A Segmentation-Based Stereovision Approach for Assisting Visually Impaired People

Hao Tang and Zhigang Zhu

Department of Computer Science, CUNY City College
Convent Avenue and 138th Street, New York, NY 10031
{tang,zhu}@cs.ccny.cuny.edu

Abstract. An accurate 3D map, automatically generated in real-time from a camera-based stereovision system, is able to assist blind or visually impaired people to obtain correct perception and recognition of the surrounding objects and environment so that they can move safely. In this paper, a segmentation-based stereovision approach is proposed to rapidly obtain accurate 3D estimations of man-made scenes, both indoor and outdoor, with largely textureless areas and sharp depth changes. The new approach takes advantage of the fact that many man-made objects in an urban environment consist of planar surfaces. The final outcome of the system is not just an array of individual 3D points. Instead, the 3D model is built in a geometric representation of plane parameters, with geometric relations among different planar surfaces. Based on this 3D model, algorithms can be developed for traversable path planning, obstacle detection and object recognition for assisting the blind in urban navigation.

1 Introduction

Using portable or wearable systems to assist blind or visual impaired for navigation attracts more and more attention during last decade. The algorithms can be categorized into three main groups: Electronic travel aids (ETAs), electronic orientation aids (EOAs) and position locator devices (PLDs). We are mostly interested in ETAs that could be used in a GPS denied environment.

In this paper, we propose a rapid segmentation-based stereovision approach to generate dense 3D maps. It is efficient since it is a feature-based matching approach. The dense 3D map is accurate because it is propagated from accurate 3D measurements of some well related salient features. The outcome of the system is not just an array of individual 3D points that are usually produced by a typical stereovision system. Instead, it is a geometric representation of plane parameters, with geometric relations among neighboring planar surfaces.

The 3D maps should be transduced to users, blind/or visually impaired, thorough auditory description and other types of "displays", such as vibrotactile or Braille, so that they can make a decision for navigation. Therefore, it is useful to provide uncertainty measurements of those planar regions to tell users how reliable the 3D map is.

K. Miesenberger et al. (Eds.): ICCHP 2012, Part II, LNCS 7383, pp. 581–587, 2012.

The paper is organized as the following. Section 2 discusses a few closely related works. In Section 3, we present our segmentation-based stereo vision approach. Section 4 provides some experimental results, with discussions in transducing stereovision results to some novel simulation devices. Finally we conclude our work in Section 5.

2 Related Work

There are various sensors used in ETA systems, such as video cameras [2], [10], [11], [15], ultrasound rangers [12], [20], and sonars [1], [3], [5]. Video camera become popular sensors in such systems due to the availability of low-cost, small CCD or CMOS cameras and recent advance in the computer vision algorithms. Coughlan et al. [7], [8] propose systems for helping visually impaired to find a path to a machine-readable sign using a cellphone camera. Using stereo cameras [2], [10], depth maps are produced to aid navigation. Staircases [14], [16], [17] and zebra-crossings [18] are detected using stereo cameras. However, above methods require a textured environment, and will not work well in textureless area because matching in textureless areas is ambiguous. Though stereovision using global optimization frameworks [4] may obtain more robust results, the computation is expensive and is not suitable for this application.

3 Our Algorithm

Before we go into more details of the algorithm, here is an overview. For a pair of stereo images, the left view is used as the reference view, color segmentation is performed on this image, and the so-called natural matching primitives (Fig.1a, details explained later) are extracted. Multiple natural matching primitives are defined for each homogeneous color image patch, which approximately corresponds to a planar patch in 3D. Then the matches of those natural matching primitives are searched for in the right image, a plane is fitted for each patch, and its planar parameters are estimated. To improve the robustness of stereo matching, each planar patch is warped between views to evaluate the matching accuracy, and uncertainty values are generated.

There are three major steps in our algorithm for the segmentation-based stereo matching: (1) matching primitive extraction; (2) patch-based stereo matching and plane fitting; and (3) plane merging, splitting and refinement. We will discuss them in the next three subsections.

3.1 Matching Primitive Extraction

First, the reference image is segmented, using a mean-shift based approach [6]. The segmented image consists of image patches with homogeneous colors, and each of them is assumed to be a planar patch in 3D space. For each patch, its boundary is extracted as a closed curve. Then we use a line fitting approach to extract feature points for stereo matching. The boundary of each patch is first fitted with connected straight-line segments using an iterative curve splitting method. The connecting points (with large curvature) between line segments are defined as interest points around which the natural matching primitives are defined (Fig. 1a).The representations are effective for urban scenes with objects of largely textureless regions and sharp depth boundaries.

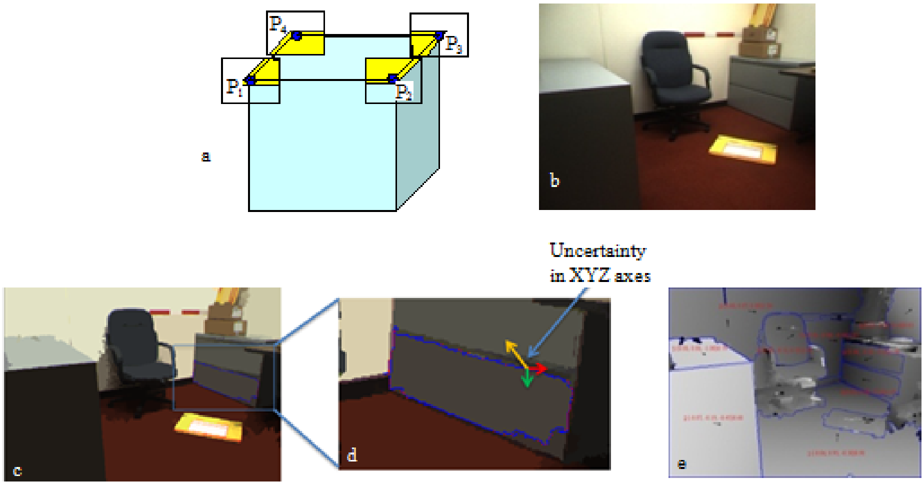

Fig. 1. a. Natural matching primitives around four interest points on a patch (the top of a rectangular object in this illustration); b: The left stereo image (reference image); c: The segmented image of the reference image (b); d: In a cropped window, boundary and features are drawn in a patch on a cabinet. Gaussian uncertainty measurement of a feature is drawn. e. Depth image with the boundaries and plane parameters of some patches overlaid.

3.2 Stereo Matching and Plane Fitting

On each patch of the reference image, the match for each interest point is searched along its epipolar line in the target image. We define a matching mask centered at each interest point, which only includes points on the patch (Fig.1a). The size of the mask is adaptively changed depending on the actual size of the patch. A few more pixels (e.g., 1-2) around the region boundary (but not belonging to the region) are also included so that we have sufficient salient image features to match. A sub-pixel search is performed in order to improve the accuracy of 3D reconstruction. For an interest point in the reference image, (x_l, y_l), we use the epipolar geometry to find its possible matches, using correlation. Then based on the correlation curve (in both the epipolar line and a few pixel above and below), we find a range of possible matches in the target image, and we then sample a number of points $\{(x_r, y_r)\}$. Finally we obtain the 3D coordinates $\{(X_i, Y_i, Z_i)\}$ of the points, and fit a 3D Gaussian distribution so the mean and variance values in X, Y, and Z can be obtained (Fig. 1d).

Assuming that each homogeneous color region is a planar patch in 3D, a plane

$$aX + bY + cZ + d = 0 \tag{1}$$

is fitted to each patch after obtaining the 3D coordinates of the interest points of the patch. We use a robust RANSAC method to fit a plane. In the voting step, interest points with higher confidences are able to vote more tickets, the uncertainty values of the interest points are also updated using the plane fitting result. The result of a patch after the above steps is in the form of a 3D planar equation and the boundary of each patch with 3D coordinates and their uncertainties. The process is very efficient, particularly for a large textureless region, like the surfaces of a desk, walls and doors. The interest points of a patch that lie on the image borders are not taken into account (marked with very large uncertainty) therefore partially visible regions can also be correctly handled.

3.3 Plane Merging, Splitting and Parameter Refinement

One real-world planar surface may have been segmented to several sub-regions during segmentation. In order to recover meaningful surface structure, we try to combine them back into one surface. On the other hand, a patch may include multiple planar surfaces due to lack of texture. To solve the problem, first we perform a modified version of the neighboring plane parameter hypothesis approach [21] to infer better plane estimates. The main modifications are: first the plane hypothesis is only provided by patches with small uncertainty. Second, the neighboring regions sharing the same or very close plane parameters are merged into one larger region. This procedure is performed recursively till no more merging or spitting occurs.

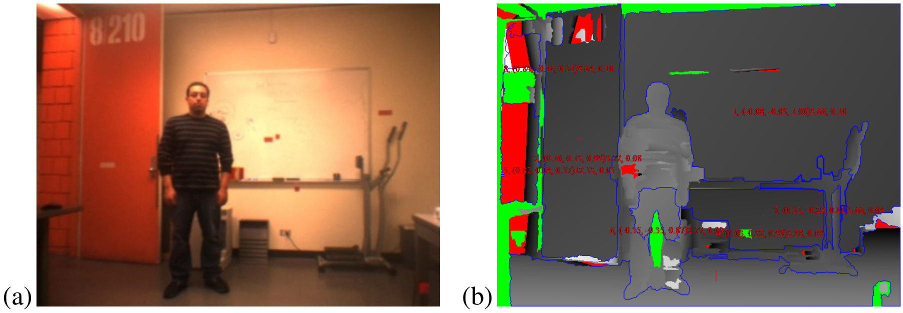

Fig. 2. (a) The left stereo image (reference image); (b) Depth image with the boundaries and plane parameters of some patches overlaid, regions with large uncertainty (wrong estimate) are marked in color

4 Experimental Results

Experiments have been performed to test our approach. Image sequences were captured by the stereovision head Bumblebee. The baseline distance between the left and the right cameras is 12 cm, and focal length of each is 3.8 mm. The stereo system has been pre-calibrated and image pairs rectified.

In Fig. 1b, all objects (a table, a chair, a cabinet, walls and some boxes) are about 1-4 meters away from the camera. A depth map (the brighter, the closer) rendered from the results of the plane parametric estimation has been shown in Fig. 1e. Several large surfaces are annotated by their normalized surface normal vectors – using arrows and values (a,b,c), and distances of the centers of the patches to the camera (d, in meters), labeled with their boundaries. These plane estimation results are consistent with the results measured by hand. For the textureless regions in the experiment, e.g. the doors, the box and even the ground surface, full 3D results are also obtained. Fig. 2 shows another example of indoor scene, including a person sitting in front of camera, 3D models of large surface/objects (walls, door, ground are person) are corrected recovered.

One of our ongoing works is to apply the stereo vision results into visual prosthetic approaches. A retinal implant is used to partially restore vision for blind and visually impaired, especially who lost their vision due to retinitis pigmentosa or macular degeneration. Currently the state of art retina implants have limited resolution (60 – 100 channels) [22]. As another example, Brainport technique [23] invented by Wicab Inc. captures an image and processes the image by converting it into impulses which are sent via electrode array on the tongue (tongue simulation) to brain that is able to interpret the impulses into visual signals. The tongue simulation has 400 channels (20x20). Both methods are facing a problem of low resolutions. If we simply subsample an original image into a 20x20 or lower resolution array to drive the retinal or tongue stimulation, is would be hard to identify small objects that are close to the user.

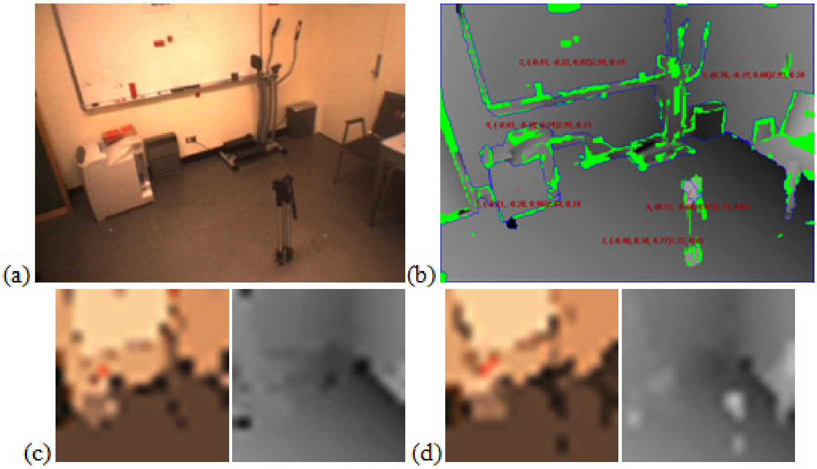

Fig. 3. (a) an indoor scene (left view) is captured in an office with a number of objects (note: a tripod is in a close range); (b) 3D depth map of the indoor scene; pixels with large uncertainty are marked in green; (c) sampling results of 2D image and 3D depth map using uniform sampling method: the tripod is missing after regular sampling; (d) sampling results of 2D image and 3D depth map using smart sampling method: the tripod is kept after sampling.

Therefore we applied a special sub-sampling method (refer smart sub-sampling) to sample 2D images and 3D depth maps and transduce the sampled results into the above devices that have limited resolutions so blind or visual impaired people can better ‘see’ surrounding environments through alterative perceptions. Instead of uniformly sampling, the smart sub-sampling method can convey more important information with a limited resolution. The main idea is to use both the segmentation and 3D reconstruction results to keep the small but close objects in the sub-sampled images. Fig. 3 shows an example after applying smart sub-sampling. Fig. 3a and 3b are left view of a stereo images and recovered depth map using the proposed method, respectively; Fig. 3c shows that the tripod, which is about 1.55 meters from the user, is missing after uniform sampling, but it is still preserved using the proposed smart sub-sampling (Fig 3d).

5 Conclusion

In both indoor and outdoor urban environments, most of the surfaces of man-made objects are planes; therefore a 3D reconstruction method that directly produces plane surfaces is an appropriate approach to solve the navigation problem for blind or visually impaired individuals. In this paper, we have proposed a segmentation-based stereo approach that features natural matching primitives, three-step efficient matching and accuracy parametric 3D estimation. Planar surfaces are represented by their plane parameters (orientations, distances, boundaries), and their relations, Based on this 3D model, algorithms in traversable path planning, obstacle detection and object recognition will be developed for assisting the blind in urban navigation. In our future work, we will test our segmentation-based stereovision approach for both 3D reconstruction and transducing with visually impaired users.

Acknowledgments. This work is supported by US National Science Foundation Emerging Frontiers in Research and Innovation Program under Award No. EFRI-1137172, and City Seeds: City College 2011 President Grant for Interdisciplinary Scientific Research Collaborations.

References

1. Aguerrevere, D., Choudhury, M., Barreto, A.: Portable 3D sound / sonar navigation system for blind individuals. In: The 2nd LACCEI Int. Latin Amer. Caribbean Conf. Eng. Technol. Miami, FL, June 2–4 (2004)
2. Audette, R., Balthazaar, J., Dunk, C., Zelek, J.: A stereo-vision system for the visually impaired, Sch. Eng., Univ. Guelph, Guelph, ON, Canada, Tech. Rep. 2000-41x-1 (2000)
3. Bouzit, M., Chaibi, A., De Laurentis, K.J., Mavroidis, C.: Tactilefeedback navigation handle for the visually impaired. In: ASME Int. Mech. Eng. Congr. RD&D Expo., Anaheim, CA, November 13–19 (2004)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Patten Analysis and Machine Intelligence 23(11) (November 2001)

5. Cardin, S., Thalmann, D., Vexo, F.: A wearable system for mobility improvement of visually impaired people. Vis. Comput. 23(2), 109–118 (2007)
6. Comanicu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Patten Analysis and Machine Intelligence (May 2002)
7. Coughlan, J., Manduchi, R., Shen, H.: Cell phone-based wayfinding for the visually impaired. In: 1st International Workshop on Mobile Vision (2006)
8. Manduchi, R., Coughlan, J., Ivanchenko, V.: Search Strategies of Visually Impaired Persons Using a Camera Phone Wayfinding System. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 1135–1140. Springer, Heidelberg (2008)
9. Dakopoulos, D., Bourbakis, N.: Wearable obstacle avoidance electronic travel aids for blind: a survey. IEEE Trans. on Systems, Man, and Cybernetics, January 1 (2010)
10. Gonzalez-Mora, J.L., Rodrıguez-Hernandez, A., Rodrıguez-Ramos, L.F., Dıaz-Saco, L., Sosa, N.: Development of a new spaceperception system for blind people, based on the creation of a virtual acousticspace. Tech. Rep., May 8 (2009)
11. Hub, A., Diepstraten, J., Ertl, T.: Design and development of an indoor navigation and object identification system for the blind. In: Proc. ACMSIGACCESS Accessibility Computing, vol. 77–78, pp. 147–152 (September 2003/January 2004)
12. Ifukube, T., Sasaki, T., Peng, C.: A blind mobility aid modeled after echolocation of bats. IEEE Trans. Biomed. Eng. 38(5), 461–465 (1991)
13. Liu, J., Cong, Y., Li, X., Tang, Y.: A stairway detection algorithm based on vision for UGV stair climbing. In: IEEE Networking, Sensing and Control (2008)
14. Lu, X., Manduchi, R.: Detection and localization of curbs and stairways using stereo vision. In: IEEE International Conference on Robotics and Automation, ICRA (2005)
15. Meijer, P.B.L.: An experimental system for auditory image representations. IEEE Trans. Biomed. Eng. 39(2), 112–121 (1992)
16. Pradeep, V., Medioni, G., Weiland, J.: Piecewise planar modeling for step detection using stereo vision. In: Workshop on Computer Vision Applications for the Visually Impaired (2008)
17. Se, S., Michael, B.: Vision-based Detection of Stair-cases. In: Asian Conference on Computer Vision, ACCV (2000)
18. Se, S.: Zebra-crossing detection for the partially sighted. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR (2000)
19. Shah, C., Bouzit, M., Youssef, M., Vasquez, L.: Evaluation of RU-netra–tactile feedback navigation system for the visually impaired. In: Proc. Int. Workshop Virtual Rehabil, New York, pp. 71–77 (2006)
20. Shoval, S., Borenstein, J., Koren, Y.: Mobile robot obstacle avoidance in a computerized travel aid for the blind. In: Proc, IEEE Int. Conf. Robot. Autom., San Diego, CA, May 8–13, pp. 2023–2029 (1994)
21. Tao, H., Sawhney, H.S., Kumar, R.: A global matching framework for stereo computation. In: Proc. Int. Conf. Computer Vision (2001)
22. Second Sight, `http://2-sight.eu/en/home-en` (last visited April 2012)
23. BrainPort Vision Technology, `http://vision.wicab.com/technology` (last visited April 2012)