Image and Vision Computing 29 (2011) 653–665

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Maximum likelihood autocalibration ☆

Stuart B. Heinrich [a,*], Wesley E. Snyder [a], Jan-Michael Frahm [b]

[a] Department of Electrical and Computer Engineering, NC State University, Box 7911, Raleigh, NC 27695-7911, United States
[b] Department of Computer Science, Campus Box 3175, Brooks Computer Science Building, 201 South Columbia Street, UNC-Chapel Hill, Chapel Hill, NC 27599-3175, United States

ARTICLE INFO

Article history:
Received 24 December 2010
Received in revised form 23 April 2011
Accepted 11 July 2011

Keywords:
Autocalibration
Self-calibration
Maximum likelihood
Absolute dual quadric
Structure from motion

ABSTRACT

This paper addresses the problem of autocalibration, which is a critical step in existing uncalibrated structure from motion algorithms that utilize an initialization to avoid the local minima in metric bundle adjustment. Currently, all known direct (not non-linear) solutions to the uncalibrated structure from motion problem solve for a projective reconstruction that is related to metric by some unknown homography, and hence a necessary step in obtaining a metric reconstruction is the subsequent estimation of the rectifying homography, known as autocalibration. Although autocalibration is a well-studied problem, previous approaches have relied upon heuristic objective functions, and have a reputation for instability. We propose a maximum likelihood objective and show that it can be implemented robustly and efficiently and often provides substantially greater accuracy, especially when there are fewer views or greater noise.

1. Introduction

This paper addresses the general problem of reconstructing, from a collection of corresponding image points identified in uncalibrated images, all of the camera parameters (position, orientation, focal length, etc.) and 3D coordinates of points in the scene. This problem is most often referred to as uncalibrated structure from motion (SFM).

A metric reconstruction is one that differs from the true configuration only by the choice of coordinate system; in other words, there is some unknown rotation, translation and scale [1]. It is well known that metric reconstruction is not possible from projective constraints alone [1–3] because the solution is ambiguous up to multiplication by some arbitrary homography. Thus, a reconstruction obtained from projection constraints alone is referred to as a projective reconstruction.

Given any additional constraints on the intrinsic camera parameters (e.g., that the images are not skewed, that pixel aspect ratio is known, that the center of projection is in the center of the image, or that multiple images were produced by the same physical camera), the ambiguity can be resolved. In practice, some of these constraints will always be available. However, incorporating these constraints directly into an initial estimate is difficult: an efficient solution is only possible in the simplest minimal case of two views with fully calibrated cameras [4]. A more general solution for two uncalibrated cameras was recently proposed in Hartley and Kahl [5], although the time complexity of this latter solution was prohibitively high.

In contrast, techniques for computing a projective reconstruction are much more efficient, so the usual approach is to compute an initial projective reconstruction minimally or linearly. This initial projective solution can be refined to a maximum likelihood (ML) projective reconstruction using projective bundle adjustment [1,6,7], which can be multiplied by the rectifying homography to yield an initial metric solution, and finally refined to a maximum likelihood metric reconstruction using metric bundle adjustment. The estimation of the rectifying homography is known as autocalibration (a.k.a. self-calibration).

In projective bundle adjustment, camera views are parameterized by projection matrices and the projection equation is simple, with the only nonlinearity being due to the perspective division. The basin of attraction for projective bundle adjustment is relatively large, and convergence is fast and reliable. In contrast, the presence of rotation matrices significantly complicates the projection equation in metric bundle adjustment, making the linearized update approximations less accurate. As a result, we observe a much smaller basin of attraction with less reliable convergence in metric bundle adjustment. Thus, even when an initial metric estimate is available, one may still prefer to do bundle adjustment in projective space in order to avoid local minima, and this would necessitate the use of autocalibration to map the result back into a metric space.

A plethora of approaches to autocalibration have been presented in the literature (see Section 1.1), but autocalibration has a reputation for instability, and obtaining robust results in the presence of realistic levels of measurement noise can often be difficult. This has motivated a recent trend towards approaches that use more computationally expensive global optimization methods, under the assumption that the instabilities are due to getting stuck in local minima. However, we

☆ This paper has been recommended for acceptance by Peter Sturm.
* Corresponding author. Tel.: +1 802 922 0731.
E-mail address: sbheinri@ncsu.edu (S.B. Heinrich).

0262-8856/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2011.07.003

will show that the heuristic objectives that are optimized by these global approaches are still fundamentally sensitive to noise.

In this paper, we formulate a maximum likelihood objective for autocalibration and show that it can be optimized efficiently and robustly. Using the maximum likelihood method avoids the sensitivity to noise and can give considerably more accurate results, especially for small numbers of views or high levels of relative measurement error (equivalent to more distant point clouds) where the autocalibration problem is more difficult.

We begin by summarizing previous autocalibration approaches in Section 1.1, and then derive our maximum likelihood objective in Section 2. A method for efficient optimization is described in Section 3. In Section 4 we devise a framework for objective evaluation of autocalibration performance, and identify a set of representative autocalibration algorithms to compare against. The results of our experiments are presented in Section 5, demonstrating the efficiency, robustness and quality of the proposed ML method, with some concluding remarks in Section 6.

Finally, we refer the interested reader to the appendix, where we explain the geometric relationships between all previous autocalibration constraints in Appendix B, and show that all previous constraints are enforced by our ML method. In Appendix A we provide insight into the fundamental instability and limitations behind the heuristic maximum a priori objectives (not to be confused with maximum likelihood), which are considered the current state of the art.

1.1. Background

The first known method of autocalibration was based on the Kruppa equations [8–11], now understood to be an algebraic representation of the correspondence of epipolar lines tangent to the dual image of the absolute conic (DIAC).

It was shown in [12] that an equivalent constraint to the Kruppa equations is that the essential matrix between any view pair must have two equal non-zero singular values, called the rigidity constraint. This is the fundamental principle behind several autocalibration approaches that theoretically work for two views when focal length is the only unknown [3,13–16], although they are highly sensitive to noise.

When more than two views are considered, autocalibration via the Kruppa equations requires finding the simultaneous solutions to many quadratic equations, which has not been regarded as a promising approach [1], but has been attempted using homotopy continuation [17], nonlinear methods [18,19], and more recently using globally convergent interval analysis [20]. Because the Kruppa equations do not enforce all of the calibration constraints that are now understood, such as the common support plane for the plane at infinity, these methods are subject to singularities that can lead to instabilities.

In [21], the modulus constraint on the plane at infinity was introduced, which is complementary to the Kruppa equations because it enforces constraints on the common plane at infinity without enforcing constraints on the DIAC. A unifying framework for these entities was presented with the absolute dual quadric (ADQ) [22], a fixed entity in space that encodes for both the plane at infinity and absolute dual conic (ADC) and projects to the DIACs. The ADQ is useful because all autocalibration constraints can be translated onto it.

The ADQ can be estimated using linear and nonlinear least squares [1,22–25], sometimes weighted according to prior assumptions as in [26]. Unfortunately, both of these variations are often unstable in practice [14]. It has been commented [27] that the main reason for instability of the linear method is that the rank and positive-semidefinite constraints of the ADQ are not enforced. However, we believe that the greater issue with the linear method is that the constraint equations do not directly correspond to the parameters they are intended to constrain in the presence of noise.

The nonlinear method has no singularities and enforces all known constraints, but still does not have any geometric meaning [1] and frequently produces unstable results in practice. We speculate from the recent trend towards more global approaches that minimize essentially the same cost function that the instability of the nonlinear method has been largely attributed to difficulties in obtaining a good initialization.

For example, in Hartley's stratified approach [28], chirality constraints [29] are used to solve for a finite bounding volume for the plane at infinity and then this space is explored with a brute force search. From each candidate location, the infinite homography constraint is used to linearly estimate the ADC from any desired calibration constraints, the best plane is taken as the one that minimizes the least squares residual, and finally the result is improved nonlinearly. Unfortunately, this brute force search can be slow, and we have observed that the minimum is often so pointlike that the basin of attraction is not reliably found using any reasonably spaced discretization. Additionally, it has been pointed out [30] that a single outlier can cause the chirality constraints to have no solution, or to not contain the correct solution.

More recently, the issue of discretization has been addressed by globally convergent methods. For example, interval analysis (IA) with branch and bound was used to minimize a heuristic based on the essential matrix constraint in Fusiello et al. [20]. Unfortunately, the method was not very efficient, having computation times of about 1.5 h for a problem with 40 views. IA was used again in [27], but the parameterization that was used only works for constant focal length and does not evenly distribute error. Computation times were improved in this latter method, but were still on the order of a minute for 20 views, which is too slow for many applications.

Under the constraint of zero skew (which can always be assumed in practice) and known principal point (which can be guessed but is often not known exactly), semidefinite programming was used to globally minimize a heuristic cost function in Agrawal [31], which was extended with a brute force search for principal point in Agrawal [32]. These methods enforced the internal ADQ constraints, but neglected the constraints on aspect ratio and always assumed that principal point is constant, which makes them applicable to video but not photo collections.

Convex relaxation was used with branch and bound to identify the plane at infinity that globally minimizes a heuristic cost associated with the modulus constraint in the recent stratified approach of Chandraker et al. [33,34],but the heuristic is not ideal because it does not consider constraints on the DIAC in the search for the plane at infinity. Similar techniques were used to estimate the ADQ directly using all known constraints in Chandraker et al. [35], which makes it perhaps the most generally applicable global approach.

In general, the globally convergent approaches are very difficult to implement and not very efficient. As an alternative, the dual stratified approach of estimating the plane at infinity from known calibration matrix, first proposed in Bougnoux [14], has recently been revived with a closed form solution from a view pair in Gherardi and Fusiello [36]. The advantage of the dual stratified approach is that prior knowledge may be used to restrict the search into a very narrow plausible region, rather than exhaustively searching through all of parameter space for the plane at infinity. This leads to an algorithm that is simple, fast and robust. However, it lacks in precision, and still minimizes a heuristic objective that is not geometrically meaningful. As a result, attempting to further minimize the heuristic using nonlinear methods can result in divergence.

The fundamental limitation of all previous algorithms is that the objective being minimized is a heuristic with no particular geometric meaning, and these heuristics do not always work as well as one would hope. This becomes especially apparent for projective reconstructions with greater noise (or equivalently, more distant geometry) and for short reconstructions which are commonly on

the verge of a Critical Motion Sequence (CMS) [27] for which the problem is ill-posed.

Currently, the most meaningful heuristic objectives minimize a weighted sum of squared errors between the rectified intrinsic parameters and an assumed mean of zero. If calibration parameters are all independent and normally distributed, then the homography that minimizes this error is a maximum a priori rectifying homography [30], where the weights (chosen heuristically) implicitly correspond to inverse variance of some assumed prior distribution. However, the tenability of this prior model has never been justified, and it has been primarily adopted out of convenience as a substitute for likelihood. In the next section, we show that in fact likelihood can be used as the objective, and we will present an algorithm to maximize likelihood robustly and efficiently.

2. Maximum likelihood autocalibration

Consider a set of n homogeneous structure points $\overline{\mathbf{X}}_i, i = 1\ldots n$ in the projective space $\mathbb{P}^3$, viewed by a set of m cameras having 3×4 projection matrices

$$\overline{\mathbf{P}}^j = \overline{\mathbf{K}}^j\left[\overline{\mathbf{R}}^j|\overline{\mathbf{t}}^j\right], j = 1\ldots m, \tag{1}$$

where $\overline{\mathbf{R}}^j$ is a rotation matrix and $\overline{\mathbf{K}}^j$ is a non-singular upper triangular calibration matrix with positive diagonal elements. Note that any projection matrix can be uniquely factored into these components [37]. We refer to the combined set of this information as the *true configuration*, denoted by

$$\overline{\Theta} = \left\{\overline{\mathbf{X}}_i, \overline{\mathbf{P}}^j | \forall i,j\right\}, \tag{2}$$

and any estimate $\widehat{\Theta}$ of the configuration from some measurements as a *reconstruction* of the configuration.

The perspective projection of a homogeneous structure point $\mathbf{X}\in\mathbb{P}^3$ as viewed by a camera with projection matrix $\mathbf{P}$ is accomplished by multiplication, yielding a homogeneous image point $\mathbf{x}\in\mathbb{P}^3$,

$$\mathbf{x}\propto\mathbf{PX}. \tag{3}$$

Let the measured coordinates of the image of the ith structure point in the jth image be denoted by $\tilde{\mathbf{x}}_i^j$. If we assume, as is commonly done [38], that measurement error is normally distributed with standard deviation σ, then the probability (or likelihood) of a measurement is

$$P\left(\tilde{\mathbf{x}}_i^j|\overline{\Theta}\right) = \frac{1}{2\pi\sigma^2}\exp\left(-d\left(\tilde{\mathbf{x}}_i^j,\overline{\mathbf{x}}_i^j\right)^2/\left(2\sigma^2\right)\right), \tag{4}$$

where $\overline{\mathbf{x}}_i^j$ is the true image of $\overline{\mathbf{X}}_i$ in the jth view, and $d(\mathbf{a},\mathbf{b})$ is the Euclidean distance between the inhomogeneous points represented by homogeneous points $\mathbf{a}$ and $\mathbf{b}$. The log-probability of a measurement is

$$\log P\left(\tilde{\mathbf{x}}_i^j|\overline{\Theta}\right) = -\frac{1}{2\sigma^2}d\left(\tilde{\mathbf{x}}_i^j,\overline{\mathbf{x}}_i^j\right)^2 + \underbrace{\log\left(1/\left(2\pi\sigma^2\right)\right)}_{constant}, \tag{5}$$

and therefore the maximum likelihood (ML) projective reconstruction $\widehat{\Theta}_{ML}$ from measurements $\left\{\tilde{\mathbf{x}}_i^j\right\}$ is given by

$$\widehat{\Theta}_{ML} = \underset{\Theta}{\operatorname{argmax}}\prod_{i,j}P\left(\tilde{\mathbf{x}}_i^j|\Theta\right) \tag{6}$$

$$= \underset{\Theta}{\operatorname{argmax}} -\frac{1}{2\sigma^2}\sum_{i,j}d\left(\tilde{\mathbf{x}}_i^j,\mathbf{x}_i^j\right)^2 \tag{7}$$

$$= \underset{\Theta}{\operatorname{argmin}}\sum_{i,j}d\left(\tilde{\mathbf{x}}_i^j,\mathbf{x}_i^j\right)^2. \tag{8}$$

The distance $d\left(\tilde{\mathbf{x}}_i^j,\mathbf{x}_i^j\right)$ is known as *reprojection error*, so the ML reconstruction can be found by minimizing the sum of squared reprojection errors. This nonlinear minimization is known as *bundle adjustment* [1,6,7], and can be parameterized either in terms of projection matrices, or more explicitly in terms of camera calibration, rotation, and translation matrices.

Without any additional constraints, it is well known that the solution is ambiguous up to some arbitrary homography. This may be easily recognized by observing that, from any reconstruction $\{\mathbf{P}^j, \mathbf{X}_i|\forall i,j\}$, an alternative reconstruction with the same likelihood is given by choosing any invertible $\mathbf{H}$ in

$$\mathbf{x}_i^j\propto\mathbf{P}^j\mathbf{X}_i\propto\left(\mathbf{P}^j\mathbf{H}\right)\left(\mathbf{H}^{-1}\mathbf{X}_i\right)\propto\mathbf{P}^{j\prime}\mathbf{X}_i^{\prime}, \forall i,j. \tag{9}$$

A reconstruction that suffers from this ambiguity is known as a projective reconstruction. Autocalibration is an attempt to resolve this ambiguity by finding a rectifying homography that removes the projective distortion by bringing a projective reconstruction to within a similarity transformation of the true configuration.

As will be discussed in Section 2.1, there are additional (partially problem-dependent) constraints on the calibration matrices of real cameras. Previous autocalibration algorithms have attempted to resolve the ambiguity by minimizing some heuristic objective based on these calibration constraints, or more formally by assuming a prior distribution for calibration parameters and then maximizing prior probability. However, this maximum a priori approach relies on several flawed assumptions that make it highly sensitive to noise, which we discuss in detail in Appendix A.

It is normally thought that likelihood cannot be used as an objective for autocalibration, because from Eq. (9), it is clear that multiplication by a homography does not change likelihood. However, the calibration constraints actually reduce the degrees of freedom (DOF) of camera views: whereas a general projection matrix has 11 DOF, after taking into account calibration constraints, this is typically reduced to just 7 DOF for pose and focal length.

This fact has several consequences. First, there is the pedantic recognition that any general reconstruction parameterized in terms of projection matrices (or equivalently, parameterized by fully general calibration matrices) will violate the calibration constraints, and hence does not represent a plausible reconstruction. Thus, minimizing reprojection error under such a parameterization does *not* yield a true maximum likelihood estimate.

A second, more practical consequence, is that if one uses a parameterization that admits only plausible solutions (i.e., enforces calibration constraints), then the maximum likelihood solution does *not* suffer from the projective ambiguity (i.e., it is a metric reconstruction).

Because a projective reconstruction has too many DOF, there will not generally exist any homography that will rectify a projective reconstruction and cause all calibration constraints to be exactly satisfied. Thus, in order to obtain a plausible metric reconstruction from a projective reconstruction, it will always be necessary to 'project' the solution from the implausible space to the lower dimensional plausible space *after* multiplying by the rectifying homography.

This projection step increases reprojection error and is affected by the choice of rectifying homography. Therefore, we propose to seek the homography that leads to a reconstruction with maximal likelihood. Specifically, the maximum likelihood rectifying homography is given by

$$\widehat{\mathbf{H}}_{ML} = \underset{\mathbf{H}}{\operatorname{argmin}}\sum_{i,j}d\left(\tilde{\mathbf{x}}_i^j,\mathbf{P}_C^j\mathbf{H}^{-1}\mathbf{X}_i\right)^2, \tag{10}$$

where $\left\{\mathbf{P}_C^j|\forall j\right\}$ are the 'closest' projection matrices to $\left\{\mathbf{P}^j\mathbf{H}|\forall j\right\}$ that are *plausible* (i.e., exactly satisfy the problem-dependent metric constraints). There are multiple ways for accomplishing this projection into plausible space which we will discuss in Section 2.2.

Considering that previous methods have attempted to maximize prior probability, and we have shown that it is also possible to maximize likelihood, it is natural to wonder if it would not be better to devise a maximum *a posteriori* (MAP) objective that takes into account *both* prior probability and likelihood. However, as discussed in Appendix A, the prior distribution model that was assumed by previous approaches is not actually an accurate representation of the available prior knowledge, which for the most part takes the form of hard constraints that are already exactly satisfied by the projection step in our ML approach.

With that said, one could still form meaningful prior distributions for focal length and/or principal point, but our analysis of variance has shown that the variance in these parameters of a projective reconstruction is dominated more by propagated measurement noise than actual uncertainty in the true configuration. This propagated variance depends on the unknown camera configuration in metric space, and hence there is no way to accurately predict the scale in the prior model. Even if the scale were known, there would be little benefit because of the large degree of uncertainty in these parameters.

We conclude this section by reminding the reader that: (a) a projective reconstruction that maximizes likelihood (i.e., minimizes reprojection errors) is not a true ML reconstruction, although we refer to it as a 'ML projective reconstruction' in keeping with existing literature, and for lack of a better term; (b) the ML rectifying homography for any arbitrary projective reconstruction takes that reconstruction as close as possible to the ML reconstruction, although there are not enough DOF to actually obtain a ML reconstruction by multiplication with any homography; (c) the only practical method for obtaining a ML reconstruction is by using metric bundle adjustment with a parameterization that takes into account all problem-dependent calibration constraints, although this is a non-linear method that requires a good initialization to avoid local minima (hence, the need for projective reconstruction and autocalibration).

2.1. Metric constraints

Theoretically, any constraints or prior assumptions about the true configuration could be exploited by an autocalibration algorithm in order to assist in resolving the projective ambiguity and obtaining a metric reconstruction. For example, the fact that real cameras can only image what is in front of them results in an inequality constraint for each measured image point. These inequality constraints, known as *chirality constraints* [29], can be used to restrict the location of the plane at infinity (which partially defines the rectifying homography) to a convex polytope [28].

Not much more can be said about the point cloud or camera poses without some loss of generality, although assumptions of this sort are sometimes used. For example, if the images are known to come from a high frame-rate video source, the change in camera position and rotation between successive frames may be assumed small, and this assumption was found useful in Alcantarilla et al. [39]. Additionally, one might assume that the reconstructed point cloud should have some significant volume (i.e., it should not be embedded in a line or plane), and should occupy some relatively compact volume of space. These assumptions were implicitly used in the stratified method of Hartley et al. [28] by normalizing the principal components to make the reconstructed point cloud roughly spherical.

In contrast, much more definitive assumptions can be made about the calibration matrices, because these represent intrinsic parameters of the cameras independent of pose. The geometric meaning of each calibration parameter is made clear by the parameterization of Faugeras [40],

$$\mathbf{K} = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u \\ & \alpha_v / \sin\theta & v \\ & & 1 \end{bmatrix}, \tag{11}$$

where $\theta \in (0, \pi)$ is the image skew angle, α_u and α_v are the image scaling factors in the skewed basis, and (u, v) are the coordinates of the principal point. This can be reduced to the simpler parameterization used by Hartley and Zisserman [1],

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & u \\ & \alpha_y & v \\ & & 1 \end{bmatrix}, \tag{12}$$

where $s \in (-\infty, \infty)$ is the skew parameter and α_x and α_y are the horizontal and vertical image scaling factors. Additionally, one may write $\alpha_x = f$ and $\alpha_y = fr$, where f is focal length and r is pixel aspect ratio.

Because the images taken by any real camera will be unskewed and have a known (usually 1:1) pixel aspect ratio (i.e., the images are generally not stretched or squished either horizontally or vertically), this means that in the true configuration, $s = 0$ and $\alpha_x = \alpha_y$. When it is known that the principal point is in the center of the image, the image coordinate system can be chosen so that $u = v = 0$. For any two projection matrices that are known to correspond to the same camera, the calibration parameters will be equal (except perhaps for focal length, which might be assumed to be variable).

We propose an additional inequality constraint based on the fact that a camera field of view must be in the range $(0, \pi)$, and in practice a much less conservative range can usually be assumed, which we denote $(\theta_{min}, \theta_{max})$. Thus, focal length must be in the range

$$f \in \left(\frac{S_w}{2 \tan\left(\frac{\theta_{max}}{2}\right)}, \frac{S_w}{2 \tan\left(\frac{\theta_{min}}{2}\right)} \right), \tag{13}$$

where S_w is the width of the image in pixels. For a standard 35 mm camera using an 18–55 mm lens, this means that focal length is approximately in the range of 1–3 image widths.

Several other constraints can be found in the literature on autocalibration, but it is important to understand that any additional constraints can all be derived from the above-mentioned calibration constraints (see Appendix B). This includes the rank (i.e., common plane at infinity π_∞) and positive-semidefinite constraints of the absolute dual quadric $\mathbf{Q}^*_\infty$ [22], the linear and nonlinear constraints on $\mathbf{Q}^*_\infty$ [23], the infinite homography constraint on the absolute dual conic Ω^*_∞ [1], the Kruppa equations [8], the essential matrix and rigidity constraints [12], and the modulus constraint [41]. Thus, although Eq. (10) is not specifically formulated in terms of these previously used constraints, it does not neglect any of them.

2.2. Projection into plausible space

Given a rectifying homography $\mathbf{H}$, we have identified two reasonable ways to accomplish the 'projection' of an implausible projective reconstruction into the space of plausible metric reconstructions. The simplest and most direct method is to factor the projection matrices [37] to extract calibration matrices,

$$\mathbf{P}^j\mathbf{H} \rightarrow \mathbf{K}^j \left[\mathbf{R}^j \,|\, \mathbf{t}^j \right] \forall j, \tag{14}$$

then minimally augment the calibration matrices to fit into the parameterization defined by the constraints of the plausible metric space (e.g., if skew is zero and pixel aspect ratio is unity, set $s \Leftarrow 0$ and $\alpha_y \Leftarrow \alpha_x$),

$$\left\{ \mathbf{K}^j \,|\, \forall j \right\} \rightarrow \left\{ \mathbf{K}'^j \,|\, \forall j \right\}, \tag{15}$$

and then reform the projection matrices using the augmented calibration matrices,

$$\mathbf{P}_C^j \leftarrow \mathbf{K}^j \left[\mathbf{R}^j \,|\, \mathbf{t}^j\right] \forall j. \tag{16}$$

Although this direct projection is simple and efficient, it does not necessarily yield the projection matrices that minimize reprojection error. Thus, it may sometimes be preferable to resection the set of cameras from the rectified structure point cloud using a parameterization that implicitly enforces calibration constraints. That is,

$$\left\{\mathbf{P}_C^j \,|\, \forall j\right\} = \underset{\left\{\mathbf{K}^j\left[\mathbf{R}^j \,|\, \mathbf{t}^j\right] \,|\, \forall j\right\}}{\operatorname{argmin}} \sum_{i,j} d\left(\tilde{\mathbf{x}}_i^j, \mathbf{K}^j \left[\mathbf{R}^j \,|\, \mathbf{t}^j\right] \mathbf{H}^{-1} \mathbf{X}_i\right)^2, \tag{17}$$

where the $\mathbf{K}^j$ are parameterized to enforce calibration constraints. If the calibration matrices are independent (that is, if they do not correspond to the same physical camera), then the cameras can be resectioned independently and Eq. (17) becomes equivalent to

$$\mathbf{P}_C^j = \underset{\mathbf{K}^j\left[\mathbf{R}^j \,|\, \mathbf{t}^j\right]}{\operatorname{argmin}} \sum_i d\left(\tilde{\mathbf{x}}_i^j, \mathbf{K}^j \left[\mathbf{R}^j \,|\, \mathbf{t}^j\right] \mathbf{H}^{-1} \mathbf{X}_i\right)^2 \forall j. \tag{18}$$

The minimization of Eq. (17) or Eq. (18) can be carried out using Levenberg–Marquardt [42]. If one assumes that images are taken by different cameras, and that the center of projection is in the center of the images, then each camera can be parameterized by 1 parameter for focal length, 4 parameters for quaternion rotation, and 3 parameters for position.

3. Implementation

As discussed in Section 2.1, the constraints that can be incorporated into Eq. (10) depend upon the type of problem and ones knowledge of the cameras that were used (if it is a photo collection, video, etc.). However, the most generally available constraints, which apply to all cameras that take unskewed and unstretched images, are that skew is zero and pixel aspect ratio is unity. Therefore, we parameterize calibration matrices in the plausible metric space by

$$\mathbf{K}^j = \begin{bmatrix} f^j & 0 & u \\ & f^j & v \\ & & 1 \end{bmatrix}. \tag{19}$$

When projecting into this parameterization we prefer to use the direct method from Section 2.2 for reasons of efficiency. We further incorporate chirality constraints by modifying the objective function of Eq. (10) to minimize

$$\gamma \sum_{i,j} \delta_i^j + \sum_{i,j} d\left(\tilde{\mathbf{x}}_i^j, \mathbf{P}_C^j \mathbf{H}^{-1} \mathbf{X}_i\right)^2, \tag{20}$$

where γ is a penalty cost and δ_i^j is an indicator function that is 1 when the ith point is *behind* the jth camera, and zero otherwise. Using $\gamma = \infty$ would cause strict enforcement of the chirality constraints, but this could introduce problems because, as pointed out by Nister [30], a single outlier can cause the chirality constraints to have no solution, or for the solution polytope to not contain the true plane at infinity. Therefore, in order to remain robust to these potential outliers, we prefer to use a finite penalty of $\gamma = 100$.

In order to minimize Eq. (20) we have designed the Dual Stratified Monte Carlo Maximum Likelihood (DS-MC-ML) autocalibration algorithm (Algorithm 1). We obtain a first guess estimate of **H** by linearly estimating $\mathbf{Q}_\infty^*$ using the method of Pollefeys et al. [23] (line 1). This initial linear estimate is efficient and provides an exact solution in the absence of noise, although we do not expect the estimate to be accurate under realistic levels of noise.

Therefore, our primary method of initialization is to use a random sampling approach that takes advantage of the well-understood prior distribution for calibration matrices. This is done by randomly selecting two projection matrices $\mathbf{P}_1$ and $\mathbf{P}_2$ (line 4) and then randomly sampling a hypothesis for **K** from the prior distribution (line 5). For the prior distribution, we assume focal length is uniformly distributed in the plausible range (Eq. (13)) and that principal point is in the center of the image. If one has greater uncertainty in principal point, it can simply be sampled from a normal distribution centered around the image center.

We then use the closed form dual stratified (DS) autocalibration method of [36] to calculate an exact solution for **H** from the view pair using **K** (line 6). If the sampled **K** was correct, then **H** will also be correct. We evaluate the likelihood of each candidate homography (line 7) according to Eq. (20), and continue the random sampling process, keeping track of the solution with maximal likelihood.

Algorithm 1. DS-MC-ML Autocalibration

Require: A projective *reconstruction* from ≥2 views, and a prior model for the distribution of K.
Ensure: $\hat{\mathbf{H}}$ is the rectifying homography that takes the input reconstruction as close as possible to the ML metric reconstruction.
1: $\hat{\mathbf{H}} \leftarrow$ LinearAutocalibrate(*reconstruction*)
2: $\epsilon_{min} \leftarrow$ MeanProjectionError($\hat{\mathbf{H}}$, *reconstruction*)
3: **repeat**
4: $\{\mathbf{P}_1; \mathbf{P}_2\} \leftarrow$ select two projection matrices at random.
5: $\mathbf{K} \leftarrow$ sample from prior distribution.
6: $\mathbf{H} \leftarrow$ AutocalibratePair($\mathbf{P_1}$, **K**, $\mathbf{P_2}$, **K**)
7: $\epsilon \leftarrow$ MeanReprojectionError(**H**, *reconstruction*)
8: **if** $\epsilon < \epsilon_{min}$ **then**
9: $\hat{\mathbf{H}} \leftarrow \mathbf{H}$
10: **if** $\epsilon < \epsilon_{thresh}$ **then**
11: **break**
12: **end if**
13: $\epsilon_{min} \leftarrow \epsilon$
14: $count \leftarrow 0$
15: **else**
16: $count \leftarrow count + 1$
17: **end if**
18: **until** $count \geq stopCount$
19: $\hat{\mathbf{H}} \leftarrow$ NonlinearImprove($\hat{\mathbf{H}}$, *reconstruction*)

The beauty of sampling from the prior distribution is that the search space is inherently restricted to solutions that are, by definition, within the high density region of the assumed probability distribution. Also, as this is essentially only a 1-dimensional search for focal length within a restricted range, we do not expect a large number of random trials to be necessary before finding a very good solution.

However, the number of trials that are needed in order to find a solution that is in the basin of attraction of the global minimum is dependent on the relative pose of camera views, the number of views, and the assumed prior distribution. If the epipolar geometry is near a Critical Motion Sequence (CMS) [27], or if there is a large number of views, or if one uses a prior distribution with unusually large uncertainty, then more samples will be needed. Therefore, rather than choosing a fixed number of iterations up front, we prefer to use an adaptive strategy that terminates the sampling process after $stopCount = 300$ iterations have elapsed without further improvement (line 18). This allows the algorithm to remain efficient for simple problems, while naturally scaling up to use more iterations on more difficult problems.

In addition, we incorporate a threshold for early termination based on the mean reprojection error value (line 10). For example, if the mean reprojection error is less than $\epsilon_{thresh} = 1$ pixel, then the solution is already quite accurate and the search can be terminated. The ability to do early termination has not been possible using previous

autocalibration algorithms because the previously used heuristics have no geometric meaning like reprojection error.

Once the random sampling process has completed, we will have obtained a robust and fairly accurate estimate of the rectifying homography. From this estimate of **H** we further minimize Eq. (20) using Levenberg–Marquardt [42] with numerical differentiation. Although a reduced parameterization for **H** is possible (Eq. (B.12)), we prefer to use a free parameterization, except we fix the fourth column to $(0, 0, 0, 1)^T$ because it has no bearing on the result.

The final output of Algorithm 1 is a robust and precise estimate of the rectifying homography, but obtaining a metric estimate requires one final projection into plausible metric space. This final projection can be accomplished by again using the direct projection method, or by using the more computationally expensive resectioning method (Section 2.2), which results in lower reprojection error.

4. Experimental methods

We compare the proposed maximum likelihood method against several other representive autocalibration algorithms by generating synthetic configurations and then projecting structure points to obtain image points (a.k.a., image correspondences). The correspondences are corrupted by adding normally distributed noise, and then projective bundle adjustment is used to find the maximum likelihood projective reconstruction. Finally, we use each autocalibration algorithm to rectify the projective reconstruction and make an objective comparison to the true configuration.

4.1. Algorithms compared

Fundamentally, the rectifying homography **H** is encoded by the absolute dual quadric $\mathbf{Q}_\infty^*$, which is defined by the plane at infinity π_∞ and absolute dual conic Ω_∞^*. These relationships (explained in detail in Appendix B) define a natural categorization of autocalibration algorithms.

In linear methods, $\mathbf{Q}_\infty^*$ is estimated directly. In nonlinear methods **H** is improved nonlinearly. In stratified methods a brute force search is first used to identify π_∞ (after computing a finite bounding volume via chirality constraints) and then Ω_∞^* is estimated in a second phase. Finally, in the dual stratified method **K** (which is directly related to Ω_∞^*) is first guessed based on prior knowledge and then π_∞ is estimated in a second phase.

Therefore, we compare against representative algorithms from each category. The specific algorithms we compare against are:

Linear Method (L). We use the linear method of Hartley and Zisserman [1], Pollefeys et al. [23] to estimate $\mathbf{Q}_\infty^*$ using the symmetric parameterization of $\mathbf{Q}_\infty^*$. We have found that this is much more reliable than using the reduced parameterization. The constraints we use are based on the assumptions of zero skew, unit aspect ratio, and zero principal point, with experimentally determined weighting coefficients of $w_s = 1, w_r = 1, w_u = w_v = 0.2$.

Linear with Nonlinear Method (L + NL). A nonlinear improvement is also given in Hartley and Zisserman [1], Pollefeys et al. [23] which can be parameterized by the rectifying homography to implicitly enforce the rank and positive-semidefinite constraints. It is also necessary to parameterize $\omega^{*j} \forall j$ using some problem-dependent constraints. We found that this method has a tendency to become extremely unstable if the parameterization of ω^{*j} allows for nonzero principal point or varying focal length, so we use only a single parameter for constant focal length.

Stratified Method (S). We used Hartley's stratified approach [28] with a brute force search for π_∞ using the recommended discretization of $100 \times 100 \times 100$. Additionally, we augment the search space with one additional point representing the location of π_∞ that would be found by the linear algorithm. When solving for the absolute conic, we include constraints for zero skew, constant focal length, unit aspect ratio and constant principal point.

Stratified with Nonlinear Method (S + NL). Hartley [28] calls for a nonlinear improvement after the initial stratified search so we follow by using the method of Pollefeys et al. [23].

Dual Stratified Method (DS). We have used the method of Gherardi and Fusiello [36], with the recommended parameters of 50 samples for focal length, $w_{sk} = 1/0.01$, $w_{ar} = 1/0.2$, $w_{u0} = 1/0.1$, $w_{v0} = 1/0.1$.

Dual Stratified with Nonlinear Method (DS + NL). It is recommended in Gherardi and Fusiello [36] to follow up with a nonlinear improvement so we use the method of Pollefeys et al. [23].

(NEW) Maximum Likelihood Method (ML). We estimate the rectifying homography using DS-MC-ML (Algorithm 1).

(NEW) ML with Resection Method (ML + R). We estimate the rectifying homography using DS-MC-ML (Algorithm 1), and finally project the reconstruction into the plausible metric space by using the resectioning method (18).

4.2. Objective evaluation

A reconstruction includes structure points, camera poses, and intrinsic camera parameters. Due to the large number of different types of parameters, identifying a good objective evaluation can be challenging.

It is common to see autocalibration algorithms compared purely on the basis of how accurately specific intrinsic camera parameters have been reconstructed. However, each camera has 5 intrinsic parameters, and the relative importance of each parameter is unclear, so there is no objective way to combine the various intrinsic parameters into a single measure of reconstruction quality. One could report the error for each parameter independently, but usually a reduction of the error in any one parameter forces an increase in the error of some other parameter. Thus, there would still be no clear objective way for the reader to assess which method was better.

A much more objective evaluation is to look at the structure of the reconstruction, because all of the structure points and camera centers in a perfect reconstruction should be within a similarity transformation of the true configuration. Therefore, when the true configuration is known one can simply factor out this similarity transformation to align the two point clouds and then measure the sum of squared distances between them. A linear solution for finding the similarity transformation that minimizes sum of squared distance is given in Umeyama [43].

We have observed that all autocalibration algorithms tend to have much higher error in the reconstructed camera centers than in the reconstructed structure point cloud. Therefore, we use only the structure point cloud to compute the alignment, and then use the mean squared error of reconstructed camera centers as the objective error. Any error in the intrinsic camera parameters necessitates shifting the camera center to compensate (e.g., a reduction in focal length moves the cameras closer), and thus by measuring error in camera center we obtain a truly objective measure of reconstructed camera accuracy that is independent of the specific objectives being optimized by the autocalibration algorithm.

4.3. Experiments

We generate random configurations with 2000 structure points distributed uniformly on the surface of a cube of width 100 centered at the origin. There are 10 cameras arranged on a circle of radius 1500 at 10° increments, with a positional jitterof ± 10, looking at a random point in a cube of width 40 centered at the origin. All cameras have a constant focal length in the range of 600–800 (relative to an image size of 640×480), zero skew, unit aspect ratio and zero principal point.

For each level of noise, we generate a set of 100 random configurations, project the image points and perturb with additive gaussian noise using $\sigma=0$, $\sigma=1$, or $\sigma=3$ pixels. Then we run projective bundle adjustment to obtain the maximum likelihood projective reconstruction and attempt autocalibration.

The case of zero noise is not realistic, but validates that the algorithms have been implemented correctly. The case of $\sigma=1$ represents a realistic level of noise for a typical subpixel matching algorithm after outliers have been removed using MLESAC [44] or some other variation of RANSAC [45]. The case of $\sigma=3$ represents a larger level of noise that might be obtained by using less accurate multi-scale features, such as SIFT [46] features.

For each algorithm and for each noise level, we computed the Empirical Cumulative Distribution Function (ECDF) of the objective error measure after autocalibrating all 100 configurations.

5. Results

For the unrealistic case of $\sigma=0$ (i.e., the when there exists a homography that exactly rectifies the projective solution with no noise), we found that all methods performed extremely well, with the bulk of objective error (as described in Section 4.2) being approximately in the range of 10^{-27} to 10^{-15} for all algorithms. Because the stratified and dual stratified algorithms use discretized searches, their precision was worse before nonlinear improvement.

The ECDFs at a more realistic noise level of $\sigma=1$ pixel are shown in Fig. 1. Here we see that many of the previous algorithms have difficulty achieving robust and accurate results. It is clear that our ML+R and ML algorithms have superior precision and robustness to all of the other algorithms compared. The DS algorithm is a close runner up about 85% of the time, but gives unstable results the remaining 15% of the time. Surprisingly, the nonlinear improvement actually tends to worsen the DS algorithm at this level of noise, which we speculate is due to inherent ability of the DS algorithm to guarantee that at least two cameras have calibration matrices that meet our expectations, whereas the nonlinear method has the potential to diverge.

We show the ECDFs at a larger noise level of $\sigma=3$ pixels in Fig. 2. At this high level of noise, the DS algorithm becomes very unreliable. Only our ML and ML+R algorithms continue to provide robust results, and the benefit of the final resectioning stage is more pronounced.

In Fig. 3, we evaluate autocalibration performance of the ML method at $\sigma=1$ using 2,3,4,5,6 and 10 views. We have plotted the median error with interquartile range (IQR), and both asymptotically approach zero as the number of views increases, demonstrating stability for larger problems.

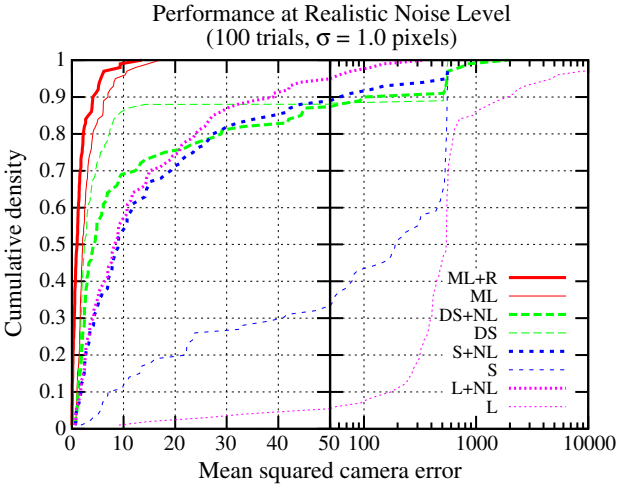

Fig. 1. Empirical cumulative distribution of camera errors from 100 random configurations using each method of autocalibration. The initial projective reconstruction is obtained by projective bundle adjustment from image point measurements with normally distributed noise having $\sigma=1.0$ pixels. The x-axis uses a log-scale for $x>50$.

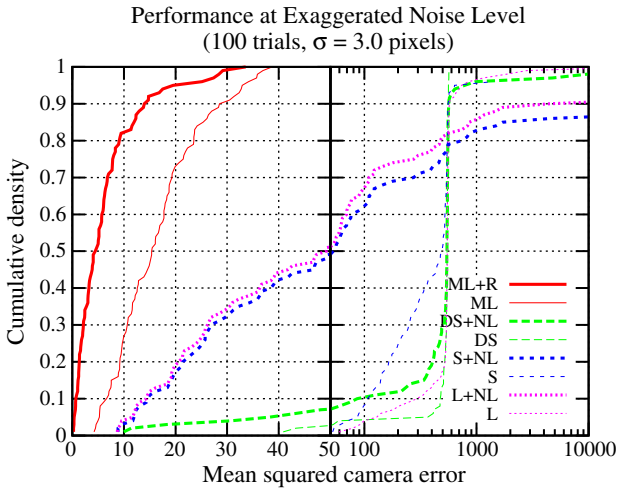

Fig. 2. Empirical cumulative distribution of camera errors from 100 random configurations using each method of autocalibration. The initial projective reconstruction is obtained by projective bundle adjustment from image point measurements with normally distributed noise having $\sigma=3.0$ pixels. The x-axis uses a log-scale for $x>50$.

Although the ML+R method performs slightly better it is omitted from Fig. 3 for clarity because the median performance is almost the same. It should also be noted that lower error could be achieved with the same number of views by using a wider baseline, moving the points closer to the cameras, or reducing the added noise.

We demonstrate the runtime performance of our ML and ML+R autocalibration algorithms in Fig. 4, using constraints on aspect ratio, skew and principal point during the resectioning step. The timings are shown with 95% confidence intervals from 25 repetitions, and indicate that performance scales linearly with the number of views. Runtime is about 1.2 s for 1000 points in 64 views using the ML method, or 6 s when using the ML+R method, on a Core i7 920 processor.

5.1. Examples on real data

In order to demonstrate our method on real data we obtained some reconstructions that were created using existing structure from motion systems. We will show that the method is general enough to be applied on any type of problem, such as a photo collection from different cameras, a photo collection from the same camera, and a video reconstruction. Our general approach is to take an existing

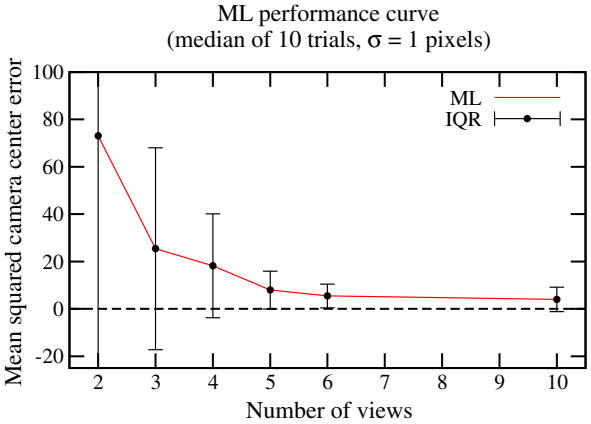

Fig. 3. Objective autocalibration performance using 2,3,4,5,6 and 10 views at $\sigma=1$. The median of 10 trials with interquartile range is plotted.

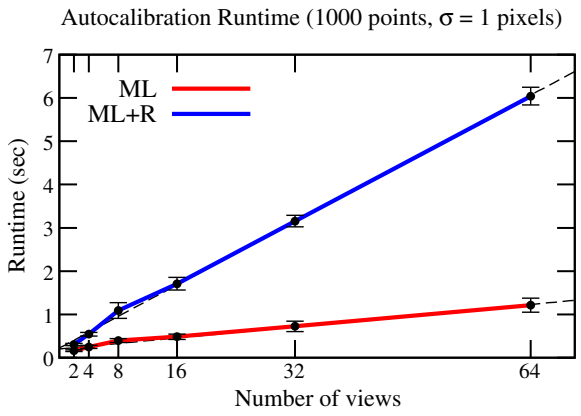

Fig. 4. Runtime performance of ML and ML + R autocalibration routines, shown with 95% confidence intervals from 25 repetitions. Performance scales linearly with the number of views.

metric reconstruction and then apply projective bundle adjustment to corrupt it, giving us a ML projective reconstruction from which we attempt autocalibration. We have not used any metric bundle adjustment to improve the results.

It should be noted that this is a much more realistic test of autocalibration than simply multiplying an existing metric reconstruction by an arbitrary homography and then trying to recover this homography; as we have shown in the Results section, all of the tested algorithms are capable of solving that problem extremely well. The difficulty in autocalibration is entirely due to the violation of metric constraints during projective bundle adjustment.

The success of an autocalibration algorithm can be assessed visually by looking at the reconstructed point cloud (Fig. 5). The point cloud of an arbitrary projective reconstruction (e.g., as obtained after projective bundle adjustment) is unbounded (Fig. 5b), and it is impossible to discern any meaningful structure. If π_∞ has been identified approximately correctly but Ω^*_∞ has not, then the convex hull will be bounded but the reconstruction will appear with a roughly affine skew as in Fig. 5c, and this is called a quasi-affine reconstruction. On the other hand, if Ω^*_∞ is identified correctly but π_∞ is not as in Fig. 5e, then the reconstruction will be unbounded and typically have a 'bow-tie' shape that spans from $-\infty$ to ∞, with some discernible structure near the origin. Only in a metric reconstruction are straight lines and angles preserved, and thus it is easy to identify a correct autocalibration of a cube, as computed by our ML algorithm in Fig. 5d, by observing 90° angles between adjacent faces. It is important to note that beyond correcting for an overall warping factor, autocalibration cannot be expected to remove noise from the individual structure points or camera positions.

In our first example, we used a reconstruction of *Brandenburg Gate* in Berlin, Germany that has been reconstructed by Frahm et al. [47] from a collection of 5514 photos gathered from the internet. The reconstruction consists of 19,963 structure points and has a total of 1,006,741 observations. From the existing image correspondences we ran projective bundle adjustment to obtain a ML projective reconstruction, and then autocalibrated using our ML + R method assuming zero skew and aspect ratio, obtaining a mean squared reprojection error of 0.79 pixels relative to image dimensions of 1000×1000. A view of the reconstruction after our autocalibration is shown in Fig. 6, and does not appear to suffer from any overall distortion, indicating that autocalibration was successful.

The next example is a reconstruction of the *Piazza dei Signore* in Verona, Italy that was reconstructed by the SAMANTHA [48] pipeline from a collection of 1144×856 photos taken by the same physical camera. The reconstruction consists of 2971 structure points, 39 views, and had an initial mean squared reprojection error of 0.330401 pixels which was reduced to 0.253106 after our projective bundle adjustment. We autocalibrated using our ML method, assuming a search range for focal length in the range of 1–3 screen widths, and initially assumed a principal point at (572, 428) (the center of the image) for the dual-stratified search. After the nonlinear improvement, the principal point was corrected to (554.836, 452.762). After factoring out the similarity transformation, the mean squared difference between the structure points in our metric rectified reconstruction and the original was just 2.73426×10^{-005}, which agrees quite closely with the original. An orthographic view of the reconstructed point cloud after our autocalibration is shown from a top down perspective in Fig. 7, where it can be verified from the density of points on the vertical walls that they are parallel.

Finally, we demonstrate autocalibration of a video reconstruction that was made by [49] using a parallel real-time visual SLAM method. This reconstruction consists of a total of 23 keyframes that were selected out of a 300 frame video with a resolution of 1224×1024. There are a total of 1473 structure points and 17,077 observations. The mean squared reprojection error of the reconstruction was 4.78 pixels, which we reduced to 0.529021 pixels using our projective bundle adjustment. Then we autocalibrated using our ML + R method assuming constant focal length, constant principal point, zero skew, and zero aspect ratio, which raised the mean square reprojection error only slightly to 0.530253 pixels. A view of the reconstruction after our autocalibration is shown in Fig. 8.

6. Conclusions

It has been thought that likelihood cannot be exploited during autocalibration because multiplication by any homography does not change reprojection error, and all previous autocalibration algorithms

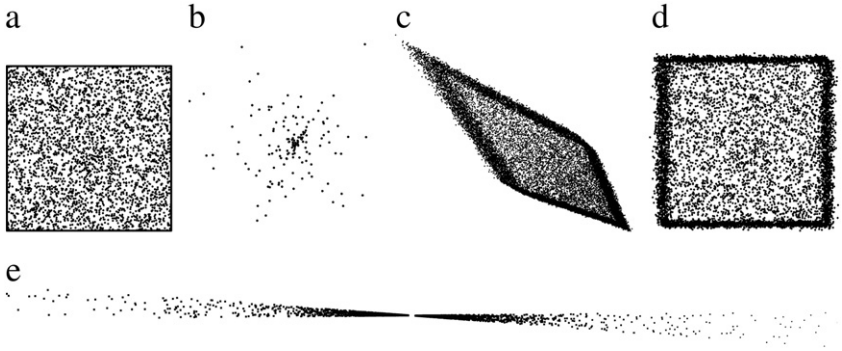

Fig. 5. Example point clouds viewed from the top. (a) Points on the surface of a cube in the true configuration without noise. (b) A subset of the (unbounded) ML projective reconstruction after bundle adjustment. (c) A partially successful autocalibration has obtained a quasi-affine reconstruction where at least π_∞ does not intersect the convex hull. (d) A successful autocalibration is visually identified by preserving right-angles. (e) A failed autocalibration attempt where π_∞ intersects the convex hull, sending reconstructed points to infinity and producing a characteristic 'bow-tie' shape.

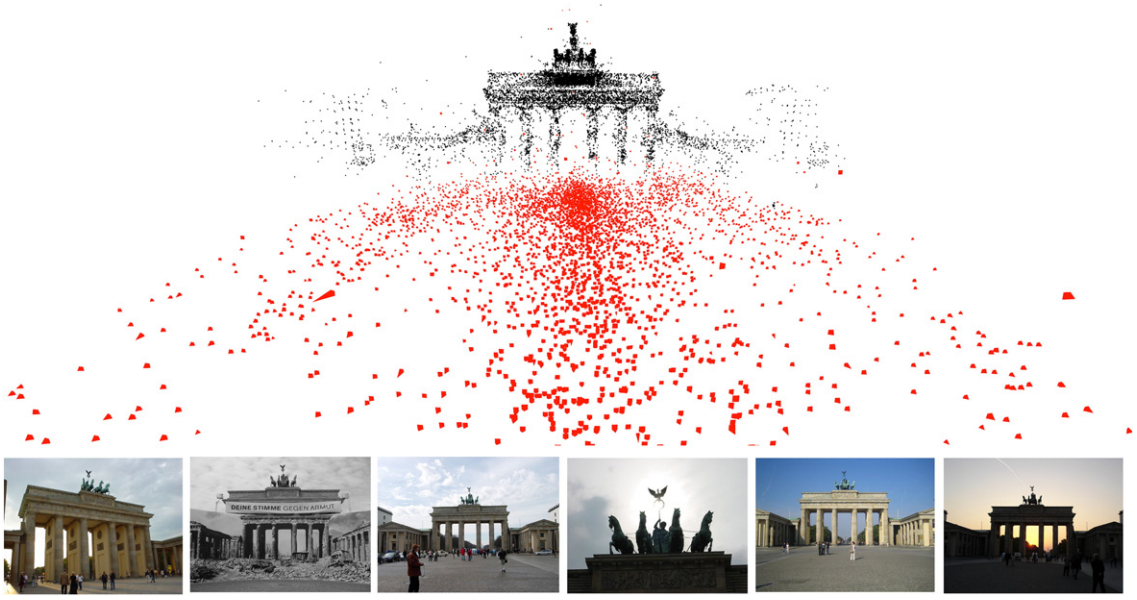

Fig. 6. Autocalibrated reconstruction from a collection of 5514 web photos (taken by different cameras) of *Brandenburg Gate* in Berlin, Germany. The reconstruction consists of 19,963 structure points (black dots), and cameras are shown as red pyramids. Some representative images used in the reconstruction are shown along the bottom.

have instead opted to minimize various heuristics based on manipulating the algebraic constraints that arise based on the assumption of zero skew, unit aspect ratio, or constant principal point/focal length. However, these heuristic cost functions are unstable because even the ideal ML projective reconstruction is not exactly within a homography from a true metric reconstruction.

We have shown that by taking into account metric constraints, a likelihood can be associated with any homography, thereby allowing one to seek the maximum likelihood rectifying homography. The ML homography can be found reliably by using a dual-stratified initialization followed by nonlinear improvement, and this method is more robust and accurate than any of the other algorithms we have tested for both small and large problems.

The advantages of maximizing likelihood as opposed to minimizing some arbitrary heuristic are many. First, the solution is invariant to the initial projective ambiguity; unlike all previous approaches, the projective reconstruction can be multiplied by any homography without changing the ML rectified result. Second, it is very robust to noise because it minimizes reprojection errors, which are geometrically meaningful. Third, because the error is geometrically meaningful, it is simple to incorporate early termination for improved performance whenever a reasonable error tolerance has been met. Fourth, likelihood can be used as an objective way to compare the performance of various autocalibration algorithms when a ground truth reconstruction is not available. And finally, it does not require any configuration-dependent weighting coefficients.

Acknowledgments

We thank Dr. Andrea Fusielli and Dr. Riccardo Gherardi for providing the *Piazza dei Signore* reconstruction, Dr. Enrique Dunn for providing the video reconstruction and Richard Steffen for providing

Fig. 7. Autocalibrated reconstruction from a collection of photos (taken by the same camera) of the *Piazza dei Signore* in Verona, Italy. The reconstruction is shown from a top down orthographic perspective. The reconstruction consists of 39 views and a total of 2971 structure points. Some images from representative views are shown along the bottom (the aerial view was not used in the reconstruction and is presented only for reference).

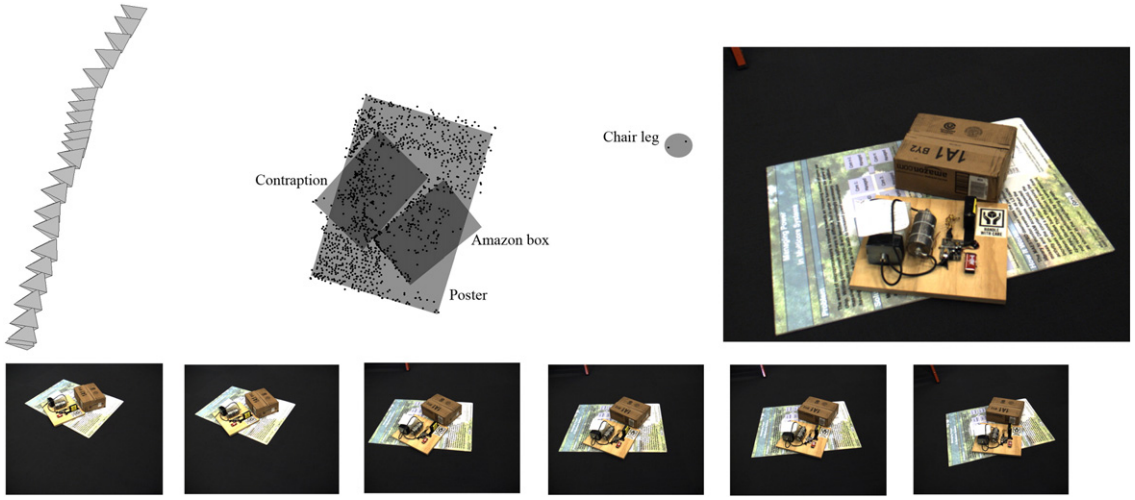

Fig. 8. A view of the point cloud of the autocalibrated reconstruction from a video reconstruction with 23 views, 1473 structure points and 17,077 observations. Some representative views are shown along the bottom. The approximate location of the scene elements was determined based on point elevations and is pictured here using opacity mapped squares for reference.

the *Brandenburg Gate* reconstruction. We thank Dr. Margaret J. Eppstein for her comments throughout, as well as the anonymous reviewers for their many specific suggestions. SBH was sponsored in part by the Army Research Office under grant W911NF-09-1-0458.

Appendix A. Limitations of maximum a priori autocalibration

We denote the meaningful elements of a calibration matrix $\mathbf{K}^j$ by a *calibration vector*, $\mathbf{k}^j = \left(\alpha_x^j - \alpha_y^j, s^j, u^j, v^j\right)^{\mathrm{T}}$. Note that the first element of the calibration vector is the difference between horizontal and vertical image scaling factors, because we generally expect these to have some unknown but equal value.

If one *assumes* that these calibration vectors are each distributed according to independent multivariate normal distributions, then maximizing prior probability leads to the maximum a priori rectifying homography, which is found by minimizing the sum of squared Mahalanobis distances,

$$\widehat{\mathbf{H}}_{MAP} = \underset{\mathbf{H}}{\operatorname{argmax}} \prod_{j=1}^{m} \frac{1}{\sqrt{(2\pi)^N |\Sigma_{\mathbf{k}}|}} \exp\left(-\frac{1}{2}\left(\mathbf{k}^j - \mu_{\mathbf{k}}\right)^{\mathrm{T}} \Sigma_{\mathbf{k}}^{-1} \left(\mathbf{k}^j - \mu_{\mathbf{k}}\right)\right) \tag{A.1}$$

$$= \underset{\mathbf{H}}{\operatorname{argmin}} \sum_{j=1}^{m} \left(\mathbf{k}^j - \mu_{\mathbf{k}}\right)^{\mathrm{T}} \Sigma_{\mathbf{k}}^{-1} \left(\mathbf{k}^j - \mu_{\mathbf{k}}\right), \tag{A.2}$$

where $N = 4$ is the length of $\mathbf{k}^j$, and $\mu_k = (0, 0, 0, 0)^{\mathrm{T}}$ and $\Sigma_{\mathbf{k}}$ define the mean and covariance of the assumed prior distribution for $\mathbf{k}^j$, [30]. If one further *assumes* that all parameters are independent, then Eq. (A.2) reduces to a weighted sum of squared errors,

$$\widehat{\mathbf{H}}_{MAP} = \underset{\mathbf{H}}{\operatorname{argmin}} \sum_{j} w_u \left(u^j\right)^2 + w_v \left(v^j\right)^2 + w_\alpha \left(\alpha_x^j - \alpha_y^j\right)^2 + w_s \left(s^j\right)^2. \tag{A.3}$$

Most autocalibration algorithms strive to optimize an objective of this form (or some close approximation), where the weighting coefficients w_u, w_v, w_α, w_s are determined using various guesses [26,27,36] or more commonly just omitted [1,23,27,28], which is equivalent to assuming they are all equal. It has been thought that the ideal way to choose these weighting coefficients is by looking at the distribution of intrinsic parameters in real cameras [26,30].

However, with the exception of focal length and possibly principal point, the uncertainty of these parameters in real cameras is negligible because modern cameras are manufactured to not produce skewed or stretched images. In other words, the prior model for these parameters should essentially be a delta function, and the parameters of any rectified result will never fall within the high density region of such a distribution.

In practice, one observes significant variance in these parameters, but the dominant source of this uncertainty is not due to uncertainty in the prior model, but rather is propagated from uncertainty in the measured image correspondences. Because projection matrices are over-parameterized (a homogeneous projection matrix has 11 DOF whereas a metric camera has only 7 DOF for pose and focal length), metric constraints will be happily violated by projective bundle adjustment in order to achieve a solution with lower reprojection error. The rectifying homography does not have enough DOF to restore these metric constraints, and hence the error is propagated into the intrinsic parameters of the metric rectified result.

Thus, the distribution of calibration parameters is configuration dependent, and the assumptions of Eq. (A.3) that parameters are normally distributed and independent are not valid. Furthermore, the propagated variance causes the expected value of a calibration vector to be shifted from zero. If one were to use this objective anyway, the optimal choice of weighting coefficients w_u, w_v, w_α, w_s would ideally be determined as the inverse of propagated variance, rather than being hard-coded based on some prior model of real cameras.

Because variance must be propagated through projective reconstruction, then though autocalibration, and finally through the RQ factorization to get calibration parameters, we have not had much success using analytical approximations. However, we have managed to obtain accurate estimations of the propagated variance by using Monte Carlo methods. Each trial consists of perturbing the original image point measurements with Gaussian noise and then repeating the entire process of projective reconstruction and autocalibration. An example of the sample covariance matrix containing all intrinsic parameters for a system of 6 views is shown in Fig. A.9, calculated from 1000 random trials.

The most distinct observation from Fig. A.9 is that the matrix is not even remotely close to being diagonal, as was assumed by the independence assumptions in Eq. (A.3). Furthermore, we observe that the greatest amount of uncertainty is propagated into focal length, and secondly into principal point. This confirms the already known facts that neither the estimation of principle nor focal length are well-

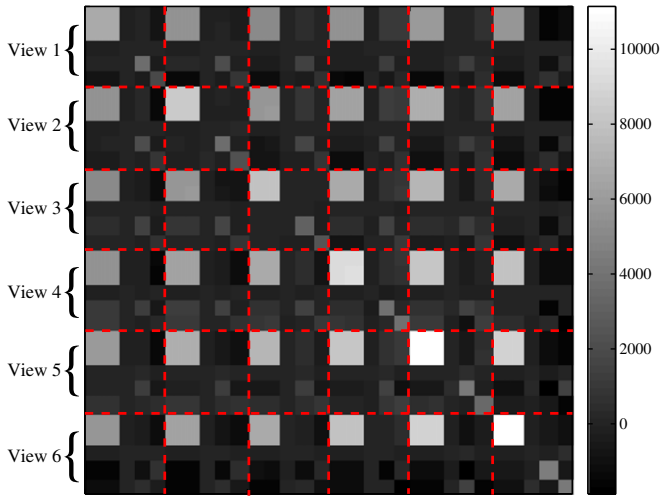

Fig. A.9. Covariance matrix of intrinsic parameters for a set of six views. The parameters are ordered as (α_x^1, α_y^1, s^1, u^1, v^1,..., α_x^6, α_y^6, s^6, u^6, v^6). The delineation between parameters of the same view is indicated by dotted grid lines.

conditioned problems [14]. The large covariance between horizontal and vertical focal lengths shows that the aspect ratio constraint is relatively good. In some views, there is a noticeable positive or negative correlation between focal length and principal point. Additionally, we see that there is very strong dependence between the focal lengths of different views. In other words, the assumption of independence in Eq. (A.3) is incorrect.

In order to get a more qualitative idea about how to best choose weighting coefficients, we calculated the constraint power (inverse variance) from 100 random configurations and show the results relative to skew, with 95% confidence intervals, in Table A.1. These results indicate that, for the range of random configurations tested, the optimal overall choice for weighting coefficients in Eq. (A.3) is $w_s = 1$, $w_\alpha = 1.01577$, $w_u = 0.0217334$, and $w_v = 0.0219337$. In other words, skew and aspect ratio constraints should be weighted approximately equally, whereas principal point should have a very negligible weight. In practice, we find that using any non-zero weight on principal point or focal length tends to magnify the effects of noise and provide a worse result.

However, we stress that the optimal choice of weighting coefficients in Eq. (A.3) is still highly configuration dependent. In order to demonstrate this we have generated several random configurations and then autocalibrated using all possible relative weightings between skew and aspect ratio. The objective measure of autocalibration accuracy is shown as a 2D surface for each configuration in Fig. A.10. One sees that the location of the minima is different for each problem, which is a clear indication that the optimal choice of weights is configuration dependent.

Table A.1
Power of calibration constraints relative to a constraint on skew, calculated as the inverse variance of each parameter in the calibration vector from 100 random configurations, and shown with 95% confidence intervals.

Constraint	Relative Power
s	1
$\alpha_x - \alpha_y$	1.01577 ± 0.0170464
α_x	0.0169471 ± 0.00156958
α_y	0.0169465 ± 0.00157017
u	0.0217334 ± 0.00202811
v	0.0219337 ± 0.00200886

Appendix B. Relationship to previous autocalibration constraints

Any 3×4 projection matrix $\mathbf{P}$ can be uniquely factored [37] into a rotation matrix $\mathbf{R}$, translation $\mathbf{t}$, and calibration matrix $\mathbf{K}$ as

$$\mathbf{P} \propto \mathbf{K}[\mathbf{R}|\mathbf{t}]. \tag{B.1}$$

Define $\tilde{\mathrm{I}} = diag(1,1,1,0)$. Then, observe that the rotation and translation components can be canceled out by

$$\mathbf{P}\tilde{\mathrm{I}}\mathbf{P}^{\mathrm{T}} \propto \mathbf{K}[\mathbf{R}|\mathbf{t}]\tilde{\mathrm{I}}[\mathbf{R}|\mathbf{t}]^{\mathrm{T}}\mathbf{K}^{\mathrm{T}} \propto \mathbf{K}[\mathbf{R}|0]\begin{bmatrix}\mathbf{R}^{\mathrm{T}}\\ 0\end{bmatrix}\mathbf{K}^{\mathrm{T}} \propto \mathbf{K}\mathbf{K}^{\mathrm{T}}. \tag{B.2}$$

Any projective camera $\mathbf{P}^j$ in the reconstruction is related to a metric camera (denoted by subscript M) via the rectifying homography $\mathbf{H}_M$,

$$\mathbf{P}^j_{\mathrm{M}} \propto \mathbf{P}^j\mathbf{H}_{\mathrm{M}} \forall j. \tag{B.3}$$

Substituting Eq. (B.3) into Eq. (B.2), we obtain

$$\mathbf{P}^j\mathbf{H}_{\mathrm{M}}\tilde{\mathrm{I}}\mathbf{H}_{\mathrm{M}}^{\mathrm{T}}\mathbf{P}^{j\mathrm{T}} \propto \mathbf{K}^j\mathbf{K}^{j\mathrm{T}} \forall j. \tag{B.4}$$

Using Eq. (B.4), prior constraints on the calibration matrices can be translated into nonlinear constraints on $\mathbf{H}_M$. This makes it the most fundamental equation of autocalibration, from which all other constraints that are based on calibration matrices can be derived. Making the substitutions of $\mathbf{Q}_\infty^* = \mathbf{H}_{\mathrm{M}}\tilde{\mathrm{I}}\mathbf{H}_{\mathrm{M}}^{\mathrm{T}}$ and $\omega^{*j} = \mathbf{K}^j\mathbf{K}^{j\mathrm{T}}$, Eq. (B.4) is usually written as a constraint on $\mathbf{Q}_\infty^*$ and ω^{*j},

$$\mathbf{P}^j\mathbf{Q}_\infty^*\mathbf{P}^{j\mathrm{T}} \propto \omega^{*j} \forall j. \tag{B.5}$$

In the literature $\mathbf{Q}_\infty^*$ is known as the *absolute dual quadric* (ADQ), and ω^{*j} is a *dual image of the absolute conic* (DIAC). The relationship between these entities is depicted graphically in Fig. B.11.

The advantage of using $\mathbf{Q}_\infty^*$ is that some constraints can be translated into linear constraints on $\mathbf{Q}_\infty^*$, allowing linear least squares method to be used as an initialization. However, there are additional internal constraints on $\mathbf{Q}_\infty^*$ that cannot be enforced by a linear solution. By construction, $\mathbf{Q}_\infty^*$ must be symmetric, rank 3 and positive-semidefinite (Theorem 1) and all ω^{*j} must be symmetric and positive-semidefinite (Theorem 2).

Because $\mathbf{Q}_\infty^*$ is singular, it is a degenerate (dual) quadric, meaning that it actually represents a dual conic embedded in some plane. This dual conic, called the *absolute dual conic* (ADC) and denoted by Ω_∞^*, is encoded in the upper 3×3 portion of $\mathbf{Q}_\infty^*$. The plane it lives in is called the *plane at infinity*, denoted by π_∞ and encoded by the null space of $\mathbf{Q}_\infty^*$. Geometrically, Eq. (B.5) shows us that ω^{*j} is the projection of $\mathbf{Q}_\infty^*$ onto the image plane of $\mathbf{P}^j$. Therefore, Ω_∞^* is also the projection of $\mathbf{Q}_\infty^*$ by the canonical projection matrix, $\mathbf{P} = [\mathrm{I}|0]$.

In a metric reconstruction $\mathbf{H}_M = \mathrm{I}$, in which case one observes that $\mathbf{Q}_\infty^* M = \tilde{\mathrm{I}}$, $\Omega_\infty^* M = \mathrm{I}$, and $\pi_\infty M = (0,0,0,1)^{\mathrm{T}}$, which does not correspond to any real plane equation. By definition, a point $\mathbf{X}$ lies on a plane π iff $\pi^{\mathrm{T}}\mathbf{X} = 0$, and therefore the only points lying on $\pi_\infty \mathrm{M}$ are homogeneous points of the form $(a, b, c, 0)^{\mathrm{T}}$. These are all points at infinity, hence why we call π_∞ the plane at infinity. Under the action of a homography $\mathbf{H}$ a plane π transforms to $\pi' = \mathbf{H}^{-\mathrm{T}}\pi$, and hence π_∞ could be any real plane in the projective reconstruction.

Appendix B.1. The infinite homography

As a consequence of the fact that $\mathbf{Q}_\infty^*$ is degenerate, its projection ω^{*j} can alternatively be computed via a planar homography transfer of Ω_∞^*. The homography that transfers from $\pi_\infty = (\mathbf{p}^{\mathrm{T}}, 1)^{\mathrm{T}}$ to the image

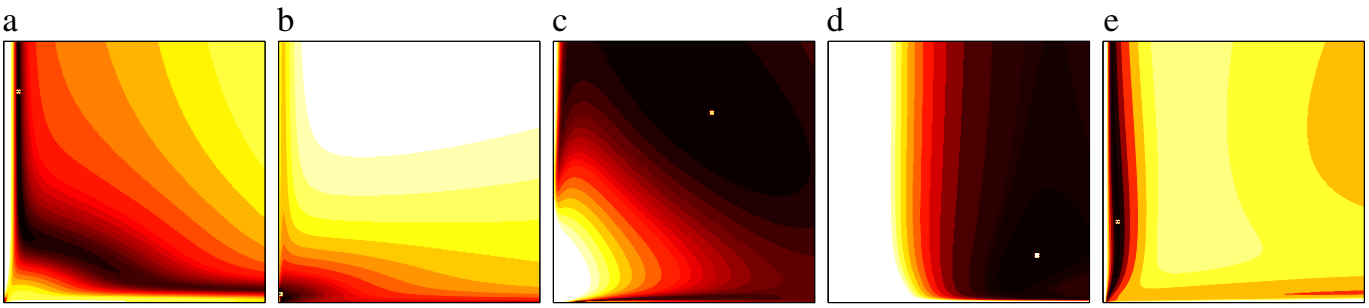

Fig. A.10. Example cost surfaces demonstrating that the optimal choice of weights is configuration dependent. Each surface corresponds to a different configuration, and the intensity at each point on the surface indicates the objective reconstruction quality as a function of the relative weighting between skew (x-axis) and aspect ratio (y-axis) constraints. The weights corresponding to the most accurate reconstruction is marked, and changes significantly with each configuration.

plane of view $\mathbf{P}^j = \left[\mathbf{A}^j \,|\, \mathbf{a}^j\right]$ is called the *infinite homography*, denoted by $\mathbf{H}^j_\infty$, and given by

$$\mathbf{H}^j_\infty \propto \mathbf{A}^j - \mathbf{a}^j \mathbf{p}^{\mathrm{T}} \forall j. \tag{B.6}$$

Under the action of a homography $\mathbf{H}$, a dual conic ω^* transforms to $\omega^{*\prime} = \mathbf{H}\omega^*\mathbf{H}^{\mathrm{T}}$. Thus, we obtain the *infinite homography constraint* on Ω^*_∞,

$$\mathbf{H}^j_\infty \Omega^*_\infty \mathbf{H}^j_\infty \mathrm{T} \propto \omega^{*j} \forall j. \tag{B.7}$$

The infinite homography constraint is a pairwise constraint between image planes (see Fig. B.11), a special case of Eq. (B.4) that does not enforce the common support plane for π_∞.

Appendix B.2. The kruppa equations

Let $\mathbf{e}^{ij}$ denote the image of the ith camera center in view j (an epipole). There is a corresponding skew-symmetric matrix $\left[\mathbf{e}^{ij}\right]_\times$ [1], and multiplying both sides of Eq. (B.7) by $\left[\mathbf{e}^{ij}\right]_\times$ leads to

$$\left[\mathbf{e}^{ij}\right]_\times \omega^{*j} \left[\mathbf{e}^{ij}\right]_\times \propto \left(\left[\mathbf{e}^{ij}\right]_\times \mathbf{H}^j_\infty\right) \Omega^*_\infty \left(\mathbf{H}^j_\infty \mathrm{T} \left[\mathbf{e}^{ij}\right]_\times\right) \tag{B.8}$$

$$\propto \mathbf{F}_{ij} \Omega^*_\infty \mathbf{F}^{\mathrm{T}}_{ij} \forall ij, \tag{B.9}$$

where $\mathbf{F}_{ij}$ is the fundamental matrix between views i and j. The set of equations in (B.9) are equivalent to the original Kruppa equations [8], expressing constraints on Ω^*_∞ in the form of corresponding epipolar lines tangent to ω^{*j}. Thus, the Kruppa equations are just a special case of the infinite homography constraint.

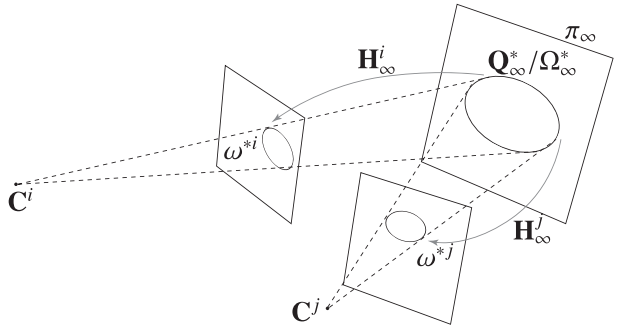

Fig. B.11. The absolute dual quadric $\mathbf{Q}^*_\infty$ encodes for the absolute dual conic Ω^*_∞, as well as the plane at infinity π_∞ that Ω^*_∞ is embedded in. A dual image of the absolute conic ω^{*j} is the image of Ω^*_∞ as seen by camera $\mathbf{P}^j$ having focal point at $\mathbf{C}^j$, and can be obtained either by projection of $\mathbf{Q}^*_\infty$, or by using the infinite homography $\mathbf{H}^j_\infty$ to map Ω^*_∞ from π_∞ to the image plane of $\mathbf{P}^j$.

Appendix B.3. The rigidity constraint

It has been shown [12] that the Kruppa equations are also equivalent to a constraint that the essential matrix has two identical singular values and one zero singular value. The essential matrix is related to the fundamental matrix by

$$\mathbf{E}_{ij} = [\mathbf{t}]_\times \mathbf{R} = \mathbf{K}^{jT} \mathbf{F}_{ij} \mathbf{K}^i, \tag{B.10}$$

where $\mathbf{t}$ and $\mathbf{R}$ represent the translation and rotation between the camera pair. This constraint may also be stated as

$$det\left(\mathbf{E}_{ij}\right) = 0 \wedge 2tr\left(\left(\mathbf{E}_{ij}\mathbf{E}^{\mathrm{T}}_{ij}\right)^2\right) - \left(tr\left(\mathbf{E}_{ij}\mathbf{E}^{\mathrm{T}}_{ij}\right)\right)^2 = 0, \tag{B.11}$$

and is called the *rigidity constraint* on $\mathbf{E}_{ij}$ because it is a result of the rigid motion between cameras.

Appendix B.4. The modulus constraint

Without loss of generality, we can align the projective reconstruction such that $\mathbf{P}^i = [\mathrm{I}|0]$, and choose our metric reconstruction such that $\mathbf{P}^i_M = \mathbf{K}^i[\mathrm{I}|0]$. Then, because $\mathbf{P}^i_M = \mathbf{P}^i\mathbf{H}_M$, it can be verified that $\mathbf{H}_M$ is of the form

$$\mathbf{H}_M = \begin{bmatrix} \mathbf{K}^i & 0 \\ \mathbf{v}^{\mathrm{T}} & 1 \end{bmatrix}. \tag{B.12}$$

In this case, $\Omega^*_\infty = \mathbf{K}^i\mathbf{K}^i\mathrm{T}$, and $\mathbf{v} = -\mathbf{K}^i\mathrm{T}\mathbf{p}$. Substituting Eq. (B.12) into Eq. (B.3), we see that

$$\mathbf{K}^j\left[\mathbf{R}^j \,|\, \mathbf{t}^j\right] \propto \left[\left(\mathbf{A}^j - \mathbf{a}^j\mathbf{p}^{\mathrm{T}}\right)\mathbf{K}^i \,|\, \mathbf{a}^j\right] \tag{B.13}$$

$$\mathbf{K}^j\mathbf{R}^j \propto \left(\mathbf{A}^j - \mathbf{a}^j\mathbf{p}^{\mathrm{T}}\right)\mathbf{K}^i \tag{B.14}$$

$$\mathbf{K}^j\mathbf{R}^j\mathbf{K}^{i-1} \propto \mathbf{A}^j - \mathbf{a}^j\mathbf{p}^{\mathrm{T}} \tag{B.15}$$

Thus, if $\mathbf{K}^j = \mathbf{K}^i$ then $\mathbf{H}^j_\infty = \mathbf{A}^j - \mathbf{a}^j\mathbf{p}^{\mathrm{T}}$ is similar (a.k.a. conjugate) to a rotation, and has eigenvalues proportional to $\{e^{i\theta}, e^{-i\theta}, 1\}$. In other words, the modulus of the first two eigenvalues are equal. This is known as the *modulus constraint* on $\mathbf{p}$. Note that the modulus constraint is usually enforced as a constraint on the coefficients of the characteristic polynomial of $\mathbf{H}^j_\infty$, as in Chandraker et al. [34], Pollefeys and Van Gool [41].

Assuming all calibration matrices are equal, enforcing the modulus constraint between all pairs of views would ensure a common support plane for π_∞, and is therefore also a special case of (B.5).

Theorem 1. Let $\mathbf{H}$ be any real matrix. Then $\mathbf{H}\mathbf{H}^{\mathrm{T}}$ is positive-semidefinite.

Proof. A square matrix $\mathbf{M}$ is positive-semidefinite if and only if $\mathbf{z}^T\mathbf{M}\mathbf{z} \geq 0$, for any non-zero $\mathbf{z}$. It holds that $\mathbf{z}^T\mathbf{H}\mathbf{H}^T\mathbf{z} = \|\mathbf{H}^T\mathbf{z}\|^2 \geq 0$. □

Theorem 2. Let $\mathbf{Q}$ be an $n \times n$ positive-semidefinite matrix, and $\mathbf{P}$ be any $m \times n$ matrix. Then $\mathbf{P}\mathbf{Q}\mathbf{P}^T$ is also positive-semidefinite.

Proof. The factorization $\mathbf{Q} = \mathbf{L}\mathbf{L}^T$ must exist because $\mathbf{Q}$ is positive-semidefinite. Let $\mathbf{H} = \mathbf{P}\mathbf{L}$. Then $\mathbf{H}\mathbf{H}^T = \mathbf{P}\mathbf{L}\mathbf{L}^T\mathbf{P}^T = \mathbf{P}\mathbf{Q}\mathbf{P}^T$ is positive-semidefinite by Theorem 1. □

References

[1] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second edn., Cambridge U. Press, 2003.

[2] O.D. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, ECCV '92: Proceedings of the Second European Conference on Computer Vision, Springer-Verlag, London, UK, ISBN: 3-540-55426-2, 1992, pp. 563–578.

[3] R.I. Hartley, Estimation of relative camera positions for uncalibrated cameras, ECCV '92: Proceedings of the Second European Conference on Computer Vision, Springer-Verlag, London, UK, ISBN: 3-540-55426-2, 1992, pp. 579–587.

[4] D. Nistér, An efficient solution to the five-point relative pose problem, IEEE Trans. Pattern Anal. Mach. Intell. 26 (6) (2004) 756–777 ISSN 0162–8828, doi:http://dx.doi.org/10.1109/TPAMI.2004.17.

[5] R.I. Hartley, F. Kahl, Global optimization through rotation space search, Int. J. Comput. Vis. 82 (1) (2009) 64–79 ISSN 0920–5691, doi:http://dx.doi.org/10.1007/s11263-008-0186-9.

[6] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment — a modern synthesis, in: B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice, vol. 1883 of *Lecture Notes in Computer Science*, Springer-Verlag, 2000, pp. 298–372, URL http://lear.inrialpes.fr/pubs/2000/TMHF00.

[7] M. Lourakis, A. Argyros, The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg–Marquardt algorithm, Tech. Rep., 340, Institute of Computer Science — FORTH, Heraklion, Crete, Greece, 2004, available from http://www.ics.forth.gr/lourakis/sba+.

[8] E. Kruppa, Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung, Other J. (1913) 1939–1948.

[9] O.D. Faugeras, Q.-T. Luong, S.J. Maybank, Camera self-calibration: theory and experiments, ECCV '92: Proceedings of the Second European Conference on Computer Vision, Springer-Verlag, London, UK, 1992, ISBN 3-540-55426-2, 321–334.

[10] T. Vieville, O. Faugeras, Q. Luong, Motion Of Points And Lines In The Uncalibrated Case, Int. J. Comput. Vis. 17 (1) (1996) 7–41.

[11] R. Hartley, Kruppa's Equations Derived from the Fundamental Matrix, IEEE Trans. Pattern Anal. Mach. Intell. 19 (2) (1997) 133–135.

[12] T.S. Huang, O.D. Faugeras, Some properties of the e matrix in two-view motion estimation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (12) (1989) 1310–1312 ISSN 0162–8828, doi:http://dx.doi.org/10.1109/34.41368.

[13] P. Mendonça, R. Cipolla, A simple technique for self-calibration, CVPR99, 1999, pp. 500–505, I.

[14] S. Bougnoux, From projective to Euclidean space under any practical situation, a criticism of self-calibration, ICCV '98: Proceedings of the Sixth International Conference on Computer Vision, IEEE Computer Society, Washington, DC, USA, 790, ISBN: 81-7319-221-9, 1998.

[15] W. Lao, Z. Cheng, A. Kam, T. Tan, A. Kassim, Focal length self-calibration based on degenerated Kruppa's equations: method and evaluation, Image Processing, 2004. ICIP '04. 2004 International Conference on, Vol. 5, 2004, pp. 3391–3394, doi:10.1109/ICIP.2004.1421842, vol. 5, ISSN 1522–4880.

[16] Y. Gao, H. Radha, A multistage camera self-calibration algorithm, Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on, vol. 3, 2004, pp. iii-537–iii-540, doi:10.1109/ICASSP.2004.1326600, vol. 3, ISSN 1520–6149.

[17] Q. Luong, matrice fundamentale et autocalibration en vision par ordinateur, Ph.D. thesis, Universite de Paris-Sud, Paris, France, 1992.

[18] C.Zeller, Projective, Affine and Euclidean Calibration in Computer Vision and the Application of Three Dimensional Perception, Ph.D. thesis, RobotVis Group, INRIA Sophia-Antipolis, 1996.

[19] Q. Luong, O.D. Faugeras, Self-calibration of a moving camera from point correspondences and fundamental matrices, Int. J. Comput. Vis. 22 (3) (1997) 261–289 ISSN 0920–5691, doi:http://dx.doi.org/10.1023/A:1007982716991.

[20] A. Fusiello, A. Benedetti, M. Farenzena, A. Busti, Globally convergent autocalibration using interval analysis, pattern analysis and machine intelligence, IEEE Trans. 26 (12) (2004) 1633–1638, doi:10.1109/TPAMI.2004.125 ISSN 0162–8828.

[21] M. Pollefeys, L. Van Gool, A. Oosterlinck, The modulus constraint: a new constraint self-calibration, Pattern Recognition, 1996., Proceedings of the 13th International Conference on, vol. 1, 1996, pp. 349–353, doi:10.1109/ICPR.1996.546047, vol. 1.

[22] B. Triggs, Autocalibration and the absolute quadric, CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), IEEE Computer Society, Washington, DC, USA, 609, ISBN: 0-8186-7822-4, 1997.

[23] M. Pollefeys, R. Koch, L. Van Gool, Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters, Computer Vision, 1998. Sixth International Conference on, 1998, doi:10.1109/ICCV.1998.710705, 90 –95.

[24] J. Oliensis, Fast and accurate self-calibration, ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2, IEEE Computer Society, Washington, DC, USA, 745, 1999, ISBN 0-7695-0164-8.

[25] J. Ponce, On computing metric upgrades of projective reconstructions under the rectangular pixel assumption, SMILE '00: Revised Papers from Second European Workshop on 3D Structure from Multiple Images of Large-Scale Environments, Springer-Verlag, London, UK, 2001, pp. 52–67, ISBN 3-540-41845-8.

[26] M. Pollefeys, F. Verbiest, L.J.V. Gool, Surviving dominant planes in uncalibrated structure and motion recovery, ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part II, Springer-Verlag, London, UK, 2002, pp. 837–851, ISBN 3-540-43744-4.

[27] B. Bocquillon, A. Bartoli, P. Gurdjos, A. Crouzil, On constant focal length self-calibration from multiple views, computer vision and pattern recognition, IEEE Computer Society Conference on, 0, 2007, pp. 1–8, doi:http://doi.ieeecomputersociety.org/10.1109/CVPR.2007.3830 66.

[28] R.I. Hartley, E. Hayman, L. de Agapito, I. Reid, Camera calibration and the search for infinity, computer vision, IEEE International Conference on, 1, 1999, p. 510, doi:http://doi.ieeecomputersociety.org/10.1109/ICCV.1999.7912 64.

[29] R.I. Hartley, Chirality, Int. J. Comput. Vis. 26 (1) (1998) 41–61 ISSN 0920–5691, doi:http://dx.doi.org/10.1023/A:1007984508483.

[30] D. Nister, Calibration with robust use of cheirality by quasi-affine reconstruction of the set of camera projection centres, IEEE International Conference on Computer Vision (ICCV 2001), vol. 2, 2001, pp. 116–123.

[31] M. Agrawal, On automatic determination of varying focal lengths using semidefinite programming, Image Processing, 2004. ICIP '04. 2004 International Conference on, Vol. 5, 2004, pp. 3379–3382, doi:10.1109/ICIP.2004.1421839, vol. 5, ISSN 1522–4880.

[32] M. Agrawal, Practical camera auto calibration using semidefinite programming, Motion and Video Computing, 2007. WMVC '07. IEEE Workshop on, 2007, doi:10.1109/WMVC.2007.39, 20 –20.

[33] M. Chandraker, S. Agarwal, D. Kriegman, S. Belongie, Globally optimal affine and metric upgrades in stratified autocalibration, ICCV07, 2007, pp. 1–8.

[34] M. Chandraker, S. Agarwal, D. Kriegman, S. Belongie, Globally optimal algorithms for stratified autocalibration, Int. J. Comput. Vis. 90 (2010) 236–254, ISSN 0920–5691, URL http://dx.doi.org/10.1007/s11263-009-0305-2, doi:10.1007/s11263-009-0305-2.

[35] M. Chandraker, S. Agarwal, F. Kahl, D. Nister, D. Kriegman, Autocalibration via rank-constrained estimation of the absolute quadric, CVPR07, 2007, pp. 1–8.

[36] R. Gherardi, A. Fusiello, Practical autocalibration, ECCV 2010: Proceedings of the 11th European Conference on Computer Vision, 2010.

[37] G.H. Golub, C.F. Van Loan, Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition), 3rd edn., The Johns Hopkins University Press0801854148, 1996.

[38] R.I. Hartley, P. Sturm, Triangulation, Computer Vision and Image Understanding 68 (2) (1997) 146 – 157, ISSN 1077–3142.

[39] P. Alcantarilla, L. Bergasa, F. Dellaert, Visual odometry priors for robust EKF-SLAM, Robotics and Automation (ICRA), 2010 IEEE International Conference on, ISSN 1050–4729, 2010, pp. 3501–3506, doi:10.1109/ROBOT.2010.5509272.

[40] O. Faugeras, Stratification of 3-D vision: projective, affine, and metric representations, J. Opt. Soc. Am. A 12 (1995) 465–484.

[41] M. Pollefeys, L. Van Gool, Stratified self-calibration with the modulus constraint, IEEE Trans. Pattern Anal. Mach. Intell. 21 (8) (1999) 707–724 ISSN 0162–8828, doi:http://dx.doi.org/10.1109/34.784285.

[42] D. Marquardt, An Algorithm for Least-Squares Estimation of Nonlinear Parameters, SIAM J. Appl. Math. 11 (2) (1963) 431–441.

[43] S. Umeyama, Least-squares estimation of transformation parameters between two point patterns, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 376–380 ISSN 0162–8828, doi:http://dx.doi.org/10.1109/34.88573, URL http://dx.doi.org/10.1109/34.88573.

[44] P.H.S. Torr, A. Zisserman, MLESAC: a new robust estimator with application to estimating image geometry, Comput. Vis. Image Underst. 78 (1) (2000) 138–156 ISSN 1077–3142, doi:http://dx.doi.org/10.1006/cviu.1999.0832.

[45] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, Vol 24, Issue 6, 1981.

[46] D.G. Lowe, Object recognition from local scale-invariant features, ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2, IEEE Computer Society, Washington, DC, USA, 1150, ISBN: 0-7695-0164-8, 1999.

[47] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Ragauram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, M. Pollefeys, Building Rome on a Cloudless Day, ECCV 2010: Proceedings of the 11th European Conference on Computer Vision, 2010, pp. 368–381.

[48] M. Farenzena, A. Fusiello, R. Gherardi, Structure-and-motion pipeline on a hierarchical cluster tree, Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on, 2009.

[49] B. Clipp, J. Lim, J.-M. Frahm, M. Pollefeys, Parallel, Real-Time Visual SLAM, Computer Vision and Pattern Recognition, 2010, Proceedings. IEEE Conference on, 2010.