

Relative pose problem for non-overlapping surveillance cameras with known gravity vector

Branislav Micusik
AIT Austrian Institute of Technology
branislav.micusik@ait.ac.at

Abstract

We present a method for estimating the relative pose of two calibrated or uncalibrated non-overlapping surveillance cameras from observing a moving object. We show how to tackle the problem of missing point correspondences heavily required by SfM pipelines and how to go beyond this basic paradigm. We relax the non-linear nature of the problem by accepting two assumptions which surveillance scenarios offer, i.e. the presence of a moving object and easily estimable gravity vector. By those assumptions we cast the problem as a Quadratic Eigenvalue Problem offering an elegant way of treating nonlinear monomials and delivering a quasi closed-form solution as a reliable starting point for a further bundle adjustment. We are the first to bring the closed form solution to such a very practical problem arising in video surveillance. Results in different camera setups demonstrate the feasibility of the approach.

1. Introduction

The Structure from Motion (SfM) problem has been intensively studied in computer vision and it has become one of the problems that has reached a certain maturity. Robustness, accuracy, and scalability allow to employ the algorithms in many large scale applications, e.g. as a pre-step for dense multi-view reconstructions, for visual odometry, arranging hundred thousands images in 3D space, etc.

To come up with a valid solution of camera poses, SfM algorithms and multi-view geometry techniques heavily rely on point correspondences across the field of views [10]. This narrows the use of classical SfM algorithms to cameras with a significant overlap. Surveillance is an example where this assumption is mostly violated. The cameras are in most cases distributed and placed such that an observed area is maximized resulting in a very tiny or no overlap of camera field of views. A typical setup is shown in Fig. 1 where two cameras observe a moving point, e.g. a person, and where no point correspondences can be established.

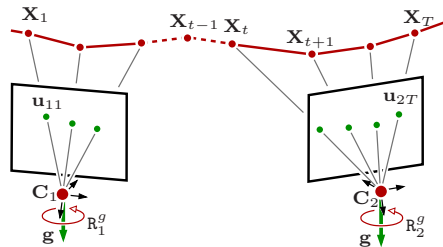


Figure 1. Surveillance camera setup with two non-overlapping cameras observing a moving object. The paper addresses the estimation of all in red plotted elements, i.e. the camera centers C_i , camera rotations R_i^g and 3D points X_t simultaneously, while given in green plotted measurements u_{it} and the gravity vector g .

This paper focuses on such a scenario and introduces a solution for estimating the relative pose of two non-overlapping calibrated or uncalibrated cameras in the case of missing point correspondences. We deliver a quasi¹ closed-form solution just from detecting and tracking a moving object. This brings a novel solution which enriches the group of algorithms for SfM beyond the overlapping field of views and brings a significant contribution to the problem of localizing non-overlapping cameras. Until now there had been no closed-form solution to this problem with unknown relative rotation neither for calibrated nor for uncalibrated case.

We focus on a problem with important practical implications for surveillance and on a problem which has been considered in the SfM community only marginally. With increase of security level at public areas, the surveillance camera networks with slightly or non overlapping field of view cameras have been obtaining more attention. In surveillance community the camera poses are typically assumed a priori known in order to increase the performance of high level visual tasks like detecting and tracking people. To come to the point of knowing the camera poses we propose a method which automatically estimates two camera relative pose and could be further utilized in a large camera network cali-

¹'quasi' while the solution yields SVD which is in fact an iterative method but requiring no starting point and delivering an optimal solution.

bration where standard manual calibration methods become soon infeasible or simply impossible.

The contribution of the paper is the following. We show how to relax the original non-linear ill-posed problem of relative camera pose estimation in the non-overlapping case. We cast the problem as a rectangular Quadratic Eigenvalue Problem (QEP), which solves in a quasi closed-form for parameters appearing in non-linear monomials in solution equations. The closed-form solution for both the calibrated and uncalibrated case is possible thanks to the assumption that a person moves linearly and to a known gravity vector for both cameras. The gravity vector can be automatically estimated either from a vertical vanishing point in the calibrated case or given by an inertial sensor in the uncalibrated case.

2. Related work

A lot of work has been done in the classical SfM for relative and absolute pose problems (PnP), also utilizing the gravity vector [16, 12, 8, 14]. However, all the methods require point correspondences.

The missing point correspondences as a result of non-overlapping camera setup have been tackled in the past by various strategies. Seeing the 3D points as points on a smooth trajectory and enforcing their temporal consistency has appeared in [4]. Linear extrapolation of trajectories into the invisible image that hallucinates the missing correspondences has been proposed for uncalibrated [11] and calibrated cameras [6]. Attempts to use polynomials for extrapolation have recently been made [1]. A related approach [18, 19] estimates camera topology in terms of a graph where the vertices express the cameras and the edges express spatial proximity. Topology contains less information than a geometric camera model, however, topology can be a prior for camera localization [5]. Obtaining the missing correspondences by using planar mirror and their reflections has been tackled in [15]. A non-linear optimization method to estimate poses of non-overlapping cameras from local camera measurements enforcing linear Gaussian Markov dynamics along the trajectory is proposed in [24, 26]. Thanks to special assumptions on cameras, the problem reduces to search for 1D rotation and 2D center of each camera. Not surprisingly, in the general case of cameras that we tackle, there is no convergence due to many local minima of the criterion function and a good initial starting point is necessary.

To get the initial starting point the work [22] as an augmentation of [25] to non-overlapping cameras made a first step, outlined in Sec. 4. The camera rotations and internal calibrations are assumed to be known which is a strong assumption and hence impractical. In this paper we avoid this assumption and deliver a solution for unknown rotations and internal camera calibration, described in Sec. 5.

3. Notation

Let us assume a 3D point \mathbf{X} to be observed by the i th camera at time instance t on its trajectory as \mathbf{u}_{it} , see Fig. 1. For simplicity of notation, we consider only one trajectory, however, the formulation holds for an arbitrary number of trajectories. The standard camera perspective equation [10] reads as

$$\lambda_{it}\mathbf{u}_{it} = K_i[\mathbf{R}_i \ \mathbf{t}_i] (\mathbf{X}_t^\top \ 1)^\top, \quad (1)$$

with calibration matrices K_i , camera rotations \mathbf{R}_i , translations \mathbf{t}_i , 3D points \mathbf{X}_t , scale λ_{it} and known observations \mathbf{u}_{it} . We assume two synchronized cameras observing a linearly moving object with roughly constant velocity.

4. Known rotation

The rotations \mathbf{R}_i and internal camera calibration matrices K_i are assumed *known*. This case has been studied in [25] for overlapping and in [22] for non-overlapping case, known as the Direct Reference Plane (DRP) method. By rearranging Eq. (1) one can write

$$\underbrace{\lambda_{it} \mathbf{R}_i^\top K_i^{-1} \mathbf{u}_{it}}_{\mathbf{d}_{it}} = \mathbf{X}_t - \mathbf{C}_i, \quad (2)$$

with $\mathbf{C}_i = -\mathbf{R}_i^\top \mathbf{t}_i$ as the camera center and \mathbf{d}_{it} as the ray passing through \mathbf{C}_i and the image measurement \mathbf{u}_{it} . Enforcing same directions of both vectors,

$$\|\mathbf{d}_{it} \times (\mathbf{X}_t - \mathbf{C}_i)\|_2 = E_{dat}(i, t) \rightarrow 0, \quad (3)$$

eliminates λ_{it} and defines an algebraic error called the data term. Observing 3D points by at least two cameras, one can estimate their locations and simultaneously localize the cameras by minimizing the error $E_{dat}(i, t)^2$ from Eq. (3) in a closed form [25]. In the non-overlapping case the problem is ill-posed with not enough constraints. However, the smoothness temporal consistency of the 3D points can complement for that to come to a linear problem [22] or to a convex program [21]. The technique in [22], further referred as the DRP method, has been observed to be especially good for linear trajectories rather than for loops [21]. However, the big disadvantage is the necessity of knowing the camera rotations.

5. Unknown rotation

The rotations \mathbf{R}_i are assumed *unknown*. The solution to this problem defines the contribution of this paper. This case has not been solved yet in a closed-form due to the non-linear nature of the problem which the unknown rotations bring. We show how to relax the problem yielding the closed-form for calibrated as well as for uncalibrated cameras.

The basic idea is to split the rotations R_i into two components such that one component can be estimated from the images while the second one is estimated simultaneously with all other unknowns. Let us re-write Eq. (1) as

$$\lambda_{it} \mathbf{u}_{it} = K_i [R_i^v R_i^g \mathbf{t}_i] (\mathbf{X}_t^\top \mathbf{1})^\top, \quad (4)$$

where the rotation R_i is split into two components R_i^v and R_i^g . The rotation R_i^v transforms the world coordinate system having aligned one of its axis with the gravity vector \mathbf{g} to the i th local camera coordinate system. The rotation R_i^g represents the 1D rotation around the gravity vector, see Fig. 1. The rotation R_i^v is given by a gravity vector \mathbf{g} which can easily be estimated from the images of a vertical vanishing point given the camera calibration or from the gravity vector measured by the Inertial Measurement Unit (IMU). Using the gravity vector is a reliable and safe relaxation of the problem allowing to compute the remaining 1 DOF simultaneously with other parameters in a closed form.

The vertical vanishing points are typically easy to be estimated in surveillance applications thanks to walking people which typically stay straight and/or from the vertical lines present in the environments. Assuming all three vanishing points to be known to fix the full rotation, as in [22], is often too restrictive and even impossible to be estimated in non-Manhattan or very cluttered environments.

Alternative to the vertical vanishing point is the gravity vector measured by an IMU. The IMUs are very common nowadays in many devices such as smart cameras and smart phones. They measure full rotation w.r.t. to the earth magnetic field. However, we rely on the gravity vector only, hence measuring the elevation is reliable and accurate everywhere (indoor/outdoor) while measuring azimuth, like a standard compass does, is very sensitive to electro-magnetic noise present in many indoor environments.

5.1. Known focal length

Re-writing Eq. (4) yields

$$\lambda_{it} \underbrace{R_i^v \top K_i^{-1} \mathbf{u}_{it}}_{\mathbf{d}_{it}} = R_i^g \mathbf{X}_t + R_i^v \top \mathbf{t}_i, \quad (5)$$

with the unknowns R_i^g , \mathbf{X}_t , and \mathbf{t}_i . Notice the bilinear relation between the rotation and the 3D point coordinates. The 1D rotation around the gravity vector \mathbf{g} by the angle ϕ reads as

$$R_i^g = \begin{bmatrix} \cos \phi_i & 0 & -\sin \phi_i \\ 0 & 1 & 0 \\ \sin \phi_i & 0 & \cos \phi_i \end{bmatrix}. \quad (6)$$

We set $q_i = \tan \frac{\phi_i}{2}$ which gives $\cos \phi_i = \frac{1-q_i^2}{1+q_i^2}$ and $\sin \phi_i = \frac{2q_i}{1+q_i^2}$ [14]. Substituting that into the rotation R_i^g yields

$$(1+q_i^2)R_i^g = \begin{bmatrix} 1-q_i^2 & 0 & -2q_i \\ 0 & 1+q_i^2 & 0 \\ 2q_i & 0 & 1-q_i^2 \end{bmatrix}. \quad (7)$$

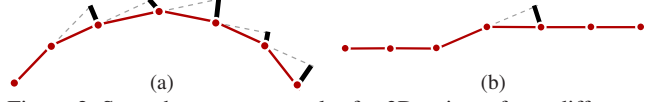


Figure 2. Smoothness term penalty for 3D points of two different trajectories. The lengths of thick black lines correspond to the cost paid for each triplet of the 3D points.

Eliminating λ_{it} in Eq. (5) in the similar manner as in Eq. (3) yields

$$\mathbf{d}_{it} \times \left((1+q_i^2)R_i^g \mathbf{X}_t + \underbrace{(1+q_i^2)R_i^v \top \mathbf{t}_i}_{\mathbf{p}_i} \right) = 0, \quad (8)$$

which expands to

$$\begin{bmatrix} d_{itx} \\ d_{ity} \\ d_{itz} \end{bmatrix} \times \left(\begin{bmatrix} 1-q_i^2 & 0 & -2q_i \\ 0 & 1+q_i^2 & 0 \\ 2q_i & 0 & 1-q_i^2 \end{bmatrix} \mathbf{X}_t + \begin{pmatrix} p_{ix} \\ p_{iy} \\ p_{iz} \end{pmatrix} \right) = 0. \quad (9)$$

We set the world coordinate system into the first camera, then $q_1 = 0$, $R_1^g = \mathbf{I}_{3 \times 3}$, we choose $\mathbf{X}_1 = \mathbf{0}$, and $\mathbf{p}_1 = (p_x \ p_y \ 1)^\top$ for fixing the scale and re-denote $q_2 = q$. Then the first and third linearly independent equations from Eq. (9) read as

$$\begin{aligned} -d_{itz}(p_{iy} + X_{ty} + q^2 X_{ty}) + d_{ity}(p_{iz} + X_{tz} + 2qX_{tx} - q^2 X_{tz}) &= 0 \\ d_{itx}(p_{iy} + X_{ty} + q^2 X_{ty}) - d_{ity}(p_{ix} + X_{tx} - 2qX_{tz} - q^2 X_{tx}) &= 0 \end{aligned} \quad (10)$$

Remark that all parameters except for the image measurements $d_{it\{x,y,z\}}$ are unknown and subject to be estimated. Notice the difficulty of the problem since there are some monomials, e.g. $2qX_{tx}$, $q^2 X_{tz}$, composed of a product of two unknowns. To solve for all the unknowns we propose the following.

Observing the structure of the equations in Eq. (10) we see that it can be re-written in the matrix form as

$$(D_1 + D_2 q + D_3 q^2) \mathbf{x} = \mathbf{0}, \quad (11)$$

where

$$\mathbf{x} = (X_{2x}, X_{2y}, X_{2z} \dots X_{Ty}, X_{Tz}, p_{1x}, p_{1y}, 1, p_{2x}, p_{2y}, p_{2z})^\top \quad (12)$$

is a vector of unknowns and D_1 , D_2 , D_3 are known rectangular matrices. Eq. (11) represents a Quadratic Eigenvalue Problem (QEP), well studied in mathematics. The relative pose problem for overlapping cameras requiring point correspondences has already been formulated as QEP, e.g. epipolar geometry with unknown radial distortion [7] or unknown omnidirectional camera model [20], 5pt and 6pt relative pose problem [13], essential matrix estimation with a single unknown focal length [3]. Here we show that non-overlapping case yields QEP as well.

There are $3(T-1) + 6$ unknown parameters in total and 2 equations for each visible point from Eq. (10) hence $2T$ equations. Clearly, in non-overlapping case the system is under-constrained. We must add additional equations, hence we enforce zero second derivative along the trajectory as

$$\|\mathbf{X}_{t-1} - 2\mathbf{X}_t + \mathbf{X}_{t+1}\|_2 = E_{sm}(t) \rightarrow 0, \quad (13)$$

see Fig. 2 for its visualization. Minimization of $E_{sm}(t)$ within Eq. (11) is enforced by splitting the Eq. (13) into three parts, added as three separate rows into the matrix D_1 , e.g. for the x -coordinate,

$$\gamma(X_{(t-1)x} - 2X_{tx} + X_{(t+1)x}) = \gamma E_{sm}(t)_x \rightarrow 0. \quad (14)$$

Since Eq. (11) is being solved subject to $\|\mathbf{x}\|^2 = 1$, in the end, a squared norm $E_{sm}^2(t) = E_{sm}^2(t)_x + E_{sm}^2(t)_y + E_{sm}^2(t)_z$ is minimized as required by Eq. (13). The smoothness term brings us $3T - 6$ equations. It implies that minimal case is when 5 points are visible in total, but at least 2 points in each view. Constructing the matrices $D_{\{1,2,3\}}$ from image measurements and solving Eq. (11) gives us all the unknowns and $\mathbf{t}_i = \mathbf{R}_i^v \mathbf{p}_i / (1 + q_i^2)$. Solving the rectangular QEP is discussed later in Sec. 5.3. The parameter γ weights the smoothness term.

Note that the solution of the problem does not change if there is an overlap and we have point correspondences. They simply deliver more constraints yielding thus much better estimate.

5.2. Unknown focal length

Re-writing Eq. (4) in a different way than in the previous section yields

$$\lambda_{it} K_i^{-1} \mathbf{u}_{it} = \mathbf{R}_i^v \mathbf{R}_i^g \mathbf{X}_t + \mathbf{t}_i. \quad (15)$$

Let us assume the calibration matrices K_i to have square pixels with known principal points which shift image measurements \mathbf{u}_{it} to \mathbf{d}_{it} . The calibration matrices K_i become then $K_i = \text{diag}(f_i, f_i, 1)$. Eliminating the scale by the cross product as in Eq. (9) and by multiplication of the vector \mathbf{d}_{it} by f_i yields

$$f_i K_i^{-1} \mathbf{d}_{it} \times \left(\mathbf{R}_i^v (1 + q_i^2) \mathbf{R}_i^g \mathbf{X}_t + (1 + q_i^2) \mathbf{t}_i \right) = 0. \quad (16)$$

Substituting the 1D rotation \mathbf{R}_i^g parametrized in the same way as in Eq. (7) gives

$$\begin{bmatrix} d_{itx} \\ d_{ity} \\ f_i \end{bmatrix} \times \left(\mathbf{R}_i^v \begin{bmatrix} 1 - q_i^2 & 0 & -2q_i \\ 0 & 1 + q_i^2 & 0 \\ 2q_i & 0 & 1 - q_i^2 \end{bmatrix} \mathbf{X}_t + (1 + q_i^2) \begin{pmatrix} t_{ix} \\ t_{iy} \\ t_{iz} \end{pmatrix} \right) = 0. \quad (17)$$

There are three equations, two of them are linearly independent. The first and third equations are

$$\begin{aligned} -f_i d_{itz} & \left[(R_{i21}(1 - q_i^2) + 2R_{i23}q_i)X_{tx} + (R_{i22}(1 + q_i^2))X_{ty} + \right. \\ & \left. + (R_{i23}(1 - q_i^2) - 2R_{i21}q_i)X_{tz} + (1 + q_i^2)t_{iy} \right] + \\ & + d_{ity} \left[(R_{i31}(1 - q_i^2) + 2R_{i33}q_i)X_{tx} + (R_{i32}(1 + q_i^2))X_{ty} + \right. \\ & \left. + (R_{i33}(1 - q_i^2) - 2R_{i31}q_i)X_{tz} + (1 + q_i^2)t_{iz} \right] = 0 \end{aligned} \quad (18)$$

and

$$\begin{aligned} -d_{ity} & \left[(R_{i11}(1 - q_i^2) + 2R_{i13}q_i)X_{tx} + (R_{i12}(1 + q_i^2))X_{ty} + \right. \\ & \left. + (R_{i13}(1 - q_i^2) - 2R_{i11}q_i)X_{tz} + (1 + q_i^2)t_{ix} \right] + \\ & + d_{itx} \left[(R_{i21}(1 - q_i^2) + 2R_{i23}q_i)X_{tx} + (R_{i22}(1 + q_i^2))X_{ty} + \right. \\ & \left. + (R_{i23}(1 - q_i^2) - 2R_{i21}q_i)X_{tz} + (1 + q_i^2)t_{iy} \right] = 0. \end{aligned} \quad (19)$$

Note that Eq. (19) has interesting structure and there is no focal length. Let us set $\mathbf{X}_1 = \mathbf{0}$, $q_1 = 0$, $q_2 = q$. We can again re-write it into the form of QEP in Eq. (11) but with

$$\mathbf{x} = (X_{2x}, X_{2z} \dots X_{Tx}, X_{Tz}, t_{1x}, t_{1y}, t_{2x}, t_{2y})^\top. \quad (20)$$

Moreover, the smoothness term enforced on the trajectory is in the same way encoded into the matrices $D_{\{1,2,3\}}$ as in the previous section in Eq. (14).

Unlike the case with known focal length there are still too many unknowns to come up with a solution. It is very important to emphasize that without additional constraints the QEP in Eq. (11) is ill-posed and not solvable. It is reflected by rank deficiency of the matrix D_1 . We propose therefore to bring the missing constraint by enforcing all the points in one track to have the y -coordinates X_{ty} equal. Remark that our world coordinate system is aligned such that its y coordinate is co-linear with the gravity vector hence the points within one track have equal this coordinate. Enforcing the equal y -coordinate is a safe assumption in surveillance since we rely on observing and tracking people and they normally move on the plane. For more than one trajectory, all are required to lie in parallel planes. So every trajectory except for the first one (there $X_y = 0$) would have one extra parameter X_y for all the points within the same track.

Each 3D point gives us one Eq. (19), so T equations, and the smoothness brings us $2T - 4$ equations. Since there are $2(T - 1) + 5$ unknowns in Eq. (19), the minimal configuration is 7 points to be seen, at least 2 and 3 points in the first and the second camera, respectively. Having solutions for q and unknowns in Eq. (20), the rest unknown parameters, i.e. f_i, p_{iz} , can be estimated from Eq. (18) in a linear manner as a right null space of the design matrix A_i composed of already known parameters and image observations, i.e. solving the linear system $A_i(f_i, p_{iz}, 1)^\top = \mathbf{0}$.

5.3. Rectangular QEP

When the design matrices $D_{\{1,2,3\}}$ in Eq. (11) are square, they yield a square QEP and there are many solvers available, e.g. `polyeig` in MATLAB. Alternatively the QEP can be easily converted to a generalized eigenvalue problem (GEP) and for which even more solvers exist. The QEP as an interesting solvable non-linear algebraic problem which has appeared in many scientific disciplines and many of such problems are summarized in [9].

If the matrices are rectangular, Eq. (11) can be squared by left multiplication by D_1^\top , as suggested by [7]. This trick has virtue of preserving the true solution in the noiseless case. However, such a solution suffers from a significant bias and variance in the presence of noise. We follow the work of [2] and to find a global optimum for q we suggest the following solution. First, we transform the QEP in Eq. (11) to the GEP to the form

$$(\mathbf{E} - \mathbf{F}q)\tilde{\mathbf{x}} = \mathbf{0}, \quad (21)$$

to be consistent with [2], where \mathbf{E} and \mathbf{F} are rectangular matrices, composed of matrices $D_{\{1,2,3\}}$ and $\tilde{\mathbf{x}} = (\mathbf{x}^\top, q \mathbf{x}^\top)^\top$, see *e.g.* [7].

The iterative technique in [2] is based on perturbation of the rectangular matrices. They start from the solution given by solving first the GEP: $(\mathbf{F}^\top \mathbf{E} - \mathbf{F}^\top \mathbf{F}q)\tilde{\mathbf{x}} = \mathbf{0}$. In noisy situations this is not suitable and often far from optimal solution. To cope with local minima and to find a global solution we take advantage of knowing that q is bounded since θ is between $[-\pi, \pi]$ deg. We design the following algorithm:

1. Discretize the space of q such that $q_j = \tan(\frac{-\pi + \Delta: \pi}{2})$
2. Compute the smallest singular value of the rectangular matrix $[\mathbf{E} - \mathbf{F}q_j]$ as a solution vector $\tilde{\mathbf{x}}_j$.
3. Evaluate the following residuum [2]

$$\epsilon_j = \|(\mathbf{E} - \mathbf{F}q_j)\tilde{\mathbf{x}}_j\|_2^2 / (1 + q_j^2) \quad (22)$$

4. Find local minima of ϵ_j , keep them as $q_l, \tilde{\mathbf{x}}_l$.
5. For each $q_l, \tilde{\mathbf{x}}_l$ as a possible solution compute all the rest parameters and evaluate the smoothness cost E_{sm} in Eq. (13) and the re-projection error E_{rep} with L_2 norm. Choose only one solution having the sum $E_{rep} + \gamma E_{sm}$ minimal while fulfilling the cheirality constraint that all the measured points are in front of the cameras which see them.
6. For the selected solution $q, \tilde{\mathbf{x}}$ run a simple local minimization analogously to gradient descent, *i.e.* iterate steps 2, 3 by changing q by adding or subtracting Δ which adaptively changes over the runtime depending on decreasing the residuum.

The value Δ was set in all our experiments to 10deg. Typically the final local minimization in the step 6 takes about 30 iterations. Solving the QEP reduces to null space computation of a rectangular matrix which can be done efficiently by SVD. Since the matrices $D_{\{1,2,3\}}$ are sparse and so \mathbf{E} and \mathbf{F} , fast sparse solvers for SVD can be utilized for significant speeding up. If there is no local minimum after the step 4, it indicates no solution. The reason is that the noise level in image observations \mathbf{u}_{it} is too high or the true trajectory is far from being straight.

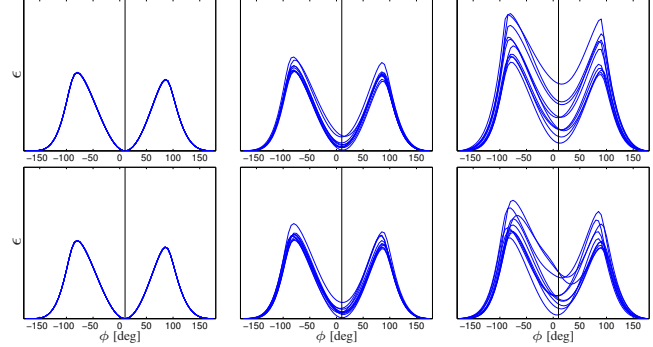


Figure 3. Noise influence on residuum ϵ , Eq. (22), for solving over-constrained rectangular QEP with known (*top row*) and unknown focal length (*bottom row*). Ground truth q is marked by the black vertical line. Two type of noise was added; the noise rotating a gravity vector with std of 0.5 deg and the noise on image measurements with std of 0 pxl (*left*), 1 pxl (*middle*), 2 pxl (*right*). Ten runs are shown to see better the deviations.

We did investigate the shape and noise sensitivity of the residuum function from Eq. (22) in Fig. 3 of the SETUP B shown in Fig. 4. Positive is that the function is multi-modal, typically with one minimum. Noise does not change much the shape of the function, but it can shift the minimum and can sometimes introduce other minima. To find the global minimum we therefore propose, first, to localize all the local minima by the strategy of equal sampling in the step 1. Second, to check the cheirality constraint, and third, to run the local optimizer for the selected minimum. Notice that by the QEP we minimize the algebraic error, but for selection the final solution we evaluate the re-projection error as statistically optimal criterion. Not surprisingly, the calibrated case is less sensitive to noise than the uncalibrated one.

5.4. Degenerate cases

The first problem is in parametrization of the 1D rotation matrix in Eq. (7) by $q = \tan \frac{\phi}{2}$. It introduces a degenerate case when $\phi = \pi$ as tangent goes to infinity. This would correspond to the case when two cameras are mounted on opposite walls observing the same part of the scene. It can cause numerical instability hence no feasible solution. This can be easily resolved by squeezing a rotation matrix $\tilde{\mathbf{R}}$ for the second camera between \mathbf{R}_2^v and \mathbf{R}_2^g in Eq. (4) as $\mathbf{R}_2^v \tilde{\mathbf{R}} \mathbf{R}_2^g$. The matrix $\tilde{\mathbf{R}}$ has the same form as \mathbf{R}_i^g where the angle ϕ is set as least to three different values. Running the algorithm three times independently with $\mathbf{R}_2^g := \tilde{\mathbf{R}} \mathbf{R}_2^g$, at least two times, the results must be identical up to numerical accuracy and thus correspond to a valid solution.

The second issue is a potential rank deficiency of D_1 in the case of the unknown focal length, Sec. 5.2, when the cameras have optical axis perpendicular to the gravity vector. The matrix elements R_{13}, R_{23} are 0 which cancels the constraints on X_{tz} in Eq. (19) turning the QEP unsolvable. This is very rare situation in surveillance as the cameras are

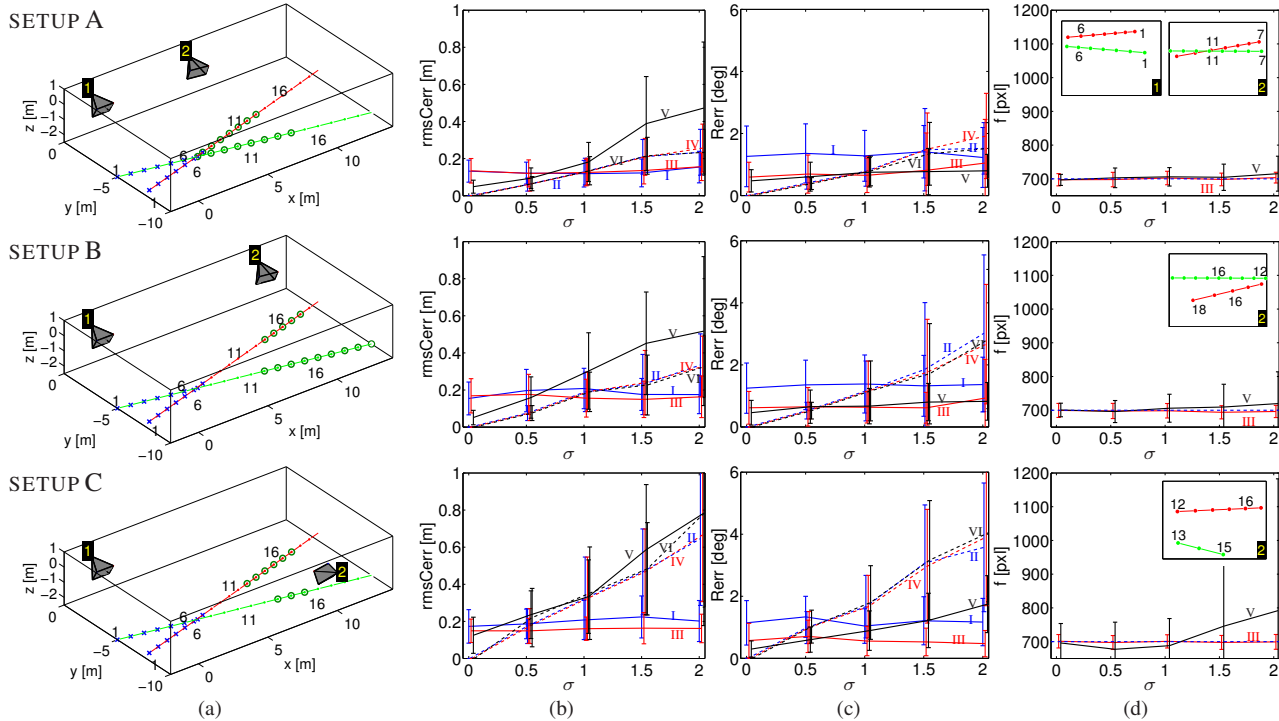


Figure 4. Synthetic experiment. (a) Setup configuration. Error in (b) camera center, (c) rotation, and (d) focal length estimation. Camera views are in-plotted in (d). I -DRP (blue), III - DRPupR (red), V -DRPupRf (black); II, IV, VI - after BA.

mostly mounted near ceilings looking down, observing the ground plane, and therefore a minor issue.

6. Experiments

6.1. Synthetic data

We simulated three different setups typically present in the video surveillance, see Fig. 4 a. *SETUP A* has a slight overlap and three points are in correspondence near the image boundaries. *SETUP B* and *SETUP C* are fully non-overlapping cases. We added Gaussian noise to image measurements with standard deviation σ and performed 100 trials at each σ value. To simulate real situations we extra added at each iteration noise to focal length (2%), rotation of the gravity vector in random direction (0.5deg) and rotation around gravity vector (1.5deg). It reflects the fact that we typically measure these entities with such order of accuracy. Two different amounts of noise to rotation come from the fact that measuring the gravity vector is more accurate, typically 3 times than measuring the earth north pole direction.

We evaluated three different methods. The method [22] denoted as DRP, the proposed method with known camera focal lengths, DRPupR, and unknown focal lengths, DRPupRf. For each method we run the bundle adjustment (BA) hence six results in total. In Fig. 4 we show the statistics for estimation of camera center, rotation and focal length of

the second camera. The error in rotation is defined as the minimal angle needed to rotate the camera to the ground truth orientation. The gauge freedom between the estimate and the ground truth camera positions is solved by aligning coordinate system with the first camera and enforcing the x -coordinate of the second camera center to be equal to the ground truth value.

Let us summarize the results from Fig. 4. The DRP method assumes known camera rotations and focal lengths which values are typically known up to some accuracy. At some point it is better to estimate them as we see from Fig. 4 b,c that DRPupR estimating rotation given the focal length outperforms the DRP. Estimating focal length is very difficult problem, especially when it comes to robustness. However despite that the DRPupRf estimating rotation and the focal length outperforms the DRP and also DRPupR until some σ values. The BA decreases rapidly the re-projection error, however, resulting in the increase of the error in the center and rotation estimation when image noise is to high. Generally we see that using BA initiated by the DRP, DRPupR, and DRPupRf methods gives almost identical results. That underlines the feasibility of our methods as we are comparable to the state-of-the-art DRP method while assuming much less information. From practical point of view, knowing at least the focal length is the best and the most stable case.

We use two non-collinear trajectories allowing the BA to

be effective. If one trajectory were used the BA optimizing over full camera rotations and translations would be ambiguous. One trajectory allows the BA to arbitrarily rotate the cameras around the linear trajectory without changing the re-projection error. The second trajectory fixes this freedom. Since there are no point correspondences the standard bundle adjustment cannot be used. We therefore modified BA [17] relying on point correspondences in such a way that instead of re-projection error alone the weighted sum of it and the smoothness term from Eq. (13) is minimized [23].

6.2. Real data

For better comparison to DRP we evaluated the proposed methods on the same image sequence as used in [22]. The 3 minute long footage contains a walking person in front of two slightly overlapping cameras mounted against each other similar to SETUP C. There is no single detection simultaneously seen by both cameras thus it represents a non-overlapping case.

Since our methods assume linear trajectories we manually selected 10 roughly linear subtrajectories from the entire sequence. Fig. 5 shows images with head-shoulder detections and all the ten tracks in both views. The accuracy of the Adaboost based detector is rather poor as seen in the zoomed figures. Notice that the centers of the tracked region vary within the bound of 20 pixels which makes the problem challenging.

Since we assume only linear trajectory we can benefit from that and simply detect and remove outliers and correct all the measurements in advance. We fit a line on the detections in each track and project all the points back into the line. We fit a 2x2 homography on the projected points in RANSAC style and shift all the projections accordingly. This pre-step gives us perfectly linear trajectories with geometrically validated projections. Since this pre-step is always possible to be done it has the following practical outcome. A person is being detected, tracked and if the detections can be approximated by a line, only then they are considered and fed into our method. It is a great advantage to standard SfM where we typically have no prior information about 3D point locations and outliers are pruned only when solving a relative or absolute pose problem.

We consider the sequence extremely hard in sense that no matter how far the person from the cameras is all the detections are almost co-linear and close to the horizon, see Fig. 5b,d. This is due to the fact that cameras are mounted at the same height as the walking person. As noted in Sec. 5.4 this is very close to a degenerate configuration for DRPupRf method. In practical video surveillance this never happens and therefore we consider this experiment together with very noisy detections as extremely challenging and giving us an upper bound of inaccuracy.

We evaluated the same methods as in the synthetic case.

err	C [m]	R [deg]	f_1 [%]	f_2 [%]
DRP (GT)	0.06 / 0.05	0.0 / 0.1	-	-
(IMU)	0.43 / 0.43	7.5 / 7.5	-	-
DRPupR (GT)	0.20 / 0.20	2.7 / 2.7	-	-
(IMU)	0.46 / 0.45	7.8 / 7.7	-	-
(VP)	0.39 / 0.38	5.3 / 5.3	-	-
DRPupRf (GT)	0.58 / 0.58	8.3 / 8.3	9	27
(IMU)	1.88 / 1.86	9.2 / 9.1	98	270
(VP)	1.14 / 1.13	6.5 / 6.4	30	20

Table 1. Real data. Comparison of three methods using ground truth (GT) rotation, rotation from IMU and vanishing point (VP). The numbers after slashes are the results after BA.

As inputs we use the approximated linear trajectories, however, for the following BA we switched back and use the original noisy data as shown in Fig. 5. The results are summarized in Tab. 1. To make the BA effective in each run we considered 4 trajectories simultaneously, hence 7 runs in total. Median values for error in camera center, rotation and focal lengths are reported.

The results show a great practical use of the proposed methods, especially the calibrated case DRPupR. We show the results when using the ground truth (GT) rotation, IMU measurement and vertical vanishing point. The error of 0.38m in center estimation and 5deg in rotation for DRPupR method at the camera distance of 10m is highly acceptable for the video surveillance. DRPupRf is a bit worse than DRPupR, but still delivers a reasonable estimate. The sensitivity of the focal length estimate is a direct consequence of the already mentioned almost degenerate case. It turns out that using visual information captured by the vertical vanishing point (VP) gives much better result than the hardware solution by a common IMU (even if using only gravity vector). The DRPupR using vanishing point is even better than DRP with the IMU. There is still a room for improvement, *e.g.* using more accurate detections and because of quite evident radial distortion either correcting the input images for it or incorporating it into the estimation process similar to [14].

7. Conclusion

We have moved forward the SfM by delivering a quasi closed-form solution for calibrated as well as uncalibrated relative pose problem for non-overlapping cameras observing a moving person. Until now there had been only a solution assuming known rotation and calibrated cameras. We show how to relax the original non-linear ill-posed problem with unknown rotations and focal length, yielding an elegant solution via the Quadratic Eigenvalue Problem. We have proposed fully automatic method with accuracy which is sufficient for the video surveillance despite noisy inputs. The method is suited for long time running as it can automatically select appropriate linear trajectories, estimate the unknown parameters, and integrate the results over time un-

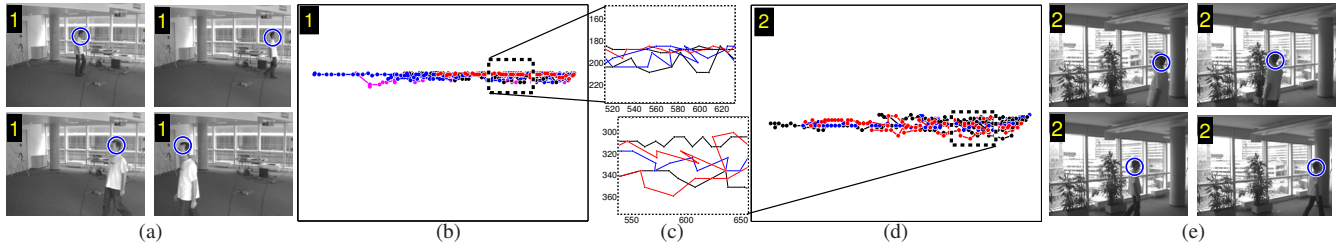


Figure 5. Real data. (a), (e) Input images with automatically established head-shoulder detections. (b), (d) Ten tracks seen by the cameras with (c) detailed views on some of them.

til a certain peakness in parameter voting space is reached.

The proposed method also offers a solution to the classical relative pose problem with point correspondences and known gravity vector. In such a case 3 point correspondences are needed for calibrated case and no requirement on linear trajectory is enforced.

Acknowledgement: I thank Roman Pflugfelder and Cristina Picus for their useful feedback and discussion. We gratefully acknowledge funding from Wiener Wissenschafts-, Forschungs- und Technologiefonds - WWTF, Project No ICT08-030.

References

- [1] N. Anjum, M. Taj, and A. Cavallaro. Relative position estimation of non-overlapping cameras. In *IEEE Intern. Conf. on Acoustics, Speech and Signal Processing*, 2007.
- [2] G. Boutry, M. Elad, G. H. Golub, P. Milanfar, and G. H. G. P. Milanfar. The generalized eigenvalue problem for non-square pencils using a minimal perturbation approach. *SIAM J. Matrix Anal. Appl.*, 27:582–601, 2005.
- [3] M. Bujnak, Z. Kukelova, and T. Pajdla. 3D reconstruction from image collections with a single known focal length. In *ICCV*, 2009.
- [4] Y. Caspi and M. Irani. Aligning non-overlapping sequences. *IJCV*, 48(1):39–51, 2002.
- [5] D. Devarajan, Z. Cheng, and R. Radke. Calibrating distributed camera networks. *Proceedings of the IEEE*, 96(10):1625–1639, 2008.
- [6] R. B. Fisher. Self-organization of randomly placed sensors. In *ECCV*, pages IV: 146–160. Springer-Verlag, 2002.
- [7] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *CVPR*, 2001.
- [8] F. Fraundorfer, P. Tanskanen, and M. Pollefeys. A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In *ECCV*, 2010.
- [9] L. Grammont, N. J. Higham, and F. Tisseur. A framework for analyzing nonlinear eigenproblems and parametrized linear systems. *Linear Algebra and its Applications*, In Press, Corrected Proof, 2010.
- [10] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.
- [11] O. Javed, Z. Rasheed, O. Alatas, and M. Shah. M-knight: A real time surveillance system for multiple and non-overlapping cameras. In *IEEE Intern. conf. on Multimedia and Expo*, 2003.
- [12] M. Kalantari, A. Hashemi, F. Jung, and Guédon. A new solution to the relative orientation problem using only 3 points and the vertical direction. *Computer Research Repository (CoRR)*, abs/0905.3964, 2009.
- [13] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *BMVC*, 2008.
- [14] Z. Kukelova, M. Bujnak, and T. Pajdla. Closed-form solutions to the minimal absolute pose problems with known vertical direction. In *ACCV*, 2010.
- [15] R. Kumar, A. Ilie, J.-M. Frahm, and M. Pollefeys. Simple calibration of non-overlapping cameras with a mirror. In *CVPR*, 2008.
- [16] J. Lobo and J. Dias. Vision and inertial sensor cooperation using gravity as a vertical reference. *PAMI*, 25(12), 2003.
- [17] M. A. Lourakis and A. Argyros. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Software*, 36(1):1–30, 2009.
- [18] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *CVPR*, pages 205–210, 2004.
- [19] D. Marinakis and G. Dudek. Self-calibration of a vision-based sensor network. *Image and Vision Computing*, 27(1-2):116–130, 2009.
- [20] B. Micusik and T. Pajdla. Structure from motion with wide circular field of view cameras. *PAMI*, 28(7), 2006.
- [21] B. Micusik and R. Pflugfelder. Localizing non-overlapping surveillance cameras under the L-inf norm. In *CVPR*, 2010.
- [22] R. Pflugfelder and H. Bischof. Localization and trajectory reconstruction in surveillance cameras with non-overlapping views. *PAMI*, 2009.
- [23] C. Picus, B. Micusik, and R. Pflugfelder. From single cameras to the camera network: an auto-calibration framework for surveillance. In *DAGM*, 2010.
- [24] A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous calibration and tracking with a network of non-overlapping sensors. In *CVPR*, pages (1):187–194, 2004.
- [25] C. Rother and S. Carlsson. Multi view reconstruction and camera recovery. In *ICCV*, 2001.
- [26] M. Rudoy and C. E. Rohrs. Simultaneous sensor calibration and path estimation. In *IEEE Asilomar Conf. on Signals, Systems, and Computers*, 2006.