Integrating a Multitouch Kiosk System with Mobile Devices and Multimodal Interaction

Simon Bergweiler
DFKI GmbH
Campus D3_2
66123 Saarbrücken, Germany
simon.bergweiler@dfki.de

Matthieu Deru
DFKI GmbH
Campus D3_2
66123 Saarbrücken, Germany
matthieu.deru@dfki.de

Daniel Porta
DFKI GmbH
Campus D3_2
66123 Saarbrücken, Germany
daniel.porta@dfki.de

ABSTRACT
We present Calisto, a service-oriented information kiosk system for public places, like museums or hotel lobbies. Calisto supports collaboration between multiple users. They can connect their mobile devices to the large public terminal and share interesting facts and media contents via intuitive multimodal interaction. The novel contribution of our work is a seamless combination of a touch-based kiosk system and mobile devices for accessing heterogeneous information services.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Design, Human Factors

Keywords: multitouch, kiosk system, interactive tabletop, multimodal interaction, mobile device, Spotlet

INTRODUCTION

In today's increasingly tight interlinked world, floods of information must be handled. On the one hand, modern kiosk systems enable intuitive touch- and gesture-based interaction and provide users with useful contextual information that normally resides in the system. On the other hand, a growing number of people carry a sensor-rich mobile device with large storage capacities as an interface to their digital life with them. However, interaction capabilities are often restricted by physical device limitations, e.g. small screen sizes. In order to combine best of both worlds, we implemented a service-oriented kiosk system that provides data exchange and access to information services at public places, like museums or hotel lobbies. Furthermore, the system allows for collaboration among users.

THE CALISTO SYSTEM

In general, Calisto, as an interactive and collaborative kiosk system, consists on the hardware-side of a customized 40'' Full HD multitouch terminal and via Wi-Fi connected mobile devices, see figure 1. On the software-side, the Calisto system hosts a service engine for retrieving information from external Web services, a device manager that handles connected mobile clients, a dialogue system that also enables multimodal interaction, and the actual GUI. The system deals with a set of structured metadata, e.g. Exif for images or ID3 tags for audio. These data structures are mapped to a unified internal representation that is used as internal annotation for an object.

Interacting with the Terminal-GUI

Each annotated visual media object is represented on the terminal GUI by a thumbnail, for audio files the album cover is used. The objects can be dragged-and-dropped and freely rotated by (multi-)touch gestures. Video/audio playback is triggered by simply tapping on an element.

The terminal GUI provides access to different information sources. Typically such information sources are implemented as complex Web services. Access to these information sources is encapsulated by dedicated interaction containers, called *Spotlets*. Hence, we denote the term Spotlet as a graphical metaphor for an orchestration of services. User interaction with a Spotlet is accomplished by simply dropping an annotated media object onto it. The Spotlet uses the underlying annotation of the dropped media object as an input for corresponding services.

For example, we implemented a *Music Search Spotlet* composing services from LastFM, MusicBrainz and SeeqPod. If a user drops a song of "Madonna" there, a similarity search is initiated. All results, song previews, are arranged around the Spotlet for further interactions. Other Spotlets encapsulate access to YouTube, Twitter and Freebase [2].

Integrating Mobile Devices

In the basic idea behind our approach, users can synchronize their mobile devices with the terminal via Wi-Fi and hence extend the small-screen device to a large tabletop surface. For this, we developed an Android-based mobile client application that allows the user for managing a shared media folder and exchanging media contents with the Calisto terminal. Technically, this is implemented as a RESTful service interface.

Once connected, references to the contents of the shared media folder are synchronized with a graphical representation of the device on the terminal, the *Device Spotlet*. Hence, the mobile client acts as a virtual media store that can be accessed by the user from the terminal.

ITS'10, November 7–10, 2010, Saarbrücken, Germany.

The user can transfer, e.g., a picture of his last visit of Berlin to the terminal GUI simply by selecting the picture on the mobile client and at the same time performing a Frisbee-throwing 3D gesture towards the terminal. This gesture is detected by interpreting the sensor data of the built-in accelerometer, as described in [1], and triggers an action that puts the user-selected media object on the terminal.

Figure 1: Calisto terminal with mobile devices

If the user found new media contents and information that he wants to take with him, he can intuitively drag-and-drop the media objects on the Device Spotlet. As a result, the media object is physically transferred to the user's mobile device by performing an HTTP/PUT request against the client-side RESTful service interface. Finally, the client's shared media folder is refreshed.

Supporting Collaboration

The Calisto system is not restricted to only one connected mobile client. Regarding the backend-side, the number is currently limited by the number of connections the Wi-Fi router can handle. However, the GUI implementation is tailored to two simultaneously connected mobile clients, which is sufficient for a proof-of-concept. So, if more than one user is connected to the Calisto system with a respective mobile client, they can collaborate with each other. Shared media objects owned by a user-1 can also be used as Spotlet input (and hence as input for a service request) via drag-and-drop by another user-2. A special use case is then the exchange of media objects across users.

For example in the hotel lobby scenario, two tourists interact with the system in order to find interesting places in Berlin. As depicted in figure 2, user-1 throws a picture of the Brandenburg Gate on the terminal GUI. User-2, who has not been there yet, can either drag-and-drop the picture to the *Map Spotlet* in order to see the exact location of the point of interest and optional directions by taking the user's current position into account. Alternatively, user-2 can drag-and-drop the picture on his personal Device Spotlet, causing the system to copy the picture to his mobile device.

Enabling Multimodal Interaction

In addition to touch- and gesture-based interaction, users can interact with the Calisto terminal system also via natural language. Having their personal mobile device, users can utter speech commands combined with optional deictic pointing gestures to media objects on the large terminal screen, e.g., "*Show me the Brandenburg Gate on the map.*" or simply "*Where is [simultaneous pointing gesture to the picture of the Brandenburg Gate] this?*" To transfer this utterance, the mobile client application implements real-time audio streaming (sending and receiving) capabilities. The speech input is sent as RTP packets to a speech recognizer that is connected to the Calisto dialogue system. The sent audio stream is interpreted, the backend processing is initiated and, as a result, the Calisto terminal displays the location of the Brandenburg Gate in the Map Spotlet. At the same time, an appropriate auditory response "*Showing the Brandenburg Gate on the map*" is sent back to the respective mobile client.

Figure 2: Frisbee and touch gesture interaction: user-1 (yellow) throws a media object on the terminal and user-2 (red) drags-and-drops the object on his Device Spotlet.

CONCLUSION & FUTURE WORK

We presented our first prototype of the Calisto system, which offers a new innovative way to interact with touch-based terminals for data exchange and access to different remote services at public places. We plan to conduct a series of user studies in order to evaluate the system concerning intuitive collaborative interaction and social aspects, regarding which kind of interaction or modality is preferred in which situational context.

ACKNOWLEDGMENTS

This research has been supported by the THESEUS Programme funded by the German Federal Ministry of Economics and Technology (01MQ07016). The responsibility for the content of this publication lies with the authors.

REFERENCES

1. Porta, D., Sonntag, D., and Neßelrath, R. 2009. New business to business interaction: shake your iPhone and speak to it. In *Proc. of the 11th Int. Conf. on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09),* ACM, New York, NY

2. Sonntag, D., Deru, M., and Bergweiler, S. 2009. Design and Implementation of Combined Mobile and Touchscreen-Based Multimodal Web 3.0 Interfaces. *Proc. of the 2009 Int. Conf. on Artificial Intelligence (ICAI'09)*, Las Vegas, NV, p. 974-979