

ROBUST PLANE-BASED CALIBRATION FOR LINEAR CAMERAS

Simon Donn ¹ Hiep Luong¹ Stijn Dhondt² Nathalie Wuyts² Dirk Inz ² Bart Goossens¹ Wilfried Philips¹

¹IMEC - IPI - Ghent University

²VIB - Center for Plant Systems Biology - Ghent University

²Department of Plant Biotechnology and Bioinformatics - Ghent University

ABSTRACT

A linear, or 1D, camera is a type of camera that sweeps a linear sensor array over the scene, rather than capturing the scene using a single impression on a 2D sensor array. They are often used in satellite imagery, industrial inspection, or hyperspectral imaging. In satellite imaging calibration is often done through a collection of ground points for which the 3D locations are known. In other applications, e.g. hyperspectral imaging, such known points are not available and annotating many different points is onerous. Hence we will use a checkerboard for calibration. The state-of-the-art method for linear camera calibration with a checkerboard becomes unstable when the checkerboards are parallel to the image plane. Our proposed method¹ yields more accurate camera calibrations without suffering from this shortcoming.

Index Terms— linear camera, camera calibration, hyperspectral imaging

1. INTRODUCTION

Linear cameras are one-dimensional imaging devices primarily used in satellite imaging and hyperspectral camera systems. They sweep a linear sensor array over the scene: at each time instance, a different slice of the scene is projected onto the sensor array as shown in fig. 1.

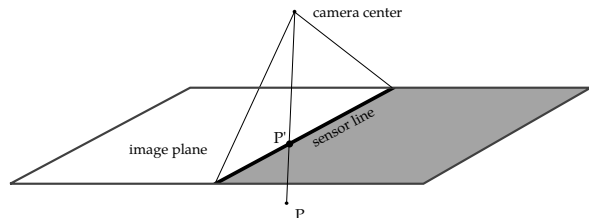


Fig. 1. The linear camera captures the scene on a linear sensor array (the X-direction), while moving along the Y-direction and projecting along the Z-direction. Usually, all of these directions are perpendicular, as in a conventional camera.

¹The source code is publicly available at telin.ugent.be/~sdonn/code/lincams.zip.

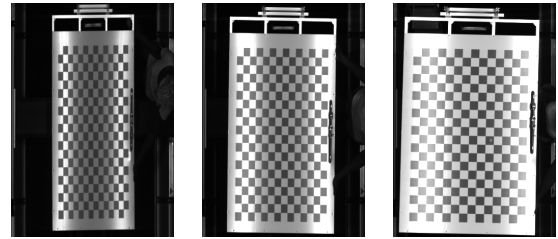


Fig. 2. The double nature of the camera model becomes apparent when imaging a checkerboard at different heights: note that the imaged width of the board depends on the distance to the camera, but the imaged length is invariant.

The camera projection is orthogonal along the movement direction but perspective along the array direction: classical camera calibration [1] is not applicable. The main drawback is that the extent of the camera's movement should be of the same magnitude as the extent of the scene.

For satellite imagery, the sweeping movement of the camera is a side-effect of the satellite's orbit; because of the astronomical scale of the orbit, it is assumed that the movement is locally linear [2]. On the other hand, a 1D sensor array is smaller and lighter compared to a 2D sensor grid: important considerations for anything launched into space.

This type of camera is also used for so-called spatial scanning in hyperspectral imaging. A strip of the scene is projected onto a slit and subsequently dispersed onto a 2D array: one axis presenting the linear sensor array and the other distinguishing between various wavelengths because of Snell's law. In this case, linear cameras are used because of the size (and cost) constraints of building the hyperspectral camera.

In many practical applications, calibrated cameras are needed: we wish to find correspondences between 3D locations and 2D observations by the linear camera. For satellite imagery, calibration is often done using ground reference points: distinct points with well known 3D locations [2]. When performing sensor fusion with hyperspectral imaging, we prefer the calibration to be as accurate as possible; manually annotating a large number of points is onerous and prone to inaccuracies because of manual measurement errors. For this reason we use checkerboards to calibrate the camera [3].

In section 2 we discuss the state-of-the-art for plane-based linear camera calibration, and highlight some important shortcomings in the existing methods. Section 3 contains an explanation of the proposed method: heavily influenced on the state-of-the-art but with some key changes to make the algorithm more generally applicable. Section 4 shows the experimental results and their discussion. We present a summary and some closing remarks in section 5.

2. EXISTING WORK

The seminal work on linear cameras, by Hartley et al. [2], is focused on satellite applications. They extensively discuss and explain the used model, and derive relations between separate linear cameras akin to the essential matrix for stereo setups of classic cameras. However, for calibration of the cameras they rely on known correspondences between 3D ground reference points and 2D camera observations [2] – an assumption that scales poorly when requiring many observations for high accuracy.

Luna et al. [4] took this one step further for lower-scale problems. Using a single line-scan of a known 3D object (two parallel planes at different heights with crossing lines) they are able to estimate the camera parameters well. Yao et al. use a similar set-up to estimate line-scan cameras [5].

Contrastingly, several methods have gone the other way around: using an auxiliary 2D camera to help calibrate the line-scan camera [6, 7, 8]. All these methods start from the assumption that a linear camera movement is difficult to achieve in practice: they use a series of single line-scans rather than 2D images. After rigidly coupling a pre-calibrated 2D camera with the linear camera, Sun et al. use the 2D camera to ensure a good distribution of correspondence points during calibration [7]. Using bundle adjustment refinement [6, 7] and distortion estimation [7], the linear camera is calibrated.

Drareni et al. [3] have found inspiration in classical camera calibration: using a checkerboard, they are able to easily generate large numbers of point correspondences with the known object, improving the accuracy of the calibration as much as possible. Using a linear set of equations, the cameras are calibrated, and bundle adjustment is used to locally refine the estimate. However, the method cannot handle some specific orientations of the planar pattern as it entails dividing by elements of rotation matrices. By avoiding such divisions, our proposed method sidesteps the degenerate positions while at the same time yielding a more accurate calibration in other situations.

3. PROPOSED METHOD

The outline of the method is based on that of Drareni et al. [3], with the aforementioned note that we avoid dividing by elements of rotation matrices. The result is that our proposed

method handles formerly degenerate plane positions, while yielding more accurate results in other situations.

3.1. The Camera Model

As given in [3], the undistorted projection (u, v) of a point (X, Y, Z) in the camera coordinate system is given by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \begin{bmatrix} fX + u_0Z \\ sYZ \\ Z \end{bmatrix} = \underbrace{\begin{bmatrix} f & 0 & u_0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix}}_K \begin{bmatrix} X \\ YZ \\ Z \end{bmatrix}, \quad (1)$$

where K represents the intrinsic camera matrix with focal length f , optical center u_0 and movement speed s . Note the non-linearity due to the YZ term, which precludes the use of classical pin-hole calibration methods.

3.2. The Planar Grid

In the i th image, the n th point $[a_n, b_n, 0]^T$ on the planar pattern is transformed to the camera space by a rotation R_i and a translation \mathbf{t}_i . In favour of legibility, we omit the subscripts image and point indices in matrix elements. The matrix names indicate whether they depend on the image and/or point index. The expression becomes

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = [R_i | \mathbf{t}_i] \begin{bmatrix} a \\ b \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} ar_{11} + br_{12} + t_1 \\ ar_{21} + br_{22} + t_2 \\ ar_{31} + br_{32} + t_3 \end{bmatrix} \quad (2)$$

If we express the projected point $[u, v, 1]^T$ in terms of the location on the planar pattern $[a, b, 0]^T$, we encounter quadratic terms in a and b . To circumvent this, we express $[X, YZ, Z]^T$ in terms of the Veronese mapping of the planar locations, resulting in the projection equation (3).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim K \underbrace{\begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21}t_3 + r_{31}t_2 & r_{22}t_3 + r_{32}t_2 & t_2t_3 \\ r_{31} & r_{32} & t_3 \end{bmatrix}}_{P_{i,1}} \begin{bmatrix} a \\ b \\ 1 \end{bmatrix} + \underbrace{K \begin{bmatrix} 0 & 0 & 0 \\ r_{21}r_{31} & r_{22}r_{32} & r_{21}r_{32} + r_{22}r_{31} \\ 0 & 0 & 0 \end{bmatrix}}_{P_{i,2}} \begin{bmatrix} a^2 \\ b^2 \\ ab \end{bmatrix}. \quad (3)$$

3.3. Homography Estimation

The homography $H_i \sim K [P_{i,1} | P_{i,2}]$ maps the points on the grid on their images in the camera. It is a camera-dependent matrix whose 12 non-zero entries (see equation (6)) we estimate as follows. To estimate the homography, we recall that

$$[u, v, 1]^T \sim H_i [a, b, 1, a^2, b^2, ab]^T. \quad (4)$$

$$\begin{bmatrix} a & b & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -au & -bu & -u \\ 0 & 0 & 0 & a & b & 1 & a^2 & b^2 & ab & -av & -bv & -v \\ -av & -bv & -v & au & bu & u & a^2u & b^2u & abu & 0 & 0 & 0 \end{bmatrix} \mathbf{h} = \mathbf{0} \quad (5)$$

$$H = \frac{1}{t_3} \begin{bmatrix} fr_{11} + u_0r_{31} & fr_{12} + u_0r_{32} & ft_1 + u_0t_3 & 0 & 0 & 0 \\ s(r_{21}t_3 + r_{31}t_2) & s(r_{22}t_3 + r_{32}t_2) & st_2t_3 & sr_{21}r_{31} & sr_{22}r_{32} & s(r_{21}r_{32} + r_{22}r_{31}) \\ r_{31} & r_{32} & t_3 & 0 & 0 & 0 \end{bmatrix} \quad (6)$$

By multiplying the above previous equation with the cross product skew matrix of $[u, v, 1]^T$, we end up with the linear homogeneous system in H_i 's entries from equation (5). In this system, $\mathbf{h} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{24}, h_{25}, h_{26}, h_{31}, h_{32}, h_{33}]^T$. In order to get rid of the unknown scale factor, we propose dividing H_i by h_{33} : a value which cannot be zero because that would mean that the planar pattern contained the camera location, a physical improbability. The result is that H_i is of the form (6). This step is not present in [3], but does make future notations easier and fixes the unknown scaling factor. At this point, we have estimated the homography H_i for each image i , with unknown image-dependent scale factors $1/t_{i,3}$.

3.4. Intrinsic parameter estimation

Next, we express the first two columns of R_i in terms of H_i :

$$\begin{aligned} [R_{i,1}|R_{i,2}] &= \frac{t_{i,3}}{sf} L_i M_i, \\ L_i &= \begin{bmatrix} s & 0 & -su_0 \\ 0 & f/t_{i,3} & 0 \\ 0 & 0 & sf \end{bmatrix}, \\ M_i &= \begin{bmatrix} h_{11} & h_{12} \\ \frac{h_{33}h_{21} - h_{31}h_{23}}{h_{33}^2} & \frac{h_{33}h_{22} - h_{32}h_{23}}{h_{33}^2} \\ h_{31} & h_{32} \end{bmatrix}, \\ &= \begin{bmatrix} h_{11} & h_{12} \\ h_{21} - h_{31}h_{23} & h_{22} - h_{32}h_{23} \\ h_{31} & h_{32} \end{bmatrix}. \end{aligned} \quad (7)$$

Here, our proposed formulation of M_i differs from the existing method: Drareni et al. use $[h_{24}/h_{31}, h_{25}/h_{32}]$ for the second row of M_i . Yet their approach entails dividing by entries of the rotation matrix that will be very close to 0 when the planar pattern is close to parallel to the image plane. This does crop up in practical situations, and we have to take pains to avoid this (in our experiments this was always the case).

We introduce the notation

$$\begin{aligned} X_i = L_i^T L_i &= \begin{bmatrix} s^2 & 0 & -s^2u_0 \\ 0 & f^2/t_{i,3}^2 & 0 \\ -s^2u_0 & 0 & s^2(u_0^2 + f^2) \end{bmatrix} \\ &= \begin{bmatrix} x_1 & 0 & x_2 \\ 0 & x_{i,4} & 0 \\ x_2 & 0 & x_3 \end{bmatrix}. \end{aligned} \quad (8)$$

This allows us to solve the orthogonality constraint of R_i to the values x_1, x_2, x_3 and $x_{i,4}$. Of these, only $x_{i,4}$ depends on the scaling factor $t_{i,3}$ and hence the image index i . The orthogonality constraint of R_i is rewritten to

$$\begin{cases} (M_i^T X_i M_i)_{12} = 0 \\ (M_i^T X_i M_i)_{11} = (M_i^T X_i M_i)_{22} \end{cases} \quad (9)$$

Gathering the orthogonality constraints for each rotation matrix R_i , we end up with a system of $2I$ equations in $3 + I$ parameters. After solving the system up to scale, we have estimates for u_0 and f :

$$u_0 = -x_2/x_1 \quad \text{and} \quad f = \sqrt{\frac{x_3}{x_1} - u_0^2}. \quad (10)$$

We now use the diagonal elements of $M_i^T X_i M_i$, equating $s^2 f^2 / t_{i,3}^2$, to yield a set of equations we solve for $t_{i,3}^2$ and s^2 . These equations appear to be nonlinear, but they are actually linear in $1/t_{i,3}^2$ and $1/s^2$. Because $t_{i,3}$ represents the movement away from the camera, it is necessarily positive. Inverting the sign of S is equivalent to inverting the Y -axis of all camera coordinate systems. The choice is therefore arbitrary, and we select it to be positive.

There is one caveat with this method. In the case of a perfectly flat position of the plane, $h_{31} = h_{32} = 0$, solving (9) does not yield a unique solution (the null space has dimension 3). However, we now have a shortcut to the estimation of the scanning speed, as this also implies that $r_{13} = r_{23} = 0$, so that $h_{21}^2 + h_{22}^2 = 1$. Estimation of u_0 and f is done from the non-degenerate equations in (9), after we continue in the discussed fashion. In the extreme case one sees that an offset in u_0 is equivalent to one in t_1 , as is the scaling of f and t_3 .

3.5. Extrinsic parameter estimation and non-linear optimization

In the original paper, Drareni et al. go on to show how to extract the scaling factors and movement speed as well as the extrinsic parameters through another instance of the orthogonality constraint of R_i . We have already estimated the values of s and $t_{i,3}$, so now the transformation matrices from (2) are straightforward to extract from the homography (6).

Similar to classical camera calibration, we perform non-linear optimization of the initial estimates: bundle adjustment. Although the form of the Jacobian differs from that of pinhole cameras, the same principle remains valid.

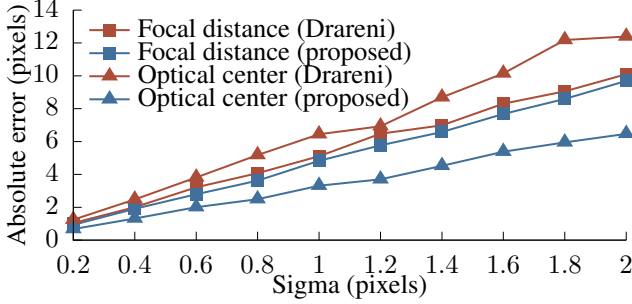


Fig. 3. Focal length and optical center errors w.r.t. the noise level in the point measurements.

4. EXPERIMENTS AND RESULTS

We will now evaluate both the proposed method and the original method by Drareni et al. on synthetic test data. In the experimental set-up the Drareni method yields nonsense results as the planar patterns are all parallel to the image plane: degenerate orientations for that method.

4.1. Synthetic data

We performed several tests of our proposed algorithm using synthetic data, with the same settings as [3]. A planar calibration grid of 10×10 corners is captured by a virtual camera with a 1000×1000 image resolution, $f = 1000$ and $u_0 = 500$. The scanning speed is 50 pixels.

4.1.1. Noise level sensitivity

For 10 planes oriented randomly, additive white Gaussian noise pollutes the measurement. We perform 1000 independent runs for σ from 0.2 to 2.0 and report the average absolute errors for both the focal length and the optical center, as in [3]. We see in fig. 3 that the error increases linearly with rising σ . While our results for the Drareni method match those reported in [3], the proposed method finds better optima.

4.1.2. Sensitivity to the number of planes and their distance

Next, we evaluate the performance of the method with respect to the amount of data available, i.e. the number of images captured with the camera. As fig. 4 shows, the accuracy increases drastically up to about 20 planes, after which the marginal gain levels off. Once again, we see that our proposed method performs better than the existing approach.

4.2. Experimental set-up

For the real-world evaluation of our proposed method, we calibrate a hyperspectral Specim camera. This camera is in an imaging chamber controlled by a robotics system installed at the VIB Center for Plant Systems Biology in Ghent, Belgium.

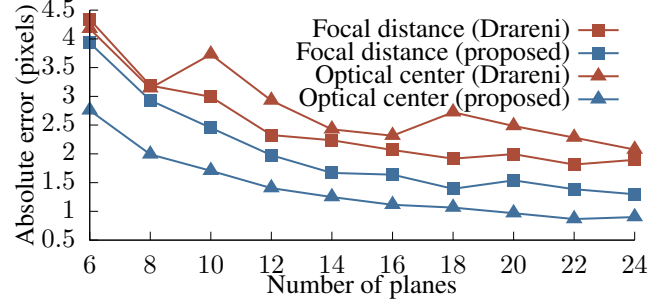


Fig. 4. Focal length and optical center errors w.r.t. the number of measurements available.

The WIWAM conveyor belt system is only able to raise or lower objects, or rotate them in the horizontal plane. Hence, the planar patterns are always parallel to the image plane. As discussed earlier this implies an equivalency between the optical center u_0 and displacements along the Y -axis, and between the focal distance and a scaling along the Z -axis.

As we need to find out the accurate position of the camera with respect to the robot, we pass the algorithm the manufacturer-specified parameters: a focal distance of 500 pixels, and the optical center at 160 pixels. Because of this, we supply the algorithm with the known positions of the planes: four scans at two different heights, a displacement of roughly 0.6 times the plane height (200mm). Using this information, our proposed algorithm estimates the position distance to be 198.81 mm, an error of less than 1%, using only a tiny amount of information. This calibration allows us to accurately relate 3D reconstructions built from other cameras to the hyperspectral scans. In contrast, the method proposed by Drareni et al. could not calibrate this set-up at all, yielding nonsense results due to the division by rotation matrix elements discussed in section 3.4.

5. CONCLUSION

In this paper, we provide an improved version of Drareni's method for calibrating linear cameras [3]. We show how to reformulate the intrinsic parameter estimation so that the method becomes robust to formerly degenerate plane positions. It is shown that the proposed methods performs better than the existing approach in general situations, and how it can be used to calibrate a formerly degenerate rig without issues. The Matlab code is publicly available.

Acknowledgement

This work has in part been performed in the ICON BAHAMAS, a project realized in collaboration with iMinds (imec) and VIB. Simon Donné is also a grantee of the Belgian Research Funds under grant number 01D21213.

6. REFERENCES

- [1] Zhengyou Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [2] Richard I. Hartley and Rajiv Gupta, *Linear pushbroom cameras*, pp. 555–566, Springer Berlin Heidelberg, Berlin, Heidelberg, 1994.
- [3] Jamil Draréni, Sébastien Roy, and Peter Sturm, “Plane-based calibration for linear cameras,” *International Journal of Computer Vision*, vol. 91, no. 2, pp. 146–156, 2011.
- [4] Carlos A Luna, Manuel Mazo, José Luis Lázaro, and Juan F Vazquez, “Calibration of line-scan cameras,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 8, pp. 2185–2190, 2010.
- [5] Ming Yao, Zuyun Zhao, and Bugao Xu, “Geometric calibration of line-scan camera using a planar pattern,” *Journal of Electronic Imaging*, vol. 23, no. 1, pp. 013028–013028, 2014.
- [6] Bingwei Hui, Gongjian Wen, Peng Zhang, and Deren Li, “A novel line scan camera calibration technique with an auxiliary frame camera,” *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 9, pp. 2567–2575, 2013.
- [7] Bo Sun, Jigui Zhu, Linghui Yang, Shourui Yang, and Zhiyuan Niu, “Calibration of line-scan cameras for precision measurement,” *Applied Optics*, vol. 55, no. 25, pp. 6836–6843, 2016.
- [8] Dongdong Li, Gongjian Wen, and Shaohua Qiu, “Cross-ratio-based line scan camera calibration using a planar pattern,” *Optical Engineering*, vol. 55, no. 1, pp. 014104–014104, 2016.