Using Computer Vision Technologies to Make the Virtual Visible

Sebastian Büttner

Mobile Life Centre

Stockholm University

DSV, Forum 100

16440 Kista, Sweden

sebastian@mobilelifecentre.org

Tengjiao Cai

Mobile Life Centre

Uppsala University

S:t Olofsgatan 10

75105 Uppsala, Sweden

caitengjiao1987@hotmail.com

Henriette Cramer

Mobile Life Centre

SICS

Isafjordsgatan 22

16440 Kista, Sweden

henriette@mobilelifecentre.org

Mattias Rost

Mobile Life Centre

SICS

Isafjordsgatan 22

16440 Kista, Sweden

rost@sics.se

Lars Erik Holmquist

Mobile Life Centre

SICS

Isafjordsgatan 22

16440 Kista, Sweden

leh@sics.se

Abstract

Augmented reality (AR) applications typically overlay information about the user's environment in their mobile phone's camera view. Rather than only using the camera view as a backdrop for information presentation, however, AR applications could also benefit from using the camera as a sensor to a greater extent. Beyond using visual data for markerless tracking, AR applications could recognize objects and provide users with information based on these objects. We present two applications that use the camera as a sensor: *Pic-in* and *SubwayArt*. The first allows users to check-in on location-sharing service foursquare by taking a picture of the venue they are at. The second provides users with information about artworks in the Stockholm subway system by combining localization and computer vision techniques.

Keywords

Mobile augmented reality, computer vision, camera as a sensor

ACM Classification Keywords

H5.m. Information interfaces and presentation: Miscellaneous.

General Terms

Design, Human Factors

Introduction

In the last years, commercial mobile augmented reality (AR) application gained ground. Services like Layar, the Wikitude World Browser and Junaio create virtual layers on top of the real world and provide users with locative media inside the created hybrid space. In these environments the mobile phone camera is used to capture visual information of the real world, which is augmented by virtual objects and displayed to the user.

While the presentation of information integrates virtual with physical world showing this mixed reality to the user by overlaying views, most commercial services are not bridging the gap between physical and virtual objects when it comes to making information visible that relates to the real-world objects in the camera view. The selection of information in the mentioned services is mainly based on a choice of the information source (e.g. Wikipedia or foursquare) and position and direction. Visual data from the camera is ignored for selecting information in many commercial systems even though it is available. Therefore existing commercial applications are able to show directions to objects but ignore the possibility of augmenting objects that are in the view of a user.

As an example the screenshot from the Wikitude World Browser in figure 1 shows the directions of different venues taken from the database of location-sharing service

foursquare. All virtual information close to the user's position is projected into the view. For certain use cases, e.g. location-sharing, this might be confusing since

figure 1: foursquare venues in Wikitude World Browser

the virtual overlay presents information outside of the visible scope of the user.

Other systems use visual data for recognizing movements in the physical space (markerless tracking), but physical objects are often simply used as 'natural markers' without making object-related information visible. However we envision mobile AR systems where users do not need to predefine their wishes of information. Information could be selected based on objects recognized in the camera view, making digital information visible that relates to them.

In this paper we state our position that the visual data captured by the mobile phone camera can be used and processed with computer vision technologies to find and select information about real-world objects that are visible to the user. We describe our earlier explorations about bridging the gap between physical and virtual world and describe our experience in the field of computer vision. We present two recently implemented applications that use the camera as a sensor to provide users with relevant information. The two applications show the capabilities that mobile phones nowadays already have.

We would like to engage a discussion, how future mobile AR systems can be designed to use the camera as a sensor. We envision that future applications will not only receive a video stream that is augmented and looped through to the user, but also be used to make virtual representations of objects visible to the user.

Related Work on Computer Vision

Computer vision technologies have been used since years in AR to enable markerless tracking, e.g. in the work of Neumann and You [6]. Basis for this tracking as well as for recognizing objects are local features in the pictures, e.g. points or regions that distinct from the other parts of the image. These features can be mathematically described

and matched to features from other pictures, which allows detection of movement or recognition of objects. Two examples for algorithms achieving this feature description are Lowe's [5] SIFT and Bay et al.'s [1] SURF algorithm.

There have been earlier systems in research that make information about objects in the camera view available: Cuellar et al. [3] present a system that shows tourist sight information when users point their camera phone to it. Their 2D AR system recognizes local features based on a combination of visual and positioning data [8]. Omerčević and Leonardis [7] have been working on a system that is identifying objects based on their visual appearance and presenting them in a 2D AR view. A study of their system showed positive reactions of users even though the system did not work in real-time and took 15-50 seconds to return results [7]. With our implementations we show that using these techniques is now actually feasible in customary mobile phones and close to real-time.

Our Computer Vision Explorations

We will now present two applications that use the camera as a sensor to retrieve and show information about the user's environment. We implemented the applications recently to explore the possibilities of using this visual data in connection with location data for object and place recognition on mobile devices. Even though the applications are not AR applications in the common sense, they are able to make the virtual world more visible and demonstrate possibilities for AR applications to select information based on objects recognized in the camera view. Both applications were entered in the Ericsson Application Awards Competition (ericssonapplicationawards.com). In the competition among 158 applications, Pic-In made the 3rd place in the company section and SubwayArt reached the semi-final round (top 7) of the student section.

SubwayArt
Our first exploration is the application SubwayArt, which is shown in figure 2. Users can take a picture of any of the art pieces in the Stockholm subway system to retrieve information about it. The service uses GSM net based positioning to cut down the object recognition problem. In a first evaluation our application
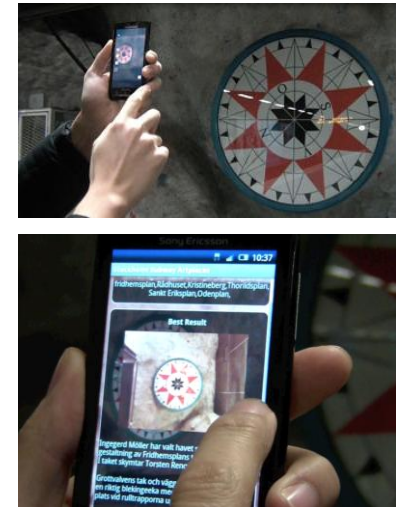

figure 2: SubwayArt

showed a reliable and fast (less than a second) recognition and we are optimistic that the application could be adapted to recognize art pieces in real-time from a video stream within an AR environment. A demo video can be found at vimeo.com/22601310

Pic-In
In previous work [2] we explored how we could link virtual and real venues in location-sharing. We used 2D barcodes to enable people to check-in by scanning the visual tag. We now aim to skip this middleman and directly use the camera as a sensor.
Pic-in is a system that allows users to check-in to location-sharing service foursquare by taking a picture of a location. The application is shown in figure 3. It combines the location data with the image data from the camera to determine the semantically named place of a user. The system is trained and improved using crowd-sourcing:
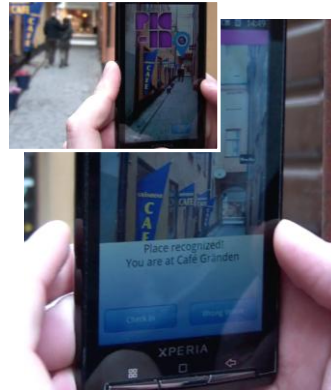
figure 3: Pic-In

Users can correct wrong information or add new information, if the system is not able to determine a venue. In this way the system makes not only 'invisible' information visible but also allows users to affect the invisible data. The application will be launched end of June in the Android market to allow an evaluation in a large scale. A demo video can be found at vimeo.com/22229315

Conclusion and Challenges

We believe that using the camera as a sensor to capture information about the physical environment can further merge physical with virtual world in a mobile AR environment. We believe that future AR applications will recognize physical objects based on their visual appearance and present information based on these objects. Indeed, we presented two applications that already take advantage of these possibilities.

We would like to engage discussions on different issues that come up with the use of visual data for physical selection of information in an AR environment:

▪ How can we design systems that are making more sense out of the objects that are around the user and make information visible based on those objects?

▪ How can we allow not only visualization of the 'hidden' information about the user's environment within AR

applications, but also design interactions that allow users to change this information in an engaging way?

▪ AR is now mostly focused on the visual dimension, how can we use other modalities? Are there other ways of presenting the 'hidden' information about the user's environment and allowing users to interact with this information?

References

[1]   Bay, H., Tuytelaars, T., and Van Gool, L. SURF: Speeded Up Robust Features. In *Lecture Notes in Computer Science*, vol. 3951 (2006), 404-417.

[2]   Büttner, S., Cramer, H., Rost, M., Belloni, N., and Holmquist, L. E. (2010). φ²: Exploring physical Check-Ins for Location-Based Services. In *Ext. Abstracts UbiComp 2010.*

[3]   Cuellar, G., Eckles, D., and Spasojevic, M. Photos for Information: A Field Study of Cameraphone Computer Vision Interactions in Tourism. In *Proc. CHI 2008*.

[4]   Höller, N., Geven, A., Tscheligi, M., Paletta, K., Amlacher, P. L., and Omerčević, D. Exploring the urban environment with a camera phone: Lessons from a user study. In *Proc. MHCI 2009*.

[5]   Lowe, D. G. Object Recognition from Local Scale-Invariant Features. In *Proc. ICCV* 1999.

[6]   Neumann, U., and You, S. Natural feature tracking for augmented reality. In *Transactions on Multimedia*, vol. 1, no. 1 (1999), 53-64.

[7]   Omerčević, D., and Leonardis, A. Hyperlinking reality via camera phones. In *Machine Vision and Applications*, vol. 22, no. 3 (2010), 512-534.

[8]   Takacs, G., Chandrasekhar, V., Gelfand, N., Xiong, Y., Chen, W., Bismpigiannis, T., Grzeszczuk, R., Pulli, K., and Girod, B. Outdoors augmented reality on mobile phone using loxel-based visual feature organization. In *Proc. MIR 2008*.