Camera Pose Estimation with Unknown Principal Point

Viktor Larsson
Lund University
Lund, Sweden
viktorl@maths.lth.se

Zuzana Kukelova
FEE, Czech Technical University
Prague, Czech Republic
kukelzuz@fel.cvut.cz

Yinqiang Zheng
National Institute of Informatics
Tokyo, Japan
yqzheng@nii.ac.jp

Abstract

To estimate the 6-DoF extrinsic pose of a pinhole camera with partially unknown intrinsic parameters is a critical sub-problem in structure-from-motion and camera localization. In most of existing camera pose estimation solvers, the principal point is assumed to be in the image center. Unfortunately, this assumption is not always true, especially for asymmetrically cropped images. In this paper, we develop the first exactly minimal solver for the case of unknown principal point and focal length by using four and a half point correspondences (P4.5Pfuv). We also present an extremely fast solver for the case of unknown aspect ratio (P5Pfuva). The new solvers outperform the previous state-of-the-art in terms of stability and speed. Finally, we explore the extremely challenging case of both unknown principal point and radial distortion, and develop the first practical non-minimal solver by using seven point correspondences (P7Pfruv). Experimental results on both simulated data and real Internet images demonstrate the usefulness of our new solvers.

1. Introduction

To determine the 6-DoF rotation and translation parameters of a pinhole camera is proven indispensable for many applications ranging from structure-from-motion, to augmented reality and image based geo-localization. Given fully known intrinsic parameters, the perspective-three-point (P3P) problem arises, for which many solvers have been developed [6, 10, 8, 14]. In terms of partially calibrated intrinsic parameters, the majority of existing work assume that the principal point lies in the image center. Unfortunately, as noticed by Triggs [21], this assumption is not always true, especially in the case of cropped images. When the offset is trivial, it can be partially compensated by the translation, without severely affecting the rotation. However, in the presence of significant offset such solvers with centered principal point will give poor rotation and translation estimates.

In this paper we present the first exactly minimal solver for the case of unknown principal point and focal length, using four and a half point correspondences (P4.5Pfuv). In addition to the classical constraints on the camera matrix [5, 12], we derive novel polynomial constraints which allow us to avoid solutions with rank deficient camera matrices.

We also consider unknown aspect ratio and construct a minimal solver which uses five point correspondences (P5Pfuva). In this case the equations reduce to a simple quartic polynomial which allows for a closed form solver that is both extremely fast and stable. In experiments we show that the new solvers are superior in terms of stability and efficiency compared to the previous state-of-the-art five point solver from Triggs [21].

Later, we consider case of both unknown principal point and radial distortion. This problem is very difficult and highly non-linear due the radial distortion being centered on the unknown principal point. We develop the first practical non-minimal solver by using seven point correspondences (P7Pfruv) instead of the minimum five.

Our major contributions are:

i. To derive new polynomial constraints on the camera matrix for the case of unit aspect ratio and zero skew.

ii. To develop the first minimal P4.5Pfuv solver (unit aspect ratio and zero skew) as well as an extremely fast P5Pfuva solver (zero skew).

iii. To explore the extremely challenging case of unknown principal point and radial distortion, and develop the first practical non-minimal solver using seven point correspondences.

2. Background and Related Work

Given 2D-to-3D point correspondences, the camera pose estimation problem aims to estimate the rotation matrix R and the translation vector t, and possibly all or a subset of the intrinsic parameters

$$K = \begin{bmatrix} \alpha f & s & u \\ 0 & f & v \\ 0 & 0 & 1 \end{bmatrix}, \qquad (1)$$

where f denotes the focal length, (u, v) the principal point, α the aspect ratio, and s the skew.

When all the intrinsic parameters are unknown, the direct linear transform (DLT) [11] algorithm applies, which uses at least five and a half point correspondences. In practice, considering that a part of the intrinsic parameters are usually known a priori, the DLT algorithm suffers from over-parametrization, and usually gives inaccurate estimate because of overfitting in the presence of noisy data.

When all intrinsic parameters except the focal length are known, several nearly minimal solvers were developed by using four point correspondences (P4Pf) [2, 24, 17]. These solvers employed the distance related constraints in the 3D camera framework, and ignored one constraint in the process of solving polynomial systems. Later, Wu [23] proposed the first exactly minimal solver by using three and a half point correspondences (P3.5Pf) on the basis of a novel (but occasionally degenerate) parametrization of the camera projection matrix. Recently, Larsson *et al.* [18] proposed a degenerate-free minimal solver from three and a half point correspondences.

Since consumer photography is now dominated by mobile-phone and wide-angle action cameras (*e.g.* GoPro-type cameras), images with significant radial lens distortion are quite common. Due to its compactness and expressive power, the single-parameter division model [7] is widely used to model radial lens distortion. In the division model

$$\begin{bmatrix} x, & y, & 1 + k(x^2 + y^2) \end{bmatrix}^\top \simeq PX, \qquad (2)$$

where (x, y) are the centered image coordinates of a projected 3D point X with a camera projection matrix P, and k is the distortion parameter. Some researchers have studied camera pose estimation in the presence of unknown radial distortion. For example, Josephson and Byröd [13] simultaneously estimated the focal length and the distortion in the absolute pose estimation setup by using four point correspondences (P4Pfr). To reduce the size of the solver, Bujnak *et al.* [3] split this P4Pfr problem into the planar and non-planar case. Later, Kukelova *et al.* [15] used five point correspondences to accelerate computations. Recently, Larsson *et al.* [18] was able to significantly reduce the elimination template size for the minimal P4Pfr problem.

All of the solvers above assume that the principal point is given or centered in the image center. Triggs [21] was the first to consider estimation of the camera pose with unknown focal length and principal point. A non-minimal 5-point solver was presented, although the minimal case is with four and a half points only. As noted in [21], this 5-point solver is very sensitive to noise due to the non-minimal parametrization.

In the following, we augment the classical camera matrix constraints and develop the exactly minimal solvers for P4.5Pfuv and P5Pfuva. In addition, we explore the extreme non-linearity in the scenario of simultaneous unknown principal point and unknown radial distortion, and develop the first practical non-minimal P7Pfruv solver.

3. Unit Aspect Ratio and Zero Skew

In this section, we focus on the absolute pose problem without radial distortion. We assume that the only constraints on the intrinsic parameters are zero skew ($s = 0$) and unit aspect ratio ($\alpha = 1$). These are natural constraints which are satisfied by most consumer cameras with a modern CCD/CMOS sensor.

The constraints for a camera matrix P to admit the following factorization

$$P = K \begin{bmatrix} R & t \end{bmatrix}, \quad K = \begin{bmatrix} f & & u \\ & f & v \\ & & 1 \end{bmatrix} \qquad (3)$$

are well known and summarized in the following theorem.

Theorem 1 (Faugeras [5], Heyden [12]) *The matrix*

$$P = \begin{bmatrix} \boldsymbol{p}_1^T & p_{14} \\ \boldsymbol{p}_2^T & p_{24} \\ \boldsymbol{p}_3^T & p_{34} \end{bmatrix} \qquad (4)$$

corresponds to a perspective camera with zero skew and unit aspect ratio if and only if

$$\det \begin{bmatrix} \boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3 \end{bmatrix} \neq 0 \qquad (5)$$

and

$$(\boldsymbol{p}_1 \times \boldsymbol{p}_3) \cdot (\boldsymbol{p}_2 \times \boldsymbol{p}_3) = 0 \qquad (6)$$
$$\|\boldsymbol{p}_1 \times \boldsymbol{p}_3\|^2 - \|\boldsymbol{p}_2 \times \boldsymbol{p}_3\|^2 = 0 \qquad (7)$$

If only (6) *holds the camera has non-unit aspect ratio.*

Although formulated differently, the constraints (6) and (7) are equivalent to the ones used to create the solver from Triggs [21].

3.1. New Camera Matrix Constraints

The non-zero determinant constraint in (5) is difficult to incorporate in polynomial solvers. Ignoring this constraint adds false solutions corresponding to rank deficient camera matrices. In this section we use tools from algebraic geometry (see *e.g.* [4]) to find additional polynomial constraints which ensure that we only recover the true camera matrices.

Let I be the ideal generated by the original constraints,

$$I = \left\langle (\mathbf{p}_1 \times \mathbf{p}_3) \cdot (\mathbf{p}_2 \times \mathbf{p}_3), \|\mathbf{p}_1 \times \mathbf{p}_3\|^2 - \|\mathbf{p}_2 \times \mathbf{p}_3\|^2 \right\rangle. \qquad (8)$$

Using Macaulay2 [9] we find that this ideal is of dimension 7^1 and degree 16. This means that if we add 7 linear constraints we will in general have 16 solutions. Now some of these solutions might correspond to rank deficient camera matrices. To remove these solutions we compute the saturation of I w.r.t to the determinant, *i.e.*

$$J = \left\{ f(\mathbf{x}) \,\middle|\, \exists N \geq 0, \ \det([\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3])^N f(\mathbf{x}) \in I \right\}. \tag{9}$$

The saturated ideal J contains additional polynomial constraints which should be satisfied by the correct camera matrices. We find that this ideal is also of dimension 7 but only of degree 10. This means that in general there are 6 false solutions corresponding to rank deficient camera matrices if you only use the constraints in (6) and (7).

The ideal J is generated by the two constraints from (6) and (7), as well as 5 polynomials of degree 5. To save space we only present one of these five constraints here,

$p_{11}p_{12}p_{32}^2p_{33}+p_{11}p_{12}p_{33}^3-p_{11}p_{13}p_{32}^3-p_{11}p_{13}p_{32}p_{33}^2-p_{12}^2p_{31}p_{32}p_{33}\ +\ p_{12}p_{13}p_{31}p_{32}^2\ -\ p_{12}p_{13}p_{31}p_{33}^2\ +\ p_{13}^2p_{31}p_{32}p_{33}+p_{21}p_{22}p_{32}^2p_{33}+p_{21}p_{22}p_{33}^3-p_{21}p_{23}p_{32}^3-p_{21}p_{23}p_{32}p_{33}^2\ -\ p_{22}^2p_{31}p_{32}p_{33}\ +\ p_{22}p_{23}p_{31}p_{32}^2\ -\ p_{22}p_{23}p_{31}p_{33}^2+p_{23}^2p_{31}p_{32}p_{33} = 0.$

The remaining constraints and Macaulay2 code for generating them can be found in the supplementary material.

3.2. Building a Polynomial Solver - P4.5Pfuv

Cameras with unit aspect ratio and zero skew have 9 degrees of freedom (3 intrinsic and 6 extrinsic), making the pose estimation problem minimal with 4.5 points. Computing the three dimensional nullspace to the projection equations allow us to parametrize the camera matrix with three unknowns

$$P = \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3. \tag{10}$$

The scale can be fixed by setting $\alpha_3 = 1$. Inserting the first 3×3 block of (10) into the constraints from Section 3.1 we get 2 equations of degrees 4 and 5 equations of degree 5 in the two unknowns α_1 and α_2. Using the automatic generator from Larsson *et al.* [19] we constructed a polynomial solver with template size 11×21.

If we only use the two original constraints, (6) and (7), the automatic generator from Larsson *et al.* [19] returns a polynomial solver with template size 20×36 and if we employ the automatic saturation technique from [20] to saturate the determinant we get a template of size 34×50.

3.3. Unknown Aspect Ratio - P5Pfuva

In the case of unknown aspect ratio and zero skew we only have a single constraint (6) on the camera matrix. For

---
^{1}Note that here we only consider the first 3×3 block of the camera matrix. For the full camera matrix the dimension would be 10.

this ideal no additional constraints are yielded when we saturate the determinant. Cameras with zero skew have 10 degrees of freedom (4 intrinsic and 6 extrinsic) and the pose estimation problem becomes minimal with 5 point correspondences. The linear constraints from 5 points have a 2 dimensional nullspace, allowing us to parameterize the camera using only a single unknown,

$$P = \alpha_1 P_1 + P_2. \tag{11}$$

Inserting into the constraint (6) yields a single quartic equation in α_1 that can be efficiently solved.

3.4. Implementation Details

For the 4.5 point solver from Section 3.2 we can get 9 linear constraints from 5 point correspondences by ignoring a coordinates for one of the image points. The ignored coordinate can then be used to filter solutions. Another approach is to consider all 5 points (10 linear constraints) and compute an approximate 3 dimensional nullspace. In experiments we found that this approach is less sensitive to noise, however the runtime is slightly longer due to the need for computing an SVD to find the approximate nullspace.

For the zero-skew 5 point solver from Section 3.3 we need to find the roots to a quartic polynomial. This can be done by either computing the eigenvalues of the companion matrix, or using the closed form solution for the quartic. In experiments we found that these have similar accuracy if some care is taken to avoid cancellation errors when implementing the closed form solver.

Implemented in C++ the runtimes on a standard desktop computer are ≈ 120 μs (P4.5Pfuv) and 5 μs (P5Pfuva).

4. Radial Distortion with Unknown Center

Radial distortion adds an extra non-linearity to the projections which makes pose estimation more difficult. The problem is further complicated if the center of distortion (typically the principal point) is unknown. In this case the projection equations can be written as

$$\begin{bmatrix} x - u \\ y - v \\ 1 + kd(u,v) \end{bmatrix} \simeq \mathrm{diag}(f, f, 1) \begin{bmatrix} R & t \end{bmatrix} \begin{pmatrix} X \\ 1 \end{pmatrix} \tag{12}$$

where $d(u,v) = (x-u)^2 + (x-v)^2$. The problem contains 10 degrees of freedom and thus becomes minimal with 5 points. However the minimal problem is extremely difficult and we have not found any tractable formulation.

4.1. Seven Point Relaxation - P7Pfruv

To tackle this problem we instead consider a non-minimal relaxation using 7 points. The idea is to consider only the first two equations of (12),

$$\begin{bmatrix} x - u \\ y - v \end{bmatrix} \simeq f \begin{bmatrix} R_1 X + t_1 \\ R_2 X + t_2 \end{bmatrix} \tag{13}$$

These equations constrain the projections to lie on the lines passing through the distortion center (u, v) and the image point (x, y). This constraint is independent of the focal length and the radial distortion parameter, since they just move the projections along these lines.

In (13) we can of course ignore the focal length since it is non-zero for any interesting solution. Using 2×2 determinants we can rewrite the equations as

$$(x - u)(R_2 X + t_2) - (y - v)(R_1 X + t_1) = 0 \quad (14)$$

This relaxed problem has 7 degrees of freedom, and since each point yields a single constraint the problem becomes minimal with 7 points. Of course solving (14) only gives the orientation R, distortion center (u, v) and the first two components of the translation t_1 and t_2. The remaining unknowns in (12) can be solved for linearly, see Section 4.4.

This relaxation is similar to the one made in [15] where they solved the P4Pfr problem using five points instead of the minimal four. In [22] Tsai used a similar approach but did not enforce the constraints on the rotation and simply solved for the unknowns linearly using more points. However, in both these works the principal point is assumed to be known.

4.2. Simplifying the Equations

To solve the equations in (13) we start by doing a change of coordinate systems such that

$$x_1 = y_1 = 0, \quad X_1 = (0, 0, 0). \quad (15)$$

The equations from the first point then reduce to

$$\begin{bmatrix} -u \\ -v \end{bmatrix} \simeq \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \quad (16)$$

If we let R be a scaled rotation, we can instead fix the scale of the camera matrix by setting $t_1 = u$ and $t_2 = v$. This eliminates two unknowns and has the additional benefit that when we insert this into the equations in (14), the mixed quadratic terms in (u, v, t_1, t_2) cancel and we are left with equations which only depend linearly on (u, v).

Using the hidden variable trick [4] we can eliminate the distortion center (u, v) from our equations. Rewrite (14) as

$$M(R) \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = 0 \quad (17)$$

where $M(R)$ is a 6×3 matrix depending on the rotation R. Requiring all 3×3 determinants of this matrix to vanish, we get 20 equations of degree 3 in the elements of R.

Finally using quaternions we parameterize the scaled rotation matrix, i.e. $R(\mathbf{q}) =$

$$\begin{bmatrix} q_1^2 + q_2^2 - q_3^2 - q_4^2 & 2q_2q_3 - 2q_1q_4 & 2q_1q_3 + 2q_2q_4 \\ 2q_1q_4 + 2q_2q_3 & q_1^2 - q_2^2 + q_3^2 - q_4^2 & 2q_3q_4 - 2q_1q_2 \\ 2q_2q_4 - 2q_1q_3 & 2q_1q_2 + 2q_3q_4 & q_1^2 - q_2^2 - q_3^2 + q_4^2 \end{bmatrix}$$

This yields 20 equations of degree 6 in $\mathbf{q} = (q_1, q_2, q_3, q_4)$.

Studying this equation system in Macaulay2 [9] we find that it has 88 solutions. However 16 of these are false solution introduced in the hidden variable trick (17).

4.3. Removing Symmetries

Using the quaternion parametrization we introduce a 2-fold symmetry into our problem since $R(\mathbf{q}) = R(-\mathbf{q})$. For this problem there is another symmetry corresponding to changing the sign of the first two rows of the rotation matrix. Note that this symmetry is also present in the original (non-relaxed) problem where the focal length also changes sign.

This type of symmetry also occurred in the WPnP problem from Larsson *et al.* [17] and in the P3.5Pf formulation of [23]. In [17] they handled the symmetry by doing the following linear change of variables in the quaternion,

$$\mathbf{q} = \begin{bmatrix} 0 & -i & -i & 0 \\ -1 & 0 & 0 & 1 \\ i & 0 & 0 & i \\ 0 & -1 & 1 & 0 \end{bmatrix} \hat{\mathbf{q}}, \quad (18)$$

which reduces the symmetry into two sign symmetries in $(\hat{q}_1, \hat{q}_2)$ and $(\hat{q}_3, \hat{q}_4)$. Removing the symmetries collapses the 88 solutions into 22.

Using the automatic generator from [19] (which automatically handles these sign symmetries) we were able to construct a polynomial solver with template size 124×162.

4.4. Recovering the Full Solutions

The solver we created only returns the rotation R. To recover the remaining parameters we first solve linearly for u, v, t_1 and t_2 from (14). Since these are part of the relaxed problem we have an exact solution to this system.

To recover the remaining parameters f, k and t_3 we rewrite the equations in (12) as

$$(R_3 X + t_3) \begin{bmatrix} x - u \\ y - v \end{bmatrix} = f(1 + kd(u, v)) \begin{bmatrix} R_1 X + t_1 \\ R_2 X + t_2 \end{bmatrix} \quad (19)$$

where we can solve linearly for f, fk and t_3. Note in general there is no exact solution satisfying all $7 \times 2 = 14$ equations since we did not solve the true minimal problem. So instead we solve the linear equations in a least squares sense.

Since this does not minimize any meaningful geometric error we can refine the solutions by performing a few iterations of local optimization. Note that this can be done very quickly since we have very few unknowns and residuals. Since the division model's inverse transform is quite messy, we minimize

$$\sum_{i=1}^{7} \left\| \begin{bmatrix} x_i - u \\ y_i - v \end{bmatrix} - \frac{f(1 + kd_i(u, v))}{(R_3 X_i + t_3)} \begin{bmatrix} R_1 X_i + t_1 \\ R_2 X_i + t_2 \end{bmatrix} \right\|^2 \quad (20)$$

instead of the true reprojection error. Empirically we have seen that this approximation works well.

5. Experiments

We experimentally evaluate our new solvers on both synthetic and real image data. For P4.5Pfuv we compare both using the exact nullspace from 4.5 points (computed using QR), as well as the approximate nullspace from 5 points (computed using SVD). For P5Pfuva we compare solving the quartic equation both using the closed form solution and by computing the eigenvalues to the companion matrix. For solvers returning multiple focal length estimates (*i.e.* non-unit aspect ratio) we compute the focal length error using the geometric mean $f = \sqrt{f_1 f_2}$. For the P7Pfruv solver the results are without the non-linear refinement proposed in Section 4.4 unless otherwise noted.

For the solver from Triggs [21] we added some normalization of the coordinate systems (scaling and shifting) since it was not available in the original implementation available from the author. Experimentally we have observed that this improves the performance drastically in the presence of noise.

To save space we only show the errors in the focal length for some experiments since the other errors are qualitatively similar. More results are available in the supplementary material.

5.1. Stability

In this section we evaluate the numerical stability of the proposed polynomial solvers. We generated random but feasible noise-free synthetic problem instances. To generate the scene we uniformly sample five 3D-points from the box $[-2, 2] \times [-2, 2] \times [2, 8]$ in the camera's local coordinate system. These are then transformed with a random rotation and translation. The focal length was drawn uniformly from the interval $[200, 2000]$ and the principal point was placed randomly 500 px from the origin.

Figure 1 shows the distribution of the $\log_{10}$ relative focal length errors for 10,000 instances. We can see that all solvers are quite stable.

We ran a similar experiment but where we added radial distortion to the image points. The distortion parameter was drawn uniformly from the interval $[-0.4, 0]$. The results for the P7Pfruv solver with and without the non-linear refinement is shown in Figure 2. On a small number of instances the P7Pfruv solver had issues with numerical stability. However, these solutions were still a good enough starting guess for the non-linear refinement (Section 4.4).

5.2. Varying Noise

Next we evaluate the sensitivity to image noise. We use a similar setup as in Section 5.1 but where we fixed the focal length $f_{gt} = 1000$ and added Gaussian noise with varying standard deviation to the image points. Figure 3 shows the median relative focal length error against the noise level and Figure 4 shows the error distribution for $\sigma = 2$ px.
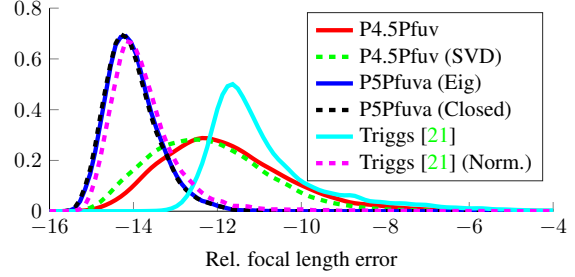

Figure 1. Relative focal length error $\frac{|f - f_{gt}|}{f_{gt}}$ for 10,000 random synthetic instances.
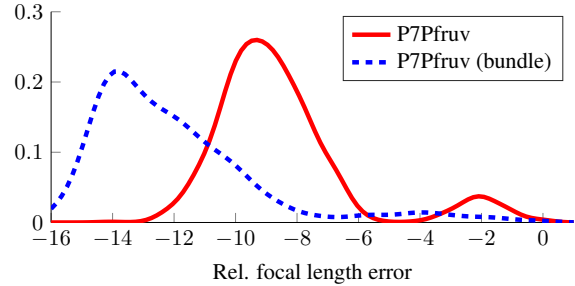

Figure 2. Relative focal length error $\frac{|f - f_{gt}|}{f_{gt}}$ for 10,000 random synthetic instances for image points with added radial distortion.

Our new solvers, especially P4.5Pfuv (SVD), performs the best in the presence of image noise. For the solver from Triggs [21], the accuracy degrades heavily for noisy image points. Additionally we can see that normalization of the image and 3D coordinate systems is essential.

Again we ran a similar experiment but with radial distortion added to the image points. The distortion parameter was fixed at $k_{gt} = -0.2$. The median relative focal length errors are shown in Figure 5.
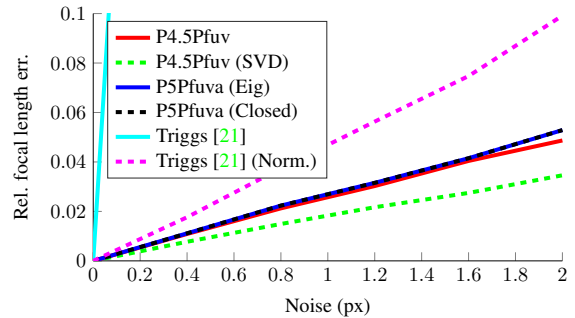

Figure 3. Median relative focal length error for varying noise.

5.2.1 Comparison to 6p DLT

We also did a comparison with standard 6 point DLT [11]. To get a fair comparison we use 6 points to compute the approximate nullspace in our P4.5Pfuv (SVD) solver. In
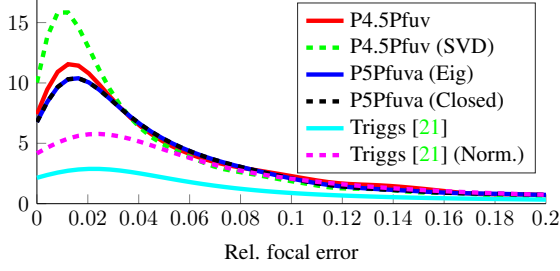
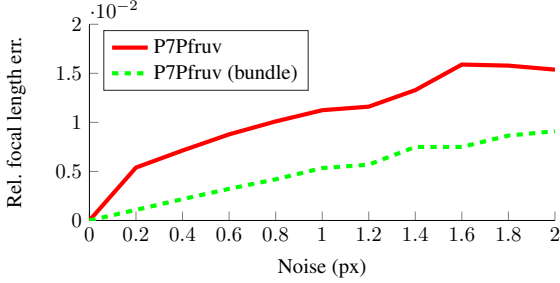Figure 4. Distribution of relative focal length error for 2 px noise.

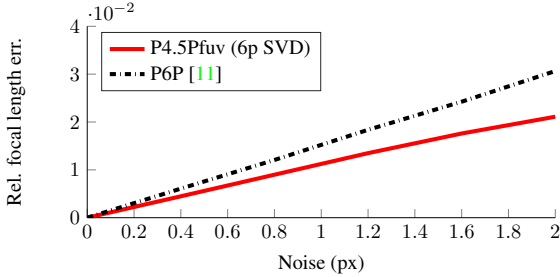Figure 5. Median relative focal length error for varying noise with radial distortion.

Figure 6. Comparison with 6 point DLT. The graph shows the median relative focal length error for varying noise levels. Note that we use 6 points in our P4.5Pfuv solver as well.

Figure 6 we can see that we are able to get better results by enforcing the correct constraints on the intrinsic parameters.

5.3. Varying Principal Point

In this section we study the effects of varying principal point. We compare our new solvers to the state-of-the-art P3.5Pf and P4Pfr solvers from Larsson *et al*. [18], which assume that the principal point is in the center of the image. We generate synthetic scenes where the distance from the origin to the principal point is varied. The ground truth focal length was $f_{gt} = 1000$ and we added small Gaussian noise ($\sigma = 0.1$ px) to the image coordinates. For the distortion solvers we also add radial distortion to the image points with $k_{gt} = -0.2$. The median errors in the focal length and rotation are shown in Figure 7. As expected ignoring the principal point makes the pose estimation significantly worse.
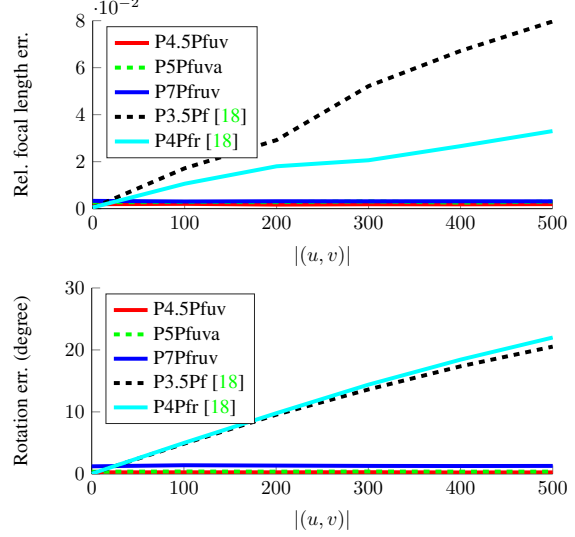

Figure 7. Varying principal point. *Top:* Median relative focal length error $\frac{|f - f_{gt}|}{f_{gt}}$. *Bottom:* Median rotation error in degrees.

5.4. Varying Radial Distortion

In this section we compare the performance of the new solvers when we add varying degrees of radial distortion to the image points. We generated synthetic scenes similarly to Section 5.2 and added varying radial distortion. The ground truth focal length was $f_{gt} = 1000$ and the principal point was chosen randomly at distance 500 px from the origin. We also added some small Gaussian noise ($\sigma = 0.1$ px) to the image coordinates. Figure 8 shows the median relative focal length and rotation errors for different radial distortion parameters. For the solvers which do not model the radial distortion the error increases drastically when distortion is added. We can also see that the P7Pfruv solver has slightly worse performance on image data with little or no radial distortion.

5.5. Real Data

We evaluated all proposed solvers on real image data and compared them with the current state-of-the-art solvers. We downloaded 101 images of the Notre Dame cathedral from the Internet. All downloaded images have a square resolution varying from 800 px × 800 px to 3000 px × 3000 px. Since the images have a square resolution, there was a higher probability that some of these images were edited or cropped and that their principal points are not in the center of the image. Some example images are shown in Figure 9.

Using the RealityCapture software [1] we built a 3D reconstruction of the scene. Since the dataset is quite challenging and it contains many manually edited images, images taken at different conditions, or images with a small overlap, the RealityCapture was only able to register 81 of the 101 images. The Notre Dame reconstruction contains
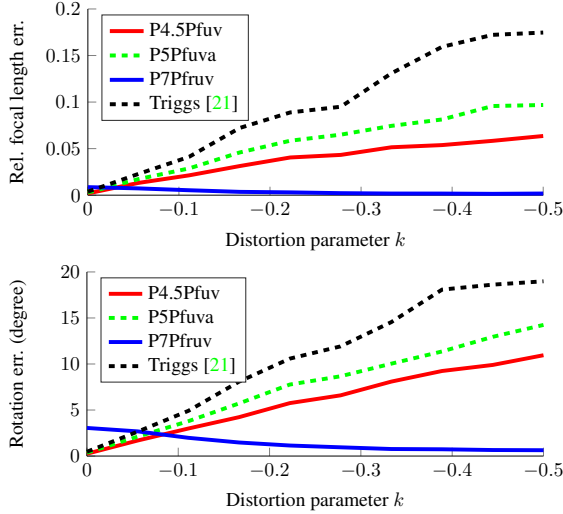
Figure 8. Varying radial distortion. *Top:* Relative focal length error $\frac{|f-f_{gt}|}{f_{gt}}$. *Bottom:* Rotation error in degrees.

24762 3D points, the average reprojection error was 0.6277 pixels and the maximum 1.997 pixels over 73543 image points. According to the principal point estimates returned by RealityCapture, 32 from 81 images have the principal point shifted by more than 6% of the image size.

We used the 3D model and 2D-to-3D correspondences returned by RealityCapture to estimate the pose of each image using the solvers in a simple RANSAC framework with the number of RANSAC iterations fixed to 1000. We used the camera intrinsic and extrinsic parameters obtained from RealityCapture as the ground truth for the experiment. Table 1 shows the errors for the focal length, the radial distortion, the camera pose, as well as the ratio of inliers obtained by different solvers for all 81 registered images. Table 2 shows the same errors for 32 images with the principal point shift larger than 6% of the image size. The full results, including medians of errors, errors for the principal point, the results for images with small principal point shifts and the results of the new P7Pfruv solver with non-linear refinement can be found in the supplementary material.

Overall, the errors are quite small and the new solvers (marked bold) perform the best. It is visible that solvers which assume the principal point in the center of the image (P3.5Pf, P4Pfr) perform significantly worse than the solvers with unknown principal point, especially on images with larger principal point shift (Table 2). Even though the images from Notre Dame dataset do not have a significant radial distortion, the new radial distortion solver helps to improve the results.

5.6. Real Images with Radial Distortion

Finally, we evaluate our solvers on real images which have both a shifted principal point and significant radial distortion. Since it is difficult to generate reliable ground truth

for this problem we took the following approach. We started with the *Rotunda* dataset which was used in [16, 18]. The dataset contains images captured with a wide-angle cameras with large radial distortion and a 3D reconstruction with cameras' intrinsic and extrinsic parameters (see [18] for more details). We used these camera parameters as a ground truth data. For each image in the dataset we cropped out 80% of the image, starting from each of the four corners. See Figure 9 for an example. This gave us a new dataset with $4 \times 62 = 248$ images which have both radial distortion and shifted principal points.

The results for running 1000 iterations of RANSAC are shown in Table 3. We can see that the best results are obtained using the new P7Pfruv solver which can model both radial distortion and shifted principal point. Note that these are the results without any non-linear refinement. For more results see the supplementary material.

6. Conclusions

In this paper, we have revisited the camera pose estimation problem with unknown principal point. We proposed effective polynomial constraints to avoid trivial rank-deficient solutions, and successfully developed the first exactly minimal solver for the case of unknown principal point and focal length by using four and a half point correspondences (P4.5Pfuv), as well as an extremely efficient variant using five point correspondences in the presence of unknown aspect ratio (P5Pfuva). We have also explored the extremely challenging case of unknown principal point and radial distortion, and developed the first practical non-minimal solver by using seven point correspondences (P7Pfruv). The applicability of these new solvers has been verified on both synthetic data and real images.

The high non-linearity in the case of both unknown principal point and radial distortion prevents us from developing an exactly minimal solver. We will continue to explore the feasibility of the exactly minimal problem in the future.

7. Acknowledgements

Figure 9. Example images. *Left:* Images from the *NotreDame* internet dataset. *Middle:* Image from the original *Rotunda* dataset [16]. *Right:* Cropped images from the *Rotunda* dataset used for the experiment in Section 5.6.

		P3.5Pf [18]	P6P [11]	P5Pfuv [21]	**P4.5Pfuv**	**P4.5Pfuv (SVD)**	**P4.5Pfuv (6pt)**	**P5Pfuva**	P4Pfr [18]	**P7Pfruv**
Focal length	mean	0.1256	0.0242	0.0234	0.0209	0.0189	**0.0142**	0.0243	0.1012	0.0211
	max	1.3042	0.1342	0.1123	0.1449	0.0962	**0.0535**	0.1176	0.9307	0.0807
Rotation	mean	2.9495	1.2067	1.2559	1.2544	1.2564	1.1228	1.2322	3.0043	**0.9005**
	max	19.5789	5.1510	4.1978	3.9726	3.8889	4.0754	4.5883	21.7933	**2.7497**
Translation	mean	0.6196	0.1163	0.1249	0.0894	0.0849	**0.0714**	0.1200	0.4941	0.1016
	max	5.4442	0.6711	1.6183	**0.5100**	0.5809	0.7289	0.7050	4.1039	0.5929
Distortion	mean	-	-	-	-	-	-	-	0.0868	**0.0474**
	max	-	-	-	-	-	-	-	**1.0509**	1.9341
Inlier (%)	mean	74.9315	88.0265	86.8770	87.8971	88.6420	89.4039	87.5297	82.1422	**93.2453**

Table 1. *NotreDame* dataset: Comparison of different solvers on 81 images with 1000 RANSAC iterations. The table shows the relative errors except for the rotation errors which are in degrees. The best results are marked bold.

		P3.5Pf [18]	P6P [11]	P5Pfuv [21]	**P4.5Pfuv**	**P4.5Pfuv (SVD)**	**P4.5Pfuv (6pt)**	**P5Pfuva**	P4Pfr [18]	**P7Pfruv**
Focal length	mean	0.2615	0.0295	0.0228	0.0185	0.0186	**0.0123**	0.0286	0.2177	0.0192
	max	1.3042	0.1342	0.0809	0.0572	0.0762	**0.0437**	0.1176	0.9307	0.0807
Rotation	mean	5.6991	1.6485	1.5334	1.3566	1.3602	1.3361	1.5225	5.9204	**0.8759**
	max	19.5789	5.1510	4.1978	3.9726	3.8889	4.0754	4.5883	21.7933	**2.7497**
Translation	mean	1.2483	0.1679	0.1528	0.1004	0.0978	**0.0699**	0.1646	1.0641	0.1140
	max	5.4442	0.6711	1.6183	0.3328	0.3557	**0.2896**	0.6612	4.1039	0.4998
Distortion	mean	-	-	-	-	-	-	-	0.1647	**0.0240**
	max	-	-	-	-	-	-	-	1.0509	**0.1082**
Inlier (%)	mean	61.7820	89.4821	87.9500	89.3505	90.0521	90.8913	88.6629	66.8490	**93.0248**

Table 2. *NotreDame* dataset: Comparison of different solvers on 32 images with principal point shift > 6% and 1000 RANSAC iterations. The table shows the relative errors except for the rotation errors which are in degrees. The best results are marked bold.

		P3.5Pf [18]	P6P [11]	P5Pfuv [21]	**P4.5Pfuv**	**P4.5Pfuv (SVD)**	**P4.5Pfuv (6pt)**	**P5Pfuva**	P4Pfr [18]	**P7Pfruv**
Focal length	mean	0.3696	0.2199	0.1667	0.1699	0.1666	0.1696	0.1928	0.0845	**0.0012**
	max	4.1083	0.5135	0.5066	0.4187	0.4461	0.4217	0.5450	0.5896	**0.0040**
Rotation	mean	14.6800	10.7499	8.4318	8.4168	8.3194	8.2880	9.7119	13.6371	**0.2558**
	max	176.5515	24.4273	22.8580	20.9626	22.7519	22.4078	27.0481	21.1274	**0.7516**
Translation	mean	0.4538	0.1687	0.1446	0.1436	0.1434	0.1450	0.1535	0.2181	**0.0043**
	max	5.9461	0.8867	0.6083	0.6217	0.6804	0.7358	0.6327	0.7070	**0.0138**
Distortion	mean	-	-	-	-	-	-	-	0.2042	**0.0067**
	max	-	-	-	-	-	-	-	2.9758	**0.0334**
Inlier (%)	mean	21.26	40.46	34.73	36.04	36.81	36.93	37.45	33.18	**96.78**

Table 3. Cropped *Rotunda* dataset: Comparison of different solvers on 248 images with radial distortion and shifted principal point. The table shows the relative errors except for the rotation errors which are in degrees. The best results are marked bold.

References

[1] RealityCapture. www.capturingreality.com. 6

[2] M. Bujnak, Z. Kukelova, and T. Pajdla. A general solution to the p4p problem for camera with unknown focal length. In *Computer Vision and Pattern Recognition (CVPR)*, 2008. 2

[3] M. Bujnak, Z. Kukelova, and T. Pajdla. New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion. In *Asian Conference on Computer Vision (ACCV)*, 2010. 2

[4] D. Cox, J. Little, and D. O'shea. *Ideals, varieties, and algorithms*, volume 3. Springer, 1992. 2, 4

[5] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993. 1, 2

[6] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. ACM*, 24(6):381–395, 1981. 1

[7] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Computer Vision and Pattern Recognition (CVPR)*, 2001. 2

[8] X. Gao, X. Hou, J. Tang, and H. Cheng. Complete solution classification for the perspective-three-point problem. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 25(8):930–943, 2003. 1

[9] D. R. Grayson and M. E. Stillman. Macaulay 2, a software system for research in algebraic geometry, 2002. 3, 4

[10] B. Haralick, C. Lee, K. Ottenberg, and M. lle. Review and analysis of solutions of the three point perspective pose estimation problem. *International Journal of Computer Vision (IJCV)*, 13(3):331–356, 1994. 1

[11] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2nd edition, 2003. 2, 5, 6, 8

[12] A. Heyden. *Geometry and Algebra of Multiple Projective Transforms*. PhD thesis, PhD thesis, Lund University, Lund Institute of Technology, Department of Mathematics, 1995. 1, 2

[13] K. Josephson and M. Byröd. Pose estimation with radial distortion and unknown focal length. In *Computer Vision and Pattern Recognition (CVPR)*, 2009. 2

[14] L. Kneip, D. Scaramuzza, and R. Siegwart. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *Computer Vision and Pattern Recognition (CVPR)*, 2011. 1

[15] Z. Kukelova, M. Bujnak, and T. Pajdla. Real-time solution to the absolute pose problem with unknown radial distortion and focal length. In *International Conference on Computer Vision (ICCV)*, 2013. 2, 4

[16] Z. Kukelova, J. Heller, M. Bujnak, A. Fitzgibbon, and T. Pajdla. Efficient solution to the epipolar geometry for radially distorted cameras. In *International Conference on Computer Vision (ICCV)*, 2015. 7, 8

[17] V. Larsson and K. Åström. Uncovering symmetries in polynomial systems. In *European Conference on Computer Vision (ECCV)*, 2016. 2, 4

[18] V. Larsson, Z. Kukelova, and Y. Zheng. Making minimal solvers for absolute pose estimation compact and robust. In *International Conference on Computer Vision (ICCV)*, 2017. 2, 6, 7, 8

[19] V. Larsson, K. Åstrom, and M. Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 3, 4

[20] V. Larsson, K. Åstrom, and M. Oskarsson. Polynomial solvers for saturated ideals. In *International Conference on Computer Vision (ICCV)*, 2017. 3

[21] B. Triggs. Camera pose and calibration from 4 or 5 known 3d points. In *International Conference on Computer Vision (ICCV)*, 1999. 1, 2, 5, 6, 7, 8

[22] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal on Robotics and Automation*, 3(4):323–344, 1987. 4

[23] C. Wu. P3.5p: Pose estimation with unknown focal length. In *Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 4

[24] Y. Zheng, S. Sugimoto, I. Sato, and M. Okutomi. A general and simple method for camera pose and focal length determination. In *Computer Vision and Pattern Recognition (CVPR)*, 2014. 2