

A Versatile Method for Trifocal Tensor Estimation

Bogdan Matei
ECE Department

Bogdan Georgescu
CS Department

Peter Meer
ECE Department

Rutgers University, Piscataway, NJ 08854, USA

matei, georgesc, meer@caip.rutgers.edu

Abstract

Reliable estimation of the trifocal tensor is crucial for 3D reconstruction from uncalibrated cameras. The estimation process is based on minimizing the geometric distances between the measurements and the corrected data points, the underlying nonlinear optimization problem being most often solved with the Levenberg-Marquardt (LM) algorithm. We employ for this task the heteroscedastic errors-in-variables (HEIV) estimator and take into account both the singularity of the multivariate tensor constraint and the bifurcation which can appear for noisy data. In comparison to the Gold Standard method, the new approach is significantly faster while having the same performance, and it is less sensitive to initialization when the data is close to degenerate. Analytical expressions for the covariances of the parameter and corrected image point estimates are available for the HEIV estimator, and thus the confidence regions of the corrected measurements can be delineated in the images.

1. Trifocal Tensor

The trifocal tensor describes the intrinsic projective properties of a group of three images taken with uncalibrated cameras. The role of trifocal tensor in the projective reconstruction of 3D structures is extensively discussed in the literature, and we refer to the recent book [5, pp.355–378] for an excellent treatment of all the relevant topics, and to [16] for an comprehensive discussion of the involved optimization methods.

In this paper will focus on the problem of estimating the trifocal tensor from 3-view point correspondences, i.e., from the matched images of 3D points. We will assume that all the correspondences are correct. While the estimation method presented here can be easily robustified similar to [14], we concentrate on issues related to the behavior of the estimation process, i.e., on numerical robustness.

Will start by reviewing the geometric relations needed in the sequel. Given three cameras characterized by unknown projective matrices \mathbf{P} , \mathbf{P}' , \mathbf{P}'' , the images of a 3D point \mathbf{X} in each view will be denoted \mathbf{x} , \mathbf{x}' , \mathbf{x}'' . In homogeneous coordinates $\mathbf{x} = [x_1, x_2, x_3]^T$ and similarly for the points in the other two images. The projective ambiguity allows to express the camera matrices as $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$, $\mathbf{P}' = [\mathbf{A}|\mathbf{e}']$, $\mathbf{P}'' = [\mathbf{B}|\mathbf{e}'']$, where \mathbf{e}' and \mathbf{e}'' are the projection of the first camera center \mathbf{C} in the second and third image (Figure 1).

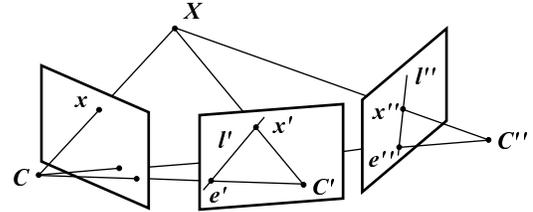


Figure 1: Point-point-point correspondence. Definition of the geometric elements used in the paper.

Thus \mathbf{e}' and \mathbf{e}'' are the epipoles corresponding to \mathbf{C} .

The $3 \times 3 \times 3$ trifocal tensor \mathbf{T} describes a 3D incidence relation through a trilinearity among related points (lines) in the three images. The trilinearities can be written for the point correspondences as

$$[\mathbf{x}']_{\times} \left(\sum_i x_i \mathbf{T}_i \right) [\mathbf{x}'']_{\times} = \mathbf{0}_{3 \times 3} \quad (1)$$

where $[\mathbf{v}]_{\times}$ is the skew-symmetric matrix such that $\mathbf{v} \times \mathbf{u} = [\mathbf{v}]_{\times} \mathbf{u}$, and the 3×3 matrices

$$\mathbf{T}_i = \mathbf{a}_i \mathbf{e}''^T - \mathbf{e}' \mathbf{b}_i^T \quad (2)$$

are the correlation slices [3] of the trifocal tensor. Note that \mathbf{T}_i depends only on the parameters of the projection matrices \mathbf{P}' , \mathbf{P}'' , and has rank two. It can be shown that only four of the nine relations captured in (1) are linearly independent [5, pp.417–421]. We will use those in the upper left 2×2 block in (1).

The epipolar lines in the second and third view of the image point \mathbf{x} in the first view will be defined as

$$\mathbf{l}' = [\mathbf{e}']_{\times} \mathbf{x}' \quad \mathbf{l}'' = [\mathbf{e}'']_{\times} \mathbf{x}'' . \quad (3)$$

Multiplying (2) left and right respectively with \mathbf{l}'^T and \mathbf{l}''^T yields due to the vanishing cross products

$$\mathbf{l}'^T \mathbf{T}_i \mathbf{l}'' = 0 . \quad (4)$$

The relation (4) captures the degeneracy of epipolar line transfer [5, pp.374–375]. Another important algebraic property of the trifocal tensor is that the linear combination $\sum_i x_i \mathbf{T}_i$ has also rank two. The left and right null-vectors are the epipolar lines \mathbf{l}' and \mathbf{l}'' [5, pp.363–364]

$$\mathbf{l}'^T \left(\sum_i x_i \mathbf{T}_i \right) = \mathbf{0}_3^T \quad \text{and} \quad \left(\sum_i x_i \mathbf{T}_i \right) \mathbf{l}'' = \mathbf{0}_3 . \quad (5)$$

2. Estimation Methods

From (1) it can be seen that the tensor is defined only up to a constant, i.e., there are 26 unknown parameters. From geometric considerations, however, it is easy to prove that the tensor can have only 18 degrees of freedom (d.f.). Indeed, the three cameras have $3 \times 11 = 33$ parameters from which we must subtract the inherent ambiguities introduced by the 3D projective transformation (15 d.f.).

The additional constraints which have to be satisfied by the 26 parameters in order to represent a trifocal tensor are complicated polynomial expressions. Papadopoulos and Faugeras [12] were the first to obtain 12 such (not independent) relations, while the minimal set of 8 conditions was recently derived by Canterakis [3]. The additional constraints being higher degree polynomials they can be imposed only after an estimate of the 26 parameters becomes available through unconstrained minimization. To take into account the underlying geometry the tensor (1) must be reparametrized. An often used reparametrization is by the 24 parameters of the projection matrices \mathbf{P}' , \mathbf{P}'' based on (2), but other possibilities were also investigated [14].

Without loss of generality we can consider that \mathbf{x} , \mathbf{x}' , \mathbf{x}'' are affine coordinates, i.e., $\mathbf{x} = [x_1, x_2, 1]^\top$, etc., where x_1, \dots, x_2'' are the measurements in the three images. Assuming independent normally distributed measurement errors, the optimal maximum likelihood estimate (MLE) is obtained by minimizing

$$\mathcal{J}_G = \sum_{j=1}^n d(\mathbf{x}_j, \hat{\mathbf{x}}_j)^2 + d(\mathbf{x}'_j, \hat{\mathbf{x}}'_j)^2 + d(\mathbf{x}''_j, \hat{\mathbf{x}}''_j)^2 \quad (6)$$

i.e., the sum of squared geometric distances between the measurements \mathbf{x}_j , \mathbf{x}'_j , \mathbf{x}''_j and the corrected data points $\hat{\mathbf{x}}_j$, $\hat{\mathbf{x}}'_j$, $\hat{\mathbf{x}}''_j$, the latter obeying the trilinear constraints (1) for the estimated tensor $\hat{\mathbf{T}}$. Note the distinction between the noisy measurements and the geometric elements (the true or equivalently the corrected data points). In Section 1 \mathbf{x} was a geometric element.

Since the corrected data points must be in correspondence, they are the projections on the three image planes of an unknown 3D point \mathbf{X} . Thus the corrected point coordinates $\hat{x}_1, \dots, \hat{x}_2''$ are rational functions of the camera parameters and the three spatial coordinates of \mathbf{X} . Solving the nonlinear estimation problem (6) with this parametrization always yields a geometrically valid tensor. This projective bundle adjustment procedure is called the ‘‘Gold Standard’’ method in [5, pp.385]. The minimization is performed using the Levenberg-Marquardt (LM) algorithm for the $3n + 24$ unknown parameters yielded by n measurements, making the estimation process very demanding computationally. It is possible to reduce the computations by implementing a sparse LM algorithm [5, pp.571–576], but in practice most

often a first order approximation of (6) is used. The approximation is called the Sampson distance [5, pp.387] since it is based on an old technique proposed for ellipse fitting [13].

Define the vector $\mathbf{m} = [x_1, x_2, x'_1, x'_2, x''_1, x''_2]^\top \in \mathbb{R}^6$. The four independent trilinear constraints from (1) can be rearranged as a homogeneous multivariate expression

$$\mathbf{f}(\hat{\mathbf{m}}, \hat{\boldsymbol{\theta}}) = \hat{\mathbf{Z}}\hat{\boldsymbol{\theta}} = \mathbf{0}_4 \quad (7)$$

linear in the 27-dimensional vector of the tensor parameters $\hat{\boldsymbol{\theta}}$. Let the 24-dimensional vector $\hat{\boldsymbol{\beta}}$ have as entities the elements of the estimated projection matrices $\hat{\mathbf{P}}'$ and $\hat{\mathbf{P}}''$. Then (2) can be rewritten as $\hat{\boldsymbol{\theta}} = \mathbf{g}(\hat{\boldsymbol{\beta}})$. The 4×27 matrix $\hat{\mathbf{Z}}$ has each element a product of some of the corrected point coordinates $\hat{x}_1, \dots, \hat{x}_2''$. Thus, $\hat{\mathbf{Z}} = \Phi(\hat{\mathbf{m}})$ and let $\varphi_k(\hat{\mathbf{m}})$ be the k -th row of $\Phi(\hat{\mathbf{m}})$. The constraint (7) can be also written as

$$\mathbf{f}(\hat{\mathbf{m}}, \hat{\boldsymbol{\beta}}) = \Phi(\hat{\mathbf{m}})\mathbf{g}(\hat{\boldsymbol{\beta}}) = \mathbf{0}_4. \quad (8)$$

The 6×4 Jacobian matrix of the constraint with respect to the noisy measurements \mathbf{m} is defined following [2]

$$\mathbf{J}_{f|m} = \frac{\partial \mathbf{f}(\mathbf{m}, \hat{\boldsymbol{\beta}})^\top}{\partial \mathbf{m}} \quad (\mathbf{J}_{f|m})_{ik} = \frac{\partial \varphi_k(\mathbf{m}, \hat{\boldsymbol{\beta}})}{\partial m_i}. \quad (9)$$

In the first order approximation of (6) the constraint (8) is linearized, as will be shown in Section 4. In [14] the linearization was performed around the measurements \mathbf{m}_j , however, we will show that linearizing around the corrected data points $\hat{\mathbf{m}}_j$ is more advantageous.

The expression of the Sampson distance is

$$\mathcal{J} = \sum_{j=1}^n \mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}})^\top \left(\mathbf{J}_{f|m_j}^\top \mathbf{J}_{f|m_j} \right)^+ \mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}}) \quad (10)$$

where \mathbf{K}^+ stands for the pseudoinverse of the matrix \mathbf{K} . The quantity $\mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}}) = \mathbf{Z}_j \hat{\boldsymbol{\theta}}$ is the value of the multivariate constraint (8) computed for \mathbf{m}_j , i.e., the algebraic distance of \mathbf{m}_j from the variety in \mathbb{R}^6 defined by (8). Note that $\hat{\boldsymbol{\theta}}$ appears also in the expression of $\mathbf{J}_{f|m_j}$. To minimize the criterion (10) again the LM algorithm is employed [5, pp.387], or a modification of the Gauss-Newton technique [14].

The LM based optimization starts from the initial solution obtained by minimizing the algebraic error

$$\hat{\boldsymbol{\theta}}^{[0]} = \arg \min_{\boldsymbol{\theta}} \sum_{j=1}^n \|\mathbf{Z}_j \boldsymbol{\theta}\|^2 \quad \text{subject to} \quad \|\hat{\boldsymbol{\theta}}^{[0]}\| = 1 \quad (11)$$

which is a Total Least Squares (TLS) problem. Since \mathbf{Z}_j is not a linear function in the measurements the TLS solution is biased [8], but the normalization of the measurements [5,

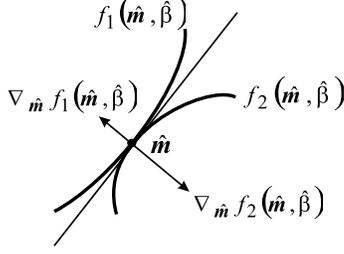


Figure 2: Singular constraint. The dimension of the space spanned by the gradients is less than the number of independent equations.

pp.91–93] helps to reduce the effect of the bias. The initial solution is usually not a geometrically valid tensor and $\hat{\theta}^{[0]}$ is corrected to obey (2) [5, pp.385].

The quality of the initial solution has a strong influence on the performance of nonlinear optimization methods. The main contribution of this paper is to propose a different way to estimate the trifocal tensor by minimizing a cost function similar to the Sampson distance. For “good” data the new approach has the same performance as the Gold Standard method (but it is significantly faster), while for “bad” data it is less sensitive and thus numerically more robust.

3. On the Singularity of $\mathbf{J}_{f|\hat{m}}$

The 6×4 Jacobian matrix (9) computed with \hat{m} and $\hat{\beta}$, $\mathbf{J}_{f|\hat{m}}$ has rank three. The property was already reported in [14] and explained using differential geometry concepts. The tensor defines a variety of dimension three in \mathbb{R}^6 (the space of the measurements, \mathbf{m}) since each point on the variety corresponds uniquely to a point in 3D. See also [15] for a detailed discussion of varieties in the context of the epipolar constraint.

A constraint having the Jacobian vanish is called singular [6, pp.131–133]. The columns of $\mathbf{J}_{f|\hat{m}}$ are the gradient vectors of the surfaces $f_k(\hat{m}, \hat{\beta}) = 0$, $k = 1, \dots, 4$, and the singularity implies that the four vectors are linearly dependent. This is illustrated in two dimensions in Figure 2. The two surfaces are tangent in the point \hat{m} and the two gradients are collinear. The point-point-point correspondence of the trifocal tensor is not the only case in computer vision when a singular constraint is met for general (nondegenerate) data. The conditions to be satisfied by an essential matrix are also singular [6, pp.336–338].

A rigorous algebraic proof of the singularity of (8) is based on the relations between the geometric elements in the three views (Figure 1), and provides another interpretation of the role of degenerate transfers between the views. To simplify the notations, for the moment will return to geometric entities (as in Section 1) and assume that \mathbf{m} satisfies the tensor constraint. The Jacobian matrix is singular when a linear combination of the four columns vanishes, i.e., sums

up to $\mathbf{0}_6$. This is equivalent to show that a linear combination of the four elements on a row is zero, using *same* coefficients for *all* the rows.

Denote the 3×3 matrix on the left side of (1) as $\mathbf{D}(\mathbf{m})$, and $\mathbf{s} = [0, 0, 1]^T$. Using \mathbf{s} we can select the four independent constraints employed in this paper, and thus the four elements of the i -th row of the Jacobian matrix are in the upper left 2×2 block of the 3×3 matrix

$$[\mathbf{s}]_{\times}^T \frac{\partial \mathbf{D}(\mathbf{m})}{\partial m_i} [\mathbf{s}]_{\times}. \quad (12)$$

We prove next that the coefficients of the sought linear combination of the four gradients are products of the components of the epipolar lines $\mathbf{l}', \mathbf{l}''$ (3). That is

$$\mathbf{l}'^T [\mathbf{s}]_{\times}^T \frac{\partial \mathbf{D}(\mathbf{m})}{\partial m_i} [\mathbf{s}]_{\times} \mathbf{l}'' = 0 \quad i = 1, \dots, 6. \quad (13)$$

When $i = 1, 2$

$$\frac{\partial \mathbf{D}(\mathbf{m})}{\partial m_i} = [\mathbf{x}']_{\times} \mathbf{T}_i [\mathbf{x}'']_{\times} \quad (14)$$

and it is easy to verify that

$$[\mathbf{x}'']_{\times} [\mathbf{s}]_{\times} \mathbf{l}'' = [\mathbf{x}'']_{\times} [\mathbf{s}]_{\times} [\mathbf{e}'']_{\times} \mathbf{x}'' = -\mathbf{x}_3'' \mathbf{l}'' . \quad (15)$$

The geometric meaning of (15) is that the point \mathbf{x}'' is on the epipolar line \mathbf{l}'' . Therefore (13) becomes (4) and vanishes. When $i = 3$

$$\frac{\partial \mathbf{D}(\mathbf{m})}{\partial m_3} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\times} \left(\sum_i x_i \mathbf{T}_i \right) [\mathbf{x}'']_{\times} \quad (16)$$

which makes (13) vanish because of (15) and (5). The cases of $i = 4, 5, 6$ can be proved the same way.

The Jacobian matrix thus is ill-conditioned near the solution. This might be the cause why the LM algorithm uses more steepest-descent than Gauss-Newton steps, as we have observed experimentally when minimizing the Sampson distance (10). Preference for the former significantly increased the time to convergence.

4. Numerically Robust Estimation

The *heteroscedastic errors-in-variables* (HEIV) model is the most general representation of the measurements, each data point \mathbf{m}_j , $j = 1, \dots, n$, having independent errors with different covariance matrices, i.e., $\delta \mathbf{m}_j \sim GI(\mathbf{0}, \sigma_v^2 \mathbf{C}_{m_j})$, where \mathbf{C}_{m_j} is known and the common noise variance σ_v^2 is to be estimated from the data. An estimator for multivariate constraints and heteroscedastic noise was developed in [9]. The technique is applied here to the trifocal tensor estimation taking into account the singularity of the problem. As the experimental results will prove, the HEIV estimator is significantly faster than the Levenberg-Marquardt algorithm and has superior performance for close to degenerate data.

4.1. HEIV Estimator

The availability of the first two moments of the error distribution allows the definition of a minimization criterion based on squared Mahalanobis distances

$$\mathcal{J}_M = \frac{1}{2} \sum_{j=1}^n (\mathbf{m}_j - \hat{\mathbf{m}}_j)^\top \mathbf{C}_{m_j}^+ (\mathbf{m}_j - \hat{\mathbf{m}}_j), \quad (17)$$

subject to (8). Note that for normally distributed errors (17) is the MLE criterion, and if the errors in the three views are uncorrelated it is identical to (6) under the Mahalanobis metric.

The minimization of (17) can be carried out by introducing the Lagrange multipliers $\boldsymbol{\eta}_j$,

$$\mathcal{J}_M = \frac{1}{2} \sum_{j=1}^n (\mathbf{m}_j - \hat{\mathbf{m}}_j)^\top \mathbf{C}_{m_j}^+ (\mathbf{m}_j - \hat{\mathbf{m}}_j) + \sum_{j=1}^n \boldsymbol{\eta}_j^\top \mathbf{f}(\hat{\mathbf{m}}_j, \hat{\boldsymbol{\beta}}).$$

Imposing $\frac{\partial \mathcal{J}_M}{\partial \hat{\mathbf{m}}_j} = \mathbf{0}$ we have

$$\hat{\mathbf{m}}_j = \mathbf{m}_j - \mathbf{C}_{m_j} \mathbf{J}_{f|\hat{\mathbf{m}}_j} \boldsymbol{\eta}_j, \quad \mathbf{J}_{f|\hat{\mathbf{m}}_j} = \frac{\partial \mathbf{f}(\hat{\mathbf{m}}_j, \hat{\boldsymbol{\beta}})^\top}{\partial \hat{\mathbf{m}}_j}.$$

The first order expansion of $\mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}})$ around $\hat{\mathbf{m}}_j$ yields

$$\mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}}) = \mathbf{f}(\hat{\mathbf{m}}_j, \hat{\boldsymbol{\beta}}) + \mathbf{J}_{f|\hat{\mathbf{m}}_j}^\top (\mathbf{m}_j - \hat{\mathbf{m}}_j)$$

and therefore the Lagrange multipliers are

$$\boldsymbol{\eta}_j = \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}}), \quad \hat{\boldsymbol{\Sigma}}_j = \mathbf{J}_{f|\hat{\mathbf{m}}_j}^\top \mathbf{C}_{m_j} \mathbf{J}_{f|\hat{\mathbf{m}}_j} \quad (18)$$

hence

$$\hat{\mathbf{m}}_j = \mathbf{m}_j - \mathbf{C}_{m_j} \mathbf{J}_{f|\hat{\mathbf{m}}_j} \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}}). \quad (19)$$

Due to the singularity of the Jacobian matrix $\mathbf{J}_{f|\hat{\mathbf{m}}_j}$, the covariance matrix of the constraint $\hat{\boldsymbol{\Sigma}}_j$ has rank three.

From (18) and (19) the cost function (17) is approximated as

$$\mathcal{J} = \frac{1}{2} \sum_{j=1}^n \mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}})^\top \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{f}(\mathbf{m}_j, \hat{\boldsymbol{\beta}}) \quad (20)$$

which represents a first order approximation of the geometric distance (6) and is similar to the Sampson distance (10). However, in (10) the matrices $\hat{\boldsymbol{\Sigma}}_j$ were computed using the error corrupted measurements \mathbf{m}_j instead of the corrected $\hat{\mathbf{m}}_j$ data points, and $\mathbf{C}_{m_j} = \mathbf{I}_6$.

Using the linearity of the constraint (7) in the tensor elements $\hat{\boldsymbol{\theta}}$ the cost function (20) can be rewritten as

$$\mathcal{J} = \frac{1}{2} \hat{\boldsymbol{\theta}}^\top \mathbf{S}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\theta}}, \quad \mathbf{S}(\hat{\boldsymbol{\theta}}) = \sum_{j=1}^n \mathbf{Z}_j^\top \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{Z}_j. \quad (21)$$

Minimizing (21) over $\boldsymbol{\theta}$ is an iterative process and yields an unconstrained $\hat{\boldsymbol{\theta}}$ which is not a valid tensor. Let $\hat{\boldsymbol{\theta}}^{[u]}$ be the estimate obtained at the u -th iteration. Our approach to compute a geometrically valid tensor is to project $\hat{\boldsymbol{\theta}}^{[u]}$ onto the space of $\boldsymbol{\beta}$, obtain $\hat{\boldsymbol{\beta}}^{[u]}$, compute $\hat{\boldsymbol{\theta}}_c^{[u]} = \mathbf{g}(\hat{\boldsymbol{\beta}}^{[u]})$ which is then used for the next iteration (see also Section 4.4).

The advantage of using $\hat{\mathbf{m}}_j$ in the Jacobian is twofold. Firstly, the matrix $\hat{\boldsymbol{\Sigma}}_j$ is known to have rank three and not just ill-conditioned, which may not be fully compensated in the estimation process by using the pseudoinverse. Secondly, the data correction alternates with the parameter estimation in a fashion similar with the expectation maximization (EM) paradigm. Indeed, given an estimate $\hat{\boldsymbol{\beta}}^{[u]}$, the corrected measurements $\hat{\mathbf{m}}_j$ are computed from (19), which in turn yields an improved $\hat{\boldsymbol{\beta}}^{[u+1]}$ found by minimizing (20) over the space of $\boldsymbol{\theta}$ and enforcing the tensor constraint. Similar two stage approach was also employed, for example, in statistics for nonlinear factor analysis [1], and in computer vision for structure from motion [7].

The minimization (21) over the unconstrained parameter space of $\boldsymbol{\theta}$ requires solving the generalized eigenvalue problem

$$\nabla_{\hat{\boldsymbol{\theta}}} \mathcal{J} \triangleq \frac{\partial \mathcal{J}}{\partial \hat{\boldsymbol{\theta}}} = [\mathbf{S}(\hat{\boldsymbol{\theta}}) - \mathbf{C}(\hat{\boldsymbol{\theta}})] \hat{\boldsymbol{\theta}} = \mathbf{0}_{27} \quad (22)$$

subject to $\|\hat{\boldsymbol{\theta}}\| = 1$, with the scatter matrix $\mathbf{S}(\hat{\boldsymbol{\theta}})$ defined in (21) and the weighted covariance matrix $\mathbf{C}(\hat{\boldsymbol{\theta}})$ defined as

$$\mathbf{C}(\hat{\boldsymbol{\theta}}) = \sum_{j=1}^n \sum_{k,l=1}^4 \eta_{kj} \eta_{lj} \left[\frac{\partial \varphi_k(\hat{\mathbf{m}}_j)}{\partial \hat{\mathbf{m}}_j} \right]^\top \mathbf{C}_{m_j} \left[\frac{\partial \varphi_l(\hat{\mathbf{m}}_j)}{\partial \hat{\mathbf{m}}_j} \right] \quad (23)$$

where η_{kj} denotes the k -th component of the Lagrange multiplier $\boldsymbol{\eta}_j$ (18). Both $\mathbf{S}(\hat{\boldsymbol{\theta}})$ and $\mathbf{C}(\hat{\boldsymbol{\theta}})$ are positive semidefinite matrices. The solution is unbiased in the first order [9].

To solve (22), at each iteration the updated estimate is the smallest eigenvector of

$$\mathbf{S}(\hat{\boldsymbol{\theta}}_c^{[u]}) \hat{\boldsymbol{\theta}}^{[u+1]} = \lambda \mathbf{C}(\hat{\boldsymbol{\theta}}_c^{[u]}) \hat{\boldsymbol{\theta}}^{[u+1]}. \quad (24)$$

See [9] for detailed description of the HEIV algorithm.

For conic fitting and fundamental matrix estimation it was found experimentally that the minimization of (21) by solving an eigenvalue problem similar to (24) is faster than the LM technique while having the same accuracy [4, 10]. The results to be described in Section 5 extend this observation for the trifocal tensor estimation.

It is important to point out that the original Sampson solution [13] solved iteratively

$$\mathbf{S}(\hat{\boldsymbol{\theta}}^{[u]}) \hat{\boldsymbol{\theta}}^{[u+1]} = \lambda \hat{\boldsymbol{\theta}}^{[u+1]}$$

which yields a biased estimate [6, p.273]. Thus the term Sampson distance implies only the cost function since the minimization is achieved with the LM algorithm [5], or by more advanced eigentechniques [4, 6, 8, 9].

4.2. Bifurcation of the Solution

Consider for the moment the algebraic error minimization (11). The solution is the eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_{j=1}^n \mathbf{Z}_j^\top \mathbf{Z}_j$. However, for noisy data it may happen that the last two eigenvalues have very similar values, i.e., the dimension of the effective null space is larger than one. Furthermore, the sought solution may no longer correspond to the smallest eigenvalue. The change of order represents a bifurcation of the objective function and can have a dramatic effect on the performance when the smallest eigenvector is not the desired solution. The use of the wrong eigenvector as the initialization of the LM algorithm leads to a drastic increase in the convergence time and possibly to an incorrect final solution.

The presence of bifurcation was described for structure from motion [7] for large measurement errors and a translation parallel to the image planes. It was noted that at a certain noise level the translation estimate suddenly changed direction with 90° . Since the effective dimension of the null space is relatively easy to determine, the possibility for a bifurcation can be recognized. The solution proposed in [7] was to retain the eigenvector which better satisfies the underlying geometrical properties of the task.

A null space with effective rank two can also appear when solving (24). The solution may be chosen as the “better” tensor, i.e., the one which gives a smaller reprojection error, however, this requires a large amount of computations. Instead, we used the *minimum norm* technique from statistics [17, pp.58–59], which beside being much faster was also found experimentally to give better results.

Let $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ be the two generalized eigenvectors spanning the null space. Then $\hat{\boldsymbol{\theta}} = \alpha_1 \hat{\boldsymbol{\theta}}_1 + \alpha_2 \hat{\boldsymbol{\theta}}_2$ is also a valid solution. To find α_1, α_2 beside the minimum norm condition another linear constraint $\hat{\boldsymbol{\theta}}^\top \mathbf{b} = 1$, where \mathbf{b} is a known vector, is employed. It can be proven that

$$\hat{\boldsymbol{\theta}} = \frac{\mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{b}}{\mathbf{b}^\top \mathbf{H}(\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{b}}, \quad \mathbf{H} \triangleq [\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2] \quad (25)$$

which depends on \mathbf{b} . In the iterative procedure an adequate choice of \mathbf{b} for the $(u+1)$ -th iteration is to find the index κ of the largest entry in absolute value of $\hat{\boldsymbol{\theta}}_c^{[u]}$ and set all the elements of \mathbf{b} to zero, except $b_\kappa = 1$. The vector $\hat{\boldsymbol{\theta}}$ is normalized to satisfy the norm one constraint.

4.3. Improved Initial Solution

Following [8] we used a generalized total least squares (GTLS) solution instead of (11). The approximate covari-

ance of the k -th row of the data matrix \mathbf{Z}_j is obtained by error propagation

$$\mathbf{C}_{kj} = \left[\frac{\partial \varphi_k(\mathbf{m}_j)}{\partial \mathbf{m}_j} \right]^\top \mathbf{C}_{m_j} \left[\frac{\partial \varphi_k(\mathbf{m}_j)}{\partial \mathbf{m}_j} \right]$$

Will assume that

$$\mathbf{C}_{kj} = \gamma_{kj} \bar{\mathbf{C}}, \quad \gamma_{kj} > 0, \quad k = 1, \dots, 4, \quad j = 1, \dots, n \quad (26)$$

where $\gamma_{kj}, \bar{\mathbf{C}}$ are unknown and determined by minimizing

$$\sum_{j=1}^n \sum_{k=1}^4 \|\mathbf{C}_{kj} - \bar{\mathbf{C}}\|_F^2 \quad (27)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix. The solution of (27) is

$$\bar{\mathbf{C}} = \frac{\sum_{j=1}^n \sum_{k=1}^4 \gamma_{kj} \mathbf{C}_{kj}}{\sum_{j=1}^n \sum_{k=1}^4 \gamma_{kj}^2}, \quad \gamma_{kj} = \frac{\text{trace}(\bar{\mathbf{C}} \mathbf{C}_{kj})}{\text{trace}(\bar{\mathbf{C}}^2)}. \quad (28)$$

To obtain for $\bar{\mathbf{C}}$ and γ_{kj} start from $\gamma_{kj} = 1$ and iterate twice (28). The relation (26) together with the simplifying assumption that the rows of the data matrix \mathbf{Z}_j are uncorrelated implies that computing $\mathbf{C}(\hat{\boldsymbol{\theta}})$ (23) using the measurements yields a matrix proportional to $\bar{\mathbf{C}}$. Under the same assumptions the scatter matrix $\mathbf{S}(\hat{\boldsymbol{\theta}})$ becomes

$$\bar{\mathbf{S}} = \sum_{j=1}^n \mathbf{Z}_j^\top \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{Z}_j, \quad \hat{\boldsymbol{\Sigma}}_j = \mathbf{diag}(\gamma_{1j}, \gamma_{2j}, \gamma_{3j}, \gamma_{4j}).$$

The unconstrained initial solution $\hat{\boldsymbol{\theta}}^{[0]}$ therefore is obtained using instead of (24)

$$\bar{\mathbf{S}} \hat{\boldsymbol{\theta}}^{[0]} = \lambda \bar{\mathbf{C}} \hat{\boldsymbol{\theta}}^{[0]} \quad (29)$$

which is a GTLS problem having closed form solution, the smallest generalized eigenvector.

4.4. Imposing the Tensor Constraint

Let $\hat{\boldsymbol{\theta}}^{[u+1]} \in \mathbb{R}^{27}$ be the current estimate, obtained by solving (24). To impose the tensor constraint on $\hat{\boldsymbol{\theta}}^{[u+1]}$ the unconstrained solution is projected into \mathbb{R}^{24} , the space of the camera parameters $\hat{\boldsymbol{\beta}}$, where the constraint is approximately obeyed. The solution is updated in this space and subsequently mapped back into the unconstrained space. The projection can be carried out optimally in the metric induced by the covariance matrix of $\hat{\boldsymbol{\theta}}^{[u+1]}$. This covariance matrix is estimated by [6, p.285]

$$\hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}^{[u+1]}} \propto \mathbf{W}^+, \quad \mathbf{W} = \left[\mathbf{S} \left(\hat{\boldsymbol{\theta}}_c^{[u]} \right) - \lambda \mathbf{C} \left(\hat{\boldsymbol{\theta}}_c^{[u]} \right) \right] \quad (30)$$

where λ is the smallest generalized eigenvalue corresponding to $\hat{\boldsymbol{\theta}}^{[u+1]}$. Because of (30) the range of the covariance matrix $\hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}^{[u+1]}}$ (and the row space of \mathbf{W}) coincides with the hyperplane in \mathbb{R}^{27} having the normal $\hat{\boldsymbol{\theta}}^{[u+1]}$, as required by any procedure seeking a satisfactory update for $\hat{\boldsymbol{\beta}}^{[u]}$. Indeed, the projector into this hyperplane is $\mathbf{P}_{\hat{\boldsymbol{\theta}}^{[u+1]}} = (\mathbf{I}_{27} - \hat{\boldsymbol{\theta}}^{[u+1]} \hat{\boldsymbol{\theta}}^{[u+1]\top})$ and

$$\mathbf{P}_{\hat{\boldsymbol{\theta}}^{[u+1]}} \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}^{[u+1]}} \mathbf{P}_{\hat{\boldsymbol{\theta}}^{[u+1]}} = \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}^{[u+1]}}$$

because of (24). To find $\hat{\boldsymbol{\beta}}^{[u+1]}$ the following nonlinear least squares problem has to be solved under the Mahalanobis metric defined by $\hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}^{[u+1]}}$

$$\hat{\boldsymbol{\beta}}^{[u+1]} = \arg \min_{\boldsymbol{\beta}^{[u+1]}} \left\| \hat{\boldsymbol{\theta}}^{[u+1]} - \mathbf{g}(\boldsymbol{\beta}^{[u+1]}) \right\|_{\hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}^{[u+1]}}}^2. \quad (31)$$

Linearizing $\mathbf{g}(\hat{\boldsymbol{\beta}}^{[u+1]})$ yields

$$\mathbf{g}(\hat{\boldsymbol{\beta}}^{[u+1]}) = \hat{\boldsymbol{\theta}}_{\mathbf{c}}^{[u]} + \mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}^{[u]}}^{\top} (\hat{\boldsymbol{\beta}}^{[u+1]} - \hat{\boldsymbol{\beta}}^{[u]}) \quad (32)$$

where the definition of $\hat{\boldsymbol{\theta}}_{\mathbf{c}}^{[u]}$ was also taken into account. The projection defined in (31) is then solved as a linear weighted least squares problem. Note that the weights \mathbf{W} are optimal by the Gauss-Markov theorem. The updated estimate is

$$\hat{\boldsymbol{\beta}}^{[u+1]} = \hat{\boldsymbol{\beta}}^{[u]} - \left[\mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}^{[u]}} \mathbf{W} \mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}^{[u]}}^{\top} \right]^+ \mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}^{[u]}} \mathbf{W} \hat{\boldsymbol{\theta}}_{\mathbf{c}}^{[u]} \quad (33)$$

where the fact that $\hat{\boldsymbol{\theta}}^{[u+1]}$ is in the null space of \mathbf{W} was also taken into account.

The above procedure is actually a gauge fixing [11]. The parametrization by $\hat{\boldsymbol{\beta}}$ is consistent, but not minimal [14] and at least in theory the two scale factors of \mathbf{P}' and \mathbf{P}'' can be also eliminated. Gauge fixing was shown to improve the computational behavior of the LM based optimizations [11].

4.5. Reliability Evaluation

The reliability of the obtained estimates $\hat{\mathbf{m}}_j$ and $\hat{\boldsymbol{\theta}}$ can be assessed by defining their confidence regions. for the parameters $\hat{\mathbf{m}}_j$ and $\hat{\boldsymbol{\theta}}$. During the estimation process the knowledge of the noise variance σ_{ν}^2 was not required, however the scale of the confidence regions depends on σ_{ν}^2 . It can be proven [9] that the estimator

$$\hat{\sigma}_{\nu}^2 = \frac{\hat{\boldsymbol{\theta}}^{\top} \mathbf{S}(\hat{\boldsymbol{\theta}}) \hat{\boldsymbol{\theta}}}{3n - 24}$$

is unbiased in the first order.

The covariance of the unconstrained tensor was computed (up to a scale) in (30). It can be shown by error propagation that the estimated covariance of the constrained tensor $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$ is $\hat{\sigma}_{\nu}^2 \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\mathbf{c}}}$ where

$$\hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\mathbf{c}}} = \mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}}^{\top} \left[\mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}} \mathbf{W} \mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}}^{\top} \right]^+ \mathbf{J}_{\mathbf{g}|\hat{\boldsymbol{\beta}}}. \quad (34)$$

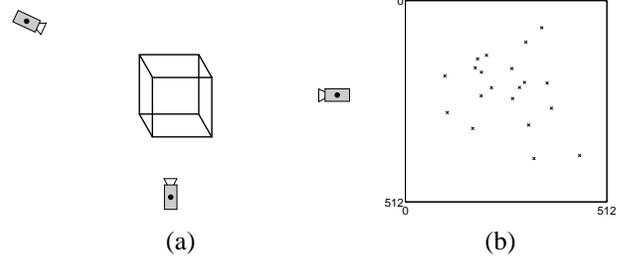


Figure 3: Generic data. (a) Spatial configuration of the cameras. (b) Typical view of the 20 points.

The covariance of the corrected measurements $\hat{\mathbf{m}}_j$, conditioned on $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$ can be obtained by error propagation from (19) as described in [9]. The *unconditioned* covariance matrix $\hat{\sigma}_{\nu}^2 \hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}$ must include the uncertainty due to the estimation of $\hat{\boldsymbol{\theta}}_{\mathbf{c}}$. It can be proven that

$$\begin{aligned} \hat{\mathbf{C}}_{\hat{\mathbf{m}}_j} &= \hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}^{(1)} + \hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}^{(2)} \\ \hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}^{(1)} &= \mathbf{C}_{m_j} - \mathbf{C}_{m_j} \mathbf{J}_{f|\hat{\mathbf{m}}_j} \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{J}_{f|\hat{\mathbf{m}}_j}^{\top} \mathbf{C}_{m_j} \\ \hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}^{(2)} &= \mathbf{C}_{m_j} \mathbf{J}_{f|\hat{\mathbf{m}}_j} \hat{\boldsymbol{\Sigma}}_j^+ \hat{\mathbf{Z}}_j \hat{\mathbf{C}}_{\hat{\boldsymbol{\theta}}_{\mathbf{c}}} \hat{\mathbf{Z}}_j^{\top} \hat{\boldsymbol{\Sigma}}_j^+ \mathbf{J}_{f|\hat{\mathbf{m}}_j}^{\top} \mathbf{C}_{m_j} \end{aligned} \quad (35)$$

The matrix $\hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}^{(1)}$ has rank three since the second expression on the right side removes the uncertainty in the subspace orthogonal to the three dimensional variety of the constraint. The second term $\hat{\mathbf{C}}_{\hat{\mathbf{m}}_j}^{(2)}$ accounts for the uncertainty of the estimation process due to the specific properties the available measurements such as number of points, spread, etc.

5. Experimental Results

We have estimated the trifocal tensor using three different approaches

- **GS**: Gold Standard algorithm [5, pp.385]. TLS initial solution (11).
- **HEIV**: minimization of the Sampson distance (20) with HEIV estimator. GTLS initial solution (29).
- **SLM**: minimization of the Sampson distance (10) with the LM procedure. TLS initial solution (11).

All three methods were implemented in MATLAB and for LM the Optimization Toolbox function was used. It is important to emphasize that the GS method is the theoretically optimal technique under normal measurement noise.

5.1. Synthetic Data

Two types of synthetic data were employed. The *generic* data consisted of 20 points uniformly distributed in a cube with the three cameras ($f=1700$ pixel units) located around

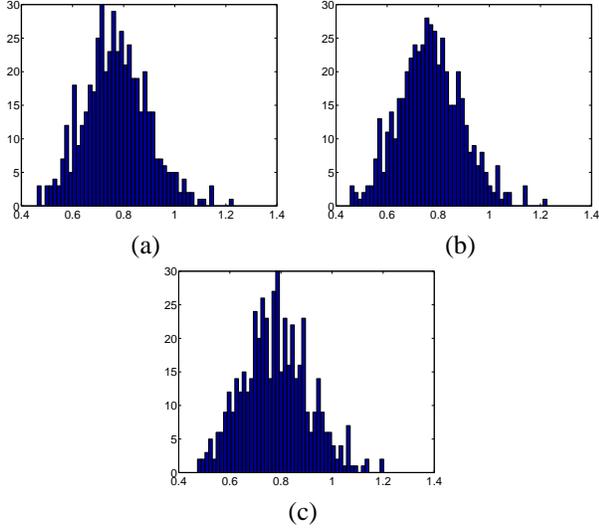


Figure 4: The histograms of the residual errors for generic data. Employed estimation method: (a) GS. (b) HEIV. (c) SLM.

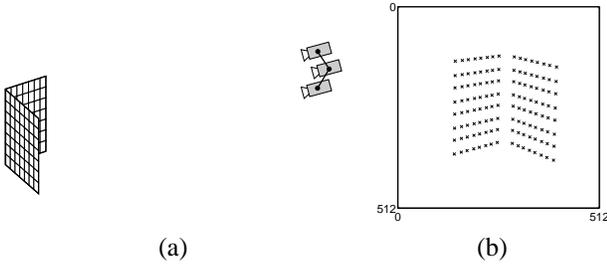


Figure 5: Difficult data. (a) Spatial configuration of the cameras. (b) Typical view of the 128 image points.

the cloud (Figure 3a). The image points were corrupted with normal noise, $\sigma = 2$ pixel units. A typical view is shown in Figure 3b. The trifocal tensor was estimated in 500 trials for each method.

As expected, for “good” data all three methods performed similarly (Figure 4). However, the time to convergence differed dramatically. Taking the HEIV as reference, GS was about three times and the SLM at least an order of magnitude slower.

We have also investigated a *difficult* data set. The three cameras ($f=1440$) are far from from a synthetic calibration grid and their positions only slightly deviate from a vertical line (Figure 5a). Note that the baseline between the cameras is very small and the definition of the trifocal plane is not firm [5, p.373]. The 128 image points in each view were corrupted with normal noise, $\sigma = 1$. The trifocal tensor was estimated in 100 trials for each method.

The performances of the GS and HEIV methods were statistically identical except for five trials in which the GS failed to converge to the correct solution (Figure 6a) due

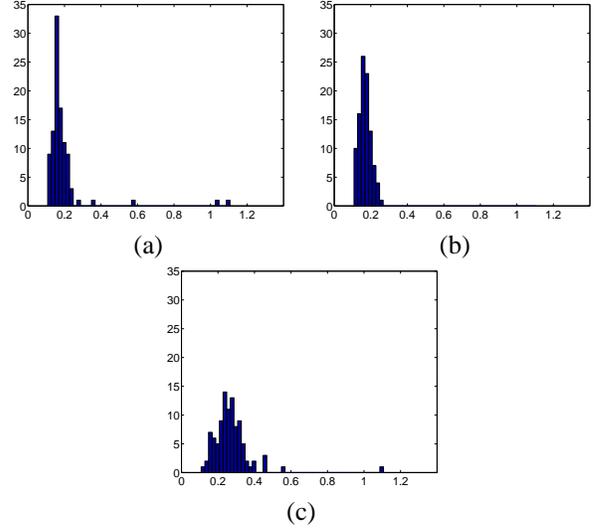


Figure 6: The histograms of the residual errors for difficult data. Employed estimation method: (a) GS. (b) HEIV. (c) SLM.

to a poor TLS initial solution generated by a bifurcation. However, using the GTLS initial solution borrowed from the HEIV procedure these errors could be eliminated, proving the importance of employing the best possible initialization.

The difficult data put a large burden on the LM algorithm which needed much more iterations than for generic data. Again using the HEIV as reference, the GS took at least ten times longer to converge, while the convergence of SLM was impractically slow. Furthermore, the performance of SLM is significantly worse than that of the two other techniques (Figure 6c). Thus, for difficult data the SLM cannot be considered as a substitute for GS [5, p.389].

We conclude that the HEIV method has a performance identical to the optimal GS technique but it is much faster. The use of GTLS initial solution is recommended for GS, though once implemented already most of the code needed for HEIV is in place. None of the algorithms were implemented optimally beyond common sense MATLAB tricks, and we have also not investigated the sparse LM algorithm [5, pp.571–576]. The sparse LM returns the same estimates as a full LM, only the amount of computations is reduced. However, its implementation is problem specific and requires considerable sophistication. The HEIV estimator (while not a trivial method, we acknowledge) is a general technique already applied to numerous computer vision problems, e.g., [8, 9, 10].

5.2. Real Data

To illustrate the effectiveness of the reliability measures (Section 4.5) we have used three images taken from the webpage of the Oxford Visual Geometry Group (Figure 7).

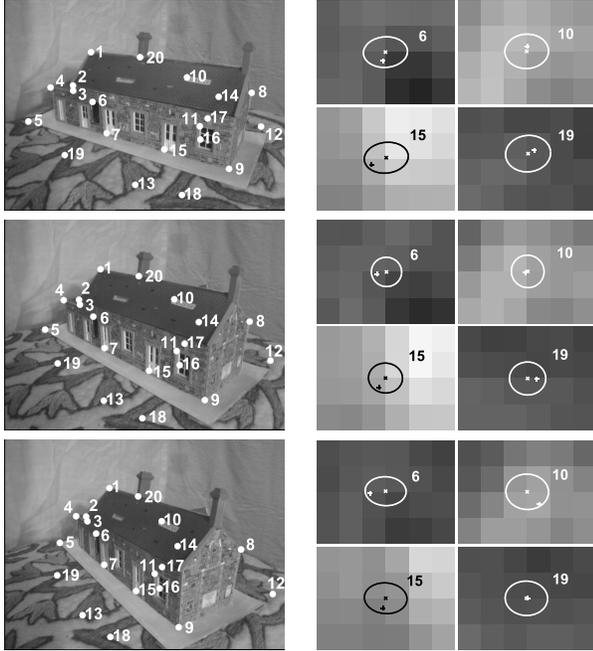


Figure 7: Experiments with real images. Left: the three views with the correspondences marked. Right: confidence regions of four points from the view at the left.

From these images 20 correspondences were chosen manually. The trifocal tensor was then estimated (residual error 0.4033 for both GS and HEIV), and the corrected measurements marked in the images. This set of 20 triplets constituted the ground truth. Normal noise, $\sigma = 0.5$, was added to the points and the trifocal tensor was estimated again with the GS and HEIV methods. They gave practically identical results with a residual error of 0.1725.

The confidence regions of the corrected image points were then obtained (for 0.95 confidence level) according to (35). At the right of Figure 7 such confidence regions are shown for four points chosen in the view at the left. The corrected measurements are the centers of the ellipses while the “true” (ground truth) points are always located inside their confidence regions. Note the differences in size and orientation between the different confidence regions.

6. Conclusion

We have presented an application of a general technique, the heteroscedastic errors-in-variables (HEIV) estimator. Heteroscedasticity inherently appears in computer vision tasks whenever a nonlinearity is present or an incidence relation is to be enforced between noisy measurements. We have shown that for the estimation of the trifocal tensor the HEIV estimator achieves the performance of the optimal Gold Standard at less computational effort. MATLAB code of the estimator is available at the website

www.caip.rutgers.edu/riul

Acknowledgments

The support of the NSF grants IIS 98-72995 and IRI 99-87695 is gratefully acknowledged. We thank Théodore Papadopoulos for the difficult data used in the experiments.

References

- [1] Y. Amemiya, “Generalization of the TLS approach in errors-in-variables problem,” in S. V. Huffel, editor, *Recent Advances in Total Least Squares Techniques and Errors-In-Variables Modeling*, Society for Industrial and Applied Mathematics, 1997, pp. 77–86.
- [2] J. W. Brewer, “Kronecker products and matrix calculus in system theory,” *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 772–781, June 1978.
- [3] N. Canterakis, “A minimal set of constraints for the trifocal tensor,” in D. Vernon, editor, *Computer Vision - ECCV 2000*, Springer, 2000, pp. 84–99.
- [4] 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. Machine Intell.*, vol. 22, pp. 1294–1303, 2000.
- [5] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- [6] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, 1996.
- [7] J. Kosecka, Y. Ma, and S. Sastry, “Optimization criteria, sensitivity and robustness,” in B. Triggs, A. Zisserman, and R. Szelisky, editors, *Vision Algorithms: Theory and Practice*, Springer, 2000, pp. 168–182.
- [8] Y. Leedan and P. Meer, “Heteroscedastic regression in computer vision: Problems with bilinear constraint,” *International Journal of Computer Vision*, vol. 37, pp. 127–150, 2000.
- [9] B. Matei and P. Meer, “A general method for errors-in-variables problems in computer vision,” in *Computer Vision and Pattern Recognition Conference*, volume II, (Hilton Head Island, SC), June 2000, pp. 18–25.
- [10] B. Matei and P. Meer, “Reduction of bias in maximum likelihood ellipse fitting,” in *15th International Conference on Computer Vision and Pattern Recognition*, volume III, (Barcelona, Spain), September 2000, pp. 802–806.
- [11] P. McLauchlan, “Gauge independence in optimization algorithms for 3d vision,” in B. Triggs, A. Zisserman, and R. Szelisky, editors, *Vision Algorithms: Theory and Practice*, Springer, 2000, pp. 183–198.
- [12] T. Papadopoulos and O. Faugeras, “A new characterization of the trifocal tensor,” in B. N. H. Burkhardt, editor, *Computer Vision - ECCV 98*, Springer, 1998, pp. 109–123.
- [13] P. Sampson, “Fitting conic sections to “Very Scattered” data: An iterative refinement of the Bookstein algorithm,” *Computer Graphics and Image Processing*, vol. 18, pp. 97–108, 1982.
- [14] P. H. S. Torr and A. Zisserman, “Robust parametrizations and computation of the trifocal tensor,” *Image and Vision Computing*, vol. 15, pp. 591–605, 1997.
- [15] P. H. S. Torr, A. Zisserman, and S. J. Maybank, “Robust detection of degenerate configurations while estimating the fundamental matrix,” *Computer Vision and Image Understanding*, vol. 71, pp. 312–333, 1998.
- [16] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, “Bundle adjustment — A modern synthesis,” in B. Triggs, A. Zisserman, and R. Szelisky, editors, *Vision Algorithms: Theory and Practice*, Springer, 2000, pp. 298–372.
- [17] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem. Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics, 1991.