Smartphone-Based Obstacle Detection for the Visually Impaired

Alessandro Caldini, Marco Fanfani(✉), and Carlo Colombo

Computational Vision Group, University of Florence, Florence, Italy
alecaldini@gmail.com, {marco.fanfani,carlo.colombo}@unifi.it

Abstract. One of the main problems that visually impaired people have to deal with is moving autonomously in an unknown environment. Currently, the most used autonomous walking aid is still the white can. Though in the last few years more technological devices have been introduced, referred to as electronic travel aids (ETAs). In this paper, we present a novel ETA based on computer vision. Exploiting the hardware and software facilities of a standard smartphone, our system is able to extract a 3D representation of the scene and detect possible obstacles. To achieve such a result, images are captured by the smartphone camera and processed with a modified Structure from Motion algorithm that takes as input also information from the built-in gyroscope. Then the system estimates the ground-plane and labels as obstacles all the structures above it. Results on indoor and outdoor test sequences show the effectiveness of the proposed method.

Keywords: Elettronic Travel Aid (ETA) · Visually impaired people · Smartphone-based vision · Gyroscope · Depth · Obstacle · Structure from motion

1 Introduction

One of the main problems that a person with visual disabilities has to deal with is the difficulty of moving in an unknown environment. More precisely, the most urgent problem is related to avoiding obstacles along the path. Currently, visually impaired people still rely almost exclusively on the white can to help themselves to detecting obstacles and finding a safe path.

Recently, more technological solutions have been developed to support the autonomous mobility of visually impaired people [2]. Yet currently, proposed methods tend to not completely fulfill all the user requirements, so that visually impaired people are usually skeptical about them, and not keen to replace traditional solutions. ETAs, to be fully accepted by the users, should be reliable, affordable, light and their usage should not be an evident mark of disability. Moreover, they should be designed so that hands and ears remain free, thus allowing users to manipulate objects and acoustically perceive their surroundings.

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part I, LNCS 9279, pp. 480–488, 2015.
DOI: 10.1007/978-3-319-23231-7_43

In the last few years an ever increasing diffusion of mobile devices, such as smartphones and tablets, has been observed. These devices are characterized by relatively high computational resources and limited dimensions. Equipped with visual and inertial sensors, they offer an optimal platform for the development of computer vision mobile applications [15]. In particular such devices can support the development and use of effective yet inconspicuous vision-based ETAs for the visually impaired [5].

In this paper, a novel vision-based ETA is proposed. Exploiting both the inertial gyroscope and the camera, nowadays available in any consumer smartphone, our system is able to compute the depth of the scene in front of the user and detect the presence of near obstacles. Next Section (Sect. 2) briefly describes related work on ETA. Then in Sect. 3 the proposed method is described. Section 4 discusses experimental results obtained on indoor and outdoor sequences, and Sect. 5 concludes the paper.

2 Related Work

To solve the problem of autonomous mobility for the visually impaired, solutions based on the Global Positioning System (GPS) cannot be considered due to their lack of accuracy and the impossibility to work in indoor scenarios. Hesch and Roumeliotis [8] propose a system that includes a pedometer and a laser scanner mounted on the white cane: While the authors show its validity, the additional hardware and the need of a precomputed map of the environment decrease the usability of this approach. In [1,4] the authors propose systems that visually detect known makers placed into the scene so to guarantee accurate localization of the user: Though effective, these approaches are limited to work only in previously structured environments. Zhang et al. [18] developed a smartphone-based system to visually localize the user. A drawback of this method is that images of the environment have to be previously collected and mapped, and heavy 3D computations must be run on a remote server. Other ETAs based on vision can provide localization information though object detection and optical character recognition (OCR) softwares, such as [17], or by exploiting visual and depth information to train a conditional random field (CRF) framework [14]. In [5] a smartphone-based application to detect and recognize bus line numbers have been developed to help visually impaired people to use public transport services.

Several systems were proposed to provide users with navigational aids and to detect obstacles exploiting a stereo camera pair. Leung and Medioni [9] propose an odometry system using stereo images and an inertial measurement unit (IMU) to reduce drift errors. Papers [11–13] describe methods focused on obstacle detection and safe path estimation, that employ a stereo pair to improve estimation accuracy. However, stereo cameras are currently quite expensive, bulky and showy, as compared to a standard mobile device.

In [10,16] ETAs based on single camera systems are introduced. In particular, Tapu et al. [16] develop a smartphone-based method for obstacle detection by computing homography relations and exploiting HOG descriptors for obstacle classification.

3 Method Description

Our method is designed to work with generic obstacles in both indoor and outdoor environments. Given an image sequence $\mathcal{I}$ captured by a calibrated smartphone camera attached to the user chest, the presented method implements a modified Structure from Motion (SfM) algorithm by taking advantage of the gyroscope installed on the mobile device. For each acquired image the system read the angular velocity $(\dot{\theta}_x, \dot{\theta}_y, \dot{\theta}_z)$ registered by the gyroscope; then by temporal integration, the rotation angles $(\theta_x, \theta_y, \theta_z)$ can be retrieved. With this information we can compute an estimation for the incremental rotation R_{ij} between two subsequent images $I_i, I_j \in \mathcal{I}$. Once the measurements have been acquired, the gyroscope status is re-initialized to limit the drift error.

3.1 Scene Reconstruction

Our modified SfM algorithm takes as input both a pair of images I_i, I_j and their relative rotation matrix R_{ij}.

At first point correspondences between I_i, I_j are computed using the FAST corner detector and the ORB feature descriptor. Once obtained the matching set, we exploit the relation on the essential matrix E [7], i.e.

$$\mathbf{x}_j^\top K^{-\top} E_{ij} K^{-1} \mathbf{x}_i = 0 \tag{1}$$

where $\{\mathbf{x}_i, \mathbf{x}_j\}$ is a match between I_i and I_j, and K is the calibration matrix. Then, since the essential matrix can be decomposed as $E = [\mathbf{t}]_\times R$, by substitution in Eq. 1 we obtain

$$\mathbf{x}_j^\top K^{-\top} [\mathbf{t}_{ij}]_\times R_{ij} K^{-1} \mathbf{x}_i = 0 \tag{2}$$

where R_{ij} is the rotation matrix and $[\mathbf{t}_{ij}]_\times$ the skew-symmetric matrix of the translation vector $\mathbf{t}_{ij}$. $\mathcal{T}_{ij} = [R_{ij}|\mathbf{t}_{ij}]$ describes the relative transformation between I_i and I_j.

Since R_{ij} is supposed to be known from gyroscope readings, we can define $\hat{\mathbf{x}}_j = K^{-1}\mathbf{x}_j$ and $\tilde{\mathbf{x}}_i = R_{ij}K^{-1}\mathbf{x}_i$ and Eq. 2 becomes

$$\hat{\mathbf{x}}_j^\top [\mathbf{t}_{ij}]_\times \tilde{\mathbf{x}}_i = 0 \tag{3}$$

Now Eq. 3 can be rewritten as a linear homogeneous equation on the elements of $\mathbf{t}_{ij}$. With at least three correspondences—or just two if the translation scale factor is fixed a priori— we can solve a linear system to estimate $\mathbf{t}_{ij}$. However, wrong matches are always present, and to avoid the introduction of outliers, we wrap the estimation process into a RANSAC framework [3]. Once the maximum consensus set is found, $\mathbf{t}_{ij}$ is refined minimizing the error on all the inlier correspondences.

Similarly to what happens with the decomposition of the essential matrix, the solution of Eq. 3 has a two-fold ambiguity on the sign of the translation. To select the correct vector, we triangulate [6] all inlier correspondences with

both candidate solutions and we retain the 3D map satisfying the positive depth constraint on most points (i.e. the 3D points must lie in front of both cameras). The computed 3D of the scene is then exploited for the detection of obstacles.

3.2 Obstacle Detection

To detect obstacles and evaluate their proximity to the user, our algorithm estimates first the scene ground-plane, and simultaneously identifies 3D points that lie on it.

The system selects a 3D point set $S_\pi = \{\mathbf{X}_p\}$ related to matched correspondences detected in the bottom part of the images (under the hypothesis that such points belong to the ground-plane). Then a robust plane estimation algorithm is executed over S_π by randomly choosing three points for each RANSAC iteration k.

In each iteration, a plane equation $\pi_k(\mathbf{n}_k, d_k)$ is evaluated, where $\mathbf{n}_k$ is the plane π_k normal vector and d_k its distance w.r.t. the origin of the coordinate frame. A 3D point $\mathbf{X}_p \in S_\pi$ is considered as inlier if $\mathbf{n}_k^\top \mathbf{X}_p + d_k < \epsilon_1$. In our experiments we set ϵ_1 to a low value (e.g. $\epsilon_1 = 0.1$) in order to perform a strict selection of inliers. Note that to evaluate inliers we use an algebraic criterion instead of a geometric point/plane distance: while the latter approach is typically more correct, in this case, since we don't have the metric scale factor of the scene, the definition of the right threshold for the geometric distance can be misleading.

Once obtained the maximum inlier set $\widetilde{S}_\pi$, the 2D correspondence set associated to $\widetilde{S}_\pi$ is used to estimate the homography transformation H_{ij} between the planar regions of I_i and I_j. Then all matches that don't already have the associated 3D point in $\widetilde{S}_\pi$ are tested: If the distance $\mathcal{D}(\mathbf{x}_i, \mathbf{x}_j)$ defined as

$$\mathcal{D}(\mathbf{x}_i, \mathbf{x}_j) = ||H_{ij}\mathbf{x}_i - \mathbf{x}_j|| \tag{4}$$

is less than ϵ_2, then the 3D point relative to the correspondence $\{\mathbf{x}_i, \mathbf{x}_j\}$ is considered as a point on the ground-plane.

Finally, all 3D points that don't belong to the ground plane are labeled as obstacles and their relative depth can be exploited to assign different *warning* level—higher for closer objects, lower for distant ones.

In order to better evaluate obstacle distances and proximity of collision, a bird's-eye view of the scene is produced. To this aim, all 3D points are at first registered with a coordinate frame with the X and Z axes aligned with the ground-plane; then all obstacle 3D points are projected onto the ground-plane.

4 Evaluation

All images that have been used in the tests have a resolution of 320×240 pixels. Processing is carried out on an Android LG Nexus 5 smartphone, equipped with a Qualcomm Snapdragon 800 quad-core processor at 2.3GHz. With this setup,

the method works at about 2 seconds per frame, which is suitable for use at standard walking speed.

Fig. 1 reports the results of an indoor test. The test sequence was recorded by moving the smartphone over a table where objects simulated a cluttered environment. For each original frame the system produced a sparse depth map and a bird's-eye view of the scene showing the ground-plane and the detected obstacles.

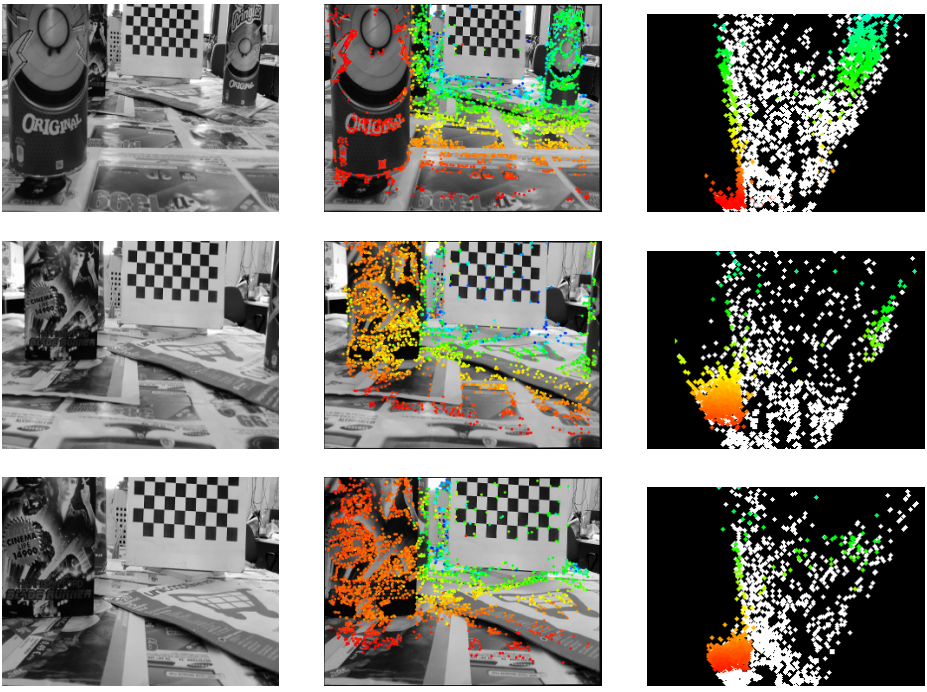

Fig. 1. Example frames of the indoor test sequence. In the first column the original images, in the second column the spare depth map computed, and finally the bird's-eye view with in white the ground-plane and obstacle colored from red to blue to represent their proximinty to the user. (Best viewed in color)

In Fig. 2 and Fig. 3 results on two different outdoor tests (respectively named *pilon* and *parking*) are reported. The tests were carried out with a walking person equipped with the smartphone held in front of his chest. Also in this case the algorithm computes correct depth values and produces a coherent bird's-eye view where ground-plane and obstacles are clearly visible.

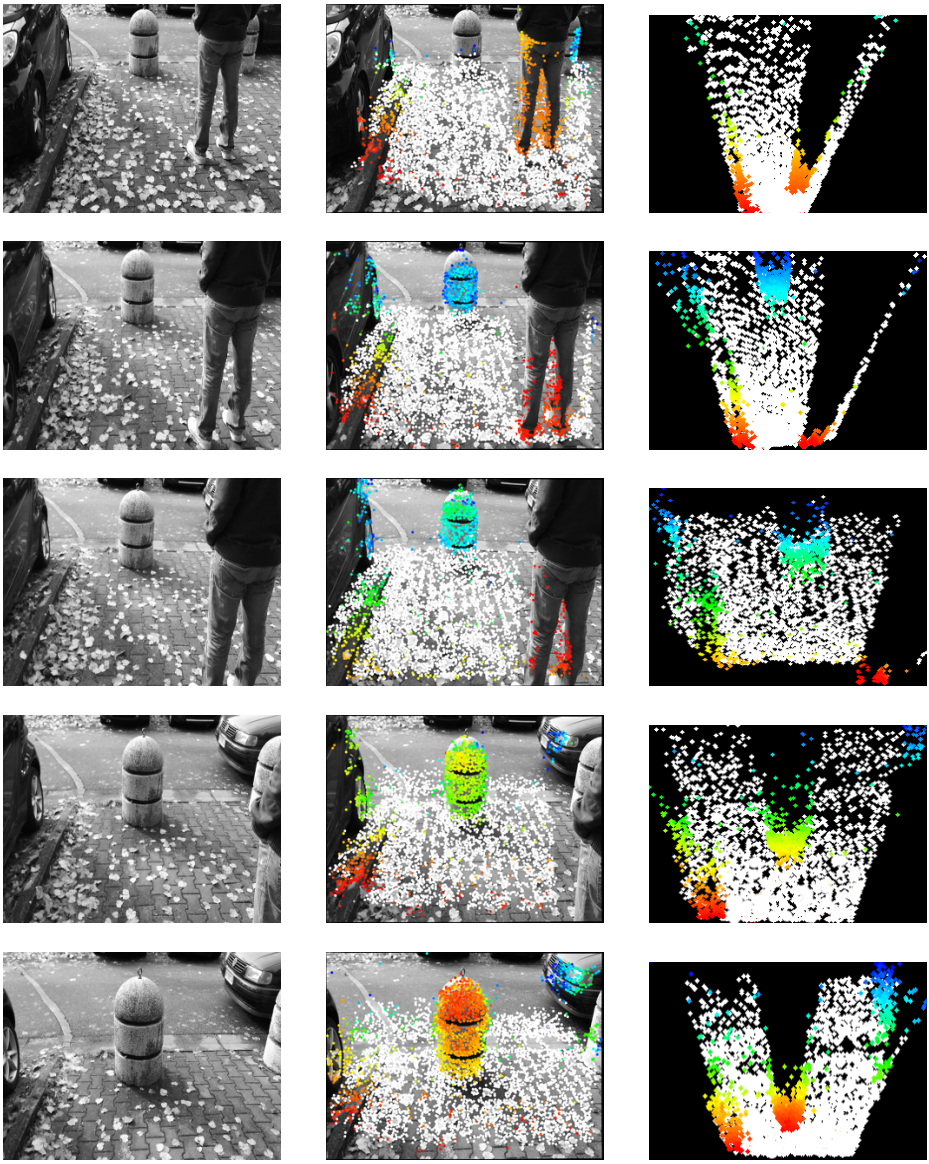

Fig. 2. Example frames of the *pilon* sequence. Again we present the original images (first column), the sparse depth map with ground-plane points in white (second column), and the bird's-eye view (third column). (Best viewed in color)

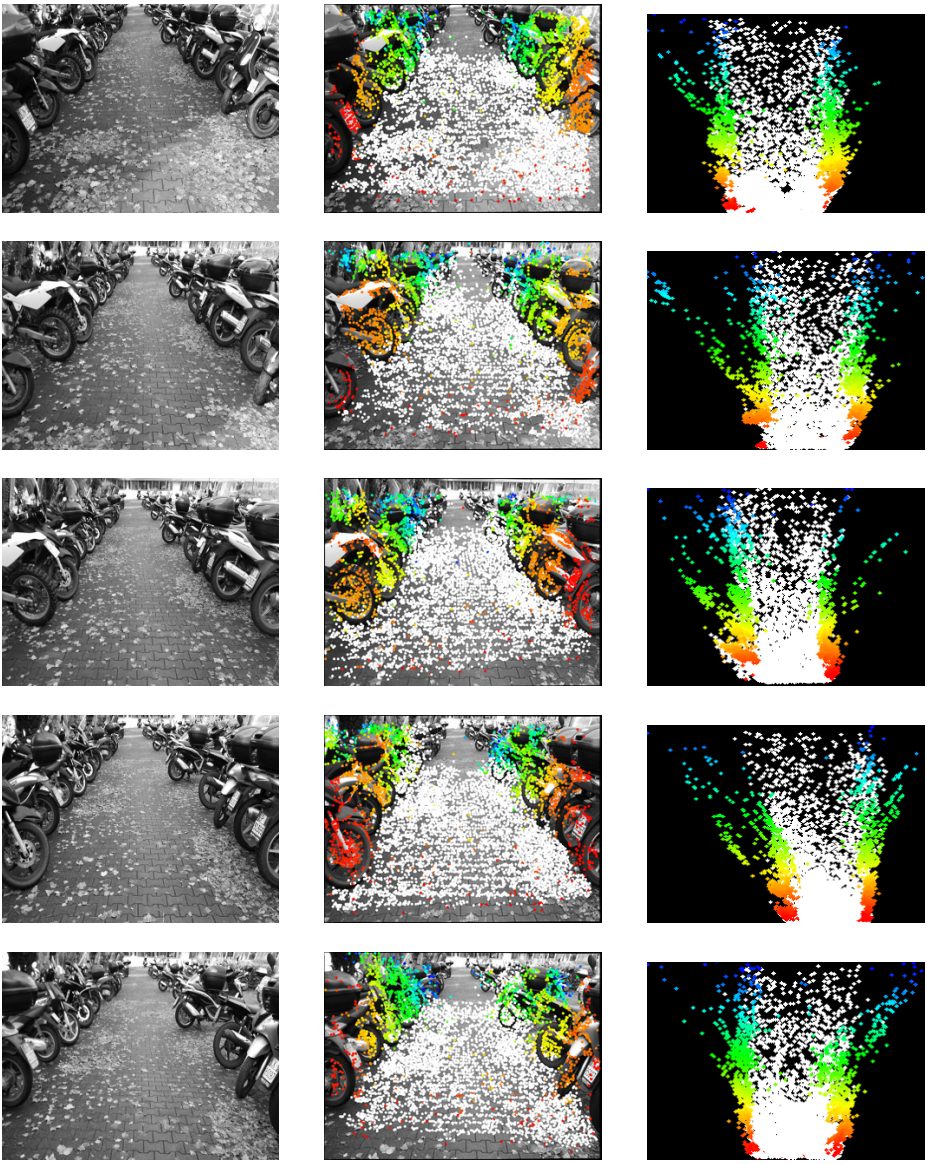

Fig. 3. Original images, depth map and bird's-eye view representation for some frame of the *parking* sequence. (Best viewed in color)

5 Conclusions and Future Work

In this paper we have presented a smartphone-based obstacle detection vision system to help visually impaired people to move autonomously in unknown

indoor and outdoor environments. The developed algorithm exploits both visual information from the camera and inertial measurements registered from the gyroscope. A sparse depth map is computed with a modified Structure from Motion approach, and obstacles are detected as they pop out the ground-plane. Results show the good performance of our method.

Future work will address the development of a tactile/acoustic interface to provide feedback to visually impaired people and alert them regarding obstacles on their path. An extensive evaluation/refinement process carried out with the help of blind users is planned, aimed at improving system performance and usefulness.

References

1. Coughlan, J., Manduchi, R., Shen, H.: Cell phone-based wayfinding for the visually impaired. In: 1st International Workshop on Mobile Vision (2006)
2. Dakopoulos, D., Bourbakis, N.: Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **40**(1), 25–35 (2010)
3. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
4. Gallo, P., Tinnirello, I., Giarr, L., Garlisi, D., Croce, D., Fagiolini, A.: ARIANNA: pAth recognition for indoor assisted navigation with augmented perception. CoRR (2013)
5. Guida, C., Comanducci, D., Colombo, C.: Automatic bus line number localization and recognition on mobile phones—a computer vision aid for the visually impaired. In: Maino, G., Foresti, G.L. (eds.) ICIAP 2011, Part II. LNCS, vol. 6979, pp. 323–332. Springer, Heidelberg (2011)
6. Hartley, R., Sturm, P.: Triangulation. Computer Vision and Image Understanding **68**(2), 146–157 (1997)
7. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004)
8. Hesch, J.A., Roumeliotis, S.I.: Design and Analysis of a Portable Indoor Localization Aid for the Visually Impaired. Int. J. Rob. Res. **29**(11), 1400–1415 (2010)
9. Leung, T.S., Medioni, G.: Visual navigation aid for the blind in dynamic environments. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2014, pp. 579–586 (2014)
10. Lin, Q., Han, Y.: Safe path estimation for visual-impaired people using polar edge-blob histogram. In: Proc. of The World Congress on Engineering and Computer Science (2013)
11. Pradeep, V., Medioni, G., Weiland, J.: Robot vision for the visually impaired. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 15–22 (2010)
12. Rodríguez, A., Yebes, J.J., Alcantarilla, P.F., Bergasa, L.M., Almazn, J., Cela, A.: Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback. Sensors **12**(12), 17476–17496 (2012)
13. Saez Martinez, J.M., Escolano Ruiz, F.: Stereo-based aerial obstacle detection for the visually impaired. In: Workshop on Computer Vision Applications for the Visually Impaired (2008)

14. Schauerte, B., Koester, D., Martinez, M., Stiefelhagen, R.: Way to go! Detecting open areas ahead of a walking person. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014 Workshops. LNCS, vol. 8927, pp. 349–360. Springer, Heidelberg (2015)
15. Tanskanen, P., Kolev, K., Meier, L., Camposeco, F., Saurer, O., Pollefeys, M.: Live metric 3d reconstruction on mobile phones. In: 2013 IEEE International Conference on Computer Vision (ICCV), pp. 65–72 (2013)
16. Tapu, R., Mocanu, B., Bursuc, A., Zaharia, T.: A smartphone-based obstacle detection and classification system for assisting visually impaired people. In: The IEEE International Conference on Computer Vision (ICCV) Workshops (2013)
17. Tian, Y., Yang, X., Yi, C., Arditi, A.: Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. Machine Vision and Applications **24**(3), 521–535 (2013)
18. Zhang, D., Lee, D.J., Taylor, B.: Seeing Eye Phone: a smart phone-based indoor localization and guidance system for the visually impaired. Machine Vision and Applications **25**(3), 811–822 (2014)