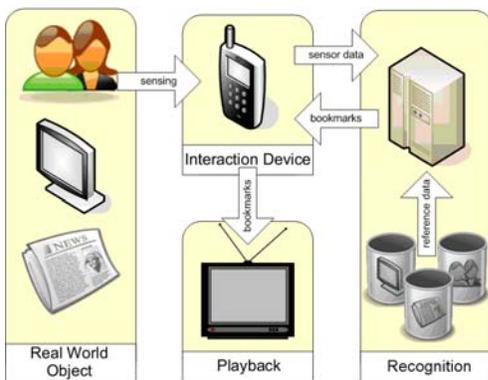
The paper is organized as follows. The next section explains the implementation of the Contextual Bookmarks system in detail. Following this we outline how the application and demonstration will look like and what an attendee will see when approaching our booth. We conclude this paper with a summary and an outlook on our future work.

## 2. CONTEXTUAL BOOKMARKS

A Contextual Bookmark is defined as a combination of a snapshot of a physical object taken with a mobile device and meta-information about the content related to this physical object. The snapshot comprises context information acquired from the environment such as time and location. Furthermore, the user's preferences and intentions may be covered. The meta-information about the physical object basically results from content analysis comprising a description of the content and links to related services. The architecture of a Contextual Bookmark system is outlined below. Afterwards the implementation of a specific subset of the concept related to our scenario is described.

### 2.1 Architecture

The concept of Contextual Bookmarks provides users with services related to physical objects. One aim of the system is to provide a flexible architecture that can integrate several recognition techniques for various physical objects and their related digital services. It consists of four basic building blocks shown in Figure 1.



**Figure 1. Central components of the Contextual Bookmark system and their interaction.**

Real world objects, or physical representations, addressed by the current implementation are videos shown on public displays and posters. Further considered examples are persons, documents, and sights. These representations can be bookmarked using a mobile device that collects information about the latter to retrieve services related to the real world object. Therefore, the object must initially be recognized. This task is performed by a dedicated recognition server that currently identifies posters and

videos but can be extended to additional recognition techniques. Consequently, a bookmark, containing information about the object and links to according services, is transmitted back to the mobile device. These services can be used on the mobile device itself, on external playback devices such as a home television, but can also be exchanged with other mobile phones.

### 2.2 Implementation

To enable users creating bookmarks using a mobile device, a Java ME (MIDP, CLDC 1.1) application running on recent camera phones (e.g. the Nokia N95 equipped with an integrated GPS receiver and a high resolution camera) was implemented. The application provides the user interface to access the intended functionalities. While aiming with the camera at a poster or a public display showing videos, a user selects the create bookmark function from the application's menu. On the phone's display the user can see the view of the camera as shown in Figure 2. When pressing the shoot soft key, a photo is created and annotated with additional contextual information (e.g. time and location). Two examples of created photos are shown in Figure 3. The photo as well as contextual information is transmitted to a recognition server via a TCP/IP connection using a wireless network (e.g. a 3G connection or Wi-Fi).



**Figure 2. User aiming at a display with a Nokia N95 to create a Contextual Bookmark.**

Hence, after having received the data collected by the phone, the recognition server is able to determine which physical object has been photographed by comparing the received data with its internal object description. For this purpose it is necessary to analyze and process the data received from the phone. The quality of the exemplary photos shown in Figure 3 is high, however, the poster appears distorted and a shadow darkens its centre. In addition, the objects are often surrounded by the street setting and it can not be ensured that the photo is in landscape or portrait format. Furthermore, objects might be rotated, partially occluded, and varied by lightning conditions. Thus, an algorithm is needed that is robust against the faced distortion. SIFT keypoints [6] are invariant or robust against most faced distortion such as scale, rotation, and light conditions.

Object descriptions including SIFT keypoints are extracted from posters and videos in advance. The extracted object descriptions are annotated with metadata such as a name, a description, and links to further services. Post processing of extracted SIFT keypoints reduces redundancies of features and improves



**Figure 3. Photo of a poster advertising the Simpsons movie and photo of a public display showing a video.**

robustness of the recognition. To compare a photo with a poster or video, SIFT keypoints of the photo are compared to SIFT keypoints of the poster or video. A Best-Bin-First heuristic [7] on a KD-tree is used to reduce the complexity. In about a second, the approach is able to compare a photo with a few minutes of video or a few dozen posters in about a second. In order to address a higher number of videos and posters, the number of considered object must inevitably be further reduced.

Taking context information recorded by the mobile phone into account is key to reduce the number of addressed objects. For example, when the location is known, the search-space is reduced to objects in the user's surrounding. Time information can be used for the same purpose, in particular for videos, where only the visible videos at the given timeframe will be analyzed. According to the recognized object, a Contextual Bookmark is created using the object's metadata. This bookmark is then transmitted back to the user's mobile phone.

The mobile phone receives the bookmark and the user can look at the related information such as the name of the advertised movie and its thumbnail. Multiple bookmarks can be created, collected, and managed on the mobile phone. In addition, the user can select services linked by the bookmark. When activating a linked video, the user can choose between playback on the device itself and playback on a remote Home Theatre PC (HTPC). The application provides usual media playback controls (e.g. play and pause). Playback functionality is encapsulated in an integrated software component. The current implementation handles links to videos and trailers. It is realized as a software player derived from the Java Media Framework, which ensures platform independency to be used on the mobile phone as well as on Home Theatre PCs at home. The playback component uses a toolkit that abstracts from concrete input devices and thus the interface on the phone controls playback on the phone as well as playback on the remote HTPC.

Sharing bookmarks between users of the system is implemented by exchanging a bookmark's unique identifiers via Short Messaging Service. From the application, a phone number, possibly from the phone's address book, can be selected. Since, the application handles the received messages automatically, the underlying process is transparent to the addressee. Finally, a message appears on her device announcing the received bookmark.

### 3. DEMONSTRATION SETUP

The demonstration will show the interaction with the Contextual Bookmark prototype. It will be presented in a supervised hands-

on setup. The system will be capable to recognize two posters that present our ongoing research. In addition, pre-processed videos shown on a computer screen can be used to create Contextual Bookmarks. Another computer will serve as a HTPC that can be used to shows bookmarked videos.



**Figure 4. The Contextual Bookmark prototype on a Nokia N95. Left shows the view of the camera, centre a list of created bookmarks, and right the details of a bookmark.**

After a very short introduction visitors will be able to create Contextual Bookmarks of posters and videos using the mobile application. The mobile device (i.e. Nokia N95 phones) running the developed application will be provided to visitors. Figure 4 shows screenshots from the N95-display: The camera image ready to take a photo, browsing of already taken bookmarks, and the details of a selected bookmark.

Like the user interacting with a basic setup of the demonstration



**Figure 5. User interacting with a basic setup of the Contextual Bookmarks system comprising a display and posters.**

shown in Figure 5, visitors can experiment with the system e.g. by taking pictures of our posters, create the bookmark including recognition of the poster-related content, and share bookmarks with other users. Activating the exchange bookmark function visitors can send Contextual Bookmarks to one of the other provided phones. In addition we will provide a video service for watching trailers related to the respective bookmark either on the phone or on the HTPC. A digital rights management mock-up service simulates the process of buying the content for viewing the according entire video.

Interacting with the video player the user can experiment with the software-based remote control that is integrated in the mobile application of the Contextual Bookmarks system.

## 4. CONCLUSION

In this paper, we presented our approach of Contextual Bookmarks that allows capturing objects in the real world using a camera phone and mapping a taken photo to a digital concept that is related to the real object. The envisaged demonstration covers the whole lifecycle of a Contextual Bookmark including creation, using, and sharing of bookmarks. It will show how users can easily bookmark real world objects and receive a mapping to digital services. Bookmarks can be exchanged and content be watched on the mobile device as well as on other device. Thus, the demonstration can provide a hands-on experience of our ongoing research.

The lifecycle of a Contextual Bookmark from creating to using and sharing bookmarks can be further extended. Creation will be extended by further recognition techniques to validate our architecture and address more types of real world objects. A promising area is face recognition to access people's website. Additional services can widen the use cases of Contextual Bookmarks. In particular, we work on the integration of a service to guide users to places advertised by bookmarks.

So far the prototype has only been demonstrated at small events. Thus, we aim to evaluate our system with users in order to compare it with existing approaches, to find issues in our current implementation, and to further analyze the user's needs. To get further insights and advice we highly anticipate the presentation of our concept to a larger group of HCI experts at Mobile HCI

## 5. ACKNOWLEDGMENTS

This work was performed as part of the Intermedia NoE [1], funded by the European Union as part of the Framework VI

research program. The authors wish to express their gratitude to the members of the Intermedia Consortium [1] for their valuable contributions.

## 6. REFERENCES

- [1] Intermedia Network of Excellence, <http://intermedia.miralab.unige.ch/>
- [2] Tim Kindberg; John Barton; Jeff Morgan; Gene Becker; Debbie Caswell; Philippe Debatty; Gita Gopal; Marcos Frid; Venky Krishnan; Howard Morris; John Schettino; Bill Serra; Mirjana Spasojevic: People, places, things: web presence for the real world. In: *Mobile Networks and Applications*, 7 (5), pages 365–376, ISSN 1383-469X, 2002.
- [3] NTT DoCoMo i-area: Location Based Services, [www.nttdocomo.co.jp/english/service/imode/site/i\\_area.html](http://www.nttdocomo.co.jp/english/service/imode/site/i_area.html)
- [4] QRCode, <http://www.qrcode.com/>
- [5] NTT DoCoMo i-mode Felica <http://www.nttdocomo.co.jp/english/service/imode/make/content/felica/>
- [6] David Lowe: Distinctive Image Features from Scale-Invariant Keypoints. In: *International Journal of Computer Vision*, 60(2), pages 91–110, 2004.
- [7] Jeffrey S. Beis; David Lowe: Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In: *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 1000–1006, 1997.