# Contextual Bookmarks

[1] Niels Henze, [2] Mingyu Lim, [3] Andreas Lorenz, [4] Michael Mueller, [5] Xavier Righetti, [4] Enrico Rukzio, [3] Andreas Zimmermann,
[2] Nadia Magnenat-Thalmann, [6] Susanne Boll, [5] Daniel Thalmann
[1] OFFIS - Institute for Information Technology, Oldenburg, Germany
[2] MIRALab - University of Geneva, Switzerland
[3] Fraunhofer-Institute for Applied Information Technology FIT, Germany
[4] Computing Department, Lancaster University, UK [5] VRlab, EPFL, Switzerland
[6] Media Informatics and Multimedia Systems, University of Oldenburg, Germany
henze@offis.de, {mingyu.lim, thalmann}@miralab.unige.ch, {xavier.righetti, daniel.thalmann}@epfl.ch, {m.mueller, rukzio}@ comp.lancs.ac.uk, {andreas.lorenz, andreas.zimmermann}@fit.fraunhofer.de, susanne.boll@informatik.uni-oldenburg.de

## ABSTRACT
The role of digital information in everyday life divides out activities in interacting with the physical and the digital world. There is no connection between these worlds that is easily accessible, even though physical objects, persons, and real world events often have digital counterparts. The physical reality is actually overlaid by an additional virtual or digital layer. As we are acting in the physical and the digital world it is desirable that we can use links that bridge the gap between both worlds. In this paper we describe our approach to narrow this gap. Starting from a scenario that shows the demand for such connections, we present an architecture that allows users to bookmark specific situations. Based on these contextual bookmarks the user can request additional digital information. Our first prototype enables the user to bookmark content shown on public displays by taking photos of the display using a mobile phone. Our system combines content analysis of the photo with context information such as position, creation time, etc. in order to form the basis to establish a link to the digital world. The presented architecture will serve as a flexible solution to find and integrate further connections between the physical and the digital world.

## 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
Algorithms, Design, Human Factors

## Keywords
Mobile interaction, user interface, contextual bookmark.

## 1. INTRODUCTION
With the advert of the era of ubiquitous computing environments, it has become common for people to have their own mobile devices (e.g. phones, PDAs, and multimedia players) and access the Internet not only at home and in the office but also from hot spots and other wireless networks offering Internet connection outside buildings. All these facilities let us face a flood of digital multimedia services through which we interact in both the digital and the real world because physical objects, public events, and people in the real world could have digital counterparts. For example, various advertisement posters in train stations or bus stops can also be found on the Internet in a digital form. We can not only enjoy a film in a cinema in the real world but also by buying and downloading the corresponding file as a digital item. However, today there is no easy accessible link that connects physical objects to digital items.

If there is a public event, which we are interested in and we want to remember, we need to record the information by writing down a memo in a calendar or a personal scheduler in a PDA or a mobile phone, which generates high demands on the people. Therefore, this paper proposes a contextual bookmark approach that aims at bridging the gap between the real and the digital world.

In this paper we introduce the concept of *contextual bookmarks* in the subsequent section. Then, section 3 motivates our approach through an illustrative usage scenario described. Section 4 describes related work regarding the identification of real-world objects. Section 5 discusses the architecture of our approach and the current state of our development efforts. Finally, section 6 concludes the paper with further work.

## 2. CONTEXTUAL BOOKMARKS
We define a contextual bookmark 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. According to and extending our scenario described below, a physical object can be a public display, an advertisement poster at a train station, or an exhibit of a museum. The meta-information about the physical object basically results from content analysis and also comprises context information acquired from the environment such as time, location or temperature. Furthermore, this meta-information may cover information about the user's preferences and intentions.

A contextual bookmark represents a physical object in a digital form. A user of our system uses his or her personalized wearable device to record and define a contextual bookmark. Currently we use a Nokia N95 which is one of the most recent camera-equipped mobile phones that is equipped with additional sensors (e.g. GPS) for acquiring context information. In a later phase this device will be extended by additional input controls (e.g. wearable buttons), output capabilities (e.g. headphone, see-through glasses), wearable storage and various sensors (e.g. a light or temperature sensor).

The contextual bookmark provides a handle to the content and the respective situation, in which it has been recorded. Browsing the list of contextual bookmarks on the mobile device enables the user to exploit services that are related to the context, in which the bookmark has been defined. Once the user selects a service associated with a contextual bookmark, he or she can, for example, access the content of the contextual bookmark via the mobile device or via a nearby output device depending on the current context and preferences (e.g. public display, TV, radio, laptop, etc.). Furthermore, the user gains access to more detailed information about the content, which was not available in the situation in which the contextual bookmark has been taken.

Although our first prototype system is specific for a movie trailer bookmark as a useful example, it can also be adapted to contextual bookmarks of various other types of physical objects such as posters advertising upcoming events, songs played on the radio, and even people who were met at a conference or in a business meeting. As such, our approach, instead of providing a solution for a single scenario, aims to support a flexible solution that enables to integrate various sensor data, media types, means of data access, and matching algorithms for different scenarios.

# 3. SCENARIO

This section presents a scenario which shows the need for and usefulness of a mobile contextual bookmark application.

Jim is on a business trip in Berlin and sits in the metro. The train is equipped with small advertising screens; suddenly, the trailer of the latest Tarantino movie pops up on the screen. Tarantino is one of Jim's favourite movie writers and he does not want to miss this movie again. He points at the screen with his mobile phone and bookmarks the trailer as a movie he likes to see.

During the day, Jim spends endless time in business meetings discussing the next release of the company's software. It is already late afternoon, when the meetings finish and he is on the way back to the hotel. He updates his digital calendar because business meetings are finished for today. The system automatically checks the cinemas near Jim's hotel and asks him if he wants to go to the movies tonight at 8:30pm to watch the movie he bookmarked in the train. Jim remembers that he has an appointment with a former class mate tonight and refuses with a sigh.

The next day, Jim is back at home. He takes his mobile device and browses the bookmarks he had taken in Berlin: There are some sights like Brandenburg Gate, the Berlin Wall and Federal Chancellery, the presentations of the business partners, and some advertisements he bookmarked to remember. There is also the movie trailer he had bookmarked in the train. The bookmark provides Jim with a link to download the trailer. When he watches the trailer he realizes, that he still had not have time to watch the movie. He decides to invite some friends to watch the movie on his home cinema system. Jim sets up everything, gets some beers and calls some friends. He selects the bookmark and advises his mobile phone to get a license for downloading the movie to his set-top box. At home, the device retrieves his personal movie pay-per-view account from his PC, requests a license from the database and asks Jim to acknowledge the payment. He confirms and the license is available.

Because some friends were late that night, they are already behind the schedule when they start watching the movie at 8:50. Jim selects the movie license; the system automatically determines the best fitting output screen and downloads the movie in highest quality to the set-top box of his home cinema system. The movie starts and everyone enjoys watching it. Unfortunately, Joe has another appointment and as they started late he cannot keep watching the movie at Jim's place. Jim acknowledges sharing the license of the movie with Joe. Joe pays a small fee for the last 20min to finish the movie on the train with his mobile phone. Joe's device requests the license from Jim's phone and downloads the last 20 min in lower resolution that fits with the capabilities of Joe's phone. Joe takes it with him while the rest of the group finishes the movie at Jim's place. They all had a great time.

# 4. RELATED WORK

There have been approaches to record and retrieve information based on markers or image analysis. However, they require external markers tagged on real objects or sophisticated content-based image analysis. These concepts are not sufficient for understanding the exact meaning of media items, which users really want in the current context. Some projects related to our contextual bookmark approach are presented in the following.

Hansen et al. introduce the term “mixed interaction space” to distinguish camera-based interaction from other types of sensor-based interaction on mobile devices [1]. Indeed, physical space plays an important role in camera-based digital interaction that is controlled by both movement and orientation of a mobile phone in the space. But there are many ways for mobile phones to interact with their environment, especially with large situated displays. Ballagas et al. classified more than 15 different interaction techniques using various communication technologies and interaction paradigms [2]. Concurrently, Ailisto et al. have established a detailed comparison table between four potential commercial technologies for physical selection [3], which comprise Visual Code, IrDA, RFID and Bluetooth.

However, Rukzio et al. note that there is very little support to build such kind of systems [4]. Hence, they present an architecture named “Physical Mobile Interaction Framework” which is based on existing standards such as the Java 2 Micro Edition (Java ME) and the Contactless Communication API. The authors plan to support all relevant interaction techniques between the device and the object by providing abstractions for the programmer, hiding technical details of the communication.

Attaching markers to real world entities makes it possible to find corresponding digital entities. Simple implementations of this approach are barcodes developed by Silver and Woodland back in 1948. Barcodes can only store a very limited amount of data but nonetheless are used to identify physical objects [6]. Thus, systems using more sophisticated markers have been developed. Examples are 2D barcodes [7] radio-frequency tags [13] and infrared tags [14]. However, all marker based approaches need a marker that is somehow attached to the object. If a user want to select a marker to acquire information about the corresponding entity the user or the device must find this marker. This is often not desirable because visual markers must be small, with long range radio frequency based approaches the user's device can not decide which marker was selected, and infrared tags are outshined by the sun. In addition, all markers can be masked by other people or objects and for some real world entities, like human beings, it is not even feasible to attach markers.

Raj et al. present a mechanism that allows users to implicitly download content available on public displays [5]. The ContentCascade framework enables the user to download either summary information or movie clips via Bluetooth. They suggest measuring the interest of the content by analyzing the user’s behaviour and the length of the interactions with the public display. Cheverst et al. developed and evaluated the Hermes photo display, a prototype that enables users with a suitable mobile phone to both send and receive pictures via Bluetooth [8]. The display uses Sun’s Bluetooth API to connect to the mobile phone and utilizes OBEX push to transmit the files. The advantage of this approach is that it requires no additional applications to be installed on the mobile phone. The Hermes door display [9] is another prototype developed by Cheverst et al. and installed on the outside of ten offices’ doors in Lancaster’s Computing department. The Hermes door displays give offices’ owners and visitors the ability to leave notes on the door using their mobile phone. During the 24 months of its use, more than 6000 notes were added. All three Bluetooth based systems need additional hardware infrastructure integrated in the public display and shares the limitation of marker based approaches described above.

Among other projects, we can also cite PhoneGuide [10], an enhanced museum guidance system that uses camera-equipped mobile phones and on-device object recognition. Föckler et al.’s main technical achievement is a light-weight object recognition algorithm that is realized with a single-layer perception neuronal network on the device itself. According to their results, over 90% of the photographed museum exhibits can be recognized without the addition of passive or active reference markers. The limitation of Föckler’s promising approach is the small number of recognizable objects and that it does not work for all types of real world objects. We believe that both can significantly be improved by taking the user’s context into account.

## 5. ARCHITECTURE AND IMPLEMENTATION

We are currently working on a first prototype which realizes a small subset of the functionality described above. The prototype will consist of multiple computers with attached large displays that show different movie trailers. Persons can use their mobile phones to define a contextual bookmark by taking a photo of the displays playing the trailer. The bookmark can be transmitted to further computers to play the connected movies.

The aim of the prototype is to provide a flexible architecture that enables the integration of further matching techniques and to evaluate the proposed interaction. The general architecture of the system integrating further extensions is outlined in Figure 1. On base of this architecture we can analyse context and content of a given snapshot of the user’s situation. In the following subsections we describe the lifecycle of a contextual bookmark from sensing the environment through matching snapshots of the sensed context to exchanging bookmarks.

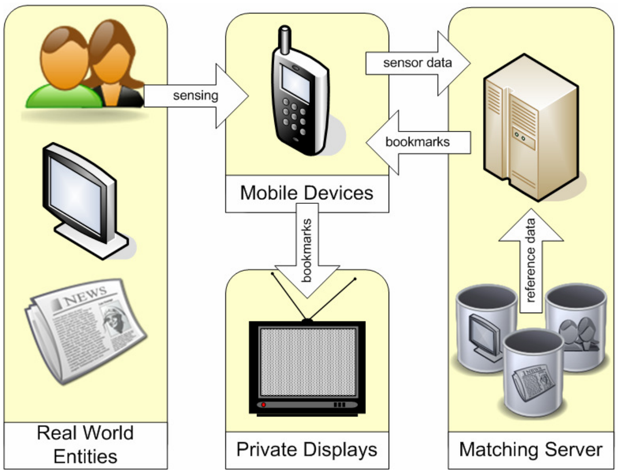


**Figure 1. Generic architecture of the prototype currently under development**

### 5.1 Sensing the Environment

When the user invokes the bookmark function on his or her mobile device a snapshot of the current context is created. The first prototype will be based on mobile phones (e.g. the Nokia N95) and Java ME, the most common platform for mobile phones. After starting the system the user can focus the mobile phone’s camera at the public display. When the user activates the bookmark function a photo is created and enriched with the actual position using the integrated GPS receiver and a timestamp. This information represents a snapshot of the user’s context at this point in time.

To ease the integration of additional sensor information and react to the respective devices individual capabilities, for instance not all mobile phones have an integrated GPS receiver, the snapshot contains self-describing sections each containing a distinct feature of the context. Thus, further extensions, e.g. recording audio, sensing the user’s emotions, or determine his or her tasks using digital agendas can be integrated easily.

### 5.2 Matching

The snapshot is transferred to a server that retrieves links to related digital items and services. Analysing the snapshot of the context is a processing power and memory consuming task. Thus, the architecture envisages implementing the matching of context to digital items and services on a separate server. The server offers its services via a web service interface. The first step of the matching process is an analysis of the kind of context data included in the received snapshot. According to this, the data is delivered to matching processes that are able to process parts of the snapshot.

For the first prototype we will develop a single matching process that receives a photo taken by the mobile device to find the according video shown on a public display using the creation time and the user’s position. The matching process needs all videos potentially shown on one of the public displays. Scale Invariant Feature Transform (SIFT) keypoints [11] are extracted from the video frames and stored in a repository. When a photo is delivered all public displays nearby the provided position are asked for a list of videos played at a timeframe around the snapshot's time of creation. The resulting video frames are compared to the photos SIFT keypoints using the Best Bin First algorithm [12].

The same matching process can be used to match photos to other visual data, e.g. posters or presentation slides. Additional matching processes will be developed, for example, for face recognition or music matching. We estimate that with the use of additional context information and users preferences the matching precision could be enhanced, for example, by assuming that it is more likely that users interested in specific topics bookmark items belonging to this topic more often.

Each item that could be a result of a match is annotated with a list of links to digital information and services. These lists are returned to the mobile device where the lists are combined with the snapshot of the context to create a contextual bookmark.

### 5.3 Exchanging Bookmarks

When the contextual bookmark is created from the snapshot of the context and links to digital information and services the user can use the bookmarks to activate the links on his mobile device, transfer them to private devices, for instance to show the movie belonging to a bookmarked trailer on his TV, or send the bookmark to friends and colleagues.

To send contextual bookmarks to another user we will provide a server based solution that enables the users to virtually find each other. To ease the handover to nearby devices and friends we will also implement a Bluetooth based solution that scans the surrounding for other devices running our system and shows them to the user. The users can select nearby devices to easily exchange bookmarks. This solution can also be used to for location based services, for example, to equip advertisement posters with the ability to proactively provide the environment with additional digital information.

## 6. Conclusion

In this paper we presented our vision of a system that narrows the gap between the digital and the physical world. Users can create contextual bookmarks using everyday devices like their personal mobile phone. We presented a scenario that motivates and justifies our work. Based on this scenario and on the analysis of the related work in this area, we outlined an architecture and its implementation that simplifies the integration of further matching techniques.

The idea presented in this paper can be extended by further matching techniques to address more types of media and real world situations. The overall aim is that the user can create contextual bookmarks of almost any real world item without much effort. For this purpose it is also necessary to address the interaction with the user. Therefore, we will explore how the user expresses the intended action when creating a contextual bookmark.

## 7. ACKNOWLEDGMENTS

This paper is supported by the European Community within the InterMedia project (project No. 038419). We thank all partners for sharing their ideas with us.

## 8. REFERENCES

[1] Hansen, T.R., Eriksson, E., and Lykke-Olesen, A., "Mixed Interaction Space – Designing for Camera Based Interaction with Mobile Devices", Conference on Human Factors in Computing Systems, pp. 1933-1936, 2005

[2] Ballagas, R., Borchers, J., Rohs, M., and Sheridan, JG, "The smart phone: a ubiquitous input device", IEEE Pervasive Computing Journal, vol. 5, pp. 70-77, 2006

[3] Ailisto, H., Korhonen, I., Plomp, J., Pohjanheimo, L., and Strömmer, E., "Realising Physical Selection for Mobile Devices", Mobile HCI 2003 Conference, Physical Interaction (PI03) Workshop on Real World User Interfaces, Udine, Italy

[4] Rukzio, E., Wetzstein, S., and Schmidt, A., "A Framework for Mobile Interactions with the Physical World", Proceedings of Wireless Personal Multimedia Communication (WPMC'05), 2005

[5] Raj, H., Gossweiler, R., and Milojicic, D., "ContentCascade incremental content exchange between public displays and personal devices", International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp. 374-381, 2004

[6] Lange, B.M., Jones, M.A., and Meyers, J.L., "Insight lab: An immersive team environment linking paper, displays, and data", Conference on Human Factors in Computing Systems, pp 550-557, 1998

[7] Rekimoto, J. and Nagao, K., "The world through the computer: computer augmented interaction with real world environments", ACM symposium on User interface and software technology, pp 29-36, 1995

[8] Cheverst, K., Dix, A., Fitton, D., Kray, C., Rouncefield, M., Sas, C., Saslis-Lagoudakis, G., and Sheridan, J.G., "Exploring Bluetooth Based Mobile Phone Interaction with the Hermes Photo Display", Conference on Human computer interaction with mobile devices and services, pp. 47-54, 2005

[9] Cheverst, K., Dix, A., Fitton, D., Kray, C., Rouncefield, M., Saslis-Lagoudakis, G., and Sheridan, J.G., "Exploring Mobile Phone Interaction with Situated Displays", PERMID Workshop, Pervasive, vol. 5, 2005

[10] Föckler, P., Zeidler, T., Brombach, B., Bruns, E., and Bimber, O., "PhoneGuide: Museum Guidance Supported by On-Device Object Recognition on Mobile Phones", Conference on Mobile and ubiquitous multimedia, pp. 3-10, 2005

[11] Lowe, D.G., "Object recognition from local scale-invariant features", Proceedings of the international Conference on Computer Vision, pp. 1150-1157, 1999.

[12] Beis, J. S. and Lowe, D.G., "Shape Indexing Using Appropriate Nearest-Neighbour Search in High-Dimensional Spaces", Proceedings of the 1997 IEEE Conference on Computer Vision and Pattern Recognition, pp 1000–1006, 1997.

[13] Want, R., Fishkin, K.P., Gujar, A., and Harrison, B.L., "Bridging physical and virtual worlds with electronic tags", Conference on Human Factors in Computing Systems, pp 370-377, 1999

[14] Want, R., Hopper, A., Falcao, V., and Gibbons, J., "The active badge location system", ACM Transactions on Information Systems, pp 91-102, 1992