University of Pennsylvania
ScholarlyCommons

Center for Human Modeling and Simulation | Department of Computer & Information Science

October 1998

Real-Time Stereo Processing, Obstacle Detection, and Terrain Estimation from Vehicle Mounted Stereo Cameras

Robert Mandelbaum
University of Pennsylvania

Luke McDowell
University of Pennsylvania

Luca Bogoni
University of Pennsylvania

Barry Reich
University of Pennsylvania

Mogens L. Hansen
University of Pennsylvania

Follow this and additional works at: http://repository.upenn.edu/hms

Recommended Citation

Mandelbaum, R., McDowell, L., Bogoni, L., Reich, B., & Hansen, M. L. (1998). Real-Time Stereo Processing, Obstacle Detection, and Terrain Estimation from Vehicle Mounted Stereo Cameras. Retrieved from http://repository.upenn.edu/hms/19

Copyright 1998 IEEE. Reprinted from *Proceedings of the 4th IEEE Workshop on the Applications of Computer Vision*, October 1998, pages 288-289.

This material is posted her with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/hms/19
For more information, please contact repository@pobox.upenn.edu.

Real-Time Stereo Processing, Obstacle Detection, and Terrain Estimation from Vehicle Mounted Stereo Cameras

Abstract
We use Sarnoff's next-generation video processor, the PVT-200, to demonstrate real-time algorithms for stereo processing, obstacle detection, and terrain estimation from stereo cameras mounted on a moving vehicle. Sarnoff's stereo processing and obstacle detection capabilities are currently being used in several Unmanned Ground Vehicle (UGV) programs, including *MDARS-E* and *DEMO III*. Sarnoff's terrain estimation capabilities are founded on a "model-based directed stereo" approach. We demonstrate ongoing collaborative research between Sarnoff and Universitat der Bundeswehr Miinchen, where we are studying vision processing for autonomous off-road navigation as part of the *AUTONAVprogram*.

Comments
Copyright 1998 IEEE. Reprinted from *Proceedings of the 4th IEEE Workshop on the Applications of Computer Vision*, October 1998, pages 288-289.

This material is posted her with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This conference paper is available at ScholarlyCommons: http://repository.upenn.edu/hms/19

Real-time stereo processing, obstacle detection, and terrain estimation from vehicle-mounted stereo cameras.

R. Mandelbaum L. McDowell L. Bogoni B. Reich M. Hansen
Vision Technologies Laboratory
Sarnoff Corporation, CN 5300, Princeton, NJ 08543
{rmandelbaum, lmcdowell, lbogoni, breich, mhansen}@sarnoff.com

Abstract

We use Sarnoff's next-generation video processor, the PVT-200, to demonstrate real-time algorithms for stereo processing, obstacle detection, and terrain estimation from stereo cameras mounted on a moving vehicle. Sarnoff's stereo processing and obstacle detection capabilities are currently being used in several Unmanned Ground Vehicle (UGV) programs, including MDARS-E *and* DEMO III. *Sarnoff's terrain estimation capabilities are founded on a "model-based directed stereo" approach. We demonstrate ongoing collaborative research between Sarnoff and Universität der Bundeswehr München, where we are studying vision processing for autonomous off-road navigation as part of the* AUTONAV *program.*

1 Introduction

Sarnoff has developed a family of video processors, called Vision Front End (VFE) systems, specifically to perform functions required of autonomous vehicle mobility. The underlying strategy behind the VFE system design is to perform front-end vision functions in hardware, thus freeing up general purpose processors to do the remainder of the video data analysis and high-level functions. The VFE 200 is the latest of these embedded platforms capable of video rate processing. It is a modular, reconfigurable, extensible system optimized for real-time pyramid-based video processing. As such, it is ideally suited for autonomous mobility image processing tasks.

The core image-processing capabilities demonstrated here are real-time *registration* and *stereo processing*. In addition, we demonstrate two examples of autonomous mobility tasks which rely on these capabilities: *obstacle detection* and *terrain estimation*.

2 Horopter-based stereo

Stereo processing uses two images taken simultaneously from a pair of cameras whose optical axes are parallel. For each pixel in the left image, a corresponding pixel is searched for in the right image. The disparity in location of the pixel in the two images is then used to triangulate the distance of the underlying point in space from the camera pair [2, 3]. An array of disparities, each value corresponding to a pixel in the image pair, is known as a *disparity map*. For typical scenes encountered during ground-based navigation, a substantial portion of the disparity map has a regular structure, corresponding to the points which lie on the ground. Obstacles which are above or below the ground plane cause deviations in the disparity map from this regular structure. Hence, an effective approach to both obstacle detection and terrain estimation is to **(a)** find the dominant regular structure in the disparity map, and **(b)** search for regions which deviate from this model. We refer to the dominant regular structure as the "horopter". In this demonstration, we use a horopter which corresponds to a flat ground plane whose normal is tilted with respect to the image plane of the stereo pair (see Figure 1). The use of a tilted horopter has two advantages: **(1)** The search for matches can be lim-

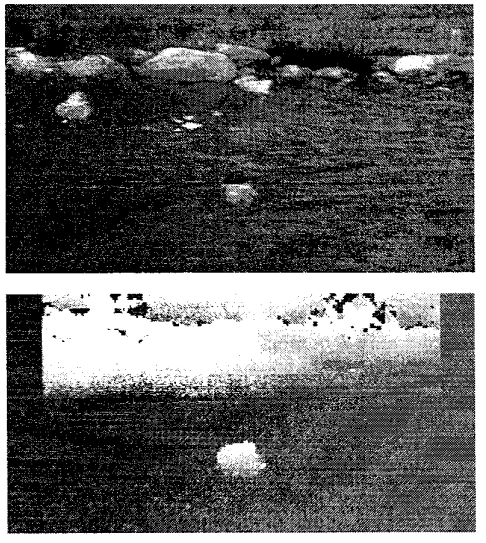

Figure 1: Example disparity map using the tilted horopter approach. The top image shows one of the two stereo images. The bottom image shows the resultant disparity map. The ground plane appears a uniform gray, regardless of the distance from the camera pair. Obstacles which are raised above the ground plane are shown with higher disparity (brighter), while negative obstacles have lower disparity than the ground plane, and would appear darker. Note the ease with which the obstacle in the foreground can be discriminated.

0-8186-8606-5/98 $10.00 © 1998 IEEE

ited to a band around the ground plane. This reduces dramatically both the computational complexity and the range over which errors and mismatches can be made. **(2)** The disparity over an entire support window is constant, which improves the sharpness of the correlation functions [2, 6].

It can be shown that for a stereo pair, the image transformation corresponding to a flat ground plane is an *affine* transform [2]. Hence, Sarnoff's approach to real-time horopter-based stereo consists of the following three major steps: **(1)** Estimate the parameters of the optimal (in a least squares sense) affine transformation mapping the right image onto the left; **(2)** Warp the right image according to this affine; and **(3)** Perform multi-resolution area-based stereo processing between the left image and the warped right image. The VFE 200 architecture enables all three steps to be performed at full frame rate (30 frames per second). The stereo demonstrated here searches for matches over a disparity range of 32 pixels on 320x240 image pairs. In addition, disparities are computed using multiple support region sizes. By adaptively combining the results into a single disparity map, the algorithm accounts for variability in scene structure and image texture.

3 Obstacle detection

Obstacle detection is commonly performed as post-processing of a disparity map generated by stereo matching [4, 5]. Horopter-based stereo processing greatly facilitates this obstacle detection since the disparity map consists of *residual* parallax relative to a flat ground plane in front of the vehicle, rather than *absolute* parallax relative to the image plane. Objects lying above the ground correspond to pixels with positive disparity, while depressions generate negative disparities. The obstacle detection algorithm demonstrated here consists of clustering pixels in the disparity map with similar non-zero disparity values. The algorithm runs at 30 frames per second and provides auxiliary information about obstacles, such as footprint size, average height, orientation, and roughness of obstacle surface.

4 Terrain estimation

A second essential image processing task for autonomous mobility is the estimation of the terrain shape. This information is used to adapt vehicle velocity according to terrain roughness, and to feed terrain navigability information to a path-planner.

Sarnoff's research in terrain estimation is performed as part an ongoing collaborative effort between Sarnoff and Universität der Bundeswehr München (UBM). This work focuses on vision processing for autonomous off-road navigation. UBM's approach to vehicle control relies on a model of the real world in space (3D) and time (1D), referred to as the "4D approach" [1]. The model is continuously refined and updated using recursive estimation techniques that use visual and other sensor measurements as the input. Figure 2 illustrates this dynamic control loop.

Sarnoff and UBM are currently working on a combination of "saccadic vision" and "time-integrated stereo processing". In this work, auxiliary information from odometry and inertial sensors is used as input to a control loop which keeps the stereo pair fixated on regions of interest in the scene such as potential obstacles. This affords multiple stereo views of the same object as the vehicle moves forward. The resultant sequence of disparity maps is then integrated over time. This "model-based directed time-integrated stereo" constitutes the next generation in stereo processing for ground vehicles.

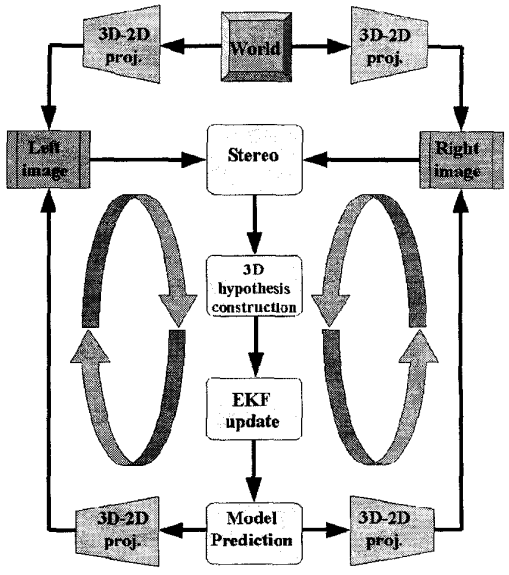

Figure 2: Dynamic control loop for the "4D" approach. The input from the auxiliary sensors is combined with the model to predict areas of particular interest in the stereo images. The stereo process is then directed towards these areas, and disparities are measured. These disparities are then converted back into 3D world coordinates and used to update the model. In this way, the model converges towards an estimate of the terrain in front of the vehicle.

References

[1] S. Baten, R. Mandelbaum, M. Lützeler, P. Burt, and E. D. Dickmanns. Area-based stereo image processing within the 4D architecture for autonomous off-road navigation. *IEEE Expert*, 1998. (to be published).

[2] P. Burt, L. Wixson, and G. Salgian. Electronically directed 'focal' stereo. *Proceedings of the Fifth International Conference on Computer Vision*, pages 94–101, June 1995.

[3] U. R. Dhond and J. K. Aggarwal. Structure from stereo — a review. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(16), November/December 1989.

[4] D. Gennery. Object detection and measurement using stereo vision. In *Proceedings of the 1980 Image Understanding Workshop*, pages 161–167, April 1980.

[5] H. Moravec. Towards automatic visual obstacle avoidance. *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, page 584, 1977.

[6] T. Williamson and C. Thorpe. A specialized multibaseline stereo technique for obstacle detection. In *Proceedings of Computer Vision and Pattern Recognition (CVPR) '98*, June 1998.