Creating Gaze Annotations in Head Mounted displays

Diako Mardenbegi
IT University of Copenhagen
Copenhagen, Denmark
dima@itu.dk

Pernilla Qvarfordt
FX Palo Alto Laboratory
Palo Alto, USA
pernilla@fxpal.com

ABSTRACT
To facilitate distributed communication in mobile settings, we developed GazeNote for creating and sharing gaze annotations in head mounted displays (HMDs). With gaze annotations it possible to point out objects of interest within an image and add a verbal description. To create an annotation, the user simply captures an image using the HMD's camera, looks at an object of interest in the image, and speaks out the information to be associated with the object. The gaze location is recorded and visualized with a marker. The voice is transcribed using speech recognition. Gaze annotations can be shared. Our study showed that users found that gaze annotations add precision and expressiveness compared to annotations of the image as a whole.

Author Keywords
Eye Tracking, Annotations, Gaze Pointing, HMD

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

INTRODUCTION
With wearable computing devices such as Google Glass or Vuzix Smart Glasses, users can communicate with remote collaborators by capturing and sharing personal views. To effectively communicate via images, annotations and tags are needed to convey a complete message. Tanaka et al. [2] built a system for creating and retrieving annotations of signs and posters using Google Glass. However, photos taken by the Glass' camera often have a wide-angle view to match the user's field of view. Hence they often contain more objects than desired which can distract from the message the user wants to communicate. Wearable computing devices often have limited input methods, which makes it hard to add precise markers in images.

In this paper we present GazeNote, a system for creating and sharing gaze enhanced annotations to photos taken by Google Glass. GazeNote can be used for training purposes, for instance a master photographer can point out aspects of a scenic view a novice can access when visiting the same spot. GazeNote can also allow for communication between different professionals, for instance inspectors could communicate issues a technician later should take care off.

ISWC '15, September 7–11, 2015, Osaka, Japan.

http://dx.doi.org/10.1145/2802083.2808404

GAZENOTE: ANNOTATING WITH GAZE
We designed GazeNote for creating, accessing and sharing annotations in HMDs. We attached an infrared camera and light source to Google Glass and connected it to the open source Haytham gaze tracker [1]. We experimented both with a wired USB camera (Genius iSlim 321R webcam, 5m cable) and a wireless camera (2.4 GHz radio night vision analog camera). Both cameras output video at 30 fps. Gaze tracking is performed using pupil and cornea reflection, and requires a 4-point calibration. The gaze estimation is performed on a server and communicated to an Android background service (GlassGaze) which routes gaze data to different applications, such as GazeNote. Due to Glass' limited battery and processing power, the gaze estimation was performed on the server.

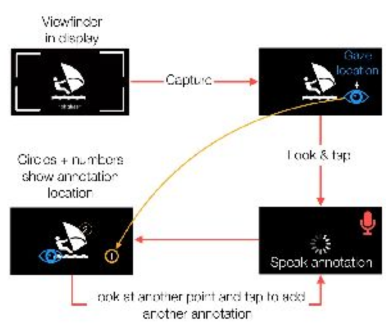

Figure 1. Interaction flow for creating a gaze annotation

Figure 2. One of the prototypes developed

To annotate a scene, the user gives a verbal command to start a new annotation (Figure 1). When the command is detected, Glass captures the scene and displays the resulting image. GazeNote superimposes the user's gaze position on the captured photo. The user decides where to add an annotation, fixates on the location, and taps on Glass. A gaze marker is inserted at the fixation point on when tapping. A prompt for voice input is shown and the user speaks the annotation. The utterance is transcribed using automatic speech recognition. We choose to transcribe the speech over attaching recorded audio to save bandwidth and to allow the gaze annotations to be searchable. GazeNote allows more than one gaze annotation per photo.

Gaze annotation sharing was implemented using a multimedia messaging service. A remote user accesses annotations by fixating on a gaze marker and tapping on Glass.

USER STUDY

To investigate users' experiences with *gaze annotations*, we compared them with simple *image annotations*. An image annotation applies to the whole image without a gaze marker referring to specific locations. It is created using the same interaction as gaze annotations; the user gives a voice command to capture a scene and taps on Google Glass to add the annotation using speech.

We recruited 7 participants, 6 men and 1 woman, for the study. In an open workspace area, a scene was created with various objects such as large Lego blocks, small boxes and toys. The task was to communicate objects' placement to an absent colleague. Before creating two annotations, the participants accessed one *image annotation* and one *gaze annotation*. The eye tracker was calibrated before accessing the gaze annotation and before creating gaze annotations.

RESULTS & DISCUSSION

In total the participants created 28 annotations, 14 gaze and 14 image annotations. To create these, the participants on average needed 1.3 attempts (SD=0.57) for gaze annotations and 1.1 (SD=0.31) for image annotations. These low numbers indicate that the participants had an easy time learning the system.

Accessing Annotations

After accessing the annotations, the participants were asked about their preferences for the two kinds of annotations along the dimensions of message precision, naturalness and expressiveness. As Figure 3 shows, the participants experienced gaze annotation messages as equally or more precise (6/7 participants), natural (6/7), and expressive (7/7) than image annotations.

Figure 4. Participants' preference of annotation type for precision, naturalness, expressiveness and ease of use

Creating Annotations

After creating gaze and image annotations, the participants' preferences for the two kinds of annotations were measured along the dimensions precision, naturalness, expressiveness and ease of use. The participants favored image annotation for naturalness and ease-of-use (see Figure 4). These results were likely heavily influenced by eye tracking inaccuracies. Two participants expressed this: "Naturalness is probably better for gaze annotation when the tracking is precise enough. As tracking wasn't perfect for me it was too hard to feel natural yet." (P7)

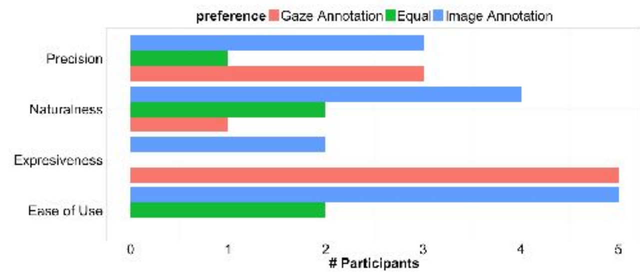

Figure 3. Participants' preferences of annotation type for message precision, naturalness and expressiveness

However, most participants (five of seven) favored gaze annotations for expressiveness (Figure 3). In particular, they expressed that their annotations more precisely identified objects of interest: "Creating image annotations is faster without targeting specific locations within the captured images, but there's less a sense that I am communicating visually what's important than with the gaze annotations." (P5)

These results show that participants felt that gaze annotations add an important dimension, as the annotator can direct attention to particular areas of the image, making the annotation more precise and expressive compared with annotating the image as a whole.

CONCLUSIONS & FUTURE WORK

Effectively identifying specific elements within a shared visual context enhances distributed communication. We created GazeNote, an HMD system enhanced with eye tracking for creating, sharing and accessing gaze annotations. Gaze annotations are created in a few simple steps; the user gives a command, the system captures and displays a photo. The user adds an annotation to an object of interest by fixating on it and taps on the side of Glass.

We performed a user study to validate that gaze annotations can enhance distributed communication. We found that users' impressions of gaze annotations were influenced by the tracker's performance, but also that gaze annotation added precision and expressiveness to the communicated message.

REFERENCES

1. Haytham Gaze tracker, http://www.itu.dk/research/eye, Sept 2014
2. Tanaka, K., Kunze, K., Iwata, M. and Kise, K. Memory Specs-An Annotation System on Google Glass using Document Image Retrieval. In *Proc. of UbiComp'14 Adjunct* (2014).