Antonius: A Mobile Search Engine for the Physical World

Markus Funk

University of Stuttgart

Pfaffenwaldring 5a

70569 Stuttgart

markus.funk@vis.uni-stuttgart.de

Albrecht Schmidt

University of Stuttgart

Pfaffenwaldring 5a

70569 Stuttgart

albrecht.schmidt@vis.uni-stuttgart.de

Lars Erik Holmquist

Yahoo! Labs

701 First Avenue

Sunnyvale, CA, USA

larserik@yahoo-inc.com

Abstract

We introduce Antonius, a wearable system that is able to detect real-world objects based on their visual appearance. Users can search for objects and be directed to their current location on a 3D map. Our formative evaluation indicates that 3D representation is preferred to 2D maps to show an object's location, and that both stationary and mobile systems would be beneficial depending on the scenario.

Author Keywords

Ubiquitous computing, visual object detection

ACM Classification Keywords

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

Introduction

In the digital world, we are used to being able to easily find information through a variety of search engines. Almost any piece of information or media can be found through a simple keyword search. In the physical world, on the other hand, we do not have similar abilities to search and locate items. Hence searching for objects in the real world is a common problem that is time consuming and may induce anxiety.

Figure 1. The wearable prototype. The cap includes a webcam for image capture and an OptiTrack marker for exact location tracking
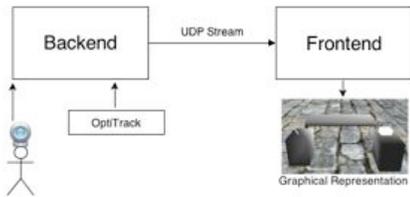

Figure 2. The architecture of mobile Antonius describing the interaction of the components.

In this work, we present a wearable prototype that allows the user to perform a keyword search on real-world entities to find their location. This would mean that you could ask a system, where you left your keys, what place a book was last seen, or how much food you have in the fridge.

Stationary Prototype

Previous systems have usually been based on tagging objects with visual or electronic tags (c.f. [6]). However, we wanted the system to integrate with the user's environment without any instrumentation of objects, which led us to select computer vision-based object detection as the technology to keep track of objects.

We built an initial prototype called Antonius[1], which uses stationary cameras that are placed in important places, e.g. the desk in an office. The system iterates over all available cameras and constantly compares the live video to previously registered objects using visual object detection. Once an object is detected, the last-seen location of the found object is written into a database. If the location of an object changes, the database entry changes accordingly. The user can query the database for the location of any registered object using a web-interface. The system uses a textual description and a 2D map to represent the location of the objects.

---

[1] We whimsically named the system after Saint Antonius, the Catholic "patron saint of lost things and missing persons" C.f. http://www.luckymojo.com/saintanthony.html (accessed May 2013)

Mobile Prototype

After evaluating the stationary system, we wanted to try a lightweight approach with less cameras. As argued in [1], objects do not move by themselves. There is always a user involved in the process of moving objects. Therefore, we posited that it would be sufficient to just instrument users instead of the whole environment. We also decided to improve the interface by creating a full 3D representation of the space as compared to the previous 2D view.

Concept

The concept of the mobile prototype is to equip the users with a wearable camera in order to detect with which objects they are interacting. Just like in the stationary case, the system compares images taken from the camera to previously registered images in order to identify those objects. Because the camera does not have a fixed location, we integrated the OptiTrack motion-capture system [2] to determine the indoor location of the camera within the testing area. All components are combined on a hat (see Figure 1).

Architecture and Technical Description

To implement the mobile version, we designed a two-tiered architecture: A backend, which does the visual object detection and tracks the position of the camera, and a frontend, which models the objects' location within a 3D representation of a testing area (see Figure 2). The backend compares the live image taken from a webcam to the registered objects' images. When a match is found, the object's ID and the current position of the object are sent to the frontend via UDP. The frontend then alters the position of the detected object in the memory and in the graphical representation according to the received UDP packet.
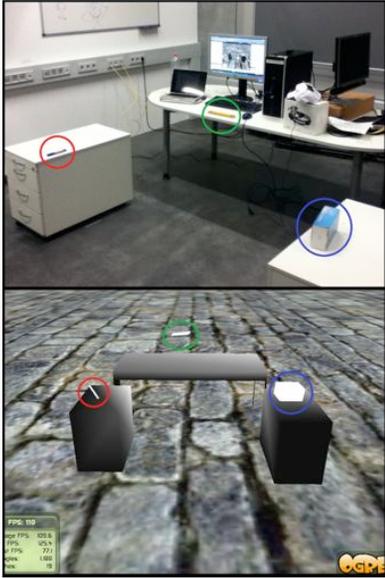
Figure 3. Top picture: The testing scenario in a lab environment. Bottom picture: The corresponding 3D representation of the testing scenario within the frontend.

Mobile Antonius uses OpenCV (Version 2.4.0) in combination with BRISK [4] to detect previously registered objects in pictures. In contrast to the stationary system, which uses SURF [5], we decided to use BRISK exclusively in this version because of its advantage in speed. Within a mobile setting speed is more important than accuracy because a registered object might only be within the field of view for just a glance of the eye.

The determination of the camera's position is done using the OptiTrack system, which is capable of tracking an OptiTrack marker (see Figure 1, where the marker is on top of the hat). Our OptiTrack system uses 6 infrared cameras, which are installed in the ceiling of the testing area. The corresponding software TrackingTools can stream the marker's location via multicast. The backend integrates the NatNet SDK (version 2.2), which is capable of receiving and processing the multicast packets.

Mobile Antonius uses the OGRE engine (version Vc10 1.8.1) [3] to display a 3D model of the testing area. Mobile Antonius uses the fixed assumption that an object is 40 cm in front and 45 cm below the marker's position. This distance was determined experimentally.

Evaluation
We conducted a formative study to evaluate the mobile prototype and to compare it to the stationary version. In particular we focused on comparing stationary and mobile concepts and the representation of object location. We recruited 5 participants through our university network (4 male, 1 female, average age 23). All participants were computer science students and would be considered as advanced technology users. As

a testing area, we used a lab setting containing two drawer cabinets and one table arranged as depicted in Figure 3. At the ceiling of the testing area 6 OptiTrack cameras were attached to a square-shaped steel stand in order to cover the whole area. The camera on the hat was connected to a computer via a 5m USB extension cable in order to support mobility of the user.

After explaining the purpose of the study, we showed pictures of the stationary system and explained the concept in detail. Then, we showed them the actual mobile system. The facilitator put on the system's hat and moved a box from the table to a drawer cabinet, and showed how the system updates. Then, he gave the hat to the subject to let them try the system themselves. A small object (a pen – see red circle in Figure 3) and a middle-sized object (a folding meter stick – see green circle in Figure 3) were introduced to show that the system is capable of multiple objects. When the group was familiar with the system, we conducted an open-ended interview.

All participants stated that they would be able to find objects based on the provided 3D representation. Compared to the 2D map in the stationary system, the participants thought that it is easier to map a 3D representation into the real world than a 2D map, presumably because the cognitive effort when linking the real world and the 3D representation is lower. Furthermore, a 3D representation provides information about the vertical positioning of an object's position within the room, which a 2D map can not. Also compared to a textual representation the 3D-representation was preferred. A textual representation only gives a rough idea, while a 3D representation provides an exact location.

When comparing the stationary Antonius system to the mobile Antonius system, the participants' views differed. Two participants preferred the mobile approach, primarily because they would have a "weird feeling" with fixed cameras in their living area. On the other hand, two other participants preferred the stationary approach, because they would not want to wear a hat all the time. One participant suggested a hybrid approach, where frequently used areas, e.g. the office, would be equipped with stationary cameras and less frequently used areas would require the user to wear the mobile system. With this hybrid approach, a user would not have to wear the system all the time.

To improve the system in the future, we want to run the 3D model on a head-mounted display or a smartphone, which might provide better directions towards finding the object based on the current location. Another improvement could be to equip the hat with multiple cameras in order to get a larger field of view. With this improvement the system could even detect objects the user is not directly interacting with. A major limitation is the OptiTrack system, which restricts the area of use. In future work it could be replaced with other indoor location techniques.

Conclusions

In this work we have shown that it is feasible to implement a wearable search engine for the physical world, which finds objects based on their optical features. In formative evaluations, we gathered valuable feedback and found that while users still find the proliferation of cameras "creepy", they could imagine using the system if privacy concerns were addressed. The evaluation shows that the users prefer a 3D representation of an object's location to a textual representation and a 2D map. Also, we found out that the users are more likely to use a mobile system in more private scenarios because the system can be undressed easily. We believe that if privacy implications are carefully considered and if the benefit of never losing an object again is valuable enough, users may be interested to adopt this system.

Acknowledgements

References

1. Pederson, T. Magic Touch: A Simple Object Location Tracking System Enabling the Development of Physical-Virtual Artefacts in Office Environments. *Personal Ubiquitous Comput.*, vol. 5, no. 1, pp. 54-57, 2001.
2. OptiTrack - http://www.naturalpoint.com/optitrack/ (accessed May 2013)
3. OGRE Engine - http://www.ogre3d.org/ (accessed May 2013)
4. Leutenegger, S., Chli, M., Siegwart, R.Y. BRISK: Binary Robust Invariant Scalable Keypoints. *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, 2011
5. Bay, H., Tuytelaars, T., Van Gool, L. SURF: Speeded Up Robust Features. *Comput. Vis. Image Underst.* 110, 3 (June 2008), 346-359.
6. Romer, K., Benedikt Ostermaier, Friedemann Mattern, Michael Fahrmair, Wolfgang Kellerer. Real-Time Search for Real-World Entities: A Survey. *Proceedings of The IEEE ,* vol. 98, no. 11, pp. 1887-1902, 2010.