STAIR DETECTION AND TRACKING FROM EGOCENTRIC STEREO VISION

Tobias Schwarze

Karlsruhe Institute of Technology
Institute of Measurement and Control Systems
Engler-Bunte-Ring 21, 76131 Karlsruhe, Germany

Zhichao Zhong*

Centrum Wiskunde & Informatica
Science Park 123
1098 XG Amsterdam, Netherlands

ABSTRACT

In this work we present a real-time approach to capture the properties of staircases from a free moving, head mounted stereo-camera. A variety of systems can profit from such ability, examples range from robots in multi-floor exploration scenarios to wearable assistance systems for the visually impaired. We introduce a light-weight method to measure the individual steps and use this information to update a minimal stair model while approaching and traversing the stair. Results are evaluated on an in- and outdoor scenario and show competitive accuracy to state of the art approaches working on precise lidar-sensors.

Index Terms— Environment perception, Scene Reconstruction, Object tracking

1. INTRODUCTION

The detection of stairways offers valuable high-level knowledge to a variety of intelligent systems. Robotic systems navigating public space can use such knowledge to expand their range of operation to multiple floors. Systems assisting visually impaired people can provide guidance towards a stairway and provide information about e.g. the number of steps. Systems relying on a flat world assumption need to detect the traversal of a stairway passage to ensure proper functionality.

In the early works of [1] and [2], stair detection is based on 2D gray-scale images from monocular cameras. Both of them used Canny edge detection along with a Hough transform or Gabor filters [3] to extract characteristic edge features of stairs. To derive a control strategy for a humanoid robot, [4] assesses the dimension of the coming step from close view stereo vision, [5] traverses descending stairways with a mobile platform and uses the stair edge orientation to control the robot's heading. Here, the initial detection is done through texture analysis, [6] and [7] treat this as a learning problem and train classifiers based on Haar features to find stairs in images.

Detection combined with 3D reconstruction from stereo vision or 3D laser scanners has been proposed in other works.
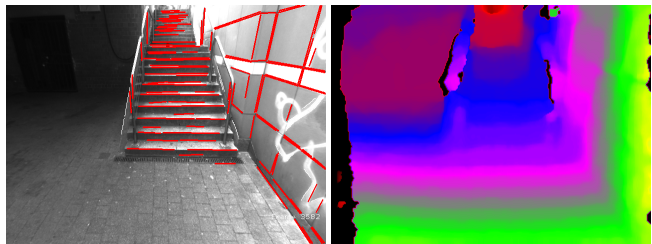

Fig. 1. The input data for stair detection and tracking: Edges detected in the gray-value camera image, and a dense, low-resolution disparity image. Additionally we estimate the ego-motion by visual odometry.

[8] proposes a curvature index to classify edges into concave and convex using a depth image and fits a minimal inclined plane model. [9] go further and fit a generative stair model into 3D edges registered in a global reference frame. As an alternative approach, often planar segmentation is used to measure individual steps by their tread plane. [10] compares two plane segmentation methods based on line segmentations and two-point random sampling in 3D lidar point clouds. [11] segments steps after normal estimation, [12] suggest a region growing strategy in 3D point clouds and connects thereby detected tread patches into a stair model.

In this work we consider vision systems that perceive the environment from an egocentric viewpoint. In our case this is a calibrated stereo camera mounted on a helmet worn on a person's head. While our work is similar to approaches detecting stair properties from a depth image, our aim is to track the model parameters while approaching and traversing the stairway. The most similar work is [9] who motivate this as a mapping problem. However, compared to sensors on robotic platforms, head mounted cameras are far less constrained in their motion and passive in that there is no way of controlling the camera heading. Furthermore there is usually no absolute orientation reference available in vision-only platforms, we tackled this problem in previous work [13] with an algorithm to keep track of the ground plane.

In this work we introduce a method that detects, models and tracks stairways based on such ground plane estimate.

---

*The author performed this work while with the Institute of Measurement and Control Systems
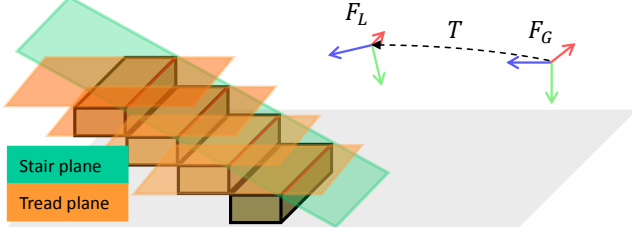
Fig. 2. Model of staircase defined by a stair plane (green) and parallel tread planes (orange) in a global frame F_G.

We extend previous approaches to deal with unconstrained camera motion and introduce a light-weight method to measure the stair properties including individual steps from low-resolution depth data. Instead of estimating the model from a single viewpoint we refine the model while crossing the stair, this increases the accuracy of parameter estimation and also enables us to determine the step count for long stairs that are not entirely visible from one single view.

2. SYSTEM

The basis of our stair estimator is a dense disparity image and the egopose T in a global reference frame F_G estimated incrementally by visual odometry. Using this input we fit a ground plane and align our stair estimation based on this plane. All information about a detected staircase are accumulated in F_G using T. Our staircase model consists of a stair plane that lies on the convex step edges (see Figure 2), as additional parameters we estimate the step height and number of steps, and – using the ground plane as reference – the slope respectively step depth.

2.1. Visual Odometry

The process of estimating the motion of a camera solely from image data is commonly referred to as visual odometry (VO). The general goal is to find the transformation in all six degrees of freedom that relates the camera poses of frame $k - 1$ and k. Various standalone implementations exist, an overview can be found in [14].

In this work we employ libViso2 [15]; the frame to frame transformation is provided as a translation vector $\mathbf{t}$ and a 3x3 rotation matrix R which can be written as an affine transformation $T = \begin{bmatrix} R & \mathbf{t} \\ 0^{\mathrm{T}} & 1 \end{bmatrix}$. Since this estimation is done incrementally it is inevitably subject to drift.

2.2. Iterative Least-Squares plane fitting

We represent a plane in $uv\delta$-space as

$$\alpha u_n + \beta v_n + \gamma - \delta = 0 \qquad (1)$$

using normalized image coordinates $u_n = \frac{u - c_u}{f}, v_n = \frac{v - c_v}{f}$ and corresponding disparity $\delta(u, v)$, camera focal length f and principal point (c_u, c_v).

An initial plane can be found using the RANSAC scheme by repeatedly sampling planes through 3 random points. A plane is evaluated by counting the support points with point-to-plane distance $|\alpha u_n + \beta v_n + \gamma - \delta|$ smaller than a margin ϵ around the plane. We optimize the initial solution using an iterative least-squares estimator

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} \mathbf{y} \qquad (2)$$

with H being the measurement matrix $[u_{n_{1..n}} \ v_{n_{1..n}} \ 1]$ of all $uv\delta$ points that satisfy the initial plane equation $|\alpha u_n + \beta v_n + \gamma - \delta| \leq \epsilon$ and $\mathbf{y}$ their corresponding disparity measurements. We apply this estimation a few iterations until no considerable update in the parameters remains.

The parameters then convert from $uv\delta$-space into a world space plane $\mathbf{p}_{F_L}$ with normal vector $\mathbf{n}$ and camera distance d through

$$\mathbf{p}_{F_L} = \begin{bmatrix} \mathbf{n}_x \\ \mathbf{n}_y \\ \mathbf{n}_z \\ d \end{bmatrix} = \frac{1}{\sqrt{\alpha^2 + \beta^2 + \gamma^2}} \begin{bmatrix} -\alpha \\ -\beta \\ -\gamma \\ Bf \end{bmatrix} \qquad (3)$$

with stereo baseline B and camera focal length f.

2.3. Plane tracking

Once an initial plane model $\mathbf{p}_{F_L}$ was fitted by RANSAC we transform and store it in the global coordinate frame as $\mathbf{p}_{F_G} = \left(T^{-1} \right)^{\mathrm{T}} \mathbf{p}_{F_L}$ with T_k being the current egopose. For the next video frame we estimate T_{k+1} and use the predicted plane $\mathbf{p}_{F_L}^- = T_{k+1}^{\mathrm{T}} \mathbf{p}_{F_G}$ as initialization for the least-squares plane optimization.

Using this plane detection and refitting scheme our system is able to keep track of the ground plane.

2.4. Stair detection and tracking

A characteristic visual feature of stairs are lines generated by the step edges. We follow the approach of Tardif [16] to detect these edge segments. After applying a Canny edge detector, edge crossings are removed and connected points extracted by flood fill region growing. These edge hypotheses are split into straight segments. The remaining segments longer than 30px are fitted with a line model and build the edge list ξ.

Our stair detector finds stair planes (compare Figure 2) by random sampling using points belonging to the edge segments ξ. A stair plane hypothesis is accepted or rejected by evaluating the number of support points and its intersection angle with the ground plane. Once a stair plane was detected
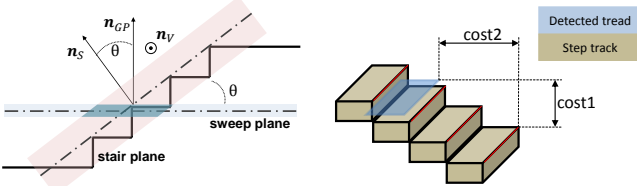
Fig. 3. Geometric relationships in the stair model. The high-lighted region of interest for sweep plane evaluation is given by intersecting the sweep plane and stair plane margin (left). The right figure shows cost calculation for data association.

we track it using the approach in Section 2.3 using only edge points belonging to the stair plane.

2.5. Estimation of step height and number of steps

To further detail the stair model we are interested in the number of steps and their height. Since staircases are often too large to be observed from a single viewpoint we complete the stair model over time while ascending the stair. We detect the tread of each step as a plane parallel to the ground plane (see Figure 2) and track them until the stair is passed.

For the detection and update of these tread planes we propose to use a sweep plane approach. We sweep the ground plane along its normal direction in discrete steps and count the support points for each position (Figure 3). The peaks of the originating profile correspond to the height over ground of each tread (see Figure 4).
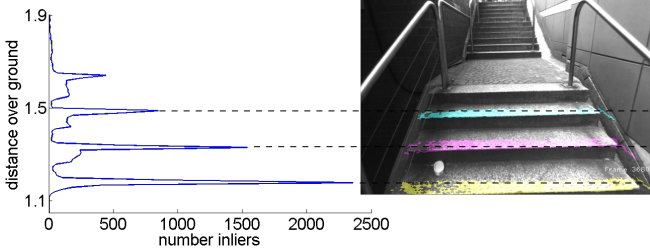

Fig. 4. Tread plane measurement and corresponding images points: the left graph shows the number of points supporting a plane swept along the ground plane normal vector w.r.t to the distance over ground. The right image overlays the support points for the detected peaks.

We track the measured tread planes and update their position in the global coordinate frame F_G. To assign the measurements to our global model we transform the existing tread tracks to the local frame F_L and apply Hungarian assignment [17]. The assignment cost is calculated as sum of the difference in height over ground and horizontal distance to the step edge (compare Figure 3 (right)). New tracks are initialized for measurements that could not be assigned because of too high assignment costs.

2.5.1. Groundplane correction

The sweep plane method requires knowledge of the ground plane normal vector, also when the ground is not visible during the stair ascend. In this situation we propagate the last valid ground plane measurement to the local frame using the estimated egopose, but since this estimation is subject to drift, also the ground plane becomes inaccurate on longer stair passages. Therefore we take indirect measurements of the ground plane normal $\mathbf{n}_{GP}$ using the stair plane normal $\mathbf{n}_S$, which is accurately measured during stair ascend. Both are related through a rotation around the intersection vector of their planes $\mathbf{n}_V$ by the stair slope θ so that $\mathbf{n}_{GP} = R(\mathbf{n}_V, \theta)\mathbf{n}_S$ (see Figure 3). The angle θ can be accurately measured during stair approach when stair and ground are visible and is kept fixed afterwards. The intersection vector is given by $\mathbf{n}_V = \mathbf{n}_S \times \mathbf{n}_{GP}$. An alternative way, which additionally compensates roll drift, is to compute $\mathbf{n}_V$ as the vanishing direction of all step edges.

We estimate this vanishing direction with the step edges ξ by minimizing the geometric cost proposed by Tardif [16]. The error for an edge ξ_j to a given vanishing direction $\mathbf{n}_V$ is defined as the orthogonal distance $D(\mathbf{n}_V, \xi_j)$ of one of the edge endpoints to the line connecting the vanishing point with the edge centroid. The vanishing direction is then updated by minimizing

$$\min_{\mathbf{n}_V} \sum_{\xi_j \in \xi} D(\mathbf{n}_V, \xi_j) \tag{4}$$

3. EXPERIMENTS AND EVALUATION

Our experimental setup consists of a calibrated stereo rig with a baseline of around 18 cm and wide-angle lenses of 3.5 mm focal length mounted on a helmet. We capture images with 640x480px at 30 fps and acquire the disparity images with OpenCV semi-global matching (SGBM) at half resolution (320x240px). Visual odometry and disparity estimation are done parallel with an output of around 20 fps on a i7 dual-core notebook. All experiments are done in real-time, i.e. frames are skipped when processing is slower than the capture rate of 30 fps.

The distance for initial detection of stair planes mainly depends on the lighting conditions, but succeeds from about 7m distance. After a stair plane has been initialized the sweep plane measurement is initiated and each step is tracked by its tread plane. In Figure 7 we show tracking results of an indoor and outdoor staircase. Videos of these sequences are available at http://t1p.de/81ku. Figure 5 shows the step heights after tracking the outdoor stair over the sequence of 120 frames. Despite that each step is tracked individually without enforcing a global step height, the estimated steps are equally spread. The averaged step height (Figure 6) during the sequence shows an average absolute error of 0.5cm.

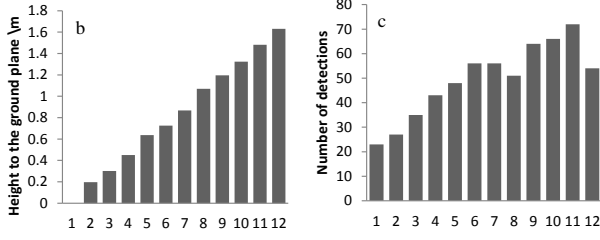A comparison of modelling errors between our method

Fig. 5. A stair model built after passing the outdoor stair in Figure 7 with 12 detected steps. Left graph shows the height of each step w.r.t. the ground plane. The number of detections per step is shown on the right. Since the lower steps move out of view as the camera is moving upstairs, they were assigned with less detections than the higher ones.
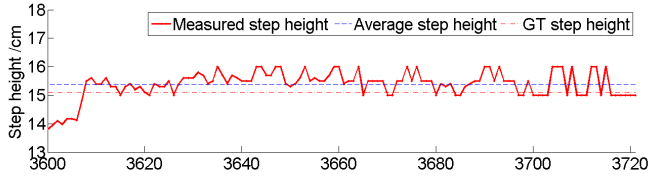

Fig. 6. Ground truth, observations and average value of average step height over 90 frame during which several steps are visible (the outdoor staircase corresponding to Figure 7).

Fig. 7. Tracking result of an indoor and outdoor staircase (170 and 120 frames): Visible step tracks are overlayed with different colors to indicate their identity. Cubes visualize position and size of steps.

Method	Angle (°) ERR± STD (%)	Step Height (cm) ERR± STD (%)
Ours (outdoor)	1.6 ± 0.5 (5.6)	0.5 ± 0.37 (3.3)
Ours (indoor)	0.6 ± 0.63 (1.7)	0.7 ± 0.77 (3.7)
Edge-based [9]	2.3 ± 1.9 (6.7)	1.7 ± 1.4 (8.9)
SLG [10]		0.42 ± 0.31 (6.0)
TPRS [10]		0.68 ± 0.54 (9.7)
Region-growing [12]		1.44 ± 0.59 (8.5)

Table 1. Error and standard deviation (ERR $\pm$ STD) as well as error percentage of stair slope and step height: The comparison is performed between our method, and reported accuracies of the 3d edge-based method in [9], scan-line grouping (SLG) and two-point random sampling (TPRS) from [10], and planar region-growing in [12].

and the reported results in [10], [9] and [12] are presented in Table 1. Our method is able to reach their accuracy despite the more noisy depth data originating from stereo-vision in our case as compared to laser scanners and RGB-D cameras. Building the model from a sequence rather than from a single viewpoint furthermore allows us to reliably count the number of steps.

If errors in our tracking process happen they are usually due to the fact that the head worn cameras are occasionally pointed too far up and the stair moves out of view temporarily. In general our approach is able to handle such situation by bridging the lacking measurements using the egomotion estimate to propagate the parameters.

The computation time of our approach during stair traversal amounts to 35-40ms after disparity estimation which allows to update the stair model with around 15fps. According to their reported runtimes it is slightly faster than the current approaches of [18] and [12].

4. CONCLUSION

We present an approach which enables robotic systems to perceive the parameters necessary to traverse a staircase using stereo vision. We treat this as a model tracking problem especially to handle the measurement during the actual crossing of the stairway. To this end we developed a method to estimate the stair inclination as well as the single steps expressed through tread planes. We make use of appearance based features in form of step edges in combination with stereo depth data to fit a stair plane that lies on the convex step edges. Afterwards we propose to use a sweep plane approach to efficiently measure the position of single treads relative to the ground plane.

A comparison with existing approaches using 3d data from laser or RGB-D scanners shows at least equal accuracy, despite that our depth data originating from stereo disparities is usually afflicted with stronger noise. The approach is lightweight enough to run with around 15fps on a notebook, including the calculation of dense disparity depth maps.

In future work we plan to evaluate the approach on descending stairs, where especially the initial detection is difficult since the ground plane blocks the view onto the stairs while approaching it.

5. REFERENCES

[1] Stephen Se and Michael Brady, "Vision-based detection of kerbs and steps," in *8th British Machine Vision Conference (BMVC)*, 2000, pp. 410–419.

[2] Nicholas Molton, Stephen Se, Michael Brady, David Lee, and Penny Probert, "Robotic sensing for the partially sighted," *Robotics and Autonomous Systems*, vol. 26, no. 2-3, pp. 185–201, 1999, Field and Service Robotics.

[3] D.C. Hernandez and Kang-Hyun Jo, "Stairway segmentation using gabor filter and vanishing point," in *International Conference on Mechatronics and Automation (ICMA)*, Aug 2011, pp. 1027–1032.

[4] J.-S. Gutmann, M. Fukuchi, and M. Fujita, "Stair climbing for humanoid robots using stereo vision," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2004, vol. 2, pp. 1407–1413 vol.2.

[5] J.A Hesch, G.L. Mariottini, and S.I Roumeliotis, "Descending-stair detection, approach, and traversal with an autonomous tracked vehicle," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct 2010, pp. 5525–5531.

[6] Sisong Wang and Han Wang, "2d staircase detection using real adaboost," in *7th International Conference on Information, Communications and Signal Processing*, Dec 2009, pp. 1–5.

[7] Young Hoon Lee, Tung-Sing Leung, and G. Medioni, "Real-time staircase detection from a wearable stereo system," in *21st International Conference on Pattern Recognition (ICPR)*, Nov 2012, pp. 3770–3773.

[8] Xiaoye Lu and Roberto Manduchi, "Detection and localization of curbs and stairways using stereo vision," in *ICRA'05*, 2005, pp. 4648–4654.

[9] J.A Delmerico, D. Baran, P. David, J. Ryde, and J.J. Corso, "Ascending stairway modeling from dense depth imagery for traversability analysis," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 2283–2290.

[10] S. Osswald, J. S. Gutmann, A. Hornung, and M. Bennewitz, "From 3d point clouds to climbing stairs: A comparison of plane segmentation approaches for humanoids," *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pp. 93, 98, 2011.

[11] Vivek Pradeep, Gerard Medioni, and James Weiland, "Piecewise planar modeling for step detection using stereo vision," in *Workshop on Computer Vision Applications for the Visually Impaired*, Marseille, France, 2008, James Coughlan and Roberto Manduchi.

[12] Alejandro Perez-Yus, Gonzalo Lopez-Nicolas, and Jose J. Guerrero, "Detection and modelling of staircases using a wearable depth sensor," in *ACVR*, 2014.

[13] T. Schwarze and M. Lauer, "Robust ground plane tracking in cluttered environments from egocentric stereo vision," in *IEEE International Conference on Robotics and Automation (ICRA)*, May 2015 (to appear).

[14] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *Robotics Automation Magazine, IEEE*, vol. 18, no. 4, pp. 80–92, Dec 2011.

[15] Andreas Geiger, Julius Ziegler, and Christoph Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[16] J.-P. Tardif, "Non-iterative approach for fast and accurate vanishing point detection," in *Computer Vision, 2009 IEEE 12th International Conference on*, Sept 2009, pp. 1250–1257.

[17] Harold W. Kuhn, "The hungarian method for the assignment problem," *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

[18] Titus Tang, Wen Lik Dennis Lui, and Wai. Li, "Plane-based detection of staircases using inverse depth," in *Australasian Conference on Robotics and Automation (ACRA)*, 2012.