An Assistive Vision System for the Blind that Helps Find Lost Things

Boris Schauerte[†⋆], Manel Martinez[†⋆], Angela Constantinescu[‡⋆], and
Rainer Stiefelhagen[†‡]

Karlsruhe Institute of Technology
[†]Institute for Anthropomatics, Adenauerring 2
[‡]Study Center for the Visually Impaired Students, Engesserstr. 4
{forename.surname}@kit.edu
76131 Karlsruhe, Germany

Abstract. We present a computer vision system that helps blind people find lost objects. To this end, we combine color- and SIFT-based object detection with sonification to guide the hand of the user towards potential target object locations. This way, we are able to guide the user's attention and effectively reduce the space in the environment that needs to be explored. We verified the suitability of the proposed system in a user study.

Keywords: Lost & Found, Computer Vision, Sonification, Object Detection & Recognition, Visually Impaired, Blind.

1 Introduction

According to recent estimates of the World Health Organization, 285 million visually impaired people live in the world of which 39 million are blind [13]. Although 80% of all visual impairment could be avoided or cured, the unfortunate fact that the majority of blind people lives in developing countries in combination with the aging global elderly population leads to a huge innovation pressure for affordable and intuitive tools that aid visually impaired people. With the decreasing costs of digital camera technologies and mobile computing power, which is closely related to the wide distribution of mobile phones[1], computer vision is an increasingly cost-effective technology that allows visually impaired people to perceive (more) visual information in their environment. Furthermore, computer and robot vision algorithms are getting more robust and thus applicable in real-world applications (see, e.g., Google Goggles [7]). Research in the area indicates that computer vision is, for example, able to help blind people navigate in urban and indoor environments [11, 3] or assist in shopping scenarios [12].

In this paper, we introduce a novel vision system that can help blind and visually impaired people find objects that were misplaced or have unexpectedly

[⋆] Equal contribution.

[1] According to strategy analysts, more than 1 billion camera phones were sold in 2011 (the first time that annual volumes have exceeded 1 billion units).

changed their location (e.g., they may have fallen to the ground or been relocated by another person). To this end, the user has to hold a small camera, which can also be attached to his wrist if he prefers to have both hands free in order to allow for unhindered grasping and haptic perception. The corresponding hand is then guided towards potential target object locations using computer vision for object detection and sonification for acoustic feedback. This way, we are able to guide the user effectively towards plausible object locations and reduce the search space. At each object location, the user can then use his accustomed senses to conclusively identify the object. Using this methodology, following our idea that we want to enhance the capabilities of the user and not replace or interfere with his intact senses, we aid the user in detecting the searched object without interfering with his sense of orientation and leave the final search strategy and decisions to the user.

2 Related Work

Most closely related to our work are the systems by Hub et al. [8], Caperna et al. [3], and Bigham et al. [2]. In 2004, Hub et al. [8] presented a system that assists blind users in orienting themselves in indoor environments. However, their system requires a world model of landmarks and objects in the target environment, because it seemed “impossible to realize object identification of arbitrary objects using systems that are only based on [...] image interpretation” [8]. Furthermore, Hub et al. do not use sonification, but rely on text-to-speech communication. Caperna et al. [3] combined a global positioning system, inertial navigation unit, computer vision algorithms, and audio and haptic interfaces. In their system, computer vision makes it possible to identify and locate objects such as signs and landmarks. To this end, they rely on the Scale-Invariant Feature Transform (SIFT) by D. Lowe (see [10]). However, the corresponding evaluation has been performed in a simplified scenario and computer vision was left as major aspect for future work. Bigham et al. [2] use Speeded Up Robust Features (SURF; see [10]) for object identification, but instead of training an object database (see, e.g., [3]), they send images with user requests (e.g., where is the object in the image) to Amazon’s Mechanical Turk [1] where humans can outline the objects. The outlines of the object can then be used to estimate the object’s location in the environment and guide the user towards the object by informing the user how close he is to the target [2].

3 Main System Components

3.1 Visual Object Detection

Specific Objects: We use SIFT (see [10]) to detect known objects. To this end, our system provides a simple training interface, which makes it possible to train new objects by holding them in front of the camera and triggering snapshots. Trained objects can then be searched for in the environment using common SIFT feature matching and classification methods (see, e.g., [4]).

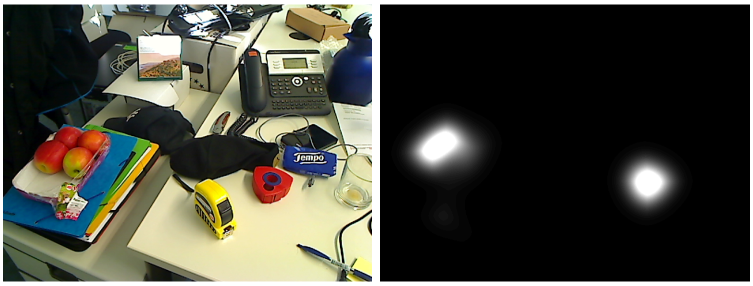

Fig. 1. Illustration of the object detection using color attributes. Image of a typical desktop environment (left) and the corresponding normalized target probability map for color "red" (right). The probability map is calculated at a lower scale than the original image to save computational resources. Here, two potential target objects are clearly identified.

Color Attributes: When using local features such as SIFT and SURF, it is only possible to detect specific, known objects (i.e., existent in the database) with a distinctive texture. As a complementary approach, we propose to use visual attributes to help find things in a broader range of scenarios; e.g., to help find a specific colored shirt in a pile of shirts or to find objects that have only been verbally described by other persons. To this end, in our prototype implementation, we use probabilistic models of the 11 basic English color terms [9], see Fig. 1, which can also be used to name the color of an object in front of the camera.

3.2 Sonification

Two sound properties – pan and pitch – are used to map the information about the object's location that is received from the vision module as follows (also see [5]): The location on the image's x-axis is mapped to pan, such that the perceived sound source location (left-front-right) corresponds with the object location relative to the image center. The location on the y-axis maps to pitch (see [6]). Here, objects located closer to the bottom of the image frame correspond to lower sounds, and objects located closer to the top of the frame correspond to higher sounds. In order to allow the user to rate how confident the system is about the detection, we map tempo to detection confidence with a more continuous sound (i.e., shorter time between "beeps") for higher detection confidence.

4 Evaluation

4.1 Procedure

To examine the suitability of the presented system, we first performed a pilot study to assess the complexity of two application scenarios with two blind users (one of which is blind from birth) and subsequently we performed our main study

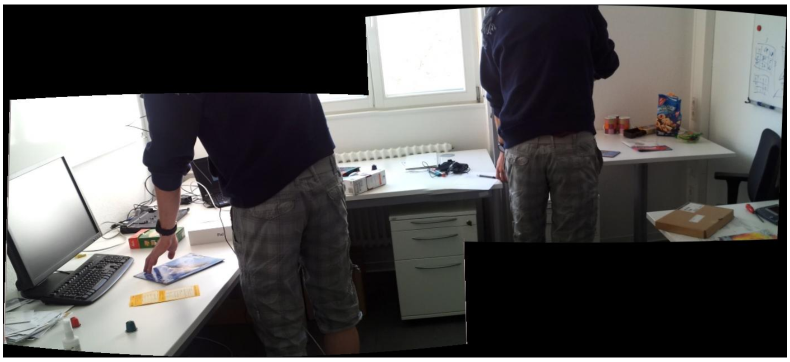

Fig. 2. Example of an evaluation trial (the participant is shown at two locations in the room), i.e. a person searching for an object inside a room. The image shows the office room and an exemplary distribution of the items. Furthermore, the image illustrates several challenges our system had to cope with such as, for example, varying lighting conditions.

with 12 users (1 blind person). In both studies, the task was to find items in an office environment, see Fig. 2. In each trial, they had to find one specific item that was placed at a random location in the environment. In our evaluation we distinguish between two scenarios: In the first scenario, the object was placed randomly inside the room, thus the user had no information about the expected location. In the second scenario, the item was placed at a random location on the desks in the room, among other distractor items, and the user was told that the object is on a desk. In this scenario, the information about the object being placed on one of the desks limits the search space substantially and allows for efficient manual, unassisted exploration of the search area. In order to accustom the users with the system, we used a single initial trial for instructions and explanations. During the tests, the users wore open headphones that leave the hearing sense mostly unaffected. As camera, we used an off-the-shelf webcam without any calibration, control of imaging features, or user intervention.

For evaluation, we recorded the time durations that were required to find the target object and performed a pre- and post-test questionnaire. The results of the second scenario's post-questionnaire (12 participants; main evaluation) is shown in Tab. 1.

4.2 Discussion

In the first scenario, if the search space is unrestricted, the system allows to rapidly find the target objects. This is especially interesting, because the users reported that they usually – i.e., without the help of our system – would have given up the search. However, in the second scenario in which the search space is restricted the search times were not always better when using the system. Nevertheless, the system was reported to be intuitive and easy to use, even

	Question	↓↑	Median	Mean	Var
1.	Which approach did you find better: searching with the system, or without it? (1: much better with, 5: much better without)	↓	2.0	2.67	1.00
2.	How easy to use did you think the system was? (1: very difficult, 5: very easy)	↑	4.0	3.75	0.21
3.	How intuitive did you find the system? (1: very intuitive, 5: very unintuitive)	↓	2.0	2.44	0.53
5.	Which approach did you find better: searching with the color search, or without it? (1: much better without, 5: much better with)	↑	4.0	4.08	1.36
6.	Which one did you think was faster, color search or searching without the system? (1: much faster with, 5: much faster without)	↓	2.5	2.58	1.54
8.	Which approach did you find better: searching with the object search, or without it? (1: much better with, 5: much better without)	↓	3.0	3.33	1.15
9.	Which one did you think was faster, object search or searching without the system? (1: much faster with, 5: much faster without)	↓	3.5	3.50	1.36
10.	Please rate the sonification (sound output), in terms of how intuitive you think it was (1: very unintuitive, 5: very intuitive)	↑	3.5	3.50	1.36
11.	Please rate how easy it was for you to interpret the sound (1: very easy, 5: very difficult)	↓	2.0	2.33	0.97
4.	Did you find the color search useful?	10× “yes”, 2× “no”			
7.	Did you find the object search useful?	conditionally*			

Table 1. Results of our post-questionnaire. Except for question 4 and 7, which allowed free answers and comments, we used an ordinal scale of $\{1, ..., 5\}$ to let the users rate specific aspects of our system. To improve the readability ↑ indicates that a higher value is better and ↓ indicates that a lower value is better. Since we have an even number of participants, there is no single middle value and we report the mean of the two middle values as median. (*) The users answered 3× “yes”, 3× that it was less useful than color search, 1× that it would help more if the latency would be lower, 1× “not really”, and 3× “no”.

though we only allowed a single training trial. Interestingly, the user reports indicated a different user experience depending on the usage of either the color attributes or the SIFT features. Due to the ambiguous results in the second scenario, we decided to further investigate it in our main evaluation.

The results of our main evaluation are shown in Tab. 1. As in our pilot study, the majority of the users reported the system as being very intuitive and easy to use (see the results for question 2 (Q2) and Q3, i.e. "How easy to use did you think the system was?" and "How intuitive did you find the system?", respectively). This is despite the fact that we only allowed the users a single trial for training and performed the post-questionnaire after three evaluation trials. Here, the chosen sonification mechanism plays a very important role and is crucial to achieve a good user experience, because it is the user's only source of information that is provided from the system, see Q10 and Q11 ("how intuitive you think [the sonification] was" and "how easy it was for you to interpret the sound", respectively). One third of the users reported that – in this scenario – they would prefer to search without the system (Q1). However, as can clearly be seen in the answers to Q5 ("with the color search, or without it?") and Q8 ("with the object search, or without it?") as well as in the answers to Q4 ("Did you find the color search useful?" – 10 out of 12 users did) and Q7 ("Did you find the object search useful?"), this depends on the features and the users prefer the color search over the search using SIFT features. As has been noted by one user in response to Q7, this is most likely caused by the fact that the SIFT approach takes more time for computation[2]. This leads to higher latencies and a decreased responsiveness, which in the end is best described as a slightly "sluggish" or "laggy" feeling when handling the system. This demonstrates that the computational complexity of algorithms and the resulting responsiveness have to be taken into account when designing and implementing such a system in order to allow for a good user experience. Using color as feature, 6 out of 12 people achieved on average better search times when using the system, which is slightly in contrast to the users' perception that they achieved better results using the system, see Q6 ("Which one did you think was faster, color search or searching without the system?"). This was likely caused by the following aspects: First, limiting the search space to the space directly above the desk surfaces made it possible for the users to rapidly detect most objects on the tables. For example, users do not have to fully orient themselves in the room and have to keep in mind all locations they already inspected and furthermore they do not need to first detect possible locations on which an object could be stored such as, for example, cupboards. Second, the users were still learning to handle the system (we observed that some users were still experimenting with features or, for example, kept misinterpreting aspects of the sonifications). Third, although seldom, false object detections did occasionally confuse the users.

[2] The SIFT feature calculation and matching is computationally more expensive than calculating the color probability maps.

5 Conclusion

We presented our current implementation of a computer vision system that is able to help visually impaired people find misplaced items. We experimentally demonstrated that the system makes it easier for visually impaired users to find misplaced items, especially if the target object is located at an unexpected location. As future work, we intend to integrate further visual attributes and, most importantly, to improve the overall system in order to reduce the average time that is required to find objects.

Acknowledgments. This work is supported by the German Research Foundation (DFG) within the Collaborative Research Program SFB 588 “Humanoide Roboter” and the VIPSAFE project by the German Federal Ministry of Education and Research (BMBF).

References

1. Amazon: Mechanical turk. `https://www.mturk.com/`
2. Bigham, J., Jayant, C., Miller, A., White, B., Yeh, T.: VizWiz::LocateIt - enabling blind people to locate objects in their environment. In: Proc. CVPR Workshop: Computer Vision Applications for the Visually Impaired (2010)
3. Caperna, S., Cheng, C., et al.: A navigation and object location device for the blind. Tech. rep., University of Maryland, College Park (2009)
4. Collet, A., Martinez, M., Srinivasa, S.S.: The MOPED framework: Object recognition and pose estimation for manipulation. Int. J. Robotics Research 30(10), 1284–1306 (2011)
5. Constantinescu, A., Schultz, T.: Redundancy versus complexity in auditory displays for object localization - a pilot study. In: Proc. Int. Conf. Auditory Display (2011)
6. Durette, B., Louveton, N., Alleysson, D., Hrault, J.: Visuo-auditory sensory substitution for mobility assistance: Testing thevibe. In: Workshop on Computer Vision Applications for the Visually Impaired (2008)
7. Google Mobile: Google Goggles. `http://www.google.com/mobile/goggles/`
8. Hub, A., Diepstraten, J., Ertl, T.: Design and development of an indoor navigation and object identification system for the blind. In: Proc. Int. ACM SIGACCESS Conf. Computers and Accessibility (2004)
9. Schauerte, B., Fink, G.A.: Web-based learning of naturalized color models for human-machine interaction. In: Proc. Int. Conf. Digital Image Computing: Techniques and Applications (2010)
10. Tuytelaars, T., Mikolajczyk, K.: Local invariant feature detectors: a survey. Found. Trends. Comput. Graph. Vis. 3, 177–280 (2008)
11. Wenqin, S., Wei, J., Jian, C.: A machine vision based navigation system for the blind. In: Proc. Int. Conf. Computer Science and Automation Engineering (2011)
12. Winlock, T., Christiansen, E., Belongie, S.: Toward real-time grocery detection for the visually impaired. In: Proc. CVPR Workshop: Computer Vision Applications for the Visually Impaired (2010)
13. World Health Organization: Visual impairment and blindness. `http://www.who.int/mediacentre/factsheets/fs282/en/`