3D reconstruction from image collections with a single known focal length

Martin Bujnak
Bzovicka 24, 85107 Bratislava
Slovakia
martinb@dataexpert.sk

Zuzana Kukelova
CMP, Czech Technical
University in Prague
kukelova@cmp.felk.cvut.cz

Tomas Pajdla
CMP, Czech Technical
University in Prague
pajdla@cmp.felk.cvut.cz

Abstract

In this paper we aim at reconstructing 3D scenes from images with unknown focal lengths downloaded from photo-sharing websites such as Flickr. First we provide a minimal solution to finding the relative pose between a completely calibrated camera and a camera with an unknown focal length given six point correspondences. We show that this problem has up to nine solutions in general and present two efficient solvers to the problem. They are based on Gröbner basis, resp. on generalized eigenvalues, computation. We demonstrate by experiments with synthetic and real data that both solvers are correct, fast, numerically stable and work well even in some situations when the classical 6-point algorithm fails, e.g. when optical axes of the cameras are parallel or intersecting. Based on this solution we present a new efficient method for large-scale structure from motion from unordered data sets download from the Internet. We show that this method can be effectively used to reconstruct 3D scenes from collection of images with very few (in principle single) images with known focal lengths [1] .

1. Introduction

Estimating relative camera pose [13] from image correspondences is an important computer vision problem. Although it is an old well studied problem [18, 9], new efficient solutions for different camera configuration appeared recently: the 5-pt relative pose problem [26, 30] for a pair of calibrated cameras, the 6-pt focal length problem [28] for camera pair with unknown but constant focal length, the 6-pt generalized camera problem [29], the 9-pt problem for estimating para-catadioptric fundamental matrices [12], the minimal problems of estimating epipolar geometry and a radial distortion parameter [19, 6].

The main application of these problems is in 3D reconstruction. Existing structure from motion (SfM) pipelines e.g. [27], are mainly based on the 5-pt algorithm [26, 30]. They are reliable and robust but they need completely calibrated cameras. Information about the calibration, i.e. the focal length, is in these pipelines usually extracted from the *jpeg-exif* headers. Unfortunately many images downloaded from photo-sharing websites do not contain the *jpeg-exif* header, the header is corrupted or the included focal length is not correct, e.g. due to the image cropping.

Figure 1. A 3D reconstruction of the Fountain di Trevi.

The 6-pt algorithm [28], which in principle could exploit cameras with unknown constant focal length, is rarely used in reality due to problems with degeneracies, e.g. when optical axes of cameras are parallel or intersecting.

In this paper we provide an efficient and robust floating point solution to the configuration with one completely calibrated camera and one camera with an unknown focal length. We show that this solution can cope with most unpleasant degeneracies and can be effectively used to reconstruct 3D scenes from collections of images with very few (in principle single) images with known focal lengths.

Although this problem looks inferior to the well known 6-pt problem [28] for two cameras with unknown but equal focal length, it has several nice and useful properties which standard 6-pt algorithm does not have. The most interesting is its resistance to several critical motions which are common in real situations, e.g. when optical axes are parallel

[1]This work has been supported by EC project FP7-SPACE-218814 PRoVisG and by Czech Government under the research program MSM6840770038.

or intersecting. These configurations are important since they appear frequently when moving around an object and taking its pictures or taking pictures while walking or from a moving car. Although these configurations are degenerate for the standard 6-pt problem setting [28] they become tractable when one of the two cameras is fully calibrated.

This problem of finding relative pose between a completely calibrated camera and a camera with unknown focal length was previously studied in [31] where a non-minimal solution was proposed. After using the 7-pt algorithm for computing the fundamental matrix, the unknown focal length was estimated in a closed-form solution using Kruppa equations.

Here we provide two new minimal solvers for this problem from six point correspondences and compare them with the existing non-minimal solution [31]. Compared to [31], our minimal solution has two advantages. 1) It needs 6 instead of 7 points, which is important for RANSAC, 2) it is more accurate in presence of noise. Our solvers are based on the Gröbner basis, resp. on the generalized eigenvalues computation. Computational complexity of these algorithms is smaller than for the 6-pt problem [28].

We show that these algorithms are useful and practical when combining calibrated images, e.g. taken by a known camera, with images from the Internet. Based on our algorithms we propose a new efficient method for large-scale SfM from unordered data sets download from the Internet e.g. from the Flickr database [11], see Figure 1.

2. Problem formulation

Consider a camera pair where the first camera is calibrated up to an unknown focal length and the second camera is completely calibrated. The constraints on corresponding image points can be written down as [13]

$$\mathbf{x}_j'^{\top} \mathtt{F}\, \mathbf{x}_j = 0, \tag{1}$$

where $\mathtt{F}$ is a 3×3 rank-2 fundamental matrix satisfying

$$\det(\mathtt{F}) = 0. \tag{2}$$

Since the first camera is calibrated up to an unknown focal length and the second camera is fully calibrated, the essential matrix [13]

$$\mathtt{E} \;=\; \mathtt{FK}, \tag{3}$$

where $\mathtt{K} \simeq diag([f\ f\ 1])$ is a diagonal calibration matrix of the first camera, containing the unknown focal length f. It is known [13] that:

$$2\,\mathtt{E}\,\mathtt{E}^{\top}\mathtt{E} - trace(\mathtt{E}\,\mathtt{E}^{\top})\,\mathtt{E} \;=\; 0. \tag{4}$$

The standard way [28] of computing the fundamental matrix $\mathtt{F}$ starts with rewriting (1) as $\mathtt{M}\,X = 0$, where the vector X contains nine elements of $\mathtt{F}$ and $\mathtt{M}$ contains image measurements from 6 image matches. Next, a three dimensional basis $\mathtt{F}_1, \mathtt{F}_2, \mathtt{F}_3$ of the null space of $\mathtt{M}$ is computed and $\mathtt{F}$ is expressed as a linear combination of the basis

$$\mathtt{F} = x\,\mathtt{F}_1 + y\,\mathtt{F}_2 + \mathtt{F}_3. \tag{5}$$

The rank (2) and trace (4) constraints on $\mathtt{E}$ are used with (3) to determine the coefficients x, y. In this case this brings ten third and fourth order polynomial equations in three unknowns x, y, and $w = f^{-2}$ in 20 monomials.

Next we describe how to solve this system of equations using the generalized eigenvalues and eigenvectors and using the Gröbner basis method.

3. Polynomial eigenvalue method

Polynomial eigenvalue solvers were used previously for solving the problem of autocalibration of one-parameter radial distortion from nine point correspondences [10] or to estimate paracatadioptric camera model from image matches [24]. In [20], a simpler, faster and numerically more stable solution to the 6-pt focal length problem [28] has been presented. Following the approach [20], we found a solution to our problem by computing generalized eigenvalues of certain matrices.

Polynomial eigenvalue problems are problems of the form

$$\mathtt{A}\,(\lambda)\,\mathbf{v} = 0, \tag{6}$$

where $\mathtt{A}\,(\lambda)$ is a square matrix, where each element is a polynomial in λ. We can expand $\mathtt{A}\,(\lambda)$ into

$$\mathtt{A}\,(\lambda) \equiv \lambda^l \mathtt{C}_l + \lambda^{l-1}\mathtt{C}_{l-1} + \cdots + \lambda \mathtt{C}_1 + \mathtt{C}_0, \tag{7}$$

in which the $\mathtt{C}_j$ are square coefficient n by n matrices.

Our problem can be formulated as the simplest "linear" eigenvalue problem

$$(\lambda\,\mathtt{B} - \mathtt{A})\;\mathbf{v} = 0, \tag{8}$$

which can be directly rewritten into the generalized eigenvalue problem (GEP) [1]

$$\mathtt{A}\,\mathbf{v} = \lambda\,\mathtt{B}\,\mathbf{v}. \tag{9}$$

Generalized eigenvalue problems (9) are well studied problems and there are efficient numerical algorithms for solving them [1]. MATLAB provides the function $eig(A, B)$ which solves the problem (8).

3.1. Polynomial eigenvalue solution

The polynomial eigenvalue solver of our problem starts with 10 equations in three unknowns x, y and $w = f^{-2}$

as described in Section 2. The system can be rewritten into the following matrix form

$$\mathtt{M}\,X = 0, \tag{10}$$

where $\mathtt{M}$ is a 10×20 coefficient matrix and $X = (wx^3, wyx^2, wy^2x, wy^3, x^3, yx^2, y^2x, y^3, wx^2, wyx, wy^2, x^2, yx, y^2, wx, wy, x, y, w, 1)^\top$ is a vector of 20 monomials.

Unknowns x and y appear in degree three but w appears only in degree one. Therefore, we can select $\lambda = w$ and rewrite these ten equations as

$$(w\,\mathtt{C}_1 + \mathtt{C}_0)\,\mathbf{v} = 0, \tag{11}$$

where $\mathbf{v} = (x^3, yx^2, y^2x, y^3, x^2, yx, y^2, x, y, 1)^\top$ is a 10×1 vector of monomials and $\mathtt{C}_1$, $\mathtt{C}_0$ are 10×10 coefficient matrices such that $\mathtt{C}_1 \equiv (m_1\ \ m_2\ \ m_3\ \ m_4\ \ m_9\ \ m_{10}\ \ m_{11}\ \ m_{15}\ \ m_{16}\ \ m_{19})$, $\mathtt{C}_0 \equiv (m_5\ m_6\ m_7\ m_8\ m_{12}\ m_{13}\ m_{14}\ m_{17}\ m_{18}\ m_{20})$, where m_j is the j^{th} column from the corresponding coefficient matrix $\mathtt{M}$.

The formulation (11) is a generalized eigenvalue problem which can be solved by MATLAB $eig(-\mathtt{C}_0, \mathtt{C}_1)$. After solving (11), we obtain 10 eigenvalues, solutions for $w = f^{-2}$ and 10 corresponding eigenvectors $\mathbf{v}$ from which we extract solutions for x and y. We do this by dividing $\mathbf{v}$ by its last coordinate which gives $x = \mathbf{v}(8)$, $y = \mathbf{v}(9)$. Then we use (5) to get solutions for the fundamental matrix $\mathtt{F}$.

Note that this solver delivers a relaxed solution to the original problem. The solution contains $\mathbf{v}$'s that automatically (within limits of numerical accuracy) satisfy the constraints induced by the problem, i.e. $\mathbf{v}(1) = \mathbf{v}(8)^3$, and additional $\mathbf{v}$'s that do not satisfy them. However, such $\mathbf{v}$'s can be eliminated by verifying the monomial dependences. Notice also that solutions satisfying the monomial constraints on $\mathbf{v}$ will be obtained for exact as well as noisy data. This is because we are solving a minimal problem and noisy data can be viewed as perfect input for a different camera configuration. Hence, we again obtain a solution satisfying the monomial constraints on $\mathbf{v}$.

4. Gröbner basis method

The Gröbner basis method is based on polynomial ideal theory and multivariate polynomial division and generates special bases of these ideals, called Gröbner bases [8] .

These bases can be used to construct special *action matrices*, which can be viewed as a generalization of a companion matrix used in solving one polynomial equation in one unknown. The solutions to the system of polynomial equations can be easily obtained from the eigenvalues and eigenvectors of this action matrix.

More on Gröbner basis methods can be found in [7, 8], and on their applications in computer vision in, e.g. in [28, 19, 5]. In [21] an automatic generator of polynomial equations solvers based on this Gröbner basis method has been proposed.

4.1. Gröbner basis solver

The Gröbner basis solver of our problem starts with 10 equations in three unknowns as described in Section 2.

To create the action matrix, we use the method described in [19, 21]. Using this method we have found that obtaining all necessary polynomials for crating the action matrix calls for generating all monomial multiples of the initial ten polynomial equations up to the total degree five. This means that we need to multiply our 4^{th} degree polynomial equations with all 1^{st} degree monomials and our 3^{rd} degree polynomial equations with all 2^{nd} degree monomials.

In this way we generated 36 new polynomials which, together with the initial ten polynomial equations, form a system of 46 polynomials in 56 monomials. Then, we removed all unnecessary polynomials by the procedure described in [4] and obtain 21 equations in 56 monomials.

After rewriting these polynomials in the matrix form and performing the Gauss-Jordan (G-J) elimination of the corresponding coefficient matrix $\mathtt{M}$, we obtained all polynomials which we need for constructing the action matrix.

Before the G-J elimination we can remove columns of the matrix $\mathtt{M}$ corresponding to the monomials that do not have impact on our solution (they do not appear in the action matrix). In this way we obtained 21×30 matrix $\mathtt{M}$.

The online solver then does only one G-J elimination of the 21×30 coefficient matrix $\mathtt{M}$. This matrix contains coefficients which arise from concrete image measurements. After G-J elimination of $\mathtt{M}$ the action matrix can be created from rows of $\mathtt{M}$. The solutions to $x, y, w = f^{-2}$ can be found from eigenvectors of this action matrix.

This online solver is exactly what can be obtained using the automatic generator from [21].

5. Critical motions

It is known that Euclidean structure can always be recovered from a pair of images acquired by a moving calibrated camera [16]. This is not the case if the cameras are calibrated only up to an unknown focal length as critical configurations with constant and varying focal lengths start appearing [16]. The critical motions for a camera pair with varying focal length appear (1) when the principal points are in epipolar correspondence, i.e. the optical axes intersect, (2) whenever the epipolar planes or optical axes are orthogonal. If either principal point coincides with an epipole, both (1) and (2) apply. Having constant focal length provides another useful constraint and hence there are less critical motions. Some of them remain however. For example, (3) arbitrary planar motions when the optical axes lie in the plane (e.g. a driving car with a forwards-pointing camera), (4) "turntable rotations" about the intersection point of the two optical axes, when these do not lie in a plane.

In [31] authors demonstrated that none out of (1), (2) and (4) results in degenerated configuration for a fully calibrated and an up to focal length calibrated camera pair. Configuration (3) works too except for the configuration in which the two optical axes are coincident, i.e. pure forward motion. It is possible to prove these results using methods from [16] but we omit this proof here due to the lack of space.

Since in [31] these critical motions were not studied in experiments, in Section 7 we show performance and comparison of our algorithms with existing ones in these configurations.

6. Reconstruction pipeline

This section presents an SfM algorithm based on our novel minimal solvers. This SFM algorithm assumes existence of very few (in principle single) cameras with known internal calibration. In Section 7 we show that assuming zero skew, aspect ratio equal to one, the principal point in the image center and the focal length extracted from the *jpeg-exif* header is sufficient.

6.1. Robust matching

Given an unordered set of images with feature points and feature descriptors, the algorithm first locates all images with available calibration information - image seeds. Then, each image seed is matched to the images in the image set using the robust matching method described below.

We implemented our minimal solver using the DEGENSAC algorithm [14]. The DEGENSAC algorithm samples 6-tuples from the set of tentative correspondences (TC), evaluates the focal length, the essential matrix and calculates the number of inliers for the estimated model in a usual way [13]. When a "better" hypothesis is sampled a degeneracy test is evaluated, i.e. test if 5 or 6 points from the sample are on a plane. Then the plane is used to split tentative correspondences into points on and off the plane. In DEGENSAC [14], only two points off-the-plane are sampled and the fundamental matrix is calculated using the plane-plus-parallax algorithm [13]. The existing off-online algorithm [31] can be used here to extract the focal length. However, we observed better results when we used clustered tentative correspondences and applied the local optimization algorithm [15] using our minimal solver with 3 off and 3 on the plane points. This way we get a more accurate solution or a solution with more inliers. Since DEGENSAC [14] is capable of detecting dominant planes, it can robustly detect when images form a panorama or observe a planar scene. We remove such degenerate pairs from further computation.

6.2. Seed reconstruction

From the previous step we get the essential matrices between the seed images and some of the remaining images and the focal lengths of these images. Not all estimated essential matrices and focal lengths are correct. Hence, unreliable geometries need to be removed. We treat each seed image separately. Fixing a seed image first, we calibrate all images where a focal length was estimated. We do this also for cameras where camera calibration is known from the *jpeg-exif* header. For such images we compare estimated focal length and the focal length from the *jpeg-exif* header. If the focal lengths difference is below a certain threshold we increase votes indicating correctness of the *jpeg-exif* value of both the seed image and the tested image. This threshold depends on the absolute focal length and the expected noise level. The acceptable difference is smaller for smaller focal lengths and greater for bigger focal lengths. We estimated these thresholds experimentally by observing deviation of estimated focal lengths from the ground truth in synthetic experiments using one pixel noise level. Next we filter unreliable essential matrices. We are using a method similar to [22], i.e. we test if rotation matrices between triplets of cameras obey transitivity.

First, we extract all rotations between seed images and the remaining images using the essential matrices [13]. Since the rotation between the i^{th} and the j^{th} image is missing we need to calculate it. Common image features between the seed, the i^{th} and the j^{th} images (common TC) are extracted and robust matching estimator is used to find common corresponding points and the essential matrix. Since both the i^{th} and the j^{th} images are already calibrated we can run calibrated 5-pt [26] solver instead of our 6-pt solver. However, we use our 6-pt since it returns focal length too. The focal is expected to be close to one since images are calibrated. Hence we compare these focal lengths as described above to detect possible inconsistences. If everything goes well and the number of inliers is greater than 80% of common TC between the i^{th} and the j^{th} image, we extract the corresponding rotation matrix. Now, rotations must obey transitivity:

$$R_{seed,j} R_{seed,i}^T \sim R_{i,j}, \tag{12}$$

where $R_{seed,*}$ is the rotation between the seed and the i^{th} resp. j^{th} image, $R_{i,j}$ is the rotation between the i^{th} and the j^{th} image and $R_1 \sim R_2$ means that the rotation $R_1 R_2^{-1}$ is small. We allow at most $5\deg$ rotation error. If the rotation is consistent, we increase votes (reliability score) for both cameras. We observed, that images with at least two votes were reliable enough, but usually this number was either zero or more than six.

After unreliable geometries were detected we extract camera pairs between the seed and the i^{th} image and triangulate 3D structure [13]. Each two-view reconstruction is

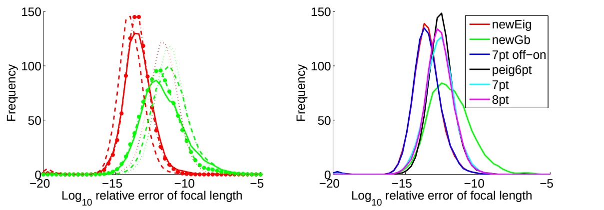

Figure 2. Left: Focal length estimation of the solvers on noise free data set for general motions (solid line), turntable rotation (dash-dot line), pure sideways translation (dashed line) and a forward translation with small sideway motion (dotted line). The generalized eigenvalue solver is shown in red and the Gröbner basis solver in green. Right: Comparison of the solvers on the general scene, see text.

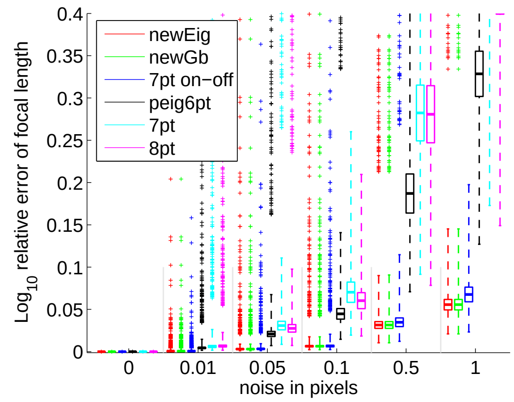

Figure 3. Performance of our solvers for general motions with growing noise level, see text.

determined up to a scale only, thus we use common 2D-3D correspondences to fix these scales. We fix the scale according to the reconstruction between the seed and the camera with the highest reliability score. A bundle adjustment [13] can be used to improve the quality of the reconstruction. However, we omitted bundle adjustment in scenes reconstructed in Section 7.

6.3. Seeds registration

After all seeds were reconstructed as described in previous section we register them together. First we identify two seeds with the highest number of common images. If more possibilities exist, we select among them the two seeds with the highest number of registered images. We transform the seed with less registered images to the coordinate system of the seed with the higher number of registered images. Rotations and translations of common cameras are "averaged" similarly as in the linear method described in [22].

When all possible seeds were merged we can run a bundle adjustment [13]. Then, cameras corresponding to the registered images with higher reliability score can be declared as calibrated and the algorithm can start again with new image seeds.

7. Experiments

In this section we evaluate both the generalized eigenvalue solution and the Gröbner basis solution of the problem and compare them to the existing non-minimal solution [31]. We study critical motions and compare the numerical stability and computational complexity of both solvers.

7.1. Synthetic data set

We study the performance of our methods on synthetically generated ground-truth 3D scenes. These scenes were generated as random points in a 3D cube. Each 3D point was projected by a camera with random parameters or parameters testing degenerate configurations (pure translation, etc.). Then, Gaussian noise with standard deviation σ was added to each image point assuming a 1000×1000 pixel image.

7.2. Numerical stability

In this synthetic experiment we study numerical stability of our solvers in various configurations and compare them with other solvers. We focus on focal length estimation because rotation and translation is usually good once we have a good focal length estimate. Figure 2 shows the performance of our solver on synthetic noise free scenes in (i) general motion, (ii) turntable configuration, (iii) pure sideways translation and (iv) forward translation with small sideway motion of cameras such that their optical axes are not coincident. We know from the previous section that (ii), (iii) and (iv) are critical configurations for camera pairs with constant or varying focal length i.e. for the 6-pt [28] and 7-pt [13] algorithms followed by a focal length extraction.

Figure 2 (left) shows that these configurations are not critical for fully calibrated and up to f calibrated camera pairs. Focal lengths estimated in "critical configurations" are equally good as those in a general configuration.

Figure 2 (right) shows the stability of the algorithms in general configurations. We compare here both our solvers (newEig, newGB) with the non-minimal off-online solver (7pt on-off) [31] and the polyeig solver of classical 6-pt problem (peig6pt) [20], the 7-pt (7pt) and the normalized 8-pt (8pt) algorithms [13]. For the 7-pt and 8-pt algorithms we first calculated fundamental matrices and then extracted focal lengths using Bougnoux method [3]. In case of the 8-pt algorithm, we used all image measurements to calculate epipolar geometry. Figure 2 (right) shows that our generalized eigenvalue and the non-minimal off-online algorithm [31] give best estimates of f but all methods perform almost equally.

We have also made a comparison of our algorithms with

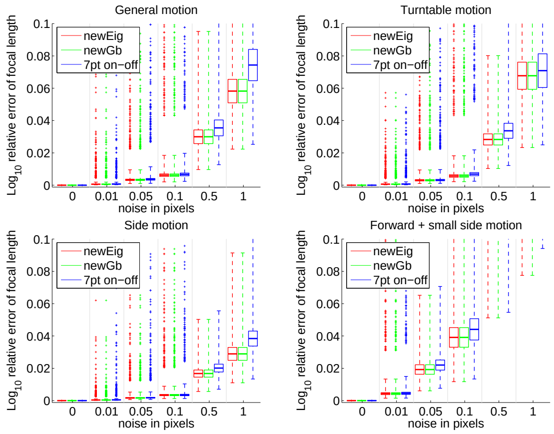

Figure 4. Performance of our solvers in the 4 studied special configurations. See text.

the non-minimal off-online algorithm for all "critical configurations" like in Figure 2 (left). The results of the non-minimal algorithm for all configurations where almost the same as the results of our generalized eigenvalue solver and therefore we didn't print them here.

In the experiment not shown here we found that the stability of both our solvers is almost independent from the true focal length. The results for all reasonable focal lengths from $25mm$ to $300mm$ were similar to the result from Figure 2 (right) which was made for the focal length $36mm$.

Next experiment shows the quality of the focal length estimation for general camera configurations when adding noise to image measurements. First, we fixed the focal lengths of the first camera to $30mm$ and of the second camera to $50mm$. Then we generated 1000 random camera poses for each tested noise. Results for the general camera motion are shown in Figure 3. Our solvers (newEig, newGb) give almost the same results here and they are performing very well even with one pixel noise level. The results of the non-minimal algorithm (7pt on-off) were for smaller noise levels similar to the results of our algorithms, however for larger noise levels the performance of our algorithms was slightly better.

The previous experiment has shown results for a general camera motion. A different situation occurs when we are testing these solvers in their critical configurations. Adding a small perturbation to the image measurements helps removing the degeneracy and lets solvers find an approximate solution. However, the experiments show that the classical 6-pt [28], 7-pt [13] and 8-pt [13] solvers failed to deliver any real solution in more than 90% of all tests. Hence Figure 4 shows plots only for our solvers and the non-minimal off-online solver in turntable motion, side motion and forward translation with small side motion and non-coincident optical axes. Our solvers failed to deliver real solutions in less than 2 cases in 1000 calls. Again the results of the

Figure 5. Correct 3D reconstruction of cameras in a "turntable configuration" by our method. This is a degenerate configuration for general 6-pt algorithms which often occurs with causal photographers taking pictures of 3D objects.

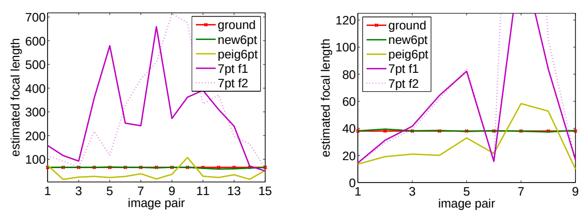

Figure 6. Estimated focal length using different algorithms for turn table sequence (left) and side motion sequence (right). For 7-pt algorithm we extract two focal lengths - 7pt f1 (solid line) and 7pt f2 (dashed line)

non-minimal algorithm were similar to the results of our algorithms for smaller noise levels and all configurations similar and for larger noise levels the performance of our algorithms was slightly better.

7.3. Real experiments - critical motion

In the real data experiments we have aimed at the critical motions described in Section 7.2. We captured a set of images of a non-planar scene in a "turntable" configuration and with a sideways moving camera. Figure 5 shows a reconstruction using our autocalibration method described in Section 6, without any additional numerical improvements.

Figure 6 (left) shows estimated focal lengths using different algorithms. Note that this is a critical configuration for the standard 6-pt an 7-pt algorithms. We obtained results for these solvers since image correspondences are not measured perfectly what helped finding solution which were close to the critical configuration. However, it can be seen from the Figure 6 (left), that estimated focal lengths are far from the ground truth.

Results for the sideways motion configuration are similar to those obtained for the "turntable" configuration. Estimated focal lengths are in Figure 6 (right). Results of the

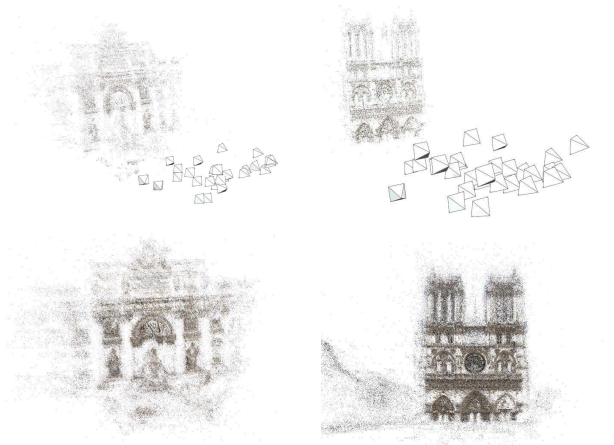

Figure 7. Single seed reconstructions on the top, full reconstruction on the bottom.

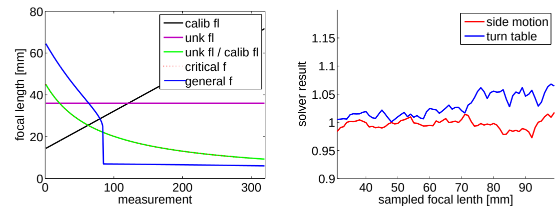

Figure 8. Left: True focal length of the calibrated camera (calib fl) is increasing. Focal length of the second camera is unknown but constant (unk fl). We have found that estimated focal lengths for critical motion (critical f) is identical to the ratio of the ground truth focal lengths (unk fl/calib fl). General motion (general f) configuration is shown in blue. Right: The "sampling" experiment on the real data sets with constant focal length camera. Graphs show the ratio of the estimated and sampled focal lengths. The ground truth ratio is equal to one.

non-minimal algorithm were similar to the results of our solvers (new6pt) and therefore are not showed in the figure.

Note that the standard 6-pt algorithm uses the same focal length in both cameras. Since the difference between estimated focal length and the ground value is high, one cannot get the correct Euclidean reconstruction. Hence, building structure from motion using such partial reconstructions is a hard problem and it is not possible to obtain a good reconstruction, e.g., by using bundle adjustment methods [13] without specifying additional constraints, e.g., constant focal length in the whole sequence. We failed to reconstruct our "turntable" sequence even with a robust state of the art systems such as PhotoSynth [23].

7.4. Real experiments – images from the Internet

In this experiment we tried to use the previous approach on a set of images downloaded from the Internet. We have downloaded 2550 images of Fountain di Trevi and 4000 images of Notre Dame de Paris from Flickr [11] database. In such a huge database of images, it is not a problem to find an image with focal stored in the *jpeg-exif* and use it as a seed.

We extracted SURF [2] feature points and descriptors of all images and used [17] to obtain image matches. For the best 50 images for each seed we extracted tentative correspondences as points where the best descriptor dominates by 20% over the second best descriptor [25]. Then we used our reconstruction pipeline from Section 6 to register images. We did not generate additional seeds and did not use bundle adjustment in this step.

The Fountain dataset contained about 10%(239) seed images with $35mm$ equivalent focal length in the *jpeg-exif*. About 14% of them did not contain Fountain scene and 11% were rejected as wrongly calibrated. Numbers for the Notre Dame sequence were similar. A 3D reconstruction of both dataset from a single seed image are shown in Figure 7 (top) and the full reconstructions in Figure 7 (bottom). We processed 4000 Flickr photos in less than a day.

7.5. Interesting questions

Here we try to answer questions you may ask yourself:

- ***Does the method work when all six point correspondences are projections of the points from a plane in the 3D space?*** No, this is a degenerated configuration [16].

- ***What happens if the calibration of the calibrated camera is inaccurate or even unknown?***

  Figure 8 (left) tries to answer this question. We have generated synthetic scenes and used our solver to find an unknown focal length (unk fl) without doing any calibration of the first camera. We were interested in the relation between the evaluated and the ground truth focal lengths.

  We found that results for general configurations are almost random. However, the experiment has shown that the estimated focal length is actually the ratio of the two ground truth focal lengths for cameras in critical configurations. This experiment shows why "sampling" approach does not help finding absolute focal length given the ratio ρ of the lengths in the critical configuration. Basically, one gets ρf for every sampled focal f. In case of the "turntable" and sideways motion sequence from our real data set, where cameras have constant focal lengths, we obtained results corresponding to a ratio close to one, Figure 8(right).

- ***Given a ratio of focal lengths, can we emulate the popular 6-pt algorithm [28] and avoid degenerated configurations ?*** Using this algorithm we can emulate the six point problem by sampling the calibrated focal length and testing if the ratio of estimated and sampled focal lengths is close to the given ratio. As expected, such algorithm will not work for critical motions as shown above.

7.6. Computational complexity

The most expensive part of both solvers is in calculating eigenvectors. The Gröbner basis solver has to perform single G-J elimination of 21×30 matrix in order to build the action matrix. We take sparseness of the matrix into account and hence elimination time is negligible comparing to the eigenvector computation. The generalized eigenvalue solver does not have to perform any elimination but has to calculate generalized eigenvectors. Still, this involves calculation of eigenvectors of 10×10 matrix. Running time of both solvers is less than $1ms$ on $3Ghz$ AMD Sempron mobile. Comparing the two algorithms, the generalized eigenvalue solver is both faster and numerically more stable.

8. Conclusions

We have presented a minimal solution to finding the relative pose between a completely calibrated camera and a camera with an unknown focal length given six point correspondences. We presented two efficient solvers to the problem based on the Gröbner basis, resp. on the generalized eigenvalues. Both algorithms are fast, numerically stable and in case of the generalized eigenvalue solver extremely simple to implement. The source codes to both solvers can be found at *http:/cmp.felk.cvut.cz/minimal*.

We have demonstrated that our solvers produce very stable results for both synthetic and the real scenes with or without noise and even for critical motions of the state of the art algorithms. Since it is often easy to get a single image from a calibrated camera, our algorithm is practical when working with unknown images, e.g., downloaded from the Internet. We have show this in experiments and presented a new efficient method for large-scale structure from motion from unordered data sets download from the Internet.

References

[1] Z. Bai., J. Demmel, J. Dongorra, A. Ruhe, and H. van der Vorst. Templates for the solution of algebraic eigenvalue problems. *SIAM 2000*. 2

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *CVIU, 110:346–359, 2008* 7

[3] S. Bougnoux. From projective to Euclid. space under a practical situation, a criticism of self-calibration. *ICCV 1998*. 5

[4] M. Bujnak, Z. Kukelova, and T. Pajdla. A general solution to the P4P problem for camera with unknown focal length. *CVPR 2008*. 3

[5] M. Byröd, K. Josephson, and K. Åström. Improving numerical accuracy of Gröbner basis polynomial equation solver. *ICCV 2007*. 3

[6] M. Byröd, Z. Kukelova, K. Josephson, T. Pajdla, and K. Åström. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. *CVPR 2008*. 1

[7] D. Cox, J. Little, and D. O'Shea. *Using Algebraic Geometry*. Springer Verlag 2005. 3

[8] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer Verlag 2007. 3

[9] O. Faugeras and S. Maybank. Motion from Point Matches: Multiplicity of Solutions. *IJCV 4(3):225-246, 1990*. 1

[10] A. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. *CVPR 2001*. 2

[11] Flickr. *http://www.flickr.com/* 2, 7

[12] C. Geyer and H. Stewenius. A 9-point algorithm for estimating paracatadioptric fund. matrices. *CVPR 2007*. 1

[13] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge Univ. Press, 2003. 1, 2, 4, 5, 6, 7

[14] O. Chum, T. Werner, and J. Matas. Two-View Geometry Estimation Unaffected by a Dominant Plane. *CVPR 2005*. 4

[15] O. Chum, J. Matas., and J. Kittler. Locally Optimized RANSAC. DAGM 2003. 4

[16] F. Kahl and B. Triggs. Critical Motions in Euclidean Structure from Motion. *CVPR 1999*. 3, 4, 7

[17] J. Knopp, J. Šivic, and T. Pajdla. Location recognition using large vocabularies and fast spatial matching. *Research Report, CTU–CMP–2009–01, 2009* 7

[18] E. Kruppa, Zur Ermittlung eines Objektes aus zwei Perspektiven mit Innerer Orientierung, *Sitz.-Ber. Akad.Wiss.,Wien, Math. Naturw. Kl., Abt. IIa., 122:1939-1948, 1918*. 1

[19] Z. Kukelova and T. Pajdla. A minimal solution to the autocalibration of radial distortion. *CVPR 2007*. 1, 3

[20] Z. Kukelova, M. Bujnak and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. *BMVC 2008*. 2, 5

[21] Z. Kukelova, M. Bujnak and T. Pajdla. Automatic Generator of Minimal Problem Solvers. *ECCV 2008*. 3

[22] D. Martinec and T. Pajdla. Robust Rotation and Translation Estimation in Multiview Reconstruction In *CVPR* 2007. 4, 5

[23] MS PhotoSynth. *http://www.photosynth.net* 7

[24] B. Micusik and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. *CVPR 2003*. 2

[25] M. Muja, and D. Lowe. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *Preprint, University of British Columbia, 2008*. 7

[26] D. Nister. An efficient solution to the five-point relative pose. *IEEE PAMI, 26(6):756–770, 2004*. 1, 4

[27] N. Snavely, S. Seitz, and R. Szeliski. Photo tourism: Exploring image collections in 3D. *SigGraph 2006*. 1

[28] H. Stewenius, D. Nister, F. Kahl, and F. Schaffalitzky. A minimal solution for relative pose with unknown focal length. *CVPR 2005*. 1, 2, 3, 5, 6, 7

[29] H. Stewenius, D. Nister, M. Oskarsson, and K. Astrom. Solutions to minimal generalized relative pose problems. *OMNIVIS 2005*. 1

[30] H. Stewenius, C. Engels, and D. Nister. Recent developments on direct relative orientation. *ISPRS J. of Photogram. and Rem. Sens., 60:284–294, 2006*. 1

[31] M. Urbanek, R. Horaud, and P. Sturm. Combining Off- and On-line Calibration of a Digital Camera. *Third Int. Conf. on 3-D Digital Imaging and Modeling, 2001*. 2, 4, 5