Proc. S+SSPR 2012, November 2012, Hiroshima, Japan, pp. 11–30.

Optimization Techniques for Geometric Estimation: Beyond Minimization

Kenichi Kanatani

Department of Computer Science, Okayama University, Okayama 700-8530, Japan
kanatani@suri.cs.okayama-u.ac.jp

Abstract. We overview techniques for optimal geometric estimation from noisy observations for computer vision applications. We first describe estimation techniques based on minimization of given cost functions: least squares (LS), maximum likelihood (ML), which includes reprojection error minimization (Gold Standard) as a special case, and Sampson error minimization. We then formulate estimation techniques not based on minimization of any cost function: iterative reweight, renormalization, and hyper-renormalization. Showing numerical examples, we conclude that hyper-renormalization is robust to noise and currently is the best method.

1 Introduction

One of the most important tasks of computer vision is to compute the 2-D and 3-D shapes of objects exploiting *geometric constraints*, by which we mean properties that can be described by relatively simple equations such as the objects being lines or planes, their being parallel or orthogonal, and the camera imaging geometry being perspective projection. We call the inference based on such geometric constraints *geometric estimation*. In the presence of noise, however, the assumed constraints do not exactly hold. To do geometric estimation "optimally" in the presence of noise, a lot of efforts have been made since 1980s by many researchers. This paper summarizes that history and reports the latest results.

2 Preliminaries

2.1 Definition of Geometric Estimation

The geometric estimation problem we consider here is defined as follows. We observe some quantity $\boldsymbol{x}$ (a vector), which is assumed to satisfy in the absence of noise an equation

$$F(\boldsymbol{x};\boldsymbol{\theta}) = 0, \tag{1}$$

parameterized by unknown vector $\boldsymbol{\theta}$. This equation is called the *geometric constraint*. Our task is to estimate the parameter $\boldsymbol{\theta}$ from noisy instances $\boldsymbol{x}_\alpha$, $\alpha = 1$, ..., N, of $\boldsymbol{x}$. Many computer vision problems are formulated in this way, and we

G.L. Gimel'farb et al. (Eds.): SSPR & SPR 2012, LNCS 7626, pp. 11–30, 2012.

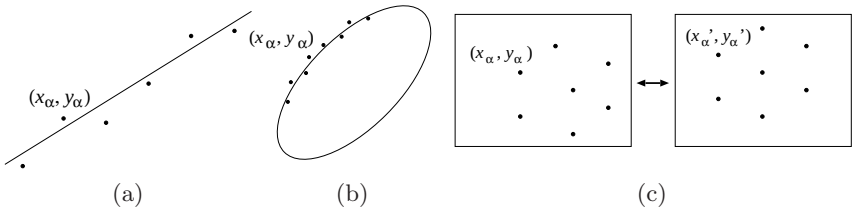


Fig. 1. (a) Line fitting. (b) Ellipse fitting. (c) Fundamental matrix computation.

can compute from the estimated $\boldsymbol{\theta}$ the positions, the shapes, and the motions of the objects we are viewing. In many problems, we can reparameterize the problem so that the constraint is liner in the parameter $\boldsymbol{\theta}$ (but generally nonlinear in the data $\boldsymbol{x}$). Then, Eq. (1) has the form

$$(\boldsymbol{\xi}(\boldsymbol{x}), \boldsymbol{\theta}) = 0, \tag{2}$$

where $\boldsymbol{\xi}(\boldsymbol{x})$ is a vector-valued nonlinear function of $\boldsymbol{x}$. In this paper, we denote the inner product of vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ by $(\boldsymbol{a}, \boldsymbol{b})$. Equation (2) implies that the scale of $\boldsymbol{\theta}$ is indeterminate, so we hereafter normalize $\boldsymbol{\theta}$ to unit norm: $\|\boldsymbol{\theta}\| = 1$.

Example 1. (Line fitting) To a given point sequence (x_α, y_α), $\alpha = 1, ..., N$, we fit a line

$$Ax + By + C = 0. \tag{3}$$

(Fig. 1(a).) If we define

$$\boldsymbol{\xi}(x, y) \equiv (x,\, y,\, 1)^\top, \qquad \boldsymbol{\theta} \equiv (A,\, B,\, C)^\top, \tag{4}$$

the line equation is written as

$$(\boldsymbol{\xi}(x, y), \boldsymbol{\theta}) = 0. \tag{5}$$

Example 2. (Ellipse fitting) To a given point sequence (x_α, y_α), $\alpha = 1, ..., N$, we fit an ellipse

$$Ax^2 + 2Bxy + Cy^2 + 2(Dx + Ey) + F = 0. \tag{6}$$

(Fig. 1(b).) If we define

$$\boldsymbol{\xi}(x, y) \equiv (x^2,\, 2xy,\, y^2,\, 2x,\, 2y,\, 1)^\top, \qquad \boldsymbol{\theta} \equiv (A,\, B,\, C,\, D,\, E,\, F)^\top, \tag{7}$$

the ellipse equation is written as

$$(\boldsymbol{\xi}(x, y), \boldsymbol{\theta}) = 0. \tag{8}$$

Example 3. (Fundamental matrix computation) Corresponding points (x, y) and (x', y') in two images of the same 3-D scene taken from different positions satisfy the *epipolar equation* [8]

$$(\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \boldsymbol{F} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}) = 0, \tag{9}$$

where $\boldsymbol{F}$ is called the *fundamental matrix*, from which we can compute the camera positions and the 3-D structure of the scene [8] (Fig. 1(c)). If we define

$$\boldsymbol{\xi}(x, y, x', y') \equiv (xx',\ xy',\ x,\ yx',\ yy',\ y,\ x',\ y',\ 1)^\top, \tag{10}$$

$$\boldsymbol{\theta} \equiv (F_{11},\ F_{12},\ F_{13},\ F_{21},\ F_{22},\ F_{23},\ F_{31},\ F_{32},\ F_{33})^\top, \tag{11}$$

the epipolar equation is written as

$$(\boldsymbol{\xi}(x, y, x', y'), \boldsymbol{\theta}) = 0. \tag{12}$$

2.2 Modeling of Noise

In the context of image analysis, "noise" means *uncertainty of image processing operations*, rather than random fluctuations over time or space as commonly understood in physics and communications. It reflects the fact that standard image processing operations such as feature extraction and edge detection are not perfect and do not necessarily output exactly what we are looking for. We model this uncertainty in statistical terms: the observed value $\boldsymbol{x}_\alpha$ is regarded as a perturbation from its true value $\bar{\boldsymbol{x}}_\alpha$ by an independent random Gaussian variable $\Delta\boldsymbol{x}_\alpha$ of mean $\boldsymbol{0}$ and covariance matrix $V[\boldsymbol{x}_\alpha]$. Furthermore, $V[\boldsymbol{x}_\alpha]$ is assumed to be known *up to scale*. Namely, we write it as

$$V[\boldsymbol{x}_\alpha] = \sigma^2 V_0[\boldsymbol{x}_\alpha] \tag{13}$$

for some unknown constant σ, which we call the *noise level*. The matrix $V_0[\boldsymbol{x}_\alpha]$, which we call the *normalized covariance matrix*, describes the orientation dependence of uncertainty in relative terms and is assumed to be known. The separation of $V[\boldsymbol{x}_\alpha]$ into σ^2 and $V_0[\boldsymbol{x}_\alpha]$ is merely a matter of convenience; there is no fixed rule. This convention is motivated by the fact that optimal estimation can be done, as shown shortly, only from the knowledge of $V_0[\boldsymbol{x}_\alpha]$.

If the observation $\boldsymbol{x}_\alpha$ is regarded as a random variable, its nonlinear mapping $\boldsymbol{\xi}(\boldsymbol{x}_\alpha)$, which we write $\boldsymbol{\xi}_\alpha$ for short, is also a random variable. Its covariance matrix $V[\boldsymbol{\xi}_\alpha] = \sigma^2 V_0[\boldsymbol{\xi}_\alpha]$ is evaluated to a first approximation in terms of the Jacobi matrix $\partial\boldsymbol{\xi}/\partial\boldsymbol{x}$ of the mapping $\boldsymbol{\xi}(\boldsymbol{x})$ as follows:

$$V_0[\boldsymbol{\xi}_\alpha] = \left.\frac{\partial\boldsymbol{\xi}}{\partial\boldsymbol{x}}\right|_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha} V_0[\boldsymbol{x}_\alpha] \left.\frac{\partial\boldsymbol{\xi}}{\partial\boldsymbol{x}}\right|^\top_{\boldsymbol{x}=\bar{\boldsymbol{x}}_\alpha}. \tag{14}$$

This expression contains the true value $\bar{\boldsymbol{x}}_\alpha$, which is replaced in actual computation by the observation $\boldsymbol{x}_\alpha$. It has been confirmed by experiments that this replacement does not practically affect the final result. It has also been confirmed that upgrading the first approximation to higher orders does not have any practical effect.

2.3 Geometric Models for Geometric Estimation

One of the most prominent distinctions of the geometric estimation from the traditional statistical estimation is that the starting equation, Eq. (1) (or Eq. (2)), which we call the *geometric model*, only specifies the necessary constraint and does not explain the mechanism as to how the data $\boldsymbol{x}_\alpha$ are generated. Hence, we cannot express $\boldsymbol{x}_\alpha$ in terms of the parameter $\boldsymbol{\theta}$ as an explicit function.

Another big difference is that while the traditional statistical estimation is based on *repeated* observations regarded as sampled from the statistical model (= probability density), and hence accuracy vs. the number N of observations in the asymptotic limit $N \to \infty$ is a major concern, geometric estimation is done from *one* set of data $\{\boldsymbol{x}_1, ..., \boldsymbol{x}_N\}$. Naturally, the estimation accuracy increases with less observation uncertainty. Hence, accuracy vs. the noise level σ in the limit of $\sigma \to 0$ is a major concern [14].

In computer vision applications, the asymptotic analysis of $N \to \infty$ does not have much sense, because the number of data obtained by image processing operations is limited in number. Usually, the output of an image processing operation is accompanied by its reliability index, and we select only those data that have high reliability indices. If we want to increase the number of data, we necessarily need to include those with low reliability, but they are often misdetections. Despite the basic differences, however, two approaches exist in both statistical and geometric estimation domains:

Minimization approach We choose the value $\boldsymbol{\theta}$ that minimizes a specified cost function. This is regarded as the standard for computer vision applications.

Non-minimization approach We compute the value $\boldsymbol{\theta}$ by solving a set of equations, called *estimating equations* [6], which need not be derivatives of some function. Hence, the solution does not necessarily minimize any cost function. In traditional statistical estimation domains, this approach is regarded as more general and more flexible with a possibility of yielding better solutions than the minimization approach, but it is not widely recognized in computer vision research.

2.4 KCR Lower Bound

For minimization or non-minimization approaches, there exists a theoretical accuracy limit. We assume that the true values $\bar{\boldsymbol{\xi}}_\alpha$ of the observations $\boldsymbol{\xi}_\alpha$ satisfy the constraint $(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{\theta}) = 0$ for some $\boldsymbol{\theta}$. If it is estimated from the observation $\{\boldsymbol{\xi}_\alpha\}_{\alpha=1}^N$ by some means, the estimate $\hat{\boldsymbol{\theta}}$ is as a function $\hat{\boldsymbol{\theta}}(\{\boldsymbol{\xi}_\alpha\}_{\alpha=1}^N)$ of $\{\boldsymbol{\xi}_\alpha\}_{\alpha=1}^N$, called an *estimator* of $\boldsymbol{\theta}$. Let $\Delta\boldsymbol{\theta}$ be its error, i.e., write $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta} + \Delta\boldsymbol{\theta}$, and define the covariance matrix of $\hat{\boldsymbol{\theta}}$ by

$$V[\hat{\boldsymbol{\theta}}] = E[\Delta\boldsymbol{\theta}\Delta\boldsymbol{\theta}^\top], \tag{15}$$

where $E[\,\cdot\,]$ denotes expectation over data uncertainty. If we can assume that

– each $\boldsymbol{\xi}_\alpha$ is perturbed from its true value $\bar{\boldsymbol{\xi}}_\alpha$ by independent Gaussian noise of mean $\mathbf{0}$ and covariance matrix $V[\boldsymbol{\xi}_\alpha] = \sigma^2 V_0[\boldsymbol{\xi}_\alpha]$, and
– the function $\hat{\boldsymbol{\theta}}(\{\boldsymbol{\xi}_\alpha\}_{\alpha=1}^N)$ is an *unbiased estimator*, i.e., $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$ identically holds for whatever $\boldsymbol{\theta}$,

then the following inequality holds [3, 11, 12, 14].

$$V[\hat{\boldsymbol{\theta}}] \succ \frac{\sigma^2}{N}\Big(\frac{1}{N}\sum_{\alpha=1}^{N}\frac{\bar{\boldsymbol{\xi}}_\alpha\bar{\boldsymbol{\xi}}_\alpha^\top}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}\Big)^-. \tag{16}$$

Here, $\boldsymbol{A} \succ \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is a positive semidefinite symmetric matrix, and $(\,\cdot\,)^-$ denotes the pseudo inverse. Chernov and Lesort [3] called the right side Eq. (16) *Kanatani-Cramer-Rao* (KCR) *lower bound*.

3 Minimization Approach

First, we overview popular geometric estimation techniques for computer vision that are based on the minimization approach.

3.1 Least Squares (LS)

Since the true values $\bar{\boldsymbol{\xi}}_\alpha$ of the observations $\boldsymbol{\xi}_\alpha$ satisfy $(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{\theta}) = 0$, we choose the value $\boldsymbol{\theta}$ that minimizes

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2 \tag{17}$$

for noisy observations $\boldsymbol{\xi}_\alpha$ subject to the constraint $\|\boldsymbol{\theta}\| = 1$. This can also be viewed as minimizing $\sum_{\alpha=1}^{N}(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2/\|\boldsymbol{\theta}\|^2$. Equation (17) can be rewritten in the form

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2 = \frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{\theta}^\top\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top\boldsymbol{\theta} = (\boldsymbol{\theta}, \underbrace{\frac{1}{N}\sum_{\alpha=1}^{N}\boldsymbol{\xi}_\alpha\boldsymbol{\xi}_\alpha^\top}_{\equiv\boldsymbol{M}}\boldsymbol{\theta}) = (\boldsymbol{\theta}, \boldsymbol{M}\boldsymbol{\theta}), \tag{18}$$

which is a quadratic form of $\boldsymbol{M}$. As is well known, the unit vector $\boldsymbol{\theta}$ that minimizes this form is given by the unit eigenvector of $\boldsymbol{M}$ for the smallest eigenvalue.

Since the sum of squares is minimized, this method is called *least squares* (*LS*). Equation (17) is often called the *algebraic distance*, so this method is also called *algebraic distance minimization*. Because the solution is directly obtained without any search, LS is widely used in many applications. However, it has been observed that the solution has a large statistical bias. For ellipse fitting in *Example 2* (Sec. 2.1), for instance, the fitted ellipse is almost always smaller than the true shape. For this reason, LS is not suited for accurate estimation. However, LS is convenient for rough estimation for guiding image processing, for the outlier-detection voting, and for initializing iterative optimization schemes.

3.2 Maximum Likelihood (ML)

If the noise in each $\boldsymbol{x}_\alpha$ is an independent Gaussian variable of mean $\boldsymbol{0}$ and covariance matrix $V[\boldsymbol{x}_\alpha] = \sigma^2 V_0[\boldsymbol{x}_\alpha]$, the *Mahalanobis distance* of the observations $\{\boldsymbol{x}_\alpha\}$ from their true values $\{\bar{\boldsymbol{x}}_\alpha\}$ is

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha)), \tag{19}$$

and the likelihood of $\{\boldsymbol{x}_\alpha\}$ is written as $Ce^{-NJ/2\sigma^2}$, where C is a normalization constant that does not depend on $\bar{\boldsymbol{x}}_\alpha$ or $\boldsymbol{\theta}$. Thus, *maximum likelihood* (*ML*) is equivalent to minimizing Eq. (19) subject to the constraint

$$(\boldsymbol{\xi}(\bar{\boldsymbol{x}}_\alpha), \boldsymbol{\theta}) = 0. \tag{20}$$

As a special case, if the noise is homogeneous, i.e., independent of α, and isotropic, i.e., independent of orientation, we can write $V_0[\boldsymbol{x}_\alpha] = \boldsymbol{I}$ (the identity), which reduces Eq. (19) to the *geometric distance*

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}\|\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha\|^2. \tag{21}$$

Minimizing this subject to Eq. (20) is called *geometric distance minimization* by computer vision researchers and *total least squares* (*TLS*) by numerical analysis researchers[1]. If $\bar{\boldsymbol{x}}_\alpha$ represents the projection of the assumed 3-D structure onto the image plane and $\boldsymbol{x}_\alpha$ is its actually observed positions, Eq. (21) is called the *reprojection error*. Minimizing it subject to Eq. (20) is often called *reprojection error minimization*.

Geometrically, ML can be interpreted to be fitting to N points $\boldsymbol{x}_\alpha$ in the data space the parameterized hypersurface $(\boldsymbol{\xi}(\boldsymbol{x}), \boldsymbol{\theta}) = 0$ by adjusting $\boldsymbol{\theta}$, where the discrepancy of the points from the surface is measured not by the Euclid distance but by the Mahalanobis distance of Eq. (19), which inversely weights the data by their covariances, thereby imposing heavier penalties on the points with higher certainty. In the field of computer vision, this approach is widely regarded as the ultimate method and often called the *Gold Standard* [8]. However, this is a highly nonlinear optimization problem and difficult to solve by a direct means. The difficulty stems from the fact that Eq. (20) is an implicit function of $\bar{\boldsymbol{x}}_\alpha$. If we could solve Eq. (20) for $\bar{\boldsymbol{x}}_\alpha$ to express it as an explicit function of $\boldsymbol{\theta}$, we could substitute it into Eq. (19) to obtain an unconstrained optimization problem for $\boldsymbol{\theta}$ alone, but this is generally not possible. In *Examples 1* (line fitting), *2* (ellipse fitting), and *3* (fundamental matrix computation) in Sec. 2.1, for instance, we cannot express (x, y) or (x, y, x', y') in terms of $\boldsymbol{\theta}$.

[1] If the data $\boldsymbol{x}_\alpha$ are 2-D positions $\boldsymbol{x}_\alpha = (x_\alpha, y_\alpha)$ and the y-coordinate alone undergoes noise, we only need to minimize $(1/N)\sum_{\alpha=1}^{N}(y_\alpha - \bar{y}_\alpha)^2$. In general, if only some components of the data $\boldsymbol{x}_\alpha$ contain noise, the problem is called *partial least squares* (*PLS*).

3.3 Bundle Adjustment

A standard technique for minimizing Eq. (19) subject to Eq. (20) is to introduce a problem-dependent auxiliary variable to each $\boldsymbol{X}_\alpha$ and express $\bar{\boldsymbol{x}}_\alpha$ in terms of $\boldsymbol{X}_\alpha$ and $\boldsymbol{\theta}$ in the form

$$\bar{\boldsymbol{x}}_\alpha = \bar{\boldsymbol{x}}_\alpha(\boldsymbol{X}_\alpha, \boldsymbol{\theta}). \tag{22}$$

Then, we substitute this into Eq. (19) and minimize

$$J(\{\boldsymbol{X}_\alpha\}_{\alpha=1}^N, \boldsymbol{\theta}) = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha(\boldsymbol{X}_\alpha, \boldsymbol{\theta}), V_0[\boldsymbol{x}_\alpha]^{-1}(\boldsymbol{x}_\alpha - \bar{\boldsymbol{x}}_\alpha(\boldsymbol{X}_\alpha, \boldsymbol{\theta}))) \tag{23}$$

over the joint parameter space of $\{\boldsymbol{X}_\alpha\}_{\alpha=1}^N$ and $\boldsymbol{\theta}$.

A typical example of this approach is 3-D reconstruction from multiple images, for which $\boldsymbol{x}_\alpha$ has the form of $\boldsymbol{x}_\alpha = (x_\alpha, y_\alpha, x'_\alpha, y'_\alpha, ..., x''_\alpha, y''_\alpha)$, concatenating the projections (x_α, y_α), (x'_α, y'_α), ..., (x''_α, y''_α) of the αth point in the scene onto the images. The unknown parameter $\boldsymbol{\theta}$ specifies the state of all the cameras, consisting of the extrinsic parameters (the positions and orientations) and the intrinsic parameters (the focal lengths, the principal points, the aspect ratios, and the skew angles). If we introduce the 3-D position $\boldsymbol{X}_\alpha = (X_\alpha, Y_\alpha, Z_\alpha)$ of each point in the scene as the auxiliary variable, the true value $\bar{\boldsymbol{x}}_\alpha$ of $\boldsymbol{x}_\alpha$ can be explicitly expressed in the form $\bar{\boldsymbol{x}}_\alpha(\boldsymbol{X}_\alpha, \boldsymbol{\theta})$, which describes the image positions of the 3-D point $\boldsymbol{X}_\alpha$ that should be observed if the cameras have the parameter $\boldsymbol{\theta}$. Then, we minimize the *reprojection error*, i.e., the discrepancy of the observed projections $\boldsymbol{\xi}_\alpha$ from the predicted projections $\bar{\boldsymbol{x}}_\alpha(\boldsymbol{X}_\alpha, \boldsymbol{\theta})$. The minimum is searched over the entire parameter space of $\{\boldsymbol{X}_\alpha\}_{\alpha=1}^N$ and $\boldsymbol{\theta}$. This process is called *bundle adjustment* [23, 32], a term originated from photogrammetry, meaning we "adjust" the "bundle" of lines of sight so that they pass through the observed points in images. The package program is available on the Web [23]. The dimension of the parameter space is $3N +$ 'the dimension of $\boldsymbol{\theta}$', which becomes very large when many points are observed.

This bundle adjustment approach is not limited to 3-D reconstruction from multiple images. In *Examples 1* (line fitting) and *2* (ellipse fitting) in Sec. 2.1, for example, if we introduce the arc length s_α of the true position $(\bar{x}_\alpha, \bar{y}_\alpha)$ along the line or the ellipse from a fixed point as the auxiliary variable, we can express (x_α, y_α) in terms of s_α and $\boldsymbol{\theta}$. Then, we minimize the resulting Mahalanobis distance J over the entire parameter space of s_1, ..., s_N and $\boldsymbol{\theta}$. Instead of the arc length s_α, we can alternatively use the argument ϕ_α measured from the x-axis [30]. A similar approach can be done for fundamental matrix computation [2].

The standard numerical technique for the search of the parameter space is the *Levenberg-Marquardt* (*LM*) *method* [27], which is a hybrid of the Gauss-Newton iterations and the gradient descent. However, depending on the initial value of the iterations, the search may fall into a local minimum. Various global optimization techniques have also been studied [7]. A typical method is *branch and bound*, which introduces a function that gives a lower bound of J over a given region and divides the parameter space into small cells; those cells which have

lower bounds that are above the tested values are removed, and other cells are recursively subdivided [7, 9]. However, the evaluation of the lower bound involves a complicated technique, and searching the entire space requires a significant amount of computational time.

3.4 Gaussian Approximation of Noise in the $\boldsymbol{\xi}$-Space

The search in a high-dimensional parameter space of the bundle adjustment approach can be avoided if we introduce Gaussian approximation to the noise distribution in the $\boldsymbol{\xi}$-space. If the noise in the observation $\boldsymbol{x}_\alpha$ is Gaussian, the noise in its nonlinear transformation $\boldsymbol{\xi}_\alpha = \boldsymbol{\xi}(\boldsymbol{x}_\alpha)$ is not strictly Gaussian, although it is expected to have a Gaussian-like distribution if the noise is small. If it is approximated to be Gaussian, the optimization computation becomes much simpler. Suppose $\boldsymbol{\xi}_\alpha$ has noise of mean $\boldsymbol{0}$ with the covariance matrix $V[\boldsymbol{\xi}_\alpha] = \sigma^2 V_0[\boldsymbol{\xi}_\alpha]$ evaluated by Eq. (14). Then, the ML computation reduces to minimizing the Mahalanobis distance

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}(\boldsymbol{\xi}_\alpha - \bar{\boldsymbol{\xi}}_\alpha, V_0[\boldsymbol{\xi}_\alpha]^{-1}(\boldsymbol{\xi}_\alpha - \bar{\boldsymbol{\xi}}_\alpha)) \tag{24}$$

in the $\boldsymbol{\xi}$-space subject to the *linear* constraint

$$(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{\theta}) = 0. \tag{25}$$

Geometrically, this is interpreted to be fitting to N points $\boldsymbol{\xi}_\alpha$ in the $\boldsymbol{\xi}$-space the parameterized "hyperplane" $(\boldsymbol{\xi}, \boldsymbol{\theta}) = 0$ by adjusting $\boldsymbol{\theta}$, where the discrepancy of the points from the plane is measured by the Mahalanobis distance of Eq. (24) inversely weighted by the covariances of the data in the $\boldsymbol{\xi}$-space. Since Eq. (25) is now "linear" in $\bar{\boldsymbol{\xi}}_\alpha$, this constraint can be eliminated using Lagrange multipliers, reducing the problem to unconstrained minimization of

$$J = \frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}. \tag{26}$$

Today, Eq. (26) is called the *Sampson error* [8] after the ellipse fitting scheme introduced by P. D. Sampson [29].

3.5 Sampson Error Minimization

Various numerical techniques have been proposed for minimizing the Sampson error in Eq. (26). The best known is the *FNS* (*Fundamental Numerical Scheme*) of Chojnacki et al. [5], which goes as follows:

1. Let $W_\alpha = 1$, $\alpha = 1$, ..., N, and $\boldsymbol{\theta}_0 = \boldsymbol{0}$.
2. Computer the matrices

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top, \qquad \boldsymbol{L} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha^2 (\boldsymbol{\theta}_0, \boldsymbol{\xi}_\alpha)^2 V_0[\boldsymbol{\xi}_\alpha]. \tag{27}$$

3. Solve the eigenvalue problem $(\boldsymbol{M} - \boldsymbol{L})\boldsymbol{\theta} = \lambda\boldsymbol{\theta}$, and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest[2] eigenvalue λ.
4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{28}$$

and go back to Step 2.

The background of FNS is as follows. At the time of convergence, the matrices $\boldsymbol{M}$ and $\boldsymbol{L}$ have the form

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} \frac{\boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{L} = \frac{1}{N}\sum_{\alpha=1}^{N} \frac{(\boldsymbol{\theta}, \boldsymbol{\xi}_\alpha)^2 V_0[\boldsymbol{\xi}_\alpha]}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})^2}. \tag{29}$$

It is easily seen that the derivative of the Sampson error J in Eq. (26) is written in terms of these matrices in the form

$$\nabla_{\boldsymbol{\theta}} J = 2(\boldsymbol{M} - \boldsymbol{L})\boldsymbol{\theta}. \tag{30}$$

It can be shown that if the above iterations converge, the eigenvalue λ must be 0. Hence, the returned value $\boldsymbol{\theta}$ is the solution of $\nabla_{\boldsymbol{\theta}} J = \boldsymbol{0}$.

Other methods exist for minimizing Eq. (26) including the *HEIV* (*Heteroscedastic Errors-in-Variables*) of Leedan and Meer [22] and Matei and Meer [24], and the *projective Gauss-Newton iterations* of Kanatani and Sugaya [18]; all compute the same solution. Note that the "initial solution" obtained in the beginning by letting $W_\alpha = 1$ coincides with the LS solution described in Sec. 3.1.

3.6 Computation of the Exact ML Solution

Since the Sampson error of Eq. (26) is obtained by approximating the non-Gaussian noise distribution in the $\boldsymbol{\xi}$-space by a Gaussian distribution, the solution does not necessarily coincide with the ML solution that minimizes the Mahalanobis distance in Eq. (19). However, once we have obtained the solution $\boldsymbol{\theta}$ that minimizes Eq. (26), we can iteratively modify Eq. (26) by using that $\boldsymbol{\theta}$ so that Eq. (26) coincides with Eq. (19) in the end. This means that we obtain the exact ML solution. The procedure goes as follows [21]:

1. Let $J_0^* = \infty$ (a sufficiently large number), $\hat{\boldsymbol{x}}_\alpha = \boldsymbol{x}_\alpha$, and $\tilde{\boldsymbol{x}}_\alpha = \boldsymbol{0}$, $\alpha = 1, ..., N$.
2. Evaluate the normalized covariance matrices $V_0[\hat{\boldsymbol{\xi}}_\alpha]$ by replacing $\boldsymbol{x}_\alpha$ by $\hat{\boldsymbol{x}}_\alpha$ in their definition.
3. Compute the following $\boldsymbol{\xi}_\alpha^*$:

$$\boldsymbol{\xi}_\alpha^* = \boldsymbol{\xi}_\alpha + \left.\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_\alpha} \tilde{\boldsymbol{x}}_\alpha. \tag{31}$$

[2] We can alternatively compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue λ in absolute value, but it has been experimentally confirmed that convergence is faster for computing the smallest eigenvalue [18].

4. Compute the value $\boldsymbol{\theta}$ that minimizes the *modified Sampson error*

$$J^* = \frac{1}{N}\sum_{\alpha=1}^{N}\frac{(\boldsymbol{\xi}_\alpha^*, \boldsymbol{\theta})^2}{(\boldsymbol{\theta}, V_0[\hat{\boldsymbol{\xi}}_\alpha]\boldsymbol{\theta})}. \tag{32}$$

5. Update $\tilde{\boldsymbol{x}}_\alpha$ and $\hat{\boldsymbol{x}}_\alpha$ as follows:

$$\tilde{\boldsymbol{x}}_\alpha \leftarrow \frac{(\boldsymbol{\xi}_\alpha^*, \boldsymbol{\theta})V_0[\boldsymbol{x}_\alpha]}{(\boldsymbol{\theta}, V_0[\hat{\boldsymbol{\xi}}_\alpha]\boldsymbol{\theta})}\left.\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_\alpha}^{\top}\boldsymbol{\theta}, \qquad \hat{\boldsymbol{x}}_\alpha \leftarrow \boldsymbol{x}_\alpha - \tilde{\boldsymbol{x}}_\alpha. \tag{33}$$

6. Evaluate J^* by

$$J^* = \frac{1}{N}\sum_{\alpha}^{N}(\tilde{\boldsymbol{x}}_\alpha, V_0[\boldsymbol{x}_\alpha]\tilde{\boldsymbol{x}}_\alpha). \tag{34}$$

   If $J^* \approx J_0$, return $\boldsymbol{\theta}$ and stop. Else, let $J_0 \leftarrow J^*$ and go back to Step 2.

Since the modified Sampson error in Eq. (32) has the same form as the Sampson error in Eq. (26), we can minimize it by FNS (or other methods). According to numerical experiments, this modification converges after four or five rounds, yet in many practical problems the first four or five effective figures remain unchanged [19, 20]. In this sense, we can practically identify the Sampson error minimization with the ML computation.

3.7 Hyperaccurate Correction of ML

It has been widely recognized that the Sampson error minimization solution, which can be practically identified with the ML solution, has very high accuracy. However, it can be shown by detailed error analysis that the solution has statistical bias of $O(\sigma^2)$ and that the magnitude of the bias can be theoretically evaluated [14]. This implies that the accuracy can be further improved by subtracting the theoretically expected bias. This process is called *hyperaccurate correction* and goes as follows [13, 14]:

1. Estimate the square noise level σ^2 from the computed solution $\boldsymbol{\theta}$ and the corresponding matrix $\boldsymbol{M}$ in Eq. (29) by

$$\hat{\sigma}^2 = \frac{(\boldsymbol{\theta}, \boldsymbol{M}\boldsymbol{\theta})}{1-(n-1)/N}, \tag{35}$$

   where n is the dimension of the vector $\boldsymbol{\theta}$.
2. Compute the correction term[3]

$$\Delta_c\boldsymbol{\theta} = -\frac{\sigma^2}{N}\boldsymbol{M}_{n-1}^{-}\sum_{\alpha=1}^{N}W_\alpha(\boldsymbol{e}_\alpha, \boldsymbol{\theta})\boldsymbol{\xi}_\alpha + \frac{\hat{\sigma}^2}{N^2}\boldsymbol{M}_{n-1}^{-}\sum_{\alpha=1}^{N}W_\alpha^2(\boldsymbol{\xi}_\alpha, \boldsymbol{M}_{n-1}^{-}V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})\boldsymbol{\xi}_\alpha, \tag{36}$$

   where $\boldsymbol{e}_\alpha$ is a vector that depends on individual problems, and $\boldsymbol{M}_{n-1}^{-}$ is the pseudoinverse of $\boldsymbol{M}$ with truncated rank $n-1$ (the smallest eigenvalue is replaced by 0 in its spectral decomposition).

[3] The first term of Eq. (36) is omitted in [13, 14].

3. Correct the ML solution $\boldsymbol{\theta}$ in the form

$$\boldsymbol{\theta} \leftarrow \mathcal{N}[\boldsymbol{\theta} - \Delta_c\boldsymbol{\theta}], \tag{37}$$

where $\mathcal{N}[\,\cdot\,]$ is the normalization operator into unit norm ($\mathcal{N}[\boldsymbol{a}] \equiv \boldsymbol{a}/\|\boldsymbol{a}\|$).

The vector $\boldsymbol{e}_\alpha$ is $\mathbf{0}$ for many problems including line fitting (*Example 1* in Sec. 2.1) and fundamental matrix computation (*Example 3* in Sec. 2.1). It is generally $\mathbf{0}$ if multiple images are involved. A typical problem of nonzero $\boldsymbol{e}_\alpha$ is ellipse fitting (*Example 2* in Sec. 2.1), for which $\boldsymbol{e}_\alpha = (1, 0, 1, 0, 0, 0)^\top$. However, the effect is negligibly small, and the solution is practically the same if $\boldsymbol{e}_\alpha$ is replaced by $\mathbf{0}$.

The above bias correction concerns geometric estimation based on the geometric model of Eq. (2). In statistics, on the other hand, it is known that ML entails statistical bias in the presence of what is known as "nuisance parameters", and various studies exist for analyzing and removing bias in the ML solution. Okatani and Deguchi [25, 26] applied them to vision problems by introducing auxiliary variables in the form of Eq. (22). They analyzed the relationship between the bias and the hypersurface defined by the constraint [25] and introduced the method of projected scores [26].

For those computer vision researchers who regarded reprojection error minimization as the ultimate method, or the *Gold Standard* [8], the fact that the accuracy of ML can be improved by the above hyperaccurate correction was rather surprising. For hyperaccurate correction, however, one first needs to obtain the ML solution by an iterative method such as FNS and also estimate the noise level σ. Then, a question arises. Is it not possible to directly compute the corrected solution from the beginning, say, by modifying the FNS iterations? We now show that this is possible if we adopt the non-minimization approach of geometric estimation.

4 Non-minimization Approach

4.1 Iterative Reweight

The oldest method that is not based on minimization is the following *iterative reweight*:

1. Let $W_\alpha = 1$, $\alpha = 1$, ..., N, and $\boldsymbol{\theta}_0 = \mathbf{0}$.
2. Computer the following matrix $\boldsymbol{M}$:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top. \tag{38}$$

3. Solve the eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda\boldsymbol{\theta}$, and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue λ.

4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{39}$$

   and go back to Step 2.

The motivation of this method is the *weighted least squares* that minimizes

$$\frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2 = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\theta}^\top \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top \boldsymbol{\theta} = (\boldsymbol{\theta}, \underbrace{\frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top}_{\equiv \boldsymbol{M}} \boldsymbol{\theta}) = (\boldsymbol{\theta}, \boldsymbol{M}\boldsymbol{\theta}). \tag{40}$$

This is minimized by the unit eigenvector $\boldsymbol{\theta}$ of the matrix $\boldsymbol{M}$ for the smallest eigenvalue. As is well known in statistics, the optimal choice of the weight W_α is the inverse of the variance of that term. Since $(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{\theta}) = 0$, we have $(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta}) = (\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{\theta}) + \cdots$, and hence the leading term of the variance is

$$E[(\Delta\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2] = E[\boldsymbol{\theta}^\top \Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top \boldsymbol{\theta}] = (\boldsymbol{\theta}, E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top]\boldsymbol{\theta}) = \sigma^2(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}). \tag{41}$$

Hence, we should choose

$$W_\alpha = \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \tag{42}$$

but $\boldsymbol{\theta}$ is unknown. So, we do iterations, determining the weight W_α from the value of $\boldsymbol{\theta}$ in the preceding step. The "initial solution" computed with $W_\alpha = 1$ coincides with the LS solution, minimizing Eq. (17) in Sec. 3.1.

If Eq. (42) is substituted, Eq. (40) coincides with the Sampson error in Eq. (26). With the iterative update in Eq. (39), it appears that Eq. (26) is minimized. However, we are computing at each step the value of $\boldsymbol{\theta}$ that minimizes the numerator part for the fixed value of the denominator terms determined in the preceding step. Hence, at the time of the convergence, the resulting solution $\boldsymbol{\theta}$ is such that

$$\frac{1}{N}\sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})} \leq \frac{1}{N}\sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta}')^2}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})} \tag{43}$$

for any $\boldsymbol{\theta}'$, but this does not mean

$$\frac{1}{N}\sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta})^2}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})} \leq \frac{1}{N}\sum_{\alpha=1}^{N} \frac{(\boldsymbol{\xi}_\alpha, \boldsymbol{\theta}')^2}{(\boldsymbol{\theta}', V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta}')}. \tag{44}$$

The fact that iterative reweight does not minimize a particular cost function has not been well recognized by vision researchers.

The perturbation analysis in [14] shows that the covariance matrix $V[\boldsymbol{\theta}]$ of the resulting solution $\boldsymbol{\theta}$ agrees with the KCR lower bound (Sec. 2.4) up to $O(\sigma^4)$. Hence, it is practically impossible to reduce the variance any further. However, it has been widely known that the iterative reweight solution has a large bias [11]. Thus, the following strategies were introduced to improve iterative reweight:

– Remove the bias of the solution.
– Exactly minimize the Sampson error in Eq. (26).

The former is Kanatani's renormalization [10, 11], and the latter is the FNS of Chojnacki et al. [5] and the HEIV of Leedan and Meer [22] and Matei and Meer [24].

4.2 Renormalization

Kanatani's renormalization [10, 11] goes as follows[4]:

1. Let $W_\alpha = 1$, $\alpha = 1, ..., N$, and $\boldsymbol{\theta}_0 = \boldsymbol{0}$.
2. Compute the following matrices $\boldsymbol{M}$ and $\boldsymbol{N}$:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top, \qquad \boldsymbol{N} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{\xi}_\alpha]. \tag{45}$$

3. Solve the generalized eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda \boldsymbol{N}\boldsymbol{\theta}$, and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue λ in absolute value.
4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{46}$$

   and go back to Step 2.

The motivation of renormalization is as follows. Let $\bar{\boldsymbol{M}}$ be the true value of the matrix $\boldsymbol{M}$ in Eq. (45). Since $(\bar{\boldsymbol{\xi}}_\alpha, \boldsymbol{\theta}) = 0$, we have $\bar{\boldsymbol{M}}\boldsymbol{\theta} = \boldsymbol{0}$. Hence, $\boldsymbol{\theta}$ is the eigenvector of $\bar{\boldsymbol{M}}$ for eigenvalue 0, but $\bar{\boldsymbol{M}}$ is unknown. So, we estimate it. Since $E[\Delta\boldsymbol{\xi}_\alpha] = 0$ to a first approximation, the expectation of $\boldsymbol{M}$ is

$$\begin{aligned} E[\boldsymbol{M}] &= E[\frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha (\bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha)(\bar{\boldsymbol{\xi}}_\alpha + \Delta\boldsymbol{\xi}_\alpha)^\top] = \bar{\boldsymbol{M}} + \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha E[\Delta\boldsymbol{\xi}_\alpha \Delta\boldsymbol{\xi}_\alpha^\top] \\ &= \bar{\boldsymbol{M}} + \frac{\sigma^2}{N}\sum_{\alpha=1}^{N} W_\alpha V_0[\boldsymbol{\xi}_\alpha] = \bar{\boldsymbol{M}} + \sigma^2 \boldsymbol{N}. \end{aligned} \tag{47}$$

Thus, $\bar{\boldsymbol{M}} = E[\boldsymbol{M}] - \sigma^2\boldsymbol{N} \approx \boldsymbol{M} - \sigma^2\boldsymbol{N}$, so instead of $\bar{\boldsymbol{M}}\boldsymbol{\theta} = \boldsymbol{0}$ we solve $(\boldsymbol{M} - \sigma^2\boldsymbol{N})\boldsymbol{\theta} = \boldsymbol{0}$, or $\boldsymbol{M}\boldsymbol{\theta} = \sigma^2\boldsymbol{N}\boldsymbol{\theta}$. Assuming that σ^2 is small, we regard it as the smallest eigenvalue λ in absolute value. As in the case of iterative reweight, we iteratively update the weight W_α so that it approaches Eq. (42).

Kanatani's renormalization [10, 11] attracted much attention because it exhibited higher accuracy than any other then known methods. However, questions

[4] This is slightly different from the description in [10], in which the generalized eigenvalue problem is reduced to the standard eigenvalue problem, but the resulting solution is the same [11].

were repeatedly raised as to what it minimizes, perhaps out of the deep-rooted preconception that optimal estimation should minimize something. Chojnacki et al. [4] argued that renormalization can be "rationalized" if viewed as approximately minimizing the Sampson error. However, the renormalization process is not minimizing any particular cost function.

Note that the initial solution with $W_\alpha = 1$ solves $\Big(\sum_{\alpha=1}^N \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top\Big)\boldsymbol{\theta} = \lambda\Big(\sum_{\alpha=1}^N V_0[\boldsymbol{\xi}_\alpha]\Big)\boldsymbol{\theta}$, which is nothing but the method of Taubin [31], known to be very accurate algebraic method without requiring iterations. Thus, *renormalization is an iterative improvement of the Taubin solution*. According to many experiments, renormalization is shown to be more accurate than the Taubin method with nearly comparable accuracy with the FNS and the HEIV. The accuracy of renormalization is analytically evaluated in [14], showing that the covariance matrix $V[\boldsymbol{\theta}]$ of the solution $\boldsymbol{\theta}$ agrees with the KCR lower bound up to $O(\sigma^4)$ just as iterative reweight, but the bias is much smaller. That is the reason for the high accuracy of renormalization.

4.3 Analysis of Covariance and Bias

Since the covariance matrix $V[\boldsymbol{\theta}]$ of the renormalization solution $\boldsymbol{\theta}$ agrees with the KCR lower bound up to $O(\sigma^4)$, the covariance of the solution cannot be substantially improved any further. Very small it may be, however, the bias is not 0. Note that the renormalization procedure reduces to iterative reweight if the matrix $\boldsymbol{N}$ is replaced by the identity $\boldsymbol{I}$. This means that the reduction of the bias is attributed to the matrix $\boldsymbol{N}$. This observation implies the possibility of further reducing the bias by *optimizing* the matrix $\boldsymbol{N}$ in the form

$$\boldsymbol{N} = \frac{1}{N}\sum_{\alpha=1}^N W_\alpha V_0[\boldsymbol{\xi}_\alpha] + \cdots, \tag{48}$$

so that *the bias is zero up to high order error terms*. Using the perturbation analysis in [14], Al-Sharadqah and Chernov [1] actually did this for ellipse fitting, and Kanatani et al. [15] extended it to general geometric estimation. Their analysis goes as follows. We write the observation $\boldsymbol{x}_\alpha$ as the sum $\boldsymbol{x}_\alpha = \bar{\boldsymbol{x}}_\alpha + \Delta\boldsymbol{x}_\alpha$ of the true value $\bar{\boldsymbol{x}}_\alpha$ and the noise term $\Delta\boldsymbol{x}_\alpha$. Substituting this into $\boldsymbol{\xi}_\alpha = \boldsymbol{\xi}(\boldsymbol{x}_\alpha)$ and expand it in the form

$$\bar{\boldsymbol{\xi}}_\alpha + \Delta_1\boldsymbol{\xi}_\alpha + \Delta_2\boldsymbol{\xi}_\alpha + \cdots, \tag{49}$$

where and hereafter the bar denotes the noiseless value and Δ_k denotes terms of $O(\sigma^k)$. We similarly expand $\boldsymbol{M}$, $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{N}$ and express the generalized eigenvalue problem in the form

$$\begin{aligned}&(\bar{\boldsymbol{M}}+\Delta_1\boldsymbol{M}+\Delta_2\boldsymbol{M}+\cdots)(\bar{\boldsymbol{\theta}}+\Delta_1\boldsymbol{\theta}+\Delta_2\boldsymbol{\theta}+\cdots)\\ &= (\bar{\lambda}+\Delta_1\lambda+\Delta_2\lambda+\cdots)(\bar{\boldsymbol{N}}+\Delta_1\boldsymbol{N}+\Delta_2\boldsymbol{N}+\cdots)(\bar{\boldsymbol{\theta}}+\Delta_1\boldsymbol{\theta}+\Delta_2\boldsymbol{\theta}+\cdots).\end{aligned}$$

Equating the terms of the same order in σ, we obtain

$$\Delta_1\boldsymbol{\theta} = -\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}\bar{\boldsymbol{\theta}}, \tag{50}$$

$$\Delta_2^\perp\boldsymbol{\theta} = \bar{\boldsymbol{M}}^-\Big(\frac{(\bar{\boldsymbol{\theta}}, \boldsymbol{T}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} - \boldsymbol{T}\bar{\boldsymbol{\theta}}\Big), \tag{51}$$

where $\bar{\boldsymbol{M}}^-$ is the pseudoinverse of $\bar{\boldsymbol{M}}$; since $\bar{\boldsymbol{M}}$ has the eigenvector $\bar{\boldsymbol{\theta}}$ of eigenvalue 0, its rank is $n-1$ (n is the dimension of $\boldsymbol{\theta}$). The symbol $\Delta_2^\perp\boldsymbol{\theta}$ denotes the component of the second order noise term orthogonal to $\bar{\boldsymbol{\theta}}$; since $\boldsymbol{\theta}$ is a unit vector, it has no error in the direction of itself, so we are interested in the error orthogonal to it. The matrix $\boldsymbol{T}$ in Eq. (51) is defined to be

$$\boldsymbol{T} \equiv \Delta_2\boldsymbol{M} - \Delta_1\boldsymbol{M}\bar{\boldsymbol{M}}^-\Delta_1\boldsymbol{M}. \tag{52}$$

From Eq. (50), we can show that the leading term of the covariance matrix of $\boldsymbol{\theta}$ has the following form [14].

$$V[\boldsymbol{\theta}] \equiv E[\Delta_1\boldsymbol{\theta}\Delta_1\boldsymbol{\theta}^\top] = \frac{\sigma^2}{N}\bar{\boldsymbol{M}}^-. \tag{53}$$

From this we observe:

- The covariance matrix $V[\boldsymbol{\theta}]$ is $O(\sigma^2)$.
- The right side of Eq. (16) agrees with the KCR lower bound.
- Eq. (53) does not contain the matrix $\boldsymbol{N}$.

Thus, we cannot change the value of Eq. (53) by adjusting the matrix $\boldsymbol{N}$. However, the root-mean-square (RMS) error of $\boldsymbol{\theta}$ is the sum of the covariance term and the bias term, and the bias term is also $O(\sigma^2)$ (the expectation of odd order noise terms is 0, so the first order bias is $E[\Delta_1\boldsymbol{\theta}] = \boldsymbol{0}$). Since the second order bias term contains the matrix $\boldsymbol{N}$, we can reduce it by adjusting $\boldsymbol{N}$. From Eq. (51), the second order bias has the following expression:

$$E[\Delta_2^\perp\boldsymbol{\theta}] = \bar{\boldsymbol{M}}^-\Big(\frac{(\bar{\boldsymbol{\theta}}, E[\boldsymbol{T}\bar{\boldsymbol{\theta}}])}{(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} - E[\boldsymbol{T}\bar{\boldsymbol{\theta}}]\Big). \tag{54}$$

4.4 Hyper-renormalization

Equation (54) implies that if we can choose an $\boldsymbol{N}$ such that

$$E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] = c\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} \tag{55}$$

for some constant c, we will have

$$E[\Delta_2^\perp\boldsymbol{\theta}] = \bar{\boldsymbol{M}}^-\Big(\frac{(\bar{\boldsymbol{\theta}}, c\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}{(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}})}\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}} - c\bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}\Big) = \boldsymbol{0}, \tag{56}$$

i.e., *the second order bias is completely eliminated*. Kanatani et al. [15] showed that if the matrix $\bar{\boldsymbol{N}}$ is defined by

$$\bar{\boldsymbol{N}} = \frac{1}{N}\sum_{\alpha=1}^{N} \bar{W}_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[\bar{\boldsymbol{\xi}}_\alpha \boldsymbol{e}_\alpha^\top] \Big) - \frac{1}{N^2}\sum_{\alpha=1}^{N} \bar{W}_\alpha^2 \Big((\bar{\boldsymbol{\xi}}_\alpha, \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha) V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha] \bar{\boldsymbol{M}}^- \bar{\boldsymbol{\xi}}_\alpha \bar{\boldsymbol{\xi}}_\alpha^\top] \Big), \tag{57}$$

then $E[\boldsymbol{T}\bar{\boldsymbol{\theta}}] = \sigma^2 \bar{\boldsymbol{N}}\bar{\boldsymbol{\theta}}$ holds, where $\boldsymbol{e}_\alpha$ is a vector that depends on individual problems (the same vector as that in Eq. (36)), and $\mathcal{S}[\,\cdot\,]$ denotes symmetrization ($\mathcal{S}[\boldsymbol{A}] = (\boldsymbol{A} + \boldsymbol{A}^\top)/2$). In actual computation, the true values in Eq. (57) are replaced by computed values. This entails errors of $O(\sigma)$, but since the expectation of odd order noise terms is 0, Eq. (56) is $O(\sigma^4)$. Thus, we obtain the following procedure of *hyper-renormalization*:

1. Let $W_\alpha = 1$, $\alpha = 1, ..., N$, and $\boldsymbol{\theta}_0 = \boldsymbol{0}$.
2. Compute the following matrices $\boldsymbol{M}$ and $\boldsymbol{N}$:

$$\boldsymbol{M} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top, \tag{58}$$

$$\boldsymbol{N} = \frac{1}{N}\sum_{\alpha=1}^{N} W_\alpha \Big(V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[\boldsymbol{\xi}_\alpha \boldsymbol{e}_\alpha^\top] \Big) - \frac{1}{N^2}\sum_{\alpha=1}^{N} W_\alpha^2 \Big((\boldsymbol{\xi}_\alpha, \boldsymbol{M}_{n-1}^- \boldsymbol{\xi}_\alpha) V_0[\boldsymbol{\xi}_\alpha] + 2\mathcal{S}[V_0[\boldsymbol{\xi}_\alpha] \boldsymbol{M}_{n-1}^- \boldsymbol{\xi}_\alpha \boldsymbol{\xi}_\alpha^\top] \Big). \tag{59}$$

   Here, $\boldsymbol{M}_{n-1}^-$ is the pseudoinverse of $\boldsymbol{M}$ with truncated rank $n-1$ (cf. Eq. (36).
3. Solve the generalized eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda \boldsymbol{N}\boldsymbol{\theta}$, and compute the unit eigenvector $\boldsymbol{\theta}$ for the smallest eigenvalue λ in absolute value.
4. If $\boldsymbol{\theta} \approx \boldsymbol{\theta}_0$ up to sign, return $\boldsymbol{\theta}$ and stop. Else, let

$$W_\alpha \leftarrow \frac{1}{(\boldsymbol{\theta}, V_0[\boldsymbol{\xi}_\alpha]\boldsymbol{\theta})}, \qquad \boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}, \tag{60}$$

   and go back to Step 2.

It turns out that the initial solution with $W_\alpha = 1$ coincides with what is called *HyperLS* [16, 17, 28], which is derived to remove the bias up to second order error terms within the framework of algebraic methods without iterations[5]. Thus, *hyper-renormalization is an iterative improvement of HyperLS*.

Standard linear algebra routines for solving the generalized eigenvalue problem $\boldsymbol{M}\boldsymbol{\theta} = \lambda \boldsymbol{N}\boldsymbol{\theta}$ assume that $\boldsymbol{N}$ is positive definite, but the matrix $\boldsymbol{N}$ in Eq. (59)

[5] The expression of Eq. (59) with $W_\alpha = 1$ lacks one term as compared with the corresponding expression of HyperLS, but the same solution is produced.

has both positive and negative eigenvalues. For renormalization, the matrix $\boldsymbol{N}$ is positive semidefinite, having eigenvalue 0. This, however, causes no trouble, because the problem can be rewritten as

$$\boldsymbol{N}\boldsymbol{\theta} = \frac{1}{\lambda}\boldsymbol{M}\boldsymbol{\theta}. \tag{61}$$

The matrix $\boldsymbol{M}$ is positive definite for noisy data, so we can use a standard routine to compute the eigenvector $\boldsymbol{\theta}$ for the eigenvalue $1/\lambda$ with the largest absolute value. If the matrix $\boldsymbol{M}$ happens to have eigenvalue 0, it indicates that the data are all exact, so the unit eigenvector for the eigenvalue 0 is the exact solution.

5 Numerical Examples

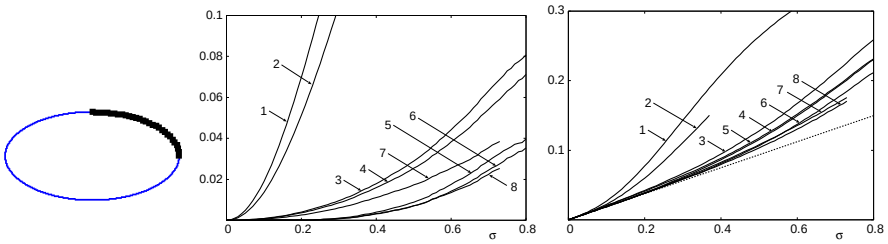
We define 30 equidistant points on the ellipse shown in Fig. 1(a). The major and minor axis are set to 100 and 50 pixels, respectively. We add random Gaussian noise of mean 0 and standard deviation σ to the x and y coordinates of each point independently and fit an ellipse to the noisy point sequence using : 1) LS, 2) iterative reweight, 3) the Taubin method, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, and 8) hyperaccurate correction of ML.

Since the computed $\boldsymbol{\theta}$ and its true value $\bar{\boldsymbol{\theta}}$ are both unit vectors, we measure the discrepancy between them by the orthogonal component $\Delta^{\perp}\boldsymbol{\theta} = \boldsymbol{P}_{\bar{\boldsymbol{\theta}}}\boldsymbol{\theta}$, where $\boldsymbol{P}_{\bar{\boldsymbol{\theta}}}$ ($\equiv \boldsymbol{I} - \bar{\boldsymbol{\theta}}\bar{\boldsymbol{\theta}}^{\top}$) is the projection matrix along $\bar{\boldsymbol{\theta}}$. We generated 10000 independent noise instances and evaluated the bias B (Fig. 1(b)) and the RMS (root-mean-square) error D (Fig. 1(c)) defined by

$$B = \Big\| \frac{1}{10000} \sum_{a=1}^{10000} \Delta^{\perp}\boldsymbol{\theta}^{(a)} \Big\|, \qquad D = \sqrt{\frac{1}{10000} \sum_{a=1}^{10000} \|\Delta^{\perp}\boldsymbol{\theta}^{(a)}\|^2}, \tag{62}$$

where $\boldsymbol{\theta}^{(a)}$ is the solution in the ath trial. The dotted line in Fig. 1(c) indicates the KCR lower bound. The interrupted plots in Fig. 2(a) for iterative reweight, ML, and hyperaccurate correction of ML indicate that the iterations did not converge beyond that noise level. Our convergence criterion is $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < 10^{-6}$ for the current value $\boldsymbol{\theta}$ and the value $\boldsymbol{\theta}_0$ in the preceding iteration; their signs are adjusted before subtraction. If this criterion is not satisfied after 100 iterations, we stopped. For each σ, we regarded the iterations as not convergent if any among the 10000 trials does not converge.

We can see from Fig. 2(a) that LS and iterative reweight have very large bias, in contrast to which the bias is very small for the Taubin method and renormalization. The bias of HyperLS and hyper-renormalization is still smaller and even smaller than ML. Since the leading covariance is common to iterative reweight, renormalization, and hyper-renormalization, the RMS error reflects the magnitude of the bias as shown in Fig. 2(b). Because the hyper-renormalization solution does not have bias up to high order error terms, it has nearly the same



(a) (b) (c)

Fig. 2. Thirty points on an ellipse (a). The bias (a) and the RMS error (b) of the fitted ellipse for the standard deviation σ of the added noise over 10000 independent trials. 1) LS, 2) iterative reweight, 3) the Taubin method, 4) renormalization, 5) HyperLS, 6) hyper-renormalization, 7) ML, 8) hyperaccurate correction of ML. The dotted line in (c) indicates the KCR lower bound.

accuracy as ML, or reprojection error minimization. A close examination of the small σ part reveals that hyper-renormalization outperforms ML. The highest accuracy is achieved, although the difference is very small, by hyperaccurate correction of ML. However, it first requires the ML solution, and the FNS iterations for its computation may not converge above a certain noise level, as shown in Figs. 2(a), (b). On the other hand, hyper-renormalization is very robust to noise. This is because the initial solution is HyperLS, which is itself highly accurate already as shown in Fig. 2. For this reason, we conclude that it is the best method for practical computations.

6 Concluding Remarks

We have overviewed techniques for optimal geometric estimation from noisy observations for computer vision applications. We first described minimization-based approaches: LS, ML, which includes reprojection error minimization (Gold Standard) as a special case, and Sampson error minimization. We then formulated non-minimization approaches: iterative reweight, renormalization, and hyper-renormalization, which can be viewed as iterative improvement of LS, the Taubin method, and HyperLS, respectively (Table 1). Showing numerical examples, we conclude that hyper-renormalization is robust to noise and currently is the best method.

Table 1. Summary of non-minimization approaches.

initial	weight update	final
LS	$\longrightarrow$	iterative reweight
Taubin	$\longrightarrow$	renormalization
HyperLS	$\longrightarrow$	hyper-renormalization

Acknowledgments: A major part of this paper resulted from the author's collaboration with Prasanna Rangarajan of Southern Methodist University, U.S.A., Ali Al-Sharadqah of the University of Mississippi, U.S.A., Nikolai Chernov of the University of Alabama at Birmingham, U.S.A., and Yasuyuki Sugaya of Toyohashi University of Technology, Japan. This work was supported in part by JSPS Grant-in-Aid for Challenging Exploratory Research (24650086).

References

1. Al-Sharadqah, A., Chernov, N.: A doubly optimal ellipse fit. Comp. Stat. Data Anal. 56(9), 2771–2781 (2012)
2. Bartoli, B., Sturm, P.: Nonlinear estimation of fundamental matrix with minimal parameters. IEEE Trans. Patt. Anal. Mach. Intell. 26(3), 426–432 (2004)
3. Chernov, N., Lesort, C.: Statistical efficiency of curve fitting algorithms. Comp. Stat. Data Anal. 47(4), 713–728 (2004)
4. Chojnacki, W., Brooks, M.J., A. van den Hengel, A.: Rationalising the renormalization method of Kanatani. J. Math. Imaging Vis. 21(11), 21–38 (2001)
5. Chojnacki, W., Brooks, M.J., van den Hengel, A., Gawley, D.: On the fitting of surfaces to data with covariances. IEEE Trans. Patt. Anal. Mach. Intell. 22(11), 1294–1303 (2000)
6. Godambe, V.P. (ed.): Estimating Functions. Oxford University Press, New York (1991)
7. Hartley, R., Kahl, F.: Optimal algorithms in multiview geometry. In: Proc. 8th Asian Conf. Comput. Vis., Tokyo, Japan, vol. 1, pp. 13–34 (November 2007)
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
9. Kahl, F., Agarwal, S., Chandraker, M.K., Kriegman, D., Belongie, S.: Practical global optimization for multiview geometry. Int. J. Comput. Vis. 79(3), 271–284 (2008)
10. Kanatani, K.: Renormalization for unbiased estimation. In: Proc. 4th Int. Conf. Comput. Vis., Berlin, Germany, pp. 599–606 (May 1993)
11. Kanatani, K.: Statistical Optimization for Geometric Computation: Theory and Practice. Elsevier, Amsterdam (1996); reprinted, Dover, New York (2005)
12. Kanatani, K.: Cramer-Rao lower bounds for curve fitting. Graphical Models Image Process. 60(2), 93–99 (1998)
13. Kanatani, K.: Ellipse fitting with hyperaccuracy. IEICE Trans. Inf. & Syst. E89-D(10), 2653–2660 (2006)
14. Kanatani, K.: Statistical optimization for geometric fitting: Theoretical accuracy analysis and high order error analysis. Int. J. Comput. Vis. 80(2), 167–188 (2008)
15. Kanatani, K.: Al-Sharadqah, A., Chernov, N., Sugaya, Y.: Renormalization Returns: Hyper-renormalization and Its Applications. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmidt. C. (eds.) ECCV 2012, Part III. LNCS, vol. 7574, pp. 348–397. Springer, Heidelberg (2012).
16. Kanatani K., Rangarajan, P.: Hyper least squares fitting of circles and ellipses. Comput. Stat. Data Anal. 55(6), 2197–2208 (2011)
17. Kanatani, K., Rangarajan, P., Sugaya, Y., Niitsuma, H.: HyperLS and its applications. IPSJ Trans. Comput. Vis. Appl. 3, 80–94 (2011)
18. Kanatani, K., Sugaya, Y.: Performance evaluation of iterative geometric fitting algorithms. Comp. Stat. Data Anal. 52(2), 1208–1222 (2007)

19. Kanatani, K., Sugaya, Y.: Compact algorithm for strictly ML ellipse fitting. In: Proc. 19th Int. Conf. Patt. Recog., Tampa, FL (December 2008)
20. Kanatani, K., Sugaya, Y.: Compact fundamental matrix computation. IPSJ Tran. Comput. Vis. Appl. 2, 59–70 (2010)
21. Kanatani, K., Sugaya, Y.: Unified computation of strict maximum likelihood for geometric fitting. J. Math. Imaging Vis. 38(1), 1–13 (2010)
22. Leedan, Y., Meer, P.: Heteroscedastic regression in computer vision: Problems with bilinear constraint. Int. J. Comput. Vis. 37(2), 127-150 (2000)
23. Lourakis, M.I.A., Argyros, A.A.: SBA: A software package for generic sparse bundle adjustment. ACM Trans. Math. Software 36(1), 2, 1–30 (2009)
24. Matei, J., Meer, P.: Estimation of nonlinear errors-in-variables models for computer vision applications. IEEE Trans. Patt. Anal. Mach. Intell. 28(10), 1537–1552 (2006)
25. Okatani, T., Deguchi, K.: On bias correction for geometric parameter estimation in computer vision. In: Proc. IEEE Conf. Comput. Vis. Patt. Recog., Miami Beach, FL, U.S.A., pp. 959-966 (June 2009)
26. Okatani, T., Deguchi, K.: Improving accuracy of geometric parameter estimation using projected score method. In: Proc. Int. Conf. Comput. Vis., Kyoto, Japan, pp. 1733-1740 (September/October 2009)
27. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C: The Art of Scientific Computing, 2nd edn. Cambridge University Press, Cambridge (1992)
28. Rangarajan, P., Kanatani, K.: Improved algebraic methods for circle fitting. Electronic J. Stat. 3, 1075–1082 (2009)
29. Sampson, P.D.: Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithm. Comput. Graphics Image Process. 18(1), 97–108 (1982)
30. Sturm, P., Gargallo, P.: Conic fitting using the geometric distance. In: Proc. 8th Asian Conf. Comput. Vis., Tokyo, Japan, Vol. 2, pp.784–795 (November 2007)
31. Taubin, G.: Estimation of planar curves, surfaces, and non-planar space curves defined by implicit equations with applications to edge and range image segmentation. IEEE Trans. Patt. Anal. Mach. Intell. 13(11), 1115–1138 (1991)
32. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.: Bundle Adjustment— A Modern Synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.), ICCV-WS 1999. LNCS, vol. 1883, pp. 298–375. Vision Algorithms: Theory and Practice, Springer, Heidelberg (2000)