In 18th Portuguese Conference on Pattern Recognition (RecPad 2012)
Coimbra, Portugal, October 2012

Effect of Image Resolution on *DLT-Lines* Camera Calibration Including Radial Distortion

Manuel Silva
manuel.silva.thesis@gmail.com
Instituto Superior Técnico / UTL, Lisbon, Portugal

Ricardo Ferreira, José Gaspar
[ricardo,jag]@isr.utl.pt
Institute for Systems and Robotics, Instituto Superior Técnico / UTL, Lisbon, Portugal

Abstract

Straight lines in 3D scenes and their images, allow to estimate the camera projection matrix and radial distortion, using the *DLT-Lines* calibration methodology. In this paper we detail the *DLT-Lines* and analyze the effect of image resolution based on a synthetic setup having calibration ground truth. More precisely, we consider noise in the 2D data of the synthetic setup, i.e. induce uncertainty in the estimated calibration parameters, and thus assess the benefit of augmenting the camera resolution.

1 Introduction

Laser range finders and color-depth (RGBD) cameras make the process of acquiring 3D data of a scene simpler. In particular, these sensors allow to obtain 3D lines/points on the scene that can be matched with lines/points imaged by a standard (RGB) video camera. These matchings have been shown to be useful for the calibration of networked RGB cameras [4, 5].

A set of 3D-to-2D point correspondences and the Direct Linear Transformation (DLT) permit to estimate the camera intrinsic parameters, orientation and position in a global frame [2]. In case of having just image-lines containing the images of the 3D points, calibration is still possible using the so called *DLT-Lines* methodology [5]. Considering the *Division Model* for radial distortion, introduced by Fitzgibbon [1], *DLT-Lines* also allows the calibration of cameras with radial distortion [5].

Given that *DLT-Lines* is based on 3D and 2D data (see example in Fig.1), it is prone to measurement errors both in the 3D and 2D data sources. In this paper we use the uncertainty caused by the 2D noise on the calibration process, to help assessing, numerically, the benefits of improving the camera resolution.

2 Camera Model and *DLT-Lines*

The perspective camera model describes the mapping of the 3D space to the 2D projective plane [2]. According to the *pin-hole* camera model, a scene point in homogeneous coordinates $M = [X\ Y\ Z\ 1]^T$ is imaged as a point $m = [u\ v\ 1]^T$:

$$m \doteq P\,M = K\,[R \quad t]M \tag{1}$$

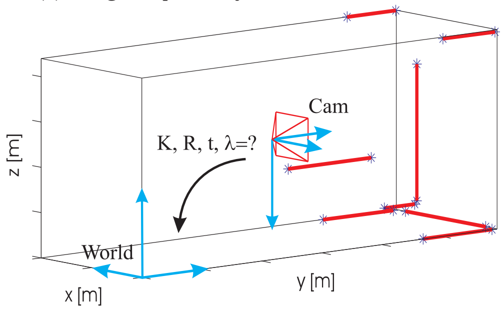
where $\doteq$ denotes equal up to a scale factor, P is a 3×4 projection matrix, K is a 3×3 upper triangular matrix containing the intrinsic parameters of the camera, R is a 3×3 rotation matrix representing the orientation of the camera and t is a 3×1 vector representing the position of the camera. The rotation and translation are defined with respect to a fixed absolute (world) coordinate frame.

The *pin-hole* parameters can be estimated using image lines and scene points [5]. The projection of a 3D line L_i on the camera image plane can be represented by the cross product of two image points, in projective coordinates, $l_i = m_{1i} \times m_{2i}$ [2]. Applying the multiplication by l_i^T on both sides of Eq.1, leads to $l_i^T\ P\,M_{ki} = 0$ where M_{ki} is a 3D point in projective coordinates lying in L_i. The properties of Kronecker product [3] allow to obtain a form factorizing the vectorized projection matrix:

$$(M_{ki}^T \otimes l_i^T)\ vec(P) = 0. \tag{2}$$

Considering $N \geq 12$ pairs (M_{ki}, l_i), one forms a matrix B, $N\times12$, by stacking the N matrices $M_{ki}^T \otimes l_i^T$. The *DLT-Lines* calibration data is illustrated in Fig.1. It consists of paired 3D points and 2D lines. Alternatively, given a 3D line L_i and its projection represented by the image line l_i, any 3D point lying on the 3D line L_i can be paired with 2D line l_i. On the other hand, any image line l_i can be paired with any 3D point lying on L_i, i.e more than one image line can be paired with a 3D point. The least squares solution, more precisely the minimizer of $\|B\ vec(P)\|^2$ subjected to $\|vec(P)\| = 1$, is the right singular vector corresponding to the least singular value of B.

(a) Image acquired by the camera to calibrate



(b) Calibration data, 3D lines

Figure 1: *DLT-Lines* calibration methodology based on 2D-to-3D lines and points correspondences. The image data (a) consists of line segments (red) represented by points (blue stars), while the scene data is formed by 3D lines/points (b). Calibration involves estimating camera pose, $[R\ t]$, intrinsic parameters, K, and radial distortion, λ.

Note that the pin-hole camera model, as presented on Eq.1, does not contain yet the radial distortion. To include radial distortion, we use Fitzgibbon's Division Model. As proposed by Fitzgibbon [1] an undistorted image point, $\hat{m}_u = [u_u\ v_u]^T$ is computed from a radially distorted image point $\hat{m}_d = [u_d\ v_d]^T$ as $\hat{m}_u = \hat{m}_d/(1+\lambda\,\|\hat{m}_d\|^2)$, where λ represents the radial distortion parameter. Fitzgibbon model allows to define a line l_{12} as the cross product of two points:

$$l_{12} = \begin{bmatrix} u_{1d} \\ v_{1d} \\ 1+\lambda s_1^2 \end{bmatrix} \times \begin{bmatrix} u_{2d} \\ v_{2d} \\ 1+\lambda s_2^2 \end{bmatrix} = \hat{l}_{12} + \lambda e_{12} \tag{3}$$

where s_i is the norm of distorted image point i, $s_i^2 = u_{id}^2 + v_{id}^2$, the distorted image line is denoted as $\hat{l}_{12} = [u_{1d}\ v_{1d}\ 1]^T \times [u_{2d}\ v_{2d}\ 1]^T$ and the distortion correction term $e_{12} = [v_{1d}s_2^2 - v_{2d}s_1^2,\ u_{2d}s_1^2 - u_{1d}s_2^2,\ 0]^T$. Applying Eq.3 on Eq.2 leads to the following equation:

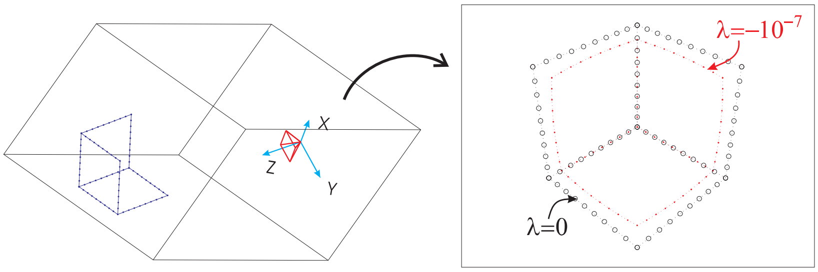
$$\left(M_{k12}^T \otimes (\hat{l}_{12} + \lambda e_{12})^T\right) vec(P) = 0, \tag{4}$$

which can be rewritten as

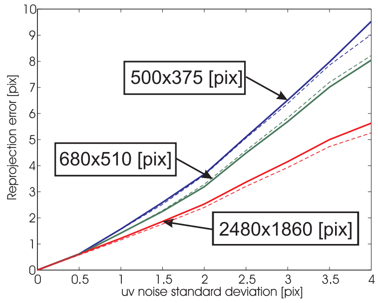
$$(B_{ki1} + \lambda B_{ki2})\ vec(P) = 0 \tag{5}$$

where $B_{ki1} = M_{k12}^T \otimes \hat{l}_{12}^T$, $B_{ki2} = M_{k12}^T \otimes e_{12}^T$ and M_{k12} denotes the k^{th} 3D point projecting to the distorted line l_{12}.

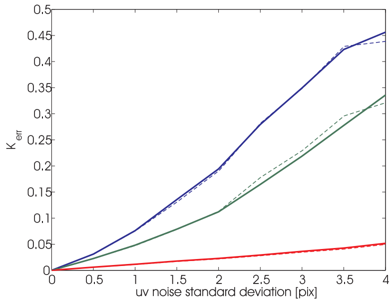
Considering $N \geq 12$ pairs $(M_{ki}, \hat{l}_i)$, where $N = k_{max} i_{max}$, one forms two $N\times12$ matrices, B_1 and B_2, by stacking matrices B_{ki1} and B_{ki2}. As proposed by Fitzgibbon, left-multiplying the stacked matrices by B_1^T results in a Polynomial Eigenvalue Problem (PEP), which can be solved for example, in Matlab using the `polyeig` function. It gives, simultaneously, the projection matrix, in the form of $vec(P)$, and the radial distortion parameter λ.



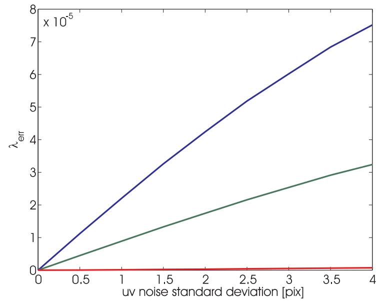
(a) Calibration Setup and image acquired by the camera



(b) Reprojection error, $\lambda \in \{0, -10^{-7}\}$



(c) K_{err}, $\lambda \in \{0, -10^{-7}\}$



(d) λ_{err}, $\lambda = -10^{-7}$

Figure 2: Effect of image resolution in *DLT-Lines*. The setup is composed by a number of 3D line segments and their images (a). Two cases considered, undistorted image (black circles) and radially distorted image (red dots), see (a)-right. The reprojection errors (b), and intrinsic errors K_{err} (c) are computed considering both the distorted (lines) and undistorted (dashed lines) cases. Radial distortion error, λ_{err} is computed only for radially distorted images (d).

3 Experiments

In order to test the effect of different camera resolutions in *DLT-Lines*, we considered three synthetic cameras with standard resolutions and common $8mm$ lens in a $1/4[in]$ CCD (intrinsic parameters). The cameras are considered to be at the world origin and aligned with the world frame ($R = I$). All image points are distorted using Fitzgibbon's radial distortion model, with $\lambda \in \{0, 10^{-7}\}$. Figure 2 illustrates the experimental setup (a).

Considering a 3D line segment, L_i represented by two 3D points (see Fig.2(a) blue dots), the corresponding image line l_i is defined as a collection of two or more points selected from the imaged line segment (red dots and lines, black circle and lines). The imaged line segment is computed using the camera projection matrix. Gaussian noise is added to the image points after computing the projected image line, and before obtaining multiple image lines, as cross products of pairs of distorted image points, which form the *DLT-Lines* input. The camera projection matrix and the distortion parameter are estimated using *DLT-Lines* (see Eq.5), and the noisy and distorted image lines according to Eq.4.

Figure 2 compares the performance of *DLT-Lines*. The comparison encompasses the mean reprojection error Fig. 2(b), the mean of the relative error of the estimated horizontal focal length Fig. 2(c)[1] and, in the case $\lambda = -10^{-7}$, the difference between λ and the estimated radial distortion parameter, λ_{err}. The plots involve 100 calibrations for each noise level (standard deviation, horizontal axis) and for each image resolution. Figures 2(b) and (c), show that in our setup the calibration errors of distorted and undistorted cameras are almost equal. In addition, the similarity of the mean reprojection error plot, Fig. 2(b) with K_{err} plot, Fig. 2(c) and λ_{err} plot, Fig. 2(d), shows that the common cost function used in nonlinear calibration, the mean reprojection error, is effectively a good indicator of the accuracy of the calibration methodology.

4 Conclusions

In this paper, we assess the benefit of image augmenting resolution in the precision of *DLT-Lines*. The assessment consists of testing the calibration methodology in a synthetic setup, encompassing a distorted or undistorted imaging system configured at different resolutions. The results show that the increase of image resolution improves the performance of *DLT-Lines*.

In terms of future work, we plan to further explore empirical and theoretical rules stating the required precisions of the 2D and 3D data in order to obtain a pre-specified precision of the results of *DLT-Lines*. These rules will serve the purpose of building user interfaces helping the *in-situ* calibration of networked cameras.

5 Acknowledgments

This work was supported by the FCT project PEst-OE / EEI / LA0009 / 2011, by the FCT project PTDC / EEACRO / 105413 / 2008 DCCAL, and by the project High Definition Analytics (HDA), QREN - I&D em Co-Promoção 13750.

References

[1] A. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. IEEE Conf. Comp. Vision and Pattern Recognition.*, volume 1, pages 125–132, 2001.

[2] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[3] H. Lütkepohl. *Handbook of Matrices*. Wiley and Sons, 1996.

[4] A. Ortega, B. Dias, E. Teniente, A. Bernardino, J. Gaspar, and J. Andrade-Cetto. Calibrating an outdoor distributed camera network using laser range finder data. In *IROS*, pages 303–308, 2009.

[5] M. Silva, R. Ferreira, and J. Gaspar. Camera calibration using a color-depth camera: Points and lines based dlt including radial distortion. In *Workshop Color-Depth Camera Fusion in Robotics, held with IROS*, 2012.

[1] The horizontal focal length relative error is defined as $K_{err} = (K(1,1) - K_e(1,1))/K(1,1)$, where K is the camera true intrinsic parameters matrix, K_e is the estimated one and $K(3,3) = K_e(3,3) = 1$. K_e can be obtained by decomposing the estimated projection matrix as $P_e = [K_e \quad 0^T]$ [2]