Piecewise Planar Modeling for Step Detection using Stereo Vision

Vivek Pradeep, Gerard Medioni, James Weiland

Piecewise Planar Modeling for Step Detection using Stereo Vision

Vivek Pradeep[1], Gerard Medioni[2], and James Weiland[1,3]

[1] Department of Biomedical Engineering*
[2] Department of Computer Science
[3] Department of Ophthalmology
University of Southern California, Los Angeles CA, USA

Abstract. A mobility aid for the visually impaired should be able to detect and warn about nearby obstacles. Reliable detection of curbs and steps is critical to meet this goal. This paper describes a stereo-vision based algorithm that estimates the underlying planar geometry of the 3D scene to generate hypotheses for the presence of steps. Tensor voting is used to calculate globally consistent normals at each data point and a clustering algorithm is described to generate a piecewise planar model of the scene. Results demonstrate the improvement in plane clustering using tensor voting and the ability of the algorithm to find sufficient evidence for the presence of curbs and steps.

1 Introduction

There are 10 million blind and visually impaired people in the United States, of which about 110,000 use long canes to get around. A significant part of the population suffers from retinal diseases (such as retinitis pigmentosa and glaucoma) that cause peripheral vision loss/tunnel vision which adversely impacts unaided mobility in indoor and outdoor environments. Recently, there has also been a lot of interest in neural stimulation devices such as the epiretinal prosthesis [1] which employs an external camera to drive an array of microelectrodes implanted onto the retina. Even in the case of such a device, surgical and technological constraints presently limit artificial vision to only the central 20 degrees field of view and therefore, the challenges to independent mobility remain.

1.1 Mobility Aids

Autonomous navigation in an unknown environment requires the ability to detect obstacles and efficiently plan routes across the terrain. For the visually impaired, these are often difficult and time-consuming tasks to perform. While the white cane and the guide dog have been popular aids, there have also been

* This material is based on work supported by the National Science Foundation under Grant No. EEC-0310723.

attempts to design devices to complement their roles. Several mobility and navigational aids leveraging the concept of sensory substitution [2] have been proposed over the years. Sonar has been extensively used ([3], [4], [5]) to provide information about the surroundings by means of auditory cues. Popular, commercial mobility aids ([6], [7]) are also sonar based, but all these devices need significant training and impose severe information-load onto the user. In [8], a robotic 'guide dog' based on RF-ID mapping and sonar obstacle detection for indoor environments is proposed.

The availability of cheap cameras and faster processors has encouraged many researchers to investigate computer vision and image processing strategies on information-rich images/videos to build more intelligent mobility aids. The vOICe [9] is an example of a vision-based device that converts images into sounds, while ASMONC [10] fuses vision and sound sensing modalities to provide scene information. Recently, [11] proposed a system inspired from SLAM literature in robotics that uses a stereo camera to provide orientation information to the visually impaired.

1.2 Step Detection

The importance of curb/step detection for autonomous mobility is well established in literature, with a diverse set of sensors (such as in [12]) being used for the purpose. In [13], an active range-finder employing a laser and camera system is described that uses Jump-Markov modeling to track distances to the ground plane and detect steps.

Solving the step-detection problem relies on obtaining accurate 3D models of the scene and one might expect ladar sensors to be well suited to the task. However, this accuracy comes at the cost of high price, weight and low speed. Furthermore, these sensors have limited field of view and low information bandwidth. A stereo camera system is a logical alternative and the acquired images enable other functionalities to be integrated.

In [14], a stereo-vision based algorithm is described that finds parallel lines using the Canny edge detector and Hough Transform to identify curbs. Step-ups and step-downs are identified by estimating the ground plane parameters on either side of the detected curb. Building on this idea, [15] combines stereo data with image intensity information to detect steps. Both the methods mentioned here rely on detecting curb-lines in the scene, which might not be trivial in cluttered environments or from certain viewpoints.

1.3 Paper Overview

We propose a stereo vision based mobility aid that performs the tasks of obstacle detection and identifying a safe path for traversal. The stereo-vision system is mounted on an eyepiece and allows the subject unrestricted head motions and auditory cues warn about nearby obstacles. Our approach is to combine SIFT features with stereo data for performing 6DOF SLAM and use the resulting 3D maps along with orientation information as input to an interpretation layer. The

tasks of obstacle, curb and stair-case detection are done in this interpretation layer by building a piecewise planar model of the scene. In this paper, we focus only on the plane-fitting module that we have developed using tensor voting. This is a more general approach to step detection and plane fitting offers further advantages in modeling the surrounding environment. In the next section, we discuss the issues in plane extraction using stereo data and present our method. We also compare with standard techniques to highlight the advantages of tensor voting for this task. Finally, as this is still a work in progress, we outline future goals to be met before this module can be be successfully integrated into the system.

2 Plane fitting and segmentation

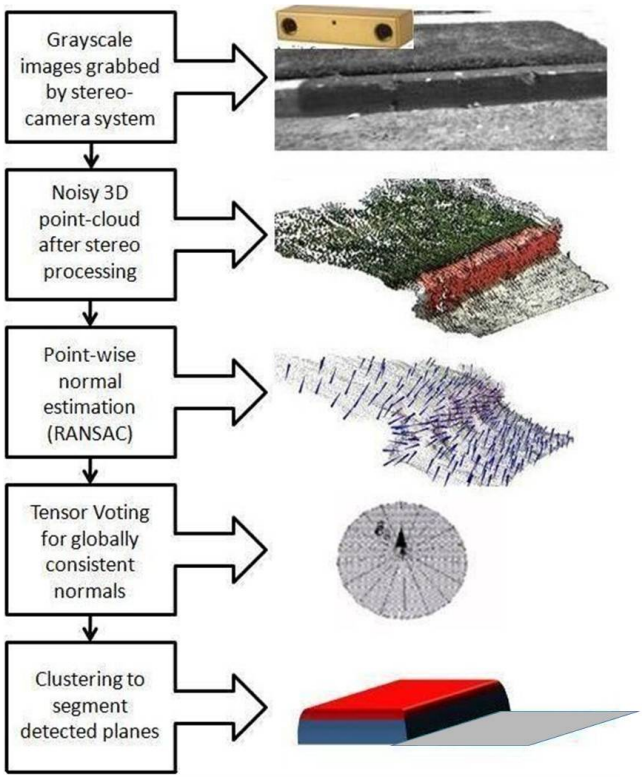

Fig. 1. Block diagram of our piecewise planar modeling algorithm. The stereo camera generates a dense 3D point-cloud, to which we fit normals by RANSAC. These normals are globally inconsistent because they are the result of local processing in each voxel of data. Tensor voting is employed to refine these normals and clustering is finally performed to segment out the fit planes

Figure 1 presents a block diagram of our piecewise planar modeling approach for step detection along with the result of each process. The stereo-camera produces a dense 3D point-cloud of the scene. A commercial system (Bumblebee by Point Grey Research), which does stereo processing 30 frames a second, is used to generate this cloud. We first fit planes to this 3D data and cluster parallel planes together. We then proceed to segment within each cluster to separate parallel but distinct planes.

2.1 Issues

Depth information obtained by stereo is inherently noisy due to several factors. As objects farther away have smaller disparities, error tends to increase with depth. Incorrect stereo matching, occlusions and shadows lead to patches where there might not be any 3D information. Given this noisy data, our task is to simultaneously fit planes to the various structures in the scene as well as segment these planes from each other. The first step to plane fitting is the estimation of a local normal at each 3D data point. One popular technique is to consider the neighborhood about each point and use PCA-based methods. For example, one may simply use SVD to calculate a plane for each voxel of the 3D point cloud. Since there might be outliers in the data, a RANSAC based algorithm can be devised for robustness. However, since this is based on local processing, the obtained normals do not end up globally consistent. The size of the voxel has a significant impact on the estimated plane. High noise content in the voxel calls for a larger size whereas, potential curvature in the data requires keeping voxel size restricted. [16] prescribes an adaptive technique to determine voxel size based on curvature and noise content. Since structures belonging to steps and curbs tend to be small (given that the camera is head-mounted and we wish to detect them from a safe distance away), curvature begins to play an early role such that we are still faced with the effects of local processing.

2.2 Tensor Voting for Normal Estimation

We employ tensor voting [17] to improve the normal estimates at each point. Specifically, we encode our 3D data as a set of points associated with 3x3 tensors. The tensor structure is obtained from initial normals evaluated by the RANSAC-SVD technique. Suppose the plane at a point is given by $P = ai + bj + ck + d$, where i, j, k are the orthogonal unit vectors. Then, a, b, c are the eigen-values along the eigen-vectors $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$ respectively. Further details of tensor voting can be found in [17]. The end result of this processing step is a set of refined normals which are more consistent with the underlying 3D structure of the data. The element of noise is taken care of by the low number of votes cast for any consistent normal by outliers.

Once accurate normals have been obtained, clustering is a trivial task. Assuming ground to be the XZ plane, we compute a histogram of the angles θ between the normals at the data points and the normal to the ground plane.

A simple Hough Voting over the space of these angles clusters points belonging to parallel planes together. For instance, all horizontal planes will vote for $\theta = 0^o$ while vertical planes vote for $\theta = 90^o$. Currently,we are assuming that the ground plane is known, given the camera height. In our final system, we aim to use the camera orientation information obtained from 6DOF visual SLAM to constrain the search for the ground plane. Parallel planes can then be easily separated from each other by casting votes in the space of their distances from the origin.

3 Comparison and Results

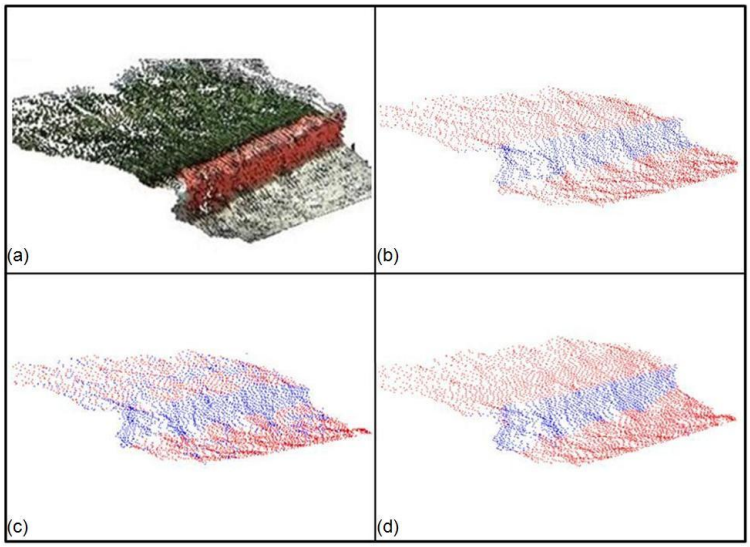

Fig. 2. Comparison of results obtained before Tensor Voting and after. (a) Original 3D point-cloud (b) Input point-cloud manually clustered into horizontal (red) and vertical (blue) planes to generate ground truth (c) Segmentation after RANSAC processing only (d) Result after tensor voting

Figure 2 presents a comparison of the results obtained after tensor voting to that using a RANSAC only based approach. We have manually clustered the points of a curb scene into horizontal and vertical planes to create a ground truth. We show the clusters obtained using the two different approaches. In this result, parallel planes have not yet been segregated and this will be done in the next step. The effect on local processing and noise can be easily seen in the RANSAC result, where globally inconsistent normals lead to several misclassifications. Figure 3 displays similar results for another scene with two steps.

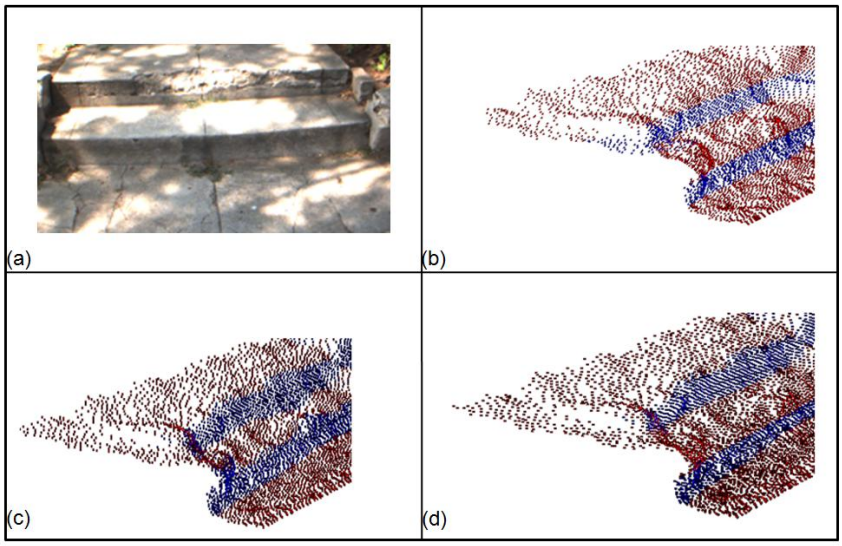

Fig. 3. Comparison of results obtained before tensor voting and after. (a) Original gray scale image (b) Input point-cloud manually clustered into horizontal (red) and vertical (blue) planes to generate ground truth (c) Segmentation after RANSAC processing only (d) Result after tensor voting

Our MATLAB implementation takes about 10 seconds per frame running on a 1.8 GHz PC with 1 GB RAM. Images are of size 320×240. We have not yet optimized our code for performance. Figure 4 shows the final plane clusters obtained for the scenes in figures 2 and 3.

To quantitatively analyze the performance of tensor voting, we have computed the fraction of misclassified pixels (e.g., if a horizontal plane pixel was classified as a vertical plane pixel) relative to the manually clustered ground truth. The values for the scenes in figures 2 and 3 are shown in Table 1.

4 Conclusion and Future work

We have presented a piecewise planar modeling algorithm that can be used in a module for curb/step detection. This is proposed as part of a stereo-vision based mobility aid that performs 6DOF SLAM in conjunction with obstacle detection. Our approach is to use tensor voting to calculate globally consistent normals and perform clustering to fit planes onto the 3D point cloud. This is important because standard, RANSAC-based approaches have been demonstrated to be insufficient given noisy stereo data. We are currently working on a module that can interpret this derived structure to recognize curbs, stairs, step-ups and step-downs. Using SLAM as the backbone, we hope to be able to track the ground plane after an initial calibration and concurrently apply the piecewise planar

modeling module to warn about step changes. Apart from a faster implementation,we also plan to incorporate multi-view integration for increased robustness. Validation on other data-sets and extensive testing on human subjects are also planned in the next few months.

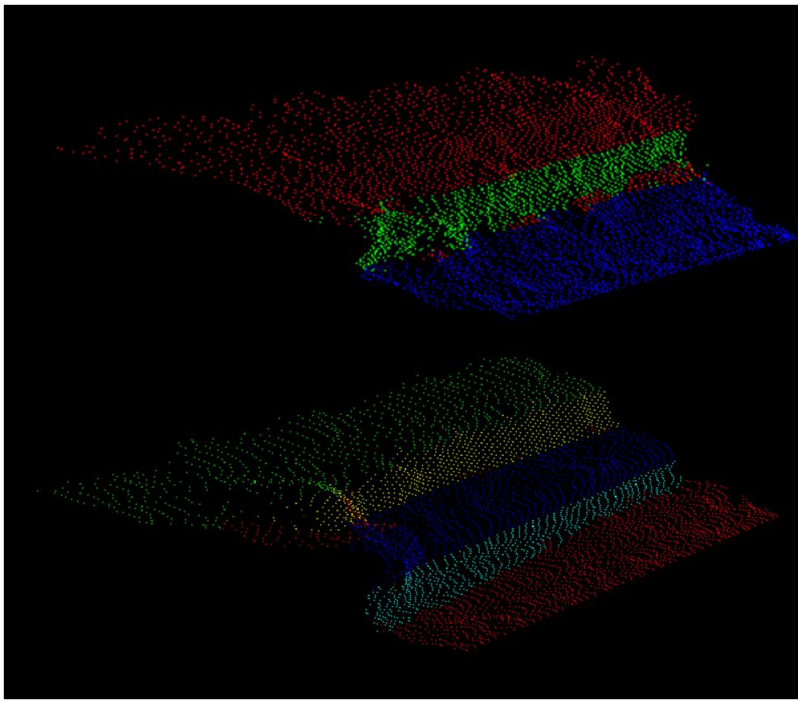

Fig. 4. Final planes segmented from the scenes in (top) figure 2 and (bottom) figure 3

Table 1. Misclassification fractions for the scenes in figures 2 and 3, before and after tensor voting. Lower misclassification fraction indicates better performance

Scene	RANSAC only	After Tensor Voting
Single Step (figure 2)	0.253	0.073
Two steps (figure 3)	0.351	0.067

References

1. Weiland, J.D., Humayun, M.S.: Intraocular retinal prosthesis. IEEE Engineering in Medicine and Biology Magazine **25** (2006) 60–66
2. Arno, P., et al.: Auditory coding of visual patterns for the blind. Perception **28** (1999) 1013–1029
3. Russel, L.: Travel path sounder. Proceedings of the Rotterdam Mobility Research Conference, New York (1965)
4. Kay, L.: An ultrasonic sensing probe as a mobility aid for the blind. Ultrasonics **2** (1964) 53–59
5. Laurent, B., Christian, T.N.A.: A sonar system modeled after spatial hearing and echolocating bats for blind mobility aid. International Journal of Physical Sciences **2** (2007) 104–111
6. Borenstein, J., Ulrich, I.: The guidecane-a computerized travel aid for the active guidance of blind pedestrians. IEEE International Conference on Robotics and Automation (1997) 1283–1288
7. Dodds, A., , Howarth, D.C.C.C.: The sonic path finder:an evaluation. Journal of Visual Impairment Blindness **78** (1984) 206–207
8. Kulyukin, V., et al.: Robot-assisted wayfinding for the visually impaired in structured indoor environments. Autonomous Robots **21** (2006) 29–41
9. Meijer, P.B.L.: Vision technology for the totally blind. (www.seeingwithsound.com)
10. Molton, N., et al: Robotic sensing for the partially sighted. Robotics and Autonomous Systems **26** (1999) 185–201
11. Saez, J.M., , Escolano, F., Penalver, A.: First steps towards stereo-based 6dof slam for the visually impaired. IEEE Conference on Computer Vision and Pattern Recognition **2** (2005) 23
12. Thorpe, C., et al.: Driving in traffic: Short-range sensing for urban collision avoidance. Proceedings of SPIE: Unmanned Ground Vehicle Technology IV **4715** (2002)
13. Yuan, D., Manduchi, R.: Dynamic environment exploration using a virtual white cane. IEEE Conference on Computer Vision and Pattern Recognition **1** (2005) 243–249
14. Se, S., Brady, M.: Vision-based detection of kerbs and steps. British Machine Vision Conference (1997) 410–419
15. Lu, X., Manduchi, R.: Detection and localization of curbs and stairways using stereo-vision. International Conference on Robotics and Automation (2005) 4648–4654
16. Mitra, N.J., Nguyen, A.: Estimating surface normals in noisy point cloud data. Proceedings of the Nineteenth Annual Symposium on Computational Geometry (2003) 322–328
17. Medioni, G., et al.: Tensor voting: Theory and applications. European Conference on Computer Vision (2002)