Approaches to Depth Estimation from Active Camera Control

D. Paulus and G. Schmidt
Lehrstuhl für Mustererkennung (Informatik 5)
Universität Erlangen–Nürnberg
Martensstr. 3, 91058 Erlangen (Germany)
{paulus,gzschmid}@informatik.uni-erlangen.de

This paper will be printed in the proceedings of the
German / Solvenian Workshop 1996

Rev. 3.3 of 9. Juli 1996
LaTeX'ed 10th July 1996

Contents

Approaches to Depth Estimation from Active Camera Control

D. Paulus and G. Schmidt

Lehrstuhl für Mustererkennung (Informatik 5)
Universität Erlangen–Nürnberg
Martensstr. 3, 91058 Erlangen (Germany)
{paulus,gzschmid}@informatik.uni-erlangen.de

Abstract

In this paper we report on theoretical results and practical experiments for depth recovery from static scenes using active camera devices. Range estimation from linear and rotational camera motion and zoom purposive variation are discussed. Qualitative as well as quantitative depth estimation techniques are applied. Feature tracking in color images is used to estimate distances. Since focus, zoom, aperture, and lens distortion are not independent, and no calibration is desired here, the major goal is simple qualitative depth estimation. In the experiments we use a cost surveillance camera Canon VC–C1 for which we show that depth can be estimated from zoom variation in a limited range.

1 Introduction

In the last decade, computer vision research concentrated mainly on the analysis of static images or recorded image sequences. Each image was analyzed as good as possible generating a symbolic description; this is quoted as the so called "Marr paradigm" [15]. The new idea of "active vision" [2, 3, 4, 22], the demand for real time systems, and the availability of computer controlled camera devices motivate computer vision with active methods and active devices.

The recovery of 3D information from 2D projections is a central goal of many computer vision systems. Obviously, biological systems can solve this task by cooperative use of eye movement, head moves, vergence (for stereoscopic images), and possibly focus information. Technically, active modification of camera parameters can also be used to reach this goal.

In this paper we report on two devices which are used for active depth recovery. We distinguish between quantitative methods and methods which check tendencies in the computed range data and yield qualitative results. The major goal here is to estimate relative distances of objects in terms of "close", "far", and relations like "in front of" and "behind".

[1]This work was funded partially by the German Research Foundation (DFG) under grant number SFB 182. Only the authors are responsible for the contents.

In Sect. 2 we survey related work and present previous results on 3D information from linear camera motion. In Sect. 3 we look at the geometry of active pan/tilt camera heads. Sect. 4 introduces the models for the active camera devices used in this system. In Sect. 5 we describe the idea how to compute depth from zoom variation. We show results and experiments in Sect. 6. A summary is given in Sect. 7 and further work is outlined in the concluding section.

2 Linear Motion

Depth recovery from monocular image sequences has been reported by several authors [5, 6, 19]. Depth from a moving camera can be computed by tracking features, if the cameras position, settings, and direction are known for each picture [5]. If the camera is moving linearly and the optical axis is perpendicular to the moving direction, the features are moving linearly, too. The camera is mounted to a robot's hand and the robot moves on a linear axis; this setup is part of a larger system [9]. Figure 1 shows an image captured with this system. Lines can easily be detected, so the depth can be computed directly from the slopes of the features' trajectory [5]. In [19], a maximum error in depth lower than 1% could be achieved using this method.

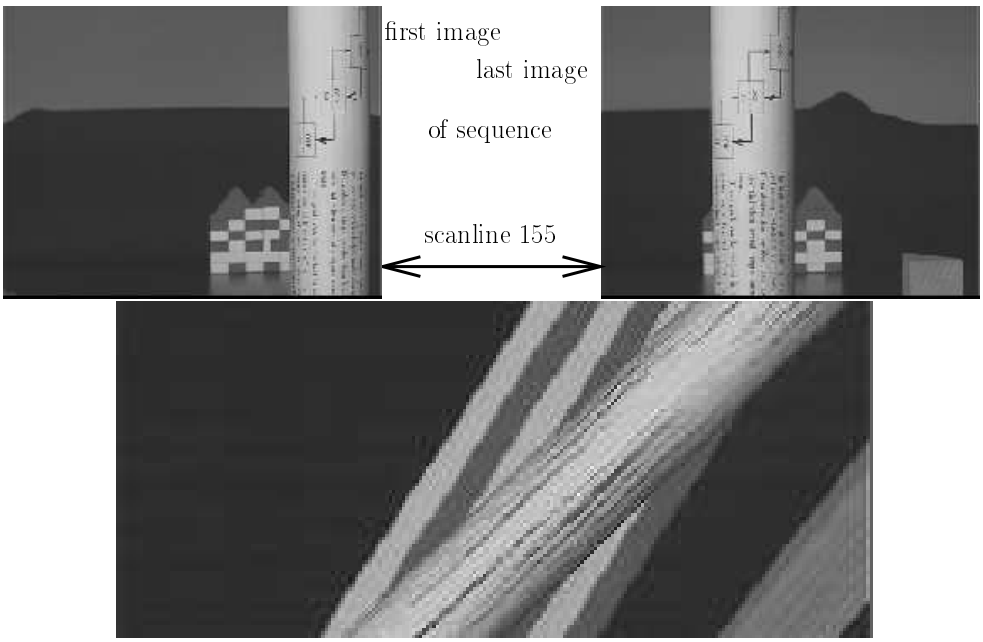
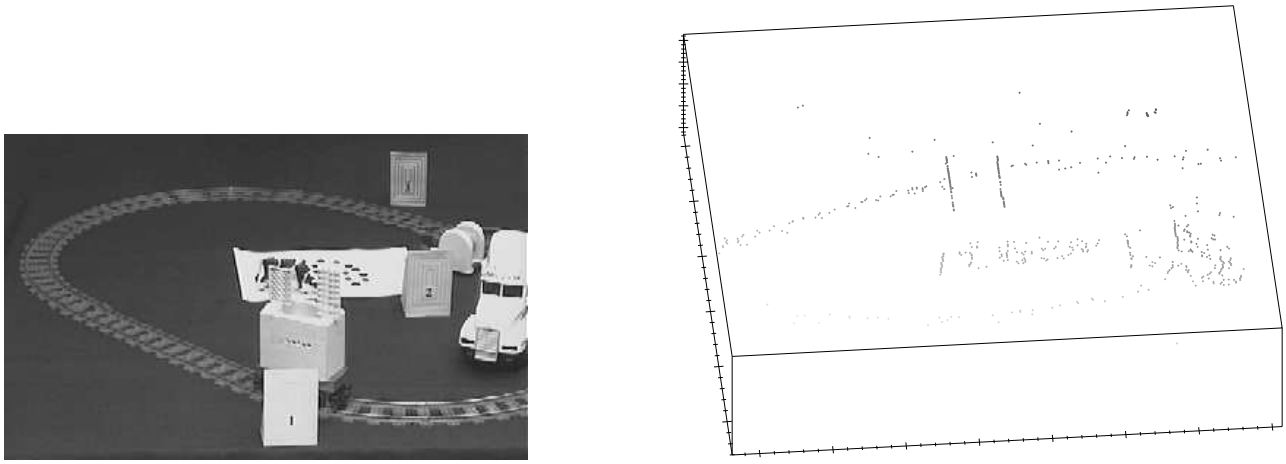


Figure 1: Two images of a sequence (top); scanline #155 for all images of the sequence combined to an image (below)

Figure 2(a) shows an experimental scene. The 3D positions of these points are shown in Figure 2(b). Rather than quantitative results, qualitative methods can also be used here. From Figure 1 below, straight line detection can be used to infer relative positions;

higher inclination characterize objects closer to the camera, line intersections relate to occlusions in the scene. Such methods are particulary useful for an uncalibrated setup. For arbitrary angles between motion direction and optical axis the trajectories will be hyperbolas. Intersections still relate to occulsions.



(a) Experimental scene (b) 3D from linear motion

Figure 2: Depth from linear motion

Axial motion stereo is another case of a moving camera with known motion parameters. Range can be recovered from such camera motion [11]. Slightly different geometries can be observed, if the camera is static but the focal length of the lens varies (e.g. in a zoom lens, [14]). In [8, 13], an uncalibrated setup was used to compute dense depth maps from zoom. We will see more about this subject in Sect. 5.

3 Trajectories from Pan and Tilt

Another idea for depth recovery from monocular image sequences requires a stationary camera which can zoom, pan, and tilt, but is stationary otherwise. Two widely used devices for this purpose are the Canon VC–C1 camera and the TRC head which is built for computer vision research purposes. Figure 3(a) shows the geometric relations of a rotation with radius r around some point outside the optical axis in 3D. The angle between the rotation and optcial axes is α. For fixed focal length f we compute the trajectory $x(\phi)$ of a point with distance d to the rotation center.

Assuming a rotation axis parallel to the image plane, we can simplify the rotation to 2D (Figure 3(b)). The basic relation is

$$x = f \cdot \tan(\alpha - \arcsin(d \sin\phi / \sqrt{r^2 + d^2 - 2rd\cos\phi})$$

If the rotation axis is on the optical axis ($r = 0$), the distance d vanishes from the equation; i.e., in order to compute depth from such trajectories, a translational displacement is required in the image coordinates. Even for rotations around some other point, the image plane displacement is relatively small for real devices. A plot of this function is

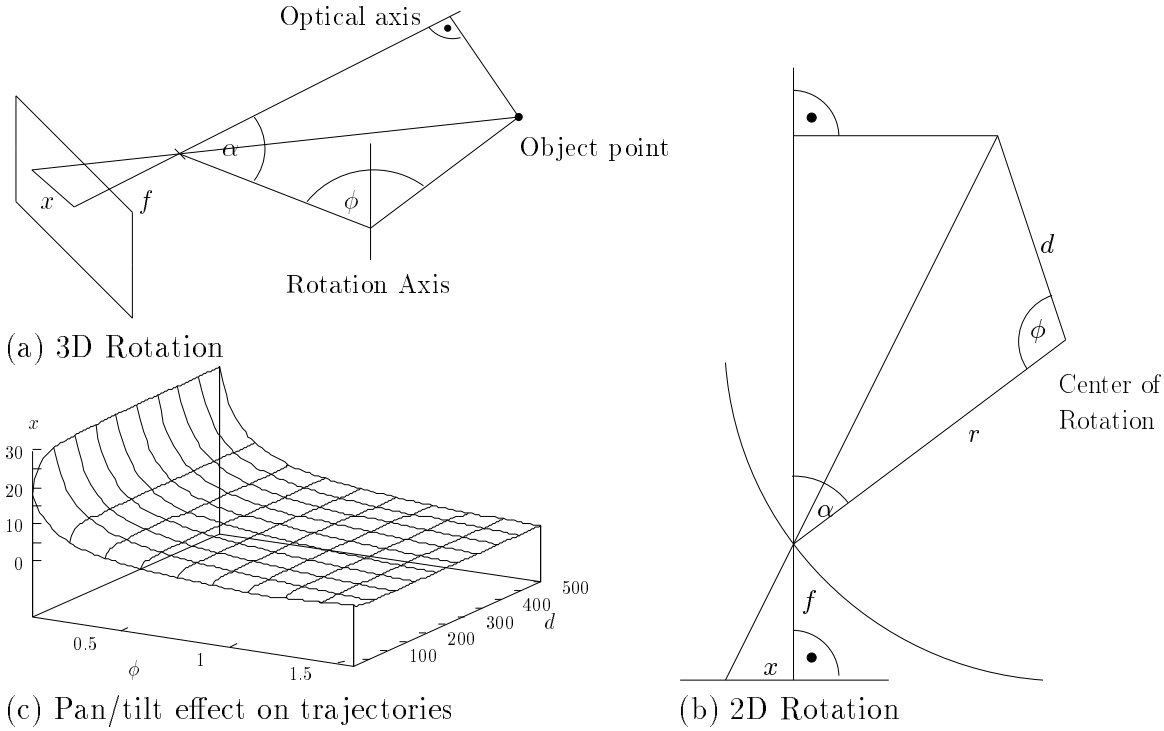


Figure 3: Geometry for rotation (pan/tilt)

shown in Figure 3(c); obviously, a change in distance d has little effect on the shape of the function. The actual values for r are approx. 3cm for the Canon and 10cm for the TRC head pan axis; the angle α is approximately 90^o for the Canon camera, and varies due to the vergence axis on the TRC head.

To conclude, for depth recovery one has to use other visual motion, such as zoom or a combination of zoom and camera motion.

4 Zoom Camera Models

The camera models used often for fixed lenses are the pinhole camera and the thin lens model [16]. These descriptions are merely models, i.e., they idealize real lenses. For exact measurements the lenses have to be calibrated, for example with Tsai's method [23] for the pinhole model.

The basic equations are $r = fR/Z$ for the pinhole model, and $1/f = 1/s + 1/s'$ for the thin lens model. To model a zoom lens the first idea is to take one of the mentioned models and change only the focal length when zooming. The question is now, how the focal length changes with zooming.

A real lens is not perfectly manufactured. The optical axis and the mechanical axis do not coincide. So the intersection of the optical axis and the image plane moves if the lens is moved along the mechanical axis. Real lenses also have a distortion which is often assumed to be radial. The area of the lens which is used for different zoom settings, differs.

The calculated distortion then relates to different parts of different size of the lens; so it is obvious that the distortion changes with zoom settings, too.

A zoom lens consists of several single lenses. These lenses need not be moved in the same way by changing lens adjustments. They can be just moved or rotated. They are also changing their direction when camera adjustments are changed monotonely. The parameters have to be recovered by calibration. In [24] a method to calibrate zoom lenses is described.

The first idea to get a zoom model is to calibrate the lens for each focal and zoom setting. Aperture may also influence the model parameters, for example radial distortion. So thousands of calibrations would have to be done. In case of the TRC head, this results in $4 \cdot 10^4$ (zoom) $\times$ $4 \cdot 10^4$ (focus settings) $= 16 \cdot 10^8$ calibrations. For the Canon camera this figure is slightly smaller, since the focus motor has an open loop controller and can thus only be positioned with moderate accuracy to 150 positions using the algorithm described in [1]; in addition, the auto–iris makes exact calibrations impossible.

The second idea is based on the assumption that the parameters change smoothly for most settings [24]. Just a few calibrations may be enough and the rest can be interpolated with a small error. This idea is sufficient for qualitative depth perception, if the parameters change monotone.

Some problems still remain. When calibrating focus, the calibration pattern should be sharply displayed for an accurate measurement. This causes different distances to the lens. We use a linearly moving skid which can be moved by 1/10mm steps. Another problem is the size of the calibration pattern for calibrating at different zoom settings. The Canon VC–C1 cameras maximum magnification at maximal focal length is eight times higher than at wide–angle zoom setting. The cameras of the TRC have a maximum magnification of 10. One small and one large calibration pattern has to be used. To automate calibration, a large pattern can be combined with a small one if they have different features which can be distinguished at tele and wide zoom setting. We use a calibration pattern with filled circles of different colors. The calibrations software then can recognize which pattern has to be applied for a special zoom setting.

5 Depth from Zoom

In [12] we examined how to compute depth from two different zoom settings f_1 and f_2. Assuming a pinhole model (see solid lines in Figure 4) we get

$$\frac{r_i}{f_i} = \frac{R}{Z - f_i}; \quad i = 1, 2 \tag{1}$$

$$Z = f_1 f_2 \frac{r_1 - r_2}{r_1 f_2 - r_2 f_1} \tag{2}$$

However, an error in localization of corresponding features of one pixel can result in an error of depth approximately as large as the depth itself [12]. This effect is shown with dotted lines in Figure 4. To get acceptable results from only two zoom settings, the localization has to be very precise.

More reliable information about depth can be obtained if more images at different zoom settings are used. Tracking a feature while zooming delivers for each zoom setting

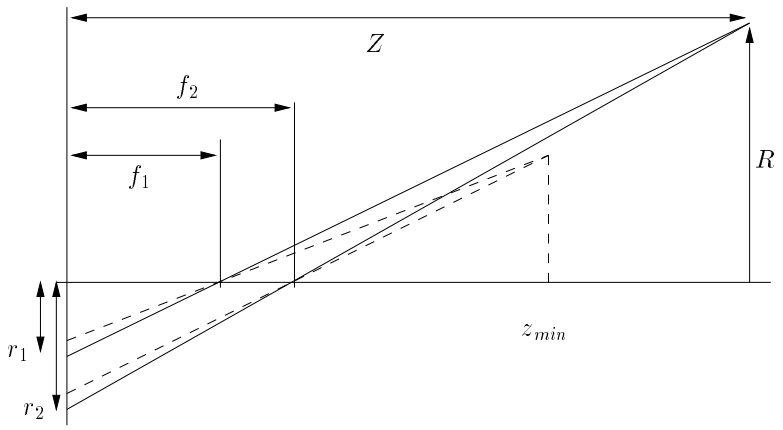


Figure 4: Depth from two zoom settings with the pinhole model [12]

a distance r of the feature to the image center. From this information the parameters Z and R can be estimated by minimizing an error criterion. Tracking features r_i at known positions f_i, we conclude from (1):

$$R = \frac{r_i}{f_i} Z - r_i. \tag{3}$$

With a simple linear regression, the parameters Z and R can be determined. The question, which features are well suited for tracking, is discussed in [21]. We select points in the color images which differ in variances of the four quadrants in a rectangular neighborhood. It can be shown that the zoom trajectory of a stationary point in 3D will be a straight line through the optical center, assuming an ideal zoom lens [17]; this simplifies tracking. This property is used to predict the next position after zooming. We use elliptical search areas for point tracking based on correlation (see below, Figure 8(right)). In this sense, no correspondence problem has to be solved, as required for stereo analysis.

Accurate determination of the origin of the coordinate system (optical center) is crucial for this method. Qualitative methods have to avoid this dependency; the tendencies of changes are analyzed and used to infer terms such as "closer" or "far".

6 Experiments and Results

We now describe solutions for a surveillance camera (Canon VC–C1). In principle, the ideas can also be used for the much more accurate TRC stereo head [7] (see Sect. 8).

Referring to Sect. 4 our first experiment was to get information about the behavior of the image center. In Figure 5 the movement of the image center while zooming is shown. The focus of expansion for zoom was taken as image center as described in [24]. Since this position not only varies with zoom but also with other camera parameter settings, quantitative results for depth recovery from zoom are unfeasible.

Figure 6(a) shows a measure for the focus for different zoom positions. The sum of the edge strength in the central part of the window is computed with auto–focus and without.

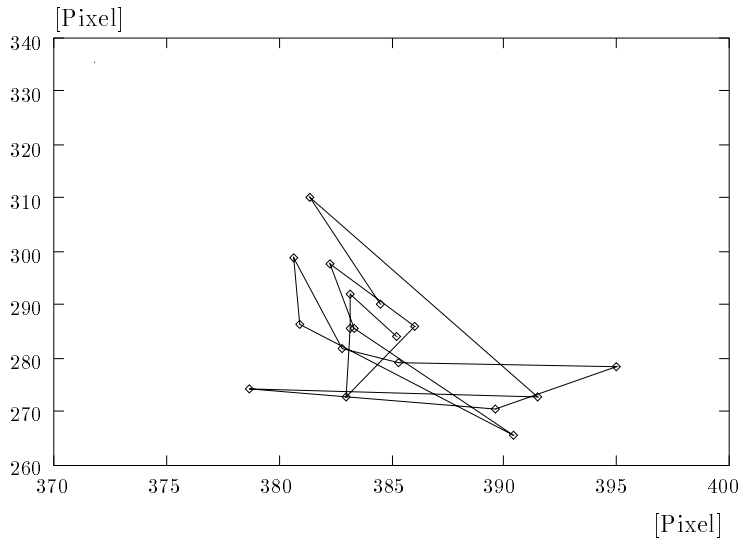
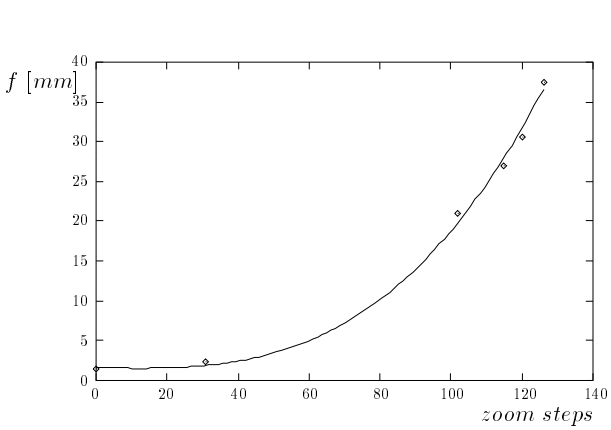
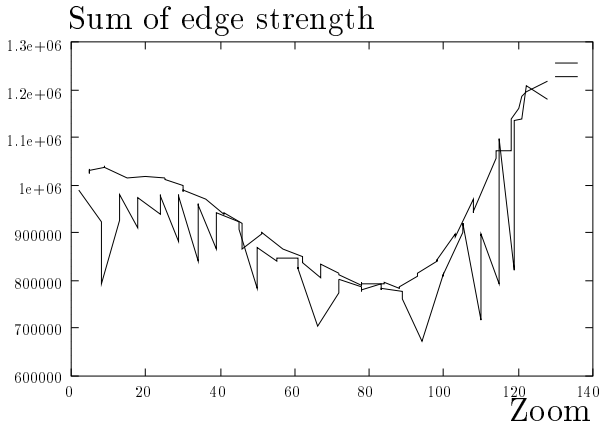


Figure 5: Image center moves while zooming for Canon VC–C1

This experiment indicates that during zoom the focus has to be adjusted which has an effect on the effective focal length. Figure 6(b) shows the results of 6 calibrations for the effective focal length for the whole zoom range of the Canon VC–C1. The focal length changes monotonely but not linearly with zoom. For approximation a cubic polynomial is used.



(a) Effective focal length for the zoom range of Canon VC–C1



(b) Focus for different zoom positions

Figure 6: Measurements for the Canon VC–C1

For small distance, range estimation based on the hyperbola (eq. 3) by regression yields reasonable results (Figure 7 right) from which relative distance (far, close) can be concluded. Points are selected in the first image on a regular grid; these points are tracked during zoom. If the correlation based search fails to find a maximum, the point is discarded. For Figure 7, the distance between object and lens was approx. 50 cm.

For larger distances, other methods have to be applied. In Figure 8(left) we show two objects with distances of approximately 1.5m and 2.5m from the lens. The selected point has the same distance to the optical axis in world coordinates. Figure 8(right) shows the corresponding elliptical search areas for tracking. The x–coordinates of the tracked points are shown in Figure 8(center). The slope of distant points (upper two lines) are smaller than that of the closer point (lower line). Depth estimation can here be done qualitatively based on the comparison of these lines.

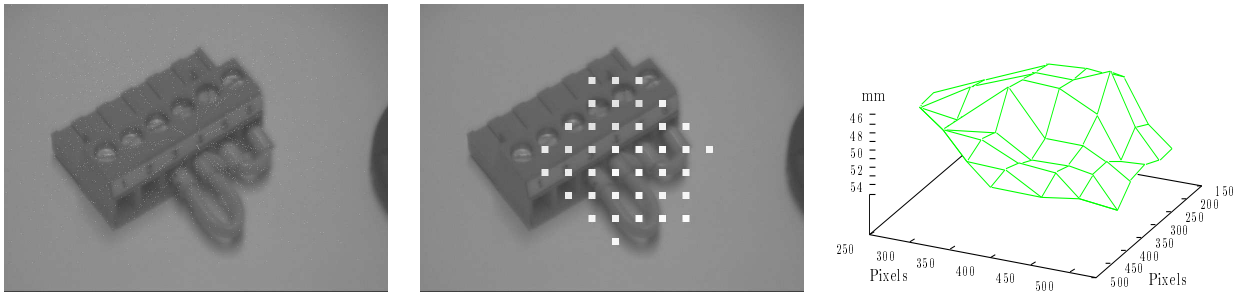


Figure 7: Object, interesting points (regular grid), estimated distances

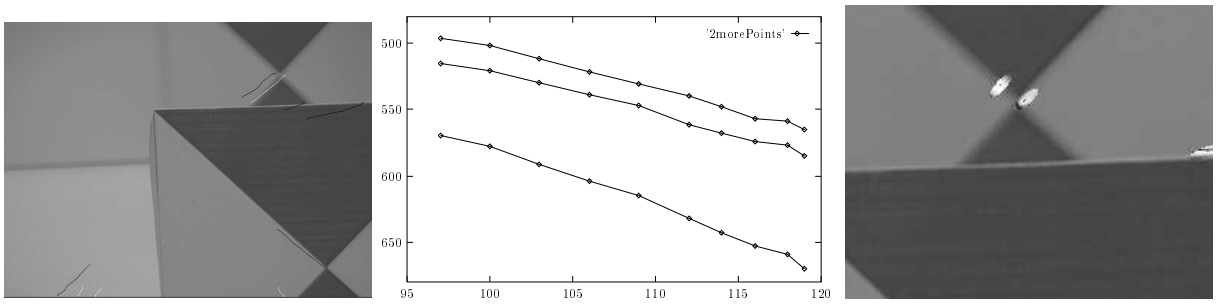


Figure 8: Distance estimation based on trajectory slope

For more complex scenes, the results were not satifactory using the Canon VC–C1. Regression for the hyperbolas yields inaccurate results, since this method relies on the distance to the optical center — which is just moving too much during zoom.

7 Conclusion

We described how distance estimation can be based on zoom variation of an active camera device. Practical experience shows that this method is only of limited value, if the camera and lens are no high quality product. In case of the Canon VC–C1 the method fails, in general.

Whereas linear motion of a camera can be used for accurate depth recovery, we showed that pan and tilt motion do not contribute to such computations, even if the rotation center is not on the optical axis.

The system described here is part of a larger project for object recognition and tracking [9]. The Canon camera is used here successfully for object tracking without explicit 3D

reconstruction. In combination with the calibration results shown above (e.g. in Figure 6), this camera is extremely useful for computer vision, although it misses the accuarcy for depth estimation [10]. All implementations in these projects are done in C++ and utilize object–oriented programming techniques [18]. More details about the particular implementation described in this contribution can be found in [20].

8 Further Work

The lenses on the TRC stereo head are of much higher quality. We hope to get more reliable results on this device using the zoom techniques described above. As we showed in principle, depth can be computed from a point trajectory along a zoom image sequence. If, however, the point to be tracked is close to the optical axis, the trajectory will be only few pixels long in the projection to the image plane. This effect can be shown in eq. 2 where the difference of r_1 and r_2 delivers approximately zero for the whole zoom range.

In this case, a camera move is required, moving the point to the border of the image plane. A subsequent zoom will then move the point on a straight line towards the optical center. Since even for unknown rotation axes the translational part of the trajectory can be ignored (cmp. Figure 3 (c)) prediction of the point's poistion is fairly simple for purposive pan and tilt changes.

Acknowledgement

The authors wish to express their thanks to Professor Stane Kovačič who read [12], then pointed to the work of Willson [24], and who also did the first experiments for point tracking with the Canon zoom lens.

References

[1] U. Ahlrichs. Sprachgesteuerte Fovealisierung und Vergenz *Fovealization and Vergence Control via Natural Language.* Diploma thesis, IMMD 5 (Mustererkennung), Universität Erlangen–Nürnberg, Erlangen, 1996.

[2] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):996–1005, 1988.

[3] R. Bajcsy and M. Campos. Active and exploratory perception. *Computer Vision, Graphics and Image Processing*, 56(1):31–40, 1992.

[4] A. Blake and A. Yuille, editors. *Active Vision.* MIT Press, Cambridge, Mass., 1992.

[5] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining struc ture from motion. *Int. Journal of Computer Vision*, 1:7–55, 1987.

[6] B.F. Buxton and H. Buxton. Monocular depth perception from optical flow by space time signal pro cessing. In *Proceedings of the Royal Society of London*, volume 218 of *B*, pages 27–47, 1983.

[7] K. Daniilidis, M. Hansen, C. Krauss, and G. Sommer. Auf dem Weg zum künstlichen aktiven Sehen: Modellfreie Bewegungsverfolgung durch Kameranachführung. In

G. Sagerer, S. Posch, and F. Kummert, editors, *Mustererkennung 1995*, pages 277–284, Berlin, September 1995. Springer.

[8] C. Delherm, J. M. Lavest, M. Dhome, and J. T. Lapresté. Dense reconstruction by zooming. *European Conference on Computer Vision*, B:427–454, 1996.

[9] J. Denzler, R. Beß J. Hornegger, H. Niemann, and D. Paulus. Learning, tracking and recognition of 3D objects. In V. Graefe, editor, *International Conference on Intelligent Robots and Systems – Advanced Robotic Systems and Real World*, volume 1, pages 89–96, 1994.

[10] J. Denzler and H. Niemann. Active rays: A new approach to contour tracking. In *German-Slowenia Workshop on Speech and Image Processing*, page to appear, Ljubljana, 1996.

[11] X. Y. Jiang and H. Bunke. Line segment based axial motion stereo. *Pattern Recognition*, 28(4):553–562, April 1995.

[12] M. Köhler. Polyedervermessung mit aktivem Sehen *Polyeder measurements by active vision*. Student's thesis, IMMD 5 (Mustererkennung), Universität Erlangen–Nürnberg, Erlangen, 1994.

[13] J.-M. Lavest, G. Rives, and M. Dhome. Three–dimensional reconstruction by zooming. *Transactions on Robotics and Automation*, 9(2):196–207, 1993.

[14] J. Ma and S. I. Olsen. Depth from zooming. *Journal of the Optical Society of America*, 7(10):1883–1890, 1990.

[15] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freemantle, San Francisco, 1982.

[16] H. Niemann. *Pattern Analysis and Understanding*. Springer, Heidelberg, 1990.

[17] D. Paulus and J. Denzler. Möglichkeiten und Grenzen aktiven Sehens mit passiven Sensoren. In G. Schmidt, editor, *9. Fachgespräch für autonome mobile Systeme*, pages 275–286, München, 1993.

[18] D. Paulus and J. Hornegger. *Pattern Recognition and Image Processing in C++*. Advanced Studies in Computer Science. Vieweg, Braunschweig, 1995.

[19] G. Schmidt. Tiefe aus linearer Kamerabewegung (*Depth from linear camera motion*). Student's thesis, IMMD 5 (Mustererkennung), Universität Erlangen–Nürnberg, Erlangen, 1995.

[20] G. Schmidt. Qualitative Tiefenermittlung aus Pan/Tilt/Zoom (*Qualitative Depth Recovery by Pan/Tilt/Zoom*). Diploma thesis, IMMD 5 (Mustererkennung), Universität Erlangen–Nürnberg, Erlangen, to appear 1996.

[21] J. Shi and C. Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

[22] M. J. Swain and M. Stricker. Promising directions in active vision. Tech. Rep. CS 91-27, University of Chicago, November 1991.

[23] R. Y. Tsai and R. K. Lenz. Real time versatile robotics hand/eye calibration using 3d machine vision. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 554–561, Philadelphia, April 1988. IEEE Computer Society Press.

[24] R.G. Willson. *Modeling and Calibration of Automated Zoom Lenses*. PhD thesis, Carnegie Mellon University, 1994.