AUTOMATIC STEREO CAMERA CALIBRATION IN REAL-WORLD ENVIRONMENTS WITHOUT DEFINED CALIBRATION OBJECTS

Thomas Marko & Wilfried Kubinger

Abstract

Using the images of a stereo camera, a 3D reconstruction of the real environment can be calculated, enabling a machine to see three-dimensionally. To calculate the 3D reconstruction, precise geometric dimensions of the stereo camera, called stereo camera parameters, are required. In the course of this work, three algorithms were realised that can calculate the stereo camera parameters from the images and motion of a stereo camera. None of the three is dependent on defined calibration objects. The algorithms are the calibration using epipolar geometry, relative world point error, and absolute world point error. These allow the calibration of a stereo camera with only the help of a measuring tape and the data sheet of the cameras. Continuous recalibration is also possible, where changes in the physical structure of the stereo camera are compensated during operation. The accuracy of the algorithms was determined by measuring the deviation of the 3D reconstruction from the real scene. The algorithms have been designed for use on a mobile robot to autonomously calibrate and recalibrate its stereo camera, constituting its visual sense.

Keywords: stereo camera; stereo camera parameters; automatic calibration; mobile robots

1. Introduction

Using the images of a stereo camera, a 3D reconstruction of the real environment can be calculated, enabling a machine to see three-dimensionally. 3D machine vision has many application examples in the field of automation. Mobile robots can map their environments in 3D, which has benefits for self-localisation or gripping and handling of objects. Self-driving cars equipped with stereo cameras like in [17] can measure the distance to obstacles such as other vehicles, natural objects, or buildings. To calculate a 3D reconstruction of a scene, precise geometric dimensions of the stereo camera, called stereo camera parameters, are required. Examples of stereo camera parameters are the focal length, the pixel size, the lens distortion, and the position and orientation of the cameras. The stereo camera parameters are nearly impossible to survey with conventional measuring tools because the cameras would have to be dismantled, a required accuracy is difficult to reach, and the effort would be immense. The solution is algorithms that can calculate the stereo camera parameters from the images of the same stereo camera. These algorithms are called stereo camera calibration algorithms and are the focus of this research paper.
2. Related Work

An already well-developed method is calibration with defined calibration objects like a printed chessboard pattern. Such an algorithm is provided by [7], [13], and [19]. Another approach is the calculation of stereo camera parameters without defined calibration objects in real-world environments. [6] and [9] use epipolar geometry and nonlinear optimisation or a kalman filter for continuous recalibration. In [21] the detection of vanishing points in the image of a mono camera is used to calibrate a roadside camera. The algorithm from [18] calculates the radial lens distortion from single images via vanishing points. In [14] bundle adjustment is applied, where the camera parameters are calculated from an image sequence containing many images of a moving stereo camera. The reprojection error and an iterated extended kalman filter is used in [2] for continuously calibrating a stereo camera from image sequences. The algorithm in [3] uses a combination of epipolar geometry, the reprojection error, and the trilinear constraint with an iterated extended kalman filter for continuous calibration.

3. Methodology

In the course of this work, two algorithms have been realised, that are based on the idea of the reprojection error. The developed equations are an alternative way to describe and use reprojection error for camera calibration and lead to a new possible application. The two algorithms are called relative world point error calibration and absolute world point error calibration throughout this paper. For parameter calculation nonlinear optimisation is used, whereby the stereo camera parameters are calculated by minimising a nonlinear error function. The theory of nonlinear optimisation can be consulted in [8].

3.1. Model definition

As a mathematical model of a stereo camera, a combination of two pinhole camera models and two radial distortion models are used. The overall model can be seen in figure 1.

![Mathematical stereo camera model](image)

Fig. 1. Mathematical stereo camera model

The projection centres of the two cameras are \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \), which are also the origins of the camera coordinate systems. The stereo camera parameters for the proposed mathematical model are the focal lengths \( f_1 \) and \( f_2 \), the pixel sizes of the optical sensors \([p_{x1}, p_{y1}]\) and \([p_{x2}, p_{y2}]\), the image main points \([c_{x1}, c_{y1}]\) and \([c_{x2}, c_{y2}]\), the parameters of the radial lens distortions of the two lenses \([k_{11}, k_{21}]\) and \([k_{12}, k_{22}]\), the position of camera 2 in regard to camera 1 \( \mathbf{t} = [t_x, t_y, t_z] \), and the orientation of camera 2 in regard to camera 1 in the form of three Euler angles \([\text{eul}_x, \text{eul}_y, \text{eul}_z]\) respectively in the form of an Euler rotation matrix \( \mathbf{R} \).

3.2. World point error

The world point error is the position error of reconstructed world points before and after a movement of the stereo camera. If the stereo camera parameters correspond to the physical structure of the real stereo camera, the reconstructed world points of a single world point before and after a movement of a stereo camera must have the same position. A stereo camera before and after a movement and with world point error can be seen in figure 2.
The two cameras with the optical centres $\mathbf{u}_1(t_1)$ and $\mathbf{u}_2(t_1)$ are the stereo camera at time $t_1$, and the two cameras with the optical centres $\mathbf{u}_1(t_2)$ and $\mathbf{u}_2(t_2)$ are the stereo camera after a movement $[\mathbf{R}_M \mathbf{t}_M]$ at time $t_2$. The world points are the intersection of the straights $\mathbf{s}_1$ and $\mathbf{s}_2$, that are shown in figure 1. $\mathbf{s}_1$ and $\mathbf{s}_2$ are defined by the two undistorted image points $\mathbf{i}_{1u}$ and $\mathbf{i}_{2u}$ and the stereo camera parameters. Consequently, the lens distortion must be removed from the known real image points $\mathbf{i}_1$ and $\mathbf{i}_2$. In this work, the radial lens distortion model is used because, according to [20], this model is more effective than other models and is suitable for use within a nonlinear optimisation process. In [10] an approximation of the inverse of the radial lens distortion equation is shown that is sufficiently precise in the case of small distortions. The lens distortion is removed with (1) and is only shown for one image point, because the procedure is the same for the others. $r_1$ is the radius from the image main point to the image point.

$$i_{1u} = i_1 \cdot \left(1 - k_{11} \cdot r_1^2 - k_{21} \cdot r_1^4\right)$$ (1)

The straights $\mathbf{s}_1$ and $\mathbf{s}_2$ can be described in parametric form and the world points can be calculated by analytical transformations. For the calculation of stereo camera parameters, the world point error is minimised with nonlinear optimisation, and for this purpose the world point error is described with four error angles that are shown in figure 3.

The four error angles, $\rho_1$, $\rho_2$, $\sigma_1$, and $\sigma_2$, are defined by the four direction vectors, $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$, and $\mathbf{h}_4$, and these vectors point against the reconstructed world points. For the calculation of $\mathbf{h}_1$ and $\mathbf{h}_4$, the reconstructed world points of the stereo camera at $t_1$ are transformed with the movement $[\mathbf{R}_M \mathbf{t}_M]$ into the coordinate systems of the stereo camera at $t_2$. 

![Fig. 2. Stereo camera at two points in time with motion and world point error](image1)

![Fig. 3. Error angles](image2)
3.3. Relative world point error calibration

In the case of the relative world point error calibration the parameters of the stereo camera at $t_2$ are calculated in regard to the parameters of the stereo camera at $t_1$. If a change occurs in the physical structure of the stereo camera between $t_1$ and $t_2$, the new parameters of the stereo camera at $t_2$ are calculated, and the quality of the stereo camera parameters, and thus the resulting 3D reconstruction, stay the same. For the relative world point error calibration, the optimisation problem is split into three parts, as shown in (2), (3), and (4). For the nonlinear optimisation, the Ceres Solver library [5] is used to minimise several error functions simultaneously, and splitting the world point error into three parts benefits the computation time.

$$
\min(f_1, c_x, c_y, k_{11}, k_{21}, R, t_M) \sum_{i=0}^{N} (\rho_{2i})^2 \quad \text{(2)}
$$

$$
\min(f_2, c_x, c_y, k_{12}, k_{22}, t_x, t_y, t_z, eul_x, eul_y, eul_z, R, t_M) \sum_{i=0}^{N} (\rho_{2i})^2 \quad \text{(3)}
$$

$$
\min(f_3, c_x, c_y, k_{11}, k_{21}, f_2, c_x, c_y, k_{12}, k_{22}, t_x, t_y, eul_x, eul_y, eul_z, R, t_M) \sum_{i=0}^{N} (\sigma_{2i} - \sigma_{1i})^2 \quad \text{(4)}
$$

$N$ is the number of reconstructed world points and thus the number of world point errors used for parameter calculation and inside the brackets, after the min()-operator, all parameters that are calculated in the course of the nonlinear optimisation are listed. In (4) $\sigma_{1i}$ is subtracted from $\sigma_{2i}$ because $\sigma_{1i}$ is dependent on the stereo camera parameters of $t_1$, and this part of the world point error is constant in the case of the relative world point error calibration.

3.4. Absolute world point error calibration

In the case of absolute world point error calibration, the stereo camera parameters of the stereo camera at $t_1$ and $t_2$ are calculated, but it is assumed that there are no changes in the physical structure of the stereo camera between $t_1$ and $t_2$.

$$
\min(f_1, c_x, c_y, k_{11}, k_{21}, f_2, c_x, c_y, k_{12}, k_{22}, R, R, t_M) \sum_{i=0}^{N} (|\rho_{1i}| + |\rho_{2i}| + |\sigma_{1i}| + |\sigma_{2i}|)^2 \quad \text{(5)}
$$

$N$ is the number of reconstructed world points, and thus the number of world point errors used for parameter calculation and inside the brackets, after the min()-operator, all parameters that are calculated in the course of the nonlinear optimisation are listed. The camera position $t$ is not calculated because this leads to unstable behaviour, and for the experimental evaluation in this work, $t$ is measured with a measuring tape. The absolute world point error calibration can improve the quality of the stereo camera parameters.

3.5. Movement fusion

The movement of the stereo camera is calculated along with the relative and absolute world point error calibration because if the movement is a constant in the nonlinear optimisation process, this causes instability. The fusion of the external movement and the movement calculated in the course of the minimisation of the relative and absolute world point error are realised with an exponential function as shown in (6).

$$
M_{\text{fuse}} = e^{\frac{|M_{\text{ext}} - M_{\text{int}}|}{\beta}} - 1 \quad \text{(6)}
$$

In (6) the absolute difference of the external integrated movement and the internal calculated movement is with the parameter $\beta$, the exponent of an e-function. $M_{\text{fuse}}$ is added to the relative and absolute world point error functions (error angles) and takes care of finding the local minima with the correct movement. The parameter $\beta$ controls the degree of influence of the external integrated movement data. (6) is calculated for all components of the movement data.

4. Realisation

For the use in automated systems, the calibration algorithms have been realised as an ROS (Robot Operating System) package (detailed information about ROS can be obtained in [15]). The inputs of the calibration package are the images of a stereo camera, algorithm parameters to control the algorithm features, an initial estimate of the stereo camera parameters, and, if required, external movement data of the stereo camera. The outputs of the package are the calculated stereo camera parameters and the related stereo image pair.
The algorithms for feature detection and matching are obtained from OpenCV [12]; for the nonlinear optimisation, the Ceres Solver [5] is used, and for all linear algebra calculations, the Eigen library [4] is used. For feature detection, it is possible to select between ORB [16] and SURF [1], and for feature matching, the brute force matcher or the FLANN based matcher from OpenCV is used. The feature pairs are filtered with a fundamental matrix filter like in [11], where a fundamental matrix is calculated with RANSAC and outliers are deleted. For the relative and absolute world point error calibration, feature quartets are needed where four features belong to the same world point. These are calculated by sorting the four feature pairs between the four images of the stereo camera before and after a movement. The resulting stereo camera parameters are stored in a YAML file.

5. Experimental Evaluation

For the experimental evaluation, a stereo camera consisting of two industrial cameras was used. The two industrial cameras are triggerable monochrome cameras with 2048 x 1536 pixel resolution, a global shutter, and a frame rate of 55fps. Both cameras were equipped with a zoom-objective, with a variable focal length from 4.4mm to 11mm.

The accuracy of the algorithms was tested by determining the deviation of the 3D reconstruction from the real-world scene. For the calculation of the 3D reconstruction, the ROS package stereo_image_proc was used, and the resulting 3D reconstruction was visualised in RVIZ. The test scene and a 3D reconstruction can be seen in figure 4.

![Test scene and 3D reconstruction](image)

Fig. 4. Test scene (left) and 3D reconstruction after the absolute world point error calibration (middle and right)

The deviation of the 3D reconstruction is the average of the differences of the measurement of five test distances with a measuring tape in the real scene and the measurements of the same five distances in the 3D reconstruction. The deviations in table 1 are the minimal, average, and maximal value after several calibrations (Matlab chess board calibration: 1 retry, Absolute world point error calibration: 29 retries, Relative world point error calibration: 11 retries). For each retry, a different stereo camera configuration was used (baseline between 70mm and 250mm, focal lengths between 4.4mm and 11mm, different camera orientations).

| Algorithm                        | Deviation before calibration [%] | Deviation after calibration [%] |
|----------------------------------|----------------------------------|---------------------------------|
| Matlab chess board calibration   | Uncalibrated                     | 2.75                            |
| Absolute world point error calibr| Uncalibrated                     | min.: 1.68; avg.: 5.55; max.: 13.04 |
| Relative world point error calibr| min.: 1.73; avg.: 4.82; max.: 10.04 | min.: 1.55; avg.: 5.25; max.: 9.61 |

Table 1. Deviation of the 3D reconstruction from the real-world scene

In the case of the chess board calibration and the absolute world point error calibration, a stereo camera with unknown parameters was calibrated. As initial guess of the stereo camera parameters for the absolute world point error calibration (which is the starting point of nonlinear optimisation), a rough guess of the focal lengths (according to the spec sheet of the objective), half of the resolution as image main point, zeroes for the lens distortion parameters, and the unit matrix as camera orientation $R$ are used. As already mentioned, the relative camera position $t$ is measured with a measuring tape and is integrated as constant, via the YAML parameter file, into the nonlinear optimisation. The relative world point error calibration was tested in three steps. Firstly, a stereo camera was calibrated with the absolute world point error calibration, a measuring tape, and the camera datasheet. Secondly, the relative world point error calibration was activated. Thirdly, a change in the physical structure of the stereo camera was performed.

6. Discussion

The average deviation of 5.55% after the absolute world point error calibration is higher than the 2.75% after the Matlab chessboard calibration, but in the case of the chessboard calibration, seventy stereo image pairs of a calibration object were used, and the absolute world point error calibration uses only two stereo image pairs of a real scene.
The average deviation of 5.25% after the compensation of a change of the physical structure of the stereo camera with the relative world point error calibration is almost the same as the average deviation of 4.82% before the change in the physical structure of the stereo camera. A compensation of changes in the camera orientations up to 3°, focal lengths up to 1 mm, and baseline up to 9 mm have been successfully tested. The algorithms were tested on a laptop with a 2.4 GHz Intel Core i7-4700MQ processor and 16 GB RAM; the calculation times were between 200 ms and 2 s and are dependent on the number of used features, the algorithm utilised, and the starting point of the nonlinear optimisation.

7. Conclusion and Outlook

With the realised ROS package, a stereo camera with unknown parameters can be calibrated with a measuring tape and the camera dataset, and the quality of the stereo camera parameters can be maintained during operation, even in the case of large changes in the physical structure of the stereo camera. The calibration is done by searching features of the same world point in the images of a moving stereo camera, and a subsequent minimisation of the world point errors of the reconstructed world points before and after a movement. The absolute world point error calibration calculates the intrinsic parameters of the pinhole camera model of both cameras, the relative extrinsic orientation of camera\textsubscript{2} regarding camera\textsubscript{1}, and the radial distortion of the two lenses. The relative world point error calibration maintains the quality of the same parameters as the absolute world point error calibration and additionally the relative extrinsic position of camera\textsubscript{2} regarding camera\textsubscript{1}, during operation.

The calibration procedure for the absolute world point error calibration in the current approach needs human intervention because the relative extrinsic camera position $t$ and the movement of the stereo camera must be measured with a measuring tape. The calibration process can be automated in combination with a mobile robot by using the odometry data as movement data for calibration, instead of the measurements with a measuring tape. The same is possible by attaching a stereo camera to the tool centre point of an industrial robot. Furthermore, the current approach uses the world point errors of only two stereo image pairs for parameter calculation. With the automation in combination with a mobile robot or an industrial robot, more than two stereo image pairs, with different movements in between, can be acquired. This enables the calculation of a larger number of world point errors in different movement situations and increases the amount of information for parameter calculation, which is beneficial for the accuracy and robustness of the calibration. Furthermore, the calculation of the relative camera position $t$ should also be possible robustly. To calculate the world point errors of more than two stereo image pairs, feature chains must be calculated. In the current approach feature quartets are used, where four features (from the four images of the two stereo image pairs) are the projection of the same world point. Feature chains contain an arbitrary number of features that are also the projection of the same world point. In future, we plan to test the absolute world point error calibration in combination with a mobile robot or an industrial robot.

The relative world point error calibration maintains the quality of all parameters of the used stereo camera model during operation. The compensation of changes of the focal lengths, of the relative extrinsic camera position $t$, and of the relative extrinsic camera orientation $R$ have been successfully tested. In future, we plan to evaluate the relative world point error calibration with an automatically variable stereo camera system, for instance equipped with autofocus cameras or with an automatically adjustable base line for high accuracies over wide depth ranges.

8. Acknowledgments

This project has been funded by MA23 - City of Vienna within the Project Call 16-02 "Photonics: Foundations and industrial applications".

9. References

[1] Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). Comput. Vis. Image Underst., vol. 110, no. 3, pp. 346–359.
[2] Dang, T. & Hoffmann, C. (2004). Stereo calibration in vehicles. IEEE Intelligent Vehicles Symposium, pp. 268–273.
[3] Dang, T., Hoffmann, C., & Stiller, C. (2009). Continuous stereo self-calibration by camera parameter tracking. IEEE Transactions on Image Processing, vol. 18, no. 7, pp. 1536–1550.
[4] Eigen.tuxfamily.org, (2018). Eigen. [Website]. Available from: <http://eigen.tuxfamily.org/index.php?title=Main_Page> [Accessed: 15.7.2018].
[5] Google, (2018). Ceres-Solver. [Website]. Available from: <http://ceres-solver.org/> [Accessed: 7.7.2018].
[6] Hansen, P., Alismail, H., Rander, P., & Browning, B. (2012). Online continuous stereo extrinsic parameter estimation. 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1059–1066.
[7] Heikkila, J. (2000). Geometric camera calibration using circular control points. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1066–1077.
[8] Jarre, F. & Stoer, J. (2004). Optimierung. Springer-Lehrbuch: Springer Berlin Heidelberg.
[9] Ling, Y. & Shen, S. (2016). High-precision online markerless stereo extrinsic calibration. 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1771–1778.
[10] Ma, L., Chen, Y., & L. Moore, K. (2004). Rational radial distortion models of camera lenses with analytical solution for distortion correction. I. J. Information Acquisition, vol. 1, no. 2, pp. 135–147.

[11] Mueller, G. R. & Wuensche, H. J. (2016). Continuous extrinsic online calibration for stereo cameras. 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 966–971.

[12] Opencv.org. (2018). OpenCV [Website]. Available from: <https://opencv.org/> [Accessed: 14.7.2018].

[13] Prokos, A., Kalisperakis, I., Peta, E., & Karas, G. (2012). Automatic calibration of stereocameras using ordinary chess-board patterns. ISPRS—International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 45–49.

[14] Rehder, E., Kinzig, C., Bender, P., & Lauer, M. (2017). Online stereo camera calibration from scratch. 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 1694–1699.

[15] ROS.org. (2018). ROS Tutorials. [Website]. Available from: <http://wiki.ros.org/ROS/Tutorials> [Accessed: 25.6.2018].

[16] Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). Orb: An efficient alternative to sift or surf. 2011 International Conference on Computer Vision, pp. 2564–2571.

[17] Weingerl, D., Kubinger, W., & Engelhardt-Novitzki, C. (2016). Preceding Vehicle Tracking in Stereo Images via 3D Feature Matching. Proceedings of the 26th DAAAM International Symposium, pp.0997-1003, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-07-5, ISSN 1726-9679, Vienna, Austria.

[18] Yang, S., Rong, J., Huang, S., Shang, Z., Shi, Y., Ying, X., & Zha, H. (2016). Simultaneously vanishing point detection and radial lens distortion correction from single wide-angle images. 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 363–368.

[19] Zhang, Z. (2000). A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330–1334.

[20] Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. [Technical Report]. Microsoft Research. Available at: <https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/> [Accessed: 28.6.2018].

[21] Zheng, Y., He, Z., Yang, W. G., & Zhang, X. F. (2014). An efficient and practical calibration method for roadside camera using two vanishing points. Proceedings 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), pp. 65–69.