

# Safe Path Estimation for Visual-Impaired People Using Polar Edge-Blob Histogram

Qing Lin and Youngjoon Han

**Abstract**—This paper presents a safe path estimation method for visual-impaired people in an outdoor sidewalk environment. Unlike many existing methods that rely on stereo-vision, the proposed method aims to detect generic obstacles in a cluttered road environment by using just single camera mounted at user's belly. One of the main difficulties of using single camera in outdoor navigation task is the discrimination of obstacles with cluttered background. To solve this problem, this paper makes use of the inhomogeneous re-sampling property of top-view transform. By mapping the original image to a top-view virtual plane using top-view transform, background edges in the near-field are sub-sampled while obstacle edges in the far-field are oversampled. Morphology filters with connected component analysis are used to enhance obstacle edges as edge-blobs with larger size, whereas sparse edges from background are filtered out. Based on the identified obstacles, safe path is estimated by tracking a polar edge-blob histogram on the top-view domain. The algorithm is tested in different sidewalk scenes with complex pavements, and its efficiency has been confirmed.

**Index Terms** — obstacle detection, monocular vision, top-view transform, polar edge-blob histogram

## I. INTRODUCTION

Authoritative statistics have shown that about 1% of the world population is visually impaired, and among them about 10% is fully blind. One of the consequences of being visually impaired is the limitations in mobility. Therefore, many electronic travel-aid systems have been developed to provide assistance to blind people in a certain local environment. Electronic travel-aid systems can be categorized depending on how to sense the environment and how to inform the blind user [1]. In general, environment can be sensed through ultrasonic sensor, laser sensor, or camera, and users can be informed via auditory or tactile sense. In recent years, camera based travel-aid systems have won much attentions due to its advantages like large sensing area, rich sensing data as well as low cost.

Most existing vision-based travel-aids systems are developed using stereo vision methods. In these systems, stereo cameras are used to create a depth map of the

surrounding environment. The distance information contained in this depth map is then quantized into certain kind of grid representation, which are converted into tactile or auditory sensing modalities so as to be perceived by the visual-impaired user. For Instance, TVS[2] and Tyflos navigator system[3][4] quantize depth map into a regular grid representation, which is converted into vibration sensing on a 2-D vibration array attached on the user's abdomen. ENVIS system [5] quantize depth map to a rectangular block representation, which maps to electrical pulses that stimulate user's fingers. In[6], depth map is quantized to a polar grid representation, which is transformed into an acoustic sound space.

Although many stereo-vision based travel-aids systems have been proved to be effective under certain environment, some problems still exist. First of all, due to the high computation cost of getting a dense depth map, most of these systems tend to directly convey the quantized depth information to the user without doing any safe path estimation process. As a result, users have to estimate a safe path themselves by sensing and judging the transformed auditory or tactile pattern from the depth map. This makes the system less easy to use and requires much user training. In addition, the accuracy of depth map is largely dependent on stereo matching, which is a challenging task in cluttered outdoor scenes.

Despite stereo-vision based system, systems using only single camera were proposed as well. Compared with stereo cameras, single camera is more compact and easier to maintain. Some of these mono-vision based systems focused on identification of object pixels among background pixels. For example, in [7], region growing segmentation is used to discriminate obstacle pixels with the background. in NAVI system proposed by Sainarayanan et al.[8], a fuzzy learning vector quantization (LVQ) neural network is trained for the classification of object pixels and background pixels. Although the obstacle detection performances for these systems in simple indoor environment is encouraging, their performances may get deteriorate in outdoor environment with various illumination changes and complex background.

In this paper, we present a monocular vision approach to do obstacle detection and safe path estimation for assisting visual-impaired users to pass through a pedestrian path. In the proposed method, a camera is attached at blind user's belly and looking downward to the road in front. The basic idea for obstacle detection is to discriminate obstacle pixels with background pixels. In contrast to Sainarayanan's method that made use of pixel-wise features, edge-based features are explored to discriminate obstacles with road

Manuscript received July 20, 2013; revised Aug. 13, 2013. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (No. 2012M3C4A7032182)).

Qing Lin is with Electronic Engineering Department, Soongsil University, Seoul, Korea (e-mail: lqsdust@163.com).

Youngjoon Han is with Electronic Engineering Department, Soongsil University, Seoul, Korea (e-mail: young@ssu.ac.kr).

pavement background. By re-sampling the original image pixels and mapping to a top-view virtual plane, clutter edges from background in the near field are suppressed, while obstacle edges in the far field are enhanced. Morphology filters are then used to enhance this inhomogeneous re-sampling effect on connectivity and scale of edges, so that obstacle edge-blobs can be identified easily by connected component analysis. To find a safe path, a polar distortion model of obstacles are built on top-view domain, based on which a polar edge-blob histogram is calculated by scanning all the polar directions to check edge-pixel accumulations that lie on each polar direction. The part of histogram where the biggest valley appears is detected and tracked to find the largest area where no obstacle edge-blob exists.

## II. OBJECT DETECTION ON TOP-VIEW DOMAIN

### A. Top-view Re-sampling and Mapping

Top-view transform is in nature an inhomogeneous re-sampling and mapping process. In this section, the re-sampling and mapping process is re-formulated in horizontal and vertical directions, and the re-sampling effect on the scale and connectivity of edges is discussed.

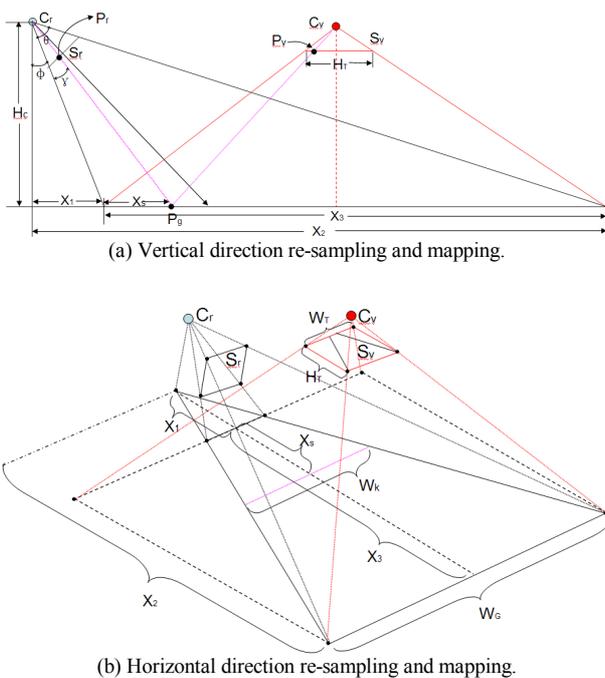


Fig. 1. Top-view re-sampling and mapping model.

The model of vertical direction re-sampling is illustrated in Fig.1 (a). In Fig.1 (a),  $C_r$  is the real camera center with  $S_r$  as its image plane, while  $C_v$  is the virtual top-view camera center with  $S_v$  as the virtual top-view plane. To figure out the re-sampling relationship between  $S_r$  and  $S_v$  plane, the only parameters that require are  $\phi$  and  $\theta$ . By the geometrical description in Fig.1 (a), for each point  $P_v$  on the virtual top-view plane  $S_v$ , its corresponding sampling point  $P_r$  on the real image plane  $S_r$  can be calculated based on their common projection point  $P_g$  on the ground plane. As in (1) shows, for each point  $i$  on top-view plane, its corresponding sampling point  $h$  on the real image plane can be obtained. The model of horizontal re-sampling is illustrated in Fig. 2 (b), for each row  $W_k$  in  $C_r$ 's field of view on the ground plane, its length

can be calculated according to the triangular similarity, and by comparing  $W_k$  with  $C_v$ 's field of view on the ground plane, the sampling ratio can be computed. As in (2) shows, the sampling ratio  $W_s$  for each row on the top-view plane can be finally calculated.

$$\begin{aligned}
 X_1 &= H_c \cdot \tan f, \quad X_2 = H_c \cdot \tan(q+f) \\
 X_3 &= X_2 - X_1, \quad X_s = i \cdot X_3 / H_T \quad (i=0..H_T) \\
 g &= \arctan[(X_1 + X_s) / H_c] - f \\
 h &= \begin{cases} H_T / 2 - f \cdot \tan(q/2 - g) & (g \leq q/2) \\ H_T / 2 + f \cdot \tan(g - q/2) & (g > q/2) \end{cases} \quad (1)
 \end{aligned}$$

$$\begin{aligned}
 W_G &= X_3 \cdot (W_T / H_T), \quad X_k = X_1 + X_s \\
 W_k &= W_G \cdot X_k / X_2, \quad W_{k-p} = W_k \cdot (W_T / W_G) \\
 W_s &= W_T / W_{k-p} \quad (2)
 \end{aligned}$$

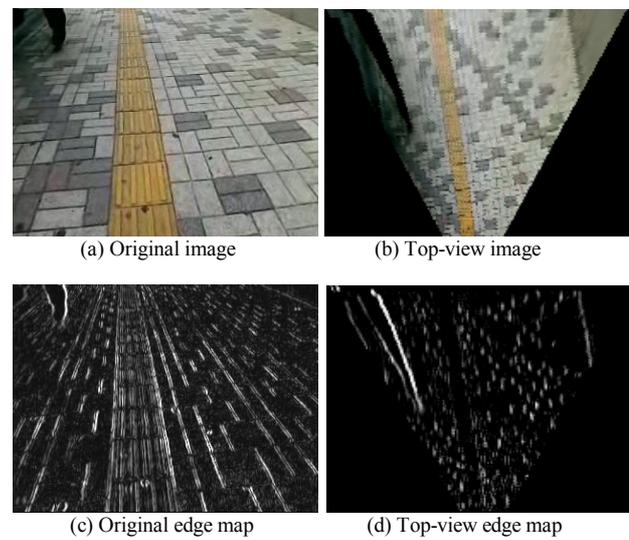


Fig. 2. Effect of top-view re-sampling and mapping.

Fig.2 shows the effect of top-view mapping by comparing original edge map with top-view edge map. On original edge map in (c), obstacle edges are mixed with clutter edges from the background, which makes it difficult to discriminate obstacle's edges with those pavement edges around. However, on top-view edge map in (d), the top-view re-sampling process enhances the scale and connectivity of obstacle edges in the far field while suppresses clutter edges in the background. Compared with the edge map of original image in (c), it is obviously much easier to discriminate obstacle edges on top-view domain in terms of their scale and connectivity.

### B. Obstacle Edge-blob Extraction

After top-view re-sampling and mapping, the obstacle edges are enhanced in terms of scale and connectivity. To further emphasize this effect, a combination of morphology operations and connected component analysis is used to extract edge-blobs with large size. These edge-blobs are regarded as candidate obstacle representations.

Here a  $3 \times 3$  rectangular structure element is used to remove pavement edge segments with an opening operation, followed by a closing operation to fill the gaps inside remaining foreground pixels. A connected component labeling operation is then applied to group the connected

foreground pixels into blobs. Blobs with size smaller than a pre-defined threshold are discarded. As shown in Fig.3(c), many small edge-blobs from pavement are eliminated by opening operation, and closing operation fixes the shape of foreground blobs. Finally, as shown in Fig.3(d), only two major edge-blobs are selected, which correspond to possible obstacle regions on the top-view.

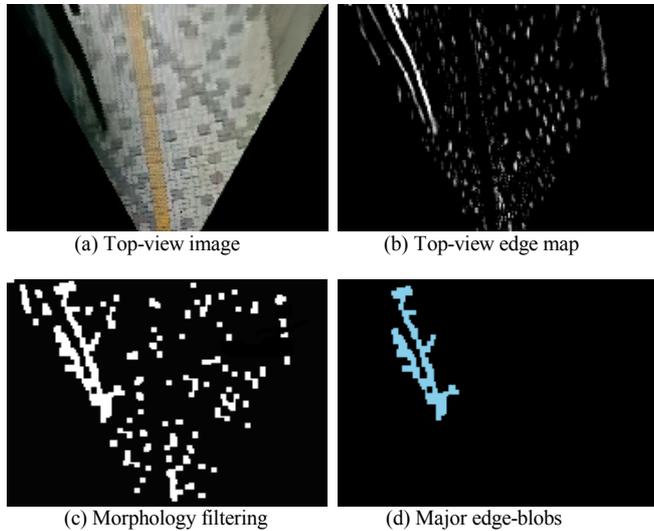


Fig. 3. Edge-blob extraction

### III. SAFE PATH ESTIMATION

#### A. Obstacle Projection Model on Top-view

Due to the top-view re-sampling effect, the shape of a generic obstacle with quasi-vertical boundaries will be distorted on the top-view domain. Here an interesting property of this distortion is that, an obstacle which rises up from the ground surface would be elongated in the direction of an imaginary connection line joining the camera's perpendicular projection on the ground and the base point of the obstacle in top-view images, as is shown in Fig. 4.

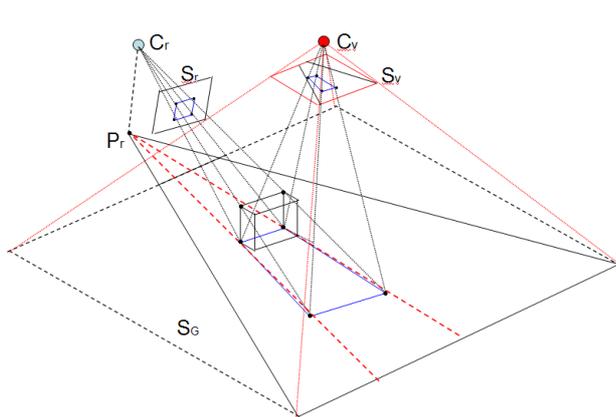


Fig. 4. Obstacle projection model on top-view domain.

This geometric property means that, obstacle edges should also lie along this connection line on top-view image. Therefore, vertical obstacle's edges should lie on series of radial orientations with respect to the camera's projection point on the top-view plane. This vertical line distortion can be partly explained by the inhomogeneous re-sampling process, while it can also be derived from formula on Inverse perspective mapping[9]. As is shown in (3), the point on the

real image plane  $S_r$  is represented in  $(u, v)$ , and point on the ground plane  $S_G$  is represented by  $(x, y, 0)$ . Vertical lines on the image plane  $S_r$  can be represented by  $v = k$ , while  $k$  is a constant value, substituting this into (3), we can get (4), where  $c_1$  and  $c_2$  are constant terms. Finally, we can obtain (5), where  $(l, d)$  represents camera center's projection point  $P_r$  on the ground plane.

$$\begin{cases} x(u, v) = h \times \cot[(q - \partial) + u \frac{2\partial}{n-1}] \times \cos[(g - \partial) + v \frac{2\partial}{n-1}] + l \\ y(u, v) = h \times \cot[(q - \partial) + u \frac{2\partial}{n-1}] \times \sin[(g - \partial) + v \frac{2\partial}{n-1}] + d \end{cases} \quad (3)$$

$$\begin{cases} x = h \times \cot[(q - \partial) + u \frac{2\partial}{n-1}] \times c_1 + l \\ y = h \times \cot[(q - \partial) + u \frac{2\partial}{n-1}] \times c_2 + d \end{cases} \quad (4)$$

$$y - d = (c_2 / c_1)(x - l) \quad (5)$$

#### B. Polar Edge-blob Histogram

Based on the obstacle projection model on top-view domain, here a polar edge histogram is constructed on the top-view plane for the estimation of safe path. As is shown in Fig.5(c), on the edge map, from the right boundary to the left boundary, polar directions (marked in red dash line) are sampled with respect to the convergence point C, which corresponds to camera's perpendicular projection point on the ground plane. For each sampled polar direction, the number of edge-blob pixels that lie along this direction is counted. By accumulating all the sampled polar directions, a polar edge-blob histogram can be constructed as shown in Fig.5(d).

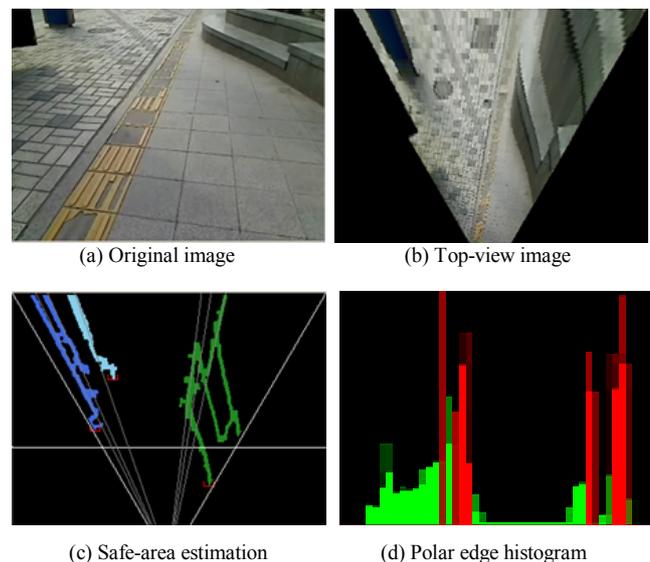


Fig. 5. Safe path estimation through polar edge-blob histogram.

In polar edge histogram, the horizontal axis represents sampled polar directions in angles, and the vertical axis is the number of edge-blob pixels that lie along each sampled direction angle. The bins with high values indicate the directions where obstacles appear, while bins with zero values correspond to the directions where no obstacles exist.

Therefore, safe-area should be estimated by the bins with zero values.

Since the camera is mounted on user's body, the camera will show some swing motions due to the movement of human body. These swing motions will bring additional noise to the safe path estimation. To estimate the safe path more steadily, the largest valley bin group on polar edge-blob histogram is tracked. For tracking initialization, consecutive zero-value bins in frame  $t$  are grouped and sorted according to their group size. Then the largest bin group is selected as the tracking group in frame  $t$ . In the following frame, the zero bin group that is closest to the tracking group in frame  $t-1$  is selected as the tracking group in frame  $t$ . If the size of tracking group is smaller than a threshold, then tracking will be stopped and re-initialized from the beginning.

#### IV. EXPERIMENTAL RESULTS

To test the performance of the algorithm, we attached a camera on a belt and fix it at user's waist, pointing a little bit downward to the road ahead of user, as is shown in Fig. 6. The camera is simply a Logitech webcam, which captures color image at  $320 \times 240$  resolution in 30 fps. The safe path algorithm is implemented using Visual C++ under MS Windows platform, which runs on a laptop computer with 1.8GHZ CPU and 2GB DDR memory. The webcam captures images of the road environment, and then processed by the path finding software which runs at the laptop computer carrying in user's backpack.

To make a top-view mapping, camera's downward viewing angle and its angular aperture should be measured. By using these two parameters and applying formula (1) and (2), a mapping table is made to store the mapping relationship from one point on the top-view domain to its corresponding location on the original-view domain. With this mapping table, top-view re-sampling and mapping can be done very efficiently.



Fig. 6. Experimental set up.

The algorithm is tested on several outdoor pedestrian path scenes, with various roadside structures and cluttered road surface. A test scene sample is shown in Fig.7, where processing results on original view and top-view are compared. In Fig.7(a), edge map on original image is examined by a vertical projection histogram. Due to the influence of clutter edges from the pavement, it is very difficult to discriminate a person located in the upper-right corner with background clutters in the image. In contrast, clutter edges are suppressed while obstacle edges are enhanced on the top-view edge map shown in Fig.7(b), and the location of the person can be clearly identified on the polar edge-blob histogram.

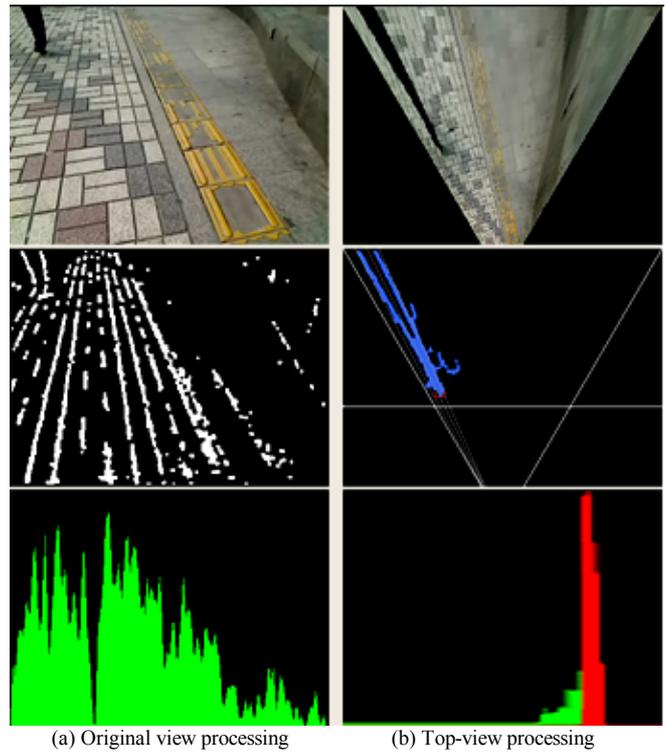


Fig. 7. Original-view and top-view processing comparison

To evaluate obstacle detection performance, the test scenes are divided into three sets, including pedestrian paths in open space scene, park scene and urban scene, some sample images from these three different test sets are shown in Fig.8.



Fig. 8. Sample images from test sets.

All the critical obstacle positions are manually labeled on the top-view images of these test sets. A true positive (TP) detection is defined to be such that the detection corresponds with an actual obstacle, and the deviation should not exceed 20% of the obstacle's size, otherwise it is considered as a false positive (FP), obstacles that have not been detected is false negative (FN). Table 1 shows the detection results on three test sets. For safe path estimation, it is very critical to control the false negative rate for sake of safe navigation. Therefore, during testing, the algorithm parameters are

tuned in such a way to achieve an acceptable true positive rate while keeping false negative rates as small as possible.

TABLE I  
OBSTACLE DETECTION RESULTS

Test sets	Obstacle	TP	FP	FN
Urban	365	314	32	12
Open	278	263	11	5
Park	212	193	13	8

Since the proposed algorithm relies on radial distribution of edges on top-view domain, when strong background edges appear in similar radial patterns with that of obstacles on top-view, they may give rise to FP cases. Moreover, small planar obstacles in the near field may be sub-sampled heavily on top-view, which makes it difficult to discriminate with ground clutters. Therefore, small holes or stones on cluttered road surface may not be properly detected, which give rise to FN cases. In the test, open space set achieves a high TP rate of 94.6%, as this set involves mainly vertical obstacles like pedestrians, and less cluttered road surface. While in urban set, only 86% TP rate is achieved, due to highly cluttered road surface as well as many planar obstacles in small size.

To show the effectiveness of the proposed algorithm on top-view domain, an obstacle detection method [10] using edge-blobs on original view is implemented and tested on the urban test set. The quantitative comparison between the two algorithms is shown in Fig. 9. The ROC curves are generated by varying the obstacle edge-blob extraction threshold in both algorithms. It can be observed that the proposed method on top-view has shown much better performance under complex background.

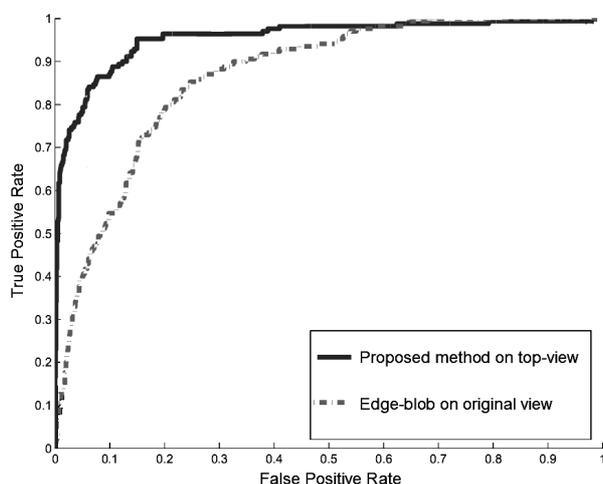


Fig. 9. ROC curve of top-view and original-view methods.

To evaluate the safe path estimation performance, simulated user walking trajectory is generated by using the estimated safe walking direction and user's walking speed. User's walking speed is recorded by using an inertial sensor attached on user's body. This simulated walking trajectory is then mapped to a top-view occupancy map generated from obstacle detection module. A segment of this synthesized map is shown in Fig. 10, which is generated by walking on a pedestrian path around our campus. Based on the simulation results, it is found that simulated walking trajectory is able to

avoid salient obstacles with vertical edges. However, as small planar obstacles in the near distance may be removed together with the clutter patterns from the pavement. It may cause problems for safe path estimation in this case. Under the experiment platform condition, the path finding algorithm can run at an average of 16.4 fps, which is fast enough to satisfy the real-time requirement of human navigation task.

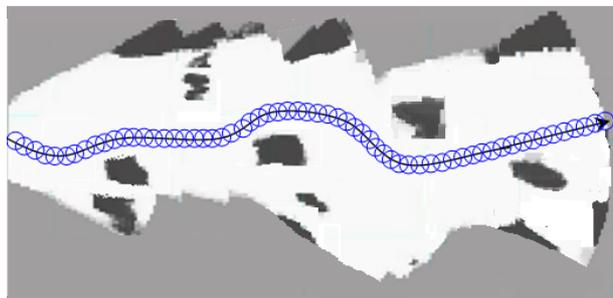


Fig. 10. Simulated walking trajectory on occupancy map.

## V. CONCLUSION

In this paper, a safe path estimation method for guiding the visual-impaired people in outdoor sidewalk environment is proposed. Rather than using stereo cameras, the proposed system handles this problem with just single camera. Compared with other single camera solutions, the proposed method takes advantage of a top-view re-sampling process to suppress and eliminate background edges. And by modelling obstacle projections on top-view domain, safe path can be estimated steadily by means of a polar edge histogram. The proposed algorithm can work efficiently in an outdoor sidewalk environment, and provide valuable information to the visual-impaired user.

## REFERENCES

- [1] D. Dakopoulos and N. G. Bourbakis, "Wearable obstacle avoidance electronic travel aids for blind: a survey," *IEEE Trans. Syst., Man, And Cybern.*, vol. 40, no.1, pp. 25–35, Jan. 2010.
- [2] L. A. Johnson and C. M. Higgins, "A navigation aid for the blind using tactile-visual sensory substitution," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, New York, pp.6268-6292, 2006.
- [3] N. Bourbakis, "Sensing 3D dynamic space for blind," *IEEE Engineering in Medicine and Biology Magazine*, vol. 27, no. 1, pp. 49-55, 2008.
- [4] D. Dakopoulos, S. K. Boddhu, and N. Bourbakis, "A 2D vibration array as an assistive device for visually impaired," in *Proc. 7th IEEE Int. Conf. Bioinf. Bioeng.*, Boston, 2007, vol. 1, pp. 930–937.
- [5] S. Meers and K. Ward, "A substitute vision system for providing 3D perception and GPS navigation via electro-tactile stimulation," in *Proc. 1st Int. Conf. Sens. Technol.*, Palmerston North, New Zealand, 2005, pp. 21–23.
- [6] A. Rodríguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazán and A. Cela, "Assisting the visually impaired: obstacle detection and warning system by acoustic feedback," *Sensors*, vol.12, no.12, pp.17476-17496, 2012.
- [7] A. Hub, J. Diepstraten, and T. Ertl, "Design and development of an indoor navigation and object identification system for the blind," in *Proc. ACM SIGACCESS Accessibility Computing*, New York, Jan. 2004, pp. 147-152.
- [8] G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy image processing scheme for autonomous navigation of human blind," *Appl. Softw. Comput.*, vol. 7, no. 1, pp. 257– 264, Jan. 2007.
- [9] M.Bertozzi and M.Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection," *IEEE Trans. Image Proc.* vol.7, no.1, pp: 62–81, 1998.
- [10] J. H. Yu, H. I. Chung, H. S. Hahn, "Walking assistance system for sight impaired people based on a multimodal transformation technique," in *Proc. ICROS-SICE Int. Joint Conf.*, Tokyo, Aug. 2009, pp: 1639-1643.