VIDEO OBJECT SEGMENTATION USING STEREO-DERIVED DEPTH MAPS[1])

Danijela Markovic and Margrit Gelautz
Interactive Media Systems Group
Institute for Software Technology and Interactive Systems
Vienna University of Technology
Favoritenstrasse 9-11/188/2, A-1040 Vienna, Austria
Email: markovic@ims.tuwien.ac.at

Abstract

In this paper, we explore possibilities to improve existing video segmentation algorithms by utilizing a stereo-derived depth map as additional source of information. The goal is to make the segmentation results more robust in order to reduce the need for user interaction in typical video editing and compositing tasks. In tests with synthetic and real video scenes, we show that a combination of the original and stereo-derived edges in conjunction with active contour models (snakes) can improve the quality of the segmentation results.

1 Introduction

Video object segmentation is an important step in many computer vision and multimedia tasks, including video editing and compositing, and the combination of real with synthetic video content. A review of image and video segmentation algorithms that have been proposed for multimedia applications can be found in [9]. The authors group the various methods into transition-based (e.g., using edges) and homogeneity-based (e.g., region growing) approaches. Only few published studies [3] have investigated the potential of stereo-derived depth maps to aid the segmentation process, which is the focus of our investigation. For example, [11] proposed the use of an MRF/GRF framework for incorporating depth information and [5] utilizes object contours in a hierarchical matching approach. Contrary to our experiments, the use of active contour models (snakes) is not considered in those studies.

In the context of an ongoing project, we use multiple video cameras to capture a dynamic scene from different points of view for subsequent 3D reconstruction. This approach is motivated by the

[1] This work was supported by the Austrian Science Fund (FWF) under project P15663.

growing availability of inexpensive video cameras and new generations of more powerful computers, which provide the basis for processing the multiple video streams in reasonable time. Examples of research laboratories that operate a large amount (e.g., 64) of synchronized video cameras in a multi-view configuration are the Robotics Institute at Carnegie Mellon University [6] and the recently established Keck Laboratory at the University of Maryland [2]. Whereas most traditional (correlation-based) matching algorithms tend to blur edges, some recent research [1] has specifically addressed the detection of depth discontinuities in stereo image pairs, which is of particular interest for extracting the video object from the background.

In our experiment, we first apply available segmentation algorithms based on edges and active contour models to the original image and stereo-derived disparity map, in order to explore possible synergisms between the results. We then demonstrate how a combination of original and stereo-derived edges can improve the quality of the segmentation results.

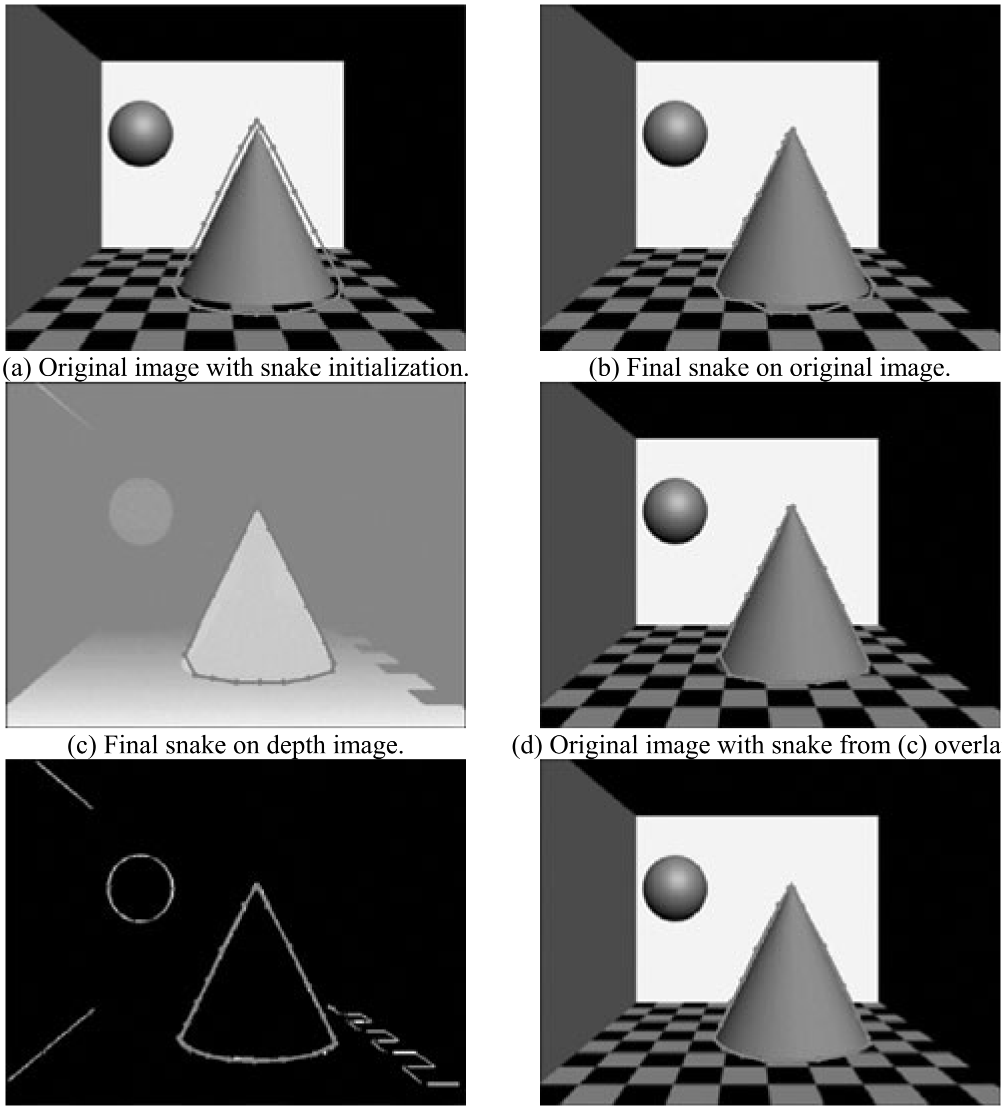
2 Test Data

We performed tests on both real and synthetic image sequences. The artificial scene shown in figures 1 and 2 was generated by using 3D Studio MAX. The frame size is 640 x 480 pixels. For the experiments with real data (figures 3 to 5), we employed a stereo configuration consisting of two Dragonfly IEEE-1394 video cameras as delivered by Pointgrey [8]. The camera set-up was calibrated using the calibration routines provided by Intel's Open Source Computer Vision (OpenCV) library [4]. For further processing, we converted the original 24 bit color images into 8 bit gray value images and transformed the stereo image pairs into epipolar geometry. An example of such a pre-processed pair of stereo video frames (size 421 x 480 pixels) is shown in figures 3 (a) and (b).

3 Algorithm, Tests, and Results

Figure 1 (a) shows a synthetic image with the initialization for a snake algorithm overlaid. In our tests, we employed both a classical snake implementation which follows the original work by [7] and a more recent snake implementation based on the *gradient vector flow (GVF)* method introduced by [10]. The illustrations in figure 1 were generated using the classical snake implementation. The final snake position, i.e. the contour found by the snake after iteration, is visible in Figure 1 (b). Note that the chessboard pattern on the ground produces errors in the snake result. The original image from (a) along with a corresponding stereo partner (not shown here) was used to compute the disparity map shown in (c). We employed the matching algorithm proposed by

[1], which has shown good results in preserving the edges. The final snake obtained from the disparity image is overlaid in (c). The deviations between the computed snake and the actual object contour can be recognized more clearly in subfigure (d).



(a) Original image with snake initialization.

(b) Final snake on original image.

(c) Final snake on depth image.

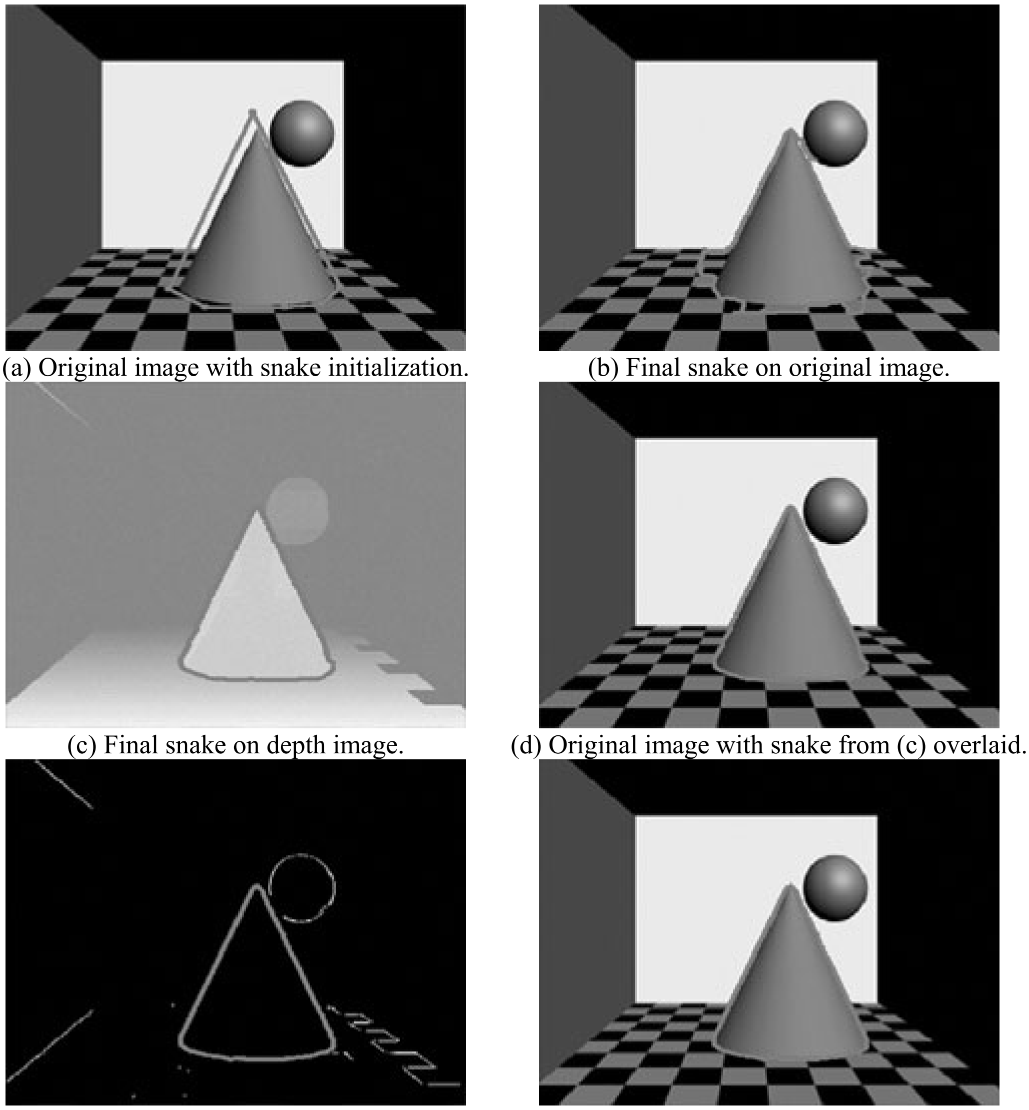
(d) Original image with snake from (c) overlaid.

(e) Final snake on edge combination image.

(f) Original image with snake from (e) overlaid.

Figure 1 Experimental results from synthetic test frame 1 with traditional snake.

After analyzing the errors in (b) and (d), we implemented an algorithm in which we first apply a Canny edge detector to both the original image and corresponding depth map and then compute a so-called *edge combination image*, which contains only those edges of the original image that are also present in the disparity image. The implementation had to account for minor deviations between corresponding edges in the original and disparity image which were caused by non-perfect edge localization of the stereo matcher. The computed edge combination image along with the corresponding snake result can be seen in subfigure (e). The good fit of the edge-derived snake is also apparent in (f). A comparison of subfigures (b), (d), and (f) confirms the improvement achieved by incorporating the stereo-derived edges into the segmentation process. Similar results were obtained from the same test images when using the GVF snake.



(a) Original image with snake initialization.

(b) Final snake on original image.

(c) Final snake on depth image.

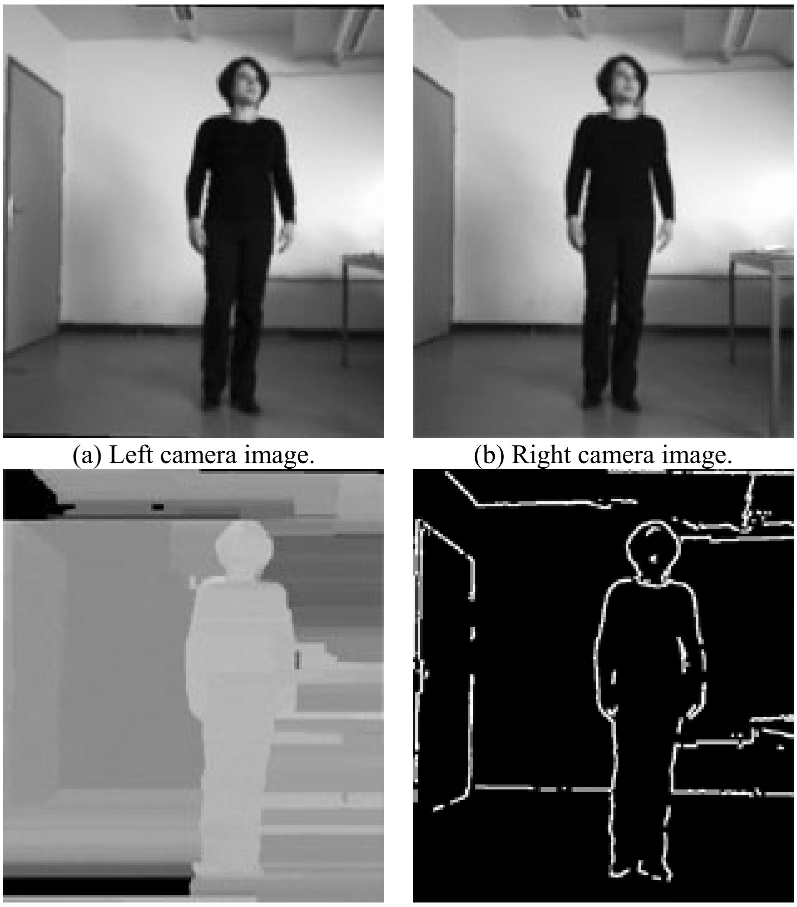
(d) Original image with snake from (c) overlaid.

(e) Final snake on edge combination image.

(f) Original image with snake from (e) overlaid.

Figure 2 Experimental results from synthetic test frame 2 with GVF snake.

In the test scene in figure 2, the sphere has moved closer to the cone. The proximity of the two objects causes perturbations in the snake on the original image (b), which are not present in the depth-derived results in (d) and the edge-combination result in (f). The most prominent errors in (b), however, are caused by the edges of the chessboard pattern on the ground. This effect is more pronounced in figure 2 due to the use of the GVF method than in the result obtained from the classical snake in figure 1. Again, the edge combination image (f) delivered the best results, although in this case the improvement over (d) is only minor, due to the high quality of the stereo-derived depth map in (c).

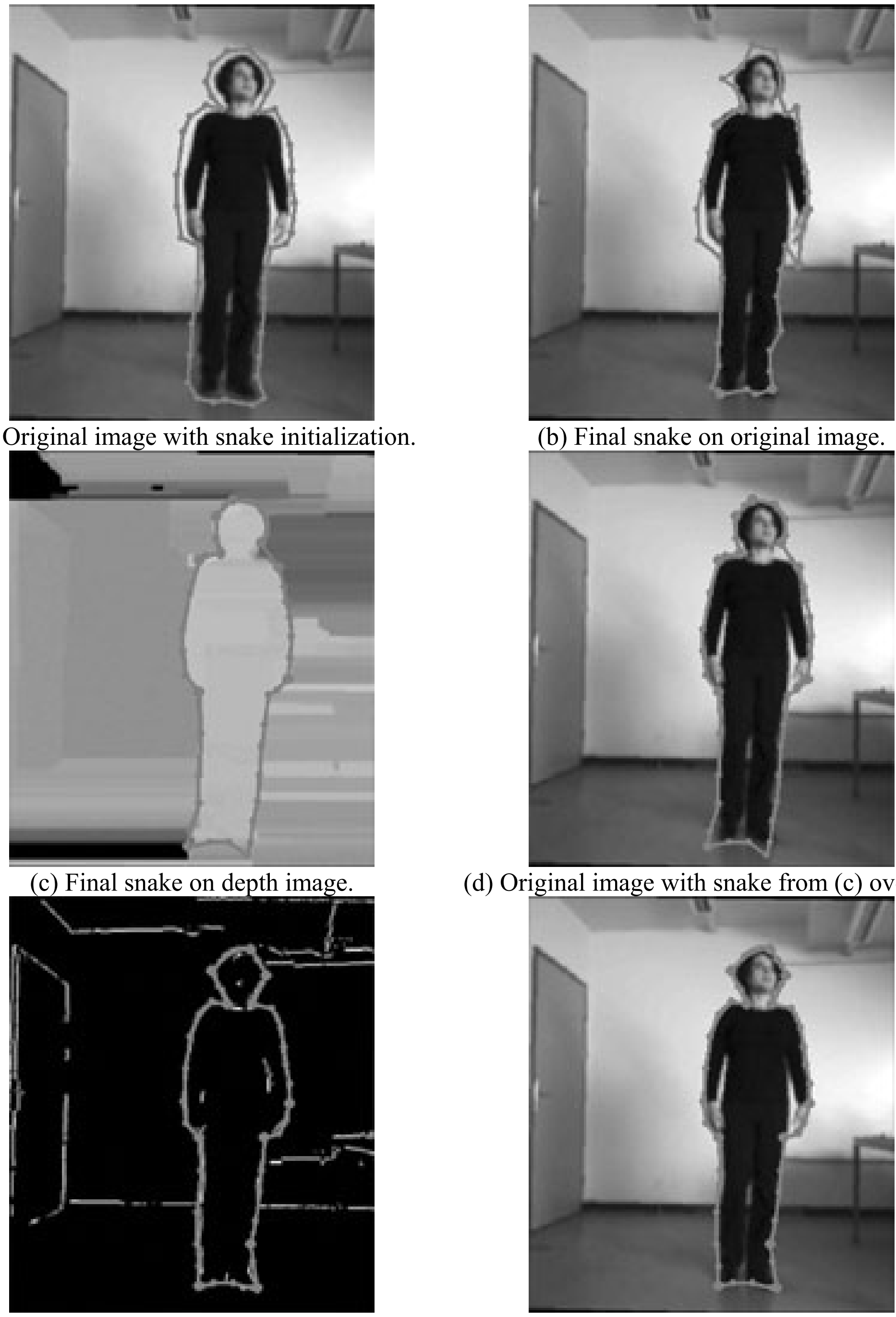


(a) Left camera image. (b) Right camera image.

(c) Depth image. (d) Edge combination image.

Figure 3 Stereo video frames with computed depth map and edge combination result.

Figures 3, 4, and 5 demonstrate the results obtained from experiments on real video sequences. Note that despite obvious errors in the stereo-derived depth map in figure 3 (c), the computed combination edges in figure 3 (d) show good correspondence with the outlines of the objects in the 3D scene. In figure 4, one can recognize how the traditional snake on the original image in subfigure (b) is influenced by intensity edges outside the object of interest. These errors are suppressed in the edge combination result in figure 4 (f).



(a) Original image with snake initialization.

(b) Final snake on original image.

(c) Final snake on depth image.

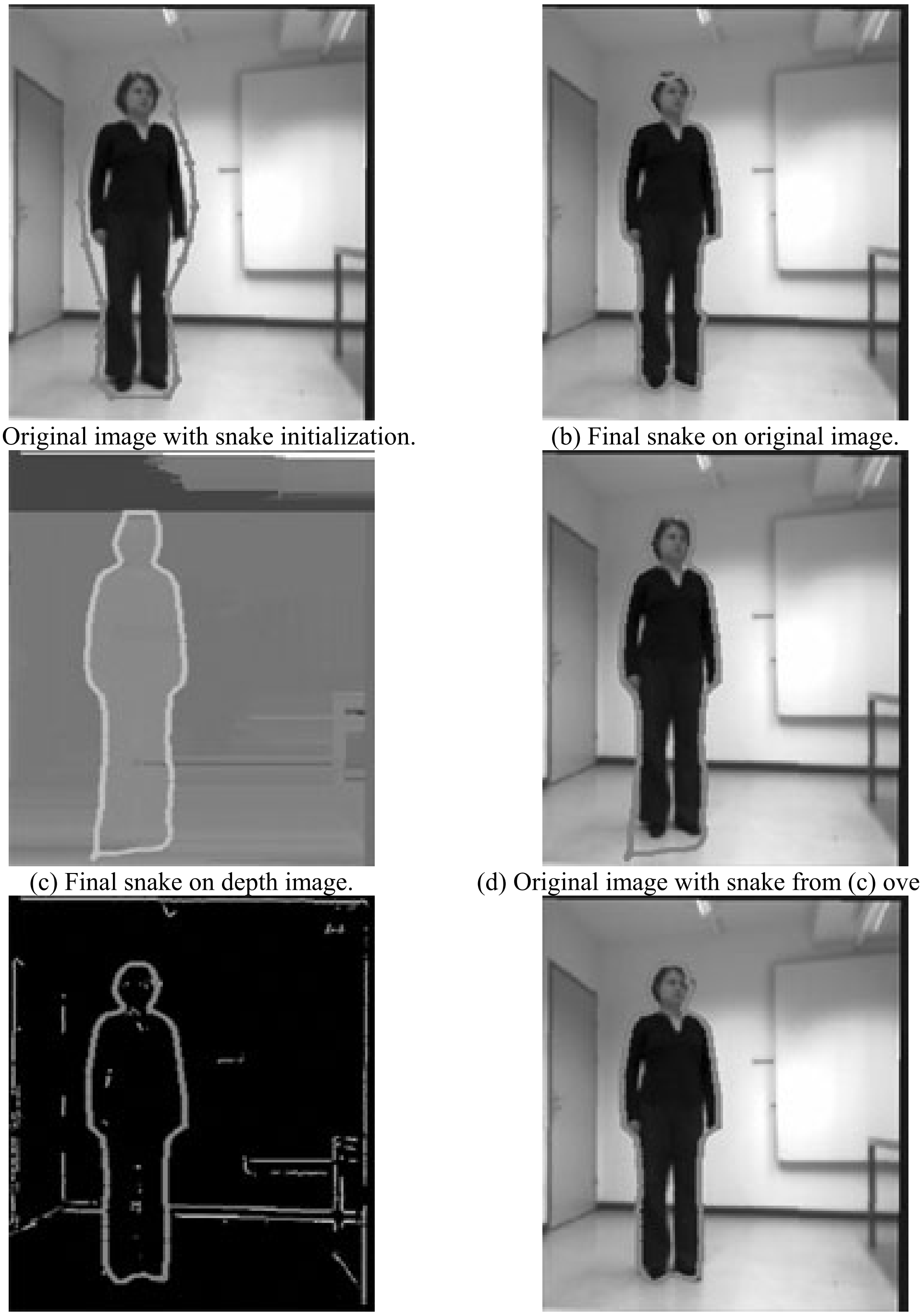
(d) Original image with snake from (c) overlaid.

(e) Final snake on edge combination image.

(f) Original image with snake from (e) overlaid.

Figure 4 Experimental results obtained from the real video frames in figure 3 with traditional snake.

Further experiments are illustrated in figure 5. The GVF snake on the original intensity image (b) shows errors in the region of the head and knees of the test person. The most prominent errors in the depth-derived result (d) occur around the feet. The errors are clearly reduced in the edge combination result in (f).



(a) Original image with snake initialization.

(b) Final snake on original image.

(c) Final snake on depth image.

(d) Original image with snake from (c) overlaid.

(e) Final snake on edge combination image.

(f) Original image with snake from (e) overlaid.

Figure 5 Results obtained from another test frame with GVF snake.

4 Summary and Outlook

In experiments on synthetic and real images, we have demonstrated how stereo-derived depth maps can be utilized to improve the segmentation results obtained from an active contours algorithm. We suggested the computation of a so-called "edge combination image" which combines edges from the intensity image with the location of discontinuities in the stereo-derived depth map.

As a next step, we plan to perform experiments on a larger variety of test scenes in order to study in more detail the advantages and possible limitations of the method (e.g., in areas where errors in the depth map coincide with intensity edges.) Furthermore, we will investigate the user interaction required for the snake initialization as well as the possible exploitation of inter-frame redundancies for a more efficient implementation.

References

[1] Birchfield, S. and C. Tomasi, Depth Discontinuities by Pixel-to-Pixel Stereo, International Journal of Computer Vision, vol. 35, no. 3, pp. 269-293, 1999.

[2] Cutler, R., R. Duraiswami, J. Qian, and L. Davis, Design and Implementation of the University of Maryland Keck Laboratory for the Analysis of Visual Movement, Technical Report CS-TR-4329/CS-TR-4329, University of Maryland, 2001.

[3] Darrell, T., G. Gordon, M. Harville, and J. Woodfill, Integrated Person Tracking Using Stereo, Color, and Pattern Detection, International Journal of Computer Vision, vol. 37, no. 2, pp. 175-185, 2000.

[4] Intel Open Source Computer Vision Library, version 3.1, http://www.intel.com/research/mrl/research/opencv/, 2003.

[5] Izquierdo, E., Disparity/Segmentation Analysis: Matching with an Adaptive Window and Depth-Driven Segmentation, IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 4, pp. 589-607, 1999.

[6] Kanade, T., H. Saito, and S. Vidula, The 3D Room: Digitizing Time-varying 3D Events by Synchronized Multiple Video Streams, Technical Report CMU-RI-TR-98-34, Carnegie Mellon University, 1998.

[7] Kass, M., A. Witkin, and D. Terzopoulos, Snakes: Active Contour Models, International Journal of Computer Vision, vol. 1, no. 4, pp. 321–331, 1987.

[8] Pointgrey Research Inc., http://www.ptgrey.com, 2003.

[9] Salembier, P. and F. Marques, Region-based Representations of Image and Video: Segmentation Tools for Multimedia Services, IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 8, pp. 1147-1169, 1999.

[10] Xu, C. and J.L. Prince, Gradient Vector Flow: A New External Force for Snakes. Proc. IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR), Los Alamitos: Comp. Soc. Press, pp. 66-71, 1997.

[11] Woo, W., N. Kim, and Y. Iwadate, Object Segmentation for Z-keying Using Stereo Techniques, Proceedings of ICSP 2000, pp. 1249-1254, 2000.