CLOSED FORM MONOCULAR RE-PROJECTIVE POSE ESTIMATION

G. Pisinger

University of Passau
D-94032 Passau, Germany

T. Hanning

University of Passau
D-94032 Passau, Germany

ABSTRACT

In this article we introduce a closed form estimation of the pose determination problem. Unlike most other approaches our method minimizes a Euclidean error to re-projected image points. If we know the distances between these point we can reconstruct the 3D position of the points completely. If the exact distance is unknown the reconstruction is correct up to a scale factor. We compare our approach to several methods to estimate the pose and orientation of a planar pattern observed by a calibrated camera. All compared approaches are closed form solutions and take only one image for the pose estimation. The gain of the proposed method is not only a better starting value for non-linear optimizations but also its applicability for mobile solutions on constrained hardware. Therefore, we compare the error of the estimated pose to the ground truth for the investigated methods.

***Index Terms*—** pose estimation, planar pattern, Euclidean error, re-projection

1. INTRODUCTION

The purpose of this work is to estimate the position of points w. r. t. a camera. We assume that we know that the observed points are equidistant and collinear. The standard approach to this problem depends on the so called pinhole assumption. Distorted images must be undistorted before applying this method, which uses the fact that every observed 3D movement is determined by a homography. The determination of this homography implies often a non-linear optimization itself (see [1]). For time-critical applications such a non-linear optimization can be too time-consuming. In this article we present two re-projective methods that do not need any non-linear optimization. Both methods exploit the simple structure of the given problem. The collinearity of the observed points simplifies the the search for the optimal rotation and allows a pure algebraic approach to the problem. Of course, both methods determine no optimal solution to the (projective) pose determination problem.

2. THE CAMERA MODEL

2.1. The projective camera mapping

In our context the term "camera" consists of a camera (including the lens), a frame-grabbing device and the displayed image. The camera mapping $\mathcal{K} : \mathbb{R}^3 \to \mathbb{R}^2$ defines the way by which an object point $p \in \mathbb{R}^3$ will be displayed in the image. The common way to model the camera mapping $\mathcal{K} : \mathbb{R}^3 \to \mathbb{R}^2$ in computer vision is to do a coordinate transformation T followed by a dimension reducing mapping $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$, resulting in $\mathcal{K} = \Pi \circ T$.

This work was supported by the German Ministry for Research and Education (BMBF) under Grant 16SV2267

The so called *extrinsic parameters* T define the transformation from a given reference coordinate system to the camera coordinate system. Since we are interested in the pose and orientation of a planar prototype w. r. t. to the camera only, in our context it is $T = \mathrm{id}$.

Π is a central projection followed by a coordinate system transformation. For theoretical analysis many authors consider only this pure pinhole camera model. For a realistic camera modeling a distortion mapping in the image plane has to be considered, which leads to a pinhole camera model with distortion (see e. g. [1]).

The first part of Π is the projection of a 3D-point on the camera plane. Let $P_z : \mathbb{R}^3 \setminus \{z = 1\} \to \mathbb{R}^2$ denote the central projection w. r. t. the z-coordinate: $P_z((x, y, z)^t) = (\frac{x}{z}, \frac{y}{z})^t$. The distortion mapping $\delta : \mathbb{R}^2 \to \mathbb{R}^2$ is defined in the camera plane w. r. t. the camera coordinate system. The most common distortion model is the one of radial distortion:

$$\delta_R\left(\begin{pmatrix} u \\ v \end{pmatrix}\right) := \begin{pmatrix} u + u \sum_{i=1}^{D} k_i (u^2 + v^2)^i \\ v + v \sum_{i=1}^{D} k_i (u^2 + v^2)^i \end{pmatrix} \tag{1}$$

with parameters $k_1, \ldots, k_D$ where in most cases it is $D = 2$.

The last step of the projection part Π accomplishes the change of the camera coordinate system to the (computer) image coordinate system (a change of units from metric to pixel). We set

$$P : \mathbb{R}^2 \to \mathbb{R}^2 : \begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} \alpha & \gamma \\ 0 & \beta \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} u_0 \\ v_0 \end{pmatrix},$$

where (u_0, v_0) is the projection of the optical center in pixel coordinates. All parameters defining the mapping Π are called the *intrinsic camera parameters*.

Recapitulating the parameterization of a camera mapping $\mathcal{K}$ splits into the parameterization of P, δ and T with $\mathcal{K} = P \circ \delta \circ P_z \circ T$. With $T = \mathrm{id}$ we are only interested in the intrinsic camera parameters $\Pi = P \circ \delta \circ ZProj$. For a pure pinhole camera it is $\delta = \mathrm{id}$ in this formula. See [2] for more details.

2.2. Re-projection of image points

As mentioned before each point in the image plane determines a straight line in the reference coordinate system intersecting the pinhole of the camera by re-projection. Let $A \subset \mathbb{R}^2$ be the image plane of the camera, then the re-projected ray of a point $i \in A$ is defined by the pre-image of i under $\mathcal{K}$. In our camera model $\mathcal{K}^{-1}(\{i\})$ is a straight line (the so called *viewing ray*).

This means $\Pi^{-1} = (P \circ \delta \circ P_z)^{-1} = P_z^{-1} \circ \delta^{-1} \circ P^{-1}$ is not well defined: The set $P_z^{-1}(\{(u, v)^t\}) = \{s(u, v, 1)^t | s \in \mathbb{R}\}$, for $(u, v) \in \mathbb{R}^2$, is a straight line in $\mathbb{R}^3$ with direction $(u, v, 1)^t$ containing the origin. In order to construct a well defined function

P_z^{-1} we choose a suitable representative of $P_z^{-1}(\{(u,v)\})$ by setting $P_z^{-1}((u,v)^t) := \frac{1}{\sqrt{u^2+v^2+1}}(u,v,1)^t$. Note that this representative has norm 1. It is the direction of the viewing ray.

3. EUCLIDEAN POSE DETERMINATION

For the re-projective pose estimation we are able to formulate a closed form solution of the pose determination problem, which is optimal in the Euclidean sense. This means that the distance of the determined pose of the prototype to the re-projected observed points is minimal.

Let $\mathbf{P} = \{p_1, \dots, p_m\} \subset \mathbb{R}^3$ be a finite set of collinear points w. r. t. the reference coordinate system. $\mathbf{P}$ is called *prototype*. Without loss of generality let $p_j = (x_j, 0, 0) \in \mathbb{R}^3$ for all $p_j \in \mathbf{P}$. For every $p_j \in \mathbf{P}$ we denote $i_{p_j} \in A$ for the observed projection of p_j in the image plane $A \subset \mathbb{R}^2$ w. r. t. the image coordinate system and define $n_j = P_z^{-1}(i_{p_j})$. Furthermore, we assume that not all i_{p_j} are equal.

For a direction $n \in S_2 := \{x \in \mathbb{R}^3 \,|\, \|x\| = 1\}$ we define $L_n := \{\alpha n \,|\, \alpha \in \mathbb{R}\}$ as the *line with direction n containing the origin*. It is easy to see that for every point $p \in \mathbb{R}^3$

$$\operatorname{dist}(p, L_n)^2 = \|(I - N_n)p\|^2 \tag{2}$$

holds, where N_n is the observed line-of-sight projection matrix defined as $N_n = n\, n^t$ and I is the 3×3 identity matrix. The Euclidean pose estimation problem is to obtain $R \in \{U \in \mathbb{R}^{3\times3} \,|\, \det(U) = 1 \wedge UU^t = I\}$ and $t \in \mathbb{R}^3$ minimizing the least-squares sum

$$\sum_{j=1}^{m} \operatorname{dist}(Rp_j + t, L_{n_j})^2.$$

Since $Rp_j + t = x_j \cdot r_1 + t$, where $R = (r_1, r_2, r_3)$, and using (2) we get

$$\begin{aligned}
&\sum_{j=1}^{m} \|(I - N_{n_j})(x_j r_1 + t)\|^2 \\
&= \sum_{j=1}^{m} (x_j r_1 + t)^t (I - N_{n_j})(x_j r_1 + t) \\
&= r_1^t \left(\sum_{j=1}^{m} x_j^2 (I - N_{n_j})\right) r_1 + 2 r_1^t \left(\sum_{j=1}^{m} x_j (I - N_{n_j})\right) t + \\
&\quad t^t \left(\sum_{j=1}^{m} (I - N_{n_j})\right) t
\end{aligned}$$

Using the abbreviations

$$M_1 = \sum_{j=1}^{m} (I - N_{n_j}),\; M_2 = \sum_{j=1}^{m} x_j (I - N_{n_j})$$

and

$$M_3 = \sum_{j=1}^{m} x_j^2 (I - N_{n_j})$$

the Euclidean pose determination problem is to obtain $r_1 \in S_2$ and $t \in \mathbb{R}^3$ minimizing

$$r_1^t M_3 r_1 + 2 r_1^t M_2 t + t^t M_1 t. \tag{3}$$

Since this minimization problem is quadratic in t, given a fixed vector $r_1 \in S_2$, the optimal value for t can be computed in closed form as

$$t = -M_1^{-1} M_2\, r_1. \tag{4}$$

For (4) to be well-defined, M_1 must be positive definite, which can be verified as follows:

For any $x \in \mathbb{R}^3 \setminus \{0\}$, it can be shown that

$$\begin{aligned}
x^t M_1 x &= x^t \left(\sum_{j=1}^{m} (I - N_{n_j})\right) x = \\
&= \sum_{j=1}^{m} (\|x\|^2 - x^t N_{n_j} x) = \sum_{j=1}^{m} (\|x\|^2 - x^t N_{n_j}^t N_{n_j} x) = \\
&= \sum_{j=1}^{m} (\|x\|^2 - \|N_{n_j} x\|^2)
\end{aligned} \tag{5}$$

While $\|x\|^2 - \|N_{n_j} x\|^2$ can be greater than or equal to zero, not all summands can be equal to zero unless all image points i_{p_j} are equal. Since this case is excluded, (5) is strictly greater than zero in every case. Therefore follows the positive definiteness of M_1.

Given the optimal translation as a function of r_1 (3) can be rewritten as

$$\min_{r_1 \in S_2} r_1^t (M_3 - M_2 M_1^{-1} M_2) r_1. \tag{6}$$

Since M_1, M_2 and M_3 are symmetric matrices the matrix $M_3 - M_2 M_1^{-1} M_2$ is also symmetric. Therefore, (6) is an eigenvector problem, where a normalized eigenvector to the smallest eigenvalue of $M_3 - M_2 M_1^{-1} M_2$ is a solution (c.f. [3]).

4. NON EUCLIDEAN POSE ESTIMATION

4.1. Standard pose estimation

A standard technique to obtain a starting value for the non-linear pose calculation problem is to exploit the observed homography. Since the points of the prototype $\mathbf{P}$ are collinear the following approach can be applied:

By abuse of notation, we still use p_j to denotes a point on the prototype, but $p_j = (x_j)$ since the second and third coordinate are always equal to 0. In turn, $\widetilde{p_j} = (x_j, 1)^t$. Let $H \in \mathbb{R}^{3\times2}$ be the matrix describing the movement of the prototype $\mathbf{P}$ to the observed image points, i.e. H minimizes the error function

$$\sum_{j=1}^{m} \|P_z(H\widetilde{p_j}) - i_{p_j}\|^2. \tag{7}$$

Obviously H can be determined only up to a scalar factor. The computation of H requires a non-linear optimization to achieve an appropriate solution. This requires an initial matrix, which can be obtained as follows.

Let $\overline{h_i} \in \mathbb{R}^2$ be the i-th row of H. Instead of minimizing (7) the problem

$$\min_{\|H\|=1} \|H\widetilde{p_j} - (\overline{h_3}\widetilde{p_j})\widetilde{i_{p_j}}\|^2$$

is considered ($\|H\|$ denotes the Frobenius norm of the matrix H, $\|H\| = \sqrt{\sum_{i=1}^{3}\sum_{j=1}^{2} h_{ij}^2}$). This can be rewritten as

$$\min_{x \in \mathbb{R}^6, \|x\|=1} \|Lx\| \tag{8}$$

with the matrix

$$L = \begin{pmatrix} \widetilde{p_1}^t & 0 & -u_1\widetilde{p_1}^t \\ 0 & \widetilde{p_1}^t & -v_1\widetilde{p_1}^t \\ \vdots & \vdots & \vdots \\ \widetilde{p_m}^t & 0 & -u_m\widetilde{p_m}^t \\ 0 & \widetilde{p_m}^t & -v_m\widetilde{p_m}^t \end{pmatrix} \in \mathbb{R}^{2m\times 6}$$

where $i_{p_j} = (u_j, v_j)$. The rows of the matrix H can be found in the vector x.

The solution of (8) is well known to be the eigenvector associated with the smallest eigenvalue of $L^tL \in \mathbb{R}^{6\times 6}$. With this starting value the minimization problem (7) is solved with the Algorithm of Levenberg and Marquardt.

For pinhole cameras it is $\mu H = \tilde{P}(r_1, t)$ for a $\mu \in \mathbb{R}$, where $R = (r_1, r_2, r_3)$ and t determine the transformation of the observed prototype $\mathbf{P}$. For a calibrated camera we know

$$\tilde{P} = \begin{pmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

yielding $(r_1, t) = \mu\tilde{P}^{-1}H =: M$. Since the first column of (r_1, t) must be a unit vector and $t_z \geq 0$ should hold, we set

$$c = \sqrt{M_{1,1}^2 + M_{2,1}^2 + M_{3,1}^2}$$

and

$$\tilde{M} = \begin{cases} \frac{1}{c}M & \text{, if } M_{3,2} > 0 \text{ holds} \\ -\frac{1}{c}M & \text{, otherwise} \end{cases} .$$

The first column of R and t can be obtained by the equality $(r_1, t) = \tilde{M}$. Let r_2 be a unit vector with $r_1^t r_2 = 0$ and $r_3 = r_1 \times r_2$. Then $R = (r_1, r_2, r_3)$ is the rotation matrix searched.

4.2. A re-projective algebraic approach

In [4] we presented an approach based on the theorem of intersecting lines. For each point $p \in L_{n_1}$ we determined two points $q_{1,2}$ on an other viewing ray L_{n_2} with a fixed given distance (see Fig. 1). The two points p, q_1, respectively p, q_2, define a line in space. If we now know that a third point r_1, respectively r_2, is collinear to these two points at a given distance, we can determine the distance of this point to its viewing ray L_{n_3} using this collinearity constraint. If we predict the first point along its viewing correctly this last distance should be zero. This distance can be described by a function of the form $F_{1,2}(x) = ax^2 \pm bx\sqrt{-cx^2 + d^2} + e$ by the position of the first point (along its re-projected viewing ray). The result is a root or a minimum of the distance function. In [4] we observed that the determination of the minimum leads to more stable algorithm.

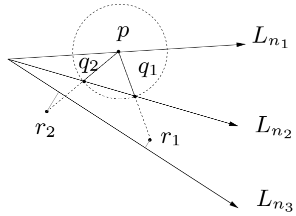

Fig. 1. Construction of p, $r_{1,2}$ and $q_{1,2}$. The distances dist(r_1, L_{n_3}) and dist(r_2, L_{n_3}) should be minimized.

Of course, this approach minimizes an algebraic error which may lead to a result which may not be optimal with respect to the Euclidean distance.

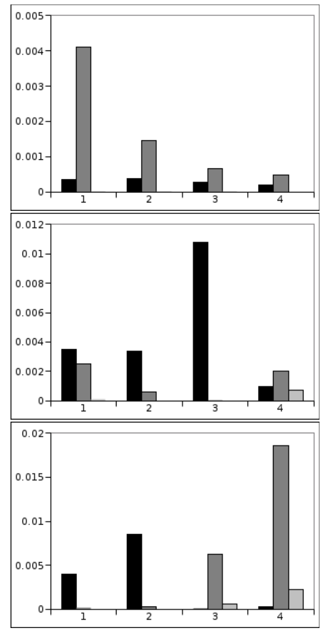

Fig. 2. Reconstruction errors for the 6 mm setup for 12 different positions, where the prototype plate is nearly parallel to the camera plane (top). The middle and bottom diagram are associated with a little and big skewness between prototype plate and camera plane. Left columns: standard pose estimation, middle columns: re-projective algebraic approach, right columns: proposed method

5. EXPERIMENTAL RESULTS

For our experiments we chose a standard CCD camera with a 1/3" chip with low cost 6 mm, 8 mm and 12 mm lenses. To calibrate our cameras we use the calibration algorithm described by Zhang ([1]) including radial distortion parameters k_1, k_2 determining the radial distortion $\delta(u, v) := (u + u(k_1r^2 + k_2r^4), v + v(k_1r^2 + k_2r^4))$ with $r^2 := u^2 + v^2$. We use a 9×5 grid of equidistant (5 cm) tiny points, which define our prototype $\mathbf{P}$. The points were extracted by an algorithm described in [5], which bases on an approximation by polynomials of total degree 2 in a 3×3 pixel neighborhood.

We reconstruct an observed grid by determining the pose of equidistant collinear points on two perpendicular lines on our prototype. A third direction can be obtained as the cross product of the directions of the first two lines. So we are able to estimate the whole transformation R, t of the prototype's coordinate system to the reconstructed coordinate system. The ground truth is defined by the transformation minimizing $(R, t) \mapsto \sum_{p\in\mathbf{P}} \|i_p - \Pi(Rp + t)\|^2$ which we determine using the non-linear optimization method of Levenberg and Marquardt. In our coordinate system the x and y axes are the axes of the camera plane. The z-axis (depth) is perpendicular to them. For each setup we used twelve different positions of the prototype $\mathbf{P}$. To compare the closed form solution of the standard pose estimation, the re-projective algebraic approach and the new method we use 3 collinear points $\{p_1, p_2, p_3\}$ from the prototype. Using this 3 points rotation and translation are computed applying the three methods. Let R_s, t_s, R_a, t_a and R_n, t_n be the solution of the starting value of the standard pose estimation, the algebraic approach and the proposed technique, respectively. The reconstruction error is defined as $\frac{1}{3}\sum_{i=1}^{3}\|(Rp_i + t) - (R_x p_i + t_x)\|^2$ for $x \in \{s, a, n\}$.

Fig. 2 - 4 shows the reconstruction error for the standard technique, the algebraic approach and the proposed method for three different lens setups (6 mm, 8 mm, 12 mm). For every setup we obtain the parameters in the matrix $\tilde{P}$ and the radial distortion parameters by the calibration algorithm due to Zhang [1]. The estimated radial distortion parameters are: $k_1 = -0.2046$, $k_2 = 0.1673$ for the 6 mm setup, $k_1 = -0.19612$, $k_2 = 0.17438$ for the 8 mm setup and $k_1 = -0.0703$, $k_2 = -0.0549$ for the 12 mm setup.

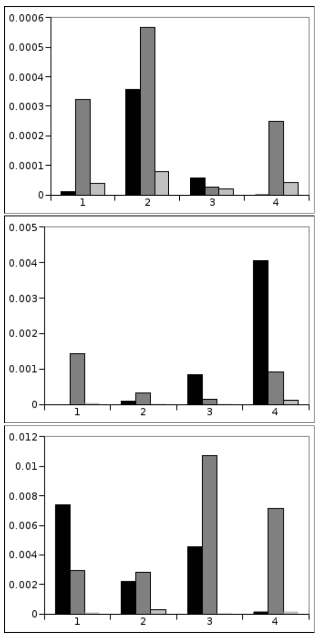

Fig. 3. Reconstruction errors for the 8 mm setup for 12 different positions, where the prototype plate is nearly parallel to the camera plane (top). The middle and bottom diagram are associated with a little and big skewness between prototype plate and camera plane. Left columns: standard pose estimation, middle columns: re-projective algebraic approach, right columns: proposed method

One can see that for almost all cases the proposed method leads to a rotation matrix and a translation vector such that the transformed points $\{p_1, p_2, p_3\}$ are closer to the ground truth points than the projective reconstruction and the algebraic approach does. The values show that the reconstruction error is increasing if the angle between the z-axis and the camera plane is decreasing.

6. CONCLUSION

In this article we presented a closed form solution for the pose estimation problem for collinear points with a known distance to each other. An advantage of the re-projective closed form solutions for 3D-pose estimation is that it's re-projective nature includes the distortion function of the camera mapping. But in contrast to prior publications the proposed method minimizes the Euclidean distance to the re-projected observations. So, in contrast to the algebraic approach every point is weighted uniformly. Errors in the extraction of

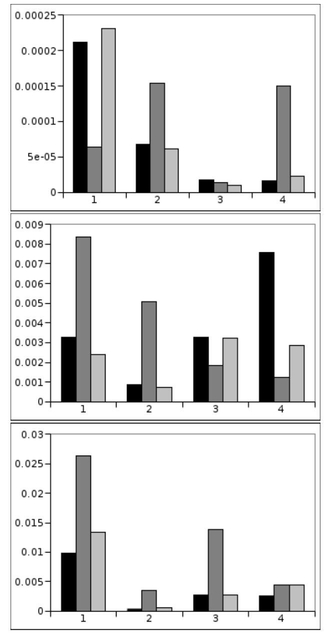

Fig. 4. Reconstruction errors for the 12 mm setup for 12 different positions, where the prototype plate is nearly parallel to the camera plane (top). The middle and bottom diagram are associated with a little and big skewness between prototype plate and camera plane. Left columns: standard pose estimation, middle columns: re-projective algebraic approach, right columns: proposed method

the observation point in the actual image will be smoothed out. The simplicity of the proposed solution w. r. t. the Euclidean error to the re-projections - i. e. the solution of a simple eigenvalue problem of a 3×3-matrix. Therefore, this solution is particularly suitable for simple hardware, which is used in autonomous solutions.

It should be mentioned that all analyzed camera setups were calibrated w. r. t. the projective error. A re-projective calibration is not common in computer vision, but can also be advantageous in some situations (see [2]). For such calibrated cameras an even better performance of the proposed solution can be expected.

7. REFERENCES

[1] Z. Zhang, "A Flexible new technique for camera calibration," Tech. Rep., Microsoft Research, 1998, Technical Report MSR-TR-98-71.

[2] T. Hanning, S. Graf, and M. Kellner, "Re-projective vs. projective camera calibration: Effects on 3d-reconstruction," in *International Conference of Image Processing (ICIP)*, Genova, Italy, 2005, vol. II, pp. 1170 – 1173.

[3] G. H. Golub and C. F. Loan, *Matrix Computations*, John Hopkins UP, 1989.

[4] T. Hanning, R. Schöne, and S. Graf, "A closed form solution for monocular re-projective 3d pose estimation of regular planar patterns," in *International Conference of Image Processing (ICIP)*, Atlanta, Georgia, 2006, pp. 2197 – 2200.

[5] Jürgen Haas, *Echtzeit-Korrespondenzprobleme in Bildsequenzen und Subpixelgenauigkeit*, Ph.D. thesis, Universität Passau, 2000.