

Pervasive Imaging: a Capture and Access Perspective

Renan G. Cattelan
U. São Paulo, Brazil
renan@icmc.usp.br

Valter R. Inácio Jr
U. São Paulo, Brazil
inacio@icmc.usp.br

Maria da Graça Pimentel
U. São Paulo, Brazil
mgp@icmc.usp.br

ABSTRACT

In this short paper, we present the pervasive imaging theme from the capture and access problem perspective. We present our findings and share our experiences on providing support for mobile automated capture, concerning annotated images and video, for field trip activities. We describe the use of our approach on two different scenarios: a visit to the zoo and a historical farm field trip.

Categories and Subject Descriptors

H.4.0 [Information Systems Applications]: General; H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems

General Terms

Ubiquitous computing

Keywords

Image capturing, capture and access applications, annotation

1. INTRODUCTION

As envisioned by Weiser [9], the *ubiquitous computing* paradigm poses a number of challenges concerning the proliferation of devices and their $n : 1$ relation with humans. The recent wave of camera phones, PDAs, Tablet PCs, digital cameras and other similar silicon gadgets that took over people's lives confirms such prediction.

Being usually widespread everywhere, available at hand and ready to use, these devices can be seen as ubiquitous capture agents, allowing the registry of everyday experiences in the form of digital artifacts such as images and short videos. In this context, it is worth investigating the development of simple, efficient and useful mechanisms to properly acquire, storage and make those registers available for review.

In previous work, we have developed capture and access applications for several domains. Inspired by *eClass* — built to capture information from a traditional lecture (slides, electronic ink, audio and video) toward automatically producing Web documents [1, 2, 8] — we have built *iClass* [3], which allowed us to investigate several design alternatives such as navigation recommendation and automatic linking [6].

Exploiting HP tablet PCs, we built *wiClass* [7], a mobile version of *iClass*, and *M4Note*, a MultiMedia MultiModal Annotation Tool [4, 5]. With respect to pervasive image capturing and sharing, both *wiClass* and *M4Note*, exploiting the integration capture images and pen-based interaction, allowing a flexible experience by using of mobile and portable digital devices (e.g. digital cameras, Tablet PCs and webcams).

In this paper, we present our findings and share our experiences on providing support for mobile automated capture, concerning annotated images and video, for field trip activities. We describe the use of our approach on two different scenarios: a historical farm field trip and a visit to the zoo.

2. DESCRIPTION OF THE SYSTEMS

Our prototypes are built with independent capture components responsible for grabbing information for specific media and devices. For instance, a whiteboard component in a Tablet PC allows a user to annotate on top of captured images. We also have capture components for video, audio, Web browsing and plain text. Another important component allows the online recognition of handwriting or special marks when used with the whiteboard component.

The capture components handle all the data being produced by users and generate XML documents. By combining the required components, a capture and access application can be easily and rapidly prototyped. Next we describe the *wiClass* and *M4Note* prototypes.

2.1 wiClass

wiClass [7] was built to observe and evaluate the impact of using mobile devices by instructors and students in traditional classrooms, lab sessions and field visits.

By using a Tablet PC and an electronic pen, the user can annotate the images with digital ink. The system allows the annotation on top of pre-prepared slides or on top of images captured and imported on the fly; the images can

be navigated forward and backward. Its also possible to capture audio from the tablet PC built-in microphone.

At the end of the capture, a XML document is generated and processed against XSL stylesheets to produce versions in derivative formats more appropriate for review, such as XHTML and SMIL. A Java applet allows the playback of the captured session by parsing the XML document according to its timestamped tags.

2.2 M4Note

To investigate the integration of mobile capture devices and natural user interaction, we have built *M4Note* (MultiMedia MultiModal Annotation Tool) [4, 5]. *M4Note* is a multi-modal application that supports audio/video capturing and annotation of images obtained as video frames extracted on-the-fly from a video stream.

M4Note was developed to run on tablet PCs, allowing capturing and annotation to be carried out during live experiences that require mobility — for instance, during a visit to a zoo or a museum, or in a family party.

The *M4Note*'s multimodal interface allows the annotations to be made via digital ink or speech, being in both cases recognized and associated with the multimedia stream. It also supports the definition of marks and gestures that can have specific semantics, leading users to create a personalized hierarchy of signs that maximize the meaning of the annotation. The approach provides annotation as metadata for indexing, retrieval and semantic processing as well as content enrichment, and uses an underlying model for structured multimedia descriptions and annotations, allowing the establishment of spatial, temporal and linking relationships.

At the end of the process, *M4Note* automatically generates a structured document referencing all the captured content. Later, this document can be used for playing back the audio, video and ink.

3. SCENARIOS OF USE

In this section, we show the use of *M4Note* and *wiClass* on two different scenarios, respectively: a historical farm field trip and a visit to a zoo. We have also present some general impressions we could notice from the users' experience with the systems.

3.1 Pinhal Farm Field Trip

The *wiClass* system has been used to support a field trip by a group with more than 30 people from several disciplines (including architecture, chemistry, cinema and history) who spent 4 days working in a farm that holds important information with respect to the history of São Carlos, in the State of São Paulo, Brazil. The farm is known as "Fazenda do Pinhal".

The task of the group was to gather data such as copies of photographs (that cannot leave the on site library), copies of historical documents that report on the development of coffee plantations and other agricultural information, floor plants relative to the buildings built by slaves as well as obtain new photographs, films and correct measures relative to the place.

Portions of original floor plants were scanned before being imported in *wiClass* on a Tablet PC: the aim was to annotate correct measures on top of the original image (Figure 1). Alternatively, pictures taken with digital cameras were imported to register and mark details of the building's architecture.

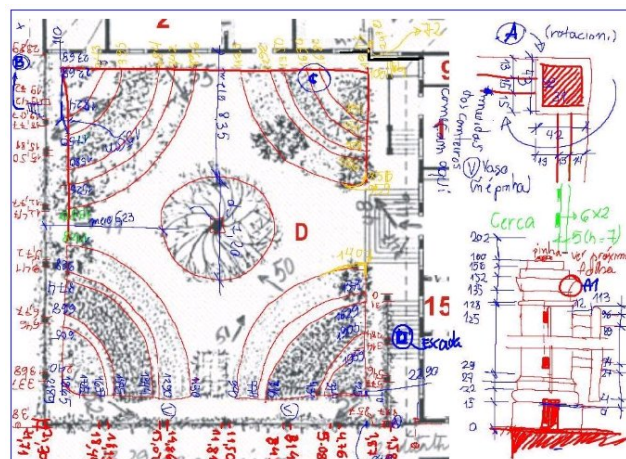


Figure 1: Floor plant, obtained on site from a digital image, is annotated with digital ink.

The group, with students and faculty from different areas, has built a novel way of using the technology when working together. A faculty said "I have totally changed the way I think about computers", and another said "I have never seen such a multidisciplinary team working so effectively together".

During the 4 days visit, over 1.5 GB of quality information in digital format was produced — including sketches, ink notes, photographs and videos — impressive considering that much of the information produced in the past year's effort, when the activity was first held and no automatic capture system was used, is still in paper format. All these digital artifacts, available via a Web repository, are used in learning activities.

3.2 A Visit to the Zoo

A scenario that illustrates the utility of the *M4Note* tool is that of a visit of a user to a zoo: users may not only record important points of the visit by means of capturing video but can also associate annotations to the video stream at the time of the capture — while the user has the focus of attention on the visit itself.

The visitor uses the *M4Note* tool running on a Tablet PC attached to a conventional WebCam (Fig. 2:left). The Tablet PC provides mobility, allowing the user to move in the environment while recording a video stream with particular points of interest. When capturing video, the user selects a video frame by clicking on the video stream. The selected frame is then converted to a JPEG image and presented as the background of a whiteboard: image annotation can be achieved by marking, handwriting or audio (Fig. 2:middle). The ink can be recognized as marks that have semantics for the particular user (Fig. 2:right).

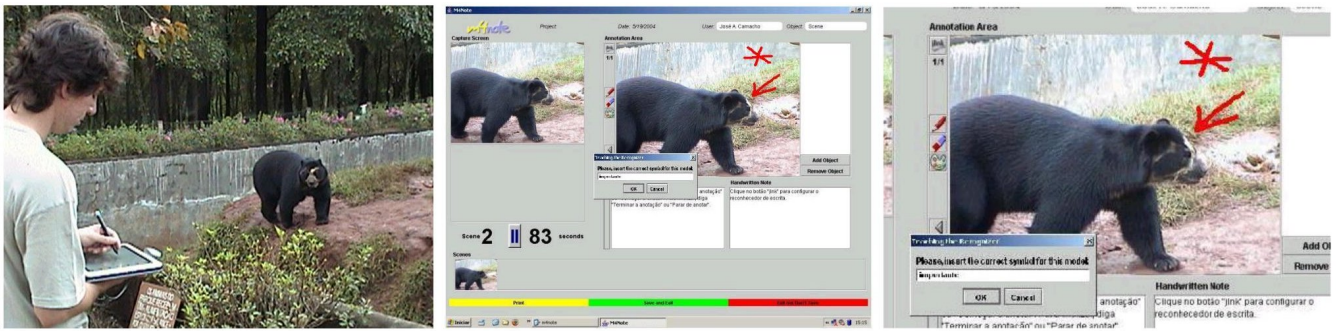


Figure 2: Using a camera attached to a Tablet PC running *M4Note* (left), the user captures video and annotates on top images extracted from the video (middle); special ink marks can be recognized automatically and given special meaning when recognized (right).

Annotations typically comprise explanations and comments that should be made to groups of visitors during a guided visit, or general notes and personal feelings a ordinary user wants to register.

3.3 The User Experience

From the users' interactions with the prototyped systems, some general impressions could be noticed:

(a) the registering of the natural interaction that occurs in live sessions is important for users that need to review the contents or want to understand the rationale behind the capture process;

(b) the use of pen-based devices, like Tablet PCs, associated with the presentation of images and ink annotation provide a user model that is easy to understand and facilitates the work of multidisciplinary teams;

(c) the capture process produces review documents at a low authoring cost and the resulting documents are already in digital form and accessible via Web in a centralized repository.

4. FINAL REMARKS

We discussed the area of pervasive image capture and sharing from the capture and access perspective. Two prototyped systems were described and situated on proper scenarios of use. We have also pointed out some general impressions the users demonstrated while experimenting with the systems.

The capture components are being adapted to communicate in a peer-to-peer fashion so they become independent of centralized services and fixed, server-based infrastructures.

Acknowledgments. We would like to thank FAPESP, CNPq and Hewlett-Packard for their financial support, and the many authors of *M4Note* and *iClass*.

5. REFERENCES

[1] G. Abowd. Classroom 2000: An experiment with the instrumentation of a living educational environment. *IBM Systems Journal*, 38(4):508–530, 1999.

[2] J. A. Brotherton and G. D. Abowd. Lessons learned from eclass: Assessing automated capture and access in the classroom. *ACM TOCHI*, 11(2):121–155, 2004.

[3] R. Cattelan, L. Baldochi, and M. Pimentel. Processing and storage middleware support for capture and access applications. In *Companion Proceedings of the 2003 ACM/IFIP/USENIX International Middleware Conference*, page 315, 2003.

[4] R. Goularte, J. A. Camacho-Guerrero, V. R. Inácio, R. G. Cattelan, and M. G. C. Pimentel. M4note: a multimodal tool for multimedia annotation. In *Proceedings of the 2nd IW3C2 Latin American Web Congress*, pages 142–149, 2004.

[5] R. Goularte, R. G. Cattelan, J. A. Camacho-Guerrero, J. Valter R. Inacio, and M. da Graça C. Pimentel. Interactive multimedia annotations: enriching and extending content. In *ACM DocEng '04*, pages 84–86, 2004.

[6] A. A. Macedo, J. A. Camacho-Guerrero, R. G. Cattelan, V. R. Inácio, Jr., and M. da Graça Campos Pimentel. Interaction alternatives for linking everyday presentations. In *Proceedings of the fifteenth ACM conference on Hypertext & hypermedia*, pages 112–113. ACM Press, 2004.

[7] M. Pimentel, L. Baldochi, and R. Cattelan. Prototyping applications to document the human experience. *submitted to IEEE Pervasive Computing*, 3(4).

[8] M. Pimentel, Y. Ishiguro, B. Kerimbaev, G. Abowd, and M. Guzdial. Supporting educational activities through dynamic web interfaces. *Interacting with Computers*, 13(3):353–374, 2001.

[9] M. Weiser. The computer for the 21st century. *Scientific American*, 265(3):94–104, 1991.