Revisiting Hartley's Normalized Eight-Point Algorithm

Wojciech Chojnacki, Michael J. Brooks, Anton van den Hengel, and Darren Gawley

Abstract—**Hartley's eight-point algorithm has maintained an important place in computer vision, notably as a means of providing an initial value of the fundamental matrix for use in iterative estimation methods. In this paper, a novel explanation is given for the improvement in performance of the eight-point algorithm that results from using normalized data. It is first established that the normalized algorithm acts to minimize a specific cost function. It is then shown that this cost function is statistically better founded than the cost function associated with the nonnormalized algorithm. This augments the original argument that improved performance is due to the better conditioning of a pivotal matrix. Experimental results are given that support the adopted approach. This work continues a wider effort to place a variety of estimation techniques within a coherent framework.**

Index Terms—**Epipolar equation, fundamental matrix, eight-point algorithm, data normalization**

I. INTRODUCTION

IN a landmark paper, Longuet-Higgins [1] proposed the *eight-point algorithm*—a simple direct method for computation of the *essential matrix*. The algorithm extends straightforwardly to computation of the *fundamental matrix*, the uncalibrated analogue of the essential matrix [2], [3]. While simple and fast, the algorithm is very sensitive to noise in the specification of the image coordinates serving as input for computation and, as such, is of limited use. Many alternative methods have been advanced since Longuet-Higgins' proposal, including more sophisticated and computationally intensive iterative algorithms [4], [5]. Hartley [6] discovered that the accuracy of the eight-point algorithm can be greatly improved if, prior to applying the method, a simple normalization of image data is performed. This fundamental modification dramatically extended the applicability of the algorithm, and, in particular, rendered it an excellent tool for generation of initial estimates for iterative methods.

Hartley attributed the improved performance of the normalized eight-point algorithm to the better numerical conditioning of a pivotal matrix used in solving an eigenvalue problem. In this paper, we analyze the normalized eight-point algorithm and offer a new insight into the working of the method. A crucial observation is that the estimate produced by the normalized eight-point algorithm can be identified with the minimizer of a cost function. The minimizer can be directly calculated by solving a generalized eigenvalue problem. We confirm experimentally that the estimate obtained as a solution

The authors are with the School of Computer Science, University of Adelaide, SA 5005, Australia.
Email: {wojtek, mjb, hengel, dg}@cs.adelaide.edu.au

of the generalized eigenproblem coincides with the estimate generated by Hartley's original method. Exploiting the cost function, we propose an alternative explanation of the improved performance of the normalized eight-point algorithm, based on a certain statistical model of data distribution. Under this model, the summands of the cost function underlying the normalized eight-point algorithm turn out to be more balanced in terms of spread than the summands of the cost function underlying the standard eight-point algorithm. Summation of more balanced terms leads to a statistically more appropriate expression for minimization, and this in turn translates into a more accurate estimator. The proposed approach continues a line of research due to Torr [7], Mühlich and Mester [8], and Torr and Fitzgibbon [9], in which variants of the normalized eight-point algorithm are analyzed statistically. The work presented here also forms part of a wider effort to place a variety of estimation techniques within a coherent framework (e.g., see [10], [11], [12], [13]).

II. ESTIMATION PROBLEM

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_1, m_2) of coordinates, or equivalently, by the "homogeneous" vector $\boldsymbol{m} = [m_1, m_2, 1]^T$. 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 $(\boldsymbol{m}, \boldsymbol{m}')$, this pair 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 [4], [5]. In addition to (1), $\boldsymbol{F}$ is subject to the *singularity constraint* (or, equivalently, the *rank-2 constraint*)

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

Using $\boldsymbol{x} = [m_1, m_2, m_1', m_2']^T$ as a compact descriptor of the single *image datum* $(\boldsymbol{m}, \boldsymbol{m}')$, the estimation problem associated with (1) and (2) can be stated as follows: Given a collection $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ of image data and a meaningful *cost function* that characterizes the extent to which any particular $\boldsymbol{F}$ fails to satisfy the system of the copies of equation (1) associated with $\boldsymbol{x} = \boldsymbol{x}_i$ ($i = 1, \ldots, n$), find an *estimate* $\widehat{\boldsymbol{F}} \neq \boldsymbol{0}$ satisfying (2) for which the cost function attains its minimum. Since (1) and (2) do not change when $\boldsymbol{F}$ is multiplied by a nonzero scalar, $\widehat{\boldsymbol{F}}$ is to be found only up to scale. If the singularity constraint is set aside, then the estimate associated

with a particular cost function $J = J(\boldsymbol{F}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ is defined as the *unconstrained* minimizer $\widehat{\boldsymbol{F}}$ of J,

$$\widehat{\boldsymbol{F}} = \arg \min_{\boldsymbol{F} \neq \boldsymbol{0}} J(\boldsymbol{F}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n).$$

Such an estimate can further be converted to a nearby rank-2 fundamental matrix by applying one of a variety of methods [4], [12]. In this paper, we shall confine our attention to the pivotal component of this overall process that determines exclusively the unconstrained minimizer, as this will prove critical to rationalizing the Hartley method. For alternative *integrated* approaches to computing a constrained minimizer, see the CFNS method [14], [15] or the Gold Standard Method [4].

III. ALGEBRAIC LEAST SQUARES

A straightforward estimation method employs the cost function

$$J_{\text{ALS}}(\boldsymbol{F}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \|\boldsymbol{F}\|_F^{-2} \sum_{i=1}^{n} (\boldsymbol{m}_i'^T \boldsymbol{F} \boldsymbol{m}_i)^2 \quad (3)$$

with $\|\boldsymbol{F}\|_F = (\sum_{i,j} f_{ij}^2)^{1/2}$ the Frobenius norm of $\boldsymbol{F}$. Here, $\boldsymbol{m}_i'^T \boldsymbol{F} \boldsymbol{m}_i$ is the signed *algebraic distance* between the individual datum $\boldsymbol{x}_i$ and the candidate matrix $\boldsymbol{F}$. The *algebraic least squares* (ALS) *estimate*, $\widehat{\boldsymbol{F}}_{\text{ALS}}$, is defined as the minimizer of J_{ALS}.

A practical means for finding $\widehat{\boldsymbol{F}}_{\text{ALS}}$ is conveniently derived based on an alternative expression for $\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m}$. For an $m \times n$ matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n]$ with $\boldsymbol{a}_j$ the jth column vector of length m, denote by $\text{vec}(\boldsymbol{A})$ the *vectorization* of $\boldsymbol{A}$, i.e., the column vector of length mn defined by $\text{vec}(\boldsymbol{A}) = [\boldsymbol{a}_1^T, \ldots, \boldsymbol{a}_n^T]^T$. Let $\boldsymbol{\theta} = \text{vec}(\boldsymbol{F}^T)$ and $\boldsymbol{u}(\boldsymbol{x}) = \text{vec}(\boldsymbol{m}\boldsymbol{m}'^T)$. Then, by applying standard rules for matrix operations [16], we find that

$$\begin{aligned} \boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m} &= \text{tr}(\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m}) = \text{tr}(\boldsymbol{m}\boldsymbol{m}'^T \boldsymbol{F}) \\ &= \text{vec}(\boldsymbol{F}^T)^T \text{vec}(\boldsymbol{m}\boldsymbol{m}'^T) = \boldsymbol{\theta}^T \boldsymbol{u}(\boldsymbol{x}), \end{aligned} \quad (4)$$

where tr denotes trace. With this formula, J_{ALS} can be written as

$$J_{\text{ALS}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \|\boldsymbol{\theta}\|^{-2} \boldsymbol{\theta}^T \boldsymbol{A} \boldsymbol{\theta}, \quad (5)$$

where

$$\boldsymbol{A} = \sum_{i=1}^{n} \boldsymbol{u}(\boldsymbol{x}_i) \boldsymbol{u}(\boldsymbol{x}_i)^T \quad (6)$$

and $\|\boldsymbol{\theta}\| = (\theta_1^2 + \cdots + \theta_9^2)^{1/2}$. Now, using (5) to evolve a variational equation for the minimizer, $\widehat{\boldsymbol{\theta}}_{\text{ALS}}$ can be characterized as an eigenvector of $\boldsymbol{A}$ associated with the smallest eigenvalue [10]. This eigenvector can be found in practice by performing *singular value decomposition* (SVD) of the matrix

$$\boldsymbol{M} = [\boldsymbol{u}(\boldsymbol{x}_1), \ldots, \boldsymbol{u}(\boldsymbol{x}_n)]^T \quad (7)$$

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 estimator is essentially identical to the eight-point algorithm of Longuet-Higgins [1]. For $\widehat{\boldsymbol{\theta}}_{\text{ALS}}$ to be uniquely determined, the null space of $\boldsymbol{A}$ cannot be more than one-dimensional. Given

that $\boldsymbol{A}$ is the sum of the rank-1 matrices $\boldsymbol{u}(\boldsymbol{x}_i)\boldsymbol{u}(\boldsymbol{x}_i)^T$, this is secured, generically, by the condition $n \geq 8$, hence the qualification "eight-point".

IV. HARTLEY'S APPROACH

Let $\overline{\boldsymbol{m}}$ and $\overline{\boldsymbol{m}}'$ be the *centroids*, or "centers of mass", of the $\boldsymbol{m}_i$ and the $\boldsymbol{m}_i'$, respectively, defined by

$$\overline{\boldsymbol{m}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}_i \quad \text{and} \quad \overline{\boldsymbol{m}}' = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}_i'. \quad (8)$$

Let $\overline{\boldsymbol{m}} = [\overline{m}_1, \overline{m}_2, 1]^T$, $\overline{\boldsymbol{m}}' = [\overline{m}_1', \overline{m}_2', 1]^T$, $\boldsymbol{m}_i = [m_{1,i}, m_{2,i}, 1]^T$, $\boldsymbol{m}_i' = [m_{1,i}', m_{2,i}', 1]^T$ ($i = 1, \ldots, n$). Following Hartley [6], let us shift the image coordinate systems to the respective centroids. In coordinates associated with the transformed systems, the points of the ith image datum can be written $[m_{1,i} - \overline{m}_1, m_{2,i} - \overline{m}_2, 1]^T$ and $[m_{1,i}' - \overline{m}_1', m_{2,i}' - \overline{m}_2', 1]^T$. Let

$$\begin{aligned} s &= \left(\frac{1}{2n} \sum_{i=1}^{n} \|\boldsymbol{m}_i - \overline{\boldsymbol{m}}\|^2 \right)^{1/2} \\ &= \left(\frac{1}{2n} \sum_{i=1}^{n} (m_{1,i} - \overline{m}_1)^2 + (m_{2,i} - \overline{m}_2)^2 \right)^{1/2}, \\ s' &= \left(\frac{1}{2n} \sum_{i=1}^{n} \|\boldsymbol{m}_i' - \overline{\boldsymbol{m}}'\|^2 \right)^{1/2} \\ &= \left(\frac{1}{2n} \sum_{i=1}^{n} (m_{1,i}' - \overline{m}_1')^2 + (m_{2,i}' - \overline{m}_2')^2 \right)^{1/2}. \end{aligned} \quad (9)$$

Define the *normalized* data as

$$\begin{aligned} \tilde{\boldsymbol{m}}_i &= [(m_{1,i} - \overline{m}_1)/s, (m_{2,i} - \overline{m}_2)/s, 1]^T, \\ \tilde{\boldsymbol{m}}_i' &= [(m_{1,i}' - \overline{m}_1')/s', (m_{2,i}' - \overline{m}_2')/s', 1]^T. \end{aligned}$$

This definition ensures that the root mean square distance of the $\tilde{\boldsymbol{m}}_i$ to the origin of the corresponding coordinate system is equal to $\sqrt{2}$ and, likewise, for the $\tilde{\boldsymbol{m}}_i'$. The normalized data can be alternatively defined by $\tilde{\boldsymbol{m}}_i = \boldsymbol{T}\boldsymbol{m}_i$ and $\tilde{\boldsymbol{m}}_i' = \boldsymbol{T}'\boldsymbol{m}_i'$, where

$$\begin{aligned} \boldsymbol{T} &= \begin{bmatrix} s^{-1} & 0 & -s^{-1}\overline{m}_1 \\ 0 & s^{-1} & -s^{-1}\overline{m}_2 \\ 0 & 0 & 1 \end{bmatrix}, \\ \boldsymbol{T}' &= \begin{bmatrix} s'^{-1} & 0 & -s'^{-1}\overline{m}_1' \\ 0 & s'^{-1} & -s'^{-1}\overline{m}_2' \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Let $\tilde{\boldsymbol{x}}_i = [\tilde{m}_{1,i}, \tilde{m}_{2,i}, \tilde{m}_{1,i}', \tilde{m}_{2,i}']^T$. Denote by $\widehat{\tilde{\boldsymbol{F}}}_{\text{ALS}}$ the minimizer of the ALS cost function seeded with the normalized data, that is, the minimizer of the function $\boldsymbol{F} \mapsto J_{\text{ALS}}(\boldsymbol{F}; \tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n)$. Let $\boldsymbol{F} \mapsto \tilde{\boldsymbol{F}}$ be the mapping defined by

$$\tilde{\boldsymbol{F}} = \boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1}. \quad (10)$$

Clearly, if $\tilde{\boldsymbol{m}} = \boldsymbol{T}\boldsymbol{m}$ and $\tilde{\boldsymbol{m}}' = \boldsymbol{T}'\boldsymbol{m}'$, then $\boldsymbol{m}'^T \boldsymbol{F} \boldsymbol{m} = \tilde{\boldsymbol{m}}'^T \tilde{\boldsymbol{F}} \tilde{\boldsymbol{m}}$. Accordingly, the image of $\widehat{\tilde{\boldsymbol{F}}}_{\text{ALS}}$ by the inverse mapping $\tilde{\boldsymbol{F}} \mapsto \boldsymbol{F}$ can be viewed as a genuine estimate of

$\boldsymbol{F}$. We term this the *Hartley* (HRT) *estimate* of $\boldsymbol{F}$ and write $\widehat{\boldsymbol{F}}_{\mathrm{HRT}}$; it is explicitly given by

$$\widehat{\boldsymbol{F}}_{\mathrm{HRT}} = \boldsymbol{T}'^T \widehat{\boldsymbol{F}}_{\mathrm{ALS}} \boldsymbol{T}. \tag{11}$$

The introduction of $\widehat{\boldsymbol{F}}_{\mathrm{HRT}}$ is motivated by the fact that if the *modified condition number* of a nonnegative definite matrix defined as the ratio of the greatest to the second smallest eigenvalues is large, then the two least eigenvalues are relatively close to one another; this makes the corresponding eigenvectors "wobbly", whereby a small perturbation of the matrix entries is conducive to a significant change of the eigenvectors, as these can fall anywhere within a vicinity of a two-dimensional eigenspace associated with a virtual degenerate eigenvalue. The matrix $\tilde{\boldsymbol{A}} = \sum_{i=1}^{n} \boldsymbol{u}(\tilde{\boldsymbol{x}}_i)\boldsymbol{u}(\tilde{\boldsymbol{x}}_i)^T$ serving to calculate $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$ is, in practice, much better conditioned (in the above sense) than the matrix $\boldsymbol{A}$ with which $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$ is calculated. As a result, Hartley's method is better conditioned (less sensitive to errors in data) and, in this sense, more advantageous than the ALS method.

V. Normalized Algebraic Least Squares

We now identify the Hartley estimate as a minimizer of a cost function. To this end, we introduce

$$J_{\mathrm{NALS}}(\boldsymbol{F}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$$
$$= \|\boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1}\|_F^{-2} \sum_{i=1}^{n} (\boldsymbol{m}_i'^T \boldsymbol{F} \boldsymbol{m}_i)^2. \tag{12}$$

The minimizer of J_{NALS} we call the *normalized algebraic least squares* (NALS) *estimate* of $\boldsymbol{F}$ and write $\widehat{\boldsymbol{F}}_{\mathrm{NALS}}$. The precise sense in which the expressions entering J_{NALS} are normalized will be revealed later. We claim that

$$\widehat{\boldsymbol{F}}_{\mathrm{HRT}} = \widehat{\boldsymbol{F}}_{\mathrm{NALS}}. \tag{13}$$

Indeed, if $\tilde{\boldsymbol{F}}$ and $\boldsymbol{F}$ are related by (10), then $\boldsymbol{m}_i'^T \boldsymbol{F} \boldsymbol{m}_i = \tilde{\boldsymbol{m}}_i'^T \tilde{\boldsymbol{F}} \tilde{\boldsymbol{m}}_i$ for each $i = 1, \ldots, n$, and so

$$J_{\mathrm{NALS}}(\boldsymbol{F}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \|\tilde{\boldsymbol{F}}\|_F^{-2} \sum_{i=1}^{n} (\tilde{\boldsymbol{m}}_i'^T \tilde{\boldsymbol{F}} \tilde{\boldsymbol{m}}_i)^2$$
$$= J_{\mathrm{ALS}}(\tilde{\boldsymbol{F}}; \tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n).$$

From this identity it follows that the function $\boldsymbol{F} \mapsto J_{\mathrm{NALS}}(\boldsymbol{F}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ attains a minimum at a particular $\boldsymbol{F}$ precisely when the function $\tilde{\boldsymbol{F}} \mapsto J_{\mathrm{ALS}}(\tilde{\boldsymbol{F}}; \tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_n)$ attains a minimum at the image $\tilde{\boldsymbol{F}}$ of this $\boldsymbol{F}$ by the mapping $\boldsymbol{F} \mapsto \tilde{\boldsymbol{F}}$. In other words, $\boldsymbol{T}'^{-T} \widehat{\boldsymbol{F}}_{\mathrm{NALS}} \boldsymbol{T}^{-1} = \widehat{\tilde{\boldsymbol{F}}}_{\mathrm{ALS}}$, which, clearly, is equivalent to $\widehat{\boldsymbol{F}}_{\mathrm{NALS}} = \boldsymbol{T}'^T \widehat{\tilde{\boldsymbol{F}}}_{\mathrm{ALS}} \boldsymbol{T}$. Comparing this with (11) establishes the claim.

The formula for J_{NALS} can be rewritten similarly to that for J_{ALS}. The starting point is the identity

$$\|\boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1}\|_F^2 = \boldsymbol{\theta}^T \boldsymbol{C} \boldsymbol{\theta}, \tag{14}$$

where $\boldsymbol{C}$ is the 9×9 matrix given by

$$\boldsymbol{C} = \boldsymbol{T}'^{-1} \boldsymbol{T}'^{-T} \otimes \boldsymbol{T}^{-1} \boldsymbol{T}^{-T}. \tag{15}$$

Here, $\otimes$ stands for the *Kronecker product*, or *tensor product*, of matrices defined as follows: If $\boldsymbol{A} = [a_{ij}]$ is an $m \times n$ matrix

and $\boldsymbol{B}$ is a $p \times q$ matrix, then $\boldsymbol{A} \otimes \boldsymbol{B}$ is the $mp \times nq$ matrix given by

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11} \boldsymbol{B} & \ldots & a_{1n} \boldsymbol{B} \\ \vdots & & \vdots \\ a_{m1} \boldsymbol{B} & \ldots & a_{mn} \boldsymbol{B} \end{bmatrix}.$$

To establish (14), note that

$$\|\boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1}\|_F^2$$
$$= \mathrm{tr}(\boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1} (\boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1})^T)$$
$$= \mathrm{tr}(\boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1} \boldsymbol{T}^{-T} \boldsymbol{F}^T \boldsymbol{T}'^{-1})$$
$$= \mathrm{tr}(\boldsymbol{T}'^{-1} \boldsymbol{T}'^{-T} \boldsymbol{F} \boldsymbol{T}^{-1} \boldsymbol{T}^{-T} \boldsymbol{F}^T)$$
$$= \mathrm{vec}(\boldsymbol{F}^T)^T (\boldsymbol{T}'^{-1} \boldsymbol{T}'^{-T} \otimes (\boldsymbol{T}^{-1} \boldsymbol{T}^{-T})^T) \, \mathrm{vec}(\boldsymbol{F}^T)$$
$$= \mathrm{vec}(\boldsymbol{F}^T)^T (\boldsymbol{T}'^{-1} \boldsymbol{T}'^{-T} \otimes \boldsymbol{T}^{-1} \boldsymbol{T}^{-T}) \, \mathrm{vec}(\boldsymbol{F}^T).$$

In view of (14), we can rewrite (12) as

$$J_{\mathrm{NALS}}(\boldsymbol{\theta}; \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = \frac{\boldsymbol{\theta}^T \boldsymbol{A} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \boldsymbol{C} \boldsymbol{\theta}}. \tag{16}$$

One consequence of this formula is that $\widehat{\boldsymbol{\theta}}_{\mathrm{NALS}}$ is a solution of the generalized eigenvalue problem

$$\boldsymbol{A} \boldsymbol{\theta} = \lambda \boldsymbol{C} \boldsymbol{\theta} \tag{17}$$

corresponding to the smallest eigenvalue. Since $\boldsymbol{A}$ may be ill-conditioned, solving the above eigenvector problem directly requires a numerically stable method. Leedan and Meer [13] proposed one such method which, when applied to the problem under consideration, employs *generalized singular value decomposition* (GSVD) of a pair of matrices $(\boldsymbol{M}, \boldsymbol{N})$ satisfying $\boldsymbol{A} = \boldsymbol{M}^T \boldsymbol{M}$ and $\boldsymbol{C} = \boldsymbol{N}^T \boldsymbol{N}$. Numerical experiments show (see later) that when this method is applied, the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$, of which the first is typically ill-conditioned, lead to a solution identical with the solution obtained using the well-conditioned matrix $\tilde{\boldsymbol{A}}$—in other words, equality (13) is experimentally confirmed.

VI. Two Formulae

To proceed smoothly later, we now present two formulae. The first is the identity

$$\boldsymbol{u}(\boldsymbol{x}_i)\boldsymbol{u}(\boldsymbol{x}_i)^T = \boldsymbol{m}_i' \boldsymbol{m}_i'^T \otimes \boldsymbol{m}_i \boldsymbol{m}_i^T \tag{18}$$

resulting immediately from

$$\mathrm{vec}(\boldsymbol{m}_i \boldsymbol{m}_i'^T) \mathrm{vec}(\boldsymbol{m}_i \boldsymbol{m}_i'^T)^T$$
$$= (\boldsymbol{m}_i' \otimes \boldsymbol{m}_i)(\boldsymbol{m}_i' \otimes \boldsymbol{m}_i)^T$$
$$= (\boldsymbol{m}_i' \otimes \boldsymbol{m}_i)(\boldsymbol{m}_i'^T \otimes \boldsymbol{m}_i^T)$$
$$= \boldsymbol{m}_i' \boldsymbol{m}_i'^T \otimes \boldsymbol{m}_i \boldsymbol{m}_i^T.$$

The second is derived starting from the observation that

$$\boldsymbol{T} = \begin{bmatrix} s^{-1} & 0 & 0 \\ 0 & s^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -\overline{m}_1 \\ 0 & 1 & -\overline{m}_2 \\ 0 & 0 & 1 \end{bmatrix}.$$

This leads to

$$\boldsymbol{T}^{-1} = \begin{bmatrix} 1 & 0 & \overline{m}_1 \\ 0 & 1 & \overline{m}_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

and further

$$\boldsymbol{T}^{-1}\boldsymbol{T}^{-T} = \begin{bmatrix} 1 & 0 & \overline{m}_1 \\ 0 & 1 & \overline{m}_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s^2 & 0 & 0 \\ 0 & s^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \overline{m}_1 & \overline{m}_2 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} s^2 + \overline{m}_1^2 & \overline{m}_1\overline{m}_2 & \overline{m}_1 \\ \overline{m}_2\overline{m}_1 & s^2 + \overline{m}_2^2 & \overline{m}_2 \\ \overline{m}_1 & \overline{m}_2 & 1 \end{bmatrix}.$$

The latter identity can succinctly be written as

$$\boldsymbol{T}^{-1}\boldsymbol{T}^{-T} = s^2\boldsymbol{I}^* + \overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}^T,$$

where $\boldsymbol{I}^* = \mathrm{diag}(1,1,0)$. Analogously,

$$\boldsymbol{T}'^{-1}\boldsymbol{T}'^{-T} = s'^2\boldsymbol{I}^* + \overline{\boldsymbol{m}}'\,\overline{\boldsymbol{m}}'^T.$$

Introducing

$$\boldsymbol{C}(s,s',\overline{\boldsymbol{m}},\overline{\boldsymbol{m}}')$$
$$= (s'^2\boldsymbol{I}^* + \overline{\boldsymbol{m}}'\,\overline{\boldsymbol{m}}'^T) \otimes (s^2\boldsymbol{I}^* + \overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}^T) \quad (19)$$

and taking into account (15), we finally obtain

$$\boldsymbol{C} = \boldsymbol{C}(s,s',\overline{\boldsymbol{m}},\overline{\boldsymbol{m}}'). \quad (20)$$

VII. STATISTICAL JUSTIFICATION

To substantiate the normalized eight-point algorithm, we shift the focus from matrices involved in the computation of estimates (which may be well or ill-conditioned) to cost functions. It is not *a priori* clear why J_{NALS} should be preferable to J_{ALS}. We now present some explanation based on a statistical argument. Our reasoning will also provide the promised justification of the label "normalized" for the terms forming J_{NALS}.

For each $i = 1, \ldots, n$, let r_i be the ith *residual* defined as

$$r_i = \boldsymbol{m}_i'^T \boldsymbol{F} \boldsymbol{m}_i,$$

with $\boldsymbol{F}$ normalized for convenience. It is a fundamental observation that different residuals may carry different statistical weight. When $\boldsymbol{m}_i$ and $\boldsymbol{m}_i'$ are treated as sample values of independent multivariate random variables, the r_i are sample values of (typically) a *heteroscedastic* set of random variables, i.e., with member variables having different variances. The larger the variance of a particular r_i, the less reliable this residual is likely to be, and the more it should be devalued. Therefore, to account for heteroscedasticity, it is natural to replace the simple cost function $\sum_{i=1}^n r_i^2$, effectively J_{ALS}, by the more complicated cost function $\sum_{i=1}^n r_i^2/\mathrm{var}\,[r_i]$, where $\mathrm{var}\,[r]$ denotes the variance of r. The latter function is closer in form to a natural cost function derivable from the principle of maximum likelihood (cf. [10], [17]). We show that under a certain statistical model of data distribution, $\sum_{i=1}^n r_i^2/\mathrm{var}\,[r_i]$ can be identified with J_{NALS}.

Assume that, for each $i = 1, \ldots, n$, the observed location $\boldsymbol{m}_i$ is a realization of a random variable $\mathbf{m}_i = \overline{\boldsymbol{m}} + \Delta\mathbf{m}_i$, where $\overline{\boldsymbol{m}} = [\overline{m}_1, \overline{m}_2, 1]^T$ is a fixed, nonrandom location and $\Delta\mathbf{m}_i = [\Delta\mathsf{m}_{1,i}, \Delta\mathsf{m}_{2,i}, 0]^T$ is a random perturbation. Likewise, assume that $\boldsymbol{m}_i'$ is a realization of a random variable $\mathbf{m}_i' = \overline{\boldsymbol{m}}' + \Delta\mathbf{m}_i'$ with nonrandom $\overline{\boldsymbol{m}}' = [\overline{m}_1', \overline{m}_2', 1]^T$

and random $\Delta\mathbf{m}_i' = [\Delta\mathsf{m}_{1,i}', \Delta\mathsf{m}_{2,i}', 0]^T$. Suppose that the following conditions hold:

- $\Delta\mathbf{m}_i, \Delta\mathbf{m}_j'$ $(i,j = 1, \ldots, n)$ are independent;
- $\mathsf{E}\,[\Delta\mathbf{m}_i] = \mathsf{E}\,[\Delta\mathbf{m}_i'] = \mathbf{0}$ for each $i = 1, \ldots, n$;
- there exist $\sigma > 0$ and $\sigma' > 0$ such that

$$\mathsf{E}\,[\Delta\mathbf{m}_i(\Delta\mathbf{m}_i)^T] = \sigma^2\boldsymbol{I}^*,$$
$$\mathsf{E}\,[\Delta\mathbf{m}_i'(\Delta\mathbf{m}_i')^T] = \sigma'^2\boldsymbol{I}^* \quad (21)$$

for each $i = 1, \ldots, n$.

Here, E denotes expectation. Note that, effectively, *all* the $\mathbf{m}_i$ have common mean value $\overline{\boldsymbol{m}}$ and *all* the $\mathbf{m}_i'$ have common mean value $\overline{\boldsymbol{m}}'$. It is helpful to view $\overline{\boldsymbol{m}}$ and $\overline{\boldsymbol{m}}'$ as the centroids of some individual "true" nonrandom locations $\overline{\boldsymbol{m}}_i$ and $\overline{\boldsymbol{m}}_i'$ that are not explicitly introduced, but are present in the background. An immediate upshot of this type of modeling is that the random perturbations $\Delta\mathbf{m}_i$ and $\Delta\mathbf{m}_i'$ cannot be regarded as small in typical situations—the magnitude of $\Delta\mathbf{m}_i$ and $\Delta\mathbf{m}_i'$ has to be big enough to account for the disparity between $\overline{\boldsymbol{m}}$ and the $\overline{\boldsymbol{m}}_i$, and $\overline{\boldsymbol{m}}'$ and the $\overline{\boldsymbol{m}}_i'$.

Let $\mathbf{r}_i = \mathbf{m}_i'^T \boldsymbol{F} \mathbf{m}_i$ be the stochastic version of the ith residual. We calculate the variances

$$\mathrm{var}\,[\mathbf{r}_i] = \mathsf{E}\,[(\mathbf{m}_i'^T \boldsymbol{F} \mathbf{m}_i)^2] - (\mathsf{E}\,[\mathbf{m}_i'^T \boldsymbol{F} \mathbf{m}_i])^2$$

within our model. In so doing, we exploit an additional assumption that $\overline{\boldsymbol{m}}$ and $\overline{\boldsymbol{m}}'$ are "true" locations bound by $\boldsymbol{F}$, in the sense that

$$\overline{\boldsymbol{m}}'^T \boldsymbol{F} \overline{\boldsymbol{m}} = \boldsymbol{\theta}^T \mathrm{vec}(\overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}'^T) = 0. \quad (22)$$

Writing

$$\mathsf{E}\,[\mathbf{m}_i\mathbf{m}_i'^T] = \overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}'^T + \mathsf{E}\,[\Delta\mathbf{m}_i]\,\overline{\boldsymbol{m}}_i'^T$$
$$+ \mathbf{m}_i\mathsf{E}\,[\Delta(\mathbf{m}_i')^T] + \mathsf{E}\,[\Delta\mathbf{m}_i(\Delta\mathbf{m}_i')^T],$$

noting that

$$\mathsf{E}\,[\Delta\mathbf{m}_i]\,\overline{\boldsymbol{m}}_i'^T = \mathbf{m}_i\mathsf{E}\,[(\Delta\mathbf{m}_i')^T] = \mathbf{0}$$

and that

$$\mathsf{E}\,[\Delta\mathbf{m}_i(\Delta\mathbf{m}_i')^T] = \mathsf{E}\,[\Delta\mathbf{m}_i]\,\mathsf{E}\,[(\Delta\mathbf{m}_i')^T] = \mathbf{0}, \quad (23)$$

which involves the independence of $\Delta\mathbf{m}_i$ and $\Delta\mathbf{m}_i'$, we find that

$$\mathsf{E}\,[\mathbf{m}_i\mathbf{m}_i'^T] = \overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}'^T. \quad (24)$$

Hence,

$$\mathsf{E}\,[\mathbf{m}_i'^T \boldsymbol{F} \mathbf{m}_i] = \boldsymbol{\theta}^T \mathrm{vec}(\mathsf{E}\,[\mathbf{m}_i\mathbf{m}_i'^T]) = \boldsymbol{\theta}^T \mathrm{vec}(\overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}'^T).$$

This jointly with (22) implies that

$$\mathsf{E}\,[\mathbf{m}_i'^T \boldsymbol{F} \mathbf{m}_i] = 0. \quad (25)$$

Now, in view of (4) and (18),

$$\mathsf{E}\,[(\mathbf{m}_i'^T \boldsymbol{F} \mathbf{m}_i)^2] = \boldsymbol{\theta}^T (\mathsf{E}\,[\mathbf{m}_i'\mathbf{m}_i'^T \otimes \mathbf{m}_i\mathbf{m}_i^T])\boldsymbol{\theta}.$$

By the independence of $\Delta\mathbf{m}_i$ and $\Delta\mathbf{m}_i'$,

$$\mathsf{E}\,[\mathbf{m}_i'\mathbf{m}_i'^T \otimes \mathbf{m}_i\mathbf{m}_i^T] = \mathsf{E}\,[\mathbf{m}_i'\mathbf{m}_i'^T] \otimes \mathsf{E}\,[\mathbf{m}_i\mathbf{m}_i^T].$$

Furthermore, a calculation similar to that leading to (24), in which (21) is used instead of (23), shows that

$$\mathsf{E}\left[\mathbf{m}_i\mathbf{m}_i^T\right] = \sigma^2 \boldsymbol{I}^* + \overline{\mathbf{m}}\,\overline{\mathbf{m}}^T,$$
$$\mathsf{E}\left[\mathbf{m}_i'\mathbf{m}_i'^T\right] = \sigma'^2 \boldsymbol{I}^* + \overline{\mathbf{m}}'\overline{\mathbf{m}}'^T.$$

Recalling the definition (19), we obtain

$$\mathsf{E}\left[\mathbf{m}_i\mathbf{m}_i^T \otimes \mathbf{m}_i'\mathbf{m}_i'^T\right] = \boldsymbol{C}(\sigma,\sigma',\overline{\mathbf{m}},\overline{\mathbf{m}}').$$

Consequently,

$$\mathsf{E}\left[(\mathbf{m}_i'^T\boldsymbol{F}\mathbf{m}_i)^2\right] = \boldsymbol{\theta}^T\boldsymbol{C}(\sigma,\sigma',\overline{\mathbf{m}},\overline{\mathbf{m}}')\boldsymbol{\theta},$$

which together with (25) yields

$$\mathrm{var}\left[\mathbf{r}_i\right] = \boldsymbol{\theta}^T\boldsymbol{C}(\sigma,\sigma',\overline{\mathbf{m}},\overline{\mathbf{m}}')\boldsymbol{\theta}.$$

This final formula says, remarkably, that all the residuals $\mathbf{r}_i$ have common variance $v = \boldsymbol{\theta}^T\boldsymbol{C}(\sigma,\sigma',\overline{\mathbf{m}},\overline{\mathbf{m}}')\boldsymbol{\theta}$. Thus $\sum_{i=1}^n \mathbf{r}_i^2/\mathrm{var}\left[\mathbf{r}_i\right]$, the random version of the cost function introduced earlier, can simply be written as $v^{-1}\sum_{i=1}^n \mathbf{r}_i^2$ with v^{-1} a common normalization of all the residuals. Treating (8) and (9) as formulae for *estimates* of the parameters $\overline{\mathbf{m}}$, $\overline{\mathbf{m}}'$, σ, σ' used in our statistical model, replacing $\boldsymbol{C}(\sigma,\sigma',\overline{\mathbf{m}},\overline{\mathbf{m}}')$ with $\boldsymbol{C}$ given by (20), and replacing the random residuals $\mathbf{r}_i$ with the nonrandom ones r_i, we arrive at the expression $(\boldsymbol{\theta}^T\boldsymbol{C}\boldsymbol{\theta})^{-1}\sum_{i=1}^n r_i^2$ which, in view of (4), (6), and (16), is identical with J_{NALS}. In this way, J_{NALS} is finally justified and its building blocks, the "algebraic least squares" $(\boldsymbol{\theta}^T\boldsymbol{C}\boldsymbol{\theta})^{-1}r_i^2$, are found to be appropriately normalized.

VIII. Related Work

A different approach to the validation of the eight-point algorithm was earlier proposed by Mühlich and Mester [8]. It rests upon a statistical description of a *small* stochastic perturbation of a symmetric matrix that does not change the mean value of the eigenvector associated with the minimum eigenvalue. Applied to a measurement error model whereby small noise is present in *one* image only, this result leads to the development of a technique akin to Hartley's method. The conceptual framework for Mühlich and Mester's work is different from that underpinning the present paper in that different noise models are used (in our model noise appears in both images and is not necessarily small), and that different statistical principles are invoked: Mühlich and Mester's idea is to search for an estimator that is unbiased to first order, whereas the idea underlying our work is to search for an estimator more resembling the maximum likelihood estimator.

Torr [7] proposed a simple modification of the nonnormalized eight-point algorithm, producing in practice improvements similar to those of Hartley's method. It is essentially the ALS method with the matrix $\boldsymbol{A}$ replaced by the matrix $\boldsymbol{A}_\zeta = \sum_{i=1}^n \boldsymbol{u}_\zeta(\boldsymbol{x}_i)\boldsymbol{u}_\zeta(\boldsymbol{x}_i)^T$, where $\boldsymbol{u}_\zeta(\boldsymbol{x}) = \mathrm{vec}(\boldsymbol{m}_\zeta\boldsymbol{m}_\zeta'^T)$, $\boldsymbol{m}_\zeta = [m_1, m_2, \zeta]^T$, with ζ the average of the coordinates of the images' centers (in pixels). Here, the third co-ordinate ζ serves to level the varying order of magnitude of the entries of the $\boldsymbol{u}(\boldsymbol{x}_i)$ and, hence, to improve the conditioning of $\boldsymbol{A}$.

Yet another variation on the eight-point algorithm has recently been advanced by Torr and Fitzgibbon [9]. The estimator developed by these authors has the property that

if a coordinate system is subjected to a Euclidean transformation in one or both of the two underlying images, then the fundamental matrix estimate obtained from the transformed corresponding points is exactly the result of an application of the same transformation(s) to the estimate based on the original corresponding points. The estimate produced by this method turns out to be the solution of the familiar generalized eigenvalue problem (17) corresponding to the smallest eigenvalue, with $\boldsymbol{C}$ taken to be $\boldsymbol{I}^* \otimes \boldsymbol{I}^*$. Interestingly, since

$$(ss')^{-2}\boldsymbol{C}(s,s',\overline{\boldsymbol{m}},\overline{\boldsymbol{m}}')$$
$$= (\boldsymbol{I}^* + s'^{-2}\overline{\boldsymbol{m}}'\,\overline{\boldsymbol{m}}'^T) \otimes (\boldsymbol{I}^* + s^{-2}\overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}^T)$$

converges to $\boldsymbol{I}^* \otimes \boldsymbol{I}^*$ as $\|\overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}^T\|/s^2 \to \infty$ and $\|\overline{\boldsymbol{m}}'\,\overline{\boldsymbol{m}}'^T\|/s'^2 \to \infty$ and since the eigenvector defined in (17) is not affected by multiplication of $\boldsymbol{C}$ by a scalar factor, it follows that the Torr–Fitzgibbon technique can be viewed as a limit case of the NALS method for some image data configurations. The outcomes of the two methods become increasingly indiscernible whenever $\|\overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}^T\| \ll s^2$ and $\|\overline{\boldsymbol{m}}'\,\overline{\boldsymbol{m}}'^T\| \ll s'^2$, which happens, for example, when the centroids of the image points are close to the origins of the respective image frames (so that $\|\overline{\boldsymbol{m}}\,\overline{\boldsymbol{m}}^T\| \sim 1$ and $\|\overline{\boldsymbol{m}}'\,\overline{\boldsymbol{m}}'^T\| \sim 1$) and the spread of the points around the centroids is significant (so that $s^2 \gg 1$ and $s'^2 \gg 1$).

IX. Experimental Results

To assess whether the theoretical identity $\widehat{\boldsymbol{F}}_{\mathrm{HRT}} = \widehat{\boldsymbol{F}}_{\mathrm{NALS}}$ holds in practice, a series of simulations were run using synthetic data. The simulations were based on a set of "true" pairs of corresponding points generated by selecting a realistic stereo camera configuration, randomly choosing many 3D points, and projecting the 3D points onto two image planes. Image resolution was chosen to be 1000×1000 pixels.

Two tests were conducted, each comprising $10,000$ trials. At each trial:

- the "true" corresponding points were perturbed by homogeneous Gaussian jitter to produce noisy points;
- three fundamental matrices were generated from the noisy corresponding points using the nonnormalized algebraic least-squares method (ALS), the normalized algebraic least-squares method (NALS), and Hartley's method (HRT);
- and the estimates were compared in the way described below.

The standard deviation of the noise was fixed at $\sigma = 1.0$ pixels (tests run with other levels of noise produced similar results).

In our experiments, the ALS estimate was computed by performing SVD of $\boldsymbol{M}$ given in (7) and taking the minimum right singular vector. Similarly, the Hartley estimate was computed by performing SVD of the matrix $\tilde{\boldsymbol{M}} = [\boldsymbol{u}(\tilde{\boldsymbol{x}}_1), \ldots, \boldsymbol{u}(\tilde{\boldsymbol{x}}_n)]^T$ and applying the back transformation prescribed by (11) to the minimum right singular vector (a standard SVD-correction step to produce a usable, rank-2 fundamental matrix before back-transforming was ignored). The NALS estimate was computed by employing Leedan–Meer's method based on the GSVD of $(\boldsymbol{M}, \boldsymbol{N})$, with $\boldsymbol{M}$
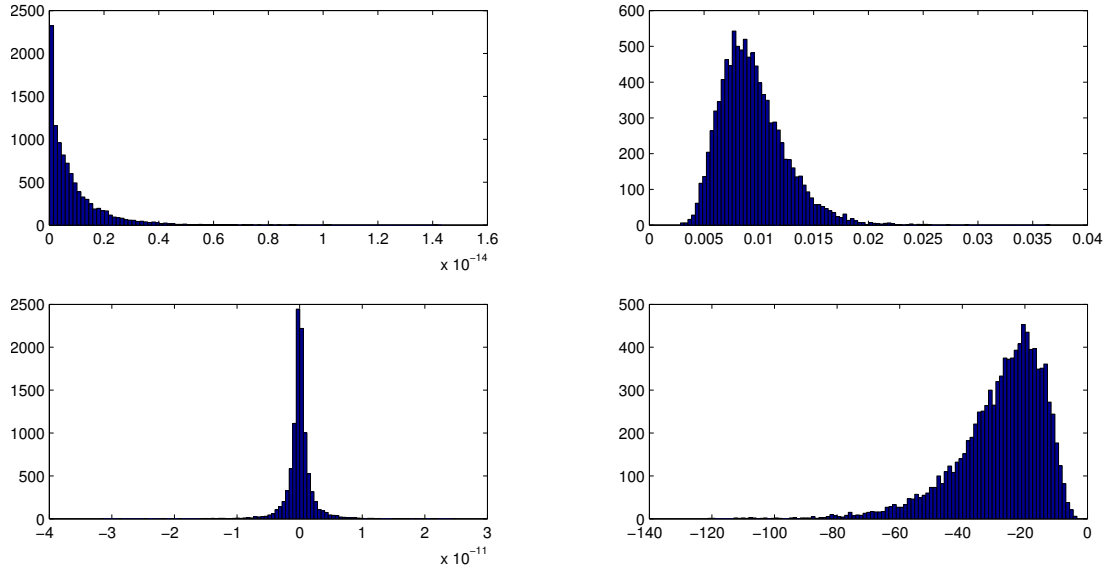
Fig. 1. Histograms of values of d_1 (upper left), d_2 (upper right), d_3 (lower left), and d_4 (lower right) for $10,000$ tests.

given by (7) and $\boldsymbol{N} = (s'\boldsymbol{I}^* + e\overline{\boldsymbol{m}}'^T) \otimes (s\boldsymbol{I}^* + e\overline{\boldsymbol{m}}^T)$, where $\boldsymbol{e} = [0, 0, 1]^T$.

In the first test, comparison of the estimates involved calculating two distances $d_1 = \min\|\widehat{\boldsymbol{F}}_{\mathrm{HRT}} \pm \widehat{\boldsymbol{F}}_{\mathrm{NALS}}\|_F$ and $d_2 = \min\|\widehat{\boldsymbol{F}}_{\mathrm{HRT}} \pm \widehat{\boldsymbol{F}}_{\mathrm{ALS}}\|_F$, with $\widehat{\boldsymbol{F}}_{\mathrm{HRT}}$, $\widehat{\boldsymbol{F}}_{\mathrm{NALS}}$, and $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$ having unit Frobenius norm. The first of these measures quantifies the discrepancy between the HRT and NALS estimates, the second informally gauges the significance of the values of the first. All results are plotted in Fig. 1. The histogram of d_1 values shows that $\widehat{\boldsymbol{F}}_{\mathrm{HRT}}$ and $\widehat{\boldsymbol{F}}_{\mathrm{NALS}}$ are almost identical, with all values of d_1 less than 1.5×10^{-14}. The significance of this may be gauged by noting that the d_2 histogram, capturing differences between the HRT and ALS estimates, exhibits values that are all greater than 1.5×10^{-3}.

The second test involved calculating the signed distances $d_3 = J_{\mathrm{AML}}(\widehat{\boldsymbol{F}}_{\mathrm{HRT}}) - J_{\mathrm{AML}}(\widehat{\boldsymbol{F}}_{\mathrm{NALS}})$ and $d_4 = J_{\mathrm{AML}}(\widehat{\boldsymbol{F}}_{\mathrm{HRT}}) - J_{\mathrm{AML}}(\widehat{\boldsymbol{F}}_{\mathrm{ALS}})$, where

$$J_{\mathrm{AML}}(\boldsymbol{F}) = \sum_{i=1}^n \frac{(\boldsymbol{m}_i'^T \boldsymbol{F} \boldsymbol{m}_i)^2}{\boldsymbol{m}_i^T \boldsymbol{F} \boldsymbol{I}^* \boldsymbol{F}^T \boldsymbol{m}_i + \boldsymbol{m}_i'^T \boldsymbol{F}^T \boldsymbol{I}^* \boldsymbol{F} \boldsymbol{m}_i'}$$

is the approximated maximum likelihood cost function commonly underlying more sophisticated iterative methods, associated with the default covariance $\boldsymbol{I}^*$ (e.g., see [10], [12], [17]). The d_3 histogram exhibits extremely small values centered on zero, confirming once again the practical equivalence of estimates $\widehat{\boldsymbol{F}}_{\mathrm{HRT}}$ and $\widehat{\boldsymbol{F}}_{\mathrm{NALS}}$. In contrast, the d_4 histogram shows differences in $\widehat{\boldsymbol{F}}_{\mathrm{HRT}}$ and $\widehat{\boldsymbol{F}}_{\mathrm{ALS}}$ that are very much larger.

X. CONCLUSION

A novel explanation has been presented for the improvement in performance of the normalized eight-point algorithm that results from using normalized data. It relies upon identifying a cost function that the algorithm effectively seeks to minimize. The advantageous character of the cost function is justified within a certain statistical model. The explanation avoids making any direct appeal to problem conditioning. Experimental results are presented that support the proposed approach.

REFERENCES

[1] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, no. 10, pp. 133–135, Sept. 1981.
[2] O. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig?" in *Computer Vision—ECCV '92, Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19–22, 1992*, ser. Lecture Notes in Computer Science, G. Sandini, Ed., vol. 588. Berlin: Springer-Verlag, 1992, pp. 563–578.
[3] R. I. Hartley, "Estimation of relative camera positions for uncalibrated cameras," in *Computer Vision—ECCV '92, Second European Conference on Computer Vision, Santa Margherita Ligure, Italy, May 19–22, 1992*, ser. Lecture Notes in Computer Science, G. Sandini, Ed., vol. 588. Berlin: Springer-Verlag, 1992, pp. 579–587.
[4] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
[5] O. Faugeras, Q.-T. Luong, and T. Papadopoulo, *The Geometry of Multiple Images*. Cambridge, Mass.: MIT Press, 2001.
[6] R. Hartley, "In defense of the eight-point algorithm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 6, pp. 580–593, 1997.
[7] P. H. S. Torr, "Outlier detection and motion segmentation," Ph.D. dissertation, Department of Engineering Science, University of Oxford, 1995.
[8] M. Mühlich and R. Mester, "The role of total least squares in motion analysis," in *Computer Vision—ECCV'98, Fifth European Conference on Computer Vision, Freiburg, Germany, June 2–6, 1998*, ser. Lecture Notes in Computer Science, H. Burkhardt and B. Neumann, Eds., vol. 1407. Berlin: Springer-Verlag, 1998, pp. 305–321.
[9] P. H. S. Torr and A. W. Fitzgibbon, "Invariant fitting of two view geometry or "In defiance of the 8 point algorithm"," Microsoft Research, Tech. Rep. MSR-TR-2002-50, May 2002.
[10] 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.*, vol. 22, no. 11, pp. 1294–1303, 2000.
[11] W. Chojnacki, M. J. Brooks, and A. van den Hengel, "Rationalising the renormalisation method of Kanatani," *J. Math. Imaging Vision*, vol. 14, no. 1, pp. 21–38, 2001.

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

[13] Y. Leedan and P. Meer, "Heteroscedastic regression in computer vision: problems with bilinear constraint," *Int. J. Comput. Vision*, vol. 37, no. 2, pp. 127–150, 2000.

[14] W. Chojnacki, M. J. Brooks, A. van den Hengel, and D. Gawley, "A new approach to constrained parameter estimation applicable to some computer vision problems," in *Proceedings of the Statistical Methods in Video Processing Workshop, Copenhagen, Denmark, June 1–2, 2002*, D. Suter, Ed. Clayton, Australia: Department of Electrical and Computer Systems Engineering, Monash University, 2002, pp. 43–48.

[15] A. van den Hengel, W. Chojnacki, M. J. Brooks, and D. Gawley, "A new constrained parameter estimator: experiments in fundamental matrix computation," in *Proceedings of the 13th British Machine Vision Conference, Cardiff, England, 2-5 September, 2002*, P. L. Rosin and D. Marshall, Eds., vol. 2. BMVA Press, 2002, pp. 468–476.

[16] H. Lütkepol, *Handbook of Matrices*. Chicester: John Wiley & Sons, 1996.

[17] Z. Zhang, "On the optimization criteria used in two-view motion analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 717–729, 1998.