3D calibration and stereoscopic view

February 20, 2015

1 Tsai model for camera calibration

Camera calibration consists in the estimation of a model relating the physical coordinates (x, y, z) to the image coordinates (X', Y'). We use the classical pinhole perspective projection model which depends on eleven parameters. The transform is performed as follows (done by the function *px_ XYZ.m* in the package uvmat):

1. A rotation and translation to express position in the 3D coordinates (x_c, y_c, z_c) linked to the camera sensor, with origin at the center of the optical axis on the image sensor, and $z_c$along the optical axis outward (see sketch below).

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \quad (1)$$

2. A projection on the sensor plane.

$$\begin{aligned} X &= x_c/z_c \\ Y &= y_c/z_c \end{aligned} \quad (2)$$

   Those correspond to the tangent of the viewing angle.

3. A rescaling factor and nonlinear quadratic distortion to express the coordinates X', Y' on the sensor in pixels.

$$\begin{aligned} X' &= f_x\,[1 + k_c(X^2 + Y^2)]X + C_x \\ Y' &= f_y\,[1 + k_c(X^2 + Y^2)]Y + C_y \end{aligned} \quad (3)$$

   The 'focal length' is expressed in units of pixel size on the sensor, so it can take a different value f_x and f_y along each axis for non-square pixels. For a focus at infinity, it should fit with the true focal length of the objective lens (normalized by the sensor pixel size), but slightly higher for a focus at close distance. A geometric distortion has been introduced as a first order quadratic correction assumed axisymmetric around the optical axis,

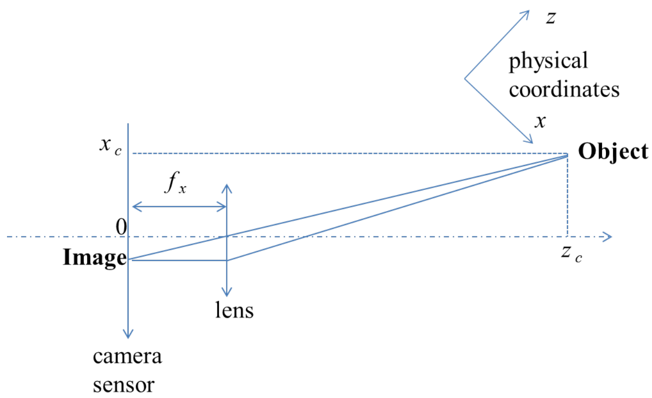

Figure 1: sketch of the pinhole camera model

with coefficient k_c. The parameters C_x and C_y represents a translation of the coordinate origin form the optical axis to the image lower left corner, so it must be equal to the half of the pixel number in each direction for a well centered sensor.

The transform is therefore defined by 17 parameters, among which 5 are intrinsic $(f_x, f_y, k_c, C_x, C_y)$, as they depend only on the optical system, and the other ones are extrinsic, as they depend on the rotation and translation of the camera with respect to its environment. Note that the rotation matrix r_i depends only on 3 independent parameters, which are the rotation angles, so there are 6 extrinsic parameters.

2 From image to physical coordinates

2.1 The general reverse transform

Retrieving the three physical coordinates from the two image coordinates of course requires additional information. The simplest case is to know the z position, or more generally the plane in which the object lies. The other possibility is to observe the same point with the two cameras, so that four numbers are available, the coordinates in each image, to determine the three physical coordinates.

In all cases, the equations (2) can be expressed as the linear system $x_c - Xz_c = 0$, $y_c - Yz_c = 0$, which writes, using (1):

$$\begin{aligned} A_{11}x + A_{12}y + A_{13}z &= XT_z - T_x \\ A_{21}x + A_{22}y + A_{23}z &= YT_z - T_y \end{aligned} \tag{4}$$

where

$$\begin{cases} A_{11} = r_1 - r_7 X\,, \quad A_{12} = r_2 - r_8 X \; A_{13} = r_3 - r_9 X \\ A_{21} = r_4 - r_7 Y\,, \quad A_{22} = r_5 - r_8 Y \; A_{23} = r_6 - r_9 Y \end{cases} \tag{5}$$

and X and Y can be obtained from the image coordinates X' and Y' by solving the equation system 3 which depends only on the intrinsic parameters. Since the quadradic deformation is weak, it can be first inversed linearly as

$$\begin{cases} X \simeq (X' - C_x) f_x^{-1} \\ Y \simeq (Y' - C_y) f_y^{-1} \end{cases} \tag{6}$$

Then in a second step, using these values of X and Y to estimate the quadratic correction,

$$\begin{cases} X = (X' - C_x) f_x^{-1} \left[1 + k_c f_x^{-2} (X' - C_x)^2 + k_c f_y^{-2} (Y' - C_y)^2\right]^{-1} \\ Y = (Y' - C_y) f_y^{-1} \left[1 + k_c f_x^{-2} (X' - C_x)^2 + k_c f_y^{-2} (Y' - C_y)^2\right]^{-1} \end{cases} \tag{7}$$

By plugging these results into (4), we get a linear system of two equations for the unkwown x, y, z.

2.2 Case of points in a known plane

In the case of a known plane, of equation $z = ax + by + c$, the system (4) reduces to the 2D system:

$$\begin{array}{c} A'_{11}\, x + A'_{12}\, y = X T_z - c A_{13} - T_x \\ A'_{21}\, x + A'_{22}\, y = Y T_z - c A_{23} - T_y \end{array} \tag{8}$$

with the definitions,

$$\begin{cases} A'_{11} = A_{11} + a A_{13}\,, \quad A'_{12} = A_{12} + b A_{13} \\ A_{21} = A_{21} + a A_{23}\,, \quad A'_{22} = A_{22} + b A_{23} \end{cases} \tag{9}$$

whose solution is

$$\begin{cases} x = \frac{-A'_{22}(X T_z - T_x) + A'_{12}(Y T_z - T_y) + c(A'_{22} A'_{13} - A'_{12} A'_{23})}{A'_{11} A'_{22} - A'_{12} A'_{21}} \\ y = \frac{-A'_{21}(X T_z - T_x) + A'_{11}(Y T_z - T_y) + c(A'_{21} A'_{13} - A'_{11} A'_{23})}{A'_{11} A'_{22} - A'_{12} A'_{21}} \end{cases} \tag{10}$$

2.3 Stereoscopic view

Now we assumed that we have identified the same points in the two images. This can be done by identifcation of specific features, like a grid of projected dots, or by image correlation between the two images. The latter is possible only to measure small displacements with respect to a reference plan. Once the points are identified, we get the position $(X', Y') = (X'_a, Y'_a)$ on image a and position $(X', Y') = (X'_b, Y'_b)$ on image b, for the same physical position (x, y, z). We then get a set of 4 linear equations of the type (4) which determines the 3

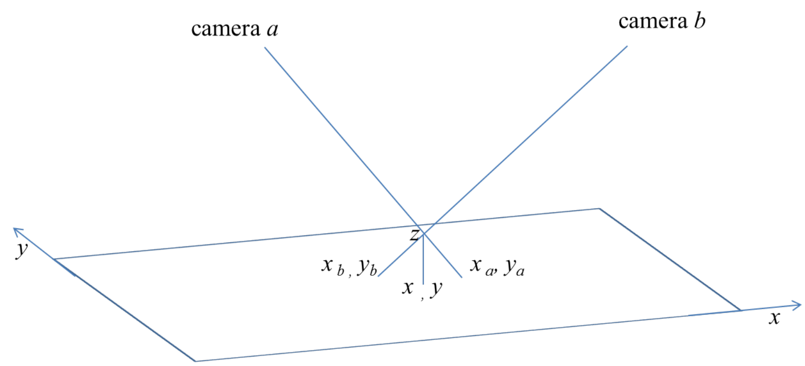

Figure 2: equivalent physical coordinates on the reference plane

physical coordinates (with redondancy). Each set of two equations determines a line along the sight of view, and their intersection gives the actual position.

When using image correlation, we define a grid of measurement points, center of a correlation box in the image, at which we get the displacement between the two images by optimizing the correlation. Since image correlation works for small displacements, it is needed to first transform each image in physical coordinates, assuming that all the points are in the reference plane, as described in section 2.2. This defines equivalent physical coordinates (x_a, y_a) , which are the x, y position after projection on the reference plane along the direction of vision (see Fig. 2). They satisfy (8) so that, introducing a subsrcipt a to specify the calibration parameters belonging to camera a, and assuming that the reference plane is $z = cte = c$ for simplicity,

$$\begin{array}{c} A_{11}\, x_a + A_{12}\, y_a = X_a T_{za} - cA_{13} - T_{xa} \\ A_{21}\, x_a + A_{22}\, y_a = Y_a T_{za} - cA_{23} - T_{ya} \end{array} \tag{11}$$

We similarly define (x_b, y_b) associated with camera b with translation constants T_{xb}, T_{yb}, T_{zb} and rotation matrix s , with coefficients B defined like A ,

$$\begin{array}{c} B_{11}\, x_b + B_{12}\, y_b = X_b T_{zb} - cB_{13} - T_{xb} \\ B_{21}\, x_b + B_{22}\, y_b = Y_b T_{zb} - cB_{23} - T_{yb} \end{array} \tag{12}$$

Comparing these relations to (4), we get

$$\begin{array}{c} A_{11}\,(x - x_a) + A_{12}\,(y - y_a) + A_{13}(z - c) = 0 \\ A_{21}\,(x - x_a) + A_{22}\,(y - y_a) + A_{23}(z - c) = 0 \end{array} \tag{13}$$

from which it results

$$\begin{cases} x - x_a = D_{xa}(z - c) \\ y - y_a = D_{ya}(z - c) \end{cases} \tag{14}$$

where

$$\begin{cases} D_{xa} = \frac{A_{12}A_{23}-A_{22}A_{13}}{A_{11}A_{22}-A_{12}A_{21}} \\ D_{ya} = \frac{A_{21}A_{13}-A_{11}A_{23}}{A_{11}A_{22}-A_{12}A_{21}} \end{cases} \tag{15}$$

with similar relations for camera b. It results that the observed displacement between the two images is

$$\begin{cases} x_b - x_a = (D_{xb} - D_{xa})(z - c) \\ y_b - y_a = (D_{yb} - D_{ya})(z - c) \end{cases} \tag{16}$$

so that the displacement $z - c$ can be in principle determined by the paralax effect in either x or y directions, providing the corresponding coefficients are non-zero. Since $z - c$ is determined by 2 relations, a condition of solvability is required on the data,

$$(D_{xb} - D_{xa})(y_b - y_a) - (D_{yb} - D_{ya})(x_b - x_a) = 0 \tag{17}$$

This is not satisfied exactly in general, so that we introduce a small error ϵ_x and ϵ_y on $x_b - x_a$ and $y_b - y_a$ respectively, so that (16) is replaced by

$$\begin{cases} \epsilon_x = (D_{xb} - D_{xa})(z - c) - (x_b - x_a) \\ \epsilon_y = (D_{yb} - D_{ya})(z - c) - (y_b - y_a) \end{cases} \tag{18}$$

and we seek the displacement $z - c$ which minimizes the quadratic error $\epsilon_x^2 + \epsilon_y^2$. The condition of vanishing derivative leads to

$$(D_{xb} - D_{xa})\epsilon_x + (D_{yb} - D_{ya})\epsilon_y = 0 \tag{19}$$

so that the two errors are linearly related by

$$\begin{aligned} \epsilon_x &= -\lambda(D_{yb} - D_{ya}) \\ \epsilon_y &= \lambda(D_{xb} - D_{xa}) \end{aligned} \tag{20}$$

where λ is a constant. With this result, (16) is replaced by

$$\begin{cases} x_b - x_a - \lambda(D_{yb} - D_{ya}) = (D_{xb} - D_{xa})(z - c) \\ y_b - y_a + \lambda(D_{xb} - D_{xa}) = (D_{yb} - D_{ya})(z - c) \end{cases} \tag{21}$$

The error factor λ can be eliminated by taking the appropriate linear combination of these two relations,

$$z - c = \frac{(D_{xb} - D_{xa})(x_b - x_a) + (D_{yb} - D_{ya})(y_b - y_a)}{(D_{xb} - D_{xa})^2 + (D_{yb} - D_{ya})^2} \tag{22}$$

which uses the informations on the observed displacements in the x and y directions in proportion to their respective sensitivity to the z displacement. By another linear combination, we find the corresponding error estimate

$$\lambda = \frac{(D_{yb} - D_{ya})(x_b - x_a) - (D_{xb} - D_{xa})(y_b - y_a)}{(D_{xb} - D_{xa})^2 + (D_{yb} - D_{ya})^2} \tag{23}$$

from which $\epsilon_x$and ϵ_y are obtained by (20). Then the rms error can be estimated as

$$E = [(\epsilon_x^2 + \epsilon_y^2)/2]^{1/2} = \frac{1}{\sqrt{2}} \frac{|(D_{yb} - D_{ya})(x_b - x_a) - (D_{xb} - D_{xa})(y_b - y_a)|}{[(D_{xb} - D_{xa})^2 + (D_{yb} - D_{ya})^2]^{1/2}} \tag{24}$$

This error is in physical length units so it has to be translated into pixel units. The scaling factor (in pixels/length unit) can be estimated as $(f_x^2 + f_y^2)^{1/2}/(2T_z)$, obtained by assuming $z_c \sim T_z$ in (1) to (3), and taking an average focal length. Then the error in pixel units can be estimated as

$$E' = \frac{(f_x^2 + f_y^2)^{1/2}}{2\sqrt{2}T_z} \frac{|(D_{yb} - D_{ya})(x_b - x_a) - (D_{xb} - D_{xa})(y_b - y_a)|}{[(D_{xb} - D_{xa})^2 + (D_{yb} - D_{ya})^2]^{1/2}} \tag{25}$$

Since image correlation provides measurments with precision in principle better than 1/2 pixel, it is possible to exclude 'false' displacements as inconsistent when E' is beyond a threshold or order unity.

3 Stereoscopic PIV:

3.1 Geometric transform for small displacements:

We now assume that particles are close to the reference plane, and we want to get the three velocity components by comparing the displacements from two successive times viewed by each camera, observed in image coordinates. To avoid interpolation procedures on the images, we keep the image coordinates instead of the transformed coordinates (x_a, y_a). From the general relation (4), a small displacement $(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$, is related to the image displacement by differentiation of (4),

$$\begin{aligned} A_{11}\mathrm{d}x + A_{12}\mathrm{d}y + A_{13}\mathrm{d}z = T_a \mathrm{d}X_a \\ A_{21}\mathrm{d}x + A_{22}\mathrm{d}y + A_{23}\mathrm{d}z = T_a \mathrm{d}Y_a \end{aligned} \tag{26}$$

for a first camera denoted by subscript a. We have used the expression $\mathrm{d}(A_{11}x) = A_{11}\,\mathrm{d}x + \mathrm{x}\,\mathrm{d}A_{11} = A_{11}\,\mathrm{d}x - r_7 x\,\mathrm{d}X_a$ and similar ones for the other terms of (4), leading to the right hand term of (26) with the notation $T_a = r_7 x + r_8 y + r_9 z + T_{za}$. A similar relation is obtained for camera b, with a rotation matrix s, and coefficients B_{ij} defined like A_{ij}.

$$\begin{cases} B_{11} = s_1 - s_7 X_b\,, & B_{12} = s_2 - s_8 X_b \; B_{13} = s_3 - s_9 X_b \\ B_{21} = s_4 - s_7 Y_b\,, & B_{22} = s_5 - s_8 Y_b \; B_{23} = s_6 - s_9 Y_b \end{cases} \tag{27}$$

This leads to the second set of two equations,

$$\begin{aligned} B_{11}\mathrm{d}x + B_{12}\mathrm{d}y + B_{13}\mathrm{d}z = T_b \mathrm{d}X_b \\ B_{21}\mathrm{d}x + B_{22}\mathrm{d}y + B_{23}\mathrm{d}z = T_b \mathrm{d}Y_b \end{aligned} \tag{28}$$

leading to a system of 4 equations with 3 unknown, which has to be solved to get the physical displacement components from the image displacements $\mathrm{d}X_a, \mathrm{d}Y_a, \mathrm{d}X_b, \mathrm{d}Y_b$.

Note that the actual image displacements $\mathrm{d}X'_a, \mathrm{d}Y'_a, \mathrm{d}X'_b, \mathrm{d}Y'_b.$, expressed in pixels, are related to $\mathrm{d}X_a, \mathrm{d}Y_a, \mathrm{d}X_b, \mathrm{d}Y_b$ by (3) whose differenciation yields

$$\begin{array}{r}[1+3k_cX^2+k_cY^2]\mathrm{d}X+2k_cXY\mathrm{d}Y=f_x^{-1}\mathrm{d}X' \\ 2k_cXY\mathrm{d}X+[1+3k_cY^2+k_cX^2]\mathrm{d}Y=f_y^{-1}\mathrm{d}Y'\end{array} \tag{29}$$

which can be reversed as a solution of the linear system

$$\begin{cases}\mathrm{d}X=[-f_y(1+3k_cY^2+k_cX^2)\,\mathrm{d}X'+2k_cXY\,f_{\mathrm{x}}\mathrm{d}Y']\,D^{-1} \\ \mathrm{d}Y=[-2f_yk_cXY\,\mathrm{d}X'+(1+3k_cX^2+k_cY^2)f_x\,\mathrm{d}Y']\,D^{-1}\end{cases} \tag{30}$$

with the determinant

$$D=[1+3k_cX^2+k_cY^2][1+3k_cY^2+k_cX^2]-4k_c^2X^2Y^2 \tag{31}$$

and (X, Y) are obtained for each camera *a,b* from the image coordinates (X', Y') by the (approximate) reverse relation (7).

3.2 Deducing physical displacements from image displacements

To solve the system of 4 equations with 3 unknown we have a condition of solvability on the image displacements $\mathrm{d}X_a, \mathrm{d}Y_a, \mathrm{d}X_b, \mathrm{d}Y_b$, in the form of a linear combination of these quantities. In practice this is never quite satisfied due to measurement errors, so that we introduce a small error on these quantities, replacing them by $\mathrm{d}X_i+\epsilon_i$ respectively in the equations. We minimise $\sum \epsilon_i^2$ with

$$\begin{array}{l}\epsilon_{xa}=\tilde{A}_{11}dx+\tilde{A}_{12}dy+\tilde{A}_{13}dz-dX_a \\ \epsilon_{ya}=\tilde{A}_{21}dx+\tilde{A}_{22}dy+\tilde{A}_{23}dz-dY_a \\ \epsilon_{xb}=\tilde{B}_{11}dx+\tilde{B}_{12}dy+\tilde{B}_{13}dz-dX_b \\ \epsilon_{yb}=\tilde{B}_{21}dx+\tilde{B}_{22}dy+\tilde{B}_{23}dz]-dY_b\end{array} \tag{32}$$

where $\tilde{A}_{ij}=A_{ij}/T_a$ and $\tilde{B}_{ij}=B_{ij}/T_b$. The measurement error due to pixel discretisation should be in fact expressed in terms of the pixel displacement dX' instead of the angular displacement dX, which would give a different weight to the different ϵ_i in the minimisation prodedure. However these weights are nearly equal since the focal lengths f_x and $f_y$are close (generally equal) and the nonlinear deformation weak.

We have the partial derivatives

$$\begin{array}{l}\frac{1}{2}\frac{\partial}{\partial(dx)}(\epsilon_{xa}^2+\epsilon_{ya}^2+\epsilon_{xb}^2+\epsilon_{yb}^2)=\tilde{A}_{11}\epsilon_{xa}+\tilde{A}_{21}\epsilon_{ya}+\tilde{B}_{11}\epsilon_{xb}+\tilde{B}_{21}\epsilon_{yb} \\ \frac{1}{2}\frac{\partial}{\partial(dy)}(\epsilon_{xa}^2+\epsilon_{ya}^2+\epsilon_{xb}^2+\epsilon_{yb}^2)=\tilde{A}_{12}\epsilon_{xa}+\tilde{A}_{22}\epsilon_{ya}+\tilde{B}_{12}\epsilon_{xb}+\tilde{B}_{22}\epsilon_{yb} \\ \frac{1}{2}\frac{\partial}{\partial(dz)}(\epsilon_{xa}^2+\epsilon_{ya}^2+\epsilon_{xb}^2+\epsilon_{yb}^2)=\tilde{A}_{13}\epsilon_{xa}+\tilde{A}_{23}\epsilon_{ya}+\tilde{B}_{13}\epsilon_{xb}+\tilde{B}_{23}\epsilon_{yb}\end{array} \tag{33}$$

The condition of error minimisation is obtained by setting to zero each partial derivative, which yields a linear system of 3 equations

$$\begin{cases}D_{11}dx+D_{12}dy+D_{13}dz=S_1 \\ D_{21}dx+D_{22}dy+D_{23}dz=S_2 \\ D_{31}dx+D_{32}dy+D_{33}dz=S_3\end{cases} \tag{34}$$

with the symmetric matrix ($D_{ij} = D_{ji}$) defined by

$$
\begin{aligned}
D_{11} &= \tilde{A}_{11}^2 + \tilde{A}_{21}^2 + \tilde{B}_{11}^2 + \tilde{B}_{21}^2 \\
D_{12} &= \tilde{A}_{11}\tilde{A}_{12} + \tilde{A}_{21}\tilde{A}_{22} + \tilde{B}_{11}\tilde{B}_{12} + \tilde{B}_{21}\tilde{B}_{22} \\
D_{13} &= A_{11}\tilde{A}_{13} + \tilde{A}_{21}\tilde{A}_{23} + \tilde{B}_{11}\tilde{B}_{13} + \tilde{B}_{21}\tilde{B}_{23} \\
D_{22} &= \tilde{A}_{12}^2 + \tilde{A}_{22}^2 + \tilde{B}_{12}^2 + \tilde{B}_{22}^2 \\
D_{23} &= \tilde{A}_{12}\tilde{A}_{13} + \tilde{A}_{22}\tilde{A}_{23} + \tilde{B}_{12}\tilde{B}_{13} + \tilde{B}_{22}\tilde{B}_{23} \\
D_{33} &= \tilde{A}_{13}^2 + \tilde{A}_{23}^2 + \tilde{B}_{13}^2 + \tilde{B}_{23}^2
\end{aligned}
\tag{35}
$$

and the source terms

$$
\begin{cases}
S_1 = \tilde{A}_{11}dX_a + \tilde{A}_{21}dY_a + \tilde{B}_{11}dX_b + \tilde{B}_{21}dY_b) \\
S_2 = \tilde{A}_{12}dX_a + \tilde{A}_{22}dY_a + \tilde{B}_{12}dX_b + \tilde{B}_{22}dY_b) \\
S_3 = \tilde{A}_{13}dX_a + \tilde{A}_{23}dY_a + \tilde{B}_{13}dX_b + \tilde{B}_{23}dY_b)
\end{cases}
\tag{36}
$$

The displacements (dx, dy, dz) are then obtained as solution of the linear system (34), and the corresponding velocity components after division by the time interval Dt.

3.3 Practical implementation of stereoscopic PIV:

We first do usual PIV in each image series, using *civ_series*. We can use the automatic regular measurement grid in image coordinates.

We then reconstruct the physical velocities using the function *civ2vel_3C*. This requires the introduction of a regular physical grid on the reference plane (x, y) (introduced as a 'projection object'). The function interpolates the z displacements obtained by *stereo_civ* on this grid if the corresponding netcdf file is introduced. Otherwise it just assumes $\delta z = 0$ (usual stereo PIV in a laser sheet). The function creates the corresponding grids in image coordinates for each view, and interpolates the PIV data on this grid, leading to $\mathrm{d}X_a, \mathrm{d}Y_a \mathrm{d}X_b \mathrm{d}Y_b$. It then obtains the corresponding physical displacement by solving (4) at each point of the physical grid. The method gives also an error estimate

$$E' = [(f_x^2 + f_y^2)^{1/2}/4](\epsilon_{xa}^2 + \epsilon_{ya}^2 + \epsilon_{xb}^2 + \epsilon_{yb}^2)^{1/2} \tag{37}$$

after multiplication by the focal length (mean between f_x and f_y to be general) and a normalisation factor to express the error in units of pixel for each displacement measurement. This can be used to eliminate false vectors, characterized by a threshold of order unity for E' since PIV is supposed to give a precision better than 1/2 pixel.