

# Camera calibration under optimal conditions

Carlos Ricolfe-Viala\* and Antonio-Jose Sanchez-Salmeron

*Instituto de Automática e Informática Industrial, Universitat Politècnica de València, Camino de Vera s/n, Valencia, 46022, Spain*

*\*cricolfe@isa.upv.es*

**Abstract:** Different methods based on photogrammetry or self-calibration exist to calibrate intrinsic and extrinsic camera parameters and also for data pre- and post-processing. From a practical viewpoint, it is quite difficult to decide which calibration method gives accurate results and even whether any data processing is necessary. This paper proposes a set of optimal conditions to resolve the calibration process accurately. The calibration method uses several images of a 2D pattern. Optimal conditions define the number of points and the number of images to resolve the calibration accurately, as well as positions and orientations from where images should be taken.

©2011 Optical Society of America

**OCIS codes:** (150.0155) Machine vision optics; (150.1135) Algorithms; (150.1488) Calibration.

---

## References and links

1. P. D. Lin and C. K. Sung, "Comparing two new camera calibration methods with traditional pinhole calibrations," *Opt. Express* **15**(6), 3012–3022 (2007).
2. M. Bauer, D. Griebbach, A. Hermerschmidt, S. Krüger, M. Scheele, and A. Schischmanow, "Geometrical camera calibration with diffractive optical elements," *Opt. Express* **16**(25), 20241–20248 (2008).
3. K. S. Choi, E. Y. Lam, and K. K. Y. Wong, "Automatic source camera identification using the intrinsic lens radial distortion," *Opt. Express* **14**(24), 11551–11565 (2006).
4. J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(10), 965–980 (1992).
5. S. Koppurapu and P. Corke, "The effect of noise on camera calibration parameters," *Graph. Models* **63**(5), 277–303 (2001).
6. J. Lavest, M. Viala, and M. Dhome, "Do we really need accurate calibration pattern to achieve a reliable camera calibration," in *Proceedings of European Conference on Computer Vision 1*, (1998) 158–174.
7. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).
8. D. Huynh, R. Hartley, and A. Heyden, "Outlier correction in image sequences for the affine camera," in *Proceedings of 9th IEEE International Conference on Computer Vision 1*, (2003) 585–591.
9. G. Stewart, "Perturbation theory for the singular value," University of Maryland, (*Tech. Report TR90 -124*, 1990).
10. R. Hartley, "In defence of the eight point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(6), 580–593 (1997).
11. R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-self TV camera lenses," *IEEE J. Robot. Autom.* **RA-3**, 323–344 (1997).
12. J. Salvi, X. Armangué, and J. Batlle, "A Comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognit.* **35**(7), 1617–1635 (2002).
13. W. Sun and J. Cooperstock, "An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques," *Mach. Vis. Appl.* **17**(1), 51–67 (2006).

---

## 1. Introduction

In the camera calibration process, it is essential to consider the quality of results, which obviously depends on data accuracy and the computed model. Some studies have been made on calibration errors committed due to precision of the computed model [1]. Also, lens distortion model is included in the calibration process to improve the computed camera model [2,3]. Distortion model can be calibrated alone or together with the pin-hole model. To calibrate the distortion model alone, geometric invariants such straight lines, vanishing points,

images of a sphere or correspondences between points in different images from multiple views are used. To calibrate the distortion model together with the pin-hole model, the pin-hole calibration process is extended with lens distortion model parameters. In this case, since the distortion is coupled with the intrinsic and extrinsic camera parameters, methods which extend the calibration of the pin-hole model to obtain the camera distortion parameters result in high errors on the internal parameters. If high distorted images are used, calibrate pin-hole and lens distortion models together may result in an absurd solution [4]. Other studies define errors from imprecise measurement of the image plane or calibration template [5–7] (to cite a few). The calibration process is also affected by erroneous association of one point in reality with a point in the image. Some authors use statistical tools to detect these anomalies [8]. Computing errors due to instabilities of the mathematical tool should also be taken into account [9]. In these cases, data normalization improves the robustness of the algorithm and gives more accurate results [10]. Another significant case is the number of parameters of the computed camera model. A very complex model can complicate the algorithm and the obtained result will not be much better than that obtained with a simpler model. A very complicated model may produce instabilities in the process of questing and produce absurdities [4].

The state of the art of calibration provides some guide about efficiency of camera calibration in all situations. The Tsai method [11] represents a classical calibration process based on the measurements of the 3D points in the template taking a fixed reference. This method has been widely used in the past. Salvi [12] compares the calibration methods developed between 1982 and 1998, with the Tsai method showing better performance, despite the fact that it requires high precision in input data. On the other hand, Zhang's method [7], which is not included in Salvi's comparison [12], represents a new era in the camera calibration process. This method uses images of a 2D template taken from different camera positions and orientations. In this way, the advantages of camera self-calibration are combined with the point coordinate-based calibration. This calibration method is highly flexible, since the camera and the template can be moved freely and also as many images as are required can be taken without measuring any position of the template. Sun [13] compares the Tsai method with Zhang's method. On one hand, Tsai produces a precise estimation of camera parameters if the input data have not been corrupted. Since 100 points in the template are necessary and the coordinates should be referred to a fixed origin, careful design of the calibration template and a very accurate coordinate measurement are necessary. Nevertheless, errors are too easily committed and in practice these results are not as accurate as expected, as shown by Sun [13]. On the other hand, Zhang's method based on a 2D template requires neither a special design nor precise point measurement. Sun obtains camera calibration with a hand-made template, and better results are computed using Zhang's method. Furthermore, the sensibility of the calibration algorithm to errors in the measures can be improved by increasing the spotted number in the template, by simply printing a chessboard with more corners. The results of the comparison show the flexibility and adaptability of Zhang's method, as it can be performed on any scene. Considering the results of these two authors, Zhang's method is used as a reference for camera calibration.

Camera calibration is a two step process where first a linear algebraic approximation is followed by a non linear searching. Since camera parameters are coupled, non linear searching is poorly conditioned and local solution is reached easily. Consequently, linear algebraic approximation is crucial to avoid divergences to local solution with the non linear searching. This paper proposes the optimum conditions from the viewpoint of number of points, number of images and location of the camera for taking the images, to improve the camera calibration method using 2D templates. These definitions obtain a calibration process with a linear algebraic approximation well-conditioned which overcomes existing calibration methods.

## 2. Optimal conditions for camera calibration

To calibrate the pin-hole camera model, Zhang [7] describes a method based on the homographies between a planar calibration pattern and its images from several camera locations. For each homography two homogeneous equations arise as:

$$\begin{bmatrix} v_{11}^T - v_{22}^T \\ v_{12}^T \end{bmatrix} \mathbf{b} = 0 \quad (1)$$

where

$$v_{ij}^T = [h_{i1}h_{j1} \quad h_{i1}h_{j2} + h_{i2}h_{j1} \quad h_{i2}h_{j2} \quad h_{i3}h_{j1} + h_{i1}h_{j3} \quad h_{i3}h_{j2} + h_{i2}h_{j3} \quad h_{i3}h_{j3}] \quad (2)$$

$h$  represents elements of the homography  $\mathbf{H}$  and  $\mathbf{b} = [b_{11}, b_{12}, b_{22}, b_{13}, b_{23}, b_{33}]$ , where  $b_{ij}$  represents the element  $ij$  of the symmetric matrix  $\mathbf{K}^T \cdot \mathbf{K}^{-1}$ .  $\mathbf{K}$  contains intrinsic parameters of pin-hole model. If  $m$  images of the calibration object are observed,  $m$  equations such as Eq. (1) arise, giving  $\mathbf{V} \cdot \mathbf{b} = \mathbf{0}$ , where  $\mathbf{V}$  is a  $2m \times 6$  matrix. At least three images are necessary  $m \geq 3$  in order to obtain a unique solution. The closed-form solution is given by the eigenvector of  $\mathbf{V}^T \mathbf{V}$  associated with the smallest eigenvalue. Once  $\mathbf{b}$  is estimated, intrinsic camera parameters can be computed. When  $\mathbf{K}$  is known, extrinsic parameters for each image are computed when the corresponding homography is known. See [7] for details.

Optimal conditions for camera calibration are defined to reduce errors when computing vector  $\mathbf{b}$ . Since vector  $\mathbf{b}$  is formed with intrinsic parameters, optimal conditions reduce errors of intrinsic parameters. However, since extrinsic parameters are computed from intrinsic parameters, both intrinsic and extrinsic parameters are improved if camera is calibrated under optimal conditions.

### 2.1 Camera calibration scene

To define the optimal positions for image capture, a calibration scene is defined.  $(\mathbf{o}, \{x_w, y_w, z_w\})$  is 3D scene coordinate system and  $(\mathbf{o}', \{x_c, y_c, z_c\})$  is camera coordinate system located as shown in Fig. 1. The centre of the template is situated in the origin of coordinates of the scene and the camera is always translated with a negative z-coordinate. The orientation of the camera is defined using the position. This means that the camera optical axis crosses the centre of the template always as shown on Fig. 1. Moreover,  $X_c$  axis of the camera coordinate system is always parallel to plane  $X_w$ - $Y_w$  of the scene. From a practical point of view, the template is located on the floor of the calibration scene and the camera is located on a tripod, which keeps the upper border of the image parallel to the floor. The origin of the scene coordinate system is the template centre. To relate orientation with position of the camera, two rotations are defined. First camera and scene coordinate systems coincide. Then, the camera is rotated at angle  $\beta$  with respect to the  $Z_c$  axis as shown in Fig. 1 (centre). Second, in relation to the  $X_c$  axis, the camera is rotated an angle  $\alpha$  as shown on Fig. 1 (right). These two rotations keep the  $X_c$  axis parallel to plane  $X$ - $Y$  of the calibration scene. These angles are a function of the camera position:

$$\cos \beta = \frac{t_y}{m_1} \quad \sin \beta = \frac{t_x}{m_1} \quad \cos \alpha = \frac{|t_z|}{m_2} \quad \sin \alpha = \frac{m_1}{m_2} \quad m_1 = \sqrt{t_x^2 + t_y^2} \quad m_2 = \sqrt{t_x^2 + t_y^2 + t_z^2} \quad (3)$$

where  $m_1$  represents the module of camera translation in plane  $X_w$ - $Y_w$ , and  $m_2$  is the distance of the camera to the origin. The origin of coordinates of the image is in the centre and will therefore be pixels with negative coordinates. Another supposition refers to the focal lengths  $\alpha_u$  and  $\alpha_v$  in image axes. Pixels are supposed squared and therefore  $\alpha_u$  and  $\alpha_v$  are equal to  $\alpha$ .

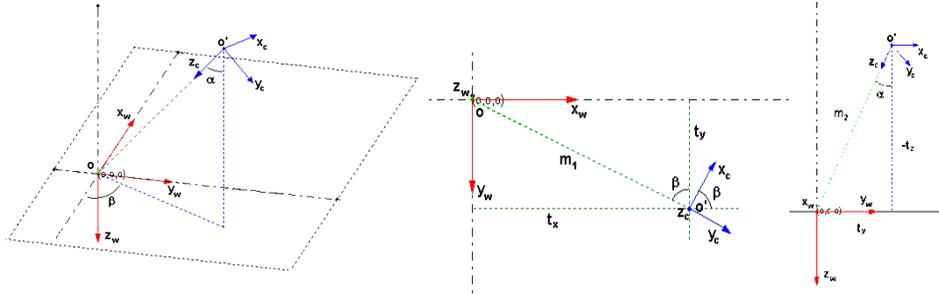


Fig. 1. (left) Scene and camera coordinate systems. Orientation is computed starting with camera location. Optical axis goes through origin of coordinates in the scene (middle) First rotation angle (right) second rotation angle.  $x_c$  remains parallel to the X-Y plane of the scene.

## 2.2. Condition of camera calibration

As known in a general framework, a matrix  $M$  has  $r$  eigen vectors and  $r$  eigen values, where  $r$  is the rank of matrix  $M$ . Eigen vectors are orthogonal to one another. Eigen vectors modules represent the number of vectors in the matrix  $M$  oriented in the direction defined by each eigen vector. The module of one eigen vector increases if there are more vectors in matrix  $M$  oriented in this eigen vector direction. The ratio between the bigger module and the smaller module is called the condition number of the matrix. To obtain a well-conditioned system, the condition number should be one. This means there are vectors in  $M$  which cover all  $R$ -dimensional space. Therefore, if a parameter vector is estimated with this matrix  $M$ , information from the data affects all dimensions of the parameter vector. In the event of a badly conditioned matrix, some parameters are not influenced by the information contained in matrix  $M$  and therefore such estimation will be erroneous.

If this theory is particularized to linear camera calibration, the camera parameters are in the eigen vector associated with the smaller eigen value of matrix  $V$  defined by expression (1). Matrix  $V$  is formed from the elements of homographies  ${}^lH$ . Since the condition number of matrix  $V$  should be one, all vectors of matrix  $V$  should have the same module, and they should be orthogonal to one another. Vectors of  $V$  depend on the homographies  ${}^lH$ , whereas  ${}^lH$  depends on the locations from which images are taken. Therefore, vectors of  $V$  depend indirectly on positions in which the camera is located. Consequently, we can say that locations from where images are taken influence the condition of matrix  $V$ . As a result, to obtain a well-conditioned matrix  $V$  several locations for the camera should be defined.

Dimension of  $V$  is  $2m \times 6$ . Therefore the minimum number  $m$  of homographies to resolve the system is  $m \geq 3$ . Within this framework, a well-conditioned matrix  $V$  will arise with vectors  ${}^l v_{ij}$ , which are orthogonal to one another, with equal modules. Sub index  $ij$  represents vector  $v$  formed from the columns  $i$ -th and  $j$ -th of homography  $l$ . Thus, homographies forming a set of vectors  ${}^l v_{ij}$  orthogonal to one another and with identical modules are required. In consequence care must be taken when dealing with homography elements.

To obtain a well-conditioned matrix  $V$ , following expressions should be true for all  $i, j, l$ .

$$\left( {}^l v_{11}^T - {}^l v_{22}^T \right) \cdot {}^l v_{12} = 0 \quad (4)$$

$$\left( {}^l v_{11}^T - {}^l v_{22}^T \right) \left( {}^l v_{11} - {}^l v_{22} \right) - {}^l v_{12}^T \cdot {}^l v_{12} = 0 \quad (5)$$

First expression represents an orthogonal condition and the second stands for module equality.  $i, j$ , have values 1 and 2 since they correspond with the two left-hand columns of the homography  $l$ . The number of homographies  $l$  should be at least 3. If 3 homographies are established whose vectors are orthogonal and with equal modules, a well-conditioned matrix  $V$  will be obtained with minimum information. Restrictions for three homographies:

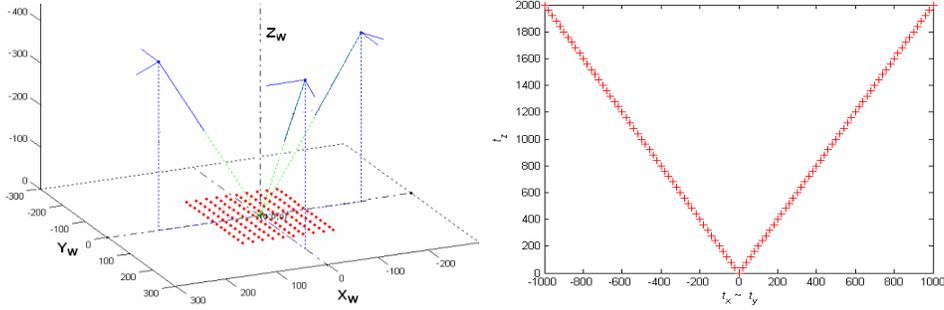


Fig. 2. Optimal positions for taking images of the planar template. Camera is moved along X or Y scene coordinates axis system and its altitude in the Z axis is defined with the expression (17).

$$\left({}^1v_{11}^T - {}^1v_{22}^T\right) \cdot {}^1v_{12} = 0 \quad (6)$$

$$\left({}^1v_{11}^T - {}^1v_{22}^T\right) \left({}^1v_{11} - {}^1v_{22}\right) - \left({}^1v_{12}^T \cdot {}^1v_{12}\right) = 0 \quad (7)$$

$$\left({}^2v_{11}^T - {}^2v_{22}^T\right) \cdot {}^2v_{12} = 0 \quad (8)$$

$$\left({}^2v_{11}^T - {}^2v_{22}^T\right) \left({}^2v_{11} - {}^2v_{22}\right) - \left({}^2v_{12}^T \cdot {}^2v_{12}\right) = 0 \quad (9)$$

$$\left({}^1v_{11}^T - {}^1v_{22}^T\right) \cdot {}^2v_{12} = 0 \quad (10)$$

$$\left({}^1v_{11}^T - {}^1v_{22}^T\right) \left({}^2v_{11} - {}^2v_{22}\right) = 0 \quad (11)$$

$${}^1v_{12}^T \cdot \left({}^2v_{11}^T - {}^2v_{22}^T\right) = 0 \quad (12)$$

### 2.3. Locations for image capturing

Analysing expressions from Eqs. (6)–(12) they are satisfied when a camera is located with null coordinates  ${}^l t_x$  or  ${}^l t_y$ . If  ${}^l t_x$  is zero,  ${}^l t_y$  and  ${}^l t_z$  will not be zero. Also, if  ${}^l t_y$  is zero,  ${}^l t_x$  and  ${}^l t_z$  will not be zero.  ${}^l t_x$  or  ${}^l t_y$  are set different to zero moving the camera along the X or Y scene axis. The camera position in the Z axis of the scene  ${}^l t_z$  is computed with the following expressions depending on whether the camera has been located with coordinate  ${}^l t_x$  or  ${}^l t_y$  different to zero:

$${}^l t_z = \pm \sqrt{-\frac{2 \cdot {}^l t_x^2 \left(-1 - 2 \cdot {}^l t_x^2 - 2 \cdot {}^l t_x^4 + \sqrt{(-3 - 4 \cdot {}^l t_x^2 + 12 \cdot {}^l t_x^4)}\right)}{1 + 3 \cdot {}^l t_x^2 + {}^l t_x^4}} \quad {}^l t_z = \pm \sqrt{-\frac{2 \cdot {}^l t_y^2 \left(-1 - 2 \cdot {}^l t_y^2 - 2 \cdot {}^l t_y^4 + \sqrt{(-3 - 4 \cdot {}^l t_y^2 + 12 \cdot {}^l t_y^4)}\right)}{1 + 3 \cdot {}^l t_y^2 + {}^l t_y^4}} \quad (13)$$

From a practical point of view, the camera is moved along one of the axes representing the X or Y axis of the coordinate system of the scene and then the altitude of the camera in the Z axis is defined using the expression (18). After this, the camera must be oriented bearing in mind that the optical axis passes through the origin of coordinates in the scene. The camera must be moved along the X or Y axis exactly the same distance. It is important to note that expressions (13) do not depend on intrinsic parameters. Therefore camera locations do not depend on camera features and can be used with any camera. Figure 2 shows these locations.

### 3. Experimental results

To test the performance of the camera calibration conditions a simulated camera is similar to Zhang in [3]. The calibration template is a chessboard of  $10 \times 14 = 140$  corners of

180x250 mm. In this case, it is situated on the floor and all corners have coordinates  $z_w = 0$ . To calibrate the camera under optimal conditions, the simulated camera is located at  $t_1 = (200, 0, 401.8)$ ,  $t_2 = (-100, 0, 200.9)$  and  $t_3 = (0, 300, 602.1)$ . Also, camera is calibrated using images from random positions  $t_1 = (150, 200, 580)$ ,  $t_2 = (-50, 250, 880)$  and  $t_3 = (100, -20, 820)$ . Results are shown in Fig. 3 for intrinsic camera parameters focal length  $\alpha_u$ , and principal point  $u_0$  only. Similar performance has been computed for the remaining camera parameters. Camera calibration is solved using the linear and the non-linear process. In all cases, computed camera parameters are improved if optimal conditions are used.

The number of points necessary for the calibration process can be defined analysing experimental results. If images are taken taking into account optimal conditions, more than 70 does not reduce parameters errors significantly. Obviously, since constructing a calibration template is an easy task, more than 70 points can be used to improve the results.

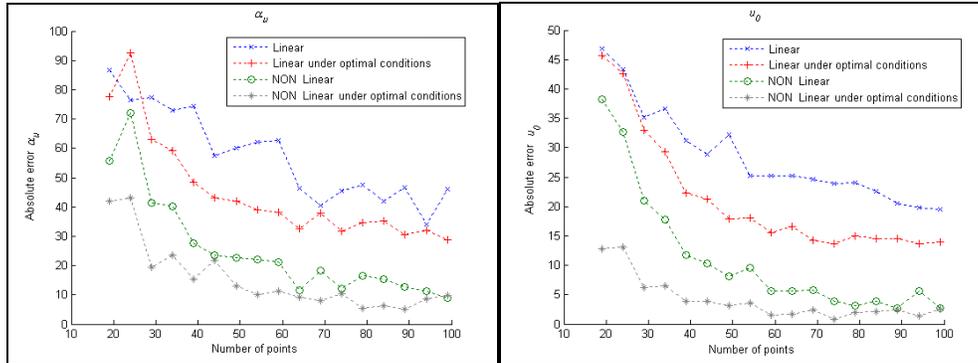


Fig. 3. Comparison of intrinsic camera parameters calibration errors under optimal and non-optimal calibration process conditions changing the number of points.

## 7. Conclusion

Optimal conditions for camera calibration using a 2D pattern have been defined. Camera calibration is a two step process where first a linear algebraic approximation is followed by a non linear searching. Since camera parameters are coupled, non linear searching is poorly conditioned and local solution is reached easily. Consequently, linear algebraic approximation is crucial to avoid divergences to local solution with the non linear searching. Linear algebraic approximation of camera parameters are computed with the eigen vector associated with the smaller eigen value of the matrix composed with elements of several homographies. The calibration process will therefore be more stable if the condition number of this matrix is close to 1. To obtain this condition number closer to 1, elements of the homographies should be taken into account. Therefore, images of the template should be captured from specific locations to obtain a well-conditioned calibration process. To define the optimal locations from where to take template images, the condition number of this matrix has been analysed. The camera is located taking into account that the altitude should be twice its separation from the origin of coordinates in the scene. Also, the camera should be located along the X or Y axes of the scene coordinate system. The camera orientation is defined assuming that the optical axis goes through the calibration template. Finally, point coordinates in the image and the template should be referred to the centre of the image and the template centre, respectively. Although camera calibration using a 2D pattern was intended for self-calibration in which images can be taken from anywhere, here we propose a useful guide to improve the calibration results.

## **Acknowledgments**

This work was partially funded by the Universidad Politécnica de Valencia research funds (PAID 2010-2431 and PAID 10017), Generalitat Valenciana (GV/2011/057) and by Spanish government and the European Community under the project DPI2010-20814-C02-02 (FEDER-CICYT) and DPI2010-20286 (CICYT).