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 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 perspective 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 \(m = [m_1, m_2, 1]^T\). 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 \((m, m')\), this pair satisfies the epipolar constraint

\[ m'^T F m = 0, \]

where \(F = [f_{ij}]\) 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), \(F\) is subject to the singularity constraint (or, equivalently, the rank-2 constraint)

\[ \det F = 0. \]

Using \(x = [m_1, m_2, m_1', m_2']^T\) as a compact descriptor of the single image datum \((m, m')\), the estimation problem associated with (1) and (2) can be stated as follows: Given a collection \(\{x_1, \ldots, x_n\}\) of image data and a meaningful cost function that characterizes the extent to which any particular \(F\) fails to satisfy the system of the copies of equation (1) associated with \(x = x_i (i = 1, \ldots, n)\), find an estimate \(\hat{F} \neq 0\) satisfying (2) for which the cost function attains its minimum. Since (1) and (2) do not change when \(F\) is multiplied by a nonzero scalar, \(\hat{F}\) is to be found only up to scale. If the singularity constraint is set aside, then the estimate associated
with a particular cost function \( J = J(F; x_1, \ldots, x_n) \) is defined as the unconstrained minimizer \( \hat{F} \) of \( J \),

\[
\hat{F} = \arg\min_{F \in \mathbb{R}^{mn}} J(F; x_1, \ldots, x_n).
\]

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

\[
J_{\text{ALS}}(F; x_1, \ldots, x_n) = \| F \|_F^2 - \sum_{i=1}^{n} (m_i^T F m_i)^2
\]

with \( \| F \|_F = \left( \sum_{i,j} f_{ij}^2 \right)^{1/2} \) the Frobenius norm of \( F \). Here, \( m_i^T F m_i \) is the signed algebraic distance between the individual datum \( x_i \) and the candidate matrix \( F \). The algebraic least squares (ALS) estimate, \( \hat{F}_{\text{ALS}} \), is defined as the minimizer of \( J_{\text{ALS}} \).

A practical means for finding \( \hat{F}_{\text{ALS}} \) is conveniently derived based on an alternative expression for \( m_i^T F m_i \). For an \( m \times n \) matrix \( A = [a_1, \ldots, a_n] \) with \( a_j \) the \( j \)th column vector of length \( m \), denote by \( \text{vec}(A) \) the vectorization of \( A \), i.e., the column vector of length \( mn \) defined by \( \text{vec}(A) = [a_1^T, \ldots, a_n^T]^T \). Let \( \theta = \text{vec}(F^T) \) and \( u(x) = \text{vec}(m m^T) \). Then, by applying standard rules for matrix operations [16], we find that

\[
m_i^T F m_i = \text{tr}(m_i^T F m_i) = \text{tr}(m m^T F) = \text{vec}(F^T)^T \text{vec}(m m^T) = \theta^T u(x),
\]

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

\[
J_{\text{ALS}}(\theta; x_1, \ldots, x_n) = \| \theta \|^{-2} \theta^T A \theta,
\]

where

\[
A = \sum_{i=1}^{n} u(x_i) u(x_i)^T
\]

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

\[
M = [u(x_1), \ldots, u(x_n)]^T
\]

and taking for the desired output the right singular vector of \( M \) 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 \( \hat{\theta}_{\text{ALS}} \) to be uniquely determined, the null space of \( A \) cannot be more than one-dimensional. Given that \( A \) is the sum of the rank-1 matrices \( u(x_i) u(x_i)^T \), this is secured, generically, by the condition \( n \geq 8 \), hence the qualification “eight-point”.

IV. HARTLEY’S APPROACH

Let \( \overline{m} \) and \( \overline{m}' \) be the centroids, or “centers of mass”, of the \( m_i \) and the \( m_i' \), respectively, defined by

\[
\overline{m} = \frac{1}{n} \sum_{i=1}^{n} m_i \quad \text{and} \quad \overline{m}' = \frac{1}{n} \sum_{i=1}^{n} m_i'.
\]  

Let \( \overline{m} = [m_1, m_2, \ldots, m_n]^T \), \( \overline{m}' = [m_1', m_2', \ldots, m_n']^T \), \( m_i = [m_{i1}, m_{i2}, \ldots, m_{in}]^T \), \( m_i' = [m'_{i1}, m'_{i2}, \ldots, m'_{in}]^T \). Let

\[
s = \left( \frac{1}{2n} \sum_{i=1}^{n} \| m_i - \overline{m} \|^2 \right)^{1/2} = \left( \frac{1}{2n} \sum_{i=1}^{n} (m_{i1} - m_{11})^2 + (m_{i2} - m_{12})^2 \right)^{1/2},
\]

\[
s' = \left( \frac{1}{2n} \sum_{i=1}^{n} \| m_i' - \overline{m}' \|^2 \right)^{1/2} = \left( \frac{1}{2n} \sum_{i=1}^{n} (m'_{i1} - m'_{11})^2 + (m'_{i2} - m'_{12})^2 \right)^{1/2}.
\]

Define the normalized data as

\[
\tilde{m}_i = [(m_{i1} - m_{11})/s, (m_{i2} - m_{12})/s, 1]^T,
\]

\[
\tilde{m}_i' = [(m'_{i1} - m'_{11})/s', (m'_{i2} - m'_{12})/s', 1]^T.
\]

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

\[
T = \begin{bmatrix}
s^{-1} & 0 & -s^{-1} \tilde{m}_1 \\
0 & s^{-1} & -s^{-1} \tilde{m}_2 \\
0 & 0 & 1
\end{bmatrix},
\]

\[
T' = \begin{bmatrix}
s'^{-1} & 0 & -s'^{-1} \tilde{m}'_1 \\
0 & s'^{-1} & -s'^{-1} \tilde{m}'_2 \\
0 & 0 & 1
\end{bmatrix}.
\]

Let \( \hat{x}_i = [\tilde{m}_{i1}, \tilde{m}_{i2}, \hat{m}_{i1}', \hat{m}_{i2}', \tilde{m}_i, \tilde{m}_i']^T \). Denote by \( \hat{F}_{\text{ALS}} \) the minimizer of the ALS cost function seeded with the normalized data, that is, the minimizer of the function \( F \mapsto J_{\text{ALS}}(F; \hat{x}_1, \ldots, \hat{x}_n) \). Let \( F \mapsto \hat{F} \) be the mapping defined by

\[
\hat{F} = T^{-T} F T^{-1}.
\]

Clearly, if \( \hat{m} = T m \) and \( \hat{m}' = T' m' \), then \( m^T F m = m'^T \hat{F} \hat{m} \). Accordingly, the image of \( \hat{F}_{\text{ALS}} \) by the inverse mapping \( F \mapsto \hat{F} \) can be viewed as a genuine estimate of
we term this the Hartley (HRT) estimate of $F$ and write $\hat{F}_{\text{HRT}}$; it is explicitly given by

$$\hat{F}_{\text{HRT}} = T^T \hat{F}_{\text{ALS}} T. \quad (11)$$

The introduction of $\hat{F}_{\text{HRT}}$ 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 $\tilde{A} = \sum_{i=1}^n u(\tilde{x}_i)u(\tilde{x}_i)^T$ serving to calculate $\hat{F}_{\text{ALS}}$ is, in practice, much better conditioned (in the above sense) than the matrix $A$ with which $\hat{F}_{\text{ALS}}$ 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

$$J_{\text{NALS}}(F; x_1, \ldots, x_n) = \|F - T^TFT^{-1}\|_F^2 = \sum_{i=1}^n (m_i^T F m_i)^2. \quad (12)$$

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

$$\hat{F}_{\text{HRT}} = \hat{F}_{\text{NALS}}. \quad (13)$$

Indeed, if $\hat{F}$ and $F$ are related by (10), then $m_i^T F m_i = m_i^T \hat{F} m_i$ for each $i = 1, \ldots, n$, and so

$$J_{\text{NALS}}(F; x_1, \ldots, x_n) = \|\hat{F}\|_F^2 = \sum_{i=1}^n (m_i^T \hat{F} m_i)^2 = J_{\text{ALS}}(\hat{F}; \tilde{x}_1, \ldots, \tilde{x}_n).$$

From this identity it follows that the function $F \mapsto J_{\text{NALS}}(F; x_1, \ldots, x_n)$ attains a minimum at a particular $F$ precisely when the function $\hat{F} \mapsto J_{\text{ALS}}(\hat{F}; \tilde{x}_1, \ldots, \tilde{x}_n)$ attains a minimum at the image $\hat{F}$ of this $F$ by the mapping $F \mapsto \hat{F}$. In other words, $T^T \hat{F}_{\text{NALS}} T^{-1} = \hat{F}_{\text{ALS}}$, which, clearly, is equivalent to $\hat{F}_{\text{NALS}} = T^T \hat{F}_{\text{ALS}} T$. Comparing this with (11) establishes the claim.

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

$$\|T^{-T}F T^{-1}\|_F^2 = \theta^T C \theta,$$

where $C$ is the $9 \times 9$ matrix given by

$$C = T^{-T}T^{-T} \otimes T^{-1}T^{-T}. \quad (15)$$

Here, $\otimes$ stands for the Kronecker product, or tensor product, of matrices defined as follows: If $A = [a_{ij}]$ is an $m \times n$ matrix and $B$ is a $p \times q$ matrix, then $A \otimes B$ is the $mp \times nq$ matrix given by

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

To establish (14), note that

$$\|T^{-T}F T^{-1}\|_F^2 = \text{tr}(T^{-T}FT^{-1}T^{-T}F T^{-1}) = \text{tr}(T^{-T}T^{-1}T^{-T}FT^{-1}) = \text{tr}(T^{-T}T^{-T}T^{-1}F T^{-1})$$

$$= \text{vec}(F^T)(T^{-1}T^{-T} \otimes T^{-1}T^{-T})\text{vec}(F T^{-1}) = \text{vec}(F^T)(T^{-1}T^{-T} \otimes T^{-1}T^{-T})\text{vec}(F T^{-1}).$$

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

$$J_{\text{NALS}}(\theta; x_1, \ldots, x_n) = \frac{\theta^T A \theta}{\theta^T C \theta}. \quad (16)$$

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

$$A \theta = \lambda C \theta$$

corresponding to the smallest eigenvalue. Since $\lambda$ 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 $(M, N)$ satisfying $A = M^T M$ and $C = N^T N$. Numerical experiments show (see later) that when this method is applied, the matrices $A$ and $C$, of which the first is typically ill-conditioned, lead to a solution identical with the solution obtained using the well-conditioned matrix $\tilde{A}$—in other words, equality (13) is experimentally confirmed.

VI. TWO FORMULAE

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

$$u(x_i)u(x_i)^T = m_i^T m_i^T \otimes m_i m_i^T \quad (18)$$

resulting immediately from

$$\text{vec}(m_i^T m_i^T) = \text{vec}(m_i \otimes m_i)^T = (m_i \otimes m_i)(m_i^T \otimes m_i)^T = m_i^T m_i \otimes m_i m_i^T.$$ 

The second is derived starting from the observation that

$$T = \begin{bmatrix} s^{-1} & 0 & 0 & 1 & 0 & -m_1 \\ 0 & s^{-1} & 0 & 0 & 1 & -m_2 \end{bmatrix}.$$ 

This leads to

$$T^{-1} = \begin{bmatrix} 1 & 0 & m_1 \\ 0 & 1 & m_2 \\ 0 & 0 & 1 \end{bmatrix}.$$ 

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

$$J_{\text{NALS}}(\theta; x_1, \ldots, x_n) = \frac{\theta^T A \theta}{\theta^T C \theta}. \quad (16)$$

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

$$A \theta = \lambda C \theta$$

corresponding to the smallest eigenvalue. Since $\lambda$ 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 $(M, N)$ satisfying $A = M^T M$ and $C = N^T N$. Numerical experiments show (see later) that when this method is applied, the matrices $A$ and $C$, of which the first is typically ill-conditioned, lead to a solution identical with the solution obtained using the well-conditioned matrix $\tilde{A}$—in other words, equality (13) is experimentally confirmed.
and further
\[
T^{-1}T^T = \begin{bmatrix}
1 & 0 & m_1 \\
0 & 1 & m_2 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
s^2 & 0 & 0 \\
0 & s^2 & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
m_1 & m_2 & 1
\end{bmatrix}
\]
\[
= \begin{bmatrix}
s^2 + m_1^2 & m_1 m_2 & m_1 \\
m_2 m_1 & m_2^2 & m_1 \\
m_1 & m_2 & 1
\end{bmatrix}
\]

The latter identity can succinctly be written as
\[
T^{-1}T^T = s^2 I^* + \overline{m} \overline{m}^T,
\]
where $I^* = \text{diag}(1, 1, 0)$. Analogously,
\[
T'^{-1}T'^T = s'^2 I^* + \overline{m'} \overline{m'}^T.
\]

Introducing
\[
C(s, s', \overline{m}, \overline{m'}) = (s^2 I^* + \overline{m} \overline{m}^T) \otimes (s'^2 I^* + \overline{m'} \overline{m'}^T)
\]
and taking into account (15), we finally obtain
\[
C = C(s, s', \overline{m}, \overline{m'}).
\]

\[\text{VII. Statistical Justification}\]

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 $J_{\text{NALS}}$ should be preferable to $J_{\text{ALS}}$. 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 $J_{\text{NALS}}$.

For each $i = 1, \ldots, n$, let $r_i$ be the $i$th residual defined as
\[r_i = m_i^T F m_i,\]
with $F$ normalized for convenience. It is a fundamental observation that different residuals may carry different statistical weight. When $m_i$ and $m'_i$ 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 $J_{\text{ALS}}$, by the more complicated cost function $\sum_{i=1}^n r_i^2 / \text{var}[r_i]$, where $\text{var}[r]$ 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, $\sum_{i=1}^n r_i^2 / \text{var}[r_i]$ can be identified with $J_{\text{NALS}}$.

Assume that, for each $i = 1, \ldots, n$, the observed location $m_i$ is a realization of a random variable $m_i = \overline{m} + \Delta m_i$, where $\overline{m} = [m_1, m_2, 1]^T$ is a fixed, nonrandom location and $\Delta m_i = [\Delta m_{1, i}, \Delta m_{2, i}, 0]^T$ is a random perturbation. Likewise, assume that $m'_i$ is a realization of a random variable $m'_i = \overline{m'} + \Delta m'_i$ with nonrandom $\overline{m'} = [m'_1, m'_2, 1]^T$ and random $\Delta m'_i = [\Delta m'_{1, i}, \Delta m'_{2, i}, 0]^T$. Suppose that the following conditions hold:
\begin{itemize}
  \item $\Delta m_{ij}, \Delta m'_{ij}$ ($i, j = 1, \ldots, n$) are independent;
  \item $E[\Delta m_i] = E[\Delta m'_i] = 0$ for each $i = 1, \ldots, n$;
  \item there exist $\sigma > 0$ and $\sigma' > 0$ such that
    \[E[\Delta m_i (\Delta m'_i)^T] = \sigma^2 I,\]
    \[E[\Delta m'_i (\Delta m'_i)^T] = \sigma'^2 I\]
    \end{itemize}
for each $i = 1, \ldots, n$.

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

Let $r_i = m_i^T F m_i$ be the stochastic version of the $i$th residual. We calculate the variances
\[
\text{var}[r_i] = E[(m_i^T F m_i)^2] - (E[m_i^T F m_i])^2
\]
within our model. In so doing, we exploit an additional assumption that $\overline{m}$ and $\overline{m'}$ are “true” locations bound by $F$, in the sense that
\[
\overline{m'} F \overline{m} = \theta^T \text{vec}(\overline{m} \overline{m}^T) = 0.
\]

Writing
\[
E[m_i m'_i^T] = \overline{m} \overline{m}'^T + E[\Delta m_i] \overline{m'}^T
+ m_i E[\Delta (m'_i)^T] + E[\Delta m_i (\Delta m'_i)^T],
\]
noting that
\[
E[\Delta m_i] \overline{m'}^T = m_i E[(\Delta m_i)^T] = 0
\]
and that
\[
E[\Delta m_i (\Delta m'_i)^T] = E[\Delta m_i] E[(\Delta m'_i)^T] = 0,
\]
which involves the independence of $\Delta m_i$ and $\Delta m'_i$, we find that
\[
E[m_i m'_i^T] = \overline{m} \overline{m}'^T.
\]

Hence,
\[
E[m_i^T F m_i] = \theta^T \text{vec}(E[m_i m'_i^T]) = \theta^T \text{vec}(\overline{m} \overline{m}'^T).
\]

This jointly with (22) implies that
\[
E[m_i^T F m_i] = 0.
\]

Now, in view of (4) and (18),
\[
E[(m_i^T F m_i)^2] = \theta^T (E[m_i m'_i^T \otimes m_i m'_i^T]) \theta.
\]

By the independence of $\Delta m_i$ and $\Delta m'_i$,
\[
E[m_i m'_i^T \otimes m_i m'_i^T] = E[m_i m'_i^T] \otimes E[m_i m'_i^T].
\]
Furthermore, a calculation similar to that leading to (24), in which (21) is used instead of (23), shows that
\[ E[\mathbf{m}, \mathbf{m}^\top] = \sigma^2 I + \mathbf{m}\mathbf{m}^\top, \]
\[ E[\mathbf{m}'\mathbf{m}'^\top] = \sigma^2 I + \mathbf{m}'\mathbf{m}'^\top. \]
Recalling the definition (19), we obtain
\[ E[\mathbf{m}_{i}^T \otimes \mathbf{m}_{i}'^T] = C(\sigma, \sigma', \mathbf{m}, \mathbf{m}'). \]
Consequently,
\[ E[(\mathbf{m}'^T F \mathbf{m}_i)^2] = \theta^T C(\sigma, \sigma', \mathbf{m}, \mathbf{m}')\theta, \]
which together with (25) yields
\[ \text{var} [r_i] = \theta^T C(\sigma, \sigma', \mathbf{m}, \mathbf{m}')\theta. \]
This final formula says, remarkably, that all the residuals \( r_i \) have common variance \( v = \theta^T C(\sigma, \sigma', \mathbf{m}, \mathbf{m}')\theta \). Thus \( \sum_{i=1}^{n} r_i^2 / \text{var} [r_i] \), the random version of the cost function introduced earlier, can simply be written as \( v^{-1} \sum_{i=1}^{n} r_i^2 \) with \( v^{-1} \) a common normalization of all the residuals. Treating (8) and (9) as formulae for estimates of the parameters \( \mathbf{m}, \mathbf{m}' \), \( \sigma, \sigma' \) used in our statistical model, replacing \( C(\sigma, \sigma', \mathbf{m}, \mathbf{m}') \) with \( C(\mathbf{m}, \mathbf{m}') \) given by (20), and replacing the random residuals \( r_i \) with the nonrandom ones \( u_i \), we arrive at the expression
\[ (\theta^T C(\theta)^{-1} \sum_{i=1}^{n} u_i^2 \theta \]
which, in view of (4), (6), and (16), is identical with \( J_{\text{NALS}} \). In this way, \( J_{\text{NALS}} \) is finally justified and its building blocks, the \textit{algebraic least squares} \( (\theta^T C(\theta)^{-1} \sum_{i=1}^{n} u_i^2 \theta \), are found to be appropriately normalized.

**VIII. RELATED WORK**

A different approach to the validation of the eight-point algorithm was earlier proposed by Mühlisch and Mester [8]. It rests upon a statistical description of a \textit{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 \textit{one} image only, this result leads to the development of a technique akin to Hartley’s method.

The conceptual framework for Mühlisch 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: Mühlisch 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 \( A \) replaced by the matrix
\[ A_{c} = \sum_{i=1}^{n} \mathbf{u}_{c}(x_{i}) \mathbf{u}_{c}(x_{i})^T, \]
where \( \mathbf{u}_{c}(x) = \text{vec}(\mathbf{m}_{c} \mathbf{m}_{c}^T), \mathbf{m}_{c} = \mathbf{m}_{1} \mathbf{m}_{2} \mathbf{c}_T, \mathbf{c} \) the average of the coordinates of the images’ centers (in pixels). Here, the third co-ordinate \( \mathbf{c} \) serves to level the varying order of magnitude of the entries of \( u(x_{i}) \), and, hence, to improve the conditioning of \( A \).

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 \( C \) taken to be \( I^* \otimes I^* \). Interestingly, since
\[ (s')^{-2}C(s', s', \mathbf{m}, \mathbf{m}') \]
converges to \( I^* \otimes I^* \) as \( \|\mathbf{m}\mathbf{m}'^T\|/s^2 \to \infty \) and \( \|\mathbf{m}\mathbf{m}'^T\|/s'^2 \to \infty \) and since the eigenvector defined in (17) is not affected by multiplication of \( C \) 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 \( \|\mathbf{m}\mathbf{m}'^T\| \ll s^2 \) and \( \|\mathbf{m}\mathbf{m}'^T\| \ll s'^2 \), which happens, for example, when the centroids of the image points are close to the origins of the respective image frames (so that \( \|\mathbf{m}\mathbf{m}'^T\| \sim 1 \) and \( \|\mathbf{m}\mathbf{m}'^T\| \sim 1 \)) 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 \( \hat{F}_{\text{HRT}} = \hat{F}_{\text{NALS}} \) 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 \( 1000 \times 1000 \) 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 \( \mathbf{M} \) given in (7) and taking the minimum right singular vector. Similarly, the Hartley estimate was computed by performing SVD of the matrix \( \mathbf{M} = [\mathbf{u}(\mathbf{x}_{1}), \ldots, \mathbf{u}(\mathbf{x}_{n})]^T \) 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 \( (\mathbf{M}, \mathbf{N}) \), with \( \mathbf{M} \).
given by (7) and \( N = (s^T + e m^T) \otimes (s^T + e m^T) \), where \( e = [0, 0, 1]^T \).

In the first test, comparison of the estimates involved calculating two distances \( d_1 = \min \| \hat{F}_{\text{HRT}} + \hat{F}_{\text{NALS}} \|_F \) and \( d_2 = \min \| \hat{F}_{\text{HRT}} - \hat{F}_{\text{NALS}} \|_F \), with \( \hat{F}_{\text{HRT}}, \hat{F}_{\text{NALS}} \) and \( \hat{F}_{\text{ALS}} \) 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 \( \hat{F}_{\text{HRT}} \) and \( \hat{F}_{\text{ALS}} \) are almost identical, with all values of \( d_1 \) less than \( 1.5 \times 10^{-14} \). 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 \( 1.5 \times 10^{-5} \).

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

\[
J_{\text{AML}}(F) = \sum_{i=1}^{n} \frac{(m_i^T F m_i)^2}{m_i^T F^T F m_i + m_i^T F^T I^T F m_i},
\]

is the approximated maximum likelihood cost function commonly underlying more sophisticated iterative methods, associated with the default covariance \( I^* \) (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 \( \hat{F}_{\text{HRT}} \) and \( \hat{F}_{\text{NALS}} \). In contrast, the \( d_4 \) histogram shows differences in \( \hat{F}_{\text{HRT}} \) and \( \hat{F}_{\text{ALS}} \) 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.

ACKNOWLEDGEMENTS

The support of the Australian Research Council and helpful comments of the referees are gratefully acknowledged.

REFERENCES


