# Smartphone-based Obstacle Detection for the Visually Impaired

**A. Caldini, M. Fanfani, and C. Colombo, CVG - University of Florence, Italy**
**alecaldini@gmail.com, {marco.fanfani, carlo.colombo}@unifi.it**

## Overview

Avoiding obstacles in an unknown environment is one of the main problems for visually impaired people. While the most used autonomous walking aid is still the white can, electronic travel aids (ETAs) have been introduced in the last years [1]. Although most of the proposed solutions employ bulky and showy hardware including inertial sensors and laser scanners, or need previously structured environment (such as known maker or precomputed 3D maps). More recently smartphones have been used to implement light and inconspicuous ETAs [2]. Typically such devices offer relatively high computational resources together with a monocular camera and inertial sensors.

## Method description

Employing a commercial smartphone, the porposed method exploit the built-in camera and gyroscope in order to compute the 3D structure of the scene and detect possible obstacles.

**Scene reconstruction**

A modified version of *Structure from Motion* algorithm is used in order to exploit the rotation measurement obtained by the gyroscope. For each acquired image $I_j$, its relative rotation $R_{ij}$ w.r.t. the previous frame $I_j$ is known. Matches $\{\mathbf{x}_i, \mathbf{x}_j\}$ between $I_i$ and $I_j$ are obtained using the FAST detector and the ORB descriptor. Then, knowing the calibration $K$, it holds

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

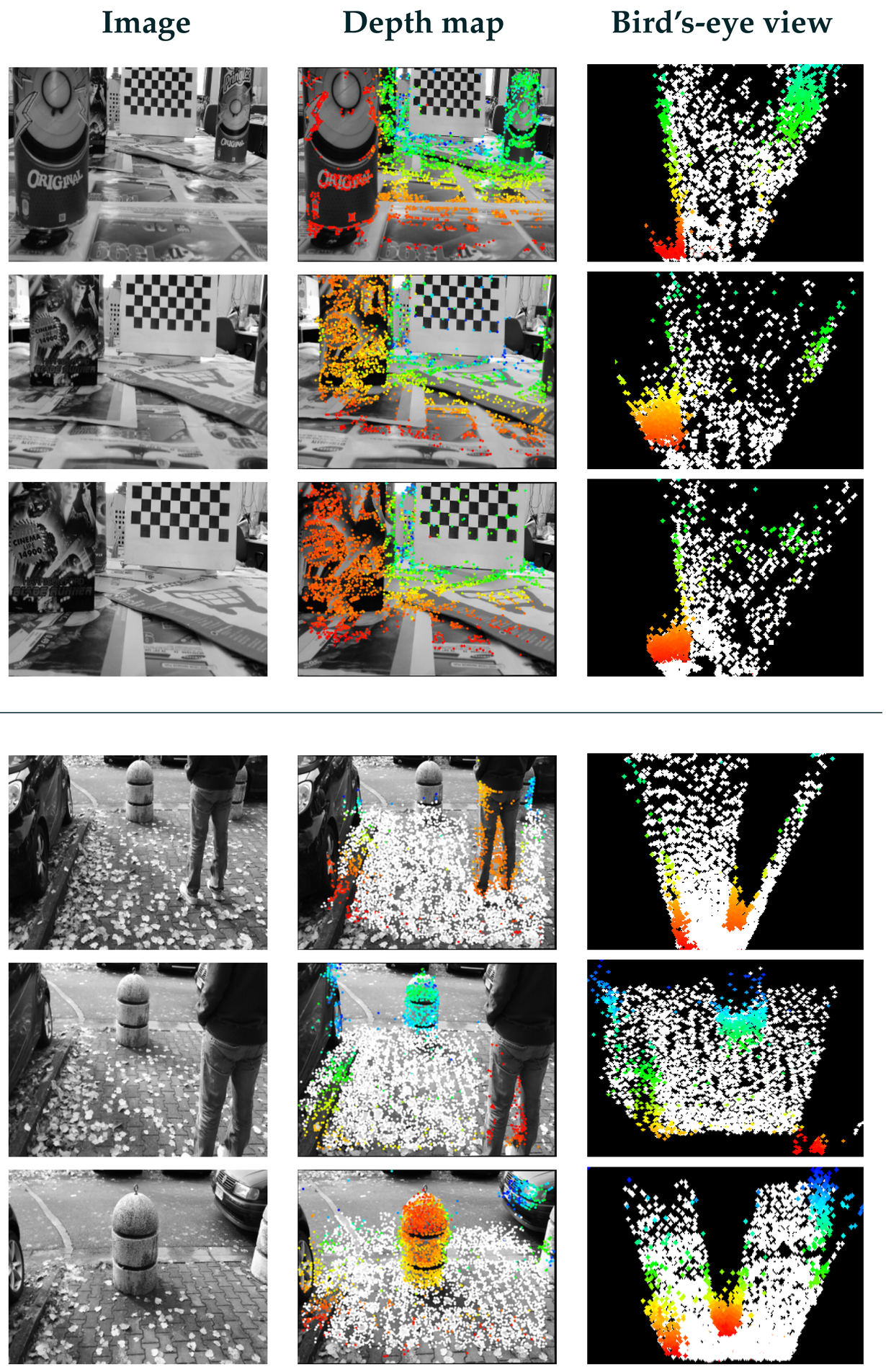
where $\hat{\mathbf{x}}_j = K^{-1}\mathbf{x}_j$ and $\tilde{\mathbf{x}}_i = R_{ij}K^{-1}\mathbf{x}_i$. With at least three correspondences the translation $\mathbf{t}_{ij}$ can be estimated.

**Obstace detection**

To detect obstacles and evaluate their proximity to the user, our algorithm estimates the scene ground-plane, and simultaneously identifies 3D points that lie on it.
Ground-plane detection is performed both in 3D and 2D space. At first a 3D plane model is estimated using RANSAC with a minimum sample of three points. A 3D $\mathbf{X}$ point is an inlier (i.e. a point on the plane) if $\mathbf{n}^\top \mathbf{X} + d < \epsilon_1$.
Image correspondences related to the 3D are used to estimate the planar homography $H_{ij}$ between $I_i$ and $I_j$. Using $H_{ij}$, all matches that have a 3D point not already labeled as plane point are evaluated and added to the plane point set if $||H_{ij}\mathbf{x}_i - \mathbf{x}_j|| < \epsilon_2$.
All 3D points that don't lie on the ground-plane are labeled as obstacle. Their depths can be exploited to assing different *warning levels*.
Finally, to better evaluate obstacle proximity bird's-eye view of the scene is produced projecting all 3D points onto the ground-plane.

## Results

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




## References

1. D. Dakopoulos, N. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews (2010).

2. R. Tapu, B. Mocanu, A. Bursuc, T. Zaharia, "A Smartphone-Based Obstacle Detection and Classification System for Assisting Visually Impaired People", In Proc. of the IEEE International Conference on Com- puter Vision (ICCV) Workshops (2013).