Visualizing Locations for a Search Engine for the Physical World

Markus Funk, Robin Boldt, Marcus Eisele, Taha Yalcin, Niels Henze, Albrecht Schmidt

Institute for Visualization and Interactive Systems, University of Stuttgart

Abstract

Today, search engines on the Internet allow us to instantly find almost any kind of information or media. Unfortunately, the same capabilities are not available in the physical world, except for in industrial settings such as automated warehouses. Therefore, we implemented a stationary prototype that supports finding objects in everyday environments. The location is represented using a web-based search interface. In a user study, we compared four different representations for our stationary object finding system. Results favor a last seen image representation and a 2d map and show that more complex representations are not suitable for representing results of a search engine for the physical world. The participants were throughout positive about using a system like this in their daily lives, although there were concerns for privacy. Based on these results, we argue that scalable physical search engines can be built for everyday environments; that searching for physical objects can be made understandable to users; and that there are many practical and commercial cases where such systems would be beneficial.

1 Introduction

In the digital world, we are used to being able to easily find information through a variety of search engines and information retrieval tools. Almost any piece of information or media created by man can now be located through a simple keyword search. In the physical world, on the other hand, we do not have similar abilities to search and locate items.

In this paper, we ask the question if it would be possible to apply the same principles that currently power digital search engines to the physical world? If so, what happens when it becomes possible to find literally everything? Imagine if the location of any physical object would be instantly available for searching, in the same way that popular Internet search engines allow us to find information on the World Wide Web today. This would mean that you could ask a system where you left your keys, or what place a book was last seen.

This could have long-ranging effects on human behavior. In the virtual domain, desktop search engines have already changed the behavior of users. For many users carefully organized hierarchical structures of folders, as well as specific Internet addresses, have become irrelevant. If the vision of a physical search engine becomes reality, it would facilitate not just the solution of trivial problems, such as finding a missing item. It could potentially change the way we interact with the world. Currently, humans use many strategies to manage and arrange the objects around us (Kirsh, 1995). If that was not necessary, we could instead organize objects according to practical or aesthetic criteria. We could even stop caring exactly where an object is, as long is it is findable – such as all those books lent to friends, which are currently to be considered lost. Furthermore, knowing about the things we use and their locations would allow completely new systems for personalization and context-aware services, as well as facilitate fundamentally new business models. We could potentially apply all the algorithms currently in use to recommend and target content on the Internet to the physical world.

For those kinds of search engines for the physical world, the representation of search results is important. In this paper, we want to find out which representation of locations is most suitable for a room-level environment like a living room or an office room. Traditionally, location is represented through 2D, 3D or textual representations, which need knowledge about the environment. Recently it has been proposed to represent indoor location through a last seen image representation. In this paper, we compare these four representations by conducting a user study inside a room-level test scenario. We show that an image-based representation of search results that does not require a model of the environment and no exact location of objects results in the same performance as more complex representations. Consequently, scalable and model-independent real-world search engines become feasible. In the following, we give an overview about related work, describe the four representations and present the results of the study.

2 Related Work

Search engines for the Internet are today a well established part of the Information Retrieval field, and have become the foundation of successful companies such as Google and Yahoo!. The vision of search in the real world has initially been tightly linked to wearable cameras and the search and indexing of personal first-person view recordings (Mann, 1998). An example of the use of video retrieval techniques in this context can be found in (Tancharoen et al., 2005). More general concepts for search engines for the physical world are a much newer concept and have emerged over the last few years with advances in camera technologies, sensors, and computer vision algorithms. On a conceptual level, they can be considered a part of the Internet of Things (IoT) vision, where real-world physical objects get connected and equipped with processing and sensing capabilities (Mattern & Floerkemeier, 2010). Typically, this has been achieved by attaching electronics or tags directly to the objects that are to be monitored. Early examples of systems that use tags to connect physical objects to digital information include Webstickers (Ljungstrand et al., 2000), where a printed barcode sticker could be attached to any object to associate it with an online URL; and the collection of systems that demonstrate how electronic tags can bridge physical and virtual worlds as seen in (Want et al. 1999) However, neither of these systems kept track of the location of tagged objects, but rather used the objects to access information or initiate actions, hence providing physical objects as links or pointers to information (Holmquist et al., 1999). This means that a coarse location of an object is known only at the time when its tag is read and the whereabouts of the reader is (Song et al., 2007).

Going beyond simple proximity information, a location technology has to be included. There are basically two options: either that tags can locate where they are, or the elements that read the tags can provide location information. Are variety of solutions are possible here. The combined use of the commercial Ubisense asset tracking system and visual markers is explored in (Song et al., 2007). Changes in the location and position of objects can be measured using passive RFID tags, which is experimentally assessed in (Fishkin et al., 2004) A reverse approach for locating objects is suggested in (Bohn, 2008), where the infrastructure (e.g. the floor) contains a dense grid of RFID tags and the object includes a reader. By reading the tags and consulting a map in which the tag locations are noted, an object can determine its location. Recently, the idea of using tags and other means to locate and search for entities in the real world has been explored in several projects. A comprehensive survey of such tag-based real-world search engines can be found in (Roemer et al., 2010).

For smart environments, a number of systems for indexing and searching objects have been explored. SearchLight presents a search function for pervasive environments using a ceiling-mounted camera/projector system (Butz et al., 2004). A system with a similar objective, but using ultrasound and RFID is presented in (Nakada et al., 2005). Another search engine for everyday objects that are tagged with RFID is presented in (Nickels et al., 2013). The paper describes a hierarchical approach to the search process and requires the instrumentation of furniture with readers. All of these approaches typically require a significant infrastructure and an augmentation of all objects with RFID or visual tags. One example that does not require tagging is Distributed Image Search (DIS) (Yan et al., 2008), which allows for searching of visual features using a submitted image. However, DIS does not build a database of known objects but acts purely by identifying visual features in the image stream. The Antonius system (Funk et al., 2013) equips the user with a wearable camera, which is constantly recognizing previously registered objects.

Location can be represented *physically*, *symbolically*, *absolutely*, or *relatively* (Hightower & Borriello, 2001). A taxonomy for visualizing location-based information is introduced by (Suomela & Lehikoinen, 2004). They distinguish between the dimensions (3D, 2D, 1D, and none) and the view: first person view and third person view - where first person is the representation of location from a user-centric view and third person is the representation of location combined with the user's current position. The taxonomy is denoted as MV (m, v), where m is the dimension and v is the view.

Different techniques to visualize an object's indoor location would require different technical implementations. Presenting the location of an object in a 3D representation, for example, requires a localization technique and a model of the environment. In contrast, presenting the location by providing an image of the surrounding of the object's current location (Funk et al., 2014) requires a camera system. Thus, it is important to learn about the differences between representation techniques for object's indoor locations to guide the technical development of object localization systems.

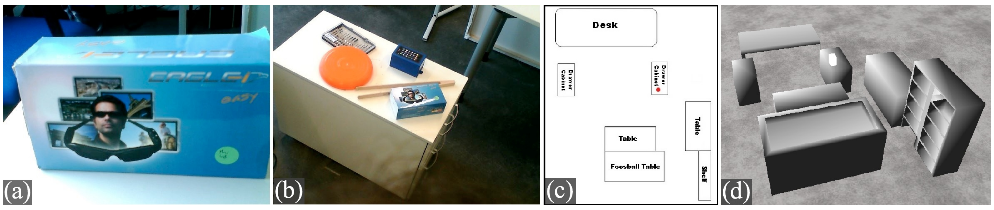

Figure 1: Representations used in our prototype: (a) Reference image that is displayed to the user to find an object and to initialize the image search. (b)–(d) Representations displayed in our prototype to find an object: (b) Last seen image representation, (c) 2D map representation with the object as red dot, and (d) full 3D representation of the demo scenario (sought object is represented as a white box).

3 Representation of Location

In our prototypical implementation, we compare four different representations (see Figure 1) of search results in context of a real-world search engine. It is designed to present the search result using one of the four representations. To describe the four representations, we use the taxonomy by (Suomela & Lehikoinen, 2004) and the categories of representing location after (Hightower & Borriello, 2001).

Textual representation

For the textual representation, we defined logical zones within the demo scenario (Kirsh, 1995). The zones are located on furniture placed in the room and named accordingly. Also, there is a huge zone, which reaches over the whole room. When zones are included in a larger zone, including zones are written after the parent zone. E.g. a shelf, which is in the lab will be represented as ”lab–>shelf”. This symbolic representation is in the category MV(0,0).

2D representation

The 2D representation uses a 2D map of the demo scenario (see Figure 2). The map is aligned north-up (up on the map corresponds to north in the environment). The location of the sought object is denoted as a red dot. The user's position is not displayed, as we use a MV(2,1) first person view with absolute representation.

3D representation

For the 3D representation, we created a 3D model of the demo scenario using Google SketchUp 8. The 3D representation is not part of the web-application as we used the OGRE engine to display the 3D model. Sought objects are displayed as a white box at the exact 3D location as in the real world. The user is able to move freely through the 3D space of this absolute representation viewing from a MV(3,1) first person view.

Last seen image

As an alternative representation, we use the last seen image. It displays the last picture, in which the sought object was recognized. The image also shows surrounding objects or furniture and gives the user contextual information about the location of the object (see Figure 2). This relative representation can be categorized as MV(3,1).

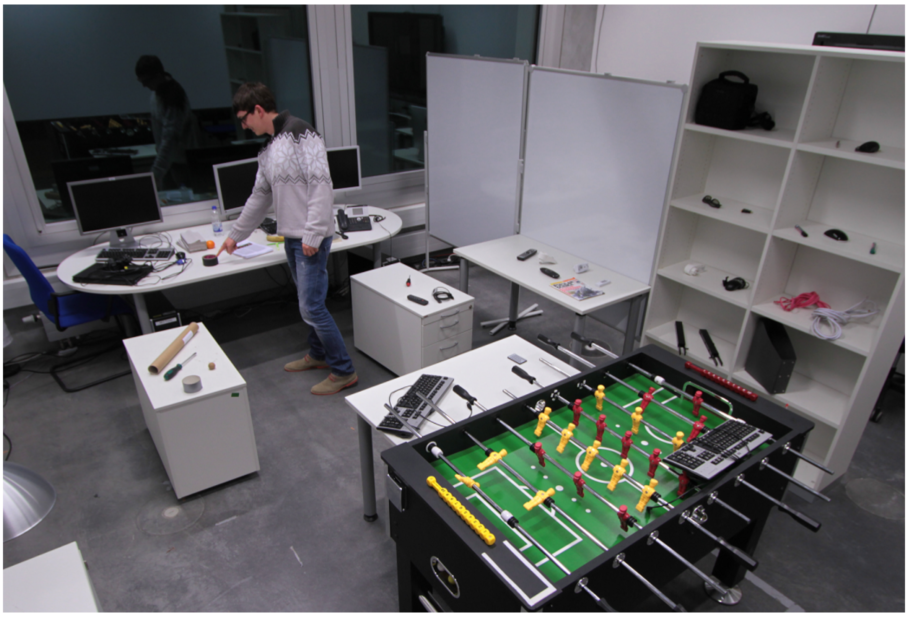

Figure 2 – A participant is trying to find an object in the demo scenario.

4 Evaluation

We conducted a controlled lab study to compare the 4 location representations in the context of a real-world search engine. In the study, the participants had to locate objects using a real-world search engine with each of the representations.

Method

We conducted the study using a repeated measures design with the 4 search layouts as independent variable. The order of the conditions was counterbalanced using Balanced Latin Square. We used objective and subjective measures as dependent variables. As objective measures, we used the time a participant needs to understand where an object is located (understanding time) and the time a participant needs to find and touch the object (task completion time). To determine the understanding time, we measured the time the participant spent in front of the PC using the search interface with the current location representation. The task completion time was measured from the moment the participant left the PC until the sought object was touched. In addition, we collected qualitative subjective feedback from the participants using the NASA TLX (Hart & Staveland, 1988) and the SUS (Bangor et al., 2008) questionnaire after each condition.

First, participants were given time to make themselves familiar with the demo scenario (see Figure 2). The scenario contained about 50 objects, which were placed onto furniture in order to make the search task harder. In addition, we used 12 searchable objects, which were not part of the scenario that was shown to the participant at the beginning. Each of the searchable objects belongs to one of the categories: small, medium or large. To prevent the participant from memorizing an object's location, a searchable object was placed into the scenario secretly after the participant was familiar with the scenario. The order of the searchable objects' sizes was also counterbalanced in each condition. For each of the four representations, the participant had to search for three objects - one belonging to each category. Therefore, the participant used our application, which displayed a reference image and one of the four location representations. After an object was found, it was removed from the scenario and a new object was placed without the participant seeing the new location. When the participants found three objects using one representation, they were asked to fill the SUS and TLX questionnaires.

We recruited 28 participants (9 female, 19 male). The participants were between 11 and 44 years old (M=26.1, SD=6,55). Most of the participants were students with various majors, e.g. computer science, international business administration or mechanics. In an initial questionnaire, 26 participants stated to be searching for something at least once per week. In fact, 17 participants stated to search for an object daily.

Results

Including welcoming the participant, a short introduction to the topic and the demo scenario and all questionnaires, the study took around 30 minutes per participant. We used an analysis of variance (ANOVA) to compare the conditions with paired t-tests as follow-up post-hoc tests. Bonferroni correction is used to prevent inflation of type I errors.

Detailed descriptive statistical results are shown in Table 1. For all conditions participants reported a low task load between 23.53 and 28.28 on the NASA TLX scale (between 0=no task load and 120=high task load). An ANOVA revealed no significant differences between the conditions for the NASA TLX (F(3, 81)=1.277, p=0.288). With scores between 74.82 and 80.71 participants rated all search layouts similarly high on the SUS (between 1=not usable and 100=highly usable) and an ANOVA revealed no significant differences (F(3, 81)=1.304, p=0.279). The average task completion time is between 5.45s and 6.00s. Again, an ANOVA did not reveal a significant difference between the conditions (F(3, 81)=0.934, p=0.428).

On average, participants needed 9.94s to understand the textual description, 10.66s to understand the image, 9.41s to understand the 2D-map, and 12.63s to understand the 3D-map. An ANOVA revealed a significant difference for the time participants needed to understand the search layouts (F(3, 81)=8.740, p<0.001). Follow-up post-hoc test revealed that participants needed significantly more time to understand the 3D scene compared to the textual description (p=0.006) and the 2D-map (p<0.001).

Representation	**TLX-Score**	**SUS-Score**	**Task Completion Time**	**Understanding Time**
Textual Description	28.28 (SD=19.16)	74.82 (SD=17.11)	5.68s (SD=1.38)	9.94s (SD=3.74)
Last Seen Image	27.67 (SD=16.59)	79.19 (SD=12.24)	5.45s (SD=1.05)	10.66s (SD=4.45)
2D-Map	23.53 (SD=14.22)	80.71 (SD=14.65)	6.00s (SD=2.49)	9.41s (SD=4.12)
3D-Scene	24.50 (SD=14.88)	78.39 (SD=14.06)	5.61s (SD=2.40)	12.63s (SD=5.65)

Table 1: Results of the user study. All scores and times are the arithmetic mean over all participants.

5 Discussion

The results of the user study reveal, that all four representations have no significant differences in task load, usability and task completion time. Only the understanding time of the 3D-scene representation proved to be significantly worse than the other representation's understanding time. The results are so far independent from the technology that is used to implement a real-world search engine. According to the results, a real-world search engine does not need to implement a 3D scene or an accurate floor plan of the area to display search results as they are needed by the 3D, 2D and textual representation. It can be argued that if a wearable camera is used to recognize an object within a room a last seen image representation would be the easiest representation to implement. It might be sufficient to provide a room-level location information combined with a last seen image representation rather than implementing a model of the environment.

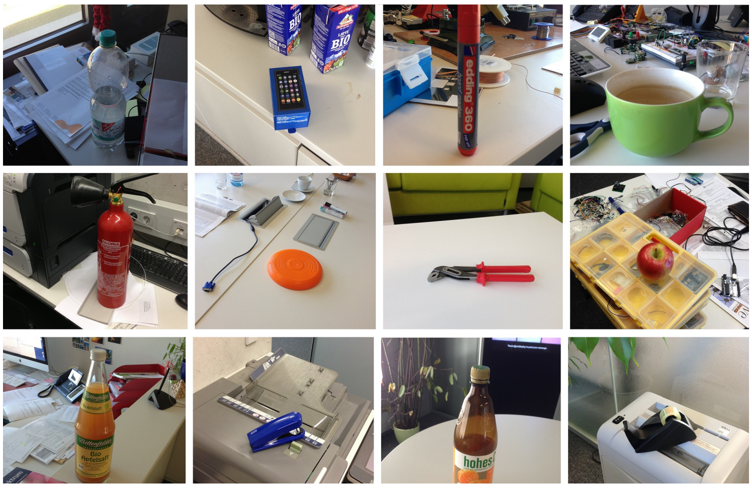

Figure 3 – An overview of all used objects in our contextual follow-up study. In the images there is enough contextual information that detailed location information might not be needed.

6 Contextual Follow-up Study

We conducted a follow-up study to further investigate how accurate the simpler image-based representation is. Therefore, we took pictures of 20 objects in an office environment. The office consists of 12 rooms. We recruited 10 participants (1 female, 9 male, average age 33.7, SD=6.37), who are familiar with the office environment, and showed them the pictures. The participants had to name the room, in which the object is located. Overall the participants could name 18,6 (SD=1.07) of the 20 rooms in the pictures correctly, which is a success rate of 93%. The most common error, which was made by 6 participants, was a picture taken in front of a shelf, where just the content of the shelf was visible. The results of the study show that participants, who are familiar with the scenario, can identify an object's position if the location information is vague, e.g. ”in this building” by using the contextual information provided in a last seen image.

7 System Recommendation

Based on the results of both of our studies, we recommend a camera-based search engine which can be installed firmly in an environment, e.g. a living room or running on a wearable computer e.g. on smart glasses. By combining a last seen image representation of search results and a room-level indoor location information e.g. WiFi-based, a wearable search engine for the real world can function in new environments without having a model of the scene and without using a precise indoor-location system.

8 Conclusion

In this paper, we compared four representations for search results of a real-world search engine. In our study participants were able to locate physical objects based on an image-based representation. In two studies we found that an image-based representation is as suitable for presenting real-world search results as 2D, 3D or textual representation. Furthermore, an image-based representation does not need any model of the environment and does not need precise location information. We believe that by using an image-based representation of search results, a scalable and model-independent real-world search engine can be built.

Acknowledgements
This project is sponsored by the German Research Foundation (DFG) in the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart.

References

Bangor, A., Kortum, P. T., and Miller, J. T. (2008) An empirical evaluation of the system usability scale. Intl. Journal of Human–Computer Interaction 24.

Bohn, J. (2008) Prototypical implementation of location-aware services based on a middleware architecture for super-distributed RFID tag infrastructures. *Personal and Ubiquitous Computing*, 12(2), 155-166.

Butz, A., Schneider, M., and Spassova, M. Searchlight–a lightweight search function for pervasive environments. In *Pervasive Computing*. 2004.

Fishkin, K., Jiang, B., Philipose, M., & Roy, S. (2004) I sense a disturbance in the force: Unobtrusive detection of interactions with RFID-tagged objects. *UbiComp 2004: Ubiquitous Computing*, 268-282.

Funk, M., Schmidt, A., and Holmquist, L. E. (2013) Antonius: A mobile search engine for the physical world. In *Adj. Proc. of Ubicomp*.

Funk, M., Boldt, R., Pfleging, B., Pfeiffer, M., Henze, N., & Schmidt, A. (2014) Representing indoor location of objects on wearable computers with head-mounted displays. *In Proc. of the 5th Augmented Human International Conference*, 18-22.

Hart, S. G., and Staveland, L. E. (1988) Development of nasa-tlx: Results of empirical and theoretical research. *Human mental workload* 1.

Hightower, J., and Borriello, G. (2001) Location systems for ubiquitous computing. *Computer 34*.

Holmquist, L.E., Redström, J. and Ljungstrand, P. (1999) Token-Based Access to Digital Information. Proc. *Handheld and Ubiquitous Computing 1999,* Springer-Verlag.

Kirsh, D. (1995). The intelligent use of space. *Artificial intelligence*, *73*(1), 31-68.

Leutenegger, S., Chli, M., & Siegwart, R. Y. (2011, November). BRISK: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (pp. 2548-2555). IEEE.

Ljungstrand, P., Redström, J. and Holmquist, L.E. (2000) WebStickers: using physical tokens to access, manage and share bookmarks to the Web. In *Proceedings of DARE – Designing Augmented Reality Environments,* ACM Press.

Mann, S. (1998) 'WearCam' (The wearable camera): personal imaging systems for long-term use in wearable tetherless computer-mediated reality and personal photo/videographic memory prosthesis. In *Wearable Computers, 1998. Digest of Papers. Second International Symposium on (pp. 124-131).* IEEE.

Mattern, F., Floerkemeier, C. (2010) From the Internet of Computers to the Internet of Things. In: Kai Sachs, Ilia Petrov, Pablo Guerrero (Eds.): *From Active Data Management to Event-Based Systems and More.* LNCS, Vol. 6462, Springer, pp. 242-259.

Nakada, T., Kanai, H., & Kunifuji, S. (2005) A support system for finding lost objects using spotlight. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services* (pp. 321-322). ACM.

Nickels, J., Knierim, P., Könings, B., Schaub, F., Wiedersheim, B., Musiol, S., & Weber, M. (2013). Find my stuff: supporting physical objects search with relative positioning. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (pp. 325-334). ACM.

Roemer, K., Ostermaier, B., Mattern, F., Fahrmair, M., Kellerer, W. (2010) Real-Time Search for Real-World Entities: A Survey. *Proceedings of The IEEE ,* vol. 98, no. 11, pp. 1887-1902.

Song, J., Haas, C. T., & Caldas, C. H. (2007). A proximity-based method for locating RFID tagged objects. *Advanced Engineering Informatics*, *21*(4), 367-376.

Suomela, R., and Lehikoinen, J. (2004) Taxonomy for visualizing location-based information. *Virtual Reality 8*.

Tancharoen, D., Yamasaki, T., & Aizawa, K. (2005). Practical experience recording and indexing of Life Log video. In *Proceedings of the 2nd ACM workshop on Continuous archival and retrieval of personal experiences* (pp. 61-66). ACM.

Want, R., Fishkin, K.P., Gujar, A. and. Harrison, B.L. (1999) Bridging physical and virtual worlds with electronic tags. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'99),* ACM Press.

Yan, T., Ganesan, D., & Manmatha, R. (2008). Distributed image search in camera sensor networks. In *Proceedings of the 6th ACM conference on Embedded network sensor systems* (pp. 155-168). ACM.

Contact information

VIS, University of Stuttgart, Pfaffenwaldring 5a, 70569, Stuttgart Germany

{markus.funk, niels.henze, albrecht.schmidt}@vis.uni-stuttgart.de

{boldtrn, eiselems, yalcinta}@studi.informatik.uni-stuttgart.de