Ubiquitous Contextual Information Access with Proactive Retrieval and Augmentation

Antti Ajanki[1], Mark Billinghurst[3], Melih Kandemir[1], Samuel Kaski[1], Markus Koskela[1], Mikko Kurimo[1], Jorma Laaksonen[1], Kai Puolamäki[2], and Timo Tossavainen[2]

[1] Aalto University, Adaptive Informatics Research Centre
[2] Aalto University, Department of Media Technology
[3] The Human Interface Technology Lab New Zealand, University of Canterbury

Abstract. In this paper we report on a prototype platform for accessing abstract information in real-world pervasive computing environments through Augmented Reality displays. Objects, people, and the environment serve as contextual channels to more information. Adaptive models will infer from eye movement patterns and other implicit feedback signals the interests of users with respect to the environment, and results of proactive context-sensitive information retrieval are augmented onto the view of data glasses or other see-through displays. The augmented information becomes part of the context, and if it is relevant the system detects it and zooms progressively further. In this paper we describe the first use of the platform to develop a pilot application, a virtual laboratory guide, and early evaluation results.

Keywords: augmented reality, gaze tracking, information retrieval, machine learning, pattern recognition

1 Introduction

We are interested in the problem of how to efficiently retrieve information in real-world environments where (i) it is hard to formulate explicit search queries and (ii) the temporal and spatial context provides potentially useful search cues. The user may not even have an explicit query in mind but the information that would be useful is likely to be related to objects in the surrounding environment or other current context cues. The system recognizes objects and people as potential search cues, uses gaze patterns and other actions to infer their relevancy, and retrieves information about them. The information is augmented on data glasses. We have developed a hardware and software platform which meets these needs and have implemented a demonstration application, *virtual laboratory guide*, which shows how the idea can be applied to help a visitor to a university department. We carry out a small-scale user study of the concept.

Previous researchers have used wearable and mobile Augmented Reality systems to display contextual cues about the surrounding environment. For example the Touring Machine [1] added virtual tags to real university buildings showing which departments were in the buildings. While several studies prove the availability of face recognition on a wearable platform [2], our research uses recognized faces to trigger contextual cues for finding context-sensitive information associated with the faces seen.

There have been studies about gaze as relevance feedback [3], but we are the first to use gaze in information filtering in an AR setup. Speech recognition has been studied a lot (for a review, see [4]). However, the combination of inferring implicitly the focus of visual attention from gaze and inferring contextual information from speech is novel.

2 Components of Contextual Information Access System

We have implemented a pilot software for on-line studies of contextual information access. Faces and objects are detected and tracked from the video feed. Their relevancy is estimated from gaze or pointing patterns. Information about the relevant objects is retrieved using the recognized objects and speech as search cue and is displayed using augmented reality rendering.

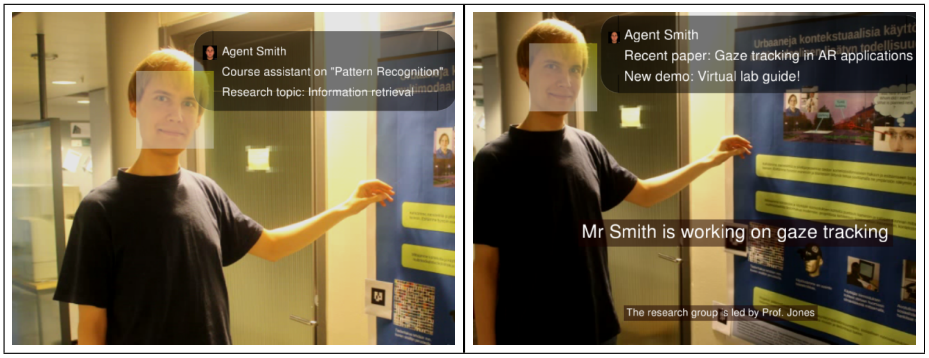
Face detection and tracking are performed with the Viola & Jones detector [5]. The detected faces are then transmitted to an image database engine for recognition. The system also detects AR markers for detecting objects. User's speech is also used as a source of contextual cues in the system. We have modified the Aalto University's online large-vocabulary speech recognizer [6] for this purpose. The word error rate of the speaker-independent recognizer in Finnish with a close-talk microphone in public places is 13.6% [7].

To determine which objects or faces should be annotated we infer their relevancy from gaze patterns. In this pilot application, we assume the relevance of an object to be proportional to the total time an object or related augmented annotations were under user's attention, within a fixed-length time window. A recent study shows that the accuracy can be further improved by using more advanced gaze features [8].

The video of the real world is augmented with textual annotations related to the recognized faces and objects (see Fig. 1). The annotations are selected so that they are relevant in the current context which is defined by the recently seen objects and recent speech recognition keywords. In the pilot application, the potential annotations are manually labeled with information about in which kind of contexts they should be displayed. The annotation that best matches the measured context is shown. The AR implementation in the pilot application is monocular video see-through. The camera captures video with 640×480 resolution at a frame rate of 15 frames per second (FPS). We use the ALVAR augmented reality library[4] for calibrating the camera and for detecting fiducial markers and determining their pose relative to the camera.

We tested our system on two output devices; (1) a head-mounted display with an integrated gaze tracker, and (2) a handheld UMPC with a camera (Fig. 2).

[4] http://virtual.vtt.fi/virtual/proj2/multimedia/alvar.html



Agent Smith
Course assistant on "Pattern Recognition"
Research topic: Information retrieval

Agent Smith
Recent paper: Gaze tracking in AR applications
New demo: Virtual lab guide!

Mr Smith is working on gaze tracking

The research group is led by Prof. Jones

Fig. 1. Screenshots from the Virtual Laboratory Guide. Left: The guide shows augmented information about a recognized person. Right: The guide has recognized that the user is interested in research-related information.

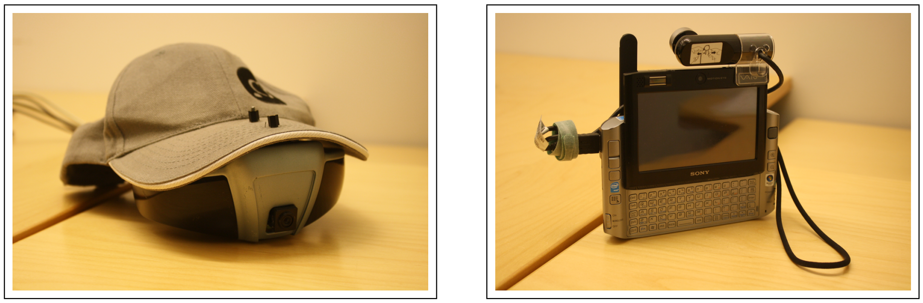


Fig. 2. The display devices. On the left, wearable near-to-eye display with integrated gaze tracker. On the right, hand-held computer with virtual see-through display.

The head-mounted device is a research prototype provided by Nokia Research Center [9]. VGA quality image is displayed with a 30 degree field of view. The hand-held PC is an ultra-mobile Sony Vaio computer with a 4.5 inch display.

3 Experiments

As a pilot application for testing the framework we have implemented an AR guide for a visitor at a university department. The *Virtual Laboratory Guide* shows relevant information and helps the visitor to navigate the department. The guide is able to recognize people, offices, research posters, and annotates them with research or teaching related information.

A small-scale study was conducted to to test the usefulness of our AR platform. The goal of the study was to evaluate the usefulness of seeing virtual contextual information and to compare two configurations: a handheld display in a UMPC (HHD) and a head-mounted display (HMD). The head-mounted display is a very early prototype which will naturally affect the results.

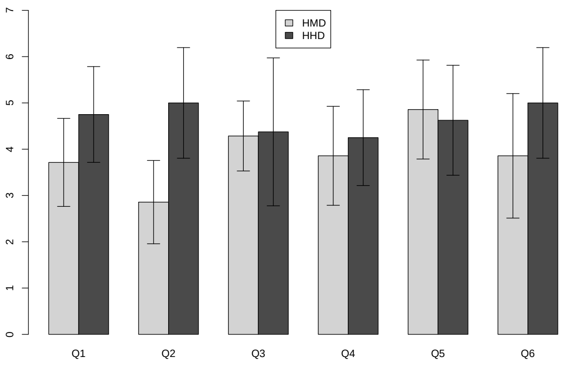


Fig. 3. Usability study of the system in handheld display (HHD) and head-mounted display (HMD) conditions. Average Likert scores with standard deviations for the usability questionary on the scale of 1 to 7. See the body text for the description of questions Q1–Q6.

The eight subjects were male university students aged from 23 to 32 years old. None of them had prior experience with the Virtual Laboratory Guide. During the experiments, AR markers were placed on research posters on the wall and on desks. Three people were available for the purpose of face recognition.

In each condition the subjects were asked to find answers to one research-related question and one teaching-related question. The answers were available through the augmented annotations. After the subject had completed one task, the experiment supervisor changed the topic of the augmented information by speaking out to the speech recognizer a sentence related to the next topic.

After each display condition the subjects were asked what they liked best and least about the condition and if they had any other comments. They also answered following questions on a 7-point Likert-scale: (Q1) How easy was it to use the application? (Q2) How easy was it to see the AR information? (Q3) How useful was the application in helping you learn new information? (Q4) How well do you think you performed in the task? (Q5) How easy was it to remember the information presented? (Q6) How much did you enjoy using the application?

3.1 Results

In general, the users were able to complete the task under both conditions and found the system a useful tool for presenting context information. Figure 3 shows the results for each of the evaluation questions averaged across all subjects. The users found the HHD easier to use (Q1, paired t-test, $p = 0.05$) and HHD display easier to read (Q2, paired t-test, $p < 0.001$).

In the interview questions the subjects reported that the most severe weakness in the head-mounted display was the quality of the image. The issues with the quality of the early prototype head-mounted display were to be expected. The subjects also felt that the handheld screen that was too small and heavy.

4 Discussion

We have proposed a novel augmented reality application which infers the interests of the user from implicit signals such as gaze and speech. The system overlays contextually relevant information about people and objects over the display of the real world.

In this paper we presented a pilot study comparing two user interfaces: a wearable display with an eye tracker, and an ultra-mobile computer. Further studies will be performed on relevance inference, face and marker recognition, speech recognition, as well as other components in the future.

Acknowledgements: This project is funded by Aalto MIDE programme, project UI-ART. We thank Toni Järvenpää for help with the data glasses research, and Tomas Günther for contributing to the augmented reality implementation.

References

1. Feiner, S., MacIntyre, B., Höllerer, T., Webster, A.: A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. Personal and Ubiquitous Computing **1**(4) (1997) 208–217
2. Starner, T., Mann, S., Rhodes, B., Levine, J., Healey, J., Kirsch, D., Picard, R.W., Pentland, A.: Augmented reality through wearable computing. Presence: Teleoperators and Virtual Environments **6**(4) (1997) 452–460
3. Ajanki, A., Hardoon, D.R., Kaski, S., Puolamäki, K., Shawe-Taylor, J.: Can eyes reveal interest? – Implicit queries from gaze patterns. User Modeling and User-Adapted Interaction **19**(4) (2009) 307–339
4. Rebman, Jr., C.M., Aiken, M.W., Cegielski, C.G.: Speech recognition in the human-computer interface. Information & Management **40**(6) (2003) 509–519
5. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Press (2001)
6. Hirsimäki, T., Pylkkönen, J., Kurimo, M.: Importance of high-order n-gram models in morph-based speech recognition. IEEE Transactions on Audio, Speech and Language Processing **17**(4) (May 2009) 724–732
7. Kallasjoki, H., Palomäki, K., Magi, C., Alku, P., Kurimo, M.: Noise robust LVCSR feature extraction based on stabilized weighted linear prediction. In: Proceedings of the 13th International Conference Speech and Computer (SPECOM). (June 2009) 221–225
8. Kandemir, M., Saarinen, V.M., Kaski, S.: Inferring object relevance from gaze in dynamic scenes. In: Short Paper Proceedings of Eye Tracking Research and Applications Symposium (ETRA), ACM Press (2010) 105–108
9. Järvenpää, T., Aaltonen, V.: Compact near-to-eye display with integrated gaze tracker. In: Photonics in Multimedia II. Volume 7001 of Proceedings of SPIE. SPIE (2008) 700106–1–700106–8