

TERRAIN ANALYSIS FOR BLIND WHEELCHAIR USERS: COMPUTER VISION ALGORITHMS FOR FINDING CURBS AND OTHER NEGATIVE OBSTACLES

James Coughlan and Huiying Shen

The Smith-Kettlewell Eye Research Institute
2318 Fillmore St.
San Francisco, CA 94115
Phone: 415-345-2146, Email: coughlan@ski.org

Abstract: We are developing computer vision algorithms for sensing important terrain features as an aid to wheelchair navigation, which interpret visual information obtained from images collected by video cameras mounted to the wheelchair. This paper focuses specifically on a novel computer vision algorithm for detecting curbs and other negative obstacles (i.e. anything below the level of the ground, such as holes and drop-offs), which are important and ubiquitous features on and near sidewalks and other walkways. The algorithm we develop extracts as much information as possible from depth information obtained from stereo video cameras (i.e. pairs of cameras mounted close together); other information (e.g. monocular cues such as intensity edges) will be incorporated in the future. We demonstrate experimental results on typical sidewalk scenes.

Keywords: blind, visually impaired, wheelchair, assistive technology, computer vision, curbs, obstacle detection

1. Introduction

Approximately one in ten blind persons uses a wheelchair, and independent travel is currently next to impossible for this population. This paper describes computer vision algorithms for detecting curbs and other negative obstacles, which are important and ubiquitous features of sidewalks and other walkways that are especially difficult for blind travelers to find. The algorithms are intended to analyze images acquired from a stereo video camera (see Figure 1a,b) mounted to a wheelchair using a portable computer (also carried on the wheelchair); the information they provide will be communicated to the traveler using synthesized speech, audible tones and/or tactile feedback, and is meant to augment rather than replace the information from existing wayfinding skills. Detecting negative obstacles is challenging since the depth cues signalling negative obstacles are noisy. Our detection algorithm addresses this problem by filtering the depth information in a way that greatly reduces the noise while preserving important information.

2. State-Of-The-Art and Related Technology

The specific problems of visually impaired wheelchair riders have received little study (Greenbaum et al, 1998). Indeed, the only commercial device targeted at this population is a version of the laser cane by Nurion Inc., mounted on the arm of a wheelchair (Gill, 2000). The laser's fixed pencil beam drastically limits its "field of view," while four added ultrasonic sensors detect only large, tall obstacles within one foot. Some technology developed in robotics and autonomous vehicle navigation research may eventually be useful in the design of navigation aids for wheelchairs, but has limitations that

prevent it from being adopted in the near future. For instance, 3-D sensing for environmental mapping in robotics is performed using a single or double axis lidar (similar to radar but using laser light rather than radio waves). Although lidars produce very accurate distance measurements, they are still expensive and bulky, which has prevented their widespread use.

Computer vision, the design of software to interpret visual information obtained from cameras, is a promising technology that overcomes many of the limitations inherent in the above modalities: it relies only on relatively inexpensive and compact hardware (digital cameras and a computer), and can sense obstacles within a wide field of view (and at distances of up to several meters or more).

3. Computer Vision Algorithms for Finding Obstacles

We begin this section with a brief overview of stereo vision (stereopsis); see Forsyth and Ponce 2002 for a comprehensive introduction. Stereo vision is a powerful computer vision method for recovering 3-D scene structure, which works by comparing the differences between images taken by two cameras placed a short distance apart (like human eyes) to estimate depth (see Figure 1a). The fixed geometric relationship between the cameras simplifies depth estimation, making it a fast and relatively robust calculation. For our application, depth estimation is used to determine the ground plane, i.e. plane that the wheelchair is rolling on, and to locate obstacles, which are points in the scene that lie significantly above or below the ground plane.

3.1 Overview of Stereo Vision

Stereo vision exploits the fact that a single point in a scene appears in slightly different locations in neighboring views of the scene. If the views are from two suitably aligned and calibrated cameras, with parallel lines of sight, a feature in the left image is horizontally shifted relative to the feature in the right image. This image shift, called the *disparity*, is directly related to the distance from the camera to the point in the scene: distant points have small disparity, while nearby points have large disparity. The following thought experiment illustrates this fact. If you alternately open and close your left and right eyes while looking at the sky, distant objects (such as stars or the sun) will appear in the same place in both eyes while nearby objects (such as your hand pointing towards the distant object) will appear in different locations in each eye.



Figure 1 (a) Stereo video camera. (b) Left image. (c) Right image. (d) Disparity map $d(x,y)$ produced by stereo algorithm: brighter green means higher disparity and thus closer to camera; black pixels indicate no estimate at that location. The disparity map clearly shows that the person in the images is closer to the camera than the walls behind him.

Stereo algorithms determine the correspondences between points in the left and right images, thereby establishing the disparity $d(x,y)$ everywhere in the image (see Figure 1b,c,d for an example). We note that the process of finding correspondences is a challenging problem that causes much of the noise and errors in stereo vision algorithms (and is an active area of research in computer vision, see Scharstein and Szeliski 2002). Geometric triangulation then yields an equation calculating the distance in terms of the disparity: the depth of any point in the scene is inversely proportional to its disparity.

A fundamental (albeit somewhat non-intuitive) result is that the disparity map corresponding to a planar surface in a scene is itself a planar (i.e. linear) function of image coordinates x,y . (We omit the derivation, which is straightforward, see Forsyth and Ponce 2002.) We will exploit this relationship to locate the dominant plane, or ground plane, in the image.

3.2 Past Applications of Computer Vision Stereo for Finding Obstacles

A variety of work has been done on computer vision algorithms for finding negative obstacles using stereo depth information, primarily in the context of autonomous vehicles (Belluta et al 2000, Labayrade et al 2002). Our work focuses on a different domain, in which the features of interest are difficult to detect from depth information alone, because the depth changes that characterize these obstacles may be small relative to the distances at which they are viewed and may be swamped by noise in the depth information estimated from computer vision stereo algorithms. (For instance, a curb is approximately 10-15 cm high but may be viewed at distances of several meters.) In contrast with recent work by Lu and Manduchi (2005), which combines depth information and monocular intensity information to infer the locations of depth discontinuities that signal the presence of curbs, we focus on extracting as much information as possible from depth information alone.

3.3 Proposed Algorithm for Finding Obstacles

We propose an algorithm for finding the depth discontinuities that signal the presence of obstacles which makes very little use of monocular cues. The key to our approach is that we smooth the noisy disparity map obtained by the stereo camera in such a way that the important discontinuities are preserved, while eliminating a lot of the noise. Our goal is not to argue that monocular cues should be avoided, but rather to demonstrate a novel technique for making the disparity information more reliable. Indeed, in future work we envision developing a hybrid approach that augments the depth information with monocular cues such as intensity edges.

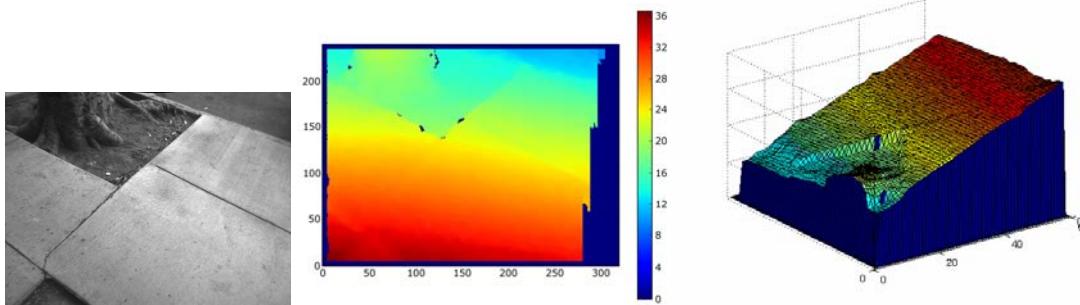


Figure 2 Disparity map of a street scene (right image of the scene shown in Fig. 2a), with the stereo camera pointed towards a sidewalk curb. In Fig. 2b the disparity map is rendered in 2-D in a false color scale: dark blue indicates no disparity estimate, and disparities increase from blue to red. In Fig. 2c the disparity map is rendered as a 3-D plot (with the same colors as in Fig. 2a and with height proportional to disparity, and rotated for ease of viewing). Notice that the overall shape of the disparity map is planar, corresponding to the ground plane, which is visible despite the high level of noise.

The main steps of our algorithm are as follows. The first – and most important – step is to apply a *median filter* (Forsyth and Ponce 2002) to the raw disparity map: Fig.'s 2 and 3 show the disparity map before and after filtering, respectively. This filter operates as follows: at each pixel, the median value of the disparities within a square neighborhood centered about the pixel (we chose a neighborhood size of 21 x 21 to apply to our 240 x 320 disparity images) is computed, and the filtered value at that pixel is then set to the median. An important property of the median filter is that it smooths out small spikes in the disparity map (visible as blue dots in Fig. 2b), while preserving important structures in the disparity map – especially step-like edges that signal discontinuities such as the transition from a sidewalk to the street.

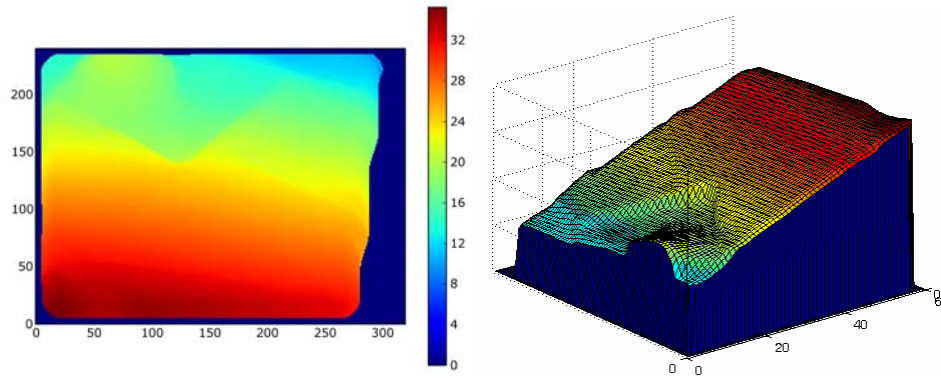


Figure 3 Disparity map from Fig. 2, after smoothing by a median filter. Note that the noise is reduced significantly, but important discontinuities are preserved. Dark blue corresponds to points for which no disparity estimate is available.

The next step is to find *edges* in the filtered disparity map, i.e. discontinuities. These are determined by estimating the magnitude of the spatial gradient (i.e. the magnitude of the vector whose components are the partial derivatives of the disparity with respect to x and y). The edge map corresponding to the filter disparity map in Fig. 3 is shown in Fig. 4a. Notice that many of these edges correspond to depth discontinuities in the scene.

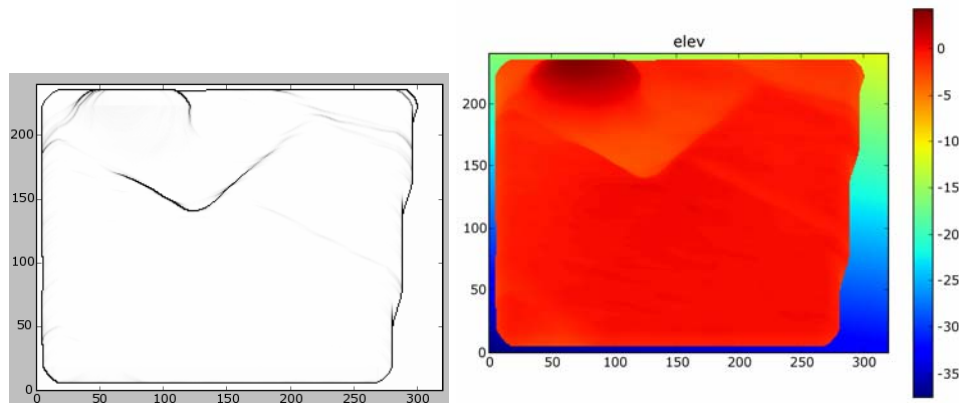


Figure 4 Edge map (a), with darkness proportional to the magnitude of the disparity gradient, and deviation from ground plane (b).

The ground plane is then determined by finding the plane that best fits as much of the disparity map as possible (allowing for the many pixels that do not lie on this plane). Once the ground plane has been determined, the difference between any pixel's disparity and the disparity of the ground plane at that pixel reflects how close the point is to the ground plane (in 3-D). This difference is shown in Fig. 4b.

The final stage of the algorithm is to apply a series of tests at each pixel to decide whether the pixel belongs to a significant depth discontinuity near the ground plane, such as a curb. First, the magnitude of the edge strength at that pixel must be above a minimum threshold; furthermore, to avoid spurious edges, the edge must not be a neighbour of a pixel without any disparity estimate (which would make the gradient difficult to estimate accurately). (Such spurious edges are common near the border of the image.) Second, the pixel must lie sufficiently close to the ground plane. Third, the pixel must be sufficiently close to the camera, both because it is hard to reliably estimate depth discontinuities at a distance, and because nearby depth discontinuities are more important than distant ones for our wheelchair application. Fourth, to eliminate spurious edges due to saturation of the image intensity (for instance, objects that are very bright, or at least much brighter than other objects in the image), which creates substantial noise in the disparity map, we eliminate from consideration any pixels whose image intensity has a saturated value (i.e. 255 for a standard 8-bit camera).

The results of this algorithm are shown in the next section.

3.4 Experimental Results

We demonstrate our algorithm on four images of sidewalks (Fig. 5). The results provide evidence that the algorithm is able to detect important depth discontinuities near the ground plane – those which are most likely to correspond to curbs and other nearby obstacles. Two of these images show curb cuts, places where a curb slopes downward to meet the ground (to provide a ramp allowing wheelchairs and other wheeled vehicles to move between the sidewalk and the street). In these images, the portions of the curb that are substantially elevated off the ground are detected as obstacle edges, while the portion that is low to the ground is not detected. In the other images, some of the borders between the sidewalk and the dirt patches (where bushes are planted) are detected, as are nearby cars.

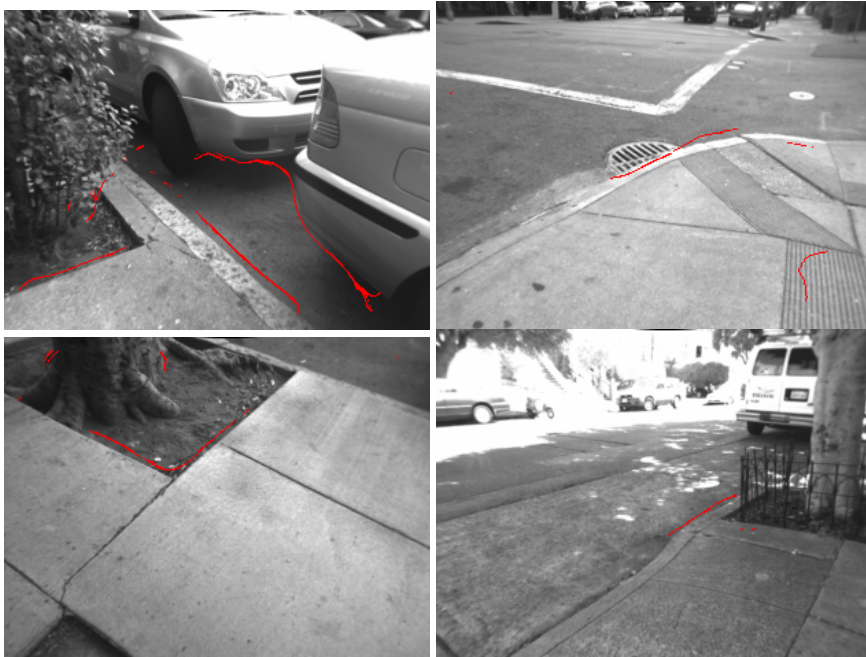


Figure 5 Experimental results: detected obstacle edges in red, superimposed on original image of scene. Result in lower left corresponds to scene in Fig.'s 2-4.

However, the algorithm in its current form has serious limitations. The result in the lower left of Fig. 5 completely misses the most distant edge of the curb (near the top of the image), because the discontinuity corresponding to that edge in the disparity map is so faint. Also, only a small number of edges from positive obstacles such as trees and bushes are detected. Finally, the alignment between the detected edges and the true edges is imprecise. All of these limitations arise from noise in the disparity maps; in order to circumvent these limitations and improve the algorithm's performance, other cues (such as monocular cues, e.g. intensity edges) should be incorporated.

4. Conclusions

We have devised a simple algorithm for finding obstacle edges using stereo vision. The key feature of the algorithm is that it smooths the depth information so as to reduce noise while preserving important information, enabling the algorithm to rely almost exclusively on depth information. Experimental results on sidewalk images demonstrate the feasibility of our approach, which in future work will be extended to include monocular image information such as intensity edges. Quantitative measures of performance, such as the fraction of curbs that are successfully detected, will need to be assessed quantitatively to guide the development of the algorithms and fully understand their strengths and weaknesses.

Ultimately we envision that computer vision algorithms will function as part of a comprehensive system for wheelchair navigation that integrates multiple sensor modalities (such as ultrasound and laser), since no one modality is reliable enough to use in isolation.

References

- Bellutta, P., R. Manduchi, L. Matthies, K. Owens and A. Rankin (2000). Terrain perception for DEMO III, *Intelligent Vehicle Symposium 2000*.
- Forsyth, D. and J. Ponce (2002). *Computer Vision: A Modern Approach*. Prentice Hall.
- Gill, J. (2000). Personal electronic mobility devices. In *Information for Professionals Working with Visually Disabled People*. <http://www.tiresias.org>
- Greenbaum, M.G., S. Fernandes and S.F. Wainapel (1998). Use of a motorized wheelchair in conjunction with a guide dog for the legally blind and physically disabled, *Arch Phys Med Rehabil.* vol. 79, no. 2, pp. 216-7.
- Labayrade, R., D. Aubert and J.P. Tarel (2002). Real time obstacle detection in stereo vision on non flat road geometry through 'V-disparity' representation, *Proceedings of IEEE Intelligent Vehicle Symposium. Versailles, France, June 18-20, 2002*.
- Lu, X. and R. Manduchi (2005). Detection and localization of curbs and stairways using stereo vision, *IEEE International Conference on Robotics and Automation (ICRA '05), Barcelona, April 2005*.
- Scharstein, D. and R. Szeliski (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, vol. 47, no. 1/2/3, pp. 7-42.

Acknowledgements: We would like to thank Roberto Manduchi for many helpful discussions. The authors were supported by the National Science Foundation (grant no. IIS0415310).