Real-time staircase detection from a wearable stereo system

Young Hoon Lee   Tung-Sing Leung   Gérard Medioni
Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, CA, USA
{lee126, tungsinl, medioni}@usc.edu

Abstract

We address the problem of staircase detection, in the context of a navigation aid for the visually impaired. The requirements for such a system are robustness to viewpoint, distance, scale, real-time operation, high detection rate and low false alarm rate. Our approach uses classifiers trained using Haar features and Adaboost learning. This first stage does detect staircases, but produces many false alarms. The false alarm rate is drastically reduced by using spatial context in the form of the estimated ground plane, and by enforcing temporal consistency. We have validated our approach on many real sequences under various weather conditions, and are presenting some of the quantitative results here.

1. Introduction

We have designed and built a real-time wearable, stereo-vision based, navigation system for the visually impaired which has proved to be well accepted by patients [8]. It consists of a head-mounted stereo camera and a vest interface with four tactile feedback effectors. The system uses a stereo camera as a data acquisition device and implements a realtime SLAM, an obstacle avoidance, and a path planning algorithm. Based on the algorithms mentioned above, an appropriate cue is generated and delivered at every frame to the tactile sensor array to alert the user to the presence of obstacles, and provide a blind user with guidance along the generated safe path. A main assumption we made here is that the ground is flat. In real scenarios, however, users have to navigate in complex environments which include stairs.

Stereo vision sensors are the primary choice for outdoor navigation. Our experiments show that, for staircases, a head mounted stereo system produces poor 3D, because of low texture, short baseline, and alignment of the stairs with the epipolar lines. Hence, we first consider staircase detection as a classical object recognition problem in 2D image analysis. For this reason, we propose to use Haar-like features with cascade classifier that is widely used in object detection because its efficiency and reliable performance. However, unlike human faces, texture, shapes, and the size of staircases have large variation among individual element that makes staircase detection a very difficult problem when using local features. Ground plane estimation from stereo vision, and smoothness of movement constraint are applied to achieve more robust and accurate detection performance at a reasonable additional computational cost. The main contribution of this work is a accurate real-time staircase detection framework. To confirm this, we have performed our experiments based on real datasets collected in different areas. Before describing the details of our algorithm in Section 3, related works are reviewed in the next section. Experimental details and results are provided in Section 4. Conclusion and future work are discussed in Section 5.

Figure 1. The wearable navigation system consists of a stereo camera with a vest-type feedback system.

2. Related work

Staircase detection for autonomous navigation system has been addressed in the literature [6, 7]. Works based on range scanners rely on accurate 3D models

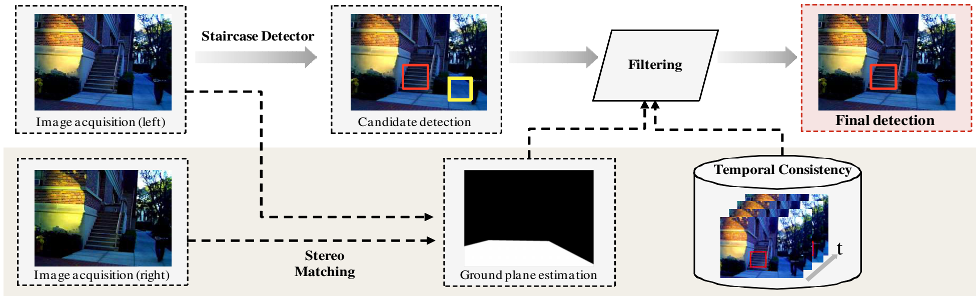


Figure 2. Algorithm flowchart is shown. Candidates are detected using Haar cascade classifiers and additional false detection(yellow boxes) removal processes are applied using the ground plane from a stereo vision system and temporal consistency.

from costly sensors [1] or inexpensive sensors, but they do not work outdoor [5]. Some researchers extended their curb detection frameworks to the staircase detection problem that make use disparity map and 3D point clouds from a stereo camera as shown in [6]. However, accuracy of 3D models from stereo algorithms severely decreases in low-texture scenes, which are quite common for staircase detection. To overcome shortcomings of previous approaches, researchers proposed algorithms that use a single camera to provide an answers to this problem. [9] uses the Gabor filter for distant detection and then finds edges that are equally spaced when approached close. Unfortunately, line extraction algorithms are brittle to lighting changes and shadow overlaid. Wang et al. proposed similar algorithm to our method in [10]. The main difference is that [10] aims to detect staircases observed only in a frontal view, thus may fail on many real world scenarios.

3. Overview of algorithm

The objective of our approach is to improve the detection accuracy and robustness. To achieve this goal, we collect 210 images containing 3 to 4 steps of a staircase taken from various conditions. 5250 positive 40×40 pixels images are prepared by creating 25 distorted or transformed images from every 210 original image. We also have 7000 negative samples that do not contain staircases in them. The Discrete AdaBoost training was performed until we obtain 18 layer cascade of classifiers with a false positive(FP) rate of 7.6×10^{-6} and true positive(TP) rate being 0.983 approximately. Note that the training is designed to have less training stages than classical applications. This is because higher detection rate is important for this staircase detection problem. However, a classifier with higher detection rate generally suffers from high FP rate which produces low precision. In the proposed detector, FPs are filtered by the estimated ground plane and temporal consistency, which enables the detector to achieve high precision.

3.1. Candidate Detection

For candidate detection, we used the cascade classifiers generated from the training stage and a 40×40 subwindow for scanning. Scale increase rate and the number of minimum neighbors are two parameters that we optimize empirically for the best performance. To deal with scale changes effectively, different sizes of scanning window must be applied. Test results indicate that increasing the window size by 30% works robustly at a faster speed for the staircase detection purpose. Haar detector generates multiple hits(detections) over a correct staircase region. On the other hands, false detections usually have a single isolated detection. In our detection algorithm, a group of detections which have more than 10 overlaps are chosen and their average rectangle region becomes the candidate. The n^{th} detected staircase in a frame is represented by its center(x_n, y_n), size(s_n), and confidence level(ρ_n). On detection, ρ_n of every candidate is initialized to 1.

3.2. False Alarm Removal

Additional false detection removal steps are applied to increase the accuracy of the detector. Recall that the main reason we approach staircase detection problem using 2D image is not because a stereo camera system is not available, but because the stereo system produces poor results specifically in the staircase detection task. However, stereo camera system can provide very meaningful information that helps decreasing the false detection rate. Furthermore, we add robustness to the algorithm by introducing smooth motion constraint.

3.3.1. Ground plane estimation Approximate ground plane assumption provides useful information and is widely used for many applications in SLAM systems that operate on the ground. We use a simple ground plane estimation method using a pair of stereo camera. For each stereo image pair, we compute a dense disparity map. We then use VDisparity algorithm [3] to locate the potential ground region within the left camera image. Prominent features are selected on these regions using [2]. If more than 8 feature points with valid disparity values are found on the largest connected region, we estimate the ground plane normal and center location by fitting the points with 3D plane using RANSAC. Staircases near the ground plane obtain higher weight because they are more critical to the blind users. For a faster computation, we divide an image into 16 by 16 disjoint subregions. Then confidence level of each candidate is determined by a modified sigmoid weight function proportional to the inverse of Euclidian distance between the ground plane and a detection. The confidence calculation by the ground plane estimation is performed only when the ground plane is extractable in a scene.

3.3.2. Temporal Consistency Appearance and motion smoothness over time is a general constraint in detection and motion tracking. Low-level association method based on affinities between two consecutive frames suggested by Huang [4] is modified and applied to corroborate temporal consistency. Confidence level from affinities between two staircases is defined by two parameters, the position and the size of a detection over two consecutive frames.

$$\rho_{n,new} = \rho_{n,old} \cdot \max(A_{pos}(S'_i, S_n) A_{size}(S'_i, S_n)) \tag{1}$$

$A_{pos}(S'_i, S_n)$ and $A_{size}(S'_i, S_n)$ represent affinities of position and size of two detected patches and ranges from 0 to 1. They are also a modified sigmoid weight functions of the inverse of Euclidian distance between S'_i and S_n. S'_i and S_n represents staircase detections from the previous frame and from the current frame, respectively. k is the number of detections from the previous frame. Detections whose ρ_n are greater than a threshold value θ_{fin} are considered to be the final detection of staircases.

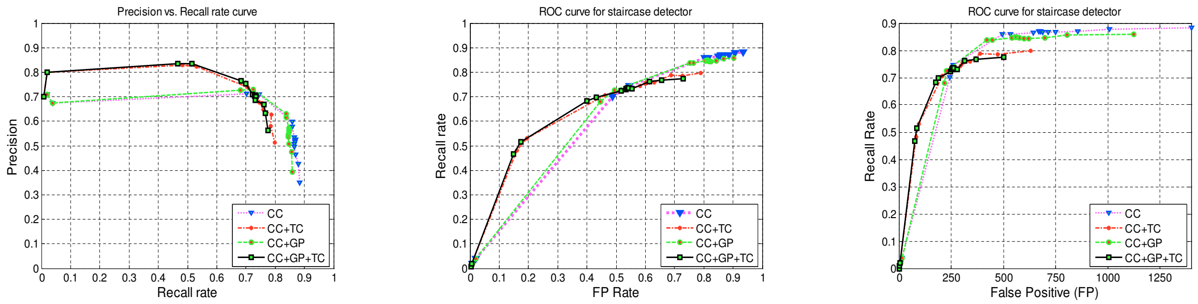
4. Experiments and Results

The following is a brief description of parameters used and H/W configuration of experiments. θ_{fin} are set to 0.8 for the best performance. The machine used has Intel(R) Xeon Quad Core @ 3.72GHz CPU, and Nvidia GeForce GTX 460 for GPU implementation. The experiments are designed to verify the accuracy of the staircase detector, and the effectiveness of false detection removal process. Processing speed is also measured to validate feasibility of the proposed algorithm as a real-time application. All experiments were performed on datasets consisting of 852 frames, 320x240 pixels, that contain staircases with various shape, texture, size, and lighting conditions. The datasets are collected from the wearable stereo camera shown in Figure 1. to contain staircases observed from many different viewpoints and distances to test the robustness of the detector. A ground truth is created by manually delineating the staircases area using polygons. Detections with more than 75% of their area overlapping with the ground truth data are defined as a TP.

Figure 3. Staircases detected using the proposed algorithm are shown. Note that the proposed algorithm can detect staircases from various view points and lighting conditions

4.1. Results

The stable performance is essential in our application, which implies the precision of the detector is important in our application. CC, TC, and GP in Figure 4. indicate cascade classifier, temporal consistency, and ground plane estimation, respectively. Figure 4.(a) shows that CC+GP+TC achieves the highest precision, 0.76 to 0.84, between 0 and 0.7 recall rate range. As provided in Figure 4.(b), the recall rates greater than 0.7 gives high FP rates but the detector does not operate stably. Hence the range of interest for our application is limited to the recall rate smaller than 0.7. Figure 4.(b) shows improvement in accuracy of detection by applying TC and GP. CC+TC, CC+GP, and CC+TC+GP has smaller FP rate in the range of interest. At recall rate of 0.7 with precision of 0.76, CC+GP+TC outperforms CC by 8% in terms of FP rate. Figure 4.(c) shows that TC, GP, and TC+GP reduces the number of FP very effectively. Experiments show that the number of FP reduces from 1400 to 501, 65% approximately. The improvement by applying the ground plane estimation is somewhat small, and certainly less than anticipated. This is due to current limitations in performance of the ground



(a) Precision vs. Recall rate curve (b) ROC curve for staircase detector (c) ROC curve for staircase detector

Figure 4. ROC curves and performance analysis of our staircase detector.

plane detector. When confidence is low, the result is discarded. Unfortunately, most previous published results on staircase detection did not provide quantitative results, hence a direct comparison to the previous approaches is not provided. The scanning process to find candidates is the biggest computational component of our detection algorithm. By applying the OpenCV GPU cascade classifier algorithm, the scanning process runs at 49.7 fps on 320x240 images, which provides speed-up factor of 18.4 (CPU: 2.7fps).

5. Conclusion

In this paper, we introduced a real-time staircase detection framework with high accuracy and low false alarm rate are essential. In order to meet the real-time requirement with accurate detection rate, the proposed algorithm uses 2D image based detection frameworks using classifiers built based on the AdaBoost learning algorithm that is widely used for object detection. For accuracy requirement, ground plane estimation step and temporal consistency validation step are added to lower the FP rate. Experimental results on real datasets confirm the accuracy and effectiveness of the method in various scenarios. The detection algorithm is designed to run either on CPU or GPU. As results suggest GPU implementation runs faster than 30fps in average which guarantees the real-time performance.

6. Acknowledgement

This research was made possible by a cooperative agreement that was awarded and administered by the U.S. Army Medical Research & Materiel Command (USAMRMC) and the Telemedicine & Advanced Technology Research Center (TATRC), at Fort Detrick, MD under Contract Number: W81XWH-10-2-0076

References

[1] G. Capi and H. Toda. A New Robotic System to Assist Visually Impaired People. In *IEEE International Symposium on Robot and Human Interactive Communication*, pages 259–263, 2011.

[2] A. Geiger, J. Ziegler, and C. Stiller. Stereoscan: Dense 3d reconstruction in real-time. In *IEEE Intelligent Vehicles Symposium*, pages 25–38, 2011.

[3] Z. Hu and K. Uchimura. U-V-disparity: an efficient algorithm for stereovision based scene analysis. In *Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, pages 48–54, 2005.

[4] C. Huang, B. Wu, and R. Nevatia. Robust object tracking by hierarchical association of detection responses. In *Proceedings of the 10th European Conference on Computer Vision: Part II*, pages 788–801, 2008.

[5] Y. H. Lee and G. Medioni. A RGB-D camera based navigation for the visually impaired. In *RSS 2011 RGB-D: Advanced Reasoning with Depth Camera Workshop*, pages 1–6, 2011.

[6] X. Lu and R. Manduchi. Detection and Localization of Curbs and Stairways Using Stereo Vision. In *IEEE International Conference on Robotics and Automation(ICRA)*, pages 4648–4654, 2005.

[7] V. Pradeep, G. Medioni, and J. Weiland. Piecewise Planar Modeling for Step Detection using Stereo Vision. In *Workshop on Computer Vision Applications for the Visually Impaired(CVAVI)*, pages 1–8, 2008.

[8] V. Pradeep, G. Medioni, and J. Weiland. Robot vision for the visually impaired. In *Computer Vision Applications for the Visually Impaired*, pages 15–22, 2010.

[9] S. Se. Vision-based detection of staircases. In *Proceedings of Asian conference on computer vision(ACCV)*, pages 1–6, 2000.

[10] S. Wang and H. Wang. 2D staircase detection using real adaboost. In *The Seventh International Conference on Information, Communications, and Signal Processing(ICICS)*, pages 1–6, 2009.