

Influence of numerical conditioning on the accuracy of relative orientation*

Siniša Šegvić, Gerald Schweighofer and Axel Pinz

Institute of Electrical Measurement and Measurement Signal Processing,
Graz University of Technology, Kopernikusgasse 24/IV, 8010 Graz, Austria

sinisa.segvic@tugraz.at gerald.schweighofer@tugraz.at axel.pinz@tugraz.at

Abstract

We study the influence of numerical conditioning on the accuracy of two closed-form solutions to the overconstrained relative orientation problem. We consider the well known eight-point algorithm and the recent five-point algorithm, and evaluate changes in their performance due to Hartley's normalization and Muehlich's equilibration. The need for numerical conditioning is introduced by explaining the known occurrence of the bias of the eight-point algorithm towards the forward motion. Then it is shown how conditioning can be used to improve the results of the recent five-point algorithm. This is not straightforward since the conditioning disturbs the calibration of the input data. The conditioning therefore needs to be reverted before enforcing the internal cubic constraints of the essential matrix. The obtained improvements are less dramatic than in the case of the eight-point algorithm, for which we offer a plausible explanation. The theoretical claims are backed up with extensive experimentation on noisy artificial datasets, under a variety of geometric and imaging parameters.

1. Introduction

We consider the recovery of relative orientation [6] (or relative pose [5]), expressing the relation between the metric coordinates (i.e. Euclidean up to an unknown scale factor) of the two camera frames. A prominent approach to achieve that in closed-form relies on the decomposition of the essential matrix into motion parameters [8]. In general, essential matrix can be recovered only if the images have been acquired with calibrated cameras, allowing the points to be directly expressed in coordinates of the underlying pinhole camera model. In this calibrated context the essential matrix is strongly related to the well-known epipo-

lar constraint. In fact, the essential matrix is uniquely constrained in many situations of practical importance by the epipolar constraint alone. This is the main idea behind the well-known eight-point (8pt) algorithm [6], in which the recovery is expressed as a solution to the linear system of homogeneous equations.

A disadvantage of the 8pt algorithm is that it does not enforce all available constraints. The 8pt algorithm requires at least 8 correspondences, while it has been long known that 5 correspondences in general position provide at most 10 solutions [6, 2]. The additional constraints express the characteristic algebraic structure of the essential matrix, and can be enforced by solving a system of cubic equations. An efficient procedure for achieving that in closed-form is at heart of the recent five-point (5pt) algorithm [10]. A major benefit of the 5pt algorithm is that it can be applied to subsets of only five correspondences. In a typical random sampling environment, this ensures faster guessing in presence of outliers and, together with a fair execution speed, can significantly release the computational burden in a real-time application.

The interest for the 8pt algorithm has been revived in the vision community after the Hartley's normalization has been introduced in [3, 1]. There it has been shown that, in the *overconstrained* case when more than 8 correspondences are available, substantial improvements are obtained after subjecting the image coordinates of the employed point correspondences to a normalizing linear transform. A different numerical conditioning scheme has been proposed by Muehlich [9], consisting in right-multiplication of the matrix of the linear system with a suitable equilibration matrix. Yet two other conditioning schemes have been proposed in [14, 15], but the reported results seem to be close to the results of Hartley's normalization.

Despite the dramatic effects of the conditioning approaches to the 8pt algorithm performance, there are no previous studies of their eventual suitability in the context of the overconstrained 5pt algorithm, to the best of our knowledge. The two algorithms are strongly related in the over-

*This work has been supported by the European MC IIF project AViC-MaL, the Austrian science fund FWF S9103-N13, and the French national project Predit Mobivip.

constrained case, since the output of the 8pt algorithm is one of the four inputs to the main procedure of the 5pt algorithm. We therefore found it intriguing that the 5pt algorithm seems to provide a fair performance without conditioning, while the experimental evidence [3, 9] implies that unconditioned data produce devastating effects on the 8pt algorithm.

In this paper, we propose some answers to the above issues, by presenting an advance towards a unified view of the influence of numerical conditioning to the recovery of the essential matrix. In order to provide a better intuition of how exactly unconditioned data spoil the accuracy of the 8pt algorithm, we provide a novel explanation of the forward bias [9] in the recovered epipole, which regularly occurs under common imaging conditions. We also show how the two conditioning schemes can be applied to the 5pt algorithm, as well as why the obtained effects are in general less prominent than in the case of the 8pt algorithm.

The paper is organized as follows. Section 2 reviews the two constraints which have been used to recover the essential matrix. The explanation of the 8pt algorithm forward bias is presented in Section 3, as a consequence of a special kind of poor conditioning in the input data. The two conditioning schemes are reviewed in Section 4, while the suitability of their application in the case of the 5pt algorithm is discussed in Section 5. The extensive experimental results on noisy artificial data, which support and extend the theoretic propositions, are presented in Section 6. Finally, a short discussion and a conclusion are provided in Section 7.

2. Recovery of the essential matrix

Let a common set of 3D points \mathbf{Q}_i be observed by two calibrated cameras C_k , $k = A, B$. Let the rigid body displacement of the camera B with respect to the referential coordinates of the camera A be given with (\mathbf{R}, \mathbf{t}) . Then, the 3D points can be expressed in coordinates of the two cameras as $\mathbf{Q}_{iA} = \mathbf{Q}_i$, and $\mathbf{Q}_{iB} = \mathbf{R}\mathbf{Q}_i + \mathbf{t}$. The relation between the corresponding homogeneous points \mathbf{q}_{ik} in two views $k = A, B$ is defined by the well known epipolar constraint: [6, 4]:

$$\mathbf{q}_{iB}^\top \cdot \mathbf{E} \cdot \mathbf{q}_{iA} = 0. \quad (1)$$

The above constraint involves the essential matrix $\mathbf{E}_{3 \times 3}$ [6] only if the two images are calibrated, i.e. their coordinates are expressed in units for which the focal length in the pin-hole camera model equals one. Let $[\mathbf{a}]_\times$ denote a skew symmetric matrix such that $[\mathbf{a}]_\times \cdot \mathbf{x} = \mathbf{a} \times \mathbf{x}$, $\forall \mathbf{a}, \mathbf{x} \in \mathbb{R}^3$. Then the relation between the essential matrix and the relative orientation is $\mathbf{E} = [\mathbf{t}]_\times \cdot \mathbf{R}$. The magnitude of the translation can not be recovered since the constraint (1) is homogeneous. Thus, the essential matrix has only 5 degrees of freedom, while the remaining 3 degrees of freedom

towards a general homogeneous matrix are enforced by the following cubic constraint [2]:

$$2 \cdot \mathbf{E}\mathbf{E}^T\mathbf{E} - \text{trace}(\mathbf{E}\mathbf{E}^T)\mathbf{E} = 0 \quad (2)$$

The algorithms for recovering the essential matrix can be divided in calibrated and uncalibrated ones, depending on whether they enforce the cubic constraint (2) or not. The eight-point (8pt) algorithm [6] is the most widely known uncalibrated procedure. It relies exclusively on the epipolar constraint which is linear in coefficients of the essential matrix. Thus, each correspondence pair defines one row of the system

$$\mathbf{A}_{n \times 9} \cdot \mathbf{e} = \mathbf{0}, \quad (3)$$

yielding the vector \mathbf{e} with elements of the essential matrix:

$$\mathbf{e} = [\mathbf{E}_{11} \ \mathbf{E}_{12} \ \mathbf{E}_{13} \ \mathbf{E}_{21} \ \mathbf{E}_{22} \ \mathbf{E}_{23} \ \mathbf{E}_{31} \ \mathbf{E}_{32} \ \mathbf{E}_{33}]^\top \quad (4)$$

The system is usually solved by enforcing $|\mathbf{e}| = 1$ using SVD [12]. Eight or more distinct correspondences uniquely define the relative orientation, except in some special arrangements of the camera and the imaged points, known as degenerate configurations [4]. The only degenerate configuration which occurs frequently in practice involves a planar scene and arbitrary positions of the two cameras, and will be further discussed in Section 3.

The recent 5pt algorithm [10] is the most evolved calibrated algorithm for recovering the essential matrix. It constructs the solution \mathbf{E} satisfying (2), as a linear combination of matrices \mathbf{E}_i , corresponding to the four lowest right-singular vectors of the matrix \mathbf{A} , \mathbf{e}_i , $i=6,7,8,9$:

$$\mathbf{E} = a \cdot \mathbf{E}_6 + b \cdot \mathbf{E}_7 + c \cdot \mathbf{E}_8 + d \cdot \mathbf{E}_9. \quad (5)$$

Thus, the procedure is applicable to sets of only five correspondences. However, due to the involved nonlinearity, the solution is not unique: the algorithm may return up to 10 motion hypotheses. Spurious hypotheses can be rejected by looking at the reprojection error [11] corresponding either to an additional (6th) point or to all available correspondences in the two images. Due to the enforcement of (2), the 5pt algorithm faces no degeneracy in the planar case. However, in the presence of planar ambiguity [7, 13], two of the returned hypotheses will satisfy all available constraints up to deviations due to noise. Without further assumptions, the disambiguation can be performed using a third view of the scene [10].

3. The forward bias of the 8pt algorithm

Assume that two cameras observe a common planar scene and denote the corresponding homography as \mathbf{H} . Then the class of matrices satisfying the epipolar constraint (1) can be concisely written as [4]:

$$\mathcal{E}(\mathbf{H}) = [\mathbf{a}]_\times \cdot \mathbf{H}, \quad \forall \mathbf{a} \in \mathbb{R}^3. \quad (6)$$

Thus, an algorithm enforcing the epipolar constraint in the planar case is confronted with a projective subspace of solutions, although at most two of them can be correct [7]. A base for the subspace can be obtained by setting \mathbf{a} in (6) equal to the three canonical unit vectors. The resulting three matrices are linearly independent (assuming that \mathbf{H} has full rank), which implies that the subspace has two dimensions.

Despite this degeneracy, previous researchers [10] have noticed an ‘‘excellent’’ performance of the unconditioned 8pt algorithm in the case of the forward motion. This anomaly is due to the known bias of the unnormalized 8pt algorithm [9]. We shall show that, in the case of a distant plane, the 8pt algorithm favours a solution with an approximately correct rotation and the epipole \mathbf{t} in the middle of the second image, regardless of the actual displacement.

Consider again the matrix \mathbf{A} of the linear system (3). The i -th row of \mathbf{A} is obtained by applying the epipolar constraint to the point correspondence $\mathbf{q}_{iA} \leftrightarrow \mathbf{q}_{iB}$, where $\mathbf{q}_{ik} = [x_{ik}, y_{ik}, 1]^T, k = A, B$:

$$\mathbf{A}_i = [x_{iB}x_{iA} \ x_{iB}y_{iA} \ x_{iB}y_{iB}x_{iA} \ y_{iB}y_{iA} \ y_{iB}x_{iA} \ y_{iB}x_{iA} \ 1]. \quad (7)$$

Assume the uniform distribution of the points in a square image and independent zero-mean noise with same variance in both images $\sigma^2 = \sigma_A^2 = \sigma_B^2$. Denote the error in variables with Δ , and label the unperturbed counterparts with a ‘‘hat’’. Then the elements of the column 1 of \mathbf{A} can be expressed as:

$$\begin{aligned} a_{i1} &= \hat{x}_{iB}\hat{x}_{iA} + \hat{x}_{iB}\Delta x_{iA} + \Delta x_{iB}\hat{x}_{iA} + \Delta x_{iB}\Delta x_{iA} \\ &\approx \hat{a}_{i1} + \hat{x}_{iB}\Delta x_{iA} + \Delta x_{iB}\hat{x}_{iA}. \end{aligned} \quad (8)$$

The above equation implies that the error in the entries of \mathbf{A} is also zero-mean [9], so that their variance is equal to:

$$\text{var}(a_{ij}) = E[(a_{ij} - \hat{a}_{ij})^2]. \quad (9)$$

If we now consider the point coordinates as random variables too, we can determine the *expected* variance of the columns of the matrix \mathbf{A} . Denote the field of view of the square image as α . Then, due to independent noise and a decorrelated point distribution, the expected variance for column 1 is:

$$\begin{aligned} E[\text{var}(a_{i1})] &\approx E[(\hat{x}_{iB}\Delta x_{iA} + \Delta x_{iB}\hat{x}_{iA})^2] \\ &= E[(\hat{x}_{iB})^2] \cdot \sigma_A^2 + \sigma_B^2 \cdot E[(\hat{x}_{iA})^2] \\ &= 2 \cdot \sigma^2 \cdot (2 \tan(\alpha/2))^2 / 12. \end{aligned} \quad (10)$$

The same result would be obtained for any other column with *quadratic* entries, $j=1,2,4,5$. On the other hand, the variance of *linear* entries is simply:

$$\text{var}(a_{ij}) = \sigma^2, \quad j = 3, 6, 7, 8. \quad (11)$$

For usual cameras with $\alpha=45^\circ$, the ratio of the two variances above is 0.11¹. Thus the columns 3,6,7 and 8 of \mathbf{A}

¹ The ratio equals one for $\alpha \approx 102^\circ$, and then the bias disappears.

will be affected most by the noise in the input data². The noise is conveniently suppressed in solutions having small values in the four involved entries of \mathbf{E} : (1,3), (2,3), (3,1), (3,2). Define the measure of the convenience as:

$$\text{conv}(\mathbf{E}) = \|[\mathbf{E}_{13} \ \mathbf{E}_{23} \ \mathbf{E}_{31} \ \mathbf{E}_{32}]\|^{-1} \quad (12)$$

We have experimentally verified that, for moderate rotations (less than $\pi/4$) around an arbitrary axis, the convenience of $[\mathbf{a}]_{\times} \mathbf{R}$ attains a maximum very close to the optical axis:

$$\mathbf{a}_m = \arg \max_a \text{conv}([\mathbf{a}]_{\times} \mathbf{R}) \approx [0 \ 0 \ 1]^T. \quad (13)$$

This does not come as a surprise since only for \mathbf{a}_m , two of the four linear components are always zeroed out.

The above discussion shows that the unconditioned linear algorithm tends to recover a forward motion whenever the solution to (3) is not distinctive enough. The disorder occurs especially in the presence of planar degeneracy, but also disturbs the correct solution in the presence of noise. In usual configurations, the target distance is substantially greater than the baseline. Under these conditions, large translation errors can be approximately compensated by slight rotation deviations. In other words, relatively small changes in the residuals are obtained in the whole spectrum of translation directions. Consequently, the system (3) can trade-off the translational accuracy in order to suppress the residuals which are most corrupted by the noise. The favoured solutions have small rotational errors, and translations giving rise to high convenience (12). Due to the property (13), this results in a bias on the epipole towards the optical axis.

4. The two numerical conditioning schemes

Solving the linear system $\mathbf{A} \cdot \mathbf{e} = 0$ is equivalent to optimizing the following algebraic criterion:

$$\min |\mathbf{A} \cdot \mathbf{e}|, \text{ subject to } |\mathbf{e}| = 1 \quad (14)$$

The optimal \mathbf{e} turns out to be a solution of $\mathbf{A}_{\mathbf{F}} \cdot \mathbf{e} = 0$, where $\mathbf{A}_{\mathbf{F}}$ is a rank-deficient matrix closest to \mathbf{A} in the sense of the Frobenius norm [4]. If the error due to noise is unevenly distributed across the entries of \mathbf{A} , the quadratical nature of the Frobenius norm can provide enough leverage to the few noisiest entries to take complete control over the output of the procedure, as shown in Section 3. Beneficial modifications to the standard procedure can therefore be achieved by manipulating either the input correspondences \mathbf{q}_{ik} or directly the matrix of the system \mathbf{A} . Note however that any conditioning approach can improve the results only in the

²A similar analysis of variances has been employed to weight contributions of individual correspondences in a statistical validation of Hartley’s normalization [1]

overconstrained case, when (14) involves an algebraic least-squares solution.

The unevenness of the noise distribution can be alleviated by a general approach for manipulating linear systems known as *equilibration* [9]. Equilibration consists in multiplying the matrix of the system \mathbf{A} with appropriately chosen weight matrices from both sides: $\mathbf{A}_{eq} = \mathbf{W}_L \cdot \mathbf{A} \cdot \mathbf{W}_R$. The new, equilibrated system is:

$$\mathbf{A}_{eq} \cdot \mathbf{e}' = 0, \text{ where } \mathbf{e}' = \mathbf{W}_R^{-1} \cdot \mathbf{e} \quad (15)$$

A theoretically sound equilibration approach defines \mathbf{W}_R in a way that the expected error of the right-singular vectors of \mathbf{A}_{eq}^\top be zero [9]. This results in enforcing an unbiased solution for \mathbf{e} . Let $\hat{\mathbf{A}}$ denote the noise-free version of \mathbf{A} , and let $\mathbf{D} = \mathbf{A} - \hat{\mathbf{A}}$. Then, an unbiased solution is ensured under an assumption of zero-mean noise by the following constraint on \mathbf{W}_R :

$$\mathbf{W}_R^\top \cdot E[\mathbf{D}^\top \cdot \mathbf{W}_L^\top \mathbf{W}_L \cdot \mathbf{D}] \cdot \mathbf{W}_R = c \cdot \mathbf{I}. \quad (16)$$

$E[\mathbf{D}^\top \mathbf{W}_L^\top \mathbf{W}_L \mathbf{D}]$ can be estimated from the input data, by an analysis similar to (10). In the absence of prior estimates of the correspondence error, \mathbf{W}_L can be set to the unit matrix. From (7) it can be seen that not all columns of \mathbf{A} have errors, which implies that $rank(\mathbf{D}) < 9$. The simplest way to resolve that is by inserting small values at the corresponding diagonal entries of $E[\mathbf{D}^\top \mathbf{W}_L^\top \mathbf{W}_L \mathbf{D}]$. The original article [9] derives the optimal equilibration matrix by assuming the error only in one image. We have constructed the equilibration matrix under a more widely used model with error in both images. Perhaps surprisingly, the two equilibration variants perform nearly the same.

Hartley's *normalization* procedure [3, 1] transforms the points in both images before solving the linear system. This results in an essential matrix $\mathbf{E}' = \mathbf{T}_2^{-\top} \mathbf{E} \mathbf{T}_1^{-1}$ relating the transformed points $\mathbf{q}'_{ik} = \mathbf{T}_k \mathbf{q}_{ik}$, $k = A, B$:

$$\mathbf{q}'_{i2}^\top \cdot \mathbf{E}' \cdot \mathbf{q}'_{i1} = 0, \quad (17)$$

A proper choice of T_1 and T_2 ensures a more uniform distribution of entries in matrix \mathbf{A} . The recommended transformation [3, 4] consists of translation and isotropic scaling such that the centroid of the transformed points is at the origin, while their root-mean-square radial distance is $\sqrt{2}$ [1]. It has been shown that the Hartley's normalization achieves a uniform variance of the n residuals of the linear system (3), in view of varying image coordinates [1].

Note that the columns of the linear system (7) obtained from normalized correspondences (17) can be expressed as a linear combination of the original columns: for each normalization there is an equivalent equilibration. On the other hand, equilibration with e.g. $\mathbf{W}_R = diag\{1, 1, \dots, 1, 1000\}$ can not be achieved by any normalization. Thus, equilibration is a proper superset of normalization.

5. Conditioning the five point algorithm

The conditioned 8pt algorithms (15) and (17) solve a modified linear system which is less susceptible to noise. The obtained solution is simply transformed back into the original problem domain before the obtained essential matrix is decomposed into motion parameters. Unfortunately, this simple procedure can not be easily applied to a calibrated algorithm which enforces the constraint (2). The 5pt algorithm can therefore only solve for genuine essential matrices, and is not applicable to the solutions of the conditioned problems.

Nevertheless, the groundtruth solution of the conditioned problem $\hat{\mathbf{e}}'$ is expected to be close to the span of the four lower right-singular vectors of the conditioned problem \mathbf{e}'_i , $i = 6, 7, 8, 9$. In the case of the right equilibration, we would have:

$$\mathbf{W}_R^{-1} \hat{\mathbf{e}} = \hat{\mathbf{e}}' \approx a \cdot \mathbf{e}'_6 + b \cdot \mathbf{e}'_7 + c \cdot \mathbf{e}'_8 + d \cdot \mathbf{e}'_9 \quad (18)$$

for some $a, b, c, d \in \mathbb{R}$.

Consequently, the real solution should not be far from the span of the transformed right-singular vectors $\mathbf{W}_R \cdot \mathbf{e}'_i$, $i = 6, 7, 8, 9$. These four vectors can be supplied to the 5pt algorithm which does not require the vectors to be orthogonal, but just finds up to 10 tuples (a, b, c, d) for which the linear combination results in an essential matrix.

6. Experimental results

We present experimental results obtained on artificial noisy data. First we describe the experimental setup. Then we show experiments with the forward bias of the 8pt algorithm, backing up the propositions from Section 3. After that, we present results explaining the relative robustness of the 5pt algorithm to the poorly conditioned input data. Finally, we evaluate the two conditioning schemes in both algorithms, for planar and non-planar scenes.

6.1. The experimental setup

The employed experimental setup is very similar to [10], however additional specifications are provided in order to ensure the repeatability of the results. Refer to Figure 1. The left-oriented coordinate system of each camera is set so that the image plane is defined by the equation $z = 1$. The default horizontal field of view is $\alpha_H = 45^\circ$, while the vertical field of view depends on image resolution (see below). The coordinate system of the camera A is set as the reference. The geometry of the camera pair is defined by the two angles θ, ϕ . The unit translation in the x-z plane is defined by θ , as the angle between the optical axis of the first camera and the baseline. The angle ϕ defines the rotation of the 2nd camera around the common y axis. Thus, $(\theta, \phi) = (0, 0)$ implies forward motion with no rotation.

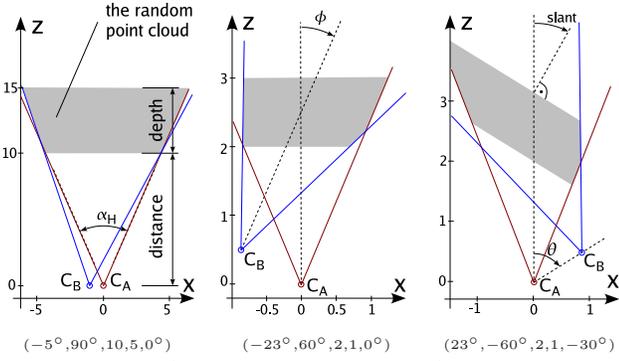


Figure 1. The default experimental setup (left), and two custom setups (middle, right), for $\alpha_H=45^\circ$. The numbers in parentheses indicate θ , ϕ , distance, depth, and slant. The baseline is always 1.

The random point cloud is instantiated in a volume visible by both cameras, between two parallel planes. The parameter *distance* denotes the distance from the first camera to the closer plane of the volume. The distance between the two planes is given by *depth*. The angle between the common normal of the two planes and the optical axis of the first camera is given by *slant*. The default values are *distance*=10, *depth*=5, *slant*= 0° . The distribution of the points in the volume is uniform. In most experiments ϕ is set in a way that the two optical axes intersect in the middle between the two planes.

The points are projected to the two images, and both coordinates of each projection are perturbed with zero-mean Gaussian noise [4]. The standard deviation of the noise is expressed in pixels of a 384×288 image obtained for $\alpha_H=45^\circ$, with the default value of $\sigma = 1.0$. In experiments with different α_H , the resolution to which σ refers changes in a way that the pixel size (and the effective noise) remain the same. Thus, for $\alpha_H=30^\circ$ and 60° , the resolution is set to 248×186 and 535×401 , respectively.

We evaluate the 5pt and the 8pt algorithm on random samples of 50 points, using normalization and equilibration (the `-hartley` and `-muehlich` variants), as well as without conditioning (the `-standard` variant). The experiments in subsections 6.2 and 6.3, as well as each datapoint in subsection 6.4 involve 10000 repetitions. The experiments were performed in Matlab and C++ using the implementations of the five-point algorithm provided at the web site of the authors³, and within the library VW34⁴ from the University of Oxford, respectively.

6.2. The bias of the 8pt algorithm

It has been shown in Section 3 that the standard 8pt algorithm is biased in the presence of degenerate or noisy data.

³<http://vis.uky.edu/~stewe/FIVEPOINT/>

⁴<http://www.doc.ic.ac.uk/~ajd/Scene/Release/vw34.tar.gz>

The bias concerns the epipole \mathbf{t} which tends to be near the optical axis regardless of the actual displacement. Some illustrative experimental results are summarized in Figure 2. The bias towards forward motion is clearly visible: the mode of the distributions is always near the z axis implying that the algorithm recovers forward motion for any motion. The figure shows that the bias is present in non-planar noisy contexts as well.

Equations (10) and (11) imply that the bias should disappear when the horizontal field of view approaches 101.5° . Figure 3 shows that the experimental results are in concordance with the theory. The bias also disappears in easier conditions (smaller noise, nearer target) or when normalized or equilibrated versions of the 8pt algorithm are used.

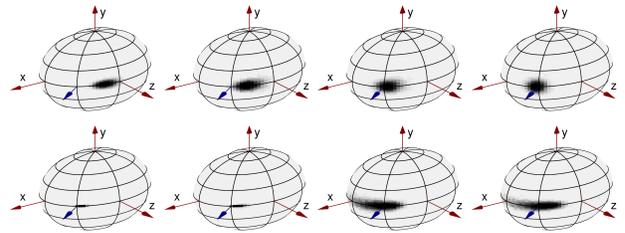


Figure 3. Frequency distributions of the epipole \mathbf{t} recovered by the 8pt algorithm. Common conditions: *distance*=10, *depth*=5, $\theta=135^\circ$, $\phi=0^\circ$, $\alpha_H=45^\circ$, $\sigma = 1$. Top: $\alpha_H=(60^\circ, 90^\circ, 100^\circ, 120^\circ)$. Bottom: $\sigma=0.2$, *distance*=3, normalization, equilibration.

Equilibrated and normalized 8pt algorithms perform quite similarly, except for approximately forward motion $|\theta| < 20^\circ$, where the advantage of the equilibrated algorithm is substantial for realistic noise levels. The advantage of the equilibrated algorithm is illustrated in Figure 4.

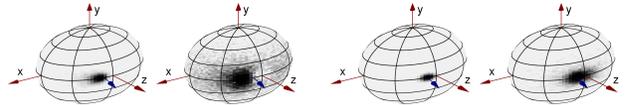


Figure 4. Frequency distributions of the epipole \mathbf{t} obtained by normalized (left group) and equilibrated (right group) 8pt algorithm, for $\sigma=0.5$ (left) and $\sigma=1.0$ (right). *distance*=10, *depth*=5, $\theta=170^\circ$, $\phi=0^\circ$, $\alpha_H=45^\circ$, $\sigma=1$.

6.3. The stability of the input to the 5pt algorithm

In order to evaluate sensitivity of the 5pt algorithm to poorly conditioned data, we analyze the stability of the right-singular vectors \mathbf{e}_i of the matrix \mathbf{A} (3). We consider the conditions in which the sensitivity is most likely to appear: sidewise motion ($\theta = 90$), considerable error ($\sigma = 1$) and a planar scene (*depth*=0). As an illustrative measure of deviation, we first calculate the deviations δ_i from the groundtruth vectors $\hat{\mathbf{e}}_i$ as:

$$\delta_i = \min(|\mathbf{e}_i - \hat{\mathbf{e}}_i|, |\mathbf{e}_i + \hat{\mathbf{e}}_i|). \quad (19)$$

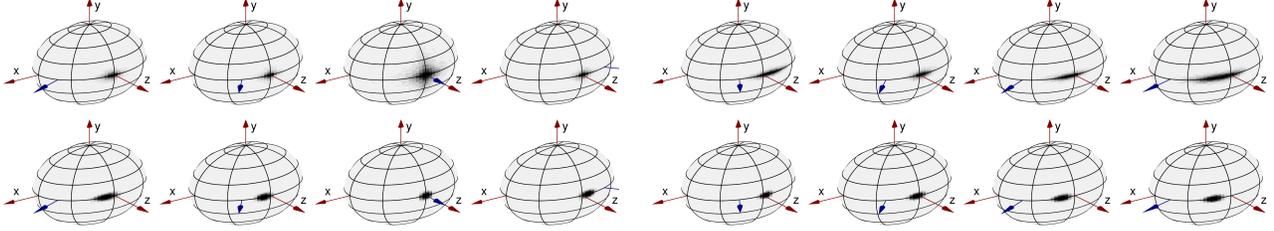


Figure 2. Frequency distributions of the epipole \mathbf{t} projected to the unit sphere, as recovered by the standard 8pt algorithm. The unlabeled arrow denotes the groundtruth \mathbf{t} . The shifted modes clearly reflect the forward bias both for planar scenes and moderate noise. Top: depth=0, $\sigma=0$. Bottom: depth=5, $\sigma=1$. Left: $\theta=(120^\circ, 150^\circ, 180^\circ, 210^\circ)$, $\phi=0^\circ$. Right: $\theta=135^\circ$, $\phi=(-20^\circ, -10^\circ, 10^\circ, 20^\circ)$. Backward motion has been specified ($|\theta| > 90^\circ$) in order to obtain \mathbf{t} with a positive z coordinate.

The results summarized in Figure 5 imply that the *span* of the last four vectors (being determined by the span of the first five vectors), is quite stable under noise, although the *individual vectors* are very sensitive. This explains why the 5pt algorithm is much more resistant to poorly conditioned data than the 8pt algorithm.

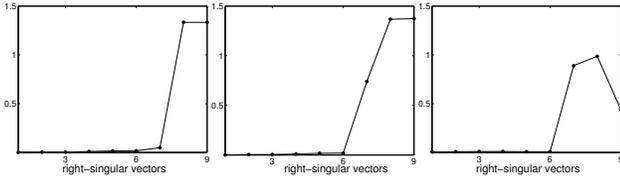


Figure 5. The deviations δ_i from (19) for default conditions $\theta = 90^\circ$, $\alpha_H = 45^\circ$, $\sigma = 1$, depth=5 (left), planar scene depth=0 (middle) and large field of view $\alpha_H = 120^\circ$ (right).

A more precise metric of the span deviation would be based on dot-products between $\hat{\mathbf{e}}_i$, $i = 1, 2, 3, 4, 5$ and \mathbf{e}_i , $i = 1, 2, 3, 4$, and vice versa. Denote the matrix of right-singular vectors as \mathbf{V} , so that $\mathbf{V} = [\mathbf{e}_i]$. Then the matrix of the dot-products can be expressed as $\mathbf{C} = \hat{\mathbf{V}}^\top \mathbf{V}$. The desired metric can be expressed as the sum of the Frobenius norms of the matrices \mathbf{D}_1 and \mathbf{D}_2 obtained as:

$$\begin{aligned} \mathbf{D}_1 &= \mathbf{C}_{[1:5, 6:9]} \\ \mathbf{D}_2 &= \mathbf{C}_{[6:9, 1:5]} \end{aligned} \quad (20)$$

The obtained values for $|\mathbf{D}_1|_F + |\mathbf{D}_2|_F$ are plotted against the horizontal field of view α_H in Figure 6, for \mathbf{e}_i obtained by standard, equilibrated and normalized procedures. The figure suggests that best results can be expected by Hartley’s normalization, until the horizontal field of view reaches 120° , when the stability of all three methods becomes roughly equal. Thus, the results confirm the previous finding that the data are much better conditioned at wider fields of view.

6.4. Performance evaluation

We evaluate the influence of the two conditioning schemes (17) and (15) on the accuracy of the recovered rela-

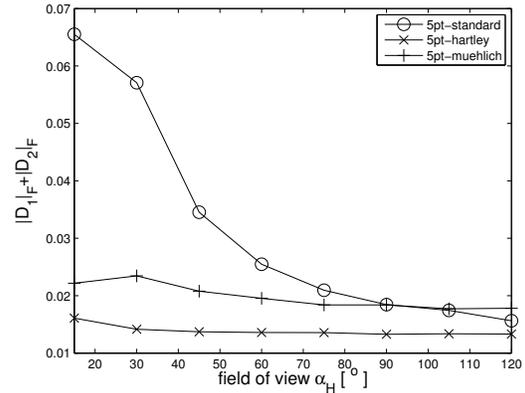


Figure 6. The dependence of $|\mathbf{D}_1|_F + |\mathbf{D}_2|_F$ from (20) on the field of view for the default conditions above.

tive pose. The experiments address the error in the recovered epipole Δt as the harder part [9, 10] of that task. The experiments consider the dependence of median Δt on the groundtruth translation direction.

6.4.1 Results for non-planar scenes

The results obtained for non-planar scenes (depth=5) are presented in Figures 7 and 8. The multiple hypotheses provided by the 5pt algorithm are disambiguated by looking at the overall reprojection error, using all correspondences. As suggested by Figure 5 in 6.3, the benefits of the conditioning are much more substantial in the case of the 8pt algorithm. As shown in Figure 4, equilibration significantly outperforms normalization for forward motion, while slightly inferior results are obtained for other motions. However, in the case of the 5pt algorithm, normalization has an edge over equilibration, as suggested by Figure 6. The results show that the 5pt algorithm is usually not the most consistent option in the overconstrained volumetric context, for common sets of parameters. This is not in concordance with claims in [10], where the conditioning of the 8pt algorithm has not been considered, especially for $\sigma < 1.0$ [13].

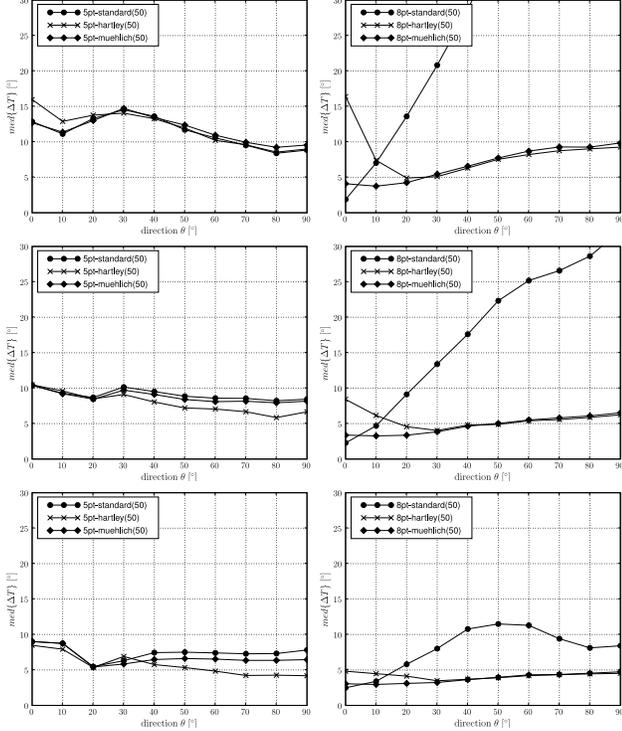


Figure 7. Angular epipole error Δt plotted against the translation direction θ , for the 5pt (left) and the 8pt (right) algorithm. $\sigma=0.5$, $\text{depth}=5$, $\alpha_H=30^\circ$ (top), 45° (middle) and 60° (bottom).

6.4.2 Results for planar scenes

In the context of planar scenes, the 5pt algorithm is compared with a specialized solution based on the decomposition of the planar homography (hg). Since the planar ambiguity [7] is more often present than not, we disregard the disambiguation issues by simply taking into account the best among the returned hypotheses. The results are summarized in Figure 9. The conditioning benefits are bigger than in the volumetric case. Despite the improvement, the homography approach performs better [13].

6.4.3 Results for other parameters of the setup

The experiments presented in the previous two paragraphs have been performed over the following parameter ranges: $\text{distance} \in \langle 2, 10 \rangle$, $\text{slant} \in \langle -30, 30 \rangle$, $\sigma \in \langle 0.2, 2.0 \rangle$, $\alpha_H \in \langle 15^\circ, 90^\circ \rangle$. The 8pt algorithm conditioning benefits are present in most parameter combinations. They are proportional to σ , inverse proportional to α_H , and vanish below $\sigma=0.2$ and above $\alpha_H=60^\circ$. Equilibration is substantially better than normalization for forward motion, while normalization is slightly better for sidewise motion. The difference between the two is again proportional to $\sigma \cdot \alpha_H^{-1}$.

The 5pt algorithm conditioning benefits are most pronounced for $|\theta| > 45^\circ$, and are *always* present in planar

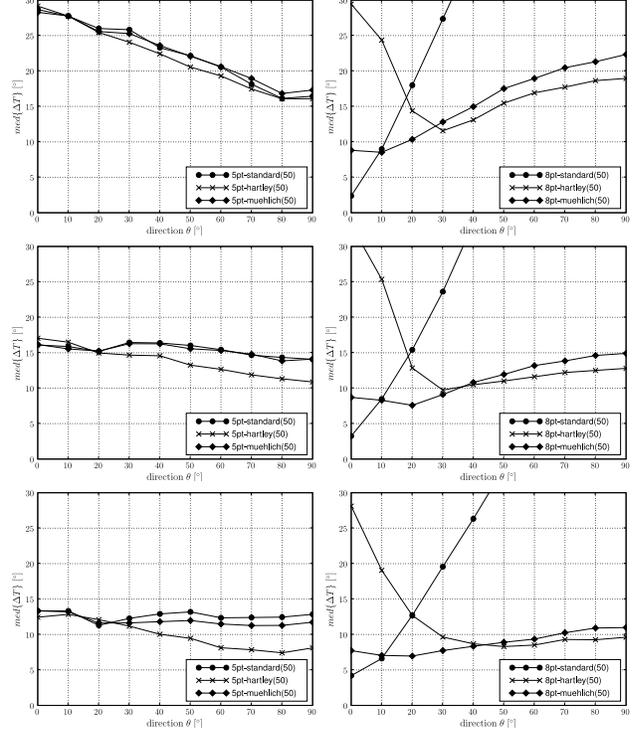


Figure 8. Angular epipole error Δt plotted against the translation direction θ , for the 5pt (left) and the 8pt (right) algorithm. $\sigma=1.0$, $\text{depth}=5$, $\alpha_H=30^\circ$ (top), 45° (middle) and 60° (bottom).

scenes. In volumetric scenes, the benefits start being noticeable at $\sigma=0.5$, but their growth stops at $\sigma > 1.0$. The benefits are present for $\alpha_H \in \langle 45^\circ, 60^\circ \rangle$, and vanish for $\alpha_H \in \{30^\circ, 90^\circ\}$. For narrow fields of view $\alpha_H < 30^\circ$, the conditioning actually worsens the results.

The variation trends for different values of distance are analogous to the variations in σ . The variations in depth produce similar results as the variations in σ^{-1} , however the 8pt algorithm is more affected than the 5pt algorithm. The variations in slant affect the 5pt algorithm and the homography more than the 8pt algorithm. Nevertheless, the qualitative overall relations are similar to what has been presented above.

7. Conclusion

The paper addressed the influence of the numerical conditioning schemes to the two prominent algorithms for recovering the relative pose. The contributions of the paper are (i) an original explanation to the forward bias of the standard 8pt algorithm, (ii) a study on the suitability of the normalization schemes for the 5pt algorithm, and (iii) performance evaluation in a rigorous artificial environment.

In the choice between 5pt and 8pt algorithms in the over-constrained non-planar case, there is a subtle trade-off. The 5pt algorithm relies on the stability of the upper 5 right-

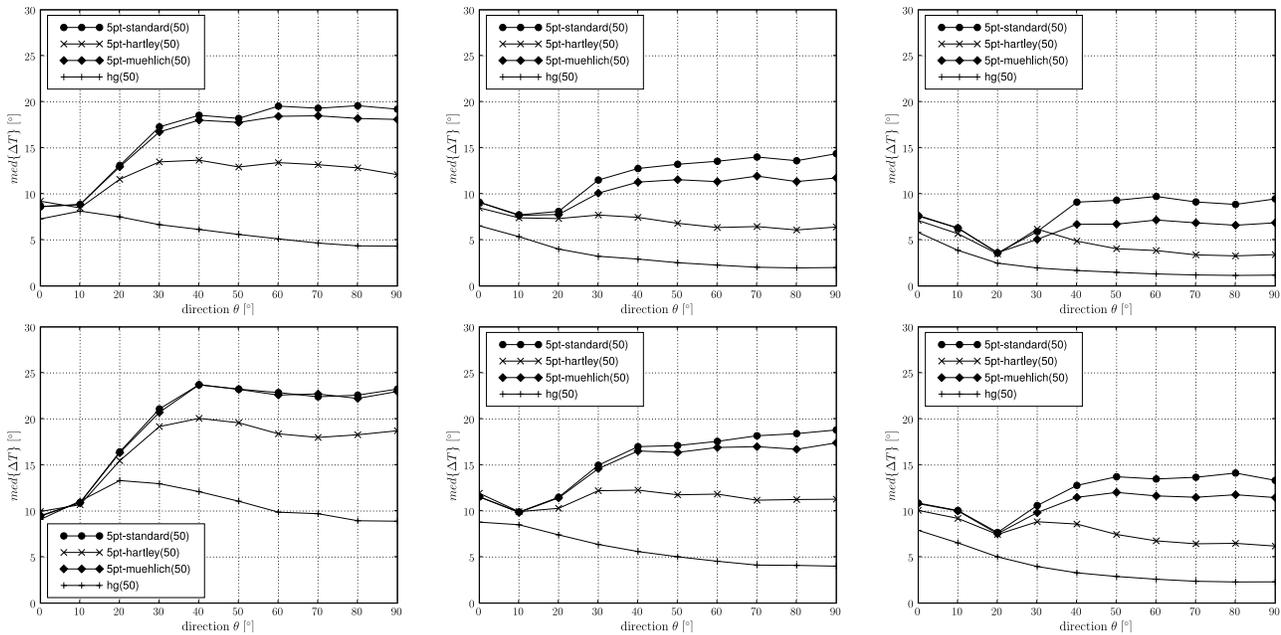


Figure 9. Angular epipole error Δt plotted against the translation direction θ , obtained with the 5pt algorithm and homography decomposition. The scene is planar, $\text{depth}=0$. $\alpha_H = 30^\circ$ (left), $\alpha_H = 45^\circ$ (middle) and $\alpha_H = 60^\circ$ (right). $\sigma=0.5$ (top), 1.0 (bottom).

singular vectors and internal constraints of the essential matrix. On the other hand, the 8pt algorithm relies on the lowest right-singular vector, which becomes much stabler with normalization. The experiments suggest that the recovery of the lowest singular vector has an advantage which becomes lower with noise: in the considered setup for $\alpha_H=45^\circ$, the break-even point is above $\sigma=1$ pixel of a 384×288 image.

The experimental results suggest that the choice of the algorithm for recovering the relative pose is highly context dependent. It therefore seems that a general solution should make an attempt to choose the best among the three options (8pt, 5pt, hg). The design of the appropriate tests in the calibrated context is an open area for future research.

References

- [1] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. Revisiting Hartley’s normalized eight-point algorithm. *IEEE Trans. PAMI*, 25(9):1172–1177, 2003. 1, 3, 4
- [2] O. D. Faugeras and S. Maybank. Motion from point matches: multiplicity of solutions. *Int. J. Comput. Vis.*, 4(3):225–246, 1990. 1, 2
- [3] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. PAMI*, 19(6):580–593, 1997. 1, 2, 4
- [4] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2004. 2, 3, 4, 5
- [5] B. K. P. Horn. Relative orientation. *Int. J. Comput. Vis.*, 4(1):59–78, 1990. 1
- [6] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981. 1, 2
- [7] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Proc. R. Soc. London*, B227(1249):399–410, 1986. 2, 3, 7
- [8] Y. Ma, S. Soatto, J. Kořecká, and S. Sastry. *An Invitation to 3-D Vision: From Images to Geometric Models*. Springer-Verlag, New York, USA, 2004. 1
- [9] M. Mühlich and R. Mester. The role of total least squares in motion analysis. In *Proc. of ECCV*, pages 305–321, Freiburg, Germany, 1998. 1, 2, 3, 4, 6
- [10] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. PAMI*, 26(6):756–770, 2004. 1, 2, 3, 4, 6
- [11] D. Nistér. Preemptive ransac for live structure and motion estimation. *Mach. Vision. Appl.*, 16(5):321–329, 2005. 2
- [12] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 1993. 2
- [13] S. Šegvić, G. Schweighofer, and A. Pinz. Performance evaluation of the five-point relative pose with emphasis on planar scenes. In *Proc. of AAPR/ÖAGM*, Schloss Krumbach, Austria, May 2007. accepted for publication. 2, 6, 7
- [14] P. H. S. Torr and A. W. Fitzgibbon. Invariant fitting of two view geometry. *IEEE Trans. PAMI*, 26(5):648–650, 2004. 1
- [15] F. Wu, Z. Hu, and F. Duan. 8-point algorithm revisited: Factorized 8-point algorithm. In *Proc. of ICCV*, pages I: 488–494, 2005. 1