Panoramic 3D Reconstruction with Three Catadioptric Cameras

Miriam Schönbein, Holger Rapp and Martin Lauer

Institute of Measurement and Control Systems
Karlsruhe Institute of Technology (KIT)
D-76131 Karlsruhe, Germany
{miriam.schoenbein, holger.rapp, martin.lauer}@kit.edu

Abstract— In this paper, we present a new system setup combining three catadioptric cameras. Our application is in autonomous vehicles. This camera setup allows for panoramic stereo vision of the environment and therefore proves to be useful for ego motion estimation and localization by 3D feature points all around the vehicle. The three catadioptric cameras are arranged in a triangle on top of the vehicle and are horizontally aligned. In the paper we discuss two methods for 3D scene point reconstruction with the system. We perform experiments for the 3D reconstruction in a simulated environment and evaluate the accuracy by means of a Monte-Carlo-Simulation. The proposed system improves the mean accuracy of the 3D reconstruction significantly compared to a system with two cameras.

I. INTRODUCTION

For intelligent robotic applications environment perception and 3D scene reconstruction are current fields of research. Especially for autonomous driving stereo vision and camera based reconstruction of a large area of the environment are very important tasks. Therefore it is desirable for camera systems utilized for autonomous vehicles to have a large monoscopic as well as stereoscopic field of view (FOV) of the environment. However, cameras typically mounted on vehicles have a limited FOV and are looking only in driving direction. But for many autonomous vehicle applications like blind spot detection or lane tracking it is necessary to perceive also the areas beside and behind the vehicle. Moreover, landmark detection on the sides of the vehicle could improve self-localization.

One common way to enhance the FOV with only one camera is using catadioptric camera systems. Catadioptric cameras are a combination of a convex mirror above a lens and provide a panoramic image of the environment at the cost of image resolution. They have been extensively studied in the past [1, 2] and are often used in indoor robotic applications for navigation and surveillance tasks. There are also some approaches using single catadioptric cameras for driver assistance application like lane tracking [3], blind spot detection [4] or ego motion estimation [5]. However, up to now the potential of stereo vision with catadioptric cameras for outdoor vehicles has not been explored in depth.

In recent years different types of omnidirectional stereo camera systems combining catadioptric cameras have been proposed. A large number of approaches uses stereo vision
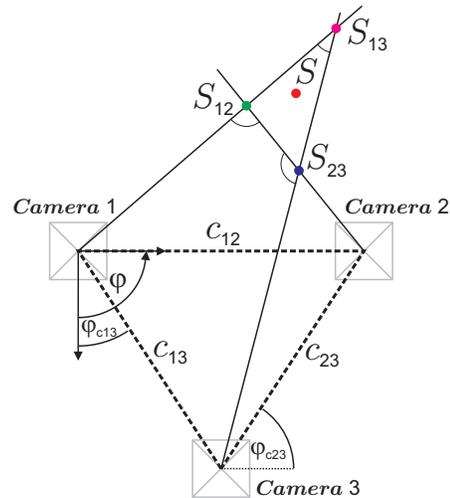

Fig. 1. Camera Setup with three catadioptric cameras. S_{12}, S_{13} and S_{23} show the 3D intersection points in each case by two cameras. S is the 3D scene point calculated by all three cameras.

with a vertical camera baseline. This can be achieved by two catadioptric cameras mounted on top of each other [6], by two mirrors mounted on top of each other with only one lens [7] or a special designed mirror with two slopes and only one lens [8]. However, vertically aligned stereo camera systems are not very useful for autonomous outdoor vehicles due to the small baseline of the camera configuration and the height of the setup. A large baseline can only be archived by a high pole mounted on the roof of the vehicle which is not suitable.

Gandhi and Trivedi [9] present a catadioptric camera system for visualizing and analyzing the nearby surrounding with two horizontally aligned cameras mounted on the left and the right side mirror of a vehicle. In [10] we discuss a similar catadioptric stereo camera setup with two catadioptric cameras mounted on the left and the right side on top of the vehicle for 3D reconstruction. This camera configuration allows for accurate stereo vision in the area in front and behind the vehicle. A drawback of such a system with horizontally aligned cameras is the dependency of the depth error from the magnitude of the intersection angle between the two rays. Thus, the advantage of a panoramic stereo view exists only in front and behind the vehicle.

In this paper we propose a new panoramic camera setup consisting of three planar aligned catadioptric cameras arranged in a triangle. With this configuration we increase the monoscopic FOV of the nearby environment by the improved position on the vehicle as well as the stereoscopic FOV. The camera configuration allows for 3D reconstruction in the whole environment around the vehicle and therefore improves many stereo based downstream application for autonomous driving. We show that three cameras overcome the problem of stereo blind spots along the baseline connecting two cameras.

The paper is divided as follows. In Section II we present the theoretical model of our catadioptric camera and the 3D reconstruction. Section III describes our new camera setup and the reconstruction with three cameras. In Section IV we show results for enhancement of the 3D reconstruction with cameras on simulated data. Finally, Section V concludes this paper and shows future work.

II. CATADIOPTRIC SCENE RECONSTRUCTION

A. Catadioptric Camera Model

For our trifocal panoramic system we use catadioptric cameras consisting of an hyperbolic mirror and a perspective lens. These catadioptric cameras fulfill the single view point condition [2]. Different projection models for single view point catadioptric cameras exist to define the relationship between a 3D world point $\mathbf{S}$ and a 2D point $\mathbf{u}$ on the catadioptric image plane ([11, 12]). In our approach we use the projection model presented by Scaramuzza [13] which takes into account natural misalignments between the camera and the mirror. This model describes the projection function as a polynomial. The dependency between an image point $\mathbf{u} = [u, v]^T$ and a vector $\mathbf{p}$, which presents a ray from the focal point of the mirror F_C to the scene point $\mathbf{S}$, is encoded by the function g.

$$\mathbf{p} = g(\mathbf{u}') = g(A\mathbf{u} + t) = P\mathbf{S} \tag{1}$$

$\mathbf{S} \in \mathbb{R}^4$ is expressed in homogeneous coordinates, $P \in \mathbb{R}^{3 \times 4}$ is the perspective projection matrix and $A \in \mathbb{R}^{2 \times 2}$ and $t \in \mathbb{R}^{1 \times 2}$ are affine transformation parameters for misalignment errors and digitizing artefacts.

The function g has the following expression

$$g(\mathbf{u}) = (u', v', f(u', v'))^T \tag{2}$$

and the polynomial f is defined as

$$f = a_0 + a_1 \cdot \rho + a_2 \cdot \rho^2 + a_3 \cdot \rho^3 + a_4 \cdot \rho^4 \tag{3}$$

with

$$\rho = \sqrt{u'^2 + v'^2}. \tag{4}$$

B. 3D Reconstruction

Out of two corresponding image points in a catadioptric image pair the according 3D world points $\mathbf{S}$ can be reconstructed by calculating the points on the mirror surfaces $\mathbf{X}_{M_i}$ corresponding to the image points $\mathbf{u}_i$ for both catadioptric cameras ($i = 1, 2$). Afterwards, we construct the rays through the focal point of the mirrors F_{C_i} and the points on the mirror
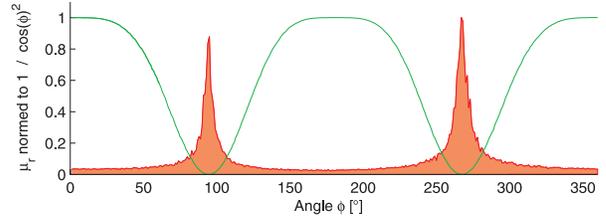

Fig. 2. Normalized mean Euclidean error μ_d as a function of angle φ of the 3D scene point (red) and the weight function $cos^2(\varphi)$ (green).

surfaces $\mathbf{X}_{M_i}$. The 3D scene point $\mathbf{S}$ is given as the midpoint of the smallest distance [14] between the two rays.

In [10] we discuss the accuracy of the 3D reconstruction with two horizontally aligned cameras i.e. Camera 1 and 2 in Fig. 1. We demonstrate that the accuracy depends on the position of the scene point $\mathbf{S}(r, \varphi, \theta)$ given in polar coordinates. The accuracy decreases from $\varphi = 0°$ to $\varphi = 90°$ with $0°$ being behind the two cameras and $90°$ being on the right side of the cameras as shown in Fig. 1. In Fig. 2 the dependency of the 3D reconstruction from the angle φ and the accuracy are displayed. The red function shows the normalized mean Euclidean error μ_d for the reconstruction of a circular simulated point cloud. The points are located in a circle with constant height around the two aligned catadioptric cameras with a radius of 8 m. Consequently, the reconstruction error is maximal at $\varphi = 90°$ and $\varphi = 270°$ which is along the baseline of the camera combination.

III. TRIFOCAL VISION

A. Camera Setup

In our approach we use three catadioptric cameras arranged in a plane parallel to the ground in a triangle as shown in Fig. 1. The cameras are mounted on top of the vehicle, two in the front on the left and right side and the third one centered in the rear. With this camera configuration all points around the vehicle are seen by at least two cameras. Moreover, points which lie on the baseline from two cameras and can not be reconstructed with the two cameras are visible by the third camera, if not covered by another object. Thus, every 3D point around the vehicle can be calculated out of the image points according to the best positioned cameras with a sufficient accuracy. In addition this camera configuration with the cameras mounted at the edges of the vehicle has the advantage that also small objects next to the vehicle can be seen with at least one camera.

B. 3D Reconstruction

First of all, for the 3D reconstruction out of three single catadioptric cameras we calculate the corresponding 3D intersection points S_{12}, S_{23} and S_{31} out of two cameras between all camera combinations as shown in Fig. 1. The straight forward approach is to determine $\mathbf{S} = \frac{1}{3} \sum \mathbf{S}_n$ with $n \in \{12, 13, 23\} =: N$ as the average of all three intersection points. Obviously, this computation is not advantageous due to the varying accuracy of the pairwise 3D reonstruction as shown in Fig. 2.

To establish the best 3D point out of the intersection points $\mathbf{S_n}$ we evaluate two different methods. Our first method calculates the best intersection point $\mathbf{S}$ by a sum over all intersection points each multiplied by a window function $w_n(\varphi)$ depending on the angle φ.

$$\mathbf{S} = \sum_{n \in N} w_n(\varphi) \cdot \mathbf{S}_n \tag{5}$$

with the following window functions $w_n(\varphi)$ for all three camera combinations.

$$w_{12}(\varphi) = \begin{cases} 1, & \text{for } \varphi \in [0°, 30°] \cup [150°, 210°] \cup [330°, 360°] \\ 0, & \text{else} \end{cases}$$

$$w_{13}(\varphi) = \begin{cases} 1, & \text{for } \varphi \in [90°, 150°] \cup [270°, 330°] \\ 0, & \text{else} \end{cases}$$

$$w_{23}(\varphi) = \begin{cases} 1, & \text{for } \varphi \in [30°, 90°] \cup [210°, 270°] \\ 0, & \text{else} \end{cases}$$

This method uses only the best intersection points depending on the angle of the 3D point.

To use all three intersection points our second method calculates the 3D points with a weighted sum over all three points.

$$\mathbf{S} = \frac{\sum g_n \cdot \mathbf{S}_n}{\sum g_n} \tag{6}$$

The weight functions g_n depend on the angle φ_{c_n} between the pairwise translation vector connecting the cameras $\mathbf{c_n} = [c_{x_n}, c_{y_n}, c_{z_n}]$ and the angle φ of the 3D point, shown in Fig. 1.

$$\varphi_{c_n} = \arccos \frac{c_{y_n}}{\sqrt{c_{x_n}^2 + c_{y_n}^2}} \tag{7}$$

For the weight function $g_n(\|\varphi_n - \varphi_{c_n}\|)$ we use

$$g_n(\|\varphi_n - \varphi_{c_n}\|) = \cos^2(\|\varphi_n - \varphi_{c_n}\|) \tag{8}$$

as depicted in green in Fig. 2. The weight function penalizes intersection points with a small spatial angle to the baseline between the cameras and rewards points perpendicular to the baseline.

IV. RESULTS

In this section the results of our approach on a simulated 3D environment are presented. We created 3D scene points in a circle ($\varphi = 0°, 1°, ..., 360°$) around the three cameras with radius 8 m and a constant height. The cameras are located at position $[0\ 0\ 0]^T$, $[0\ 1\ 0]^T$ and $[1\ 0.5\ 0]^T$ as shown in Fig. 1. The accuracy of the 3D reconstruction is evaluated by means of a Monte-Carlo-Simulation. For this purpose the circular scene points are mapped onto the image planes of the catadioptric cameras. For each of the 360 mapped image points we created 100 samples and disturbed them with Gaussian noise with a standard deviation of 1.5 pixels. Using these disturbed image points we calculated the corresponding points on the mirror surfaces and determined the 3D intersection points by reprojection with the presented methods using Eq. 5 and Eq. 6.
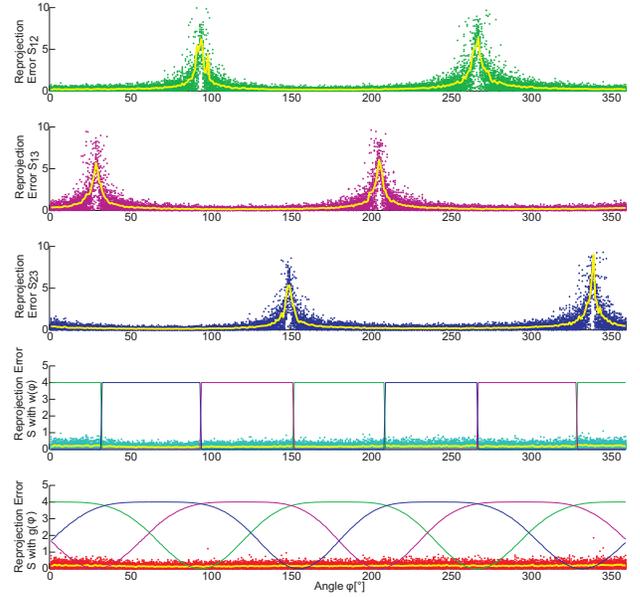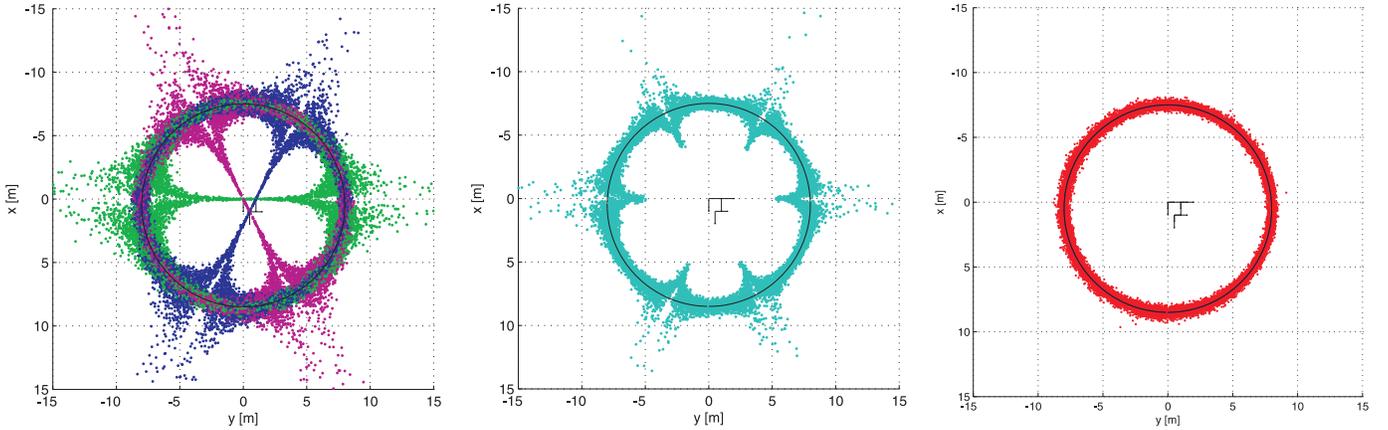

Fig. 3. Euclidean reprojection error and mean μ of the three camera combinations (S_{12} - green, S_{13} - magenta, S_{23} - blue). The fourth figure depicts the error for the 3D reconstruction calculated with Eq. 5 (cyan) and the last one the error for the 3D points calculated with Eq. 6 (red). The weight functions are shown as colored lines.

The 3D reconstruction can be judged by the reprojection error which is the Euclidean distance $\|\mathbf{X} - \mathbf{S}\|$ between the reconstructed point $\mathbf{S}$ and the ground truth scene point $\mathbf{X}$. Fig. 3 shows the error for the reconstruction with two cameras for Camera 1 and 2 (green), for Camera 1 and 3 (magenta) and for Camera 2 and 3 (blue). The dependency on the angle φ can clearly be seen. The same effect is visible in Fig. 4(a) which shows the reconstructed 3D point cloud from each camera pair in the same color. Obviously, the reconstructed points along the baselines deviate from the true scene points.

The reconstruction error for the estimated 3D point by the two computation methods are also depicted in Fig. 3. The fourth figure shows the reprojection error for the reconstruction method using Eq. 5 and the last figure displays the reprojection error for the reconstruction using Eq. 6. The weight functions $w(\varphi)$ and $g(\varphi)$ are also shown as colored lines in the figures. It can be seen that the reprojection error with both methods is smaller than for the 3D points calculated by only two cameras. Most importantly, it is independent of the angle φ.

Fig. 4 shows the reconstructed point clouds with the black circle as ground truth. In Fig. 4(b) the reprojected point cloud by the straight forward mean combination is displayed. The reconstructed point cloud by the combination of the three cameras with the weight functions (Eq. 6) is illustrated in Fig. 4(c). The reprojected point cloud with Eq. 5 looks similar to this one. It became clear that the reconstructed scene points with the weighted combinations (Eq. 5, 6) are closer to the ground truth than the mean combination of the three intersection points and are independent from the angle φ. Hence, points on all sides of the vehicle can be reconstructed with these methods with similar accuracy.

For a quantitative result we analyze the mean μ_d and

(a) Distribution for the 3D reconstruction of Camera 1 & 2 (green), Camera 1 & 3 (magenta) and Camera 2 & 3 (blue).

(b) Distribution for the 3D reconstruction calculated as average of all three intersection points.

(c) Distribution for the 3D reconstruction calculated as weighted sum of all three intersection points.

Fig. 4.   3D reconstruction for a simulated noisy circular point cloud with radius 8 m around the cameras and a constant height.

standard deviation σ_d of the error shown in Tab. I. The mean and standard deviation for the proposed methods are considerable smaller than for the intersection points calculated with two cameras. The values for the reconstruction with Eq. 6 are slightly smaller. The reconstruction with Eq. 6 also copes better with outliers because it always uses three cameras.

Setup	S_{12}	$\frac{1}{3}\sum(S_n)$	$\sum w_n S_n$	$\frac{\sum(w_n S_n)}{\sum g_n}$
μ_d [m]	0.5785	0.4273	0.1792	0.1738
σ_d [m]	0.8951	0.4282	0.0282	0.0247

TABLE I

MEAN μ_d AND STANDARD DEVIATION σ_d OF THE ERROR.

V. CONCLUSIONS AND FUTURE WORK

In this work, we have presented a new catadioptric camera system for autonomous vehicles. The proposed setup consists of three cameras aligned on a plane and arranged in a triangle on top of a vehicle. We presented two methods for the 3D reconstruction with three catadioptric cameras which achieved good results. Both methods calculate the 3D point by a weighted average, one with a window function and one with a weight function based on the cosine. We have shown that the accuracy of the 3D reconstruction with three cameras is by far better as the reprojection with two cameras and independent of the spatial angle of the expected 3D point.

Our next steps will include a validation of the 3D reconstruction with the camera setup on our experimental vehicle with real world data. Furthermore, we want to use the accurate 3D points all around the vehicle to improve the ego motion estimation and localization with catadioptric cameras. Moreover, the extrinisic calibration between the cameras can be enhanced with this 3D reconstruction.

ACKNOWLEDGMENT

REFERENCES

[1] R. Benosman and S. e. Kang, *Panoramic vision: sensors, theory, and applications*.   Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2001.

[2] S. Baker and S. K. Nayar, "A theory of single-viewpoint catadioptric image formation," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 1 – 22, 1999.

[3] S. Y. Cheng and M. M. Trivedi, "Toward a comparative study of lane tracking using omni-directional and rectilinear images for driver assistance systems," in *ICRA 2007 Workshop: Planning, Perception and Navigation for Intelligent Vehicles*, 2007.

[4] T. Ehlgen, T. Pajdla, and D. Ammon, "Eliminating blind spots for assisted driving," vol. 9, no. 4, pp. 657–665, December 2008.

[5] D. Scaramuzza and R. Siegwart, "Monocular omnidirectional visual odometry for outdoor ground vehicles," Springer Lecture Notes in Computer Science, Computer Vision Systems, 2008.

[6] J. Gluckman, S. K. Nayar, and K. J. Thoresz, "Real-time omnidirectional and panoramic stereo," in *In Proceedings of the 1998 DARPA Image Understanding Workshop*.   Morgan Kaufmann, 1998, pp. 299–303.

[7] S. Yi and N. Ahuja, "An omnidirectional stereo vision system using a single camera," in *ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition*.   Washington, DC, USA: IEEE Computer Society, 2006, pp. 861–865.

[8] J. Baldwin and A. Basu, "3d estimation using panoramic stereo," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 1, 2000, pp. 97 –100 vol.1.

[9] T. Gandhi and M. Trivedi, "Vehicle surround capture: Survey of techniques and a novel omni-video-based approach for dynamic panoramic surround maps," vol. 7, no. 3, pp. 293–308, September 2006.

[10] M. Schönbein, B. Kitt, and M. Lauer, "Environmental perception for intelligent vehicles using catadioptric stereo vision systems," in *In Proceedings of the 5th European Conference on Mobile Robots (ECMR)*, Örebro, Schweden, September 2011.

[11] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical applications," in *ECCV '00: Proceedings of the 6th European Conference on Computer Vision-Part II*.   London, UK: Springer-Verlag, 2000, pp. 445–461.

[12] T. Svoboda and T. Pajdla, "Epipolar geometry for central catadioptric cameras," *Int. J. Comput. Vision*, vol. 49, no. 1, pp. 23–37, 2002.

[13] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A flexible technique for accurate omnidirectional camera calibration and structure from motion," in *Computer Vision Systems, 2006 ICVS '06. IEEE International Conference on*, jan. 2006, p. 45.

[14] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed.   Cambridge University Press, ISBN: 0521540518, 2004.