Memory Specs–An Annotation System on Google Glass using Document Image Retrieval

Katsuma Tanaka
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
firstname.lastname@gmail.com

Motoi Iwata
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
Katsuma@m.cs.osakafu-u.ac.jp

Kai Kunze
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
masa@cs.osakafu-u.ac.jp

Koichi Kise
Graduate School of Engineering
Osaka Prefecture University
1-1 Gakuen-cho, Naka, Sakai,
Osaka, Japan
kise@cs.osakafu-u.ac.jp

UbiComp'14 Adjunct, September 13-17, 2014, Seattle, WA, USA
ACM 978-1-4503-3047-3/14/09.
http://dx.doi.org/10.1145/2638728.2638775

Abstract
We present a system working on a wearable computer that can annotate and retrieve annotations for signs, posters, public displays etc. The only limitation of the system: the annotated objects need to contain at least 3-5 lines of text with fixed layout. We evaluate our system in a poster session scenario with 5 use cases: retrieving annotations, adding notes, taking pictures and recording audio and attaching them to posters. As a ubiquitous note taking and annotation application, we believe the system can bridge online and offline discussions and give the user a way to reflect on seen information by providing them a web interface to all the annotations they took together with the source material (in our case poster and paper publications), helping to organize their thoughts and improving their recall.

Author Keywords
Annotations; memory; document image retrieval; wearable; google glass; interactive paper

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces Graphical User Interfaces;; H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Figure 1: Pictures from the system evaluation: Annotating a poster with a voice note and retrieving a video annotation associated to a text region of a poster.

Introduction

With the concepts like geo tagging and more general, augmented/mediated reality, slowly virtual and physical worlds are more and more connected. Researchers are more and more focusing on how to combine offline/online topic discussions and content [6].

To this end, we present a wearable application for annotations on ambient text. Our system can be used to annotate and retrieve annotations for any poster or other public sign that contains more than 3-6 lines of text. This text needs to be registered in a database beforehand (using a digital version of the text) or on the fly by taking a picture.

The contributions of this paper are two fold. First, we evaluate a Google Glass prototype of our system for a poster session scenario: Retrieving information, attaching notes to a poster, attaching images to a poster, attaching audio notes. A record of the annotations can be later retrieved using a web interface. We present insights from using the prototype, user opinion and an expert analysis. Second, we provide open system libraries for the use of annotations with documents, signs, posters for mobile platforms (iOS, Android and Google Glass) as well as a backend server for registering and retrieving annotations for a given document.

We believe that our system can help to merge offline and online topic discussions and enable more and more people to take part in them. Additionally, as a history of annotations and the source material can be retrieved using a web interface on the annotation server, the system can help the individual to organize their thoughts around topics and to better remember discussions and conversations.

Related Work

There are a lot of researchers focusing on combining digital and real-world using annotation technologies, especially focusing on Augmented Reality [7, 6]. Most of them use mobile devices, e.g. smart phones. The works we are aware of use SIFT or similar features, for indexing and retrieval and thus can deal only with a small number of objects/texts.

Researchers already explore annotations to support knowledge acquisition and memory recall mostly in education and work environments [5]. We build on this research in providing an viable annotation system running on a wearable computer for large scale deployments.

Other work focuses on document annotations with different application goals: interfaces for multimedia control or retrieval of travel information [2]. Additionally, researchers in AR tend to use QR codes or similar optical markers [1]. We find that for our use case these are too restrictive as it requires to prepare specific print media and limits the accuracy and place you can put annotations.

We build on our work for active reading support for documents using smart phones [4, 8].

Annotations for the Real World

Currently we focus on annotations on ambient text displayed on signs, displays, posters etc. for mainly two reasons. First, as we get a lot of information from ambient texts, we believe that this makes a good starting point for fostering offline/online discussions. Second, the technology of document image retrieval matured enough so this can be used in large scale and has not the trouble of other approaches (needing special markers, expensive feature calculation).

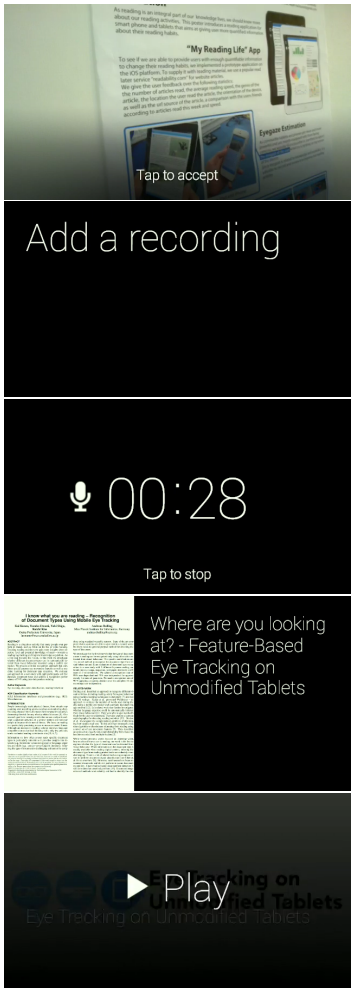

Figure 2: Screenshots from the Google Glass prototype: uploading an image from the poster text region, adding a audio note and retrieving paper and video from the poster.

We decided to use a document image retrieval method based on "Locally Likely Arrangement Hashing" (LLAH)[4]. LLAH is very fast. The feature extraction method can be implemented on todays mobile devices and run in perceived realtime. Smaller databases (a couple of documents/ambient texts) can be searched on a powerful wearable computer (e.g. Google Glass) or smart phone as well. For larger document databases we need a client server infrastructure.

Usage Scenario – Poster Annotations

To evaluate our idea of a wearable annotation system, we implement a prototype using Google Glass focusing on a poster session scenario at a conference. Poster sessions at computer science conferences are a perfect first test of our wearable annotation system, as there is a lot of ambient text and most likely also a large population of Google Glass users. In the following we describe the usage scenarios we support for our first prototype.

Retrieving information – When the poster is registered with our system. The presenter of a poster can for example link videos to their poster text. Registration of additional information can be done using Google Glass or a mobile phone or desktop client. For ease of use we recommend using a digital version of the poster and a desktop client.

Attaching notes to a poster – A user can take a picture using our Google Glass prototype and use the Google Glass voice recognition system to leave a text note to the poster.

Attaching images to a poster – A user can attach images. We think this could make it easier to remember the face of an author or presenter for a given poster or give the possibility to add interesting prototype pictures to the digital poster document.

Attaching Audio Notes – A user can also record and add audio files to a poster, so he's able to remember and recall the content of discussions related to the poster topic.

Implementation

As proof of concept, we implemented a first prototype on Google Glass supporting all use cases presented beforehand (sample screenshots of the application in Figure 2. We have working application level libraries for iOS, Android, Linux and Windows. As mentioned before, the application libraries can be used in stand-alone or client-server mode. The Google Glass prototype uses currently the client-server model.

We also show a Web interface, allowing the user to see all posters he annotated and access his annotations as well as the public annotations of his friends (see Fig. 3). So people not attending a given poster session or event can virtually take part in discussions and the user can later on have a record of his/her conversations to better recall ideas and thoughts.

Evaluation

System Test

If we run our system in client server mode a large number of documents can be supported (We used 100 Million text regions for demonstration purposes with an average retrieval time of 178 ms). Unfortunately the wait time for the user on our first prototype is longer (around 2 sec.). See the discussion later. As seen from previous

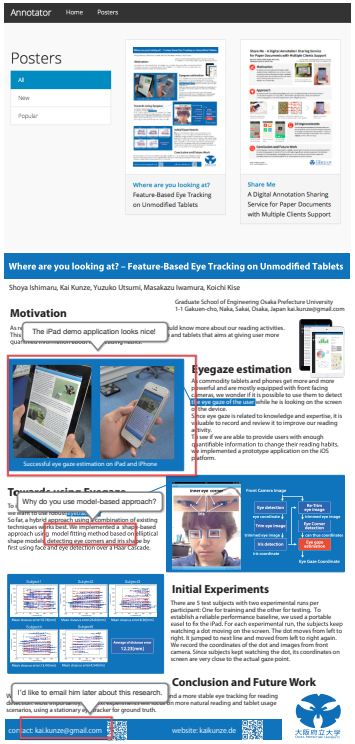

Figure 3: Screenshot from the web page, showing all annotated posters and a poster with text annotations.

publications, scaling our system to a large size is not an issue [3]. The bigger limitation is the necessity of 3-6 lines of text (at least 4-5 words per line). We are currently working on integration of other computer vision features to integrate them without loosing the scalability and performance.

Usability Impressions

For an initial test, we had 2 users with Google Glass interact with the system annotating 3 posters and retrieving information (see Figure 1). We also performed an expert evaluation with one UI researcher. In the following, we summarize the results.

Popular Features Audio notes and Video Retrieval– The two users especially liked to add quickly audio notes to a poster while discussing about the poster content. The second popular feature was retrieving a video about the poster topic was

Quick Capture – From the evaluation, we saw the need for quick capture. The main problem for annotations right now is an average wait time of 2 seconds after taking the picture, this is due to some programming limitation on Google Glass (the image should first be saved before it's send to the server). The users also raised the wish to just capture a poster picture so they can later on read the paper or watch the video on the desktop client (not adding annotations). We will add this use case (quick capture) in the next prototype.

Conclusion and Future Work

We present a framework to annotate ambient text digitally. We implemented a prototype system on Google Glass, evaluating its functionality on a poster session scenario in a small scale. As a next step, we would like to use the system in a real conference scenario to gather feedback from more users identifying use cases and improving user interface and overall usability.

References

[1] Chen, N.-S., Teng, D. C.-E., Lee, C.-H., and Kinshuk. Augmenting paper-based reading activity with direct access to digital materials and scaffolded questioning. *Comput. Educ.* (2011), 1705–1715.

[2] Erol, B., Antúnez, E., and Hull, J. J. Hotpaper: multimedia interaction with paper using mobile phones. In *Proc of MM 2008*, MM '08 (2008).

[3] Iwamura, M., Nakai, T., and Kise, K. Improvement of retrieval speed and required amount of memory for geometric hashing by combining local invariants. *Proc. of BMVC2007* (2007), 1010–1019.

[4] Kunze, K., Tanaka, K., Iwamura, M., and Kise, K. Annotate me: Supporting active reading using real-time document image retrieval on mobile devices. *UbiComp '13 Adjunct* (2013).

[5] Mackay, W. E., and laure Fayard, A. Designing interactive paper: Lessons from three augmented reality projects. In *In Proc. of IWAR98*, A K Peters, Ltd (1999), 81–90.

[6] Rekimoto, J., and Nagao, K. The world through the computer: Computer augmented interaction with real world environments. In *Proc. of UIST*, ACM (1995), 29–36.

[7] Rose, E., Breen, D., Ahlers, K. H., Crampton, C., Tuceryan, M., Whitaker, R., and Greer, D. Annotating real-world objects using augmented reality. *Computer graphics, Academic Press Ltd.* (1995).

[8] Tanaka, K., Iwata, M., Kunze, K., Iwamura, M., and Kise, K. Share me-a digital annotation sharing service for paper documents with multiple clients support. In *(ACPR*, IEEE (2013), 779–782.