CAMERA CALIBRATION AND POSE ESTIMATION USING PLANAR FEATURES AND SENSITIVITY ANALYSIS

András BÓDIS-SZOMORÚ
Advisors: Tamás DABÓCZI (BUTE-MIT), Zoltán FAZEKAS (MTA SZTAKI)

I. Introduction

In 3D computer vision, a camera model is a geometrical model that maps points of a 3D scene to corresponding image points; the latter being given in pixel coordinates. Many machine vision applications such as 3D object reconstruction, augmented reality (AR), visual servoing of robots or man-machine interfaces rely on the knowledge of this mapping [1].

In order to make use of a camera model, all or at least some of its parameters are usually calibrated off-line as a preliminary step using known world-to-image correspondences of some specific detectable features, typically corners, circular patches or linear segments.

Many object and camera tracking applications make use of the a priori knowledge of the intrinsic camera parameters and continously re-compute the extrinsic parameters over time [1]. In this case, referred as pose estimation, only the six degrees-of-freedom (DoF) of the Euclidean transformation between camera and world needs to be recovered. The precision of pose estimation in object tracking or the accuracy of a reconstructed object geometry in 3D reconstruction always coheres intensely with the accuracy of the preliminary calibration, thus highlighting the importance of an accurate calibration and a comprehensive sensitivity analysis.

This paper addresses preliminary camera calibration from static and planar features including a sensitivity analysis. Significant modifications are proposed herein to some existing methods. Two very practical applications of the proposed methods are outlined in brief. The aforementioned modifications and methods presented here are embodied in two recently developed Matlab GUI Toolboxes that are going to be freely available.

II. Calibration using planar features

A. *Camera models and parameters*

A projective (pinhole) camera can be represented as a linear mapping $\mathcal{P}$ from the projective space $\mathbb{P}^3$ to $\mathbb{P}^2$. It maps a 3D world point $\tilde{W} \in \mathbb{P}^3$ to its 2D image $\tilde{I} \in \mathbb{P}^2$: $\lambda\tilde{I} = P\tilde{W}$ where P is the 3×4 homogeneous camera matrix representation of $\mathcal{P}$ and λ is an arbitrary scale factor. In the inhomogeneous representation $W \in \mathbb{R}^3$ and $I \in \mathbb{R}^2$, the projective camera model becomes non-linear. It should be noted that lens distortions affect the inhomogeneous coordinates of the projected points. Therefore, when such distortions are considered, it is necessary to use the inhomogeneous representation. The mapping $\varphi\colon W \mapsto I$ in its general form can be written as

$$I = \varphi(\mathbf{p_{int}}, \mathbf{p_{ext}}, W), \tag{1}$$

where $\mathbf{p_{int}}$ is the vector of intrinsic parameters and $\mathbf{p_{ext}}$ is the 6-vector of extrinsic parameters representing a 6-DoF Euclidean transformation of the points from the world to the camera reference frame. This transformation is usually represented by a 3×3 rotation matrix $R = \begin{bmatrix}\mathbf{r_1} & \mathbf{r_2} & \mathbf{r_3}\end{bmatrix}$ and a translation vector $\mathbf{t}$. Since R has 9 elements but only 3-DoF, it is an over-parametrization of the model that involves 6 additional non-linear constraints induced by $R^T R = E$, where E represents the unit matrix. This is the reason why the Rodrigues-vector, quaternions or Euler-angles are commonly used to represent rotation [1]. Using the Rodrigues-vector $\mathbf{r} \in \mathbb{R}^3$, $\mathbf{p_{ext}} = (\mathbf{r}^T, \mathbf{t}^T)^T$.

B. Camera parameters from planar patterns

In case of camera calibration I_i and W_i $(i = 1, 2, \dots, n)$ correspondences are known and the parameter vector $\mathbf{p} = (\mathbf{p_{int}}^T, \mathbf{p_{ext}}^T)^T$ is to be determined. Since planar fidutials are easier to create than 3D ones, recent methods prefer using planar patterns. In the planar case, without loss of generality $W_i = (X_i, Y_i, 0)^T$ and the homogeneous mapping can be re-written as $\lambda\tilde{I}_i = \lambda(u_i, v_i, 1)^T = H(X_i, Y_i, 1)^T$; where $H = \begin{bmatrix}\mathbf{h_1} & \mathbf{h_2} & \mathbf{h_3}\end{bmatrix} = K\begin{bmatrix}\mathbf{r_1} & \mathbf{r_2} & \mathbf{t}\end{bmatrix}$. Here H is a 3×3 homography matrix mapping all W_i to I_i, and K is the 3×3 camera calibration matrix depending solely and linearly on the intrinsic parameter vector $\mathbf{p_{int}}$. Unfortunately, from a single planar arrangement it is impossible to recover all the 11 camera parameters, (i.e. the 6 extrinsic and 5 intrinsic parameters in case of no distortions). In [2], Z. Zhang presents an excellent solution to this problem by showing the same *known pattern* in m different *unknown orientations* to the camera. (Figure 1).

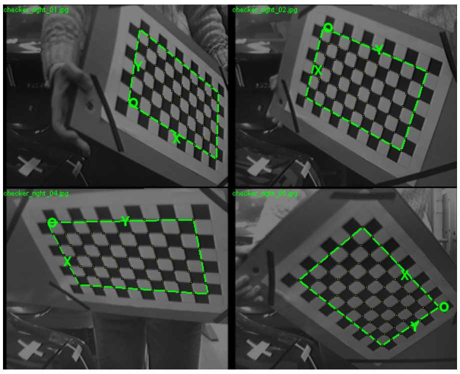

Figure 1: Calibration from planar patterns with our Matlab Toolbox

For each view, there is an individual $\mathbf{p_{ext,j}}$ $(j = 1, 2, \dots, m)$ vector of extrinsic parameters while the intrinsic parameters are common for all views. After computing the homographies H_j from the point-correspondences, based on the orthonormality of $\mathbf{r_{1,j}}$ and $\mathbf{r_{2,j}}$ (see the constraints on H above), Zhang managed to build up a set of equations involving only the intrinsic parameters. These equations are linear in the elements of $B = K^{-T}K^{-1}$ and consequently non-linear in those of $\mathbf{p_{int}}$. First, B is computed by an LS method, and then K and $\mathbf{p_{int}}$ are recovered using non-linear formulas. From the individual H_j computed beforehand, all $\mathbf{p_{ext,j}}$ can be recovered. In other words, the pose estimation problem is solved for each of the m planes. The resulting camera parameters packed in $\bar{\mathbf{p}} := (\mathbf{p_{int}}^T, \mathbf{p_{ext,1}}^T, \dots, \mathbf{p_{ext,m}}^T)^T$ are then refined by minimizing the reprojection error $C(\bar{\mathbf{p}}) = \sum_{i,j} d^2(I_{ij}, \hat{I}_{ij}) = \sum_{i,j} \|I_{ij} - \varphi(\mathbf{p_{int}}, \mathbf{p_{ext,j}}, W_{ij})\|^2$, where d denotes Euclidean distance, $\|.\|$ denotes vector 2-norm, I_{ij} is the i-th image point detected by a feature detector on the j-th image, $\hat{I}_{ij}$ is the reprojection of the corresponding 3D point W_{ij} onto the image plane using the camera model (1).

C. Sensitivity analysis

If the distribution of feature localization is a (2D) isotropic Gaussian with zero mean for all the images, and the world points W_i are precisely known, then the cost function minimized is a maximum likelihood (ML) cost function. Since the residual errors are realizations of the above-mentioned Gaussian noise, the standard deviation $\hat{\sigma}$ of all the residual coordinates is a good measure of the uncertainity in the image. When searching for the uncertainty of the camera parameters, one can back-propagate $\hat{\sigma}$ to the camera parameters through the camera model linearized at each reprojected point: $C_p = (J^T J)^{-1}\hat{\sigma}^2$, where C_p is the estimate of the $(9 + 6m) \times (9 + 6m)$ covariance matrix of all the parameters, $J = \frac{\mathrm{d}\hat{I}}{\mathrm{d}\bar{p}}$ is the stacked $2n \times (9 + 6m)$ Jacobian of the mapping evaluated at the optimal parameters and $\hat{I}$ is the vector containing all the reprojected image points $\hat{I}_{ij}$, thus having $2mn$ elements. Note that the hyper-matrix J is a sparse matrix. This is due to the fact that small perturbations in the orientation of the i-th plane affect only the reprojected points in the i-th image, though perturbations in the intrinsic parameters affect all. Finally, the deviation of the k-th camera parameter $k = 1, 2, ..., (9 + 6m)$ is estimated as $\hat{\sigma}_{p,k} = \sqrt{C_{p,kk}}$. It implies that uncertainty of the intrinsic parameters is considered independent of that of the extrinsic parameters.

A Toolbox for Matlab implementing a calibration similar to Zhang's method already exists at Caltech [3]. It is open-source and performs sensitivity analysis, as well. They back-propagate the measurement noise to all the parameters as discussed above and associate $\pm 3\hat{\sigma}_{p,k}$ uncertainty to the parameters corresponding to the 99.7% confidence level.

III. Problems and solutions

A. *Enforcing values for a subset of parameters*

In situations when exact a priori information about the mapping is available (e.g. the pixel aspect ratio is exactly known from the datasheet of the camera), it is desirable to enforce some parameters manually. Therefore, we extended the method of Zhang to handle cases when a subset of intrinsic parameters is fixed. It is important to note that the matrix B estimated with Zhang's linear method has 9 elements but only 6 or less DoF. This means that measurement noise may jeopardize the estimation in some cases. This threat increases when certain parameters are enforced. A subset of intrinsic parameters can be easily fixed using our Toolbox.

B. *Perspective-n-point problem with error in the world points*

The cost function used earlier expresses errors only in the image. It is not ML any more when significant errors are present in W_i. A practical example is when one may use Zhang's method for intrinsic calibration and an additional plane with noisy world points for a final pose estimation. Therefore, we derived the ML cost function for this case. Here, we supposed that measurement uncertainities of the 3D coordinates W_i are Gaussian and are independent of the errors in feature localization in the image. This condition is generally met in practice. The resulting cost function is:

$$C(\mathbf{p_{int}}, \mathbf{p_{ext}}) = \sum_i \|I_i - \hat{I}_i\|^2_{C_{Ii}} + \sum_i \|W_i - \hat{W}_i\|^2_{C_{Wi}}, \tag{2}$$

where $\|x\|^2_C$ denotes the Mahalanobis distance $(x^T C^{-1} x)$. Equation (2) introduces the 3D measurement errors $(W_i - \hat{W}_i)$ where $\hat{W}_i$ is the expected 3D location of W_i, while W_i is the measured one. The introduction of $\hat{W}_i$ means that the 3D points are considered as parameters and thus our parameter space is of dimension $(9 + 6m + 3n)$. Just for comparison, we also used a convenient approximation for the cost function (2) in which each $\hat{W}_i$ is estimated as the orthogonal projection of W_i to the ray through the image point I_i. This reduces the dimension of the parameter space back to $(9 + 6m)$ and gives an estimate very close to the ML one.

Also, the covariance matrices C_{Ii} and C_{Wi} appeared in (2). The first caracterizes the error of the feature detection in the image, the second, the localization errors in 3D. Therefore, this solution is only practical when C_{Ii} and C_{Wi} can be estimated. $C_{Ii} = diag(\sigma^2, \sigma^2), i = 1, 2, \ldots, n$ used to be a practical simplification. When $\mathbf{p_{int}}$ is preliminarily estimated e.g. by Zhang's method, the problem reduces to pose estimation, that is, estimating only $\mathbf{p_{ext}}$, that is, only 6 DoF needs to be recovered. A single planar arrangement with 4 points or more in general planar configuration is enough to compute an initial guess for $\mathbf{p_{ext}}$. This is called Prespective n-Point problem (PnP) for $n > 3$ [1]. The method is also extended to the multi-camera case where the 3D structure ($\hat{W}_i$) is common for all the views.

C. *Modifications to the sensitivity analysis*

In the discussed sensitivity analysis, the covariance of the errors in the image is back-propagated to all the parameters. However, when *using* the results of the calibration, generally only one reference view is important and only the corresponding sub-matrix of C_p needs to be considered.

The situation gets more complicated when errors are present in the world points of one of the calibration planes. In our Toolboxes, we use checkerboard planes with world points considered exact and an additional reference plane with errors in the world points for pose estimation. The checkerboard-based calibration and the pose estimation to the reference plane are performed in two independent steps with independent measurements. Assuming Gaussian distribution for all the measurements, we extended the sensitivity analysis for this case. Since the ML cost function introduces the estimate of the expected location of the world points as well, our overall sensitivity analysis gives the covariance of the intrinsic parameters, of all the $6m$ extrinsic parameters with respect to m checkerboards plus 6 with respect to the reference plane and of the 3D location estimates of the world points in the reference plane.

IV. Applications and results

A. Calibration of a multi-camera corneal topographer

We used our Toolbox implementing the proposed methods to calibrate the multi-camera corneal topographer described in [4]. Calibration is done by rotating a planar pattern of size 1.2×1.2 cm at the expected location of the patient's cornea. The key problem experienced during the calibration was that due to the high focal length, the perspective effect in the images taken was neglectable and therefore a high uncertainity was present in the principal point estimation. This was solved by fixing the principal point in the image center. An alternative solution is to use an affin camera model. For this purpose, the specular surface reconstruction algorithm has been adapted for an affine model, as well.

B. Calibration of a stereo lane detection system

Another interesting application of our methods is the calibration of the stereo lane detection system described in [5]. Intrinsic calibration of the cameras fixed to the vehicle's chassis is performed with a hand-held and rotated pattern (see Figure 1). An additional pose estimation is also required relative to the road. For this purpose, markers were placed in front of the vehicle and their 3D positions were measured. The measurement covariances C_{Ii} and C_{Wi} were estimated, as well. Because the 3D points W_i were measured with high uncertainty and localization errors in the image were relatively small, the proposed cost function gave significantly better results at reprojection when refining the 3D points $\hat{W}_i$ as proposed (Figure 2). Both the simplified mono version with reduced parameter space and the stereo version of the algorithm gave qualitatively the same result as shown on the right side of Figure 2. This proves the effectiveness of the approach. The estimation algorithms for an additional plane (the road's plane) are implemented in a second Toolbox supporting multiple cameras and based on the first one.

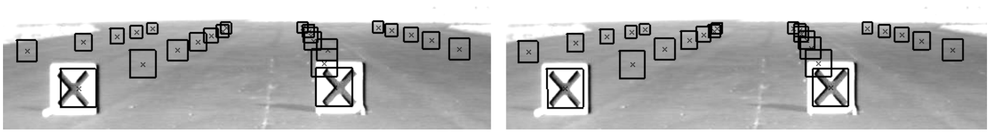

Figure 2: Pose estimation: markers reprojected using the optimal parameters when minimizing errors in the image only (left) and when using the proposed co-minimization based on (2) (right).

V. Conlclusions and future work

We proposed some important modifications to existig methods for calibrating a single or multiple perspective cameras to planes, including sensitivity analysis, parameter fixation, and errors present in the world point locations. The proposed methods are implemented in two Matlab GUI Toolboxes and have been used in two completely different multi-camera applications with success. Further work consists of a specific calibration and reconstruction method for the corneal topographer and on-line pose estimation for the lane detection system under development.

References

[1] V. Lepetit and P. Fua, "Monocular model-based 3d tracking of rigid objects: A survey," *Foundations and Trends in Computer Graphics and Vision*, 1(1):1–89, 2005.

[2] Z. Zhang, "A flexible new technique for camera calibration," Tech. Rep., Microsoft Research, Microsoft Corporation, Aug. 2002, URL: http://research.microsoft.com/∼zhang.

[3] J.-Y. Bouguet, *Camera Calibration Toolbox for MATLAB*, California Institute of Technology, Apr. 2007.

[4] A. Soumelidis, Z. Fazekas, Z. Lichtenberger, and A. Bódis-Szomorú, "Embedded computing solutions used in a multi-camera corneal topographer," in *Regional Conference on Embedded and Ambient Systems (RCEAS)*, 2007, to be published in feb. 2008.

[5] A. Bódis-Szomorú, Z. Fazekas, and T. Dabóczi, "High-accuracy calibration of lane detection systems using stereovision," in *Instrumentation and Measurement Technology Conference (IMTC)*, 2006, ISBN: 1-4244-1080-0.