Exploring the urban environment with a camera phone: Lessons from a user study

Norman Höller, Arjan Geven, Manfred Tscheligi

CURE-Center for Usability Research and Engineering

Hauffgasse 3-5, A-1110 Vienna

{hoeller, geven, tscheligi}@cure.at

Lucas Paletta, Katrin Amlacher, Patrik Luley

Joanneum Research, Steyrergasse 17-19/ 8010 Graz

{lucas.paletta, katrin.amlacher, patrick.luley}@joanneum.at

Dusan Omercevic

University of Ljubljana

Kongresni trg 12, 1000 Ljubljana

dusan.omercevic@fri.uni-lj.si

ABSTRACT

We present a study investigating two novel mobile services supporting querying for information in the urban environment using camera equipped smart phones as well as two different ways to visualize results – icon-based visualization and text-based visualization. Both applications enable the user to access information about an object by snapping a photo of it. We investigate how users would use a photo-based tourist guide in a free exploration setting in general as well as the acceptance/preference of two different ways to visualize results.

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: GUI, Interaction Styles

General Terms

Performance, Design, Experimentation, Human Factors.

Keywords

mobile devices, computer vision, augmented reality

1. INTRODUCTION

When we travel to unfamiliar cities and places, we use a tourist guide or the Internet to get information about buildings, streets, restaurants, and places to shop. This work presents two systems that provide support in ubiquitous interaction with the real world, with immediate access to virtual information spaces representing a reading-glass for stories behind the urban environment and run on a lightweight camera phone.

The two applications presented enable users to get information about things they see (e.g., buildings, neighborhoods) by simply taking a photograph of it. They return information about the photographed object and its surroundings in distinct ways. The two systems are (a) "Object Recognition" (OR) – built around geo-indexed object recognition – and (b) "Hyperlinking Reality" (HR) – built around purely image-based recognition.

2. RELATED WORK

Mobile image-based recognition and localisation have recently been proposed in terms of mobile vision services for the support of urban nomadic users. HPAT (hyper-polyhedron with adaptive threshold) indexing provided one of the first innovative attempts on building identification proposing local affine features for object matching [5]. An image retrieval methodology for the indexing of visually relevant information from the web for mobile location recognition was introduced in [6]. Exploiting knowledge about a current geo-context in a probabilistic framework using attentive, geo-indexed object recognition was done by [1]. Powerful and computationally demanding computer vision techniques based on local invariant features are described by [3]. The approach to image matching was pioneered by [4] and [3]. Similar work to this focused on learning of user behavior in regard of embodied interaction with a mobile photo-annotation system [2], evaluating it with a guided tour on campus.

MobileHCI'09, September 15 - 18, 2009, Bonn, Germany.
ACM 978-1-60558-281-8.

3. SYSTEMS UNDER EVALUATION

We evaluated two systems, Object Recognition (OR) and Hyperlinking Reality (HR).

3.1 Object Recognition (OR)

When the user takes a picture with the OR application, s/he receives a picture of that very object annotated with text and further links (Figure 1a). The server's matching algorithm cuts down the visual search space into a subset of relevant object hypothesis based on contextual processing of multi-sensor information.

3.2 Hyperlinking Reality (HR)

A picture taken with the Hyperlinking Reality (HR) application is returned with icons placed on the objects that are annotated. Selecting those objects reveals information about them (Figure 1b). Differences despite the common purpose in the two applications are summarized in Table 1.

4. STUDY DESIGN

We addressed the following research questions:

- What buildings do participants choose to photograph and how do they take the photographs of these?

- What is the preferred visualization of results?
- How much feedback do users need to receive when the system is processing the information? How important is response time for users of these systems?
- In which real-life applications can users imagine to use this technology? What are areas for improvement and further development?

To answer research questions, we created a setting in which 16 participants (9 female, 7 male; aged 22 - 30) could experiment with the technology in an urban environment measuring approximately 400 by 100 meters.

Thinking out loud (transmitted over bluetooth) and shadowing was used to understand what each user was doing and thinking during the free exploration.

A short on-site interview with each user to asses the general first impressions of the systems was conducted after exploration. Additionally two focus groups were held.

5. RESULTS

Participants took an average of 7.1 photos with the OR application and 4.6 photos with the HR application. 13 out of the 16 users reported interest in installing the application on their own mobile phone if it would be available.

The amount of time that the system needed to generate results was considered almost acceptable for OR but for some users too long for HR. Particularly the lack of feedback during the server-side processing made it difficult for the user to know what was going on and caused irritation as well.

The icon-based visualization of the results with the HR application (Figure 1b) was preferred by users over the text-based visualization from the OR application (Figure 1a). Different possible scenarios for use were mentioned during the focus groups, mainly in the context of shopping, concert tickets and tourism.

6. CONCLUSIONS

Users reacted positively on the applications and were highly motivated to take advantage of the intuitive interface, with some important remarks regarding technical features (response time, reliability), information visualization and future applications of the technology.

7. ACKNOWLEDGMENTS

This work is part of the project MOBVIS and has been partly funded by the European Commission under contract IST-FP6-511051. We would like to thank all partners of the MOBVIS consortium and the European Commission for support.

Figure 1: Annotated photos from the OR (a) and HR applications (b)

Table 1: Comparison of usability aspects between the investigated system functionalities

	OR	HR
Annotation of urban objects	*Single* objects (e.g., facades) selected by the user	*Multiple* objects in urban environment
Visualization of annotation	List of information including URLs in response message	Icon-based annotation with URLs, directly on query image
System response time	~15 sec. (1 MP image, 1 CPU)	~50 sec. (3 MP image, 8-core CPU)
Use of geo-information	Geo-indexed object recognition	Purely image based recognition
Position annotation	GPS based position without annotation.	Position and orientation of user based on single image

8. REFERENCES

[1] Amlacher, K. and Paletta, L., (2008). An Attentive Machine Interface Using Geo-Contextual Awareness for Mobile Vision Tasks, Proc. European Conference on Artificial Intelligence, ECAI 2008, pp. 601-605.

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

[3] Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. IJCV 60(2), pp. 91-110.

[4] Schmid, C. and Mohr, R. (1997). Local gray-value invariants for image retrieval. IEEE PAMI 19(5), pp. 530-535.

[5] Shao, H., Svoboda, T., and van Gool, L. (2003). HPAT indexing for fast object/scene recognition based on local appearance. *Proc. International Conference on Image and Video Retrieval*, CIVR 2003, pp. 71–80.

[6] Yeh, T., Tollmar, K., and Darrell, T. (2004). Searching the web with mobile images for location recognition. *Proc. IEEE Computer Vision and Pattern Recognition*, CVPR 2004, pp. 76