A ROBUST CAMERA CALIBRATION TECHNIQUE FOR IMAGE ACQUISITION

A. Vettore
Professor at Istituto di Topografia
Università degli Studi di Padova - Italy -
e-mail: vettoan@ux1.unipd.it

Commission VI, Working group 3

ABSTRACT

Traditional photogrammetric procedures based on metric cameras and standard photogrammetric equipment are very costly. Full automatic procedures originated in the field of computer can accomplish this task at a much lower man time/equipment costs. In this scenario though for operational convenience and accuracy of the results, one needs a robust procedure for calibrating the camera in situ. This works presents a camera calibration technique based on the use of two (or more) independent calibration planes, very robust with respect to noise and plane configurations, which is suited to be used in situ for high resolution image acquisition.

1. Introduction

This work reports about a camera calibration technique satisfying to the needed robustness and probability prerequisites needed in this task.
We adopt Tsai's model [15] which is standard in this field. In this model, radial distortion is the main disturbance but the sensitivity of the position of the image's centre with respect to the other parameters is also very important for accuracy purposes. Ideally, intrinsic and all extrinsic parameters (position and orientation of the camera with respect to the world coordinate system) must be recover simultaneously. However, because of numerical tractability, most of the proposed calibration techniques are also two steps. In the first step, one analytically finds an approximate solution to all the considered parameters with null distortion. This is typically accomplished by solving systems concerning a mixture of linear and quadratic equations (the latter concerns the rotation parameters). In the second step, the initial guess is used in order to start iterative algorithms (such as Gauss-Newton or extended Kalman filtering) aimed to evaluate the distortion parameters and to make the obtained parameters as least noise sensitive as possible [6], [15], [14], [11], [13]. A camera model formed by two planes was originally proposed by [12]. A major advantage of this arrangement is that it offers a variable numbers of parameters with the possibility of increasing their number for accuracy purpose. The major drawback of the method proposed by [12] is that the computation of the forward projection is rather difficult (contrarily to that of the backward projection). The proposed calibration method borrows the idea of the two planes from [12] but for calibration planes and uses different ideas.
In order to currently determine intrinsic and extrinsic parameters, one needs 3D features in the world coordinate space which, in our method, are obtained as sets of independent 2D features, the calibration plane. The use of these sets of 2D objects is one way for simultaneously estimating intrinsic parameters and validating the 3D reconstruction. In the proposed method, even if we use the *rigidity constraint*, the rigid transformation between calibration planes is not known and no necessarily computed. It suffices to compute the projective parameters among the planes, which can be computed with high accuracy [4].
Section 2 states the problems and presents the relationships upon which the calibration procedures of section 3 is built. Section 4 gives some simulation results and experimental results of this work. Section 5 contains the conclusions.

2. Projective and rigid transformations and problem statement

Given a set of coplanar points $\{P\}$ (called calibration plane in the following) and projected image $\{Q\}$ of these points, there exists a projective transformation π between this plane and image plane. An exact projective transformation π between object coplanar points $\widetilde{M}=(X,Y,1)$ and image points $\widetilde{m}=(x,y,1)$ can be represented by vector $\underline{\theta}$ with the following relationships:

$$\widetilde{m} = \pi \cdot \widetilde{M} \quad (1)$$

$$\pi = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ \theta_4 & \theta_5 & \theta_6 \\ \theta_7 & \theta_8 & \theta_9 \end{pmatrix} \quad (2)$$

$$x = \frac{\theta_1 X + \theta_2 Y + \theta_3}{\theta_7 X + \theta_8 Y + 1} \quad (3)$$

$$y = \frac{\theta_4 X + \theta_5 Y + \theta_6}{\theta_7 X + \theta_8 Y + 1} \quad (4)$$

The relationships between $\{P\}$ and $\{Q\}$ define π up to a 1scale factor, therefore, the 9[th] component of $\underline{\theta}$ can be set equal to 1. On the other hand rigid body transformation from the object world (X,Y,Z) to the image plane (x,y) is [15]:

$$(x - x_c) \cdot D = G_x \cdot \frac{r_{11}X + r_{12}Y + r_{13}Z + t_x}{r_{31}X + r_{32}Y + r_{33}Z + t_z} \quad (5)$$

$$(y - y_c) \cdot D = G_y \cdot \frac{r_{21}X + r_{22}Y + r_{23}Z + t_y}{r_{31}X + r_{32}Y + r_{33}Z + t_z} \quad (6)$$

with:

$$D = 1 + k_1 \cdot \left[(x - x_c)^2 + \alpha^2 \cdot (y - y_c)^2 \right] \quad (7)$$

$$\alpha = G_x / G_y$$

where R is the 3×3 rotation matrix

$$R(\theta_x,\theta_y,\theta_z)=\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}=(\underline{C}_1,\underline{C}_2,\underline{C}_3) \tag{8}$$

and $\underline{T}$ the translation vector

$$\underline{T}=\begin{pmatrix} t_x & t_y & t_z \end{pmatrix} \tag{9}$$

form one set of extrinsic parameters. Value G_x, G_y, x_c, y_c and k_1 are the intrinsic parameters. Parameters G_x and G_y are the scale factors including focal length, size of CCD elements and number of pixels. Values x_c and y_c denote the position of the image centre with respect to the optical axis, and k_1 is the radial distortion parameter of the lens which takes into account the non-linear effect far away from the centre of the lens.

In the case of a null distortion parameter ($k_1 = 0$), all parameters (extrinsic and intrinsic) are included in the θ_i's. In this case, the projective transformation π can be solved as a first step by using at least four pairs of corresponding (object and image) coplanar points ($Z = 0$, with no collinear triplet among them) by the least mean squares. The relationship between the projective transformation parameters and the extrinsic and intrinsic parameters is:

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \end{pmatrix}=\frac{1}{t_z}\begin{pmatrix} G_x\cdot r_{11}+x_c\cdot r_{31} \\ G_x\cdot r_{12}+x_c\cdot r_{32} \\ G_x\cdot t_x+x_c\cdot r_{32} \\ G_y\cdot r_{21}+y_c\cdot r_{31} \\ G_y\cdot r_{22}+y_c\cdot r_{32} \\ G_y\cdot t_y+y_c\cdot r_{32} \\ r_{31} \\ r_{32} \end{pmatrix} \tag{10}$$

From this last relationship and the orthonormality of rotation matrix R, we observe that we have to find 16 unknowns (12 for R and $\underline{T}$ and 4 for x_c, y_c, G_x, G_y) with (8 linear + 6 quadratic) 14 equations only. Therefore, in order to have as many unknowns as equations, one needs at least a second calibration plane. In this case, one has 28 unknowns and 28 equations. See in Table 1 how many unknowns follow depending on the number of planes used for calibration purpose.

3. Calibration procedure

After the estimation of vectors $\underline{\theta}^{(1)}$, $\underline{\theta}^{(2)}$, ..., $\underline{\theta}^{(n)}$ respectively representing the n projective transformations (10) between the calibration planes and the image plane, the properties of the rotation matrix R can be used in order to determine intrinsic parameters (x_c,y_c,G_x,G_y).

Namely, consider equation (10) partially rewritten as:

$$\begin{pmatrix} r_{11} \\ r_{12} \\ r_{21} \\ r_{22} \\ r_{31} \\ r_{32} \end{pmatrix}=\begin{pmatrix} (\theta_1-x_c\theta_7)\cdot t_z/G_x \\ (\theta_2-x_c\theta_8)\cdot t_z/G_x \\ (\theta_4-y_c\theta_7)\cdot t_z/G_y \\ (\theta_5-y_c\theta_8)\cdot t_z/G_y \\ \theta_7\cdot t_z \\ \theta_8\cdot t_z \end{pmatrix} \tag{11}$$

case	mono-planar	two planes	n planes ($n>2$)
number of equations $L=$ linear $Q=$ Quadratic	$14=8L+6Q$	$28=2(8L+6Q)$	$14n=n(8L+6Q)$
number of unknowns $I=$ intrinsic $E=$ extrinsic	$16=12E+4I$	$28=24E+4I$	$12nE+4I$
system	undetermined		over-determined

Table 1: Number of equations/unknowns versus the number of planes ($k_1=0$, $G_x>0$, $G_y>0$ and $t_x>0$).

By the three quadratic relations between the components of the first two columns of R ($\underline{C}_1\bullet\underline{C}_2=0$, $\|\underline{C}_1\|=1$ and $\|\underline{C}_2\|=1$) one obtains:

$$r_{11}\cdot r_{12}+r_{21}\cdot r_{22}+r_{31}\cdot r_{32}=0 \tag{12}$$

$$r_{11}^2+r_{21}^2+r_{31}^2=1 \tag{13}$$

$$r_{12}^2+r_{22}^2+r_{32}^2=1 \tag{14}$$

By substituting the values of (11) in (12), one obtains:

$$\theta_1\theta_2/G_x^2+\theta_4\theta_5/G_y^2-(\theta_1\theta_8+\theta_2\theta_7)\cdot x_c/G_x^2-(\theta_5\theta_7+\theta_4\theta_8)\cdot y_c/G_y^2+ \\ +\theta_7\theta_8\cdot\left[1+(x_c/G_x)^2+(y_c/G_y)^2\right]=0 \tag{15}$$

$$(\theta_1-x_c\theta_7)^2/G_x^2+(\theta_4-y_c\theta_7)^2/G_y^2+\theta_7^2=1/t_z^2 \tag{16}$$

$$(\theta_2-x_c\theta_8)^2/G_x^2+(\theta_5-y_c\theta_5)^2/G_y^2+\theta_8^2=1/t_z^2 \tag{17}$$

After equating (16) and (17), one obtains:

$$(\theta_1^2-\theta_2^2)/G_x^2+(\theta_4^2-\theta_5^2)/G_y^2+2\cdot(\theta_2\theta_8-\theta_1\theta_7)\cdot x_c/G_x^2+ \\ +2\cdot(\theta_5\theta_8-\theta_4\theta_7)\cdot y_c/G_y^2+(\theta_7^2+\theta_8^2)\cdot\left[1+(x_c/G_x)^2+(y_c/G_y)^2\right]=0 \tag{18}$$

Equations (15) and (18) are the relationships between intrinsic parameters of the vision system and the projective transformation coefficients which can be extracted from the estimates of parameters with at least two perspective projections (identical equations can be obtained by $\underline{\theta}^{(1)}$, $\underline{\theta}^{(2)}$, ..., $\underline{\theta}^{(n)}$. Dividing these equations by $S=\left[1+(x_c/G_x)^2+(y_c/G_y)^2\right]$, one obtains a linear system of four unknowns $w=(w_1,w_2,w_3,w_4)^T$:

$$\begin{pmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{pmatrix}=\frac{1}{S}\begin{pmatrix} 1/G_x^2 \\ 1/G_y^2 \\ x_c/G_x^2 \\ y_c/G_y^2 \end{pmatrix} \tag{19}$$

$$S = \frac{1}{1 - \frac{w_3^2}{w_1} - \frac{w_4^2}{w_2}} \tag{20}$$

which is (for a single view):

$$\underbrace{\begin{pmatrix} \theta_1\theta_2 & \theta_4\theta_5 & -(\theta_1\theta_8+\theta_2\theta_7) & -(\theta_5\theta_7+\theta_4\theta_8) \\ \theta_1^2-\theta_2^2 & \theta_4^2-\theta_5^2 & 2\cdot(\theta_2\theta_8-\theta_1\theta_7) & 2\cdot(\theta_5\theta_8-\theta_4\theta_7) \end{pmatrix}}_{a} \cdot \underline{w} = -\underbrace{\begin{pmatrix} \theta_7\theta_8 \\ \theta_7^2-\theta_8^2 \end{pmatrix}}_{\underline{b}} \tag{21}$$

By the projective transformation coefficients of multiple calibration planes, one can build matrix A, vector $\underline{b}_{2n}$ and the LMS solution $\underline{w}^*$:

$$A = \left(a^{(1)T}, a^{(2)T}, \cdots, a^{(n)T}\right)^T \qquad \underline{b}_{2n} = \left(\underline{b}^{(1)T}, \underline{b}^{(2)T}, \cdots, \underline{b}^{(n)T}\right)^T \tag{22}$$

$$A\underline{w} = \underline{b}_{2n} \Rightarrow \text{the LMS solution is: } \underline{w}^* = \left(A^T A\right)^{-1} A^T \underline{b}_{2n} \tag{23}$$

If we assume that G_x, G_y and t_z are positive, we can recover (but it's not necessary) a unique solution for all extrinsic parameters with the previous equations. In conclusion, by the use of the six quadratic relationships between rotation components, we obtained:

- 2 relationships between intrinsic parameters given by (20),
- 1 relationship in order to recover the depth (i.e. t_z given by (16) or (17)),
- 3 relationships in order to derive the third column of the rotation R (vector product $C_1 \times C_2 = 0$).

The previous estimates can be improved with respect to the noise and distortion parameter. For each perspective projection π, consider vector $\underline{\chi} = (\underline{\theta}, k_1, k_2)^T$ formed by eight projective variables and the distortion ($k_2 = k_1\alpha^2$). Equations (5) and (6) can be approximated by ignoring the dependence of (x_c, y_c) in the distortion. In this way, one decouples the estimation of (x_c, y_c) from that of k_1:

$$(x - x_c)\cdot D \approx x - x_c + x\cdot\left(k_1\cdot x^2 + k_2\cdot y^2\right) = P_x(x,y) - x_c \tag{24}$$

$$(y - y_c)\cdot D \approx y - y_c + y\cdot\left(k_1\cdot x^2 + k_2\cdot y^2\right) = P_y(x,y) - x_c \tag{25}$$

with:

$$P_x(x,y) = \frac{\theta_1 X + \theta_2 Y + \theta_3}{\theta_7 X + \theta_8 Y + 1} \tag{26}$$

$$P_y(x,y) = \frac{\theta_3 X + \theta_4 Y + \theta_6}{\theta_7 X + \theta_8 Y + 1} \tag{27}$$

A further linearization can be applied if we consider the weak perspective in the distortion effect, that is:

$$x\cdot(\theta_7 X + \theta_8 Y + 1) + k_1\cdot x^3 + k_2\cdot xy^2 - \theta_1 X - \theta_2 Y - \theta_3 = 0 \tag{28}$$

$$y\cdot(\theta_7 X + \theta_8 Y + 1) + k_1\cdot yx^2 + k_2\cdot y^3 - \theta_4 X - \theta_5 Y - \theta_6 = 0 \tag{29}$$

Then, the approximated linear system (for each $\underline{\theta}$) becomes:

$$\begin{pmatrix} -X & -Y & -1 & 0 & 0 & 0 & xX & xY & x^3 & xy^2 \\ 0 & 0 & 0 & -X & -Y & -1 & yX & yY & yx^2 & y^3 \end{pmatrix} \cdot \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \\ \theta_7 \\ \theta_8 \\ k_1 \\ k_2 \end{pmatrix} = \begin{pmatrix} -x \\ -y \end{pmatrix} \tag{30}$$

The LMS solution of the system provides better estimates of the projective transformation parameters and gives one initial estimate of the distortion parameter.

4. Performance evaluation

4.1. Simulation results

We have conducted a number of simulations with added noise for a number of 30 different space configuration planes. Table 2 and 3 shows the results obtained with additive white uniform noise with standard deviation $\varepsilon = 0.5$ and $\varepsilon = 1$ for 500 simulations and with 72 data points (see Table 2 and 3).

Intrinsic parameters	estimated value	exact value
G_x	599.921	600.000
G_y	699.921	700.000
x_c	13.069	13.000
y_c	−7.883	−8.000
k_1	6.0e − 11	1.0e − 8
Extrinsic parameters (exact value)		
t_x	−20.066	−20.000
t_y	24.980	25.000
t_z	499.982	500.000
θ_x	15.007	15.000
θ_y	−9.997	−10.000
θ_z	105.001	105.000
Mean quadratic error (pixel)	3.2e − 2	
Extrinsic parameters (second plane)		
t'_x	−20.052	−20.000
t'_y	24.946	25.000
t'_z	399.975	400.000
θ'_x	30.002	30.000
θ'_y	19.996	20.000
θ'_z	5.005	5.000
Mean quadratic error (pixel)	3.2e − 2	

Table 2: Simulated results with additive noise of standard deviation $\varepsilon = 0.5$ pixel with 500 simulated images.

It's worth noting that:

- in all simulation conditions, distortion is not well estimated and provides a bias in the estimation of all other parameters even after optimisation process,
- the optimisation process tends to reduce this bias in a significant way,
- the bias possibly induced by the added noise is not significant with respect to that of the distortion, for a great number of images samples.

Intrinsic parameters	estimated value	exact value
G_x	*599.312*	*600.000*
G_y	*699.515*	*700.000*
x_c	*12.908*	*13.000*
y_c	*−8.574*	*−8.000*
k_1	$3.0e-10$	$1.0e-8$
Extrinsic parameters (exact value)		
t_x	*−20.006*	*−20.000*
t_y	*25.145*	*25.000*
t_z	*499.846*	*500.000*
θ_x	*14.958*	*15.000*
θ_y	*−9.992*	*−10.000*
θ_z	*105.004*	*105.000*
Mean quadratic error (pixel)	$6.4e-2$	
Extrinsic parameters (second plane)		
t'_x	*−19.980*	*−20.000*
t'_y	*25.207*	*25.000*
t'_z	*399.867*	*400.000*
θ'_x	*29.962*	*30.000*
θ'_y	*19.978*	*20.000*
θ'_z	*4.983*	*5.000*
Mean quadratic error (pixel)	$6.6e-2$	

Table 3: Simulated results with additive noise of standard deviation $\varepsilon = 1$ pixel with 500 simulated images.

4.2. Experimental results

Actual experimentation by means of a SONY CCD camera XC75-CE with a focal length of 16 mm, and a frame grabber with 512×512 pixels is currently in progress for high resolution acquisition of images for tele-reality applications (see Table 4).
In order to reduce the noise effect, ten images have been grabbed for each configurations. By this way, the projective transformation is computed ten times and an average estimation $\langle \theta \rangle$ is also obtained. We can notice that deviation between estimated values of intrinsic parameters and those computed by the data manufacturer are significant, only the ratio

$$\left(G_x/G_y\right)_{estimated} \approx \left(G_x/G_y\right)_{data\ manufacturer} \approx 1.4 \qquad (31)$$

is quite constant (and independent of the focal lens value).
For all $N = 50$ configurations, the mean quadratic error ε_c is computed in the image space:

$$\varepsilon_c = \sum_{j=1}^{s} \left(x_j - \xi_j\right)^2 + \left(y_j - \upsilon_j\right)^2 \qquad (32)$$

where $\left(x_j, y_j\right)$ are the 2D image data, and $\left(\xi_j, \upsilon_j\right)$ are the estimations of the localisation of the re-projected 2D features with the projective transformation computed with the proposed method (with s features).

Intrinsic parameters	estimated value	data manufacturer
G_x	1327 ± 25	*1030.0*
G_y	1861 ± 25	*1443.0*
x_c	-4 ± 11	*0.0*
y_c	-17 ± 11	*0.0*
k_1	$3.0e-10$	—
Max. mean quadratic error $\varepsilon_{\max}$ (pixel)	*0.37*	

Table 4: Experimental results (intrinsic parameters) with fifty configurations.

$$\varepsilon_{\max} = \underset{1 \le c \le N}{Max} \left\{\varepsilon_c\right\} \qquad (33)$$

This $\varepsilon_{\max}$ value provides an estimation of the general accuracy of this method. A small value of $\varepsilon_{\max}$ ensures the constancy of all intrinsic parameters for the entire workspace enclosing the configurations chosen. This is also a validation of Tsai's model.

5. Conclusions

A procedure whose characteristics and simulations make it suitable to in situ camera calibration for high resolution image acquisition has been presented. The procedure is currently being used in the field of tele-reality applications.
Besides image mosaicing, the proposed procedure can be used for many applications such as object recognition and localisation (for which a calibrated camera allows one to use rigid transformation, reducing the correspondence search), stereo-vision (for all estimation of the Essential matrix), object modelling and tracking, non-planar trajectory estimation in robot vision and robot calibration.
Other fields of application are photogrammetry surveying: the use of this procedure avoid the need of a metric camera, so a common reflex camera can be engaged for the purpose and less expensive surveying tasks are endorsed.
Obviously the method needs to be strengthened and more precise algorithms must be developed to ensure the goodness of the results and, at the same time, a broader use of the algorithm purposed. In particular precision with regards the k_1 parameter is very important because it is the only distortion parameter estimated, so, the better is the estimation of distortion the better is the result the algorithm returns and the wider is the range of camera the algorithm can calibrate.

6. Acknowledgement

This work was partially supported by the MURST (40%) funds for the year 1997.