ON PROJECTION MATRIX IDENTIFICATION FOR CAMERA CALIBRATION

Michał Tomaszewski and Władysław Skarbek
Institute of Radioelectronics, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warszawa, Poland

Keywords: Projection matrix, homographic matrix, camera calibration, intrinsic parameters, Housholder symmetry.

Abstract: The projection matrix identification problem is considered with application to calibration of intrinsic camera parameters. Physical and orthogonal intrinsic camera models in context of 2D and 3D data are discussed. A novel nonlinear goal function is proposed for homographic calibration method having the fast convergence of Levenberg-Marquardt optimization procedure. Three models (linear, quadratic, and rational) and four optimization procedures for their identification were compared wrt their time complexity, the projection accuracy, and the intrinsic parameters accuracy. The analysis has been performed for both, the raw and the calibrated pixel data, too. The recommended technique with the best performance in all used quality measures is the Housholder QR decomposition for the linear least square method of the linear form of projection equations.

1 INTRODUCTION

Camera calibration is the fundamental generic problem in computer vision (Y. Ma, 2004). In case of pinhole camera model, the problem usually refers to estimation of camera intrinsic parameters K and to camera pose R and camera location C estimation with respect to a selected coordinates frame. Both kinds of parameters define, modulo constant factor, a projection matrix M which is the algebraic model in homogenous coordinates of the imaging geometry for the given view of 3D or 2D scene:

$$p \equiv MP,\ p = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},\ P = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad (1)$$
$$M = [M_3, m_4],\ M_3 \in \mathbb{R}^{3\times 3},\ m_4 \in \mathbb{R}^3$$

where projective relation $\equiv$ makes equivalent all points collocated in the same projective line. In algebraic notation it means that for any P there exists a scaling factor $\lambda(P)$ for which the equation is true: $p = \lambda(P)MP$.

The matrix $K \in \mathbb{R}^{3\times 3}$ of intrinsic parameters describes the transformation from scene to camera pixel coordinates on the projection plane. Since the choice of coordinate axis for the camera is not unique the K is not unique, too. However, the following decomposition formula always holds:

$$M \equiv KR^{-1}[I_3, -C] \qquad (2)$$

where the pose matrix $R = [r_x, r_y, r_z]$ consists of the camera frame axis defined by unit length vectors with coordinates wrt the scene frame $I_3 = [e_1, e_2, e_3]$, and C is the origin of the camera frame.

The matrix equivalence used in (2) is the equality modulo constant factor: $M_1 \equiv M_2$ if and only there exists $\lambda \neq 0$ such that $M_2 = \lambda M_1$.

Since any rotation in projection plane can be modelled by the matrix R, the intrinsic matrix K is the upper triangular matrix with positive elements on the diagonal. In principle there are two approaches to make the matrix K unique. In the most popular case, the requirement of orthogonality $R^t R = I_3$ makes by QR decomposition of M_3^{-1}, the unique selection of K (O. Faugeras, 2001). We call this case of calibration as *orthogonal calibration* and identify the five free parameters for the inverse matrix:

$$K_o^{-1} = \begin{bmatrix} k_1 & k_2 & k_3 \\ 0 & k_4 & k_5 \\ 0 & 0 & 1 \end{bmatrix} \qquad (3)$$

Tomaszewski M. and Skarbek W. (2007).
ON PROJECTION MATRIX IDENTIFICATION FOR CAMERA CALIBRATION.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, pages 92-97
DOI: 10.5220/0002068000920097

In the less popular case, the requirement of orthogonality is replaced by the zero condition $k_2 = 0$. The case is fully compliant with the physical model of pinhole camera having directly interpreted parameters:

$$K_p = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where for pixel size $s_x \times s_y$ and focal length f we have $f_x = f/s_x$, $f_y = f/s_y$, and (c_x, c_y) are pixel coordinates of the image center, i.e. intersection point of the projection plane by the camera z axis. Usually $c_x \approx x_{res}/2$, $c_y \approx y_{res}/2$. Having the first two columns u_x, u_y of M_3^{-1} we get $f_x = 1/\|u_x\|$, $f_y = 1/\|u_y\|$.

In case of intrinsic calibration by 2D scene views (less expensive and more accurate case) we have to estimate the 2D version of M, i.e. the homographic matrix H which relates planar points in the scene with image pixels:

$$p \equiv HP,\ p = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix},\ P = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{5}$$

The relationship of the homographic projection H with the intrinsic parameters is obtained from (2). However now, the orthogonal calibration can be only resolved from the homographic equation (5). Since then $Z = 0$ and $R^{-1} = R^t$ we have:

$$\begin{aligned} &H = [h_1, h_2, h_3] \equiv K_o R^t [e_1, e_2, -C] \\ &K_o^{-1} h_1 \equiv e_1,\ K_o^{-1} h_2 \equiv e_2 \end{aligned} \tag{6}$$

It implies the following two inherently nonlinear relationships for K_o and its vectorial representation $\vec{k} = [k_1, \ldots, k_5]^t$:

$$\begin{aligned} &h_1^t K_o^{-t} K_o^{-1} h_2 = 0,\ h_1^t K_o^{-t} K_o^{-1} h_1 = h_2^t K_o^{-t} K_o^{-1} h_2 \\ &\vec{k}^t [h_1^{\circ t} h_2^{\circ} + h_2^{\circ t} h_1^{\circ}] \vec{k} + 2 h_1(3) h_2(3) = 0 \\ &\vec{k}^t [h_1^{\circ t} h_1^{\circ} - h_2^{\circ t} h_2^{\circ}] \vec{k} + h_1^2(3) - h_2^2(3) = 0 \end{aligned} \tag{7}$$

where for the vector $h = [h(1), h(2), h(3)]^t$, a *circle operator* assigns the following matrix:

$$h^{\circ} = \begin{bmatrix} h(1) & h(2) & h(3) & 0 & 0 \\ 0 & 0 & 0 & h(2) & h(3) \end{bmatrix} \tag{8}$$

From the above introduction to camera calibration problem we see that the accuracy of the projection matrix M or H determined from 3D or 2D noisy data, is of utmost importance.

In the presented research three models (linear, quadratic, and rational) and four optimization procedures for their identification were compared wrt their time complexity, the projection accuracy, and the intrinsic parameters accuracy. The analysis has been performed for both, the raw and the calibrated pixel data, too.

2 MODELS FOR PROJECTION IDENTIFICATION

Using Kronecker's operation $\otimes$, the generic projection relation (1) can be transformed into the equation:

$$p = \lambda(P) MP = [I_3 \otimes P^t] \begin{bmatrix} \vec{m} \\ 1 \end{bmatrix} \tag{9}$$

where the matrix $M = [m_{ij}]$, with $m_{34} = 1$, has the row-wise vectorial form

$$\vec{m} = [m_{11}, m_{12}, \ldots m_{21}, m_{22}, \ldots, m_{31}, m_{32}, m_{33}]^t$$

It will be convenient to separate the matrix operator $\mathcal{A}(P) = I_3 \otimes P^t$ into three row vector operators $\mathcal{A}_1, \mathcal{A}_2, \mathcal{A}_3$:

$$\mathcal{A}(P) = I_3 \otimes P^t = \begin{bmatrix} \mathcal{A}_1(P) & 0 \\ \mathcal{A}_2(P) & 0 \\ \mathcal{A}_3(P) & 1 \end{bmatrix} \tag{10}$$

It is easy to check that the same separation is true for the homographic matrix $H = [h_{ij}]$ for which the vectorial form $\vec{h}$ has 8 elements:

$$\vec{h} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}]^t$$

2.1 Rational Model

The most close model to the projective equation (9) is the nonlinear model wrt $\vec{m}$ having the form of two rational functions:

$$\begin{aligned} E_x(\vec{m}; P) &= \frac{\mathcal{A}_1(P)\vec{m}}{\mathcal{A}_3(P)\vec{m}+1} - x \\ E_y(\vec{m}; P) &= \frac{\mathcal{A}_2(P)\vec{m}}{\mathcal{A}_3(P)\vec{m}+1} - y \end{aligned} \tag{11}$$

In order to find the projection matrix, for the spatial non-planar points P_i projected onto image pixels p_i, $i = 1, \ldots, n_p$ $(n_p \geq 6)$ we optimize by Levenberg-Marquardt method, the following nonlinear goal function:

$$\mathcal{N}(\vec{m}) = \sum_{i=1}^{n_p} \left[E_x^2(\vec{m}; P_i) + E_y^2(\vec{m}; P_i) \right]$$

The Jacobian matrix J, required by this procedure, has the compact form:

$$J(\vec{m}; P) = \begin{bmatrix} \frac{\mathcal{A}_1(P)}{\mathcal{A}_3(P)\vec{m}+1} - \frac{\mathcal{A}_1(P)\vec{m}\mathcal{A}_3(P)}{(\mathcal{A}_3(P)\vec{m}+1)^2} \\ \frac{\mathcal{A}_2(P)}{\mathcal{A}_3(P)\vec{m}+1} - \frac{\mathcal{A}_2(P)\vec{m}\mathcal{A}_3(P)}{(\mathcal{A}_3(P)\vec{m}+1)^2} \end{bmatrix} \tag{12}$$

It is interesting that the formulas (11), (12) are also valid for the homographic matrix H with $\vec{m}$ replaced by $\vec{h}$ and $P^t = [X, Y, Z, 1]$ replaced by $P^t = [X, Y, 1]$.

2.2 Linear Model

The roots of rational functions (11) are also the solutions of the following linear system of equations:

$$\mathcal{A}(P,p)\vec{m} = \begin{bmatrix} \mathcal{A}_1(P) - x\mathcal{A}_3(P) \\ \mathcal{A}_2(P) - y\mathcal{A}_3(P) \end{bmatrix} \vec{m} = \begin{bmatrix} x \\ y \end{bmatrix} \quad (13)$$

Considering n_p spatial points P_i and their images p_i, $i = 1,\ldots,n_p$ we get $2n_p$ linear equations $A\vec{m} \simeq b$ defined by the matrix A and the right hand side vector:

$$A = \begin{bmatrix} \mathcal{A}(P_1,p_1) \\ \vdots \\ \mathcal{A}(P_{n_p},p_{n_p}) \end{bmatrix},\ b = \begin{bmatrix} b(p_1) \\ \vdots \\ b(p_{n_p}) \end{bmatrix}$$

The same construction is valid for the homographic matrix H.

There are many techniques finding efficiently the minimum solution $\vec{m}^*$ of the linear least square problem $A\vec{m} \simeq b$ for the following goal function:

$$\mathcal{L}(\vec{m}) = \|A\vec{m} - b\|^2 = \sum_{i=1}^{n_p} \|\mathcal{A}(P,p) - b(p)\|2.$$

The most popular method is the pseudo-inverse matrix A^+ method (R. Klette, 1996) which is found using SVD decomposition for $A = U\Sigma V^t$:

$$A^+ = V\Sigma^+ U^t$$

where the diagonal matrix Σ^+ is the pseudo-inverse of the diagonal matrix Σ :

$$\Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_r,0,\ldots,0)$$
$$\Sigma^+ = \mathrm{diag}(1/\sigma_1,\ldots,1/\sigma_r,0,\ldots,0)$$

where r is the rank of A. Then the least square solution is given by the formula

$$\vec{m}^* = A^+ b$$

Let $r = 11$ for 3D case and $r = 8$ for 2D case. The faster algorithm for the case when $r < 2n_p$ is R-SVD algorithm. Using the complexity formulas for SVD from (G. Golub, 1989) we have the following number of flops for the pseudo-inverse method:

$$FLOPS_{PINV}(n_p,r) = 6n_p r(8r/3+1) + 20r^3 \quad (14)$$

Another technique finding the optimal solution is based on triangulation of matrix A by Housholder's symmetries $\mathcal{H}_{a_i}$, $i = 1,\ldots,r$. The process is part of QR decomposition. However we need only the triangular form T_H and concurrently transformed right hand side b_H. Then the least square solution is given by the formula:

$$\vec{m}^* = T_H^{-1} b_H$$

The exact count of flops for HS-QR approach gives the formula:

$$FLOPS_{HS}(n_p,r) = 6n_p r(r+1) + r^2(-r/2+3+5/(2r))$$

The difference of the measures shows the computational advantage of HS-QR approach.

$$FLOPS_{PINV}(n_p,r) - FLOPS_{HS}(n_p,r) = \\ = 10n_p r^2 + 41r^3/2 - 3r^2 - 5r/2$$

For 3D case, $r = 11$ and then

$$FLOPS_{PINV}(n_p,11) - FLOPS_{HS}(n_p,11) > \\ 1210n_p + 25000.$$

While in 2D case, $r = 8$ and the flops difference has the formula:

$$FLOPS_{PINV}(n_p,8) - FLOPS_{HS}(n_p,8) > \\ > 640n_p + 10000.$$

2.3 Quadratic Model

In the literature referring to camera calibration problems the quadratic model is frequently recommended (Y. Ma, 2004). It is obtained from (13) by aggregation of quadratic errors produced by each pair P_i, p_i.

Let $\vec{m'}$ extends $\vec{m}$ by $m_{3,4}$. Then, the error of reproducing p from P is described by the following matrix $\mathcal{B}(P,p)$:

$$\mathcal{B}(P,p)\vec{m'} = \begin{bmatrix} \mathcal{A}_1(P) - x\mathcal{A}_3(P) & -x \\ \mathcal{A}_2(P) - y\mathcal{A}_3(P) & -y \end{bmatrix} \vec{m'} \simeq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The same form of error matrix we obtain for 2D case extending $\vec{h}$ by $h_{3,3}$.

The total squared error leads to the quadratic form defined by the matrix $\mathcal{B}$:

$$\mathcal{L}'(\vec{m'}) = \textstyle\sum_{i=1}^{n_p} \|B(P_i,p_i)\vec{m'}\|^2 = \\ = \textstyle\sum_{i=1}^{n_p} \vec{m'}^t B^t(P_i,p_i)B(P_i,p_i)\vec{m'} = \\ = \vec{m'}^t \left[\textstyle\sum_{i=1}^{n_p} B^t(P_i,p_i)B(P_i,p_i)\right] \vec{m'} = \vec{m'}^t \mathcal{B}\vec{m'}$$

The minimization of $\mathcal{L}'(\vec{m'})$ is obtained from SVD: $\mathcal{B} = U\Sigma V^t$. Namely, from the singular base $U_r = [u_1,\ldots,u_{r+1}]$, we take the vector u_{r+1} and convert to matrix form.

The computational complexity for the quadratic method has the formula:

$$FLOPS_{SVD}(n_p,r) = 4n_p^2(r+1)^2 + 17(r+1)^3.$$

Since the above formula is quadratic in n_p, the computational advantage of HS-QR method over SVD is also quadratic. It follows from computational overhead to get the matrix $\mathcal{B}$. We could avoid this by applying SVD to the global error matrix:

$$\left[\mathcal{B}(P_1,p_n),\ldots,\mathcal{B}(P_{n_p},p_{n_p})\right]^t.$$

However, then we get a procedure equivalent to pseudo-inverse method with complexity linearly inferior in n_p to HS-QR approach.

3 INTRINSIC CALIBRATION

We have investigated this problem for both, the orthogonal and physical calibration cases, detailed in the introduction.

3.1 3D Calibration Scene

In 3D case when pixel data is obtained from images of calibration cube we start from identification of physical model.

The image center is roughly estimated from the camera image resolution and its location is accurately estimated during the lens distortion modelling. It is interesting that the remaining physical parameters f_x, f_y have simple formulas if M_3 is already estimated. Namely, if $M_3^{-1} = [u_x, u_y, u_z]$ then

$$M_3^{-1} = [u_x, u_y, u_z] = RK_p^{-1} = \\ = [r_x, r_y, r_z] \begin{bmatrix} 1/f_x & 0 & -c_x/f_x \\ 0 & 1/f_y & -c_y/f_y \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

Hence

$$f_x = \frac{1}{\|u_x\|},\ f_y = \frac{1}{\|u_y\|},\ r_x = f_x u_x,\ r_y = f_y u_y \\ r_z = \frac{u_z + c_x u_x + c_y u_y}{\|u_z + c_x u_x + c_y u_y\|} \quad (16)$$

Having parameters f_x, f_y, c_x, c_y, the calibration matrix K_p is identified according (4). The matrix K_o can be found from the matrix K_p by closed form formulas. However, we have found that QR procedure applied to M_3 :

$$M_3^{-1} \overset{QR}{=} RK_o^{-1}$$

gives K_o entries more accurate when data is noisy.

3.2 2D Calibration Scene

In 2D case when calibration is performed from homographic matrices obtained on the basis of chessboard images the error function is based on relationships (7) applied to j-th view, $j = 1, \ldots, n_v$:

$$E_1^{(j)}(\vec{k}) = \vec{k}^t A_j \vec{k} + 2h_1^{(j)}(3)h_2^{(j)}(3) \\ E_2^{(j)}(\vec{k}) = \vec{k}^t B_j \vec{k} + (h_1^{(j)})^2(3) - (h_2^{(j)})^2(3) \quad (17)$$

where the symmetric matrices A_j, B_j are defined as follows

$$A_j = (h_1^{(j)})^{\circ t}(h_2^{(j)})^{\circ} + (h_2^{(j)})^{\circ t}(h_1^{(j)})^{\circ} \\ B_j = (h_1^{(j)})^{\circ t}(h_1^{(j)})^{\circ} - (h_2^{(j)})^{\circ t}(h_2^{(j)})^{\circ}$$

In order to find the calibration matrix K_o represented by the vector $\vec{k}$, we optimize by Levenberg-Marquardt method, the following nonlinear goal function:

$$\mathcal{N}_K(\vec{m}) = \sum_{j=1}^{n_v} \left[(E_1^{(j)})^2(\vec{k}) + (E_2^{(j)})^2(\vec{k}) \right]$$

The Jacobian matrix J, required by this procedure, has a simple form for $j = 1, \ldots, n_v$:

$$J^{(j)}(\vec{k}) = 2 \begin{bmatrix} \vec{k}^t A_j \\ \vec{k}^t B_j \end{bmatrix} \quad (18)$$

Having intrinsics in orthogonal model $\vec{k}$ we can get easily the missing parameters for the physical model:

$$f_x = \frac{1}{k_1},\ f_y = \frac{1}{\sqrt{k_2^2 + k_4^2}}$$

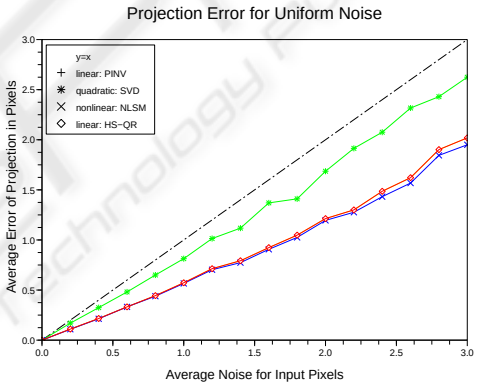

Figure 1: The average projection error for matrix identified from noisy pixels. The noise is uniform in the interval $[-\sigma, +\sigma]$, $\sigma \in [0, 3]$ is measured in pixels.

4 EXPERIMENTS

We have conducted our experiments on both, the real and the synthetic data. Real pixel data has been mainly located manually. In case of optical distortion modelling, where thousands of corners in calibration grid are required, their coordinates were detected automatically by our computer program.

For the identification of the projection matrix M and the homographic matrix H four models were compared: linear by PINV, quadratic by SVD, nonlinear by LM, and linear by HS-QR.

The accuracy of matrix identification was measured directly by Frobenius distance to the ground-truth matrix and indirectly by the average displacement of pixels projected by the identified matrix from

noisy data wrt to pixels projected by the ground truth matrix.

In Figures 1, 2 we present comparative results of projection accuracy for the four analyzed models under uniform and normal input noise and with and without pixel normalization operation. The pixel normalization is guided by the physical intrinsic parameters

$$p' = K_p^{-1} p,\ x' = (x - c_x)/f_x,\ y' = (y - c_y)/f_y.$$

The similar normalization operation is used before the calibration of extrinsic parameters (Hartley, 1997).

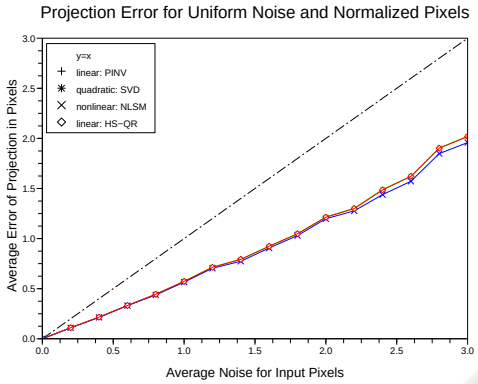

Figure 2: The average projection error for matrix identified from noisy pixels. The noise is uniform in $[-\sigma, +\sigma]$ and $\sigma \in [0,3]$ is measured in pixels. The pixel normalization is used.

We see from Figure 1 that PINV, HS-QR, and ML (initialized by PINV) have comparable accuracy (with slight advantage of nonlinear model) while SVD has significantly higher projection error. When pixel normalization is applied (in practice not always possible!) then all the methods transform input pixel noise into output noise in the same way scaling it down (cf. Figure 2) by a factor of two. The similar behavior has been observed for normal noise.

Figures 3, 4 illustrate the dependence of absolute projection matrix and intrinsic matrix (3D case) errors (per matrix element) on input pixel noise. While the relationship for projection matrix is exactly the same (the graphs were slightly shifted to distinguish them) for independently whether we use in calibration physical or orthogonal model, the accuracy of element identification for the physical intrinsic model is higher than for the orthogonal intrinsic model.

Calibration of intrinsic parameters from homographies wrt the selected computational model is analyzed in Figures 5, 6, and 7.

The advantage of linear over quadratic model is observed on all these graphs.

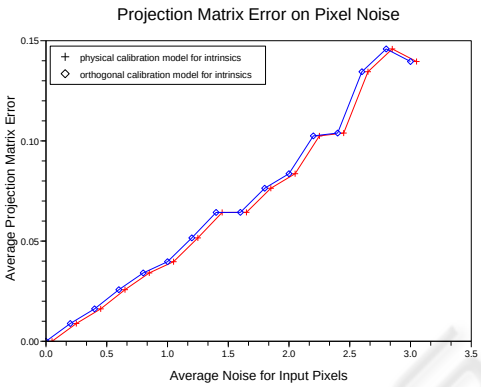

Figure 3: Average projection matrix absolute error identified from noisy pixels.

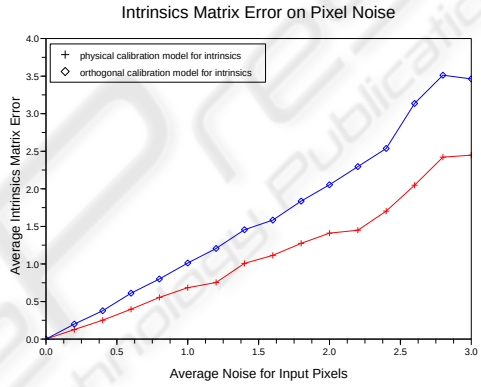

Figure 4: Average intrinsic matrix absolute error identified from noisy pixels.

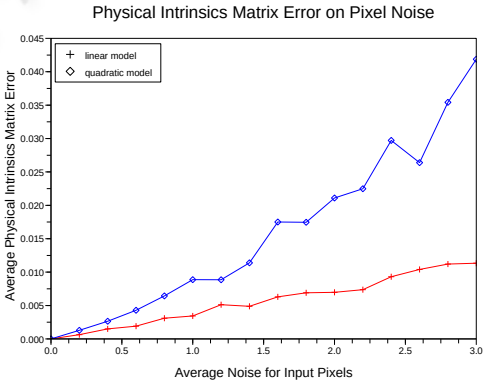

Figure 5: Average physical intrinsics matrix error for pixel noise.

In the screen shot below we have the results of intrinsic calibration for real data extracted from 16 images of chessboard.

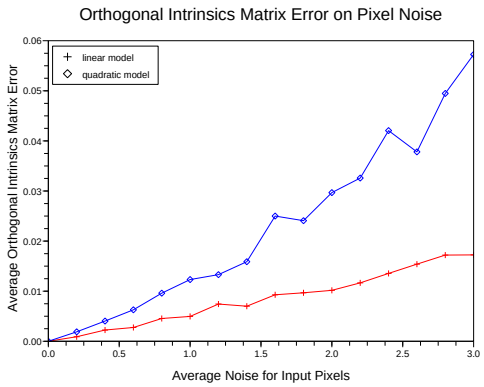

Figure 6: Average orthogonal intrinsics matrix error for pixel noise.

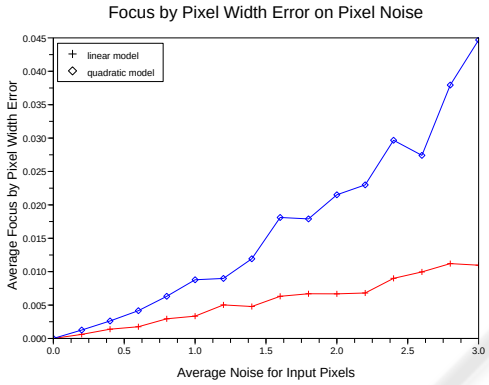

Figure 7: Average focus to pixel width error for pixel noise.

```
-->[Kp,Ko]=realKpKoByHomographies(3073/2,2305/2)

 Ko  =

    2463.0976    3.2421405    1535.4054
    0.           2473.6315    1142.4508
    0.           0.           1.

 Kp  =

    2463.0976    0.           1536.5
    0.           2473.6294    1152.5
    0.           0.           1.
```

5 CONCLUSION

The projection matrix identification problem is considered with application to calibration of intrinsic camera parameters.

Physical and orthogonal intrinsic camera models in context of 2D and 3D data are discussed.

A novel nonlinear goal function is proposed for homographic calibration method having the fast convergence of Levenberg-Marquardt optimization procedure.

Three models (linear, quadratic, and rational) and four optimization procedures for their identification were compared wrt their time complexity, the projection accuracy, and the intrinsic parameters accuracy.

The analysis has been performed for both, the raw and the calibrated pixel data, too.

The recommended technique with the best performance in all used quality measures is the Housholder QR decomposition for the linear least square method of the linear form of projection equations.

REFERENCES

G. Golub, C. L. (1989). *Matrix Computations*. The Johns Hopkins University Press, Baltimore and London, 2nd edition.

Hartley, R. (1997). In defence of the eight-point algorithm. *IEEE Transcations on Pattern Analysis and Machine Intelligence*, 19(6):580–593.

O. Faugeras, Q. L. (2001). *Geometry of Multiple Images*. The MIT Press, Cambridge, Massachusetts.

R. Klette, K. Schluns, A. K. (1996). *Computer Vision - Three-Dimensional Data From Images*. Springer, Singapore.

Y. Ma, S. Soatto, e. a. (2004). *An Invitation to 3-D Vision*. Springer, Cambridge, Massachusetts.