

CAMERA CALIBRATION METHOD FOR HIGH ACCURACY STEREOVISION

S. Nedevschi, T. Marita, R. Danescu, F. Oniga, C. Pocol¹)

Abstract

This paper presents a camera calibration method for far-distance, high accuracy stereo-vision. There are many general purpose methods found in the literature for solving the common issues of the camera calibration: intrinsic parameters estimation (focal length and principal point) and extrinsic parameters estimation (the absolute position and orientation of the cameras relative to the world coordinate system). Their results are acceptable for monocular or near range stereo vision but are improper for far range applications. For a high accuracy stereovision system the most critical parameters are the relative orientation of the two cameras positioned on the stereo-rig. Experiments proved that even a few seconds drift of one of the relative camera angles can lead to disastrous consequences in the whole stereo-vision process: incorrect epipolar lines estimations, and consequently lack of reconstructed 3D points. Therefore we propose an extrinsic parameters calibration method able to give very accurate extrinsic parameters of the stereo-rig. The main idea is to use a dedicated calibration scene, having the sizes comparable with the stereo-vision application's working range, with some control points placed in known 3D positions. The extrinsic parameters are estimated by minimizing the image projection error of the 3D control points relative to their detected 2D image positions. The obtained results are allowing the reconstruction of 3D features up to 100 m with an error bellow 1%.

1. Introduction

Camera calibration deals with the estimation of the camera parameters: intrinsic and extrinsic parameters [1]. The intrinsic parameters are describing the cameras internal optical and geometrical characteristics and are the focal length and the principal point positions, usually measured in pixels. The extrinsic parameters are describing the position and orientation of the camera in a world coordinate system. Both types of parameters must be known in order to perform any kind of vision based measurements. If a monocular system is used the system measuring capabilities are more reduced and consequently the accuracy of the required parameters is less stringent. If a stereo-vision

¹ Tech. Univ. of Cluj-Napoca, Computer Science Department, 15 C. Daicoviciu Str, 400020 Cluj-Napoca, Romania

system is used, a full 3D reconstruction of the scene is possible, and a higher accuracy is compulsory, depending on the performance requirements of the stereovision system.

For a stereovision system the parameters can be classified in the following categories:

- Internal parameters of the stereo-rig: intrinsic parameters and relative position and orientation of the two cameras (relative extrinsic parameters);
- Absolute extrinsic parameters of the stereo-rig: position and orientation of the cameras (or of a reference point of the stereo-rig – usually the left camera coordinate system) relative to a world coordinate system in which the measurements are reported.

The quality of the internal parameters is influencing the accuracy of the epipolar lines estimation [1] which is essential in the stereo correlation process. Their wrong estimation can lead to a lack of correlated points or to false correlation which is disastrous to the whole stereo reconstruction process. The quality of the absolute extrinsic parameters is influencing only the accuracy of the 3D measurements relative to the world coordinate system. The relative position and orientation of the two cameras depends on the absolute extrinsic parameters (is their vector difference).

Regarding the intrinsic parameters calibration there are many general purpose methods found in the literature [2,3,4,5]. Most of them are estimating the parameters by minimizing the projection error of a set of control points from a calibration object/pattern with known structure against the detected 2D images of the same control points. Multiple views of the calibration object are taken and the accuracy of the results usually increases with the total number of considered control points ($nr_control_points/calibration_object \times number_of_views$). Their accuracy is satisfactory for most of the vision based measurements applications. A special note must be mentioned for the Bouguet method available in the Caltech Calibration Toolbox [5], which was intensively tested and used for the estimation of the intrinsic parameters in our application. Despite the difficult control points extraction procedure, if the proper calibration methodology is fulfilled, accurate results can be obtained even for stereo-vision requirements.

Regarding the extrinsic parameters estimation, the principle is the same: minimizing the projection error of some 3D control points with known positions in the world coordinate system (measurement coordinate system). General purpose methods [5,6,7] are using the same calibration object as for the intrinsic parameters estimation, and in this case the reference point of the calibration pattern will be the origin of the world coordinate system. This approach can be satisfactory for monocular applications or near range applications (e.g. indoor visual navigation of robots), but is completely unsuited for far range stereo-vision. The main reason relies in the fact that the calibration object and the distances implied in the calibration procedure are much smaller than the working range of such an application, and therefore small estimation errors which were acceptable in the near range could

increase dramatically with the distance. Therefore the usages of a calibration scene with sizes comparable with the working range of the stereovision application have been proved a reliable solution. In [8,9] such methods used in stereo-vision based vehicle guidance are presented. Both of them are using an off-line extrinsic parameters calibration procedure, using as control points some painted markers on a flat road surface, of few meters in length. In [9] a supplementary method for on-line adjustment of the extrinsic parameters is presented. The method uses some painted markers on the car's hood and refines the calibration parameters using the same principle (minimization of the projection error of the markers) around the off-line values.

In this paper we present a method for estimating the extrinsic parameters of the cameras suited for far distance high accuracy stereo-vision. As control points 'X'-shaped targets are placed vertically in a calibration scene, built on a flat surface, up to 50m in depth. The 3D coordinates of the targets' are measured in the world coordinate system. Each camera is calibrated individually (so the method can be applied also for mono systems). The 2D image projections of the targets' central points are detected automatically and the projection error of their 3D coordinates are minimized against the position and orientation of the camera in the world coordinate system using the Gauss-Newton method. The obtained results have been proved very accurate regarding both the absolute extrinsic parameters of each camera individually and the relative extrinsic parameters in the stereo-configuration, allowing a very accurate stereo reconstruction procedure.

2. Camera model

There are two coordinate systems implied in any projection procedure from a 3D scene point to the image plane (fig. 1):

- The world coordinate system (WRF), denoted $O - XYZ$;
- The camera coordinate system (CRF), denoted $C - X_c Y_c Z_c$.

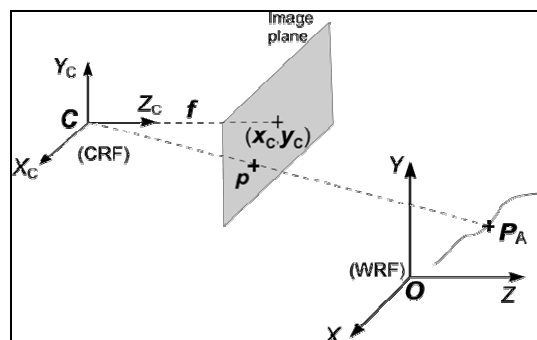


Fig. 1. Camera model.

The intrinsic parameters of the camera are the *principal point's* position (x_c, y_c) [pixels], and the *focal length* of the camera f [pixels]. The intrinsic parameters are assumed to be correctly

computed, using [5]. The extrinsic parameters are the position of the camera's optical center C in the WRF, represented by the *translation vector* $\mathbf{T}_C = [X_{cw}, Y_{cw}, Z_{cw}]^T$, and the orientation of the camera coordinates systems in the WRF, captured by the *rotation matrix* \mathbf{R}_C (1). Therefore we have a total number of 12 unknowns from \mathbf{T}_C and \mathbf{R}_C .

$$\mathbf{R}_c = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \quad (1)$$

3. Extrinsic parameters calibration

3.1. Automatic detection of the targets 2D image projection

There are two main aspects that influence the quality of the obtained extrinsic parameters: the 3D control points should be uniformly spread over the calibration field (fig. 2) and their 3D coordinates should be measured as accurate as possible and the projection of the control points onto the image plane must be detected with sub-pixel precision.

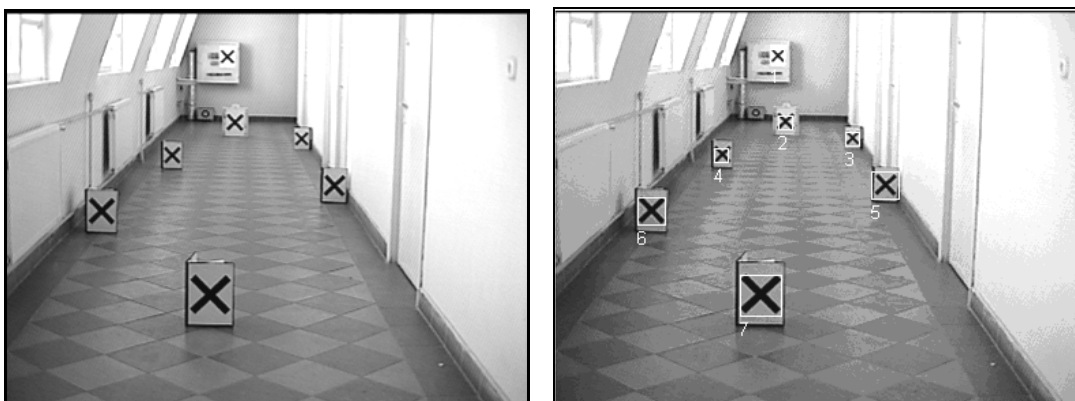


Fig. 2. *Left: calibration scenario - the 3D control points are situated in the middle of the 'X'-shaped targets; Right: Detected targets position by pattern matching*

The 'X' shape of the targets was chosen due to some major advantages in the automated detection process: firstly it has a low rate of false targets detections because most of the structures in our 3D scenarios are vertical or horizontal-like structures ('+') while the 'X' shape is based on oblique features; secondly the center of the target can be easily computed; thirdly, the centers of the vertical 'X'-shaped targets are more easier to detect than lines or other markers on the road [8,9], which are poorly visible at long distances due to the perspective projection effects, especially when the cameras are mounted at low pitch angles.

Detection of the possible targets is performed by model matching at several scales. The models used are obtained from an original ‘X’-shaped pattern, used to print the targets, by downscaling (fig. 3):



Fig. 3. The models used for targets detection by pattern matching.

To detect the sub-pixel position of the targets center, the image area bounded by the matched model-square associated to the current target (fig. 2-right) is zoomed-in with a scale factor of 20 using bi-cubic interpolation (fig. 4.a). To detect the center position of the target along the X axis, the sum of the pixel values along each column is computed (fig. 4.b). The maximum of these sums is considered (in the center of the target the number of light pixels along the row and the column is maximal). To minimize the influence of noise and image discrete nature, instead of considering exactly the position of the maximum we take the mid point of the interval with values above 90% from the maximum. The computed position is now mapped back by downscaling and translation to the original image coordinates. To detect the position of the center of the target along the Y axis the procedure is similar to the one used for the X axis, except that it’s applied along the vertical direction (fig. 4.c).

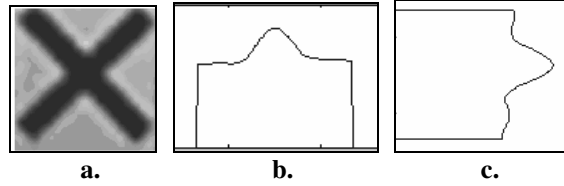


Fig. 4. a. Zoomed ‘X-shape’ image. b. The sum of intensities along the columns for each row. c. The sum of intensities along the rows for each column.

3.2. Extrinsic parameters computation

Considering a known 3D point $\mathbf{P}_A(X_A, Y_A, Z_A)$ and its image projection $\mathbf{p}(x_i, y_i)$ (fig. 1), due to perspective projection [1], the following relations are available (2), or written as in (3):

$$\begin{cases} x_i - x_c = -f \frac{r_{11}(X_A - X_{cw}) + r_{21}(Y_A - Y_{cw}) + r_{31}(Z_A - Z_{cw})}{r_{13}(X_A - X_{cw}) + r_{23}(Y_A - Y_{cw}) + r_{33}(Z_A - Z_{cw})} \\ y_i - y_c = -f \frac{r_{12}(X_A - X_{cw}) + r_{22}(Y_A - Y_{cw}) + r_{32}(Z_A - Z_{cw})}{r_{13}(X_A - X_{cw}) + r_{23}(Y_A - Y_{cw}) + r_{33}(Z_A - Z_{cw})} \end{cases} \quad (2)$$

$$\begin{cases} f_1(u) = 0 \\ f_2(u) = 0 \end{cases} \quad (3)$$

where u is a vector of the 12 unknowns $(X_{cw}, Y_{cw}, Z_{cw}, r_{ij})$, for $i, j = 1, 2, 3$.

For each 3D known point there are two equations involving the 12 unknowns. The 9 unknowns representing the rotation matrix have only 3 freedom degrees (represented by the 3D angles of rotation), thus it is necessary to add several constraints that model the dependencies between the rotation coefficients. These constraints are obtained easily considering the fact that the rotation matrix is orthogonal [1], and therefore six equations can be supplementary added:

$$\begin{cases} f_{R1}(u) = r_{11} \cdot r_{11} + r_{12} \cdot r_{12} + r_{13} \cdot r_{13} - 1 = 0 \\ f_{R2}(u) = r_{21} \cdot r_{21} + r_{22} \cdot r_{22} + r_{23} \cdot r_{23} - 1 = 0 \\ f_{R3}(u) = r_{31} \cdot r_{31} + r_{32} \cdot r_{32} + r_{33} \cdot r_{33} - 1 = 0 \\ f_{R4}(u) = r_{11} \cdot r_{21} + r_{12} \cdot r_{22} + r_{13} \cdot r_{23} = 0 \\ f_{R5}(u) = r_{11} \cdot r_{31} + r_{12} \cdot r_{32} + r_{13} \cdot r_{33} = 0 \\ f_{R6}(u) = r_{21} \cdot r_{31} + r_{22} \cdot r_{32} + r_{23} \cdot r_{33} = 0 \end{cases} \quad (4)$$

By using a number $n \geq 3$ of 3D control points, from (3) and (4) a non-linear system of $2n+6$ equations is built, where the first $2n$ equations are obtained by applying (3) for each control point, and the last 6 equations are the constraints defined in (4):

$$F(u) = 0 \quad (5)$$

To solve this system, the Gauss-Newton iterative method was used. This method starts from an initial random solution (the constraint that the initial solution must be close to the real one is not required) and, by applying iteratively correction steps, makes the initial solution to converge to the real one. Suppose that, after a number of correction steps, the current solution is u_i and, by applying another correction step, the new solution is $u_{i+1} = u_i + du$, than the correction vector du is obtained by solving the system:

$$\mathbf{J} \cdot du = F \quad (6)$$

where \mathbf{J} is the Jacobian matrix associated to equation system $\mathbf{F}(u) = 0$. To solve the over-determined system (6) a least-squares method was used:

$$du = u_{i+1} - u_i = -(\mathbf{J}(u_i)^T \cdot \mathbf{J}(u_i))^{-1} \cdot \mathbf{J}(u_i)^T \cdot \mathbf{F}(u_i) \quad (7)$$

The decision to stop the correction process after a number of iterations is taken when the norm of the correction vector du is smaller than a threshold.

4. Results

A stereo-vision system was built in general camera configuration. The intrinsic parameters were calibrated using the Bouguet method [5]. The extrinsic parameters were calibrated with the proposed method. To assess the accuracy of the detected parameters, a reference scene was built using 'X'-shaped calibration targets placed in measured 3D positions (the bird eye view of the test scenario can be seen in figure 5.a.). Their 2D image coordinates were detected and matched with sub-pixel accuracy. Using a stereo-reconstruction algorithm for general camera configuration [10] their 3D coordinates were reconstructed and compared with the measured ones. The absolute errors' plots of the reconstructed coordinates are presented in figure 5.b-d. As it can be seen from the error plots, the reconstruction error of the test targets' 3D coordinates placed in a range of 10 m in width and 40m in depth is below 4 cm for X, below 1.5 cm for Y and below 30 cm for Z, which denotes a maximum relative error below 1% for all tree coordinates.

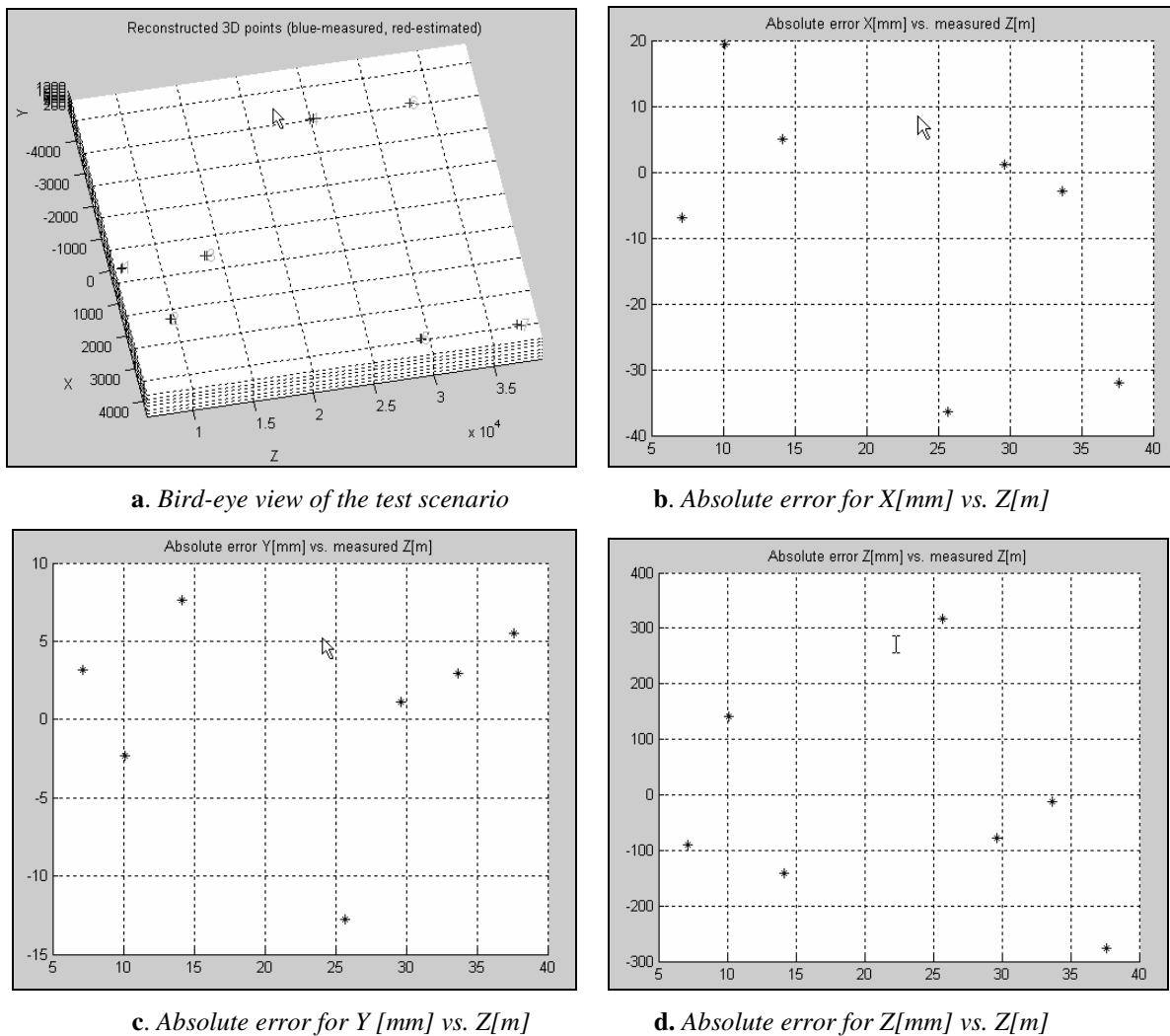


Fig.5. Error plots for the reconstructed test scenario.

5. Conclusions

A dedicated calibration method for far distance stereo-vision was developed. The method performs the estimation of the camera's extrinsic parameters (translation vector and rotation matrix relative to a world coordinate system) with high accuracy. This denotes also very accurate relative extrinsic parameters of the cameras' inside the stereo-rig, which allows a precise estimation of the epipolar lines (0 pixel drift). Therefore the method is suited for the calibration of any stereo-vision system used for far range 3D reconstruction as outdoor robot vision applications or vision based driving assistance systems.

6. References

- [1] E. Trucco, "Introductory techniques for 3D Computer Vision", *Prentice Hall*, 1998.
- [2] Tsai, R. Y.(1987), "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses". *IEEE Journal of Robotics and Automation*, RA-3(4)/1987: pp.323-344.
- [3] Zhang Z (1999), "Flexible Camera Calibration By Viewing a Plane From Unknown Orientations." *International Conference on Computer Vision (ICCV'99)*, Corfu, Greece, September 1999, pp. 666-673
- [4] Heikkila, J.; Silven, O. (1997), "A four-step camera calibration procedure with implicit image correction," *Proc. IEEE Computer Society Conf.*, 1997 pp.1106 –1112.
- [5] Jean-Yves Bouguet, Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/bouguetj/calib_doc/.
- [6] N. Kämpchen, , U. Franke, R. Ott, "Stereo vision based pose estimation of parking lots using 3D vehicle models", *IEEE Intelligent Vehicle Symposium*, 2002, pp. 459- 464.
- [7] S. Nedeveschi, T. Marita, M. Vaida, R. Danescu, D. Frentiu, F. Oniga, C. Pocol, "Camera Calibration Method for Stereo Measurements", *IEEE-TTTC International Conference on Automation, Quality and Testing, Robotics*, Cluj-Napoca, Romania, May 2002, pp.111-116.
- [8] S. Ernst, C. Stiller, J. Goldbeck, C. Roessig, "Camera Calibration for Lane and Obstacle detection", *IEEE Intelligent Transportation Systems*, 1999, pp. 356-361.
- [9] A. Broggi, M. Bertozzi, A. Fascioli "Self-Calibration of a Stereo Vision System for Automotive Applications", *IEEE Conf. on Robotics and Automation*, Seoul, Korea, May 2001, vol.4, pp. 3698-3703.
- [10] S. Nedeveschi, R. Schmidt, T. Graf, R. Danescu, D. Frentiu, T. Marita, F. Oniga, C. Pocol, "High Accuracy Stereo Vision System for Far Distance Obstacle Detection", in *Proc. of IEEE Intelligent Vehicles Symposium*, Parma, Italy, June 14-17, 2004, pp.161-166.