Università Ca' Foscari Venezia

Computer Vision

Camera calibration

Filippo Bergamasco (filippo.bergamasco@unive.it)
http://www.dais.unive.it/~bergamasco
DAIS, Ca' Foscari University of Venice
Academic year 2018/2019

Università Ca' Foscari Venezia

Camera Calibration

Accurate knowledge of camera intrinsic parameters is an essential prerequisite for any kind of quantitative geometric measurement in computer vision.

Many different approaches exist:

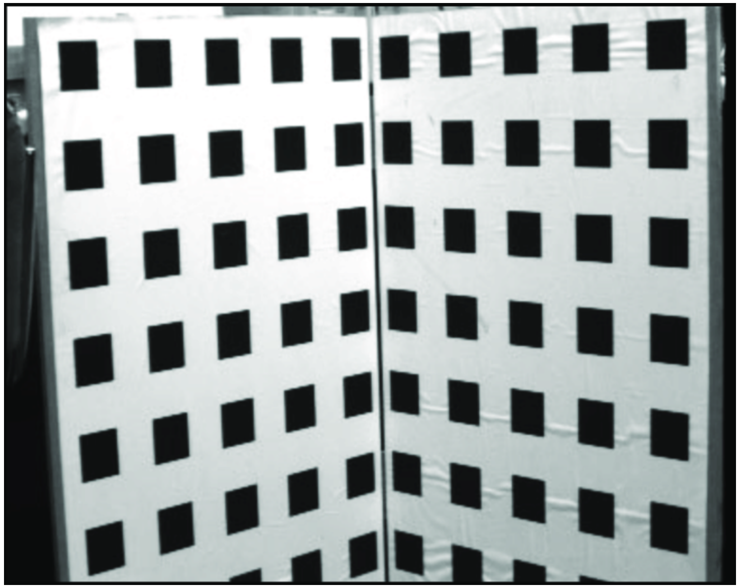
- Simple (linear) techniques allow a quick estimation of K and RT from known objects or scene itself
- Accurate camera calibration takes into account lens distortion, geometric error, imperfect calibration targets, etc.
- In general is a delicate art to master…

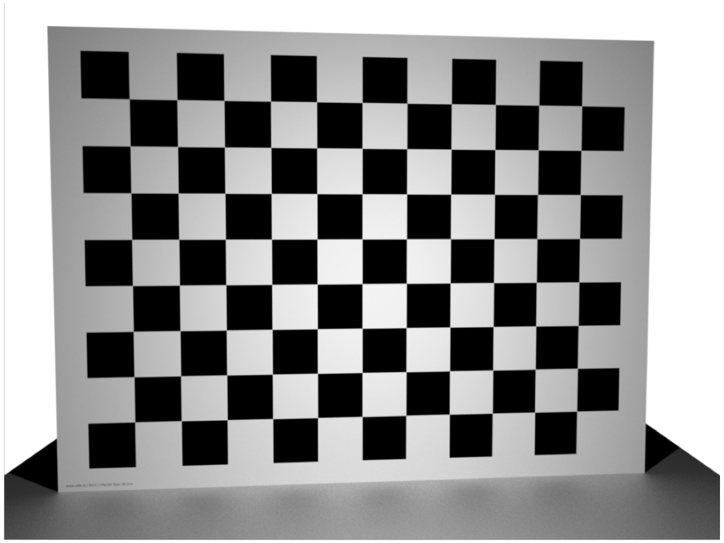
How?

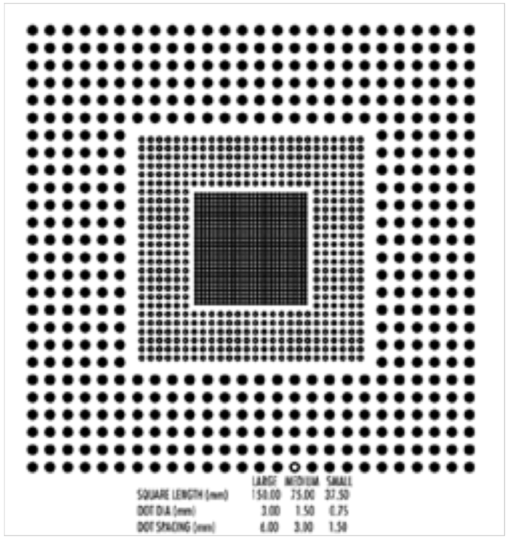
In theory parameters like focal length and principal point can be obtained by the manifacturer of the camera/lenses...

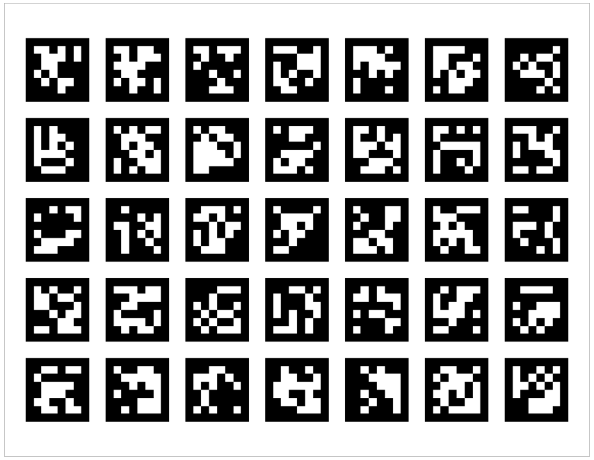
... in practice to obtain an accurate estimate of the required parameters we observe **how the imaging model behave when projecting a known object**, called «calibration target»

Calibration targets









Calibration target

The calibration target is a specially designed object in which:

- An easily distinguishable set of features is present (usually corners, circles, line intersections, etc.)
- The 3D coordinates of each feature in the target reference frame is known by design
- After projection, features are well localized in the image plane

Usually, the calibration process is performed by exposing the target multiple times (in different poses) to collect a set of 3D <-> 2D correspondences.

Computing P: basic equations

From a single target pose, the recovery of the projection matrix P is conceptually similar with the problem of computing a 2D homography H.

We have $X_1 \ldots X_n \in \mathbb{P}^3$ and $x_1 \ldots x_n \in \mathbb{P}^2$.

We want to find P such that:

$$x_i = PX_i \;\; \forall i$$

Computing P: basic equations

For each point-point correspondence we obtain a system of 2 equations in 12 unknowns:

$$\begin{pmatrix} \mathbf{0}^T & -w_i X_i^T & y_i X_i^T \\ w_i X_i^T & \mathbf{0}^T & -x_i X_i^T \end{pmatrix} \begin{pmatrix} \mathbf{P}^1 \\ \mathbf{P}^2 \\ \mathbf{P}^3 \end{pmatrix} = \mathbf{0}$$

Since the matrix P has 12 entries and 11 degrees of freedom (due to the scale), we need 11 equations to solve for P, resulting in 5 «complete» point-point correspondences plus only the x or y coordinate of the sixth image point.

Computing P: basic equations

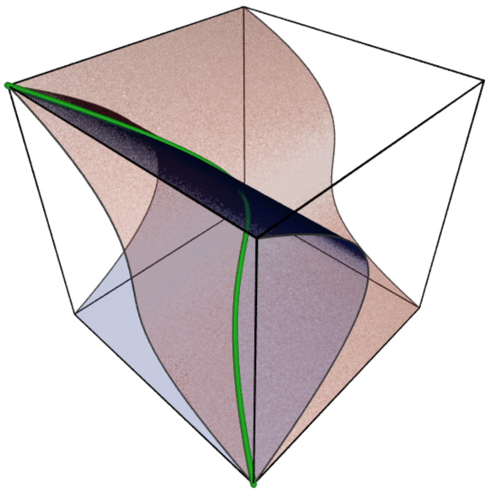
If the data is noisy we can use more than 6 correspondences to solve the system by minimizing the algebraic or geometric error.

In the former case, we need to impose a normalization constraint like $||p||^2 = 1$ and the solution can be easily obtained via SVD.

Since the algebraic error depends on the reference frame, usually the DLT is followed by a non-linear optimization to minimize the geometric error

Degenerate configurations

There are two types of configurations for which ambiguous solutions exist for P:



- Camera and all points lie on a twisted cubic

- The points all lie on the union of a plane and a single straight line containing the camera center (ie. planar targets won't work!)

Zhang's camera calibration

Z. Zhang, "*A flexible new technique for camera calibration*", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 11, pp. 1330-1334, Nov. 2000

Is a de-facto standard technique to calibrate a camera using a planar target (usually a chessboard).

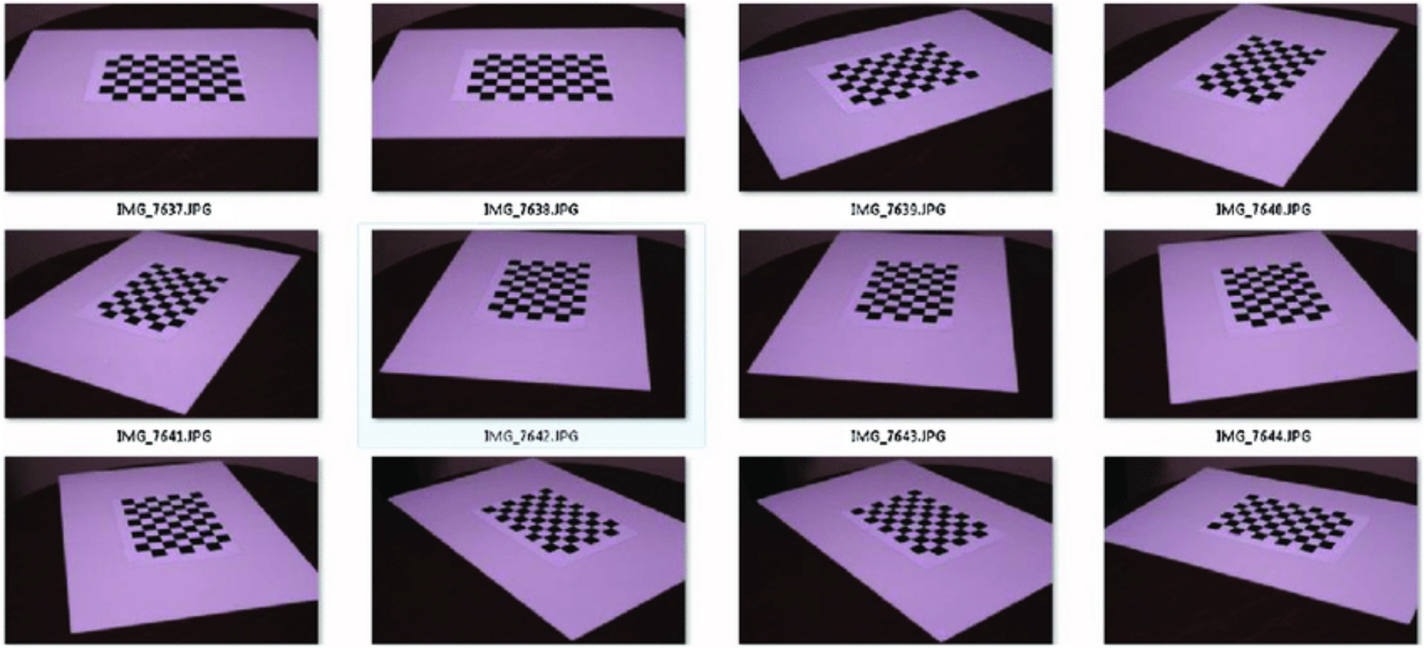
The main idea is to compute the view homographies of the target points and recover the intrinsic parameters using a closed form linear solution.

Once an initial guess is estimated, a non linear optimization refines the solution.

Data acquisition

We assume to use a planar calibration target composed by N points.

M different pictures (views) are taken by moving the calibration target in front of the camera



Homography estimation

For each picture, the homograpy H_i mapping points from the 2D target projective space to the image plane is computed.

A matrix A is composed by stacking N*2x9 matrices :

$$\begin{pmatrix} \mathbf{0}^T & -w'_i x_i^T & y'_i x_i^T \\ w'_i x_i^T & \mathbf{0}^T & -x'_i x_i^T \end{pmatrix}$$

Obtained by the constraint $x'_i \times H x_i = 0$.

The initial solution obtained by DLT is further refined to minimize the geometric error.

Structure of the homographies

Recall that, since each 3D point has coordinate Z=0, each homography has the following form:

$$H_i = \begin{pmatrix} | & | & | \\ h_{i,0} & h_{i,1} & h_{i,2} \\ | & | & | \end{pmatrix} = \lambda \mathbf{K} \begin{pmatrix} | & | & | \\ r_{i,0} & r_{i,1} & t_i \\ | & | & | \end{pmatrix}$$

Where $\mathbf{K} = \begin{pmatrix} \alpha & \gamma & u_o \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$ is the upper-triangular

matrix of intrinsic parameters common for all the views (the camera only changed its pose between the views)

Università Ca' Foscari Venezia

Structure of the homographies

Recall that, since each 3D point has coordinate Z=0, each homography has the following form:

$$H_i = \begin{pmatrix} | & | & | \\ h_{i,0} & h_{i,1} & h_{i,2} \\ | & | & | \end{pmatrix} = \lambda \mathbf{K} \begin{pmatrix} | & | & | \\ r_{i,0} & r_{i,1} & t_i \\ | & | & | \end{pmatrix}$$

Where $r_{i,0}, r_{i,1}$ are the first two columns of the rotation matrix **R** describing the camera pose in the i^{th} picture wrt. the target.

t_i is the 3D translation vector of the i^{th} pose.

Main goal

Factorize **K** out of the homographies $H_0, H_1 \dots H_M$

We do that by deriving two fundamental constraints on the intrinsic parameters for a given homography **H**

Once **K** is known, the extrinsic parameters can be estimated for each pose by multiplying **H** by $\mathbf{K}^{-1}$ and fixing the scale λ considering that R is an orthonormal matrix (more details later on...)

Constraints

$$H = \begin{pmatrix} | & | & | \\ h_0 & h_1 & h_2 \\ | & | & | \end{pmatrix} = \lambda \mathbf{K} \begin{pmatrix} | & | & | \\ r_0 & r_1 & t_i \\ | & | & | \end{pmatrix}$$

For **R** to be a valid rotation matrix, its columns must be orthonormal. This implies that:

$$r_0^T \, r_1 = r_1^T \, r_0 = 0$$

$$r_0^T \, r_0 = r_1^T \, r_1 = 1$$

Constraints

$$H = \begin{pmatrix} | & | & | \\ h_0 & h_1 & h_2 \\ | & | & | \end{pmatrix} = \lambda \mathbf{K} \begin{pmatrix} | & | & | \\ r_0 & r_1 & t_i \\ | & | & | \end{pmatrix}$$

We got that

$$h_0 = \lambda \mathbf{K} r_0$$
$$h_1 = \lambda \mathbf{K} r_1$$

$$\mathbf{K}^{-1} h_1 = \lambda r_1$$
$$\mathbf{K}^{-1} h_0 = \lambda r_0$$

$$h_0^T \mathbf{K}^{-T} = \lambda r_0^T$$
$$h_1^T \mathbf{K}^{-T} = \lambda r_1^T$$

Since:

$$(AB)^T = B^T A^T$$

Constraints

Substituting $h_0^T \mathbf{K}^{-T} = \lambda r_0^T$ and $h_1^T \mathbf{K}^{-T} = \lambda r_1^T$

into:

$$r_0^T \; r_1 = r_1^T \; r_0 = 0 \qquad\qquad r_0^T \; r_0 = r_1^T \; r_1 = 1$$

We obtain the following constraints on the intrinsic matrix **K** given H:

$$h_0^T \; K^{-T} \; K^{-1} \; h_1 = 0$$

$$h_0^T \; K^{-T} \; K^{-1} \; h_0 = h_1^T \; K^{-T} \; K^{-1} \; h_1$$

Università Ca' Foscari Venezia

Solving for K

For estimating the camera intrinsics, Zhang substitutes the expression $K^{-T}K^{-1}$ with:

$$K^{-T}\,K^{-1} = \begin{pmatrix} B_0 & B_1 & B_3 \\ B_1 & B_2 & B_4 \\ B_3 & B_4 & B_5 \end{pmatrix} = B$$

The matrix B is symmetric and composed by the following quantities:

$$B_0 = \frac{1}{\alpha^2} \quad B_1 = -\frac{\gamma}{\alpha^2\beta} \quad B_2 = \frac{\gamma^2}{\alpha^2\beta^2} + \frac{1}{\beta^2} \quad B_3 = \frac{v_0\gamma - u_0\beta}{\alpha^2\beta}$$

$$B_4 = -\frac{\gamma(v_0\gamma - u_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} \qquad B_5 = \frac{(v_0\gamma - u_0\beta)^2}{\alpha^2\beta^2} + \frac{v_0^2}{\beta^2} + 1$$

Solving for K

$$h_0^T\ K^{-T}\ K^{-1}\ h_1 = 0$$

$$h_0^T\ K^{-T}\ K^{-1}\ h_0 = h_1^T\ K^{-T}\ K^{-1}\ h_1$$

Considering the matrix B previously defined, the above constraints can be rewritten as:

$$h_0^T\ B\ h_1 = 0$$

$$h_0^T\ B\ h_0 - h_1^T\ B\ h_1 = 0$$

Solving for K

If we consider the vector $b = (B_0, B_1, B_2, B_3, B_4, B_5)^T$ we can derive the following identity:

$$h_p^T \; B \; h_q = v_{p,q}(H) b$$

p^{th} column of H (p=0,1,2)

q^{th} column of H (q=0,1,2)

Solving for K

If we consider the vector $b = (B_0, B_1, B_2, B_3, B_4, B_5)^T$

we can derive the following identity:

$$h_p^T \; B \; h_q = v_{p,q}(H) b$$

$$\boldsymbol{v}_{p,q}(\mathbf{H}) = \begin{pmatrix} H_{0,p} \cdot H_{0,q} \\ H_{0,p} \cdot H_{1,q} + H_{1,p} \cdot H_{0,q} \\ H_{1,p} \cdot H_{1,q} \\ H_{2,p} \cdot H_{0,q} + H_{0,p} \cdot H_{2,q} \\ H_{2,p} \cdot H_{1,q} + H_{1,p} \cdot H_{2,q} \\ H_{2,p} \cdot H_{2,q} \end{pmatrix}^{\mathsf{T}}$$

6-dimensional row vector obtained from
an estimated homography **H**

Università
Ca' Foscari
Venezia

Solving for K

For a particular estimated homography H, the constraints

$$h_0^T \; B \; h_1 = 0$$

$$h_0^T \; B \; h_0 - h_1^T \; B \; h_1 = 0$$

Can be reformulated as a pair of linear equations:

$$\begin{pmatrix} \boldsymbol{v}_{0,1}(\mathbf{H}) \\ \boldsymbol{v}_{0,0}(\mathbf{H}) - \boldsymbol{v}_{1,1}(\mathbf{H}) \end{pmatrix} \cdot \boldsymbol{b} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

2x6 matrix

6 coefficients

Solving for K

Considering all the homographies $H_0, H_1 \dots H_M$,
we obtain an overdetermined system of homogeneous linear equations

$$\begin{pmatrix} v_{0,1}(\mathbf{H}_0) \\ v_{0,0}(\mathbf{H}_0) - v_{1,1}(\mathbf{H}_0) \\ \hdashline v_{0,1}(\mathbf{H}_1) \\ v_{0,0}(\mathbf{H}_1) - v_{1,1}(\mathbf{H}_1) \\ \hdashline \vdots \\ \hdashline v_{0,1}(\mathbf{H}_i) \\ v_{0,0}(\mathbf{H}_i) - v_{1,1}(\mathbf{H}_i) \\ \hdashline \vdots \\ \hdashline v_{0,1}(\mathbf{H}_{M-1}) \\ v_{0,0}(\mathbf{H}_{M-1}) - v_{1,1}(\mathbf{H}_{M-1}) \end{pmatrix} \cdot \boldsymbol{b} = \begin{pmatrix} 0 \\ 0 \\ \hdashline 0 \\ 0 \\ \hdashline \vdots \\ \hdashline 0 \\ 0 \\ \hdashline \vdots \\ \hdashline 0 \\ 0 \end{pmatrix}$$

that can be solved in a least-squares sense via SVD:

$$\min \|Vb\| \quad s.t. \ \|b\| = 1$$

Recovering intrinsic parameters

Università Ca' Foscari Venezia

Once the matrix B is obtained, it must be factorized to recover the intrinsic parameters from K:

$$B = \lambda\, \mathbf{K}^{-T}\, \mathbf{K}^{-1} = (\sqrt{\lambda}\mathbf{K}^{-T})\, (\sqrt{\lambda}\mathbf{K}^{-1})$$

Where λ is an unknown scale factor (since we estimated B via SVD by imposing $\|b\| = 1$)

Since K^{-1} has the structure $K^{-1} = \begin{pmatrix} k_0 & k_1 & k_2 \\ 0 & k_3 & k_4 \\ 0 & 0 & 1 \end{pmatrix}$

We have that $\sqrt{\lambda}\mathbf{K}^{-1}$ is upper triangular and $\sqrt{\lambda}\mathbf{K}^{-T}$ is lower triangular.

Recovering intrinsic parameters

Thus, from

$$B = \lambda \, \mathbf{K}^{-T} \, \mathbf{K}^{-1} = (\sqrt{\lambda}\mathbf{K}^{-T}) \, (\sqrt{\lambda}\mathbf{K}^{-1})$$

$\sqrt{\lambda}\mathbf{K}^{-1}$ can be recovered by performing the Cholesky decomposition of B or –B depending which one is positive definite (the positive-definite check can be simply performed by looking if all diagonal elements are non-negative).

After that, the scale factor λ is recovered by normalizing $\mathbf{K}^{-1}$ so that that the lower-right element is equal to 1.

Recovering the target poses

Università Ca' Foscari Venezia

Once the camera intrinsics are recovered, the camera extrinsics (ie. the target pose) $\mathbf{R}_i$, $\mathbf{T}_i$ for each view can be calculated from the corresponding homography $\mathbf{H}_i$

$$r_0 = \lambda\, \mathbf{K}^{-1}\, h_0 \qquad r_1 = \lambda\, \mathbf{K}^{-1}\, h_1 \qquad t = \lambda\, \mathbf{K}^{-1}\, h_2$$

With the scale factor

$$\lambda = \frac{1}{\|\mathbf{K}^{-1}\, h_1\|}$$

And:

$$r_2 = r_0 \times r_1$$

Recovering the target poses

Note that, this way, there is no guarantee that r_0 and r_1 are orthogonal and hence **R** may not be a proper rotation matrix.

Zhang proposed to approximate the best rotation matrix **R** given any 3x3 matrix **Q** in sense of minimizing the Frobenius norm of the difference

$$\min_{\mathbf{R}} ||\mathbf{R} - \mathbf{Q}||_F^2 \qquad \text{subject to } \mathbf{R}^T\mathbf{R} = \mathbf{I}$$

Recovering the target poses

Università Ca' Foscari Venezia

$$\min_{\mathbf{R}} \|\mathbf{R} - \mathbf{Q}\|_F^2 \qquad \text{subject to } \mathbf{R}^T\mathbf{R} = \mathbf{I}$$

Since:

$$\begin{aligned} \|\mathbf{R} - \mathbf{Q}\|_F^2 &= \text{trace}((\mathbf{R} - \mathbf{Q})^T(\mathbf{R} - \mathbf{Q})) \\ &= 3 + \text{trace}(\mathbf{Q}^T\mathbf{Q}) - 2\text{trace}(\mathbf{R}^T\mathbf{Q}) \end{aligned}$$

We can solve the problem by maximizing trace($\mathbf{R}^T\mathbf{Q}$).

Let the SVD of Q be $\mathbf{USV}^T$. The matrix **R** maximizing trace($\mathbf{R}^T\mathbf{Q}$) is $\mathbf{R}=\mathbf{UV}^T$.

[See Zhang's technical report for details]

Final optimization

So far we assumed that the pinhole camera model holds without any distortion.

The last step of the calibration procedure is to optimize all calibration parameters (intrinsics, extrinsics and distortion) in a single non-linear system of equations starting from an initial solution obtained as previously discussed and with no distortion.

Let:

$$a = (\alpha, \beta, \gamma, u_c, v_c, k_1, k_2, k_3, p_1, p_2)$$

$$w_i = (R_i, T_i)$$

Final optimization

Given $\mathbf{x}_{i,j}$ the j[th] imaged target point in image i[th], the goal is to minimize the total reprojection error:

$$E(a, \mathbf{w}) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \|\mathbf{x}_{i,j} - P(a, w_i, X_j)\|^2$$

Where $P(a, w_i, X_j)$ is non linear function that:

- Projects X_j on the plane Z=1 to get x'_j
- Applies the distortion function mapping x'_j to its distorted coordinates
- Computes the affine 2D mapping Kx'_j

Università Ca' Foscari Venezia

Parametrizing R

Every 3D rotation matrix R consists of nine elements, despite the fact that it has only three degrees of freedom. Indeed, **R** is subject to the constraint of being an orthonormal matrix.

Such constraint is difficult to be imposed during the optimization, therefore a different parametrization is often used.

Zhang proposed to use the **Rodrigues formula**

Rodrigues formula

3D rotation is parametrized as a vector:

$$\rho = (\rho_x, \rho_y, \rho_z)$$

in which the magnitude $||\rho|| = \theta$ is equal to the rotation angle and the direction $\hat{\rho} = \frac{\rho}{||\rho||}$ is the rotation axis.

The corresponding rotation matrix R can be obtained as:

$$\mathbf{R} = \mathbf{I} + \sin\theta[\hat{\rho}]_x + (1 - \cos\theta)[\hat{\rho}]_x^2$$

where

$$[\hat{\rho}]_x = \begin{pmatrix} 0 & -\rho_z & \rho_y \\ \rho_z & 0 & -\rho_x \\ -\rho_y & \rho_x & 0 \end{pmatrix}$$

Degenerate configurations

$$h_0^T \, K^{-T} \, K^{-1} \, h_1 = 0$$

$$h_0^T \, K^{-T} \, K^{-1} \, h_0 = h_1^T \, K^{-T} \, K^{-1} \, h_1$$

Since the two equations are derived by properties of the rotation matrices, if the calibration target undergoes a pure translation no useful information is added to help with the calibration.

<u>Suggestions:</u>

- Tilt the target at every shot
- Cover all the visible area of the image with the target (avoid seeing the target always in a certain subregion of the image)
- Check the reprojection error and feel free to remove bad images