[Poster] Smartwatch-Aided Handheld Augmented Reality

Darko Stanimirovic*
Metaio GmbH

Daniel Kurz†
Metaio GmbH

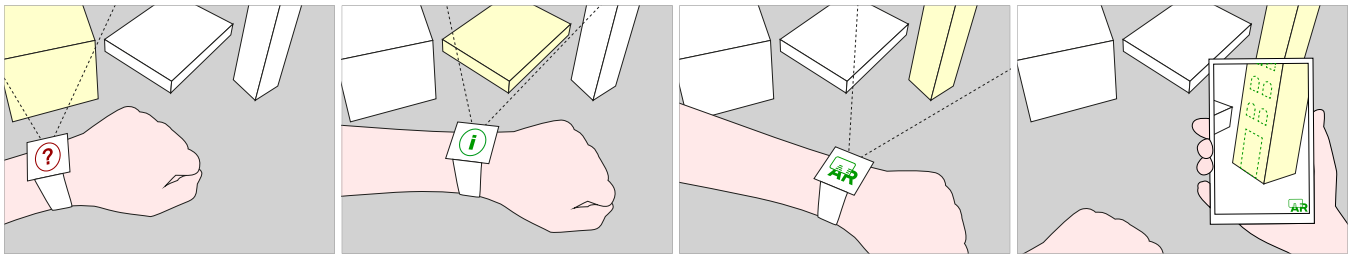

Figure 1: A conceptual view on the proposed method: A camera-equipped smartwatch is used to capture images of nearby objects. Visual Search then determines if there is no digital information available related to an object (a), or if there is information available that could be shown on the smartwatch (b), or if there exists Augmented Reality (AR) content related to the object (c). In the latter case, the smartwatch enables the user to start an AR view on a tethered smartphone (d) providing a rich presentation of related digital information.

Abstract

We propose a novel method for interaction of humans with real objects in their surrounding combining Visual Search and Augmented Reality (AR). This method is based on utilizing a smartwatch tethered to a smartphone, and it is designed to provide a more user-friendly experience compared to approaches based only on a handheld device, such as a smartphone or a tablet computer. The smartwatch has a built-in camera, which enables scanning objects without the need to take the smartphone out of the pocket. An image captured by the watch is sent wirelessly to the phone that performs Visual Search and subsequently informs the smartwatch whether digital information related to the object is available or not.

We thereby distinguish between three cases. If no information is available or the object recognition failed, the user is notified accordingly. If there is digital information available that can be presented using the smartwatch display and/or audio output, it is presented there. The third case is that the recognized object has digital information related to it, which would be beneficial to see in an Augmented Reality view spatially registered with the object in real-time. Then the smartwatch informs the user that this option exists and encourages using the smartphone to experience the Augmented Reality view. Thereby, the user only needs to take the phone out of the pocket in case Augmented Reality content is available, and when the content is of interest for the user.

1 Introduction and Related Work

The world around us is abundant with objects related to which digital information is available. There are different means to access this information, e.g. by querying the name or a description of the object at a search engine. An alternative approach is to use Augmented Reality technology to recognize the object based on a camera image and then display the digital information registered with the real object in real-time. In handheld Augmented Reality the augmentation usually takes the form of augmented 3D objects superimposed on a live video stream and shown on the display of a smartphone or tablet computer.

Even though there exists digital information related to virtually any real object, in reality only a small fraction of these are recognized by Augmented Reality browsers such as junaio[1] due to required authoring work for making a real object AR-enriched. From the point of view of a user, one of the key missing features of modern AR platforms is the ability to automatically notify the user which real-world objects in the surroundings are enriched with AR, and which are not. In most of the current implementations a user has to know a priori which objects provide AR experiences, or he or she has to continuously scan the environment with a smartphone hoping to find an AR-enriched object. This can be a cumbersome task for users, and it can in the long run decrease the appeal of Augmented Reality technology. Alternatively, real-world objects can be marked with an imprinted sign which announces the existence of an AR experience to the user, e.g. the AR logo[2]. In this case deployment of AR technology has to be carefully planned, thus limiting later addition of AR content to real objects on-the-fly.

Our proposed method relies on Visual Search, which is based on visual object recognition. State-of-the-art implementations commonly rely either on the Bag-of-Words approach [5], or on Fisher kernels [2], accompanied with a classification method, e.g. support vector machines [1]. The response of a Visual Search method after being provided with an image, usually is an identifier referring to the object or the class of the object captured in the image. After visual recognition, natural feature tracking enables determining the camera's pose relative to the object in real-time, e.g. [3, 4]. This is the basis for AR visualizations that we perform on a smartphone.

This poster does not aim to make any contributions in the fields of Visual Search or natural feature tracking, but instead combines these two existing technologies to enable a new interaction approach. This takes advantage of camera-equipped smartwatches, that recently became available off-the-shelf. To the best of the authors' knowledge, our proposed concept is the first to exploit a smartwatch to aid handheld Augmented Reality based on a smartphone, as will be explained in the following.

*e-mail:darko.stanimirovic@metaio.com
†e-mail:daniel.kurz@metaio.com

[1] http://www.junaio.com
[2] http://www.metaio.com/ar-logo

(a) (b) (c) (d)

Figure 2: Scanning an advertisement for a sofa with our prototype (a) results in brief digital information about the sofa, such as its price, being shown on the smartwatch (b). The app further offers an Augmented Reality view of the sofa (c), in which the advertisement is placed on the floor and the sofa is superimposed on top and at real scale on the display of the phone (d).

2 Proposed Approach

The advent of wearable computing devices provides new possibilities to deploy Augmented Reality technology. We propose a method that should run on a smartwatch and a smartphone simultaneously, which mutually exchange data. The main purpose of the smartwatch is to provide the user with a fast and comfortable mean to scan the environment using the built-in camera, without the need to take the smartphone out of the pocket. In the general scenario, a user wears a smartwatch and sees an object he or she is interested in. Our approach enables the user to determine if this object is AR-enriched or not. Therefore the smartwatch is used to capture an image of the object, which is automatically transferred to the smartphone via wireless link. The received image on the smartphone side is then matched against either a local or a remote Visual Search database of objects for which digital information is available. The response to the Visual Search query is transmitted back to the smartwatch and presented to the user.

For objects that are not contained in the database or which cannot be recognized, the display of the smartwatch indicates that the captured object is unknown as illustrated in figure 1 (a). If the object could be recognized and there is information related to it available, which may be provided to the user on the screen and/or by means of audio output of the smartwatch, the information is presented on the smartwatch, see figure 1 (b). Examples for such kind of information include purchase links, pricing information, brief customer reviews, or audio guides in a museum. In case an object is AR-enriched, as illustrated in figure 1 (c), the user is informed and can decide to start an AR experience using this object on the smartphone, as shown in figure 1 (d). Afterwards the user may stow the phone away and continue exploring objects with the smartwatch. There might be both digital content to be shown on the watch and AR content to be presented on the phone related to a single object. In this case the user may select which information to retrieve.

3 Implementation and Prototype

Our prototype implementation is deployed on a pair of Samsung devices: the smartwatch Galaxy Gear (SM-V700) and the smartphone Galaxy S5. Here, the computer program consists of two separate apps, executed in parallel on these two devices. The smartwatch app provides functionality for establishing communication with the smartphone app, taking pictures using the smartwatch camera, transmitting images to the smartphone app, and receiving responses from the smartphone. The smartwatch app is implemented in JavaScript and runs on the Tizen OS for Galaxy Gear. The communication relies on the Samsung Accessory Framework.

The smartphone app is responsible for establishing communication with the smartwatch app, and receiving images from the smartwatch. It further matches received images against a Visual Search database using Metaio Continuous Visual Search (CVS)[3]. It then transmits the recognition result and digital information related to it to the smartwatch. In case of visual augmentation of the object using Augmented Reality, it is further necessary to utilize camera pose tracking and rendering techniques on the phone.

Our workflow is as follows: the user triggers taking a photo by tapping on the screen and the image is then sent to the phone using Bluetooth. Once the Visual Search component receives the image, it attempts to find a correspondence in the Visual Search database containing encoded descriptions of the visual appearance of objects to search for. For AR-enriched objects, the object representations are usually linked to the corresponding tracking configurations and (3D) augmentation models. In this manner, Visual Search either gives a "not found" response, or a response containing related digital information, potentially including a configuration to track the object and virtual 3D models to be rendered registered with the object. The phone then transmits all information to be shown on the watch via Bluetooth to the smartwatch. It further keeps the remaining files, e.g. 3D models, in main memory to enable immediately running the AR experience on the smartphone.

Some impressions of the prototype implementation can be found in figure 2. Our current implementation takes on average 2.65 seconds from triggering a photo to displaying the Visual Search result on the watch. The majority of time is spent on communication (transmission of images (1.11 s), and background Android tasks (0.66 s)), which we aim to improve in future work.

4 Conclusion and Future Work

The presented combination of a smartwatch and a smartphone is the first implementation of a handheld AR experience aided by a smartwatch. Given rising popularity of such devices, it is expected that such techniques will become more common in the future. With increasing computational capabilities of smartwatches, we also expect to see a migration of certain categories of AR content from smartphones and tablet devices to smartwatches, e.g. visual guidance information and audio augmentations.

References

[1] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 1995.

[2] T. S. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *Proc. NIPS*, 1999.

[3] D. Kurz and S. Benhimane. Gravity-Aware Handheld Augmented Reality. In *Proc. ISMAR*, 2011.

[4] D. Kurz, T. Olszamowski, and S. Benhimane. Representative Feature Descriptor Sets for Robust Handheld Camera Localization. In *Proc. ISMAR*, 2012.

[5] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *Proc. ICCV*, 2003.

[3] http://www.metaio.com/visual-search