User Attention Oriented Augmented Reality on Documents with Document Dependent Dynamic Overlay

Takumi Toyama* DFKI GmbH
Wakana Suzuki† Osaka Prefecture University
Andreas Dengel‡ DFKI GmbH
Koichi Kise§ Osaka Prefecture University

ABSTRACT

When we read a document (any kind of, scientific papers, novels, etc.), we often encounter a situation that the information from the reading document is too less to comprehend what the author(s) would like to convey. In this paper, we demonstrate how the combination of a wearable eye tracker, a see-through head-mounted display (HMD) and an image based document retrieval engine enhances people's reading experiences. By using our proposed system, the reader can get supportive information in the see-through HMD when he wants. A wearable eye tracker and a document retrieval engine are used to detect which line in the document the reader is reading. We propose a method to detect the reader's attention on a word in a reading document, in order to present information at a preferable moment. Furthermore, we also propose a method to project a point of the document to a point of the HMD screen, by calculating the pose of the reading document in the camera image. This projection enables the system to overlay the information dynamically in an augmented view on the reading line. The results from the user study and the experiments show the potential of the proposed system in a practical use case.

Index Terms: H.5.2 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: User Interfaces—Input devices and strategies;

1 INTRODUCTION

When we read a document (any kind of, scientific papers, novels, etc.), we often encounter a situation that the information from the reading document is too less to comprehend what the author(s) would like to convey. In such situations, what people used to do was to go to a library and find relevant books, dictionaries, papers and refer them in order to find the information they need. The advent of digital documents eased the troubles in such situations to some extent, for example, readers can search for the meaning of unknown terms directly from texts on the internet. Today, people can easily access to a vast amount of information resources from a digital format of documents. However, it is undoubtful that we still need to rely on paper printed documents, despite the drastic growth of the digitally available documents, meaning that we still need to use other inconvenient ways to look up the information we need. Recent developments of Augmented Reality (AR) technology could propose intelligent solutions for this kind of problems, such as [4, 3, 8]. These previous works have shown the potential of AR systems for assisting people when they read paper documents.

However, one of the issues for such kind of AR applications is how to design the user input. Typical AR applications, such as Wikitude[1], Google Goggles[2], etc. overlays the information when the user directs a mobile camera integrated in a smartphone towards tagged objects. There is still room for improvements from the usability viewpoint. In this work, we propose a system that employs the user's eye movements for an input interface and presents information in a see-through head-mounted display (HMD) with an augmented reality manner by extending the work in [8]. A sample image of the proposed system is shown in Figure 1.

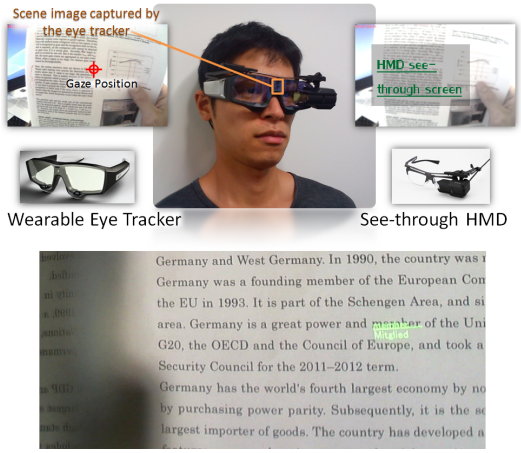

Figure 1: Top: The user wearing the proposed system. Bottom: How the world looks like from the user's perspective wearing the proposed application. The translation (German) of the gazed word is presented in the see-through display upon the original text line.

Recently, remarkable advances of an eye tracking technology have shown a large potential for gaze based human-computer interaction [1]. Human attention system has a strong relation with a gaze control [2]. Therefore, the use of eye tracker could potentially be very suitable for designing an intuitive user interface. Especially, gaze study on people's reading activity has been done over several years [6].

This paper presents a method for detection of user attention on the reading document in order to provide the information when the user needs it. For retrieval of a reading document, we use the state-of-the-art image based document retrieval method called Locally Likely Arrangement Hashing (LLAH) [5]. The user gaze on a document is analyzed and on which word in the reading document the user is attended is detected, which shall be used for a trigger event for information visualization. Furthermore, we also propose a calibration method for projection of a point of the reading document onto a point of the HMD screen. Document image retrieval using LLAH is processed amazingly fast and the estimation of the document pose in the image is done quite accurately. This feature of document retrieval facilitates the point projection quite robustly and the calibration process quickly. In the following section, we present

*e-mail:takumi.toyama@dfki.de
†e-mail:wakana@m.cs.osakafu-u.ac.jp
‡e-mail:andreas.dengel@dfki.de
§e-mail:kise@cs.osakafu-u.ac.jp

[1] http://www.wikitude.com/
[2] http://www.google.com/mobile/goggles/

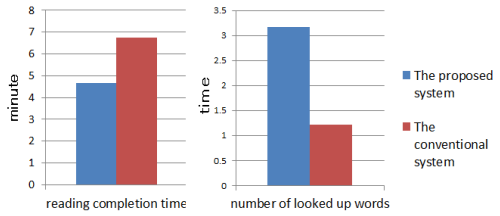

Figure 2: User study results. Average of all test persons.

the framework of the proposed system.

2 Proposed System

An image of the apparatus that we use is shown in Figure 1. We combine SMI Eye Tracking Glasses (an eye tracker) and Brother Airscouter (an HMD). As the first step before using the system, the system has to be calibrated. There are two calibration processes. One is the calibration of the eye tracker and the other is the calibration of the HMD. When the calibration processes are completed, the system is ready to use. Scene images and the user gaze position in the image are streamed to the image based document retrieval module, which returns the identity of the reading document and its pose in the given image. Based on the pose and the gaze position on the document, it detects if the user attention is on any particular word in the document. Then, the visualization module presents the translation of the word onto the HMD screen.

We adopt an image based document retrieval method proposed in [5]. This method, called LLAH (Locally Likely Arrangement Hashing) is robust to perspective distortion of an image and scale-invariant. When a scene image is given from the camera, the image is blurred by a Gaussian kernel and thresholded adaptively into a binary image in order to detect the centroid of each word region. From the arrangements of the detected centroids, affine invariant feature vectors are calculated. The retrieval process is done by matching the extracted features to the features previously stored in the database. Using a hashing technique, the retrieval can be done quite fast even with a large size of database.

By matching the features between the scene image and the retrieved database image, we also calculate the homography between them. Based on this homography, the pose of the document in the scene image is estimated. Furthermore, from the matches between two images, we can also calculate the rotation and transformation matrix of the document image to the scene image, which can be also used for the calibration of the HMD and a word point projection. The user can easily calibrate the HMD by holding a calibration sheet with a proper pose and pressing a computer key once.

The gaze position on the retrieved document image shows at which word the user is currently fixating. The system, then determines whether the user is reading (skimming) the document or looking at the word carefully based on the gaze movements, in order to trigger a translation presentation event in the HMD at a preferable moment. Because if the information is visible all the time, it is obtrusive to the user. We extend an attention detection method proposed in [7]. Instead of 3D objects used in [7], we detect the user attention on each word in a reading document.

After detecting the user attention on each word, a translation is presented as shown in Figure 1. Since the HMD screen is calibrated, the visualization can be presented right below the original reading text. Thus, the reader can keep reading without being interrupted his activity for looking up the word in a dictionary. The presentation of translation remains until the user restart reading the document again.

3 Experiments

First, we investigated the accuracy of the attended word detection. We asked 17 different persons to read one page document (A4 printout, single column) using the system. Although the accuracy of the eye tracking was dependent to the user, it was observed that there was approximately one line (0.5mm) offset in average, which is quite reasonable because it matches to the accuracy of the eye tracker (0.5 degree). Then, we also evaluated the performance of the attention detection. The system could detect the user attention quite well; however, because the gaze position is sometimes incorrect, it presented sometimes incorrect translations. To compensate the error of the eye tracker, we also implemented another visualization mode that presents a couple of possible word candidates nearby the gaze. By using this visualization mode, most of the user could correctly find the right translation.

We also conducted a user study by comparing the proposed system with a conventional translation system. As the conventional system, an ordinary online dictionary was provided with the user. The user needs to type the word manually on a keyboard to get the translation. Similar to the previous experiment, two types of one page document is given to the user and we timed how long it takes the user to read the document using each translation system. The result is shown in Figure 2. The result clearly shows that the user looks up the meaning of the word more often and the reading completion time is dramatically less using the proposed system.

4 Conclusion

This paper presented a novel AR system to enhance people's reading experiences using a wearable eye tracker, a see-through HMD and an image based document retrieval engine. We proposed a method to detect the user's attention on a particular word in a reading document and a method to calibrate the HMD quickly using the result of the document retrieval. The results from the experiments and the user study show the benefits of the system and the potential of the AR system in combination with an eye tracking system. Future work is to investigate further possibilities of gaze interaction (e.g. whole sentence inputs based on gaze gestures) and to compare the gaze input with other input modalities

References

[1] G. Buscher, R. Biedert, D. Heinesch, and A. Dengel. Eye tracking analysis of preferred reading regions on the screen. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, pages 3307–3312, Atlanta, USA, 2010.

[2] J. M. Henderson. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11):498–504, 2003.

[3] N. Henze and S. Boll. Who's that girl? handheld augmented reality for printed photo books. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part III*, pages 134–151, 2011.

[4] Q. Liu and C. Liao. Paperui. In *Camera-Based Document Analysis and Recognition*, volume 7139, pages 83–100. 2012.

[5] T. Nakai, K. Kise, and M. Iwamura. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. In *Lecture Notes in Computer Science (7th International Workshop DAS2006)*, volume 3872, pages 541–552, Feb. 2006.

[6] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, pages 372–422, 1998.

[7] T. Toyama, T. Kieninger, F. Shafait, and A. Dengel. Gaze guided object recognition using a head-mounted eye tracker. In *Proc. of the Symposium on Eye Tracking Research and Applications*, pages 91–98, 2012.

[8] T. Toyama, W. Suzuki, A. Dengel, and K. Kise. Wearable reading assist system: Augmented reality document combining document retrieval and eye tracking. In *Proceedings of the 12th International Conference on Document Analysis and Recognition*, 2013.