A Consistency Result for the Normalized Eight-Point Algorithm

Wojciech Chojnacki, Michael J. Brooks

School of Computer Science, University of Adelaide, Adelaide, SA 5005, Australia

{wojciech.chojnacki,michael.brooks}@adelaide.edu.au

Abstract

A recently proposed argument to explain the improved performance of the eight-point algorithm that results from using normalized data [IEEE Trans. Pattern Anal. Mach. Intell., 25(9):1172–1177, 2003] relies upon adoption of a certain model for statistical data distribution. Under this model, the cost function that underlies the algorithm operating on the normalized data is statistically more advantageous than the cost function that underpins the algorithm using unnormalized data. Here we extend this explanation by introducing a more refined, structured model for data distribution. Under the extended model, the normalized eight-point algorithm turns out to be approximately consistent in a statistical sense. The proposed extension provides a link between the existing statistical rationalization of the normalized eight-point algorithm and the approach of Mühlich and Mester for enhancing total least squares estimation methods via equilibration. Our contribution forms part of a wider effort to rationalize and interrelate foundational methods in vision parameter estimation.

1. Introduction

In a version adapted for computation of the *fundamental matrix*, the *eight-point algorithm* of Longuet-Higgins [12] produces significantly more accurate estimates if normalized rather than unnormalized image data are used as input for the scheme. Hartley [7], who made this fundamental discovery, attributed the improvement in performance resulting from data normalization to the better numerical conditioning of a pivotal matrix used in solving an eigenvalue problem. Recently, an alternative explanation was proposed based on a certain model for statistical data distribution [3]. Under this model, the summands of the cost function underlying the normalized version of the eight-point algorithm are more balanced in terms of spread compared with the summands of the cost function underlying the unnormalized version. The improved statistical design of the associated cost function is seen as being conducive to the normalized algorithm's gains in accuracy. Here we reinforce this line of argumentation by introducing a more refined, structured model for the data distribution. We show that under the extended model the normalized eight-point algorithm is approximately consistent in a statistical sense. This result is then directly linked to the work of Mühlich and Mester [15], where a similar argument is used to validate an algorithm akin to Hartley's. The study reported below forms part of a wider effort to rationalize and interrelate foundational methods in vision parameter estimation (e.g., see [1, 2, 4, 9, 11]).

2. Algebraic Least Squares

A 3D point in a scene perspectively projected onto the image plane of a camera gives rise to an *image point* represented by a pair (m_x, m_y) of coordinates, or equivalently, by the "homogeneous" vector $\boldsymbol{m} = [m_x, m_y, 1]^T$. A 3D point projected onto the image planes of two cameras endowed with separate coordinate systems gives rise to a pair of *corresponding points*. When represented by a pair of vectors $(\boldsymbol{m}, \boldsymbol{m}')$, a correspondence satisfies the *epipolar constraint*

$$\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m} = 0, \tag{1}$$

where $\boldsymbol{F} = [f_{ij}]$ is a 3×3 fundamental matrix that incorporates information about the relative orientation and internal geometry of the cameras [5, 8]. In addition to (1), $\boldsymbol{F}$ is subject to the *singularity constraint* (or, equivalently, the *rank-2 constraint*)

$$\det \boldsymbol{F} = 0. \tag{2}$$

Given a set $\{(\boldsymbol{m}_n, \boldsymbol{m}'_n)\}_{n=1}^N$ of *measured* correspondences, a simple "linear" estimate of $\boldsymbol{F}$ providing an approximate solution of the system

$$\boldsymbol{m}'^T_n \boldsymbol{F} \boldsymbol{m}_n = 0 \quad (n = 1, \ldots, N)$$

is obtained by minimizing the cost function

$$J_{\mathrm{ALS}}(\boldsymbol{F}) = \|\boldsymbol{F}\|_F^{-2} \sum_{n=1}^N (\boldsymbol{m}'^T_n \boldsymbol{F} \boldsymbol{m}_n)^2$$

IEEE
COMPUTER
SOCIETY

with $\|\boldsymbol{F}\|_F = (\sum_{i,j} f_{ij}^2)^{1/2}$ the Frobenius norm of $\boldsymbol{F}$. Here, $\boldsymbol{m}_n'^T \boldsymbol{F} \boldsymbol{m}_n$ is the signed *algebraic distance* between the individual correspondence $(\boldsymbol{m}_n, \boldsymbol{m}_n')$ and the candidate matrix $\boldsymbol{F}$. The minimizer of J_{ALS} constitutes the *algebraic least squares* (ALS) *estimate* and is denoted $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$. $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$ is determined only up to scale, as the value of $J_{\mathrm{ALS}}(\boldsymbol{F})$ does not change when $\boldsymbol{F}$ is multiplied by a nonzero scalar. As a rule, $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$ does not satisfy (2). However, this estimate, like many other estimates not obeying (2), can further be upgraded to a genuine rank-2 fundamental matrix by applying one of a variety of available methods [8, 9]. In this paper, only the ALS method and its variant for normalized data will be of concern and the singularity constraint will be set aside.

For the purpose of computing $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$, it is convenient to use the joint image point $\boldsymbol{x} = [m_x, m_y, m_x', m_y']^T$ as a compact descriptor of the single correspondence $(\boldsymbol{m}, \boldsymbol{m}')$. For an $m \times n$ matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ with $\boldsymbol{a}_i$ the ith column vector of length m, denote by $\mathrm{vec}(\boldsymbol{A})$ the *vectorization* of $\boldsymbol{A}$, i.e., the column vector of length mn defined by $\mathrm{vec}(\boldsymbol{A}) = [\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_n^T]^T$. Let $\boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{F}^T)$ and $\boldsymbol{u}(\boldsymbol{x}) = \mathrm{vec}(\boldsymbol{m}\boldsymbol{m}'^T)$. Then, as one easily verifies,

$$\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m} = \boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}).$$

With this formula, J_{ALS} can be written as

$$J_{\mathrm{ALS}}(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|^{-2} \boldsymbol{\theta}^T \boldsymbol{A} \boldsymbol{\theta},$$

where

$$\boldsymbol{A} = \sum_{n=1}^{N} \boldsymbol{u}(\boldsymbol{x}_n) \boldsymbol{u}(\boldsymbol{x}_n)^T$$

and $\|\boldsymbol{\theta}\| = (\theta_1^2 + \cdots + \theta_9^2)^{1/2}$. Now, an optimality condition for the minimizer of J_{ALS} (expressed as the zeroing of the gradient of the cost function) can be used to characterize $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ as the normalized eigenvector of $\boldsymbol{A}$ associated with the smallest eigenvalue (the smallest eigenvector). The eigenvector is defined uniquely up to a sign [2]. In practice $\widehat{\boldsymbol{\theta}}_{\mathrm{ALS}}$ is found by performing *singular value decomposition* (SVD) of the matrix

$$\boldsymbol{M} = [\boldsymbol{u}(\boldsymbol{x}_1), \ldots, \boldsymbol{u}(\boldsymbol{x}_n)]^T$$

and taking for the desired output the right singular vector of $\boldsymbol{M}$ associated with the smallest singular value (the minimum right singular vector). In this form, the ALS method is essentially identical to the eight-point algorithm of Longuet-Higgins [12].

3. Normalized Algebraic Least Squares

ALS estimates are highly sensitive to image noise. To curb noise-driven instabilities, Hartley proposed a simple normalization of the input data to be performed prior to running the ALS method [7]. The normalization process involves two affine transformations

$$\tilde{\boldsymbol{m}}_n = \boldsymbol{T} \boldsymbol{m}_n, \quad \tilde{\boldsymbol{m}}_n' = \boldsymbol{T}' \boldsymbol{m}_n',$$

where $\boldsymbol{T}$ and $\boldsymbol{T}'$ are 3×3 data-dependent matrices whose exact form will be presented shortly. The ALS method is applied to the normalized data $\{(\tilde{\boldsymbol{m}}_n, \tilde{\boldsymbol{m}}_n')\}_{n=1}^{N}$ and then the result is suitably back-transformed. More specifically, if, for each $n = 1, \ldots, N$, $\tilde{\boldsymbol{x}}_n = [\tilde{m}_{x,n}, \tilde{m}_{y,n}, \tilde{m}_{x,n}', \tilde{m}_{y,n}']^T$ is the joint image point corresponding to $(\tilde{\boldsymbol{m}}_n, \tilde{\boldsymbol{m}}_n')$ and $\tilde{\boldsymbol{F}}_{\mathrm{ALS}}$ is the ALS estimate based on $\{\tilde{\boldsymbol{x}}_n\}_{n=1}^{N}$, then the *normalized algebraic least squares* (NALS) estimate of $\boldsymbol{F}$, $\widehat{\boldsymbol{F}}_{\mathrm{NALS}}$, is defined by

$$\widehat{\boldsymbol{F}}_{\mathrm{NALS}} = \boldsymbol{T}'^T \widehat{\tilde{\boldsymbol{F}}}_{\mathrm{ALS}} \boldsymbol{T}.$$

The matrices $\boldsymbol{T}$, $\boldsymbol{T}'$ come in two variants depending on whether the coordinates $m_{x,n}, m_{y,n}$ and $m_{x,n}', m_{y,n}'$ are normalized separately or collectively. Let

$$\overline{\boldsymbol{m}} = [\overline{m}_x, \overline{m}_y, 1] = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{m}_n,$$

$$\overline{\boldsymbol{m}}' = [\overline{m}_x', \overline{m}_y', 1] = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{m}_n'$$

be the centroids of the data points in each image. For the data normalized separately, *anisotropic* scaling factors are introduced by means of the formulae

$$s_x = \left(\frac{1}{N} \sum_{n=1}^{N} (m_{x,n} - \overline{m}_x)^2 \right)^{1/2},$$

$$s_y = \left(\frac{1}{N} \sum_{n=1}^{N} (m_{y,n} - \overline{m}_y)^2 \right)^{1/2},$$

$$s_x' = \left(\frac{1}{N} \sum_{n=1}^{N} (m_{x,n}' - \overline{m}_x')^2 \right)^{1/2},$$

$$s_y' = \left(\frac{1}{N} \sum_{n=1}^{N} (m_{y,n}' - \overline{m}_y')^2 \right)^{1/2},$$

and the matrices $\boldsymbol{T}$ and $\boldsymbol{T}'$ are defined by

$$\boldsymbol{T} = \begin{bmatrix} s_x^{-1} & 0 & -s_x^{-1}\overline{m}_x \\ 0 & s_y^{-1} & -s_y^{-1}\overline{m}_y \\ 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{T}' = \begin{bmatrix} s_x'^{-1} & 0 & -s_x'^{-1}\overline{m}_x' \\ 0 & s_y'^{-1} & -s_y'^{-1}\overline{m}_y' \\ 0 & 0 & 1 \end{bmatrix}.$$

For the data normalized individually, *isotropic* scaling factors take the form

$$s = \left(\frac{1}{2N} \sum_{n=1}^{N} (m_{x,n} - \overline{m}_x)^2 + (m_{y,n} - \overline{m}_y)^2 \right)^{1/2},$$

$$s' = \left(\frac{1}{2N} \sum_{n=1}^{N} (m'_{x,n} - \overline{m}'_x)^2 + (m'_{y,n} - \overline{m}'_y)^2 \right)^{1/2},$$

and $\boldsymbol{T}$ and $\boldsymbol{T}'$ are given by

$$\boldsymbol{T} = \begin{bmatrix} s^{-1} & 0 & -s^{-1}\overline{m}_x \\ 0 & s^{-1} & -s^{-1}\overline{m}_y \\ 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{T}' = \begin{bmatrix} s'^{-1} & 0 & -s'^{-1}\overline{m}'_x \\ 0 & s'^{-1} & -s'^{-1}\overline{m}'_y \\ 0 & 0 & 1 \end{bmatrix}.$$

The numerical motivation for the NALS method comes from the fact that the matrix $\tilde{\boldsymbol{A}} = \sum_{n=1}^{N} \boldsymbol{u}(\tilde{\boldsymbol{x}}_n)\boldsymbol{u}(\tilde{\boldsymbol{x}}_n)^T$ serving to calculate $\widehat{\boldsymbol{F}}_{\text{ALS}}$ is better conditioned—in a certain sense—than the matrix $\boldsymbol{A}$. The improved conditioning leads to the smallest eigenvector of $\tilde{\boldsymbol{A}}$ being less sensitive to small perturbations of the matrix entries than the smallest eigenvector of $\boldsymbol{A}$ [7]. The NALS method can also be justified statistically [3]. The statistical argument exploits the fact that $\widehat{\boldsymbol{F}}_{\text{NALS}}$ is the minimizer of the *normalized algebraic least squares* (NALS) cost function

$$J_{\text{NALS}}(\boldsymbol{F}) = \|\boldsymbol{T}'^{-T}\boldsymbol{F}\boldsymbol{T}^{-1}\|_F^{-2} \sum_{n=1}^{N} (\boldsymbol{m}'^T_n \boldsymbol{F}\boldsymbol{m}_n)^2.$$

A fundamental observation is that different *residuals*

$$r_n = \boldsymbol{m}'^T_n \boldsymbol{F}\boldsymbol{m}_n \quad (n = 1, \ldots, N)$$

carry, as a rule, different statistical weight. When the $\boldsymbol{m}_n$, $\boldsymbol{m}'_n$ are treated as sample values of independent multivariate random variables, the r_n are realizations of (typically) *heteroscedastic* random variables—that is, ones having different variances. The larger the variance of a particular r_n, the less reliable this residual is likely to be, and the more it should be devalued. This suggests that, to account for heteroscedasticity, the simple cost function $\sum_{n=1}^{N} r_n^2$, which effectively is identical to J_{ALS}, should be replaced by the more complicated cost function $\sum_{n=1}^{N} r_n^2/\text{var}\,[r_n]$, where $\text{var}\,[r]$ denotes the variance of r. As it turns out, one can conceive a model for data distribution such that all the $\text{var}\,[r_n]$ are equal to $\|\boldsymbol{T}'^{-T}\boldsymbol{F}\boldsymbol{T}^{-1}\|_F^2$ and consequently $\sum_{n=1}^{N} r_n^2/\text{var}\,[r_n]$ reduces to J_{NALS}. Construed in this fashion, J_{NALS} has statistical advantage over J_{ALS}, leading to an improvement in performance of the normalized version of the eight-point algorithm over the unnormalized version.

The data distribution model that makes the above argument work is as follows. For each $n = 1, \ldots, N$, one introduces random variables

$$\mathbf{m}_n = \overline{\boldsymbol{m}} + \Delta\mathbf{m}_n, \quad \mathbf{m}'_n = \overline{\boldsymbol{m}}' + \Delta\mathbf{m}'_n,$$

where $\overline{\boldsymbol{m}}$ and $\overline{\boldsymbol{m}}'$ are fixed, nonrandom points, and $\Delta\mathbf{m}_n = [\Delta\text{m}_{x,n}, \Delta\text{m}_{y,n}, 0]^T$ and $\Delta\mathbf{m}'_n = [\Delta\text{m}'_{x,n}, \Delta\text{m}'_{y,n}, 0]^T$ are random isotropic perturbations satisfying the following conditions:

- the collection $\{\Delta\mathbf{m}_n, \Delta\mathbf{m}'_n\}_{n=1}^{N}$ is jointly independent;

- $\mathsf{E}[\Delta\mathbf{m}_n] = \mathsf{E}[\Delta\mathbf{m}'_n] = \boldsymbol{0}$ for each $n = 1, \ldots, N$;

- for some $s, s' > 0$ and each $n = 1, \ldots, N$,

$$\mathsf{E}[\Delta\mathbf{m}_n\Delta\mathbf{m}_n^T] = \text{diag}(s^2, s^2, 0),$$
$$\mathsf{E}[\Delta\mathbf{m}'_n\Delta\mathbf{m}'^T_n] = \text{diag}(s'^2, s'^2, 0).$$

Here, E denotes the expectation value. The observed locations $\boldsymbol{m}_n$ and $\boldsymbol{m}'_n$ are viewed as realizations of $\mathbf{m}_n$ and $\mathbf{m}'_n$, respectively.

4. Extended Statistical Model

A different approach to achieving improved ALS estimates was proposed by Mühlich and Mester [15, 16]. The idea is to search for 3×3 affine *weight matrices* $\boldsymbol{W}$ and $\boldsymbol{W}'$ with the property that if $\widehat{\boldsymbol{F}}_{\text{ALS}}$ is the ALS estimate based on the data obtained by applying the transformations $\tilde{\boldsymbol{m}} = \boldsymbol{W}\boldsymbol{m}$ and $\tilde{\boldsymbol{m}}' = \boldsymbol{W}'\boldsymbol{m}'$ on the raw data, then the *equilibrated ALS estimate*, $\widehat{\boldsymbol{F}}_{\text{EQL}}$, defined by

$$\widehat{\boldsymbol{F}}_{\text{EQL}} = \boldsymbol{W}'^T \widehat{\boldsymbol{F}}_{\text{ALS}} \boldsymbol{W}$$

is approximately unbiased in a statistical sense. The matrices $\boldsymbol{W}$ and $\boldsymbol{W}'$ are analogs of the normalization matrices $\boldsymbol{T}$ and $\boldsymbol{T}'$ used in the normalized eight-point algorithm. Various choices of weight matrices are possible. The resulting equilibrated ALS methods are akin to, but not identical with, Hartley's technique.

Standard weight matrices are obtained by exploiting a measurement error model whereby noise is present in *one* image only [15]. In the next section we shall establish an unbiasedness property for the normalized eight-point algorithm using a statistical model that incorporates noise in *both* images. This model is an extension of the model described in the previous section. We present the extended model and operational entities related to it in this section. The argument leading to the unbiasedness property draws upon Mühlich–Mester's approach. While tailored to elucidate the utility of Hartley normalization for the eight-point algorithm, it can also serve, with due modifications,

to buttress Hartley normalization for other ALS methods like those appropriate for conic fitting or homography estimation. A distinctive feature of our result is its asymptotic character—only if the number of data points increases will the bias in estimates decrease to zero.

We proceed to describe the extended model. A set of N image correspondences is modeled statistically using a set of pairs of random variables $\{(\mathbf{m}_n, \mathbf{m}'_n)\}_{n=1}^N$. The number of corresponding pairs is not held constant but is allowed to vary from 1 to infinity. Thus, effectively, the model involves the infinite set of pairs of random variables $\{(\mathbf{m}_n, \mathbf{m}'_n)\}_{n=1}^\infty$. The given data points $\boldsymbol{m}_n$, $\boldsymbol{m}'_n$ are viewed as samples drawn from $\mathbf{m}_n$, $\mathbf{m}'_n$, respectively. The $\mathbf{m}_n$, $\mathbf{m}'_n$ are represented as

$$\mathbf{m}_n = \mathbf{n}_n + \Delta\mathbf{m}_n, \quad \mathbf{m}'_n = \mathbf{n}'_n + \Delta\mathbf{m}'_n,$$

where $\mathbf{n}_n$, $\mathbf{n}'_n$ are random variables describing "true" locations and $\Delta\mathbf{m}_n = [\Delta m_{x,n}, \Delta m_{y,n}, 0]^T$, $\Delta\mathbf{m}'_n = [\Delta m'_{x,n}, \Delta m'_{y,n}, 0]^T$ are random variables representing noise that corrupts the "true" locations. All the $\mathbf{n}_n, \mathbf{n}'_n, \Delta\mathbf{m}_n, \Delta\mathbf{m}'_n$ are assumed mutually independent. The $\Delta\mathbf{m}_n, \Delta\mathbf{m}'_n$ encapsulate homogeneous anisotropic noise with principal components aligned with the x and y axes—they all have zero mean and satisfy

$$\mathsf{E}[\Delta\mathbf{m}_n \Delta\mathbf{m}_n^T] = \mathrm{diag}(\sigma_x^2, \sigma_y^2, 0),$$
$$\mathsf{E}[\Delta\mathbf{m}'_n \Delta\mathbf{m}_n'^T] = \mathrm{diag}(\sigma_x'^2, \sigma_y'^2, 0)$$

for some $\sigma_x, \sigma_y, \sigma'_x, \sigma'_y > 0$. The $\mathbf{n}_n$, $\mathbf{n}'_n$ capture various possible arrangements of ideal data points and are assumed to undulate anisotropically around some fixed, nonrandom "centroids" $\boldsymbol{c}$, $\boldsymbol{c}'$:

$$\mathbf{n}_n = \boldsymbol{c} + \Delta\mathbf{n}_n, \quad \mathbf{n}'_n = \boldsymbol{c}' + \Delta\mathbf{n}'_n.$$

Here $\Delta\mathbf{n}_n = [\Delta n_{x,n}, \Delta n_{y,n}, 0]^T$, $\Delta\mathbf{n}'_n = [\Delta n'_{x,n}, \Delta n'_{y,n}, 0]^T$ are zero-mean random variables satisfying

$$\mathsf{E}[\Delta\mathbf{n}_n \Delta\mathbf{n}_n^T] = \mathrm{diag}(\tau_x^2, \tau_y^2, 0),$$
$$\mathsf{E}[\Delta\mathbf{n}'_n \Delta\mathbf{n}_n'^T] = \mathrm{diag}(\tau_x'^2, \tau_y'^2, 0)$$

for some fixed $\tau_x, \tau_y, \tau'_x, \tau'_y > 0$. Note that under the above model, all the $\Delta\mathbf{m}_n, \Delta\mathbf{m}'_n, \Delta\mathbf{n}_n, \Delta\mathbf{n}'_n$ are mutually independent. Note also that the $\mathbf{n}_n$, $\mathbf{n}'_n$ being mutually independent—an assumption that is part of the model—is not compatible with the requirement that, for each N, there exists a genuine random fundamental matrix $\mathbf{F}_N$ such that $\mathbf{n}_n'^T \mathbf{F}_N \mathbf{n}_n = 0$ for all $n = 1, \dots, N$. However, the $\mathbf{n}_n$, $\mathbf{n}'_n$ being independent is consistent with the $\mathbf{m}_n$, $\mathbf{m}'_n$ being independent. As the latter property underpins the statistical model presented in the previous section, one that we aim to extend, we accept the independence of the $\mathbf{n}_n$, $\mathbf{n}'_n$ as the leading feature. Let

$$\overline{\mathbf{m}}_N = \frac{1}{N}\sum_{n=1}^N \mathbf{m}_n, \quad \overline{\mathbf{m}}'_N = \frac{1}{N}\sum_{n=1}^N \mathbf{m}'_n$$

and

$$\mathsf{s}_{x,N} = \left(\frac{1}{N}\sum_{n=1}^N (\mathsf{m}_{x,n} - \overline{\mathsf{m}}_{x,N})^2 \right)^{1/2},$$

$$\mathsf{s}_{y,N} = \left(\frac{1}{N}\sum_{n=1}^N (\mathsf{m}_{y,n} - \overline{\mathsf{m}}_{y,N})^2 \right)^{1/2},$$

$$\mathsf{s}'_{x,N} = \left(\frac{1}{N}\sum_{n=1}^N (\mathsf{m}'_{x,n} - \overline{\mathsf{m}}'_{x,N})^2 \right)^{1/2},$$

$$\mathsf{s}'_{y,N} = \left(\frac{1}{N}\sum_{n=1}^N (\mathsf{m}'_{y,n} - \overline{\mathsf{m}}'_{y,N})^2 \right)^{1/2}.$$

Let

$$\mathbf{T}_N = \begin{bmatrix} \mathsf{s}_{x,N}^{-1} & 0 & -\mathsf{s}_{x,N}^{-1}\overline{\mathsf{m}}_{x,N} \\ 0 & \mathsf{s}_{y,N}^{-1} & -\mathsf{s}_{y,N}^{-1}\overline{\mathsf{m}}_{y,N} \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{T}'_N = \begin{bmatrix} \mathsf{s}_{x,N}'^{-1} & 0 & -\mathsf{s}_{x,N}'^{-1}\overline{\mathsf{m}}'_{x,N} \\ 0 & \mathsf{s}_{y,N}'^{-1} & -\mathsf{s}_{y,N}'^{-1}\overline{\mathsf{m}}'_{y,N} \\ 0 & 0 & 1 \end{bmatrix}.$$

Here and in the following, the additional subscript N serves to emphasize the dependence on N of various quantities involved. Using the strong law of large numbers one can show that, with N increasing to infinity, the sequences $\{\mathbf{T}_N\}_{N=1}^\infty$ and $\{\mathbf{T}'_N\}_{N=1}^\infty$ converge with probability 1 to certain nonrandom matrices $\boldsymbol{T}$ and $\boldsymbol{T}'$ explicitly expressible in terms of $\sigma_x, \sigma_y, \tau_x, \tau_y$ and $\sigma'_x, \sigma'_y, \tau'_x, \tau'_y$. Now, for each $n = 1, \dots, N$, $\tilde{\mathbf{m}}_{n,N} = \mathbf{T}_N \mathbf{m}_n$ and $\tilde{\mathbf{m}}'_{n,N} = \mathbf{T}'_N \mathbf{m}'_n$ are the normalized noisy image points derived from $\mathbf{m}_n$ and $\mathbf{m}'_n$, with the normalization based on $\{(\mathbf{m}_n, \mathbf{m}'_n)\}_{n=1}^N$. We define $\tilde{\mathbf{n}}_n = \boldsymbol{T}\mathbf{n}_n$, $\tilde{\mathbf{n}}'_n = \boldsymbol{T}'\mathbf{n}'_n$ for every n ranging from 1 to infinity. Let $\mathbf{x}_n = [\mathsf{m}_{n,x}, \mathsf{m}_{n,y}, \mathsf{m}'_{n,x}, \mathsf{m}'_{n,y}]^T$ and $\mathbf{y}_n = [\mathsf{n}_{n,x}, \mathsf{n}_{n,y}, \mathsf{n}'_{n,x}, \mathsf{n}'_{n,y}]^T$ be the joint image points derived from $(\mathbf{m}_n, \mathbf{m}'_n)$ and $(\mathbf{n}_n, \mathbf{n}'_n)$, respectively. Let $\tilde{\mathbf{x}}_{n,N} = [\tilde{\mathsf{m}}_{n,N,x}, \tilde{\mathsf{m}}_{n,N,y}, \tilde{\mathsf{m}}'_{n,N,x}, \tilde{\mathsf{m}}'_{n,N,y}]^T$ and $\tilde{\mathbf{y}}_{n,N} = [\tilde{\mathsf{n}}_{n,N,x}, \tilde{\mathsf{n}}_{n,N,y}, \tilde{\mathsf{n}}'_{n,N,x}, \tilde{\mathsf{n}}'_{n,N,y}]^T$ be the corresponding transformed joint image points. Let

$$\tilde{\mathbf{A}}_N = \frac{1}{N}\sum_{n=1}^N \boldsymbol{u}(\tilde{\mathbf{x}}_{n,N})\boldsymbol{u}(\tilde{\mathbf{x}}_{n,N})^T,$$

$$\tilde{\mathbf{B}}_N = \frac{1}{N}\sum_{n=1}^N \boldsymbol{u}(\tilde{\mathbf{y}}_{n,N})\boldsymbol{u}(\tilde{\mathbf{y}}_{n,N})^T.$$

Here, the common normalizing factor $1/N$ is introduced to maintain an overall statistical balance—a quality of importance in performing revelant calculations. Crucially, this factor does not affect the eigenvectors of the matrices involved. For each $j = 1, \ldots, 9$, let $\mathbf{a}_{j,N}$ be the jth normalized eigenvector of $\tilde{\mathbf{A}}_N$ (defined uniquely up to a sign) with a corresponding eigenvalue $\alpha_{j,N}$, and let $\mathbf{b}_{j,N}$ be the jth normalized eigenvector of $\tilde{\mathbf{B}}_N$ with a corresponding eigenvalue $\beta_{j,N}$. The eigenvalues are assumed to be arranged in descending order. The random eigenvectors $\mathbf{a}_{9,N}$ and $\mathbf{b}_{9,N}$ can be restructured to form random matrices $\widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N}$ and $\tilde{\underline{\mathbf{F}}}_N$ such that

$$\mathbf{a}_{9,N} = \mathrm{vec}(\widehat{\tilde{\mathbf{F}}}^T_{\mathrm{ALS},N}), \quad \mathbf{b}_{9,N} = \mathrm{vec}(\tilde{\underline{\mathbf{F}}}^T_N).$$

Then, clearly,

$$\widehat{\mathbf{F}}_{\mathrm{NALS},N} = \mathbf{T}'^T_N \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \mathbf{T}_N$$

is the NALS estimate based on $\{\mathbf{x}_n\}_{n=1}^N$. The matrix

$$\underline{\mathbf{F}}_N = \boldsymbol{T}'^T \tilde{\underline{\mathbf{F}}}_N \boldsymbol{T}$$

will in what follows constitute a reference value against which to compare $\widehat{\mathbf{F}}_{\mathrm{NALS},N}$. Obviously, if the $\{(\mathbf{n}_n, \mathbf{n}'_n)\}_{n=1}^N$ were genuine correspondences bound by a fundamental matrix $\mathbf{F}_N$, then $\mathbf{F}_N$ would be a natural "true" value with which to gauge $\widehat{\mathbf{F}}_{\mathrm{NALS},N}$. Since, generally, no such $\mathbf{F}_N$ exists, we use $\underline{\mathbf{F}}_N$ instead. The significance of this choice can only be fully appreciated once the whole argument is complete.

With $\mathscr{A}_N = \{\mathbf{n}_n, \mathbf{n}'_n\}_{n=1}^N$, let $\mathsf{E}[\widehat{\mathbf{F}}_{\mathrm{NALS},N}|\mathscr{A}_N]$ be the *conditional expectation* of $\widehat{\mathbf{F}}_{\mathrm{NALS},N}$ given $\mathscr{A}_N$. We adopt

$$\Delta_N = \|\mathsf{E}[\widehat{\mathbf{F}}_{\mathrm{NALS},N}|\mathscr{A}_N] - \underline{\mathbf{F}}_N\|$$

as a measure of *bias* in the $\widehat{\mathbf{F}}_{\mathrm{NALS},N}$ estimates. The random variable Δ_N captures the bias conditional on any particular arrangement of the "true" data points $\{\mathbf{n}_n, \mathbf{n}'_n\}_{n=1}^N$. Assuming that all the $\{\mathbf{n}_n, \mathbf{n}'_n, \mathbf{m}_n, \mathbf{m}'_n\}_{n=1}^\infty$ are uniformly bounded random variables carried on a single probability space, we shall outline an argument to show that a certain *approximation* to Δ_N tends to 0 as $N \to \infty$ with probability 1. This asymptotic approximate unbiasedness will be referred to as *approximate consistency*. It can be viewed as one of advantageous effects of data normalization. The establishment of the property parallels similar findings related to other methods [10, 14, 18].

5. Approximate Consistency

Let $\mathscr{A}_\infty = \bigcup_{N=1}^\infty \mathscr{A}_N = \{\mathbf{n}_n, \mathbf{n}'_n\}_{n=1}^\infty$. Since both sequences $\{\mathbf{T}_N\}_{N=1}^\infty$, $\{\mathbf{T}'_N\}_{N=1}^\infty$ are uniformly bounded

(given that the family $\{\mathbf{m}_n, \mathbf{m}'_n\}_{n=1}^\infty$ is uniformly bounded) and pointwise converge to $\boldsymbol{T}$, $\boldsymbol{T}'$ almost surely[1] (a.s.), an application of the conditional form of the dominated convergence theorem [17, Sec. 2, §7, Thm. 2] implies that

$$\lim_{N \to \infty} \mathsf{E}[\mathbf{T}'^T_N \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \mathbf{T}_N - \boldsymbol{T}'^T \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \boldsymbol{T}|\mathscr{A}_\infty] = \mathbf{0}.$$

On the other hand,

$$\mathsf{E}[\mathbf{T}'^T_N \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \mathbf{T}_N|\mathscr{A}_\infty] = \mathsf{E}[\mathbf{T}'^T_N \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \mathbf{T}_N|\mathscr{A}_N]$$

$$\mathsf{E}[\boldsymbol{T}'^T \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \boldsymbol{T}|\mathscr{A}_\infty] = \boldsymbol{T}'^T \mathsf{E}[\widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N}|\mathscr{A}_N]\boldsymbol{T}$$

because $\mathbf{T}_N$, $\mathbf{T}'_N$, and $\widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N}$ being expressible in terms of $\{\mathbf{m}_n, \mathbf{m}'_n, \mathbf{n}_n, \mathbf{n}'_n\}_{n=1}^N$ are jointly independent of $\{\mathbf{n}_n, \mathbf{n}'_n\}_{n=N+1}^\infty$. Therefore the difference

$$\mathsf{E}[\widehat{\mathbf{F}}_{\mathrm{NALS},N}|\mathscr{A}_N] - \underline{\mathbf{F}}_N$$
$$= \mathsf{E}[\mathbf{T}'^T_N \widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N} \mathbf{T}_N|\mathscr{A}_N] - \boldsymbol{T}'^T \tilde{\underline{\mathbf{F}}}_N \boldsymbol{T}$$

differs from

$$\boldsymbol{T}'^T \mathsf{E}[\widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N}|\mathscr{A}_N]\boldsymbol{T} - \boldsymbol{T}'^T \tilde{\underline{\mathbf{F}}}_N \boldsymbol{T}$$

by an expression tending to zero almost surely. We see that the asymptotic behavior of Δ_N is fully controlled by the bias δ_N defined by

$$\delta_N = \|\mathsf{E}[\widehat{\tilde{\mathbf{F}}}_{\mathrm{ALS},N}|\mathscr{A}_N] - \tilde{\underline{\mathbf{F}}}_N\|$$

or, equivalently, by

$$\delta_N = \|\mathsf{E}[\mathbf{a}_{9,N}|\mathscr{A}_N] - \mathbf{b}_{9,N}\|.$$

The value of δ_N is difficult to estimate because of the lack of a simple expression for $\mathbf{a}_{9,N}$. We shall get around this problem by considering a first-order approximation to $\mathbf{a}_{9,N}$. Specifically, we shall use the following approximate formula for $\mathbf{a}_{j,N}$

$$\mathbf{a}_{j,N} = \mathbf{b}_{j,N} + \sum_{\substack{k=1 \\ k \neq j}}^9 \frac{\mathbf{b}_{k,N}^T (\tilde{\mathbf{A}}_N - \tilde{\mathbf{B}}_N)\mathbf{b}_{j,N}}{\beta_{j,N} - \beta_{k,N}} \mathbf{b}_{k,N}$$

holding for each $1 \leq j \leq 9$ [6, Sec. 7.2]. We shift the focus from the exact bias δ_N to an approximate bias δ_N^* defined by

$$\delta_N^* = \left\| \sum_{k=1}^8 \frac{\mathbf{b}_{k,N}^T (\mathsf{E}[\tilde{\mathbf{A}}_N|\mathscr{A}_N] - \tilde{\mathbf{B}}_N)\mathbf{b}_{9,N}}{\beta_{9,N} - \beta_{k,N}} \mathbf{b}_{k,N} \right\|.$$

---

[1] Recall that the phrase "almost surely" is a concise, precise way to state "except on an event of probability zero". It is equivalent to "with probability 1".

Now, it can be shown that

$$\gamma = \lim_{N \to \infty} (\beta_{8,N} - \beta_{9,N})^{-1} < \infty \quad \text{a.s.}$$

for a generic choice of $\sigma_x, \sigma_y, \tau_x, \tau_y$ and $\sigma_x', \sigma_y', \tau_x', \tau_y'$. This condition automatically implies that

$$\limsup_{N \to \infty} (\beta_{j,N} - \beta_{9,N})^{-1} \leq \gamma < \infty \quad \text{a.s.}$$

for all $1 \leq j \leq 8$. It can be also shown—and this is the main computational step, details of which we omit due to space limitations—that

$$\lim_{n \to \infty} \mathsf{E}[\tilde{\mathbf{A}}_N | \mathscr{A}_N] = \boldsymbol{I}_9 \quad \text{a.s.},$$

where, for each $n = 1, 2, \ldots, \boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. Hence, with probability 1,

$$\limsup_{N \to \infty} \delta_N^* \leq \gamma \limsup_{N \to \infty} \sum_{k=1}^{8} |\mathbf{b}_{k,N}^T (\boldsymbol{I}_9 - \tilde{\mathbf{B}}_N) \mathbf{b}_{9,N}| \|\mathbf{b}_{k,N}\|.$$

Since $\|\mathbf{b}_{k,N}\| = 1$ and

$$\mathbf{b}_{k,N}^T (\boldsymbol{I}_9 - \tilde{\mathbf{B}}_N) \mathbf{b}_{9,N} = (1 - \beta_{9,N}) \mathbf{b}_{k,N}^T \mathbf{b}_{9,N} = 0$$

for $1 \leq k \leq 8$, it follows that $\lim_{N \to \infty} \delta_N^* = 0$ with probability 1. We thus see that, when quantified with $\lim_{N \to \infty} \delta_N^*$, the asymptotic approximate bias in the normalized eight-point algorithm is zero almost surely. This is our sought-after approximate consistency result.

6. Conclusion

A new argument has been proposed to explain the improvement in performance of the eight-point algorithm that results from using normalized data. The argument relies upon adoption of a structured model for statistical data distribution. In the framework of this model, the advantage of using normalized data is attributed to a systematic increase in the consistency of estimates as the number of data points mounts. The proposed approach merges the previous statistical argument explaining the advantageous character of data normalization with the approach of Mühlich and Mester for enhancing total least squares estimation methods via equilibration. As such it contributes to a wider effort to gain a deeper understanding of the rationale underlying foundational methods in vision parameter estimation.

Acknowledgments

References

[1] W. Chojnacki, M. J. Brooks, and A. van den Hengel. Rationalising the renormalisation method of Kanatani. *J. Math. Imaging Vision*, 14(1):21–38, 2001.

[2] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. On the fitting of surfaces to data with covariances. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(11):1294–1303, 2000.

[3] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. Revisiting Hartley's normalized eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1172–1177, 2003.

[4] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley. FNS, CFNS and HEIV: A unifying approach. *J. Math. Imaging Vision*, 23(2):175–183, 2005.

[5] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, Cambridge, Mass., 2001.

[6] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, Maryland, 2nd edition, 1989.

[7] R. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(6):580–593, 1997.

[8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, 2000.

[9] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, Amsterdam, 1996.

[10] A. Kukush, I. Markovsky, and S. van Huffel. Consistent fundamental matrix estimation in a quadratic measurement error model arising in motion analysis. *Comput. Statist. Data Anal.*, 41(1):3–18, 2002.

[11] Y. Leedan and P. Meer. Heteroscedastic regression in computer vision: problems with bilinear constraint. *Int. J. Comput. Vision*, 37(2):127–150, 2000.

[12] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(10):133–135, 1981.

[13] H. Lütkepol. *Handbook of Matrices*. John Wiley & Sons, Chichester, 1996.

[14] W. J. MacLean. Removal of translation bias when using subspace methods. In *Proc. Seventh Int. Conf. Computer Vision, Corfu, Greece*, volume 2, pages 753–758, 1999.

[15] M. Mühlich and R. Mester. The role of total least squares in motion analysis. In *Proc. Fifth European Conf. Computer Vision, Freiburg, Germany*, volume 1407 of *Lecture Notes in Computer Science*, pages 305–321, 1998.

[16] M. Mühlich and R. Mester. Subspace methods and equilibration in computer vision. In *Proc. 12th Scandinavian Conf. Image Analysis, Stavanger, Norway*, pages 415–422, 2001.

[17] A. N. Shiryaev. *Probability*. Springer-Verlag, New York, 2 edition, 1996.

[18] T. Zhang and C. Tomasi. On the consistency of instantaneous rigid motion estimation. *Int. J. Comput. Vision*, 46(1):51–79, 2002.