Minimal Rolling Shutter Absolute Pose with Unknown Focal Length and Radial Distortion

Zuzana Kukelova[1]([✉]), Cenek Albl[2], Akihiro Sugimoto[3], Konrad Schindler[2], and Tomas Pajdla[4]

[1] Visual Recognition Group (VRG), Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czechia
kukelova@fel.cvut.cz
[2] Photogrammetry and Remote Sensing, ETH Zurich, Zurich, Switzerland
[3] National Institute of Informatics, Tokyo, Japan
[4] Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czechia

Abstract. The internal geometry of most modern consumer cameras is not adequately described by the perspective projection. Almost all cameras exhibit some radial lens distortion and are equipped with electronic rolling shutter that induces distortions when the camera moves during the image capture. When focal length has not been calibrated offline, the parameters that describe the radial and rolling shutter distortions are usually unknown. While, for global shutter cameras, minimal solvers for the absolute camera pose and unknown focal length and radial distortion are available, solvers for the rolling shutter were missing. We present the first minimal solutions for the absolute pose of a rolling shutter camera with unknown rolling shutter parameters, focal length, and radial distortion. Our new minimal solvers combine iterative schemes designed for calibrated rolling shutter cameras with fast generalized eigenvalue and Gröbner basis solvers. In a series of experiments, with both synthetic and real data, we show that our new solvers provide accurate estimates of the camera pose, rolling shutter parameters, focal length, and radial distortion parameters. The implementation of our solvers is available at *github.com/CenekAlbl/RnP*.

1 Introduction

Estimating the six degree-of-freedom (6DOF) pose of a camera is a fundamental problem in computer vision with applications in camera calibration [7], Structure-from-Motion (SfM) [36,37], augmented reality (AR) [32], and visual localization [33]. The task is to compute the camera pose in the world coordinate system from 3D points in the world and their 2D projections in an image.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-58558-7_41) contains supplementary material, which is available to authorized users.

Fig. 1. Removing RS and radial distortion simultaneously using our minimal absolute pose solver. The original image (left) with tentative correspondences (black) the inliers captured by P4Pfr+R6P [4] (blue) and subsequent local optimization (red) compared to inliers captured by the proposed R7Pfr (cyan) and subsequent local optimization (green). The correction using R7Pfr without local optimization (right) is better than by P4Pfr+R6P with local optimization (middle). (Color figure online)

Solvers for the camera pose are usually used inside RANSAC-style hypothesis-and-test frameworks [12]. For efficiency it is therefore important to employ *minimal* solvers that generate the solution with a minimal number of point correspondences. The minimal number of 2D-to-3D correspondences necessary to solve the absolute pose problem is three for a calibrated perspective camera. The earliest solver dates back to 1841 [14]. Since then, the problem has been revisited several times [6,12,15,16,21]. In many situations, however, the internal camera calibration is unavailable, e.g.. when working with crowdsourced images. Consequently, methods have been proposed to jointly estimate the camera pose together with focal length [8,27,38,39]. These methods have been extended to include also estimation of an unknown principal point [28], and unknown radial distortion [20,27]. The latter is particularly important for the wide-angle lenses commonly used in mobile phones and GoPro-style action cameras. The absolute pose of fully uncalibrated perspective camera without radial distortion can be estimated from six correspondence using the well-known DLT solver [1]. All these solutions assume a perspective camera model and are not suitable for cameras with rolling shutter (RS), unless the camera and the scene can be kept static.

Rolling shutter is omnipresent from consumer phones to professional SLR cameras. Besides technical advantages, like higher frame-rate and longer exposure time per pixel, it is also cheaper to produce. The price to pay is that the rows of an "image" are no longer captured synchronously, leading to motion-induced distortions and in general to a more complicated imaging geometry.

Motivation: While several minimal solutions have been proposed for the absolute pose of an RS camera with calibrated intrinsics [3–5,22], minimal solutions for uncalibrated RS cameras are missing. One obvious way to circumvent that problem is to first estimate the intrinsic and radial distortion parameters while ignoring the rolling shutter effect, then recover the 3D pose and rolling shutter parameters with an absolute pose solver for calibrated RS cameras [3,4,22]. Ignoring the deviation from the perspective projection in the first step can, however, lead to wrong estimates. For example, if the image point distribution is unfavourable, it may happen that RS distortion is compensated by an (incorrect) change of radial distortion, see Fig. 1.

Contribution: We present the first minimal solutions for two rolling shutter absolute pose problems:

1. absolute pose estimation of an RS camera with unknown focal length from 7 point correspondences; and
2. absolute pose estimation of an RS camera with unknown focal length *and* unknown radial distortion, also from 7 point correspondences.

The new minimal solvers combine two ingredients: a recent, iterative approach introduced for pose estimation of calibrated RS cameras [22]; and fast polynomial eigenvalue [11] and Gröbner basis solvers for comparatively simple, tractable systems of polynomial equations [10,26]. In experiments with synthetic and real data, we show that for uncalibrated RS cameras our new solvers find good estimates of camera pose, RS parameters, focal length, and radial distortion. We demonstrate that the new all-in-one solvers outperform alternatives that sequentially estimate first perspective intrinsics, then RS correction and extrinsics.

2   Related Work

The problem of estimating the absolute pose of a camera from a minimal number of 2D-to-3D point correspondences is important in geometric computer vision. Minimal solvers are often the main building blocks for SfM [36,37] and localization pipelines [33]. Therefore, during the last two decades a large variety of minimal absolute pose solvers for perspective cameras with or without radial distortion have been proposed.

For estimating the absolute pose of a calibrated camera, three points are necessary and the resulting system of polynomial equations can be solved in a closed form [21]. If the camera intrinsics and radial distortion are unknown, more point correspondences are required and the resulting systems of polynomial equations become more complex. The most common approach to solve such systems of polynomial equations is to use the Gröbner basis method [10] and automatic generators of efficient polynomial solvers [23,26].

Most of the minimal absolute pose solvers have been developed using the Gröbner basis method. These include four or 3.5 point minimal solvers (P4Pf or P3.5Pf solvers) for the perspective camera with unknown focal length, and known or zero radial distortion [8,27,38,39], four point (P4Pfr) solvers for perspective cameras with unknown focal length and unknown radial distortion [9,20,27], and P4.5Pfuv solver for unknown focal length and unknown principal point [28].

Recently, as RS cameras have become omnipresent, the focus has turned to estimating the camera absolute pose from images affected by RS distortion. RS cameras motion models [31] result in more complex systems of polynomial equations than perspective cameras models. Therefore, most of the existing RS absolute pose solvers use some model relaxations [3,4,22,35], scene assumptions such as planarity [2], additional information e.g. from IMU [5] or a video sequence [18], and a non-minimal number of point correspondences [2,25,30]. Moreover, all the existing solutions assume calibrated RS cameras, i.e., they assume that the camera intrinsic as well as radial distortion are known.

The first minimal solution to the absolute pose problem for a calibrated RS camera was presented in [3]. The proposed solver uses the minimal number of six 2D-to-3D point correspondences and the Gröbner basis method to generate an efficient solver. The proposed R6P is based on the constant linear and angular velocity model as in [2,19,30], but it uses the first order approximation to both the camera orientation and angular velocity, and, therefore, it requires an initialization of the camera orientation, e.g., from P3P [12]. It is shown in [3] that the proposed R6P solver significantly outperforms the P3P solver in terms of camera pose precision and the number of inliers captured in the RANSAC loop. The R6P solver was extended in [4] by linearizing only the angular velocity and also by proposing a faster solution to the "double-linearized" model. The model that linearizes only the angular velocity does not require any initialization of the camera orientation, however it results in a slower and more complicated solver. Moreover, it has been shown in [4] that such solver usually produces similar results as the "double-linearized" solver initialized with P3P [12,21].

The double-linearized model [3,4] results in a quite complex system of six quadratic equations in six unknowns with 20 solutions. The fastest solver to this problem, presented in [4], runs 0.3 ms and is not suitable for real-time applications. Therefore, a further simplification of the double-linearized model was proposed [22]. The model in [22] is based on the assumption that after the initialization with the P3P solver, the camera rotation is already close to the identity, and that in real applications, the rolling shutter rotation during the capture is usually small. Therefore, some nonlinear terms (monomials) in the double-linearized model are usually small, sometimes even negligible. Based on this assumption, a linear iterative algorithm was proposed in [22]. In the first iteration, the algorithm substitutes negligible monomials with zeros. In each subsequent iteration, it substitutes these monomials with the estimates from the previous iteration. In this way, the original, complicated system of polynomial equations is approximated with a system of linear equations. This new linear iterative algorithm usually converges to the solutions of the original system in no more than five iterations and is an order of magnitude faster than [4].

Different from the above mentioned methods for calibrated RS cameras, we combine the iterative scheme designed for calibrated RS cameras [22] with fast generalized eigenvalue and Gröbner basis solvers [26] for specific polynomial equation systems to solve the previously unsolved problem of estimating the absolute pose of an *uncalibrated* RS camera (i.e., unknown RS parameters, focal length, and radial distortion) from a minimal number of point correspondences.

3   Problem Formulation

For perspective cameras with radial distortion, the projection equation can be written as

$$\alpha_i u(\mathbf{x}_i, \boldsymbol{\lambda}) = \mathtt{K}[\mathtt{R} \mid \mathtt{C}]\mathtt{X}_i, \tag{1}$$

where $\mathtt{R} \in SO(3)$ and $\mathtt{C} \in \mathbb{R}^3$ are the rotation and translation bringing a 3D point $\mathtt{X}_i = [x_i, y_i, z_i, 1]^\top$ from the world coordinate system to the camera coordinate

system, $\mathbf{x}_i = [r_i, c_i, 1]^\top$ are the homogeneous coordinates of a measured distorted image point, $u(\cdot, \boldsymbol{\lambda})$ is an image undistortion function with parameters $\boldsymbol{\lambda}$, and $\alpha_i \in \mathbb{R}$ is a scalar.

Matrix $\mathtt{K}$ is a 3×3 matrix known as the calibration matrix containing the intrinsic parameters of a camera. Natural constraints satisfied by most consumer cameras with modern CCD or CMOS sensor are zero skew and the unit aspect ratio [17]. The principal point [17] is usually also close to the image center ($[p_x, p_y]^\top = [0, 0]^\top$). Thus the majority of existing absolute pose solvers adhere to those assumptions, and we do so, too. Hence, we adopt calibration matrix

$$\mathtt{K} = \mathrm{diag}\,(f, f, 1)\,. \tag{2}$$

For cameras with lens distortion, measured image coordinates $\mathbf{x}_i$ have to be transformed into "pinhole points" with an undistortion function $u(\cdot, \boldsymbol{\lambda})$. For standard cameras, the radial component of the lens distortion is dominant, whereas the tangential component is negligible at this stage. Therefore, most camera models designed for minimal solvers consider only radial distortion[1]. A widely used model represents radial lens distortion with a one-parameter division [13]. This model is especially popular with absolute pose solvers thanks to its compactness and expressive power: it can capture even large distortions of wide-angle lenses (e.g., GoPro-type action cams) with a single parameter. Assuming that the distortion center is in the image center, the division model is

$$u\,(\mathbf{x}_i, \lambda) = u\left(\begin{bmatrix} r_i \\ c_i \\ 1 \end{bmatrix}, \lambda\right) = \begin{bmatrix} r_i \\ c_i \\ 1 + \lambda(r_i^2 + c_i^2) \end{bmatrix}. \tag{3}$$

Unlike perspective cameras, RS cameras capture every image row (or column) at a different time, and consequently, reveal the presence of relative motion between the camera and the scene at a different position. Camera rotation $\mathtt{R}$ and translation $\mathtt{C}$ are, therefore, functions of the image row r_i (or column). Together with the calibration matrix $\mathtt{K}$ of (2) and the distortion model (3), the projection equation of RS cameras is

$$\alpha_i u(\mathbf{x}_i, \lambda) = \alpha_i \begin{bmatrix} r_i \\ c_i \\ 1 + \lambda(r_i^2 + c_i^2) \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} [\mathtt{R}(r_i) \mid \mathtt{C}(r_i)]\mathbf{X}_i. \tag{4}$$

Let $\mathtt{R}_0$ and $\mathtt{C}_0$ be the unknown rotation and translation of the camera at time $\tau = 0$, which denotes the acquisition time of the middle row $r_0 \in \mathbb{R}$. Then, for the short time-span required to record all rows of a frame (typically < 50 ms), the translation $\mathtt{C}(r_i)$ can be approximated by a constant velocity model [2,3,19, 30,31,34]:

$$\mathtt{C}(r_i) = \mathtt{C}_0 + (r_i - r_0)\mathtt{t}, \tag{5}$$

---

[1] For maximum accuracy, tangential distortion can be estimated in a subsequent nonlinear refinement.

with the translational velocity $\mathtt{t}$.

The rotation $\mathtt{R}(r_i)$, on the other hand, can be decomposed into two parts: the initial orientation $\mathtt{R}_0$ of r_0, and the change of the orientation relative to it: $\mathtt{R}_\mathtt{w}(r_i - r_0)$. In [3,30], it was established that for realistic motions it is usually sufficient to linearize $\mathtt{R}_\mathtt{w}(r_i - r_0)$ around the initial rotation $\mathtt{R}_0$ via the first-order Taylor expansion. Thereby the RS projection (4) becomes

$$\alpha_i \begin{bmatrix} r_i \\ c_i \\ 1 + \lambda(r_i^2 + c_i^2) \end{bmatrix} = \mathtt{K}\left[(\mathtt{I} + (r_i - r_0)[\mathtt{w}]_\times)\,\mathtt{R}_0 \mid \mathtt{C}_0 + (r_i - r_0)\mathtt{t}\right]\mathtt{X}_i, \qquad (6)$$

where $[\mathtt{w}]_\times$ is the skew-symmetric matrix for the vector $\mathtt{w} = [w_1, w_2, w_3]^\top$.

The linearized model (6) is sufficient for all scenarios except for the most extreme motions (which anyway present a problem due to motion blur that compromises keypoint extraction).

Unfortunately, the system of polynomial Eq. (6) is rather complex even with the linearized rolling shutter rotation. Already for calibrated RS camera and assuming Cayley parametrization of $\mathtt{R}_0$, this model results in six equations of degree three in six unknowns and 64 solutions [4].

Therefore, following [3,4], we employ another linear approximation to the camera orientation $\mathtt{R}_0$ to have the double-linearized model:

$$\alpha_i \begin{bmatrix} r_i \\ c_i \\ 1 + \lambda(r_i^2 + c_i^2) \end{bmatrix} = \mathtt{K}\left[(\mathtt{I} + (r_i - r_0)[\mathtt{w}]_\times)\,(\mathtt{I} + [\mathtt{v}]_\times) \mid \mathtt{C}_0 + (r_i - r_0)\mathtt{t}\right]\mathtt{X}_i. \quad (7)$$

This model leads to a simpler way of solving the calibrated RS absolute pose from $\geq$ six 2D-3D point correspondences than the model in [4]. However, the drawback of this further simplification is the fact that, other than the relative intra-frame rotation, the absolute rotation $\mathtt{R}_0$ can be of arbitrary magnitude, and therefore far from the linear approximation. A practical solution for calibrated cameras is to compute a rough approximate pose with a standard P3P solver [12,21], align the object coordinate system to it so that the remaining rotation is close enough to identity, and then run the RS solver [3,4,22].

The double-linearized model (7) is simpler than the original one (6), but still leads to a complex polynomial system (for calibrated RS cameras a system of six quadratic equations in six unknowns with up to 20 real solutions), and is rather slow for practical use. Therefore, further simplification of the double-linearized model was proposed in [22]. That model uses the fact that both the absolute rotation (after P3P initialisation) *and* the rolling shutter rotation $\mathtt{w}$ are small. As a consequence [22] assumes that the nonlinear term $[\mathtt{w}]_\times[\mathtt{v}]_\times$ in (7) is sufficiently small (sometimes even negligible). With this assumption, one can further linearize the nonlinear term $[\mathtt{w}]_\times[\mathtt{v}]_\times$ in (7) by approximating $[\mathtt{v}]_\times$ with $[\hat{\mathtt{v}}]_\times$, while keeping the remaining linear terms as they are; which leads to an efficient iterative solution of the original system: solve a resulting linearized system to estimate all unknowns including $[\mathtt{v}]_\times$, and iterate with updated $[\hat{\mathtt{v}}]_\times \leftarrow [\mathtt{v}]_\times$ until convergence. As initial approximation one can set $[\hat{\mathtt{v}}]_\times = \mathtt{0}$.

Algorithm 1. Iterative absolute pose solver for uncalibrated RS camera [with unknown radial distortion]

Input: $\mathbf{x}_i, \mathbf{X}_i, \{i = 1, \ldots, 7\}, k_{\max}, \epsilon_{\mathrm{err}}$
Output: $\mathbf{v}, \mathbf{C}_0, \mathbf{w}, \mathbf{t}, f, [\lambda]$
  $\mathbf{v}^0 \leftarrow \mathbf{0}, k \leftarrow 1$
  $\mathbf{R}_{\mathrm{GS}}, \mathbf{C}_{\mathrm{GS}}, f_{\mathrm{GS}} \leftarrow \mathrm{P4Pf}(\mathbf{x}_i, \mathbf{X}_j)$ [27]
  $[\mathbf{R}_{\mathrm{GS}}, \mathbf{C}_{\mathrm{GS}}, f_{\mathrm{GS}}, \lambda_{\mathrm{GS}} \leftarrow \mathrm{P4Pfr}(\mathbf{x}_i, \mathbf{X}_j)$ [27]]
  $\mathbf{X}_i \leftarrow \mathbf{R}_{\mathrm{GS}}\mathbf{X}_i$
  while $k < k_{\max}$ **do**
    $\hat{\mathbf{v}} \leftarrow \mathbf{v}^{k-1}$
    $err^k_{\max} \leftarrow \infty$
    $[\lambda_{\mathrm{RS}}], \mathbf{v}_{\mathrm{RS}}, \mathbf{C}_{0\mathrm{RS}}, \mathbf{w}_{\mathrm{RS}}, \mathbf{t}_{\mathrm{RS}}, f_{\mathrm{RS}}, \leftarrow$ solve Eq. (8)
    for $j = 1$ **to** #solutions of Eq. (8) **do**
      $err_j \leftarrow$ Residual error of Eq. (7) evaluated on $\{\mathbf{v}_{\mathrm{RS}j}, \mathbf{C}_{0\mathrm{RS}j}, \mathbf{w}_{\mathrm{RS}j}, \mathbf{t}_{\mathrm{RS}j}, f_{\mathrm{RS}j}, [\lambda_{\mathrm{RS}j}]\}$

      if $err_j < err_{\max}$ **then**
        $\{\mathbf{v}^k, \mathbf{C}^k, \mathbf{w}^k, \mathbf{t}^k, f^k, [\lambda^k]\} \leftarrow \{\mathbf{v}_{\mathrm{RS}j}, \mathbf{C}_{0\mathrm{RS}j}, \mathbf{w}_{\mathrm{RS}j}, \mathbf{t}_{\mathrm{RS}j}, f_{\mathrm{RS}j}, [\lambda_{\mathrm{RS}j}]\}$
        $err^k_{\max} \leftarrow err_j$
      end if
    end for
    if $err^k_{\max} < \epsilon_{\mathrm{err}}$ **or** $\left(|err^k_{\max} - err^{k-1}_{\max}| < \epsilon_{\mathrm{err}} \,\&\, k > 1\right)$ **then**
      return $\{\mathbf{v}^k, \mathbf{C}^k, \mathbf{w}^k, \mathbf{t}^k, f^k, [\lambda^k]\}$
    end if
    $k \leftarrow k + 1$
  end while
  return $\{\mathbf{v}^{k-1}, \mathbf{C}^{k-1}, \mathbf{w}^{k-1}, \mathbf{t}^{k-1}, f^{k-1}, [\lambda^{k-1}]\}$

Here we are interested in RS cameras with *unknown* focal length and radial distortion. In that setting (7) leads to a much more complicated system of polynomial equations that exceeds the capabilities of existing algebraic methods such as Gröbner bases [10,23,26]. We adopt a similar relaxation as in [22] and linearize $[\mathbf{w}]_\times [\mathbf{v}]_\times$ by substituting with the preliminary value $[\mathbf{v}]_\times \leftarrow [\hat{\mathbf{v}}]_\times$. Without loss of generality, let us assume that $r_0 = 0$. Then, the projection equation for this relaxed model is

$$\alpha_i \begin{bmatrix} r_i \\ c_i \\ 1 + \lambda(r_i^2 + c_i^2) \end{bmatrix} = \mathbf{K} \left[\mathbf{I} + r_i[\mathbf{w}]_\times + [\mathbf{v}]_\times + r_i[\mathbf{w}]_\times[\hat{\mathbf{v}}]_\times \mid \mathbf{C}_0 + r_i\mathbf{t} \right] \mathbf{X}_i. \quad (8)$$

4   Minimal Solvers

To develop efficient minimal solvers for uncalibrated RS cameras, we advance the idea of the calibrated iterative RS solver of [22] by combining it with a generalized eigenvalue and efficient Gröbner basis solvers for specific polynomial equation systems.

We develop two new solvers. They both first pre-rotate the scene with a rotation estimated using efficient perspective absolute pose solvers for uncalibrated

cameras, i.e. the P3.5Pf [27] and P4Pfr [27]/P5Pfr [24]. Then, they iterate two steps: (i) solve the system of polynomial equations derived from (8), with fixed preliminary $\hat{\mathbf{v}}$. (ii) update $\hat{\mathbf{v}}$ with the current estimates of the unknown parameters. The iteration is initialised with $\hat{\mathbf{v}} = \mathbf{0}$. A compact summary in the form of pseudo-code is given in Algorithm 1. Note that after solving the polynomial system (8), we obtain, in general, more than one feasible solution (where "feasible" means real and geometrically meaningful values, e.g., $f > 0$). To identify the correct one among them, we evaluate the (normalized) residual error of the original equations (7), and choose the one with the smallest error.

The described computational scheme of Algorithm 1 covers both the case where the radial distortion is known and only the pose, focal length and RS parameters must be found, and the case where also radial distortion is unknown. In the following, we separately work out the R7Pf solver for known radial distortion and the R7Pfr solver for unknown radial distortion. Both cases require seven point correspondences.

4.1   R7Pf - RS Absolute Pose with Unknown Focal Length

In the first solver, we assume that the camera has a negligible radial distortion (since known, non-zero distortion can be removed by warping the image point coordinates). This R7Pf solver follows the iterative procedure of Algorithm 1. What remains to be specified is how to efficiently solve the polynomial system (8) with $\lambda = 0$.

The R7Pf solver first eliminates the scalar values α_i by left-multiplying equation (8) with the skew-symmetric matrix $[\mathbf{x}_i]_\times$ for the vector $\mathbf{x}_i = \begin{bmatrix} r_i & c_i & 1 \end{bmatrix}^\top$. Since the projection equation (8) is defined only up to scale, we multiply the whole equation with $q = \frac{1}{f}$ ($f \neq 0$), resulting in

$$\begin{bmatrix} 0 & -1 & c_i \\ 1 & 0 & -r_i \\ -c_i & r_i & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & q \end{bmatrix} \left[\mathbf{I} + r_i[\mathbf{w}]_\times + [\mathbf{v}]_\times + r_i[\mathbf{w}]_\times [\hat{\mathbf{v}}]_\times \mid \mathbf{C}_0 + r_i\mathbf{t} \right] \mathbf{X}_i = \mathbf{0}. \quad (9)$$

(9) has 13 degrees of freedom (corresponding to 13 unknowns): $\mathbf{v}, \mathbf{w}, \mathbf{C}_0, \mathbf{t}$, and $q = \frac{1}{f}$. Since each 2D–3D point correspondence gives two linearly independent equations (only two equations in (9) are linearly independent due to the singularity of the skew-symmetric matrix), we need $6\frac{1}{2}$ point correspondences for a minimal solution.

Since half-points for which only one coordinate is known normally do not occur, we present a 7-point solver and just drop out one of the constraints in computing the camera pose, RS parameters, and focal length. The dropped constraint can be further used to filter out geometrically incorrect solutions.

After eliminating the scalar values α_i, the R7Pf solver starts with equations corresponding to the 3rd row of (9) for $i = 1, \ldots, 7$. These equations are linear in ten unknowns and do not contain the unknown $q = \frac{1}{f}$, indicating that they are independent of focal length. Let us denote the elements of unknown vectors by $\mathbf{v} = [v_1, v_2, v_3]^\top$, $\mathbf{w} = [w_1, w_2, w_3]^\top$, $\mathbf{C}_0 = [C_{0x}, C_{0y}, C_{0z}]^\top$, and $\mathbf{t} = [t_x, t_y, t_z]^\top$.

Then, the equations corresponding to the 3rd row of (9) for $i = 1, \ldots, 7$ can be written as

$$My = 0 \tag{10}$$

where M is a 7×11 coefficient matrix and y is a 11×1 vector of monomials: $y = [v_1, v_2, v_3, w_1, w_2, w_3, C_{0x}, C_{0y}, t_x, t_y, 1]^\top$. For points in the general configuration, the matrix M in (10) has a 4-dimensional null-space, so we can write the unknown vector y as a linear combination of four 11×1 basis vectors y_1, y_2, y_3, and y_4 of that null-space:

$$y = \beta_1 y_1 + \beta_2 y_2 + \beta_3 y_3 + \beta_4 y_4, \tag{11}$$

where β_j $(j = 1, \ldots, 4)$ are new unknowns. One of these unknowns, e.g.. β_4, can be eliminated (expressed as a linear combination of the remaining three unknowns $\beta_1, \beta_2, \beta_3$), using the constraint on the last element of y, which by construction is 1.

In the next step, the parameterization (11) is substituted into the equations corresponding to the 1st (or 2nd) row of (9) for $i = 1, \ldots, 6$. Note that here we use only six of seven available equations. The substitution results in six polynomial equations in six unknowns $\beta_1, \beta_2, \beta_3, C_{0z}, t_z, q$, and 10 monomials $m = [\beta_1 q, \beta_1, \beta_2 q, \beta_2, \beta_3 q, \beta_3, C_{0z} q, t_z q, q, 1]$. This is a system of six quadratic equations in six unknowns, which could be solved using standard algebraic methods based on Gröbner bases [10] and automatic Gröbner basis solver generators [23, 26]. However, in this specific case, it is more efficient to transform it to a generalized eigenvalue problem (GEP) of size 6×6, by rewriting it as

$$q A_1 [\beta_1, \beta_2, \beta_3, C_{0z}, t_z, 1]^\top = A_0 [\beta_1, \beta_2, \beta_3, C_{0z}, t_z, 1]^\top, \tag{12}$$

where A_0 and A_1 are 6×6 coefficient matrices. Equation (12) can be solved using standard efficient eigenvalue methods [11]. Alternatively, one can simplify even further by eliminating monomials $C_{0z} q$ and $t_z q$, thereby also eliminating two unknowns C_{0z} and t_z, and then solving a GEP of size 4×4. The remaining unknowns are obtained by back-substitution into (11).

4.2 R7Pfr - RS Absolute Pose with Unknown Focal Length and Unknown Radial Distortion

The R7Pfr solver finds the solution of the minimal problem with unknown absolute pose, RS parameters, focal length, and radial distortion. Compared to the first solver, there is one additional degree of freedom (14 unknowns in total); hence, we need seven 2D-to-3D point correspondences. R7Pfr follows the same iterative approach.

After eliminating the scalar values α_i by left-multiplying Eq. (8) with the skew-symmetric matrix $[u(x_i, \lambda)]_\times$ for $u(x_i, \lambda) = [r_i, c_i, 1 + \lambda(r_i^2 + c_i^2)]^\top$, and multiplying the complete system with $q = \frac{1}{f}(\neq 0)$, we obtain

$$\begin{bmatrix} 0 & d_i & c_i \\ d_i & 0 & -r_i \\ -c_i & r_i & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & q \end{bmatrix} [I + r_i[w]_\times + [v]_\times + r_i[w]_\times [\hat{v}]_\times \mid C_0 + r_i t] X_i = 0, \tag{13}$$

with $d_i = 1 + \lambda(r_i^2 + c_i^2)$ for $i = 1, \ldots, 7$. The polynomial system (13) is more complicated than that without radial distortion, but the third row remains unchanged, as it is independent not only of focal length, but also of radial distortion. We can therefore proceed in the same way: find the 4-dimensional null space of matrix M, eliminate β_4, and substitute the parametrization (11) back into the 1$^{\text{st}}$ (or 2$^{\text{nd}}$) row of (13), to obtain a system of seven quadratic polynomial equations in seven unknowns $\beta_1, \beta_2, \beta_3, C_{0z}, t_z, q, \lambda$, and 14 monomials $m = [\beta_1 q, \beta_1 \lambda, \beta_1, \beta_2 q, \beta_2 \lambda, \beta_2, \beta_3 q, \beta_3 \lambda, \beta_3, C_{0z} q, t_z q, q, \lambda, 1]$.

In the next step, we eliminate the monomials $C_{0z} q$ and $t_z q$ (and, consequently, also two unknowns C_{0z} and t_z) by simple Gauss-Jordan elimination. The resulting system of five quadratic equations in five unknowns $\beta_1, \beta_2, \beta_3, q, \lambda$, and 12 monomials has ten solutions. Different from the R7Pf case, this system does not allow a straight-forward transformation to a GEP. Instead, we solve it with the Gröbner basis method using the automatic solver generator [26].

To find a solver that is as efficient as possible, we follow the recent heuristic [29]. We generate solvers for 1000 different candidate bases and select the most efficient one among them. The winning solver performs elimination on a 26×36 matrix (compared to a 36×46 matrix if using the standard basis and grevlex monomial ordering) and eigenvalue decomposition of a 10×10 matrix. The remaining unknowns are again obtained by the back-substitution to (11).

5   Experiments

We evaluate the performance of both presented solvers on synthetic as well as various real datasets. The main strength of the presented R7Pf and R7Pfr solvers lies in the ability to handle uncalibrated data, which often occurs in the wild.

5.1   Data Setup

Synthetic Data: For the synthetic experiments, we generate random sets of seven points in the cube with side one. We simulate a camera with 60° FOV at random locations facing the center of the cube at a distance between one to four. We generate 1000 samples for each experiment, with 10 increment steps for the parameters that are being varied. We generate the camera motion using constant translational and rotational velocity model and the radial distortion using the one parameter division model. Note that even though our solvers are based on these models, they use approximations to the RS motion and therefore the data is never generated with identical model that is being solved for.

Real Data: We use altogether five datasets. Three outdoor captured by Gopro cameras, two of which are downloaded from Youtube and one was proposed in [4]. Two contain drone footage and one a handheld recording of a rollercoaster ride. We have conducted an offline calibration of the internal parameters and lens distortion of Gopro Hero 3 Black used in dataset Gopro drone 1 from [4] using the Matlab Calibration Toolbox.
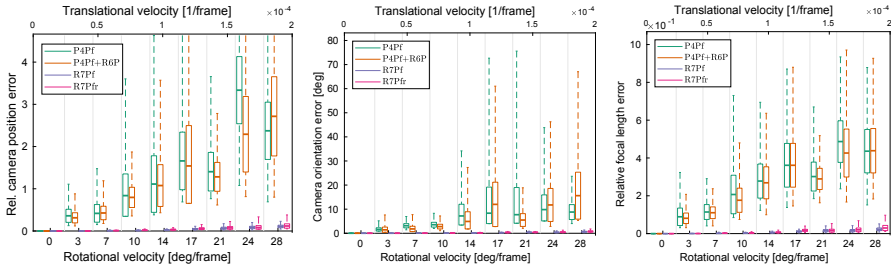
Fig. 2. Increasing camera motion (rotational+translational) on synthetic data with unknown focal length. P4Pf and P4Pf+R6P struggles to estimate the correct pose and focal length in the presence of RS distortions whereas R7Pf (gray-blue) and R7Pfr (magenta) are able to cope with RS effects. (Color figure online)

To create the ground truth we undistorted images grabbed from the entire videos and used them all in an open source SfM pipeline COLMAP [36] to reconstruct a 3D model. Of course, the images containing significant RS distortion were not registered properly or not at all, but the scene was sufficiently reconstructed from the images where RS distortions were insignificant. We then selected the parts of trajectory which have not been reconstructed well and registered the 2D features in those images to the 3D points in the reconstructed scene. This was done for datasets Gopro Drone 2 and Gopro rollercoaster. For Gopro drone 1 we had DSLR images of the scene and we could reconstruct the 3D model using those, which led to much better data overall.

Furthermore, we captured two dataset using the Xiaomi Mi 9 smartphone with both the standard and the wide FOV camera. The standard FOV camera contains virtually no radial distortion whereas the wide FOV camera has moderate radial distortion. We reconstructed the scene using static images from the standard camera and then registered sequences with moving camera to the reconstruction.

5.2   Compared Methods

When neither the focal length nor the radial distortion coefficients are known, the state-of-the-art offers a 4-point solver to absolute pose, focal length (P4Pf) [27] and radial distortion (P4Pfr) [27]. In the presence of RS distortions, one can opt for the R6P algorithm [4] which, however, needs the camera calibration. We solve the problem simultaneously for both the RS parameters, focal length, and radial distortion, which until now could be emulated by running P4Pf or P4Pfr and subsequently R6P on the calibrated and/or undistorted data. That combination is the closest viable alternative to our method, so we consider it as the state-of-the-art and compare with it.

A common practice after robust estimation with a minimal solver and RANSAC is to polish the results with local optimization using all inliers [4]. An important question is whether a simpler model, in our case the baseline
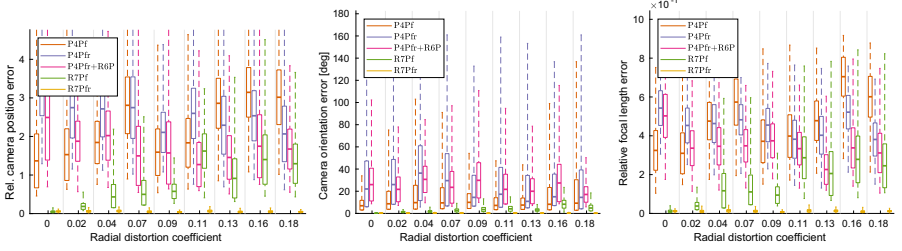
Fig. 3. Increasing radial distortion and unknown focal length. RS motion is kept constant at a value of the middle of Fig. 2. P4Pf and P4Pfr are not able to estimate good pose and focal length under RS distortions, R7Pf slowly deteriorates with increasing radial distortion and R7Pfr performs well across the entire range.

Fig. 4. (left) Original image distorted by radial and RS distortion. (middle) **Image undistorted by our R7Pfr**. (right) Image undistortion by a 3-parametric model estimated by Matlab Calibration Toolbox using a calibration board. We achieve comparable results as a method based on a calibration board.

P4Pf/P4Pfr followed by R6P could be enough to initialize that local optimization and reach the performance of the direct solution with a more complex model, i.e., the proposed R7Pf/R7Pfr. In our experiments we evaluate also the non-linear optimization initialized by RANSAC and see if our solvers outperform the non-linear optimization of the baseline approach.

5.3 Evaluation Metrics

We use various metrics to compare with the state-of-the-art and to show the benefits of our solvers. A common practice [3,5,22] is to use the number of inliers identified by RANSAC as the criterion to demonstrate the performance of minimal solvers on real data. We compare against both the state-of-the-art RANSAC output and the polished result after local optimization.

To highlight the accuracy of the estimated radial distortion and the RS parameters, we use them to remove the radial distortion and the rotational rolling shutter distortion from the images.

Due to the lack of good ground-truth for the camera poses, we evaluate them in two ways. First, we move the camera in place inducing only rotations, which resembles, e.g., an augmented or head-tracking scenario. In this case, the computed camera centers are expected to be almost static and we can show the

Table 1. Average numbers of inliers for different methods.

Dataset	P4Pfr+R6P	P4Pfr+R6P+LO	**P4Pfr+R7Pfr**	**P4Pfr+R7Pfr+LO**
Gopro drone 1	45	162	**170**	**203**
Gopro drone 2	124	130	**126**	**131**
Gopro rollerc	130	**137**	**132**	137
Xiaomi wide	58	66	**64**	67
	P4Pf+R6P	P4Pf+R6P+LO	**P4Pf+R7Pf**	**P4Pfr+R7Pf+LO**
Xiaomi standard	**72**	95	44	**110**

standard deviation from the mean as a measure of the estimated pose error. Second, we evaluate qualitatively the case where the camera moves along a smooth trajectory.

5.4   Results

The experiments on synthetic data verify that the proposed solvers are able to handle unknown focal length, radial distortion and RS distortions. Figure 2 shows results on data with unknown focal length and increasing camera motion. The state-of-the-art P4Pf solver struggles to estimate the camera pose and the focal length accurately as the RS camera rotational and translational velocity increases, resulting in mean orientation errors up to $15°$ and relative focal length error of 40% when the motion is strongest. Given such poor initial focal length estimate from P4Pfr, R6P is not able to recover the pose any better. In contrast, both R7Pf and R7Pfr are able to estimate the pose and focal length accurately, keeping the mean rotation error under $1.0°$ and the relative focal length estimate error under 3% even for the strongest motions.

Next we evaluate the effect of increasing radial distortion and the performance of our R7Pfr solver, see Fig. 3. The magnitude of the RS motion is kept constant through the experiment at the value of about the middle of the previous experiment. First thing to notice is that P4Pfr is less stable under RS distortion, providing worse estimates than P4Pf when radial distortion is close to zero. As the radial distortion increases the performance of P4Pf becomes gradually worse and is outperformed by P4Pfr in the end. R6P initialized by P4Pfr is not able to improve the poor results of P4Pfr. R7Pf slowly deteriorates with increasing radial distortion and R7Pfr provides good results under all conditions.

Additional synthetic experiments, including experiments studying the behaviour of our new solvers for increasing angular errors of the initial orientation, performance on planar scenes, convergence and run-time statistics are in the supplementary material.

The mean number of inliers on real data is summarized in Table 1 and qualitative evaluation of image undistortion is shown in Fig. 5. Camera center precision is evaluated quantitatively in Table 2 and qualitatively in the supplementary material. Our solvers achieve overall better performance in terms of number of

Table 2. Standard deviations from mean position of camera centers. The camera was purely rotating in these datasets; lower deviations mean more precise camera poses.

Dataset	P4Pfr+R6P	P4Pfr+R6P+LO	P4Pfr+R7Pfr	P4Pfr+R7Pfr+LO
Xiaomi wide	25	39	20	20
Xiaomi standard	12	14	14	10

Table 3. Mean relative errors of estimated focal length w.r.t the ground truth focal length $f_{gt} = 800px$ for the Gopro datasets for our method and the baseline and the locally optimized variants (LO).

Dataset	P4Pfr+R6P	P4Pfr+R6P+LO	**P4Pfr+R7Pfr**	**P4Pfr+R7Pfr+LO**
Gopro drone 1	10%	6.5%	**2.37**%	**1.5**%
Gopro rollerc	2.25%	1.75%	**1.75**%	**1.5**%
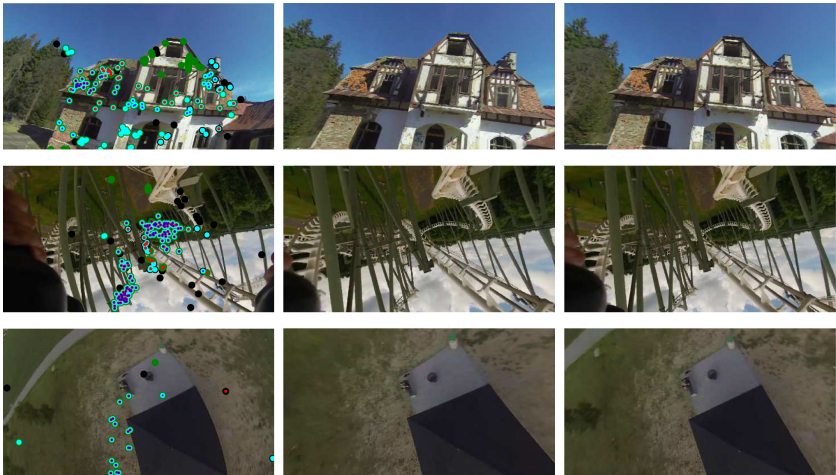
Fig. 5. Right image shows tentative matches (black), inliers found using P4Pfr followed by R7Pfr (cyan) or R6P (blue) and the inliers after subsequent local optimization (LO) of the R7Pfr result (green) and the R6P result (red). The middle image and right image shows the RS and distortion removal using the R6P parameters after LO and the R7Pfr parameters after LO respectively. (Color figure online)

RANSAC inliers and R7Pfr followed by local optimization provides the best results in all cases. The estimated radial distortion and camera motion is significantly better than that of the baseline methods and can be readily used to remove both radial and RS distortion as shown in Figs. 4 and 5. Relative focal length error compared to ground truth available in dataset Gopro drone 1 in Table 3 shows a significant improvement when using R7Pfr solver.

6   Conclusion

We address the problem of absolute pose estimation of an uncalibrated RS camera, and present the first minimal solutions for the problem. Our two new minimal solvers are developed under the same computational scheme by combining an iterative scheme originally designed for calibrated RS cameras with fast generalized eigenvalue and efficient Gröbner basis solvers for specific polynomial equation systems. The R7Pf solver estimates the absolute pose of a RS camera with unknown focal length from 7 point correspondences. The R7Pfr solver estimates the absolute pose of a RS camera with unknown focal length and unknown radial distortion; also from 7 point correspondences. Our experiments demonstrate the accuracy of our new solvers.

References

1. Abdel-Aziz, Y.I., Karara, H.M.: Direct linear transformation from comparator coordinates into object space coordinates in closerange photogrammetry. In: Symposium on CloseRange Photogrammetry (1971)
2. Ait-Aider, O., Andreff, N., Lavest, J.M., Blaise, U., Ferr, P.C., Cnrs, L.U.: Simultaneous object pose and velocity computation using a single view from a rolling shutter camera. In: European Conference on Computer Vision (ECCV), pp. 56–68 (2006)
3. Albl, C., Kukelova, Z., Pajdla, T.: R6P - rolling shutter absolute pose problem. In: Computer Vision and Pattern Recognition (CVPR), pp. 2292–2300 (2015)
4. Albl, C., Kukelova, Z., Larsson, V., Pajdla, T.: Rolling shutter camera absolute pose. In: IEEE Trans, Pattern Analysis and Machine Intelligence (PAMI) (2019)
5. Albl, C., Kukelova, Z., Pajdla, T.: Rolling shutter absolute pose problem with known vertical direction. In: Computer Vision and Pattern Recognition (CVPR) (2016)
6. Ameller, M.A., Triggs, B., Quan, L.: Camera pose revisited: new linear algorithms. In 14eme Congres Francophone de Reconnaissance des Formes et Intelligence Artificielle. Paper in French (2002)
7. Bouguet, J.Y.: Camera calibration toolbox for matlab, vol. 1080 (2008). http://www.vision.caltech.edu/bouguetj/calib_doc
8. Bujnak, M., Kukelova, Z., Pajdla, T.: A general solution to the p4p problem for camera with unknown focal length. In: Computer Vision and Pattern Recognition (CVPR), pp. 1–8. IEEE (2008)
9. Bujnak, M., Kukelova, Z., Pajdla, T.: 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)
10. Cox, D.A., Little, J., O'shea, D.: Using algebraic geometry, vol. 185. Springer, New York (2006). https://doi.org/10.1007/b138611

11. Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H., Bai, Z.: Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide. SIAM, Philadelphia (2000)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**(6), 381–395 (1981)
13. Fitzgibbon, A.W.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: Computer Vision and Pattern Recognition (CVPR) (2001)
14. Grunert, J.A.: Das Pothenotische Problem in erweiterter Gestalt nebst über seine Anwendungen in der Geodäsie (1841)
15. Guo, Y.: A novel solution to the P4P problem for an uncalibrated camera. J. Math. Imaging Vision **45**(2), 186–198 (2013)
16. Haralick, R., Lee, D., Ottenburg, K., Nolle, M.: Analysis and solutions of the three point perspective pose estimation problem. In: Computer Vision and Pattern Recognition (CVPR), pp. 592–598 (1991)
17. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, second edn.. Cambridge University Press, Cambridge (2004)
18. Hedborg, J., Ringaby, E., Forssen, P.E., Felsberg, M.: Structure and motion estimation from rolling shutter video. In: ICCV Workshops, pp. 17–23 (2011)
19. Hedborg, J., Forssén, P.E., Felsberg, M., Ringaby, E.: Rolling shutter bundle adjustment. In: Computer Vision and Pattern Recognition (CVPR). pp. 1434–1441 (2012)
20. Josephson, K., Byröd, M.: Pose estimation with radial distortion and unknown focal length. In: Computer Vision and Pattern Recognition (CVPR) (2009)
21. Kneip, L., Scaramuzza, D., Siegwart, R.: 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)
22. Kukelova, Z., Albl, C., Sugimoto, A., Pajdla, T.: Linear solution to the minimal absolute pose rolling shutter problem. In: Asian Conference on Computer Vision (ACCV) (2018)
23. Kukelova, Z., Bujnak, M., Pajdla, T.: Automatic generator of minimal problem solvers. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5304, pp. 302–315. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-88690-7_23
24. Kukelova, Z., Bujnak, M., Pajdla, T.: Real-time solution to the absolute pose problem with unknown radial distortion and focal length. In: International Conference on Computer Vision (ICCV) (2013)
25. Lao, Y., Ait-Aider, O., Bartoli, A.: Rolling shutter pose and ego-motion estimation using shape-from-template. In: European Conference on Computer Vision (ECCV), pp. 477–492 (2018)
26. Larsson, V., Åström, K., Oskarsson, M.: Efficient solvers for minimal problems by syzygy-based reduction. In: Computer Vision and Pattern Recognition (CVPR) (2017)
27. Larsson, V., Kukelova, Z., Zheng, Y.: Making minimal solvers for absolute pose estimation compact and robust. In: International Conference on Computer Vision (ICCV) (2017)
28. Larsson, V., Kukelova, Z., Zheng, Y.: Camera pose estimation with unknown principal point. In: Computer Vision and Pattern Recognition (CVPR) (2018)
29. Larsson, V., Oskarsson, M., Astrom, K., Wallis, A., Kukelova, Z., Pajdla, T.: Beyond grobner bases: basis selection for minimal solvers. In: Computer Vision and Pattern Recognition (CVPR) (2018)

30. Magerand, L., Bartoli, A., Ait-Aider, O., Pizarro, D.: Global optimization of object pose and motion from a single rolling shutter image with automatic 2d–3d matching. In: European Conference on Computer Vision (ECCV), pp. 456–469 (2012)
31. Meingast, M., Geyer, C., Sastry, S.: Geometric models of rolling-shutter cameras. arXiv cs/0503076 (2005)
32. Microsoft: Spatial Anchors (2020). https://azure.microsoft.com/en-us/services/spatial-anchors/
33. Sattler, T., Leibe, B., Kobbelt, L.: Efficient & effective prioritized matching for large-scale image-based localization. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) **39**(9), 1744–1756 (2017)
34. Saurer, O., Koser, K., Bouguet, J.Y., Pollefeys, M.: Rolling shutter stereo. In: ICCV, pp. 465–472 (2013)
35. Saurer, O., Pollefeys, M., Lee, G.H.: A minimal solution to the rolling shutter pose estimation problem. In: International Conference on Intelligent Robots and Systems (IROS) (2015)
36. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Computer Vision and Pattern Recognition (CVPR) (2016)
37. Snavely, N., Seitz, S.M., Szeliski, R.: Photo tourism: exploring photo collections in 3d. In: ACM Transactions on Graphics (2006)
38. Wu, C.: P3.5p: Pose estimation with unknown focal length. In: Computer Vision and Pattern Recognition (CVPR) (2015)
39. Zheng, Y., Sugimoto, S., Sato, I., Okutomi, M.: A general and simple method for camera pose and focal length determination. In: Computer Vision and Pattern Recognition (CVPR) (2014)