MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

13—2

Reliable Detection of Obstacles on Staircases *

Reinhold Huber, Christoph Nowak, Bernhard Spatzek, David Schreiber
Advanced Computer Vision GmbH - ACV
Tech Gate Vienna, Donau-City-Straße 1
A-1220 Vienna, Austria
E-mail: reinhold.huber@acv.ac.at
Tel.: +43 (0)1 269 62 55 - 137, Fax: +43 (0)1 269 62 55 - 200

Abstract

A stereo vision system for detecting 3D-objects is presented. The method uses a pair of externally uncalibrated cameras and is able to detect objects located on a staircase. Detection is based on differencing stereo pair images, where a warping transform is used to overlay the left onto the right image, and vice versa. The 3D camera positions are obtained through model based pose estimation and disparity is used to obtain the warping transform. The demonstrated application is staircase surveillance.

1 Introduction

This paper describes a stereo vision system for staircase monitoring. In particular, the task is to detect objects on an escalator, which can be considered as a moving background, under real-world illumination conditions. The suggested solution consists of model based background reconstruction, perspective warping of one image to the other in a stereo setup, and the final detection of differences in an image pyramid.

Detection of anomalies on a floor plane using stereo techniques for navigation, vehicle guidance and obstacle avoidance was presented by several authors [5], [1], [7], [11]. The detection of staircases as an aid for the partially sighted was discussed in [8]. We describe a combination and extension of ideas presented in those papers with respect to detection of staircases as well as objects on staircases. Specifically, we employ a model based staircase pose estimator based on grouping of line features by the use of geometric invariants. The detection is based on measuring absolute pixel differences between unwarped and warped images. Image differences are represented in an image pyramid [2] and segmented into background (staircase) and foreground (obstacles) employing the algorithm suggested in [10]. The area of application is surveillance of staircases for security reasons, i.e. detection of occupancy or disposed items. Furthermore, our approach might be useful for applications involving staircase traversal, e.g. [3].

The main components of the presented system are the acquisition part, the offline (or calibration) part and the online (or detection) part. The most interesting subparts, i.e. geometric matching in the offline part and detection in the online part, will be discussed in some detail in the following.

*This work was carried out within the K plus Competence Center ADVANCED COMPUTER VISION and was funded from the K plus program. We thank Professor Walter Kropatsch (Pattern Recognition and Image Processing Group, Technical University of Vienna) for critical comments, which helped to improve the paper.

2 Establishing Correspondences between 2D-Data and 3D-Model

In model-based pose estimation parameters describing relative orientation and position, i.e. the extrinsic camera parameters, are found using correspondence between data and model. In our case, the data are 2D lines extracted from single images and the model is a 3D wireframe object. Nearly horizontal lines are derived from the image data using standard edge detection based on directional image gradient and Hough transform techniques. On the top of Figure 1 an empty staircase image and corresponding vertical edges (derived from negative gradients) are shown. The extracted lines are shown on the bottom left, on the bottom right the matched wireframe model is projected to the image plane. To establish correspondence between data and model lines for each image in the stereo pair, and furthermore, between the two stereo pairs, the following matching procedure is applied.

2.1 Grouping based on the Cross Ratio

The first step in matching is to identify possible correspondences between data and model lines. Under perspective projection, ratios of ratios of lines and ratios of ratios of angles, the so called cross ratios, are invariant [9]. We employ cross ratios to identify groups of four lines out of a larger set of possible lines. Such a group of 4 lines, which in our case is characterized by the cross ratio obtained for the intersection points with an approximately orthogonal line, serves as a matching candidate to the staircase pattern. The definition for the cross ratio for four points $p_1, \ldots, p_4$ on a line is given as:

$$Cr(p_1, \ldots, p_4) = [(x_3-x_1)(x_4-x_2)]/[(x_3-x_2)(x_4-x_1)],$$

where the $x_1, \ldots, x_4$ are the corresponding positions of each point on the line.

The following strategy for selecting data lines which are good candidates for correspondence to model lines was employed:

- Calculate the theoretical cross ratio, e.g. for four equally spaced points on a line this is $Cr_t = 4/3$.

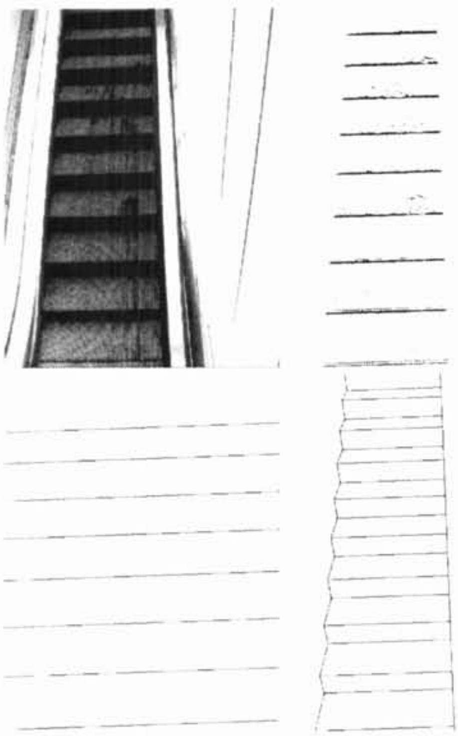

Figure 1: Empty staircase image, extracted frontal edges, fitted frontal lines and projected model.

- Detect a reasonable set L (of size N) of close to horizontal lines from the data.
- Calculate intersection points of those lines with a close to vertical line.
- Calculate all $M = \binom{N}{4}$ four-element subsets of lines $l_i \subset L, i = 1, \ldots, M$.
- Calculate all cross ratios c_i corresponding to sets l_i.
- Sort the l_i with respect to $|c_i - Cr_t|$ (in ascending order).

Only a portion of the sorted groups, corresponding to those of lower distance to Cr_t, is input to the pose estimation step, which is described below.

2.2 Estimation of Position and Orientation

Corresponding groups of lines are input to a procedure similar to RANSAC [4]. Grouping based on cross ratio delivers improved sampling for RANSAC and reduces the number of necessary iterations. The basic idea in RANSAC is, that RANSAC uses as small an initial data set as feasible and enlarges this set with consistent data when possible, see Figure 4. Figure 4 shows the enlargement of the set of lines in cases when a group of 4 lines delivers a sufficient pose estimate.

The required number of random selections n_s of samples with a size of s features is given by Fischler and Bolles as [4]:

$$n_s = log(1 - p_c)/log(1 - p_i^s),$$

where p_c is the probability, that at least one sample of $s = 4$ lines is free from outliers. The probability that any selected sample is an inlier is denoted by p_i. In our case, due to the improved sampling based on cross ratio, we can safely assume a high p_i, e.g. $p_i = 0.8$, and choosing $p_c = 0.99$, we obtain a number of necessary RANSAC iterations as low as $n_s = 9$.

Verification of the pose is based on the procedure devised by Lowe [6]. Lowe approaches the problem of derivation of object pose from a given set of known correspondences between 3D–model lines and 2D image lines by linearization of projection parameters and application of Newton's method. The result of the pose estimation step are two transformations from world to camera coordinate system, i.e. three translational and three rotational parameters for each camera.

3 Detection from Stereo Images

Detection involves detector calibration, i.e. derivation of disparity and derivation of the two–dimensional warping transform, and the detection itself, i.e. warping of one image to the other (e.g. the right image in Figure 2 to the left one), differencing of warped and unwarped images (see Figure 2 on the bottom left) and, finally, segmentation of the difference image (see Figure 2 on the bottom right) in order to obtain a decision.

Figure 2: Images and results for a person on a staircase: left and right images, absolute difference of right image warped to left image, and pyramid-based segmentation result.

3.1 Image Warping

The warping transform is found from the staircase model and the two world to camera coordinate system and projective transforms obtained by the pose estimation procedure mentioned above. A perspective warping transform provides us with two warping tables which contain the coordinate mapping for both coordinate directions in the image plane. The warping tables are calculated from disparity, which is accurately given due correspondence via the model, in straightforward fashion.

3.2 Detection of Obstacles

The main idea in detection is to warp one image, e.g. the left image to the right one, and perform some comparison. The objects on the staircase, which should be detected, have the property of being closer to the camera than the staircase on which they are placed. Therefore, objects in the image being warped appear at different positions than they appear in the unwarped image. On the other hand, disturbances like dirt or inscriptions on the staircase, appear on the same position in warped and unwarped image. Figure 2 showed initial images and some intermediate steps, i.e. the difference image generated by absolute pixel-wise differencing and the segmentation results applied to a difference image pyramid. Figure 3 shows image difference and segmentation pyramids.

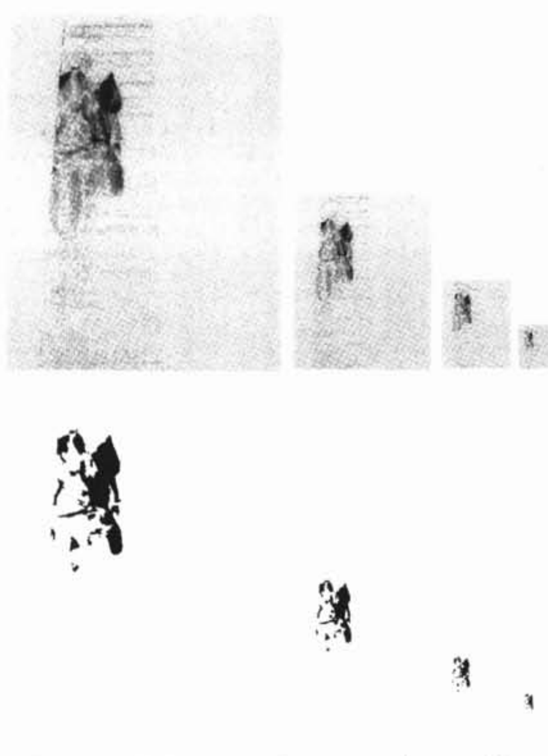

Figure 3: Difference and segmented pyramids.

4 Application and Results

We demonstrate our approach for the application of staircase surveillance, e.g. detection of persons and disposed items on a staircase of known geometry. Figure 5 shows snapshots of the staircase with and without obstacles together with the detection result, where

Reference	Detection result	
category	Obstacle	Empty
Obstacle	357	4
Empty	7	132

Table 1: Obstacle detection performance observed for 500 images.

d/d means correct detection of obstacle, **d/e** no detection of an obstacle, **e/e** means correct detection of an empty staircase, and **e/d** detection of an obstacle on an empty staircase.

Table 1 gives observed values for 500 images, 139 of which contained empty staircases and 361 show staircases occupied by various obstacles and persons. For this set of images, a rate of omission, i.e. an obstacle was not detected, in the order of magnitude of 1 percent and a rate of misdetection, i.e. an obstacle was detected in absence of an obstacle, in the order of magnitude of 5 percent was observed. Cylindrical objects down to a size of less than 15 centimeters in height were detected reliably.

5 Conclusion

An extension of stereo based obstacle detection procedures to regularly structured and non-flat background was presented. Grouping based on a cross ratio constraint improved RANSAC sampling. Pose estimation provides externally calibrated cameras, which simplify and accelerate stereo processing and the object detection task which is performed using a pyramid based segmentation procedure. A high reliability of the approach was found experimentally, i.e. a rate of omission of an obstacle in the order of magnitude of 1 percent, and a rate of false detection an obstacle in the order of magnitude of 5 percents.

References

[1] Peter Burt, Lambert Wixson, and Garbis Salgian. Electronically Directed 'Focal' Stereo. In *Proceedings of International Conference on Computer Vision*, pages 94–101, Cambridge, MA, USA, June 1995.

[2] Peter J. Burt, Tsai-Hong Hong, and Azriel Rosenfeld. Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation. *IEEE Transactions on Systems, Man and Cybernetics*, 11(12):802–809, December 1981.

[3] M. Fair and D.P. Miller. Automated Staircase Detection, Alignment & Traversal. In *Proceedings of International Conference on Robotics and Manufacturing*, pages 218–222, Cancun, Mexico, May 2001.

[4] M.A. Fischler and R.C. Bolles. Random Sample Concensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[5] M. Jenkin and A. Jepson. Detecting Floor Anomalies. In E. Hancock, editor, *Proceedings of British Machine Vision Conference*, pages 731–740, York, UK, 1994.

[6] David G. Lowe. Fitting Parameterized 3–D Models to Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):441–450, May 1991.

[7] Q.-T. Luong, J. Weber, D. Koller, and J. Malik. An Integrated Stereo-based Approach to Automatic Vehicle Guidance. In *Proceedings of the International Conference on Computer Vision*, pages 52–57, Cambridge, MA, USA, June 1995.

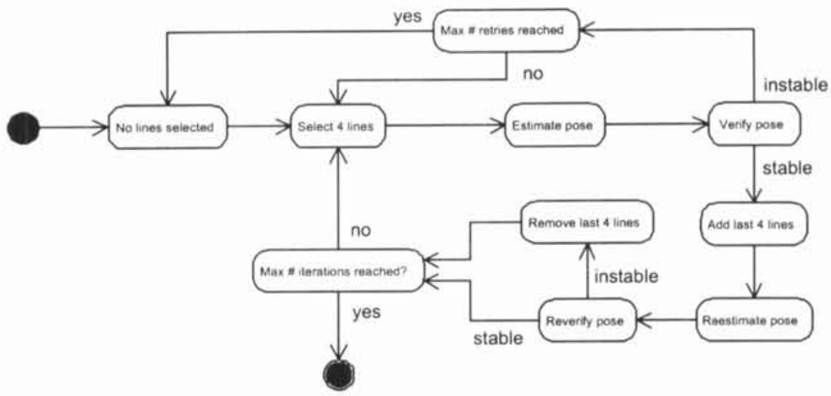

Figure 4: Dynamic behavior of the matching procedure.

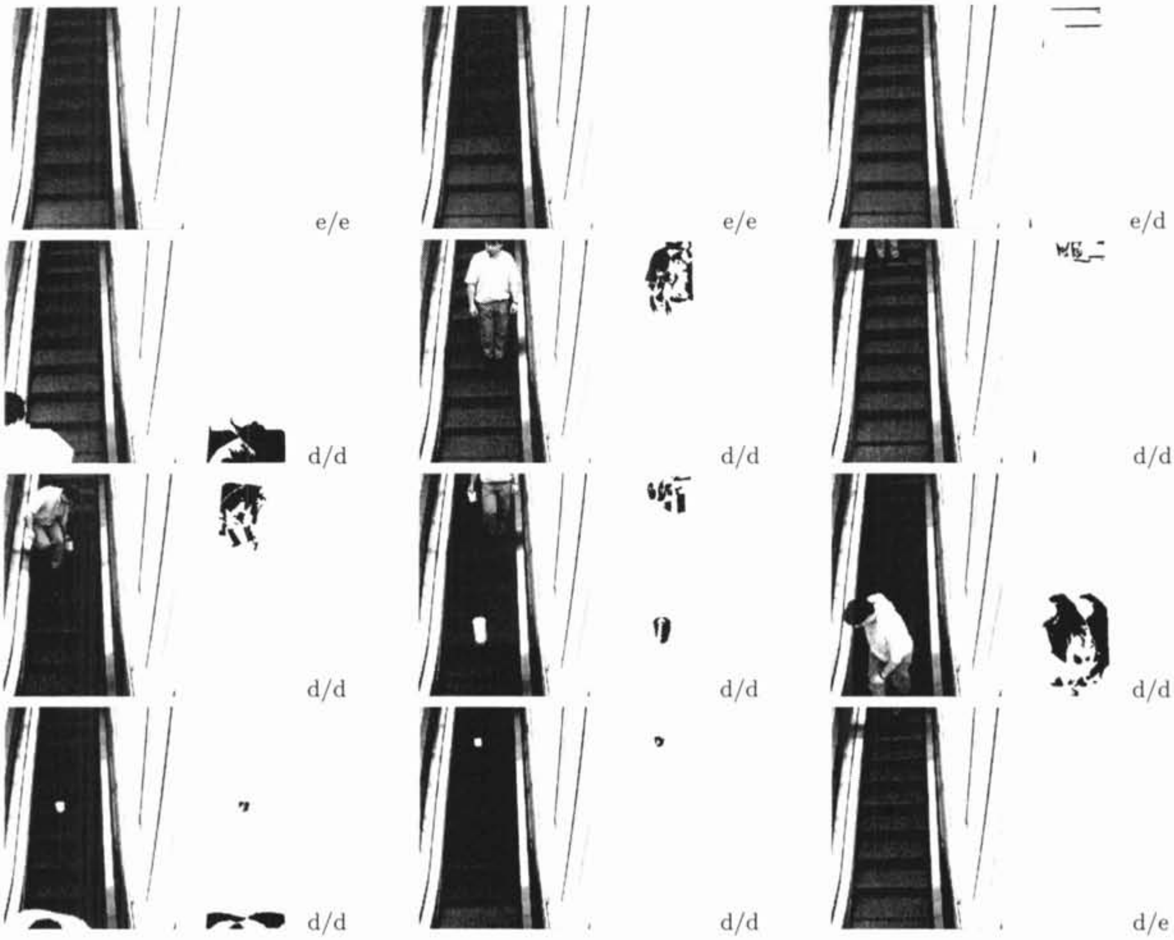

Figure 5: Images segmentation results: **d/d** means correct detection of obstacle, **d/e** missing detection, **e/e** means correct detection of an empty staircase, and **e/d** detection on an empty staircase

[8] N. Molton, S. Se, M. Brady, D. Lee, and P. Probert. Robotic Sensing for the Partially Sighted. *Robotics and Autonomous Systems*, 26(3):185–201, 1999.

[9] J. Mundy and A. Zisserman. Appendix: Projective Geometry for Machine Vision. In J. L. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 463–519. MIT Press, 1992.

[10] M. Spann and R. Wilson. A Quad–Tree Approach to Image Segmentation which Combines Statistical and Spatial Information. *Pattern Recognition*, 18(3/4):257–269, 1985.

[11] Todd Williamson and Charles Thorpe. Detection of Small Obstacles at Long Range using Multibaseline Stereo. In *Proceedings of International Conference on Intelligent Vehicles*, 1998.