A Simple Approach to Stereo Matching and Its Application in Developing a Travel Aid for the Blind

Mu-Chun Su, Yi-Zeng Hsieh, and Yu-Xiang Zhao

Department of Computer Science & Information Engineering, National Central University, Taiwan, R.O.C.
E-mail: muchun@csie.ncu.edu.tw

Abstract

In this paper, we present an idea of using stereo matching to develop a travel aid for the blind. In this approach, images are segmented into several non-overlapping homogeneous regions using a color segmentation algorithm. For each homogeneous region, a rectangular window, which is large enough to cover the region, is found. A local match with the found rectangular window size is then executed to find the disparity for the considered region. A clustering algorithm is adopted to cluster the disparities into several major different values. Finally, a piece-wise disparity map is constructed. Based on the disparity map, information about the unfamiliar environments in front of the blind will be output to them. With the information about the environment the blind will have less fear in walking through unfamiliar environments via white canes.

Keywords: stereo matching, travel aid, fuzzy c-means algorithm, assistive technology

1. Introduction

Stereo matching has traditionally been and continues to be one of the most active research areas in computer vision. Traditionally, it serves as an important step in many applications such as the navigation of an autonomous or robotic vehicle, image-based rendering, view synthesis, parts inspection, etc. In addition to the aforementioned applications, we believe that stereo matching can also be applied in developing a travel aid for the blind. Since a lot of amount of efforts and budgets have been put on the research of the navigation of an autonomous or robotic vehicle there is no reason why we cannot switch some research efforts and budgets to develop a vision-based travel aid for the blind to help them to live more independently. If a travel aid can inform the blind in time about a rough sketch of the unfamiliar environments in front of them then they will have less fear in walking through unfamiliar environments via white canes.

Depth map of a scene can provide geometrical information about objects in the scene and be estimated through stereo computation using multiple images. The goal of stereo matching is to determine the disparity map between a reference image and a matching image taken from the same scene. Disparity describes the difference in location of the corresponding pixels in the left image and the right image. It is directly related to the distance of an object normal to the image plane.

While a huge amount of algorithms for stereo matching have been proposed, the computation of accurate disparity still remains challenging in textureless regions, depth discontinuities and occluded areas. Several good reviews of stereo matching can be found in [1]-[4]. Roughly speaking, stereo algorithms can be dichotomized into three classes. The first class is the area-based algorithms. These algorithms attempt to correlate the gray levels of pixels within a finite neighboring window. A central problem of the area-based algorithms is to find the optimal size of the window. While the window size must be large enough to include enough intensity variation for matching, it must be small enough to avoid the effects of projective distortion. Adaptive-window methods have been proposed in [5]-[6]. The second class is the feature-based algorithms. They extract features of interest from the image, such as edges, and match the features. The disadvantage of the feature-based algorithms is that they usually yield sparse disparity maps. The third class is the global algorithms. The global algorithms formulate the problem of the disparity computation as an energy-minimizing problem and the smoothness assumption of the disparity map is usually made [7]-[15]. However, the smoothness assumption may not always be true in many practical cases.

Recently, several segment-based stereo algorithms have been proposed [16]-[19]. These algorithms assume that there are no large disparity discontinuities inside homogeneous color segments. Some of them segment the image into non-overlapping segments based on the color information. A local window matching method is employed to estimate a rough depth map. Then a plane is fitted for each segment using a robust surface fitting algorithm. Finally, a more accurate disparity map is computed by minimizing a kind of energy function.

We follow the assumption that there are no large disparity discontinuities inside homogeneous color segments. In our approach, the reference image is first segmented into several homogeneous color regions. A rectangular window with large enough size to cover the region to be considered is found. Then a local window matching method with the found rectangular window size is used to find the disparity of the region. Since we are only concerned about the object within a close distance from the user a clustering algorithm is employed to segment the disparities into several major ranges.

The remaining of the paper is organized as follows. Section 2 describes the stereo matching algorithm. Section 3 presents the vision-based travel aid (VITA) for the blind. Section 4 gives the conclusions.

2. The Stereo Matching

The proposed simple stereo matching also adopts the idea of color segmentation inspired by [16]-[19]. They assumed that large disparity discontinuities only exist on the boundaries of homogeneous color regions. The proposed simple stereo matching algorithm involves the following steps:

Step 1: Color segmentation:

We adopt the k-means algorithm to segment the reference image into several homogeneous color regions.

Step 2: Disparity computation:

A local window matching method was adopted by [16]-[17] to estimate a rough disparity map. Then a plane fitting was used to model continuous disparity of each segmented region. The finally disparity map is constructed by minimizing an energy function incorporated with the smoothness assumption. They claimed that their approaches are able to effectively deal with large textureless regions.

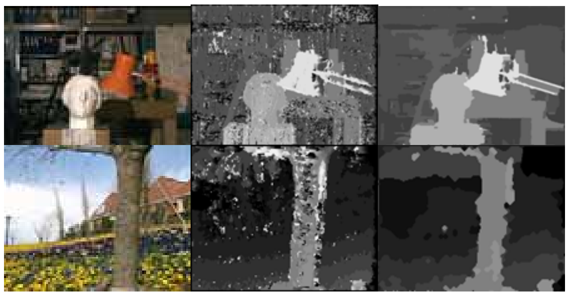
We adopt a different way to deal with the problem of textureless regions. We know that the performance of the local window matching algorithm can be improved if an appropriate window size can be adaptively selected [5]-[6]. They introduced a statistical model of disparity variation to select the window size. Since we assume that large disparity discontinuities only occur on the boundaries of homogeneous color regions, a rectangular window which covers the region to be considered will be the appropriate size to provide enough discrimination between the region and its surrounding regions. The disparity computation proceeds as follows.

For region R_i , a local window match method incorporated with the sum of absolute differences is used to compute the disparity. Each pixel point in region R_i has the same disparity as the center pixel in the region. This process is repeated until all regions are processed.

Step 3: Disparity clustering:

Smooth disparity solution is preferred to meet of the stereo matching algorithms. These algorithms usually formulate the problem as an energy-minimization problem. If gradient-based optimization methods are adopted the local minima problem will usually occur. If population-based optimization methods are adopted, lengthy convergence time will be a problem. In our approach, we just adopt a clustering algorithm such as, fuzzy c-means algorithm or k-means algorithm to cluster the estimated disparities into several different major values so regions with similar disparities will be assigned to be with the same disparity.

Some disparity maps constructed by our stereo matching algorithm are shown in Fig. 1. From Fig. 1, we find that the performance of our algorithm is acceptable.



(a) (b) (c)

Fig. 1. Results of our approach. (a) Reference view. (b) Disparity map achieved by the local window matching method. (c) Disparity map achieved by our approach.

3. The Vision-Based Travel Aid

The potential usefulness of electronic travel aids (ETAs) is unquestionable since ETAs can help the blind persons to perceive environments surrounding them. Some best known ETAs are the Guide cane [20], the sonic pathfinder [21, the NavBelt [22], the MELDOG [23], etc. Most of them are based on the use of ultrasonic sensors or laser beams. Some of them are robotics-based ETAs. Each has its own merits, limitations, and disadvantages. For example, the ultrasonic-based ETAs require the user to actively scan the environment to detect obstacles. Some ETAs are too expensive to be affordable for the blind. Although there are so many different ETAs and some of them are even commercially available, their acceptance rate by the blind persons is surprisingly low [24].

Nowadays, most of the blind persons still rely on the use of white canes. While the white cane has many appealing properties such as light weight, small size when it is folded, and low cost, its main problem is that it cannot detect obstacles beyond its reach of 1-2m.

Based on the aforementioned discussions, we propose a vision-based travel aid (VITA) to complement the white cane. The proposed aid will expand the environmental detection range of the blind persons with white canes. Similar idea of using stereo images for the visually impaired people's navigation has been proposed in [25]-[26]. Compared to these approaches, our VITA has some more appealing properties. The VITA consists of a pair of Web cameras, a portable computer, and an ear phone. The goal of the Web cameras is to build the disparity map. Via the constructed disparity map, a sketch of the environment can be provided to the user. When a blind person is in an unfamiliar place, if the user can quickly get the information of the positions and orientations of the obstacle in front of him or her then he or she can have more confidence in walking through the place via the white cane. The information about the obstacles will be output to the user via the earphone.

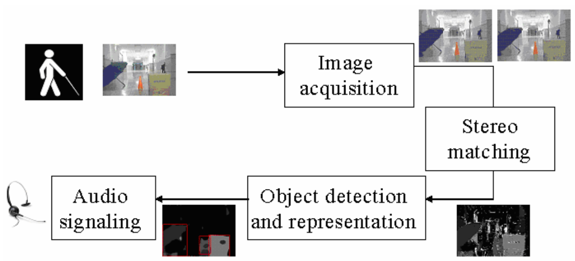


Fig.2. The diagram of the vision-based travel aid.

The VITA involves the following four stages, as shown in Fig. 2.

Step 1: Image acquisition:

The two low-cost web cameras acquire two images of the environment in front of the user.

Step 2: Stereo Matching:

The disparity map of the two images is computed at this stage.

Step 3: Object detection and representation:

Since we are only concerned about a sketch of the environment, a connected-component operator is applied to find objects with similar values of disparities. A rectangular-window is fitted to each detected object to provide a rough estimate of the object position and size. The distance and the height of the object are computed as follows:

$$r = bf/d \tag{1}$$

$$h = w_h \cdot r / f \tag{2}$$

where r is the distance of the object normal to the image plane, h is the height of the object, d is the disparity of the center of the object, b is baseline, f is the focal length of the lenses, and w_h is the height of the object.

The objects will be sorted in the increasing order of the distances. Finally, the information about the objects, such as distances, sizes, heights, and orientations, will extracted from the map.

Step 4: Audio signaling:

The information about the detect objects will be output to the user via the earphone.

Fig. 3(a) shows a corridor image in front of a user. The disparity map is shown in Fig. 3(b) and three detected objects are highlighted by rectangles in Fig. 3(c). The information of these detected objects will be output to the user via the earphone.

(a) (b) (c)

Fig. 3. Objects detected by the VITA. (a) The corridor image. (b) The disparity map. (c) The detected objects.

4. Conclusions

Stereo matching has long been applied in the navigation of autonomous or robotic vehicles. There is no reason why stereo matching cannot be utilized to help the blind to travel unfamiliar environments. In this paper, we explore the possibility of developing a vision-based travel aid for the blind. We demonstrate that this idea, if efficiently and effectively implemented, can be really helpful to the blind.

Acknowledgement

This work was partly supported by the NSC Program for Promoting Academic Excellent of Universities (Phase II) under the grant number NSC-95-2752-E-008-002-PAE, the Ministry of Economic Affairs under the 95-EC-17-A-02-S1-029, and the National Science Council, Taiwan, R.O.C, under the NSC 94-2524-S-008-002, NSC 94-2213-E-008-032, and NSC 95-2524-S-008-005-EC3.

References

[1] S. T. Barnard and M. A. Fischler, "Computation stereo," *ACM Comp. Surveys*, vol. 14, no. 4, pp. 553-572, 1982.

[2] L. G. Brown, "A survey of image registration techniques," *Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.

[3] U. R. Dhond and J. K. Aggarwal, "Structure from stereo—a review," *IEEE Trans. On Systems, Man, and Cybern.*, vol. 19, no. 6, pp. 1489-1510, 1989.

[4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Int'l J. Computer Vision*, vol. 47, no. 1, pp. 7-42, 2002.

[5] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE TPAMI*, vol. 16, no. 9, pp. 920-932, 1994.

[6] M. Okutomi and T. Kanade, "A locally adaptive window for signal matching," *IJCV*, vol. 7, no. 2, pp. 143-162, 1992.

[7] S. T. Barnard, "Stochastic stereo matching over scale," *IJCV*, vol. 3, no. 1, pp. 17-32, 1989.

[8] A. Blake and A. Zisserman, Visual Reconstruction, *MIT Press*, Cambridge, MA, 1987.

[9] P. B. Chou and C. M. Brown, "The theory and practice of Bayesian image labeling," *IJCV*, vol. 4, no. 3, pp. 185-210, 1990.

[10] D. Geiger and F. Girosi, "Parallel and deterministic algorithms for MRF's: Surface reconstruction," *IEEE TPAMI*, vol. 13, no. 5, pp. 401-412, 1991.

[11] Y. Boykov, O. Veksler and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE TPAMI*, vol. 23, no. 11, pp. 1222-1239, 2001.

[12] H. Ishikawa and D. Geiger, "Occlusions, discontinuities, and epipolar lines in stereo," *ECCV*, pp. 232-248, 1998.

[13] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," *ICCV*, vol. 2, pp. 508-515, 2001.

[14] S. Roy and I. J. Cox, "A maximum-flow formulation of the N-camera stereo correspondence problem," *ICCV*, pp. 492-499, 1998.

[15] O. Veksler, Efficient Grpah-based Energy Minimization Methods in Computer Vision, PhD thesis, Cornell University,1999.

[16] H. Li and C. George, "Segment-based Stereo Matching Using Graph Cuts," *CVPR*, 2004.

[17] H. Tao, H. S. Sawhney and R. Kumar, "A Global Matching Framework for Stereo Computation," *Proc. Int'l Conf. Computer Vision* 2002.

[18] V. Kolmogorov and R. Zabih, "Computing Visual Correspondence with Occlusions using Graph Cuts," *Proc. Int'l Conf. Computer Vision* 2001.

[19] Y. Deng, Q. Yang, X. Lin and X. Tang, "A symmetric patch-based correspondence model for occlusion handling," *ICCV* 2005.

[20] J. Borenstein, I. Ulrich and A. Arbor, "The GuideCane – a computerized travel aid for the active guidance of blind pedestrians," *International conference on Robotics and Automation*, pp. 1283-1288, 1997.

[21] S. L. Grow, "The use of the sonic pathfinder as a secondary mobility aid for travel in business environments: a single subject design," *Journal of Rehabilitation Research and Development*, vol. 36, pp. 55-59, 1999.

[22] S. Shoval, J. Borenstein and Y. Koren, "Auditory guidance with the NavBelt – a computerized travel aid for the blind based on mobile robotic technology," *IEEE Trans. on Biomedical Engineering*, vol. 45, pp. 1376-1386, 1998.

[23] H. Makino, I. Ishii and M. Nakashizuka, "Development of navigation system for the blind using GPS and mobile phone combination," *Proceedings of the 18th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society*, vol. 2, pp. 506-507, 1996.

[24] M. Sanders and E. McCormick, "Human factors in engineering and design," 7th edition, McGraw-Hill, Inc., 1993.

[25] S. Meers and K. Ward, "A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation, " The 8th International Conference on Information Visuallization, pp. 546-552, 2004.

[26] G. balakrishnan, G. Sainarayanan, R. Nagarajan, S. Yaccob, "On stereo processing procedure applied towards blind navigation aid-SVETA," The 8th International Symposium on Signal Processing and Its Applications, pp. 567-570, 2005.