Stable Intrinsic Auto-Calibration from Fundamental Matrices of Devices with Uncorrelated Camera Parameters

Torben Fetzer[1] Gerd Reis[2] Didier Stricker[1,2]

[1]Department of Computer Science, TU Kaiserslautern
[2]Department Augmented Vision, DFKI GmbH
{torben.fetzer, gerd.reis, didier.stricker}@dfki.de

Abstract

Auto-Calibration is an important task in computer vision and is necessary for many visual applications. Methods like photogrammetry, depth map estimation, metrology, augmented/mixed reality or odometry are strongly dependent on well calibrated devices. While classical calibration relies on tools like checkerboards or additional scene information, auto-calibration only takes epipolar relations into account. Classical calibration is often impractical, tends to de-adjust over time and distributes the error over the entire, limited working volume. Auto-calibration, on the other hand, does not require any information other than the image content itself, has a virtually unlimited working range and usually achieves highest accuracy at the objects' surfaces. Unfortunately, auto-calibration methods are sensitive to errors in the fundamental matrix and need good initialization to converge to the global solution. In practice this leads to difficulties if optical parameters like principal point or focal length are unconstrained. In such situations, even state-of-the-art auto-calibration methods tend to diverge and do not yield a valid calibration.

This work assesses reasons for this behavior, in particular for the initialization method of Bougnoux [3] and Lourakis' state-of-the-art auto-calibration method [21]. Based on the analysis, a more stable method is proposed. A continuous and smooth energy functional is introduced, providing superior convergence properties. I.e. it can not diverge, converges faster, and has a significantly enlarged convergence region with respect to the global minimum.

Finally, a thorough evaluation has been conducted and a detailed comparison with the state of the art is presented.

1. Introduction

Modern auto-calibration techniques aim to calibrate multiple devices without any form of user interaction. Each such calibration procedure is based on the computation of epipolar relations between the devices, which are represented as fundamental matrices. The procedures described in [32] and [15] solve these tasks sufficiently well for many situations. Methods of higher accuracy such as [11] can increase the chance of convergence of further calibration steps, even more. The correction of distortions caused by the optical system of the devices is another essential aspect. In [11] distortions are corrected while the fundamental matrices are estimated. With this approach, further auto-calibration steps no longer have to take distortions into account and also apply to cameras with a large field of view. In order to extract the extrinsic parameters of the devices (rotation and translation) from the fundamental matrices, as described in [15], the intrinsic calibration parameters are a necessary prerequisite. The most important intrinsics of an optical device are the focal length and the principal point. In order to compute Euclidean reconstructions, precise estimates are essential. In contrast, intrinsics like skew and aspect ratio are of less importance nowadays, since modern devices are equipped with square pixels. Since the entire reconstruction process, except the estimation of the fundamental matrix, depends on the intrinsic parameters, their accurate estimation is crucial and the most difficult and error-prone part.

Most state-of-the-art methods rely on either calibration tools such as checkerboards or additional scene information like planar structures or parallel lines. Although the consideration of this additional information works well, its use in practice is cumbersome and time-consuming. An auto-calibration procedure without these requirements is therefore preferable. Consequently, extensive research has been carried out in this respect in recent decades and methods have been discovered that enable intrinsic calibration directly on the basis on the fundamental matrix. Although state-of-the-art methods are theoretically sound and valuable, their practical application is often not stable and fails in many cases. Especially setups consisting of different camera models and projectors, as in the case of active scanning, can cause problems.

Famous bundle-adjustment [29] or methods like [13] have not been investigated because they are post-processing methods that require already appropriate calibration.

In this work we continue research on auto-calibration techniques. Our contributions are:

- Analysis of Bougnoux's method regarding uncertain focal length estimates (4.1)
- Proposal of a new energy functional (4.2)
- Qualitative comparison of Lourakis' and the new functional (4.4) regarding
  - Focal length estimation (4.4.1)
  - Principal point estimation (4.4.2)
- Quantitative comparison of the functionals (5)

In particular, the proposed method supports any number of devices, the method has a significantly larger region of convergence and is not biased towards larger focal lengths. In addition, the minimized energy term is smooth and nearly convex, which allows a stable estimation of the principal points, even if they are far off the image center.

2. Related work

Research on intrinsic calibration from epipolar geometry achieved a quantum leap by the theory of the absolute conic used by Faugeras *et al*. to introduce *Kruppa's Equations* into computer vision in [10]. These equations represent the basis of modern auto-calibration as they describe a direct connection between fundamental matrices and the respective intrinsic calibration matrices. Bougnoux [3] and Hartley [14] gave formulations for computing the focal lengths of two uncorrelated views given their fundamental matrix. Both approaches depend on known principal points and epipoles. Since the epipoles are usually estimated as the null-space of the fundamental matrix, they are sensitive to small inaccuracies in the fundamental matrix. This can lead to instabilities of the methods, even if correct principal points are given.

To avoid these problems, Hartley reformulated Kruppa's equations in terms of the singular value decomposition of the fundamental matrix to introduce epipole-invariant Kruppa Equations in [16]. Based on Hartleys work, Sturm [26] presented a more robust method for two views with constant focal lengths and fixed principal points. Whitehead and Roth [30] gave a more general approach for multiple devices with varying focal lengths, but still restricted to given principle points.

Although high quality cameras can be assumed to have the principal point close to the image center, this is not generally not the case. For optical components with interfaces that are not orthogonal to the principal ray, e.g. in the case of projectors, the principal point can even be outside the image. Therefore, in many practical applications, said methods are not suitable for auto-calibration.

In order to estimate both focal lengths and principal points, Pollefeys [23] presented a least-squares method based on the absolute dual quadric supporting an arbitrary number of devices. Gherardi and Fusiello [13] built on this approach and introduced a more specific energy functional with several regularizations. This method is based on a given calibration and is essentially a post-processing. In order to converge to the global minimum, both approaches require good initialization and suitable regularizers. Finally, Lourakis and Deriche [21] presented a method that minimizes pairwise differences of the epipole-invariant Kruppa equations from [16]. These differences are weighted by covariance matrices from the numerical optimization of the fundamental matrix. This method represents the current state of the art and can handle much coarser initialization than the two previous approaches. Nevertheless, this method has some weaknesses, which will be addressed in Section 3.1.2.

All three energy-based methods are likely to fail if the principal points are far off the image center or in presence of significantly differing focal lengths. The method proposed here is based on an energy derived from epipole-invariant Kruppa equations. It will converge to the global minimum under almost all reasonable initial conditions and thus significantly extends the practical applicability.

3. Background

The basis for auto-calibration was the development of the theory of the absolute conic. The main idea is that any quadric, captured by an optical device, is projected as a conic onto the image plane and the respective epipolar lines are tangential to this conic [15]. Furthermore, the dual of the image of the absolute conic is independent of the camera pose. Its computation is equivalent to the calculation of the intrinsic calibration of the device. Given the epipolar geometry between image planes I_i and I_j, represented by a fundamental matrix F_{ij}, *Kruppa's equations* use this knowledge to describe a direct connection between F_{ij} and the intrinsic calibration matrices K_i and K_j of the respective cameras C_i and C_j.

3.1. Kruppa equations

Let e_i and e_j be the left and right epipoles computed from the left and right null-space of F_{ij} and $w_i^* = K_i K_i^\mathsf{T}$ and $w_j^* = K_j K_j^\mathsf{T}$ the duals of the absolute conic. Then *Kruppa's equations* read:

$$[e_j]_\times w_j^* [e_j]_\times = F_{ij} w_i^* F_{ij}^\mathsf{T} \tag{1}$$

$$\Leftrightarrow \quad [e_i]_\times w_i^* [e_i]_\times = F_{ij}^\mathsf{T} w_j^* F_{ij} \tag{2}$$

$[\cdot]_\times$ denotes the cross-product matrix. However, solving these equations is not practicable due to the strong dependence on in generally error-prone epipole estimates.

3.1.1 Epipole-invariant Kruppa equations

Hartley [16] expressed the equations by avoiding dependencies on the epipoles:

$$\underbrace{\frac{\sigma_1^2 v_1^\mathsf{T} w_i^* v_1}{u_2^\mathsf{T} w_j^* u_2}}_{=:\,\rho_1} = \underbrace{\frac{\sigma_1\sigma_2 v_1^\mathsf{T} w_i^* v_2}{-u_2^\mathsf{T} w_j^* u_1}}_{=:\,\rho_2} = \underbrace{\frac{\sigma_2^2 v_2^\mathsf{T} w_i^* v_2}{u_1^\mathsf{T} w_j^* u_1}}_{=:\,\rho_3} \tag{3}$$

with

$$F_{ij} = USV^\mathsf{T} = (u_1, u_2, u_3) \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \tag{4}$$

Numerators and denominators of the terms in equations (3) describe the tangent lines of the image of the absolute conic in the different views. These must be identical up to scale and are therefore considered relatively. These equations are the basis of Lourakis' method [21], which is known to be the state of the art. Since the method derived in this paper addresses weaknesses of the method, we will shortly introduce its main idea.

3.1.2 State of the art: Method of Lourakis

Lourakis *et al.* [21] proposed a nonlinear approach for approximating equations (3). The least-squares energy to be minimized is defined by

$$\underset{K_l,\, l \in \{1,\dots,C\}}{\operatorname{argmin}} \sum_{\substack{(i,j) \in D_\mathcal{F} \\ (u,v) \in D_\mathcal{K}}} \frac{(\rho_u^{ij} - \rho_v^{ij})^2}{{\sigma_{uv}^{ij}}^2}, \tag{5}$$

where K_l denotes the intrinsic calibration matrices, $D_\mathcal{F}$ the set of device pairings and $D_\mathcal{K}$ the set of combinations of Kruppa terms. σ_{uv} are confidence measures calculated by estimating the fundamental matrices.

This method extends the two-view case from Section 3.1.1 to any number of C devices by considering $\frac{C(C-1)}{2}$ pairwise fundamental matrices. Each of them provides two independent constraints, which limits the number of computable camera parameters to $C(C-1)$. The number of determinable parameters per device is sufficient for most applications and increases with the number of devices used, as explained in more detail in Section 5.

The method is known to work well for three or more devices, assuming high quality epipolar relations and good initialization of the principal points and focal lengths. Nevertheless, the method may fail in many practical situations for the following reasons:

- Weak initialization of focal lengths or principal points.
- Significantly differing focal lengths.
- Bias towards larger focal lengths.
- Significantly off-center principal points.
- Singularities of the energy.

3.2. Assumptions

For all modern devices, zero skew and square pixels can be assumed. Thus the dual of the absolute conic can be written as

$$w_l^* = K_l K_l^\mathsf{T} = \begin{pmatrix} f_l^2 + x_{p_l}^2 & x_{p_l} y_{p_l} & x_{p_l} \\ x_{p_l} y_{p_l} & f_l^2 + y_{p_l}^2 & y_{p_l} \\ x_{p_l} & y_{p_l} & 1 \end{pmatrix}, \tag{6}$$

where f_l denotes the focal length and $c_{p_l} = (x_{p_l}, y_{p_l}, 1)^\mathsf{T}$ the principal point of any device C_l.

4. Improving the state of the art

A new robust energy functional is proposed. Compared to Lourakis' it has the following beneficial properties:

- Focal lengths of different scales are treated homogeneously and unbiasedly.
- The multidimensional energy field is smooth and has no discontinuities or singularities.
- A significantly larger region of convergence to the global minimum, which is finite and uniquely defined.
- The energy function is quasi-symmetric with respect to the Kruppa curves (13), and convex with respect to the principal point.

This greatly increases the stability of the numerical optimization as well as the likelihood of convergence.

4.1. Kruppa curves of focal lengths

Using the notation of Section 3.1.1 and assumptions of Section 3.2, the terms of (3) can be written as

$$\rho_1 = \frac{f_i^2\sigma_1^2(v_{11}^2 + v_{12}^2) + \sigma_1^2(c_{p_i}^\mathsf{T} v_1)^2}{f_j^2(u_{21}^2 + u_{22}^2) + (c_{p_j}^\mathsf{T} u_2)^2} \tag{7}$$

$$\rho_2 = \frac{f_i^2\sigma_1\sigma_2(v_{11}v_{21} + v_{12}v_{22}) + \sigma_1\sigma_2(c_{p_i}^\mathsf{T} v_1)(c_{p_i}^\mathsf{T} v_2)}{-f_j^2(u_{11}u_{21} + u_{12}u_{22}) - (c_{p_j}^\mathsf{T} u_1)(c_{p_j}^\mathsf{T} u_2)} \tag{8}$$

$$\rho_3 = \frac{f_i^2\sigma_2^2(v_{21}^2 + v_{22}^2) + \sigma_2^2(c_{p_i}^\mathsf{T} v_2)^2}{f_j^2(u_{11}^2 + u_{12}^2) + (c_{p_j}^\mathsf{T} u_1)^2}, \tag{9}$$

where u_{kl} and v_{kl} denote the l-th entries of vectors u_k and v_k. With the explicit formulations of (7), (8) and (9), equations (3) of any fundamental matrix F_{ij} can be written in the form

$$\frac{f_i^2 a_{i1} + b_{i1}}{f_j^2 a_{j1} + b_{j1}} = \frac{f_i^2 a_{i2} + b_{i2}}{f_j^2 a_{j2} + b_{j2}} = \frac{f_i^2 a_{i3} + b_{i3}}{f_j^2 a_{j3} + b_{j3}} \tag{10}$$

with coefficients

$$\begin{pmatrix} a_{i1} \\ a_{i2} \\ a_{i3} \end{pmatrix} = \begin{pmatrix} \sigma_1^2(v_{11}^2 + v_{12}^2) \\ \sigma_1\sigma_2(v_{11}v_{21} + v_{12}v_{22}) \\ \sigma_2^2(v_{21}^2 + v_{22}^2) \end{pmatrix}, \quad \begin{pmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{pmatrix} = \begin{pmatrix} \sigma_1^2(c_{p_i}^\mathsf{T} v_1)^2 \\ \sigma_1\sigma_2(c_{p_i}^\mathsf{T} v_1)(c_{p_i}^\mathsf{T} v_2) \\ \sigma_2^2(c_{p_i}^\mathsf{T} v_2)^2 \end{pmatrix}$$
$$\begin{pmatrix} a_{j1} \\ a_{j2} \\ a_{j3} \end{pmatrix} = \begin{pmatrix} u_{21}^2 + u_{22}^2 \\ u_{11}u_{21} + u_{12}u_{22} \\ u_{11}^2 + u_{12}^2 \end{pmatrix}, \quad \begin{pmatrix} b_{j1} \\ b_{j2} \\ b_{j3} \end{pmatrix} = \begin{pmatrix} (c_{p_j}^\mathsf{T} u_2)^2 \\ (c_{p_j}^\mathsf{T} u_1)(c_{p_j}^\mathsf{T} u_2) \\ (c_{p_j}^\mathsf{T} u_1)^2 \end{pmatrix}. \quad (11)$$

For fixed principal points, the equations define curves, known as *Kruppa curves*, which describe direct relations between the focal lengths. For each fundamental matrix F_{ij} we define coefficient vectors:

$$d_{uv}^{ij} := \begin{pmatrix} a_{iu}a_{jv} - a_{iv}a_{ju} \\ a_{iu}b_{jv} - a_{iv}b_{ju} \\ b_{iu}a_{jv} - b_{iv}a_{ju} \\ b_{iu}b_{jv} - b_{iv}b_{ju} \end{pmatrix} \quad \text{for} \quad (u,v) \in D_\mathcal{K} \qquad (12)$$

Each equation from (11) defines a two-dimensional parametric curve that can be represented by the one-dimensional functions $\mathcal{K}_{1,uv}^{ij}$ and $\mathcal{K}_{2,uv}^{ij}$:

$$\mathcal{K}_{1,uv}^{ij}(f_j) := -\frac{f_j^2 d_{uv,3}^{ij} + d_{uv,4}^{ij}}{f_j^2 d_{uv,1}^{ij} + d_{uv,2}^{ij}} \qquad (13)$$

$$\mathcal{K}_{2,uv}^{ij}(f_i) := -\frac{f_i^2 d_{uv,2}^{ij} + d_{uv,4}^{ij}}{f_i^2 d_{uv,1}^{ij} + d_{uv,3}^{ij}} \qquad (14)$$

The curves $\mathcal{K}_{1,uv}^{ij}(f_j)$ and $\mathcal{K}_{2,uv}^{ij}(f_i)$ and the coefficients d_{uv}^{ij} are obtained by resolving equations (11) with respect to f_i and f_j. Figure 1 shows the Kruppa curves for three independent fundamental matrices (from left to right), plotted as green lines. State-of-the-art two-view techniques such as Bougnoux [3] determine the intersections of the curves to estimate the focal lengths. Having said that, Bougnoux and similar methods fail in the many cases where the Kruppa curves nearly coincide. Moreover, the curves are plotted into visualizations of the top views of the energies of Lourakis (top) and the proposed method (bottom) to illustrate relationship of the methods. The color coding indicates a rather high energy (yellow) up to a low energy (blue). This may give an idea of how the methods behave during minimization.

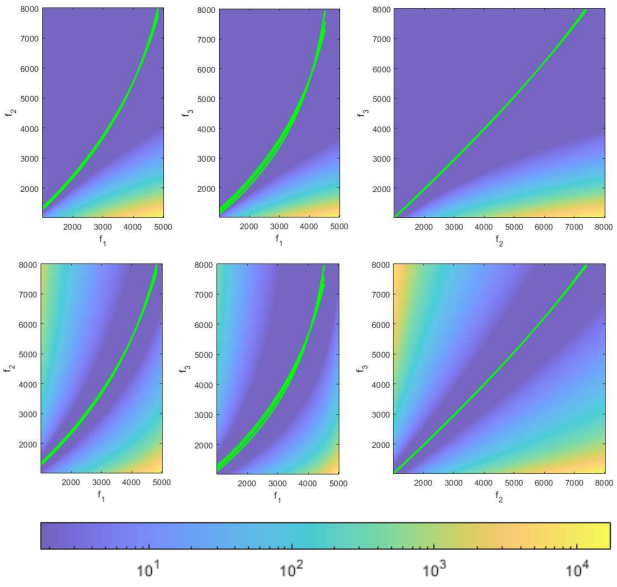

Figure 1. Top view of Lourakis' (5) (top) and the presented energy function from (15) (bottom) for several fundamental matrices. The color coding indicates a rather high energy (yellow) up to a low energy (blue) in logarithmic scale. For each fundamental matrix, the three nearly coinciding Kruppa curves are plotted in green.

4.2. Energy as relative distances to Kruppa curves

In order to establish a suitable energy term, relative Euclidean distances between focal length estimates and Kruppa curves are used, which provide a scale invariance with respect to largely different focal lengths. The new energy term reads

$$\underset{\substack{c_{p_j}, f_j \\ j \in \{1,\dots,C\}}}{\operatorname{argmin}} \sum_{\substack{(i,j) \in D_\mathcal{F} \\ (u,v) \in D_\mathcal{K}}} \left(\frac{f_i^2 - \mathcal{K}_{1,uv}^{ij}(f_j)}{f_i^2} \right)^2 + \left(\frac{f_j^2 - \mathcal{K}_{2,uv}^{ij}(f_i)}{f_j^2} \right)^2. \qquad (15)$$

By setting up the Jacobians $J_{uv,1}^{ij}$ and $J_{uv,2}^{ij}$ for each pair of energies, we can solve (15) by applying truncated Levenberg-Marquardt, with system matrix

$$A = \sum_{\substack{(i,j) \in D_\mathcal{F} \\ (u,v) \in D_\mathcal{K}}} J_{uv,1}^{ij^\mathsf{T}} J_{uv,1}^{ij} + J_{uv,2}^{ij^\mathsf{T}} J_{uv,2}^{ij} \qquad (16)$$

and inhomogenity

$$b = \sum_{\substack{(i,j) \in D_\mathcal{F} \\ (u,v) \in D_\mathcal{K}}} J_{uv,1}^{ij^\mathsf{T}} \left(1 + \frac{f_j^2 d_{uv,3}^{ij} + d_{uv,4}^{ij}}{f_i^2 f_j^2 d_{uv,1}^{ij} + f_i^2 d_{uv,2}^{ij}} \right) + J_{uv,2}^{ij^\mathsf{T}} \left(1 + \frac{f_i^2 d_{uv,2}^{ij} + d_{uv,4}^{ij}}{f_i^2 f_j^2 d_{uv,1}^{ij} + f_j^2 d_{uv,3}^{ij}} \right). \qquad (17)$$

4.3. Computational effort

Since both Lourakis' method and the proposed one are based on the singular value decomposition of $\frac{C(C-1)}{2}$ fundamental matrices, the energy functions can be set up with the same computational effort. The minimization of the energies with Levenberg-Marquardt consistently led to a faster convergence of the proposed method compared to [21], which can be explained by better condition numbers of the system matrices of the new method. Appropriate preconditioning may improve the convergence rate in both

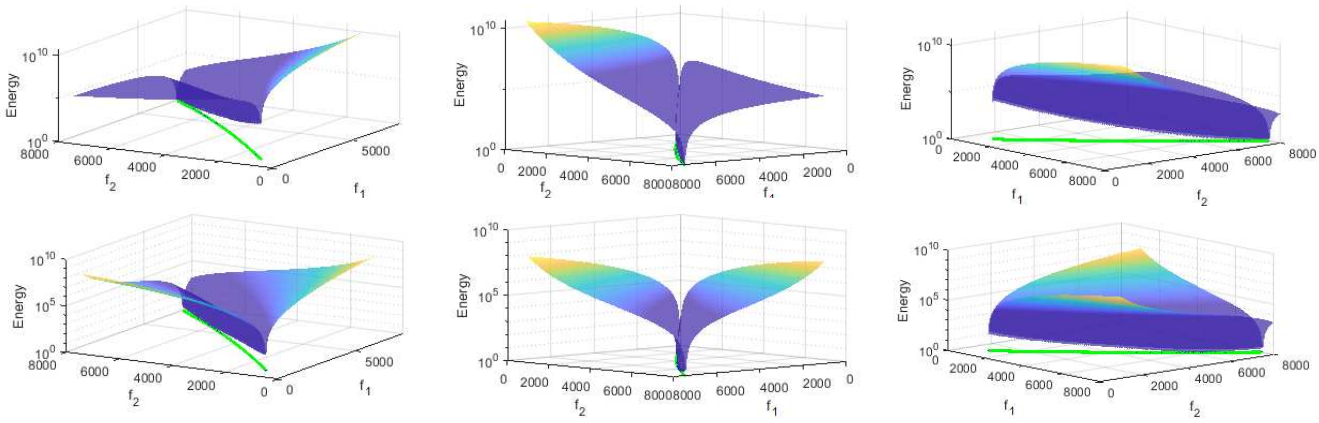

Figure 2. Kruppa curve distance energies computed from the fundamental matrix for several combinations of focal lengths $f_1, f_2 \in [1000, 8000]$ of devices with fixed principal points. The upper corresponds to Lourakis' and the lower row to the proposed method. Note that the new energy is quasi-symmetric with respect to the Kruppa curve, while the state of the art is sloped unfavorably. Plots are given in logarithmic scale.

cases. Since the running time in both cases is short and negligible compared to other calibration steps, no further investigations were performed.

4.4. Discussion

In this section we present the advantages of the proposed approach by means of visualizations of the minimized energy functional and compare it to the state of the art.

4.4.1 Case: Individual focal lengths per device

In real scenarios, the focal lengths of the devices often differ significantly. If these differences become too large, state-of-the-art methods are likely to fail if initialization is not close to the true values. Another disadvantage is the uneven slope of the gradient of Lourakis' energy in vicinity of the Kruppa curve: For small focal lengths, the slope is significantly smaller than for large ones. Therefore, a Levenberg-Marquard update will always prefer the gradient direction of the larger to the smaller focal lengths when optimizing such a system. Due to the gradient slope, the method generally tends to overestimate to focal lengths. Figure 2 compares Lourakis' energy functional (5) (top row) with the proposed one (15) (bottom row) for several combinations of focal lengths $f_1, f_2 \in [1000, 8000]$ from different perspectives. In particular, in the right subimage, the increase of the slope can be observed when increasing the values of the focal lengths. Due to the relative Euclidean distances used in (15), the new energy functional is much more homogeneous.

Moreover, it is quasi-symmetric with respect to the Kruppa curves, which avoids the preference of a particular direction over others.

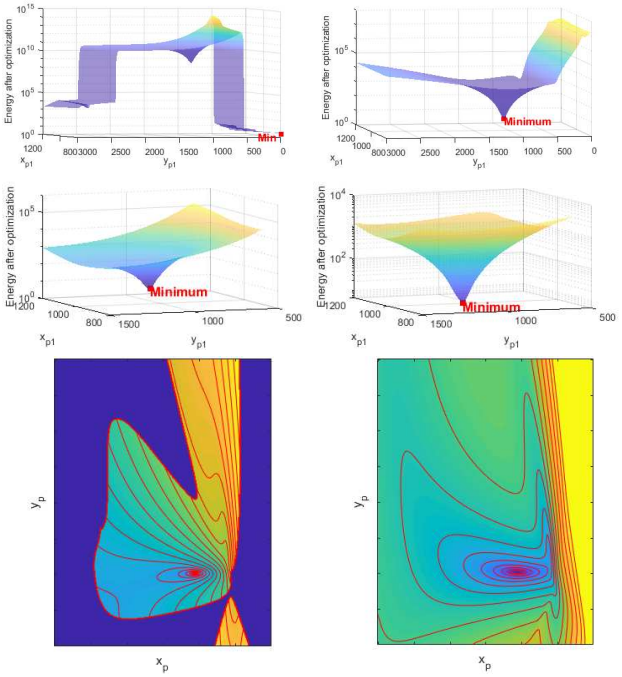

Figure 3. Energies of Lourakis (left) and the proposed Kruppa curve distance energies (right) with respect to the principal point position. While the top row shows an overview, the second row is a close-up of the area around the sought solution. Note that the location of the solution coincides for both energies. In the third row a top view of the energies of the first row is given. The region beyond the discontinuity is colored in dark blue. Please observe that the contour lines indicate significantly improved convergence properties. Plots are given in logarithmic scale for visualization.

4.4.2 Case: Principal point far off the image center

If the principal point of a device is not close to the image center, all known methods are likely to fail. In case of Lourakis' functional, the energy surface corresponding to the principal point positions has been analyzed and two main issues have been identified depicted in Figure 3 (left): The sought minimum of the energy is almost completely surrounded by a discontinuity, so that initialization beyond the discontinuity cannot converge (top left); Even initialization on the plateau, convergence cannot be guaranteed because the entire plateau is inclined. (top left). In contrast, the proposed energy functional is globally continuous, smooth and convex (Figure 3 top right). Comparing the second row of Figure 3 demonstrates that the sought minimum of both methods coincide. Although it can be assumed that the principal point of modern cameras is close to the image center, optical systems in industrial setups often have a displaced principal point. Reasons for this include additional lens assemblies, obstacles such as glass plates or liquids, and periscopic systems. Also for projection systems, e.g. used in active scanning solutions, it is not unusual for the principal point to be completely outside the image. Initialization of the principal point with the image center will often be outside the convergence plateau of Lourakis' method. The third row of Figure 3 shows slightly enlarged top views of Lourakis' and the proposed energy functional of the first row. The contour plots give a good indication of the improved convergence properties of the newly proposed functional.

5. Application and evaluation

Both the state-of-the-art method and the proposed approach, are based on Kruppa's equations. These equations provide two independent constraints for each fundamental matrix. Therefore, the number of computable parameters is determined by the number of devices (see Table 1). For two devices, only two parameters can be estimated based on the single fundamental matrix. This case is the most basic and most frequently examined system setup. With four or more devices, the problem of intrinsic calibration is well defined and theoretically all parameters can be estimated. Nevertheless, even the calibration of four devices in practice can still be a challenge. A particularly interesting case is the use of three devices, such as two cameras and a projector, as found in most active scanning setups. For all three devices, the focal length can be estimated. With the remaining constraints, the principal point of the projector can be estimated, which is usually far off the image center.

# Devices	# Basic Equations	Computable Parameters
2	2	f_1, f_2 or x_{p_l}, y_{p_l}
3	6	f_1, f_2, f_3, x_{p_1}, y_{p_1}
4	12	f_l,x_{p_l}, y_{p_l}, $l = 1, ..., 4$
$\vdots$	$\vdots$	$\vdots$
C	$C(C-1)$	f_l,x_{p_l}, y_{p_l}, $l = 1, ..., C$

Table 1. Overview of degrees of freedom in terms of the number of devices and useful calibration parameters that can be computed.

For the evaluation we consider three cases, i.e. two, three and four devices. In order to investigate the stability of the methods, we calculate probability maps that visualize the convergence chances for different initializations and thus represent the convergence regions of the methods. To calculate these probability maps we used fixed setups with two, three and four devices and fixed extrinsic and intrinsic parameters. A total of 16 different scenes were recorded with these setups. The scenarios were selected in such a way that they cover a multitude of different practical application scenarios. From the different scenes, fundamental matrices have been computed using the technique described in [11]. The matches used for the computations were previously validated to avoid falsification by outliers. Consequently, the resulting 16 fundamental matrices per setup approximate exactly the same epipolar relations with uncorrelated numerical errors because they are computed from different matches from different scenes. Applying the methods under investigation on a fundamental matrix for all combinations of initial focal lengths $f_1, f_2 \in [1, 10000]$ leads to a binary map, which indicate whether the method converged or not. The binary maps of all the fundamental matrices have been combined into probability maps depicted in Figures 4, 5 and 7. Therefore, the percentage at which convergence has been achieved color-codes the maps. Green indicates a very high probability of convergence, while red indicates either divergence or convergence to an incorrect value. Yellow depicts regions with approximately 50% chance to converge to the correct value.

Since the probability maps are not dependent on individual scenes, correspondences, or fundamental matrices, they are meaningful indicators for the convergence behavior of the procedures.

In the following, the focal lengths are given in terms of sensor pixel size. For typical devices, a plausible range would be in $[500, 15000]$. Depending on the sensor size, this would correspond to approximately $[18mm, 50mm]$. the principal points are given in terms of image pixel size, depending on the resolution.

5.1. Two-view focal length estimation

In the case of two cameras, the principal points are usually assumed to be in the image centers. Therefore, in most cases only the focal lengths are computed. Bougnoux [3] gave a famous formula to calculate the focal lengths di-

rectly:

$$f_1 = \sqrt{-\frac{c_{p_2}^\mathsf{T}[e_2]_\times \tilde{I} F_{12} c_{p_1} c_{p_1}^\mathsf{T} F_{23}^\mathsf{T} c_{p_2}}{c_{p_2}^\mathsf{T}[e_2]_\times \tilde{I} F_{12} \tilde{I} F_{12}^\mathsf{T} c_{p_2}}} \tag{18}$$

$$f_2 = \sqrt{-\frac{c_{p_1}^\mathsf{T}[e_1]_\times \tilde{I} F_{12}^\mathsf{T} c_{p_2} c_{p_2}^\mathsf{T} F_{23} c_{p_1}}{c_{p_1}^\mathsf{T}[e_1]_\times \tilde{I} F_{12}^\mathsf{T} \tilde{I} F_{12} c_{p_1}}} \tag{19}$$

where c_{p_l} and e_l denote the principal points and epipoles of camera l in homogeneous coordinates. $[.]_\times$ denotes the cross-product matrix and $\tilde{I} = \mathrm{diag}(1, 1, 0)$ is the embedding of the two-dimensional identity matrix. Unfortunately, this formula fails in many practical situations, as already mentioned in Section 4. Although it is not well suited for auto-calibration, it can still be used as initialization for iterative methods in case it is not degenerated.

For the two camera case, Bougnoux's, Lourakis' and the proposed method should be compared. Bougnoux's method is a direct one and therefore does not depend on initialization. Therefore, no region of convergence can be determined and visualized in the following. Having said this, Bougnoux's method failed in most cases during our tests, while the iterative methods could still converge when initialized accordingly.

For the current investigation, the focal lengths of the devices were chosen to be approximately equal in order to resemble the practical case of manually adjusting the cameras. Despite that, exactly the same focal lengths would lead to a degeneration that would be perfectly captured by Sturm's method [26]. However, this special case rarely occurs in practice.

Inspecting Figure 4 for the two camera case with similar focal lengths, it can be observed that the proposed method converges for nearly all initializations, while Lourakis' method only converges in a region of radius of approximately 1500 pixels relative to the true solution.

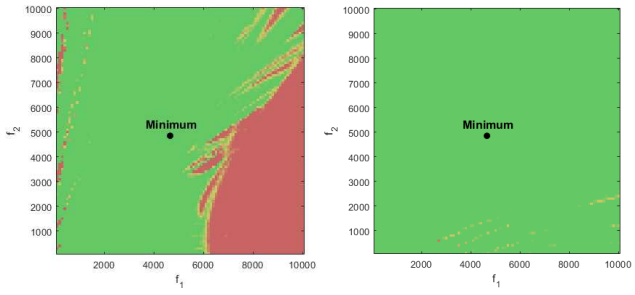

Figure 4. Comparison of the convergence behavior in the two-view case with similar focal lengths. Colors visualize the probability of successful convergence to the correct solution for different combinations of initial focal lengths. Left: Lourakis' method, right: the proposed method.

5.1.1 Strongly differing focal lengths

In the case of strongly varying focal lengths, even more benefits can be achieved. Figure 5 shows the convergence probability map of a similar configuration as in Figure 4. While the method of Lourakis converges in a region with a radius of only 500 pixels, the proposed method converges in almost all cases.

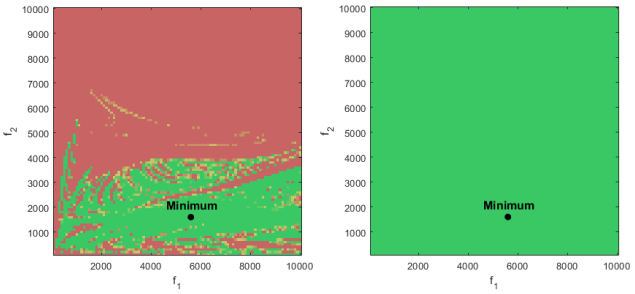

Figure 5. Comparison of the convergence behavior for focus optimization in the two-view case for dissimilar focal lengths. The axes represent the respective focal lengths. Left: Lourakis' method, right: the proposed method.

5.2. Three-view intrinsic calibration

The most interesting case for practical application is a three device setup. According to Table 1 the focal values plus the position of the principal point of one device can be estimated. This allows the calibration of setups consisting of two cameras and one projector, which is of practical importance, as it is common for modern structured light setups. In this case, it is assumed that the principal points of the cameras are in the image center, while their focus values can be very different. The projector is assumed to have a completely independent focal length and an extreme position of the principal point, usually near the image border.

We again have a system setup with fixed extrinsics and intrinsics. Fundamental matrices are computed from 16 scenes similar to the previous test. Now that we have three devices, the respective probabilistic convergence maps would be three-dimensional. In order to achieve an expressive visualization in two dimensions, the focal length of the projector was initialized by $f_3 \in \{1, 10, 100, 1000, 10000\}$ and the resulting maps averaged. Figure 7 depicts the convergence regions for Lourakis' method on the left and the proposed method on the right. As can be clearly observed, Lourakis' method does not provide a secure convergence region, i.e. a region of focal length selections that converges for an arbitrary principal point.

In order to assess the convergence behavior of both methods, a second test set was carried out. Therefore, three devices (two cameras and a projector) with focal lengths of 4000, 5600, and 5800 were selected. All principal points

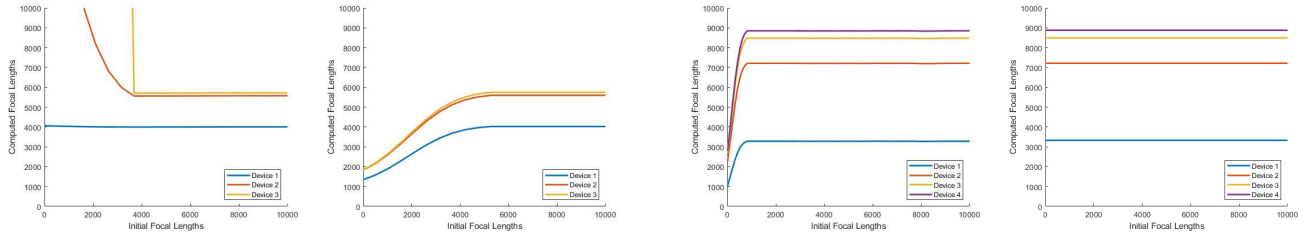

Figure 6. Comparison of the convergence behavior in the three-view case (first Louraki's, second the proposed method) and four-view case (third Louraki's, fourth the proposed method) with extreme intializations of the principal points and the focal lengths outside the convergence regions of the methods. While the state of the art diverges, the proposed method gracefully converges towards a nearby local minimum in the three-view case, providing at least a reasonable estimate for further calibration steps like bundle-adjustment. In the four-view case the newly proposed method always converges to the correct solution, while Lourakis' method may still fail.

were initialized at the image centers. The principal points of the cameras were fixed, while the principal point of the projector was subject to optimization. All devices were initialized with the same focal lengths between 1 and 10000. For initial values less than 4000, both methods fail. While the proposed method gracefully converges towards a nearby local minimum, the state of the art diverges (see Figure 6).

These results have inspired us to devise an algorithm that allows stable auto-calibration of three-device setups in nearly any practical case. It is based on Bougnoux's method for initialization and exploits the newly proposed method for stable optimization of the intrinsic camera parameters. The algorithm is outlined in the appendix.

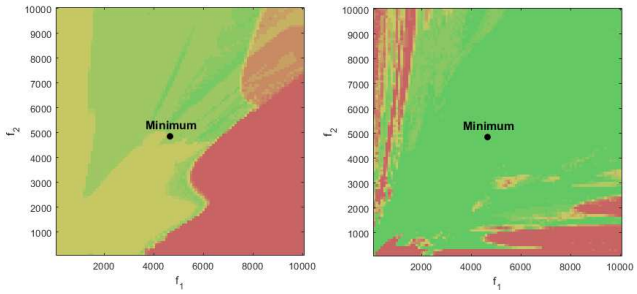

Figure 7. Comparison of the convergence behavior for focus optimization in the three-view case. The axes represent two of the three focal lengths, the third one is visualized as a mean projection along the third coordinate axis. Left: Lourakis' method, right: the proposed method.

5.3. Multi-view intrinsic calibration

In the multiview case of four or more devices, the problem is much easier to solve. Theoretically, it is possible to fully calibrate all devices, including focal lengths and principal points. In practice, we find that Lourakis' method does not converge if the focal lengths are initialized far too small, while the proposed method converges in all situations. However, we find that the focal lengths must be so small that this case can be neglected (see Figure 6 for visualization).

In the case of five or more devices, the stability of the convergence of each procedure increases. Due to the inherent difficulty of visualizing multi-dimensional data and the fact that both methods perform well in practice, we omit respective visualization.

6. Conclusions

In this paper we presented a method for robust intrinsic camera calibration from epipolar geometry of setups with at least two devices. Contrary to the state of the art, the method converges to the global solution for nearly all reasonable initializations and enables the calibration of projectors and low quality devices. It therefore has a major impact on active scanning techniques that can now be calibrated from scratch, including the active element. It has been shown that and why the applicability of the state-of-the-art method is subject to systematic limitations.

Algorithm 1 Stable Intrinsic Calibration of Structured Light Systems with Two Cameras and a Projector

Input: Fundamental matrices $F_{P,C1}$, $F_{P,C2}$ and $F_{C1,C2}$ between projector (P) and cameras (C1,C2).

1: Initialize focal lengths $f_P^{\text{init}} = f_{C1}^{\text{init}} = f_{C2}^{\text{init}} := 1000$ and principal points $c_{p_P}^{\text{init}}, c_{p_{C1}}, c_{p_{C2}}$ to image centers.

2: Apply (18) and (19) to $F_{C1,C2}$.
If successful (plausible results in range $[1; 100000]$) update f_{C1}^{init} and f_{C2}^{init}, otherwise skip.

3: Solve (15) with fixed f_{C1}^{init}, f_{C2}^{init} to update f_P^{init} and $c_{p_P}^{\text{init}}$.

4: Solve (15) to get refined values f_P, f_{C1}, f_{C2} and c_{p_P}.

5: Build intrinsic calibration matrices K_l for $l \in \{1, ..., 3\}$ from computed focal lengths and principal points.

Output: Refined intrinsic calibration matrices K_l.

This work was partially funded by the projects MARMORBILD (03VP00293) and VIDETE (01W18002) of the German Federal Ministry of Education and Research (BMBF).

References

[1] S. Agarwal, N. Snavely, S. M. Seitz, and R. Szeliski. Bundle adjustment in the large. In *European conference on computer vision*, pages 29–42. Springer, 2010.

[2] J. P. Barreto and K. Daniilidis. Fundamental matrix for cameras with radial distortion. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 1, pages 625–632. IEEE, 2005.

[3] S. Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Computer Vision, 1998. Sixth International Conference on*, pages 790–796. IEEE, 1998. 1, 2, 4, 6

[4] J. H. Brito, R. Angst, K. Köser, and M. Pollefeys. Radial distortion self-calibration. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 1368–1375. IEEE, 2013.

[5] D. C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering and Remote Sensing*, 1966.

[6] M. Byrod, Z. Kukelova, K. Josephson, T. Pajdla, and K. Astrom. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[7] X. Chen, J. Xi, Y. Jin, and J. Sun. Accurate calibration for a camera–projector measurement system based on structured light projection. *Optics and Lasers in Engineering*, 47(3-4):310–319, 2009.

[8] G. Csurka, C. Zeller, Z. Zhang, and O. D. Faugeras. Characterizing the uncertainty of the fundamental matrix. *Computer vision and image understanding*, 68(1):18–36, 1997.

[9] O. Faugeras. Stratification of three-dimensional vision: projective, affine, and metric representations. *JOSA A*, 12(3):465–484, 1995.

[10] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In *European conference on computer vision*, pages 321–334. Springer, 1992. 2

[11] T. Fetzer, G. Reis, and D. Stricker. Robust auto-calibration for practical scanning setups from epipolar and trifocal relations. In *2019 sixteenth IAPR International Conference on Machine Vision Applications (MVA)*. IEEE, 2019. 1, 6

[12] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

[13] R. Gherardi and A. Fusiello. Practical autocalibration. In *European Conference on Computer Vision*, pages 790–801. Springer, 2010. 2

[14] R. Hartley. Extraction of focal lengths from the fundamental matrix. *Unpublished manuscript*, 1993. 2

[15] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1, 2

[16] R. I. Hartley. Kruppa's equations derived from the fundamental matrix. *IEEE Transactions on pattern analysis and machine intelligence*, 19(2):133–135, 1997. 2, 3

[17] Z. Kukelova, J. Heller, M. Bujnak, A. Fitzgibbon, and T. Pajdla. Efficient solution to the epipolar geometry for radially distorted cameras. In *Proceedings of the IEEE international conference on computer vision*, pages 2309–2317, 2015.

[18] Y.-t. Kwak, J.-w. Hwang, and C.-j. Yoo. A new damping strategy of levenberg-marquardt algorithm for multilayer perceptrons. *Neural Network World*, 21(4):327, 2011.

[19] F. Li, H. Sekkati, J. Deglint, C. Scharfenberger, M. Lamm, D. Clausi, J. Zelek, and A. Wong. Simultaneous projector-camera self-calibration for three-dimensional reconstruction and projection mapping. *IEEE Transactions on Computational Imaging*, 3(1):74–83, 2017.

[20] M. I. Lourakis and R. Deriche. *Camera self-calibration using the singular value decomposition of the fundamental matrix: From point correspondences to 3D measurements*. PhD thesis, INRIA, 1999.

[21] M. I. Lourakis and R. Deriche. *Camera self-calibration using the Kruppa equations and the SVD of the fundamental matrix: The case of varying intrinsic parameters*. PhD thesis, INRIA, 2000. 1, 2, 3, 4

[22] M. O'Toole. *Optical Linear Algebra for Computational Light Transport*. PhD thesis, University of Toronto (Canada), 2016.

[23] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999. 2

[24] L. Romero and C. Gomez. Correcting radial distortion of cameras with wide angle lens using point correspondences. In *Scene Reconstruction Pose Estimation and Tracking*. In-Tech, 2007.

[25] G. P. Stein. Lens distortion calibration using point correspondences. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 602–608. IEEE, 1997.

[26] P. Sturm. On focal length calibration from two views. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2001. 2, 7

[27] G. Terzakis, P. Culverhouse, G. Bugmann, S. Sharma, and R. Sutton. A recipe on the parameterization of rotation matrices for non-linear optimization using quaternions. Technical report, Technical report MIDAS. SMSE. 2012. TR. 004, Marine and Industrial Dynamic Analysis School of Marine Science and Engineering, Plymouth University, 2012.

[28] P. H. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and vision Computing*, 15(8):591–605, 1997.

[29] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustmenta modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer, 1999. 2

[30] A. Whitehead and G. Roth. Estimating intrinsic camera parameters from the fundamental matrix using an evolutionary approach. *EURASIP Journal on Advances in Signal Processing*, 2004(8):412751, 2004. 2

[31] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 1, pages 407–411. IEEE, 1996.

[32] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International journal of computer vision*, 27(2):161–195, 1998. 1