TR-PM-01-2011
Pin-Hole Camera Reference Frame and Calibration Techniques

Paolo Medici
Dipartimento di Ingegneria dell'Informazione dell'Università di Parma

December 11, 2011

This document is really a work in progress.

1 Introduction

This is a really brief summary of calibration technique proposed in literature. Details can be found in references.

Camera calibration techniques could be divided in two approach: single pose camera calibration and multiple pose camera calibration (see section 7). In this brief document we are going to exam some techniques and give an outlook of the calibration process.

2 Pin-Hole camera model

Before introduce the calibration process, it is necessary to show the parameters that go to be calibrated.

The relationship between world coordinates and image coordinate can be expressed in matrix form using homogeneous coordinates:

$$\mathbf{p}_i = \mathbf{K}[\mathbf{R}\tilde{\mathbf{t}}_0]\mathbf{x}_i \tag{1}$$

where $\mathbf{p}_i = (u_i, v_i)^\top$ are image coordinates and $\mathbf{x}_i = (x_i, y_i, z_i)^\top$ are the corresponding world coordinates.

Relationship between image $\mathbf{p}$, camera $\mathbf{m}$ and world coordinates $\mathbf{x}$ are consequentely:

$$\begin{aligned} \mathbf{p} &= \mathbf{K}\mathbf{m} \\ \mathbf{m} &= [\mathbf{R}\tilde{\mathbf{t}}_0]\mathbf{x} \end{aligned} \tag{2}$$

Only the conversion from camera (or world) coordinates and image coordinates is not invertible.

The $\mathbf{K}$ is the intrinsic parameters matrix, with the form

$$\mathbf{K} = \begin{bmatrix} k_u & k_\gamma & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

k_u and k_v are the focal lenght expressed in pixel unit. They are close related to real f focal lenght and τ aspect ratio but this syntax is slightly more compact. $\mathbf{K}$ converts from camera coordinates to image coordinates (expressed in homogeneous form).

The $\mathbf{R}$ is the rotation matrix. Since in $\mathbb{R}^3$ a unique reference frame not exists to model rotation from 3 angles we are not to deal with the inner form of this matrix. The important stuff is that $\mathbf{R}$ is a rotation matrix or rather it is an orthonormal matrix. $\mathbf{R}$ converts from world coordinates to camera coordinates (can handle for example a permutation term in order to switch from world reference frame to camera reference frame).

$\tilde{\mathbf{t}}_0 = -\mathbf{R}\mathbf{r}_0$ is the position of camera expressed in camera coordinates. $\mathbf{t}_0$ is the position of camera expressed in world coordinates.

It is widely accepted that modern CCD/CMOS camera has zero skew factor ($k_\gamma = 0$).

It is also a bad common assumption to consider the principal point near the center of image: this assumpion is not true and could bring to wrong result in the camera pose calibration.

According to this model it is possible to explicit the projective camera function:

$$(u, v)^\top = \left(k_u \frac{r_0 x + r_1 y + r_2 z + t_x}{r_6 x + r_7 y + r_8 z + t_z} + u_0, k_v \frac{r_3 x + r_4 y + r_5 z + t_y}{r_6 x + r_7 y + r_8 z + t_z} + v_0 \right) \tag{4}$$

2.1 VisLab Reference Frame

The references frames used in our laboratory are shown in figures 1, 2 and 3.

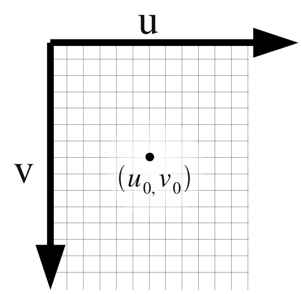

Figure 1: Image coordinates

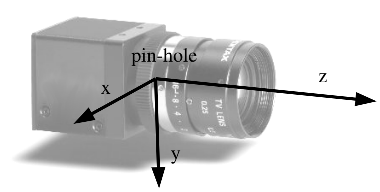

Figure 2: Camera coordinates

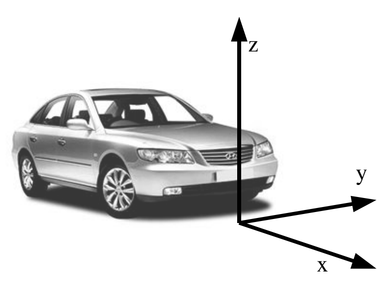

Figure 3: World coordinates ISO 8855

3 Single Pose Calibration

To calibrate a camera some constraints are required. Common constraints are usually the world points and the corrisponding image points.

Using those constraints it is possible to solve the non-linear system 4 and obtains the calibration parameters using a least squares regression:

$$\varepsilon = \sum \|\mathbf{p}_i - f(\mathbf{w}_i)\|^2 \tag{5}$$

where $\mathbf{p}_i$ is the image point (noisy), $\mathbf{w}_i$ is the world coordinate associated and f is the projective function 4. In this case this is a Maximum Likelihood Exstimator (MLE).

Using a non-linear optimization tecnhique (for example Levenberg-Marquardt) the above system is easily solved. The problem of non-convex optimization however is to feed a good starting point for the minimization. Following section try to deal with this problem.

4 Direct Linear Transformation

Equation 1 can be implicitly written as

$$\mathbf{p}_i = \mathbf{P}\mathbf{x}_i \tag{6}$$

with $\mathbf{P}$ the projection matrix, $\mathbf{p}_i = (u_i, v_i, 1)^\top$ the image point in homogeneous coordinates and $\mathbf{x}_i = (x_i, y_i, z_i, 1)^\top$ the world point. The projection matrix is a 3×4 matrix.

If all the constraint own to the same plane (for example $z = 0$) the projection transformation become an homographic transformation in the form:

$$\mathbf{p}_i = \mathbf{H}\mathbf{x}'_i \tag{7}$$

with $\mathbf{H}$ the homographic matrix and $\mathbf{x}'_i = (x_i, y_i, 1)^\top$ the coordinate on the plane. The homographic matrix is a 3×3 matrix.

In DLT proposed by Abdel-Aziz and Karara in 1971[1] the coefficients of the matrix $\mathbf{P}$ or $\mathbf{H}$ are calculated directly.

The projection matrix can be obtained solving an overdetermined homogeneous system:

$$\mathbf{p}_i \times \mathbf{P}\mathbf{x}_i = 0 \tag{8}$$

and the homographic matrix from

$$\mathbf{p}_i \times \mathbf{H}\mathbf{x}'_i = 0 \tag{9}$$

Note that the matrix $\mathbf{P}$ and $\mathbf{H}$ are known less than a multiplication factor. To recover the $\mathbf{P}$ matrix at least 6 points are required, and for $\mathbf{H}$ at least 4 points.

Not using the DLT, a Maximum Likelihood Extimator can be used to solve the problem in non-linear way.

After the implicit calibration it is possible to project world point onto image or from image point recover the world line behind the point. This is not a strong calibration and does not permit some usefull feature that explicit calibration have.

If intrinsic matric $\mathbf{K}$ is known (for example using Zhang calibration technique, section 7) it is possible to recover $[\mathbf{R}\tilde{\mathbf{t}}]$ matrix from $\mathbf{P}$ matrix:

$$[\mathbf{R}\tilde{\mathbf{t}}] = \mathbf{K}^{-1}\mathbf{P} \tag{10}$$

or directly from DLT using camera coordinates instead of image coordinates:

$$\mathbf{m}_i \times [\mathbf{R}\tilde{\mathbf{t}}]\mathbf{x}_i = 0 \tag{11}$$

The $[\mathbf{R}\tilde{\mathbf{t}}]$ has strong constraint (for example is totally known) since the $\mathbf{R}$ is an orthonormal matrix. From orthonormal constraint the multiplication factor can be recovered.

5 Calibration by Vanishing Point

Camera calibration using vanishing points or lines is one of the earliest tecnique to be proposed. From equation 6 it is possible to find in closed form the position of vanishing points in image coordinates:

$$\mathbf{p}_x = \lim_{x \to \infty} \mathbf{P}(x, y, z, 1)^\top = \mathbf{p}_1 \tag{12}$$

with $\mathbf{p}_1$ first column of $\mathbf{P}$. The same process can be applied to other 2 vanishing point.

It is clear that using vanishing point the translation part of $\mathbf{P}$ (the 4th column) cannot be recovered.

6 Lens Distortion Calibration

The pinhole camera model is valid only within lens distortion. However any commercial lens have a bit of distortion due to fabrication error.

Several distortion effects can be considered but the common are radial distorion and decentering distortion.

The undistorted point (u_u, v_u) is radial moved in the distorted point (u_d, v_d) using a only radius r_d, distance from point and distortion center, function. This function can be expressed as

$$\frac{r_u}{r_d} = f_d(r_d) \tag{13}$$

Since not exist a common model for the radial distortion function, it is been approximated from taylor series expansion(*Brown-Conrady model* [2]). The function f_d, ratio between the undistorted distance r_u and the measured distance r_d, is expanded in series as

$$f_d(r) = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots \tag{14}$$

The only 2 power series is due to the fact that f_d is an even function. Tsai, Zhang and others [3] presented several technique to recover the distortion parameters.

7 Zhang Calibration

Zhang [7] in 2000 performs an update of the calibration techniques (yet valid but developed during 80s) largely made by Tsai [5] and others [6].

Zhang uses the projection matrix $\mathbf{H}$ (see 7) generated automatically using a calibration grid and tryes to derive the intrinsic parameters explicitly. As indicated above this matrix is an homographic matrix and thus has 8 degrees of freedom. From this matrix you can put two constraints based on rotation matrix orthonormality in order to force at least two of the intrinsic parameters. The matrix $\mathbf{H}$ infact can be also written as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \tag{15}$$

Expressing orthonormality between column vectors $\mathbf{r_1}$ e and $\mathbf{r_2}$ can be forced the following two constraints:

$$\begin{array}{l} \tilde{h}_1^T \mathbf{W} \tilde{h}_2 = 0 \\ \tilde{h}_1^T \mathbf{W} \tilde{h}_1 = \tilde{h}_2^T \mathbf{W} \tilde{h}_2 \end{array} \tag{16}$$

having defined $\mathbf{W}$ (ignoring the skew factor) as

$$\mathbf{W} = \left(\mathbf{K}^{-1}\right)^\top \mathbf{K}^{-1} = \lambda^2 \begin{bmatrix} \frac{1}{k_u^2} & 0 & -\frac{u_0}{k_u^2} \\ 0 & \frac{1}{k_v^2} & -\frac{v_0}{k_v^2} \\ -\frac{u_0}{k_u^2} & -\frac{v_0}{k_v^2} & \frac{u_0^2}{k_u^2} + \frac{v_0^2}{k_v^2} + 1 \end{bmatrix} \tag{17}$$

symmetric matrix (representing the equation of a conic, the absolute conic [4]).

The 4 unknowns (or 5 considering the skew) of the matrix $\mathbf{W}$, under the two constraints 16, can be solved using at least 2 (or 3) different calibration plane or rather different matrices $\mathbf{H}$, whose columns are not linearly dependent on each other.

Given the matrix $\mathbf{W}$ a Choleski decomposition of the original matrix can be determined. However Zhang provides the equations to obtain the intrinsic parameters directly from $\mathbf{W}$. It is also possible apply the decomposition ${h_i}^\top \mathbf{W} h_j = {v_{ij}}^\top \mathbf{w}$, with appropriate values of the vector v_{ij} and with $\mathbf{w}$ the unknowns of the upper triangular matrix of $\mathbf{W}$.

The system is ill-conditioned and difficult to reach a solution with only few calibration grids.

Only one final note: Zhang and Tsai in their article to coincide with the center of the Principal Point of distortion, which is not always true.

8 Nomenclature

$\mathbf{K}$	or $\mathbf{A}$. Matrix of Intrinsic Parameters (see eq. 3).
$\mathbf{R}$	Rotation Matrix
$\mathbf{P}$	Projection Matrix (see eq. 6)
$\mathbf{H}$	Homogeneous Matrix (see eq. 7)
$\mathbf{E}$	Essential Matrix (see eq. **??**)
$\mathbf{F}$	Fundamental Matrix (see eq. **??**)
k_u, k_v	Focal Lenght in pixel unit (see eq. **??**)
k_γ	Skew Factor, rarely used
W, H	image size in pixel unit
α_u, α_v	Horizontal and vertical half Field Of View (see section **??**)
u_0, v_0	Principal Point coordinates in pixel unit
ϑ	Pitch angle
γ	Yaw angle
ρ	Roll angle

References

[1] Y. Abdel-Aziz and H. Karara. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. In *Proc. ASP/UI Symp. on Close-Range Photogrammetry*, pages 1–18, Urbana, Illinois, Jan. 1971.

[2] D. C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, 1966.

[3] F. Devernay and O. D. Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 13(1):14–24, 2001.

[4] Q.-T. Luong and O. D. Faugeras. Self-calibration of a moving camera from pointcorrespondences and fundamental matrices. *Int. J. Comput. Vision*, 22(3):261–289, 1997.

[5] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3(4):323–344, August 1987.

[6] G. Q. Wei and S. D. Ma. Implicit and explicit camera calibration: Theory and experiments. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(5):469–480, 1994.

[7] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.