International Journal of Computer Vision 37(2), 127–150, 2000

Heteroscedastic Regression in Computer Vision: Problems with Bilinear Constraint

YORAM LEEDAN
P.O.Box 730, Kfar-Vradim, 25147 Israel

PETER MEER*
Department of Electrical and Computer Engineering, Rutgers University, 94 Brett Road, Piscataway, NJ 08854-8058, USA
meer@caip.rutgers.edu

Abstract. We present an algorithm to estimate the parameters of a linear model in the presence of heteroscedastic noise, i.e., each data point having a different covariance matrix. The algorithm is motivated by the recovery of bilinear forms, one of the fundamental problems in computer vision which appears whenever the epipolar constraint is imposed, or a conic is fit to noisy data points. We employ the errors-in-variables (EIV) model and show why already at moderate noise levels most available methods fail to provide a satisfactory solution. The improved behavior of the new algorithm is due to two factors: taking into account the heteroscedastic nature of the errors arising from the linearization of the bilinear form, and the use of generalized singular value decomposition (GSVD) in the computations. The performance of the algorithm is compared with several methods proposed in the literature for ellipse fitting and estimation of the fundamental matrix. It is shown that the algorithm achieves the accuracy of nonlinear optimization techniques at much less computational cost.

Keywords: heteroscedastic regression, ellipse fitting, epipolar constraint, fundamental matrix, uncalibrated camera

1. Introduction

It is a sign of maturity for the field of image understanding that more and more often the generic problems, at the roots of several vision tasks, are recognized and addressed. In this paper we discuss such a problem, the estimation of a parameter vector from data in which each measurement is available with a different uncertainty. In this case the data is known as corrupted by heteroscedastic noise. A reliable solution for linear regression in the presence of heteroscedastic noise is clearly of importance for many vision algorithms. The same problem, however, also arises whenever a bilinear (quadratic) constraint is imposed over homoscedastic data, i.e., data with each point corrupted by the same noise process. A computationally feasible solution for estimating the parameter matrix of the bilinear form requires linearization which in turn yields a heteroscedastic regression. While the algorithm described in this paper was motivated by the bilinear problem it can be applied to any vision task in which the data is heteroscedastic.

Let $\mathbf{p}_{Aio} = [x_{Aio}, y_{Aio}, 1]^\top$ and $\mathbf{p}_{Bio} = [x_{Bio}, y_{Bio}, 1]^\top$, $i = 1 \,.\,.\, n$, be the affine coordinates of pairs of matched, *error-free* points in two images. Then, often the geometric constraints between these points can be written as a bilinear form

$$\mathbf{p}_{Aio}^\top \mathbf{F} \mathbf{p}_{Bio} = 0, \tag{1}$$

*Corresponding author.

with the 3×3 matrix $\mathbf{F}$ required to satisfy additional restrictions.

The epipolar geometry defined by two cameras is captured by (1), with $\mathbf{F}$ being of rank two. When the calibration of the cameras is not known (as in projective reconstruction), $\mathbf{F}$ is called the *fundamental* matrix. When the cameras are calibrated (as in motion estimation), $\mathbf{F}$ is called the *essential* matrix and must also have two identical nonzero singular values (Faugeras, 1993, p. 250). Taking $\mathbf{p}_{Aio} = \mathbf{p}_{Bio} = \mathbf{p}_{io}$ and limiting $\mathbf{F}$ to symmetric matrices, (1) becomes a quadratic form representing a conic in the image plane. Conics are often used to approximate the shape of 3D objects projected into 2D images, the most important case being the ellipse which appears as the projected outline of objects with rotational symmetry. The Eq. (1) represents a real conic when the signature (the number of positive and negative eigenvalues) of $\mathbf{F}$ is (2,1) or (1,2). In order to have (1) represent an ellipse, additional restrictions on $\mathbf{F}$ are required (Kanatani, 1996, p. 114). We will use both the ellipse fitting and the fundamental matrix recovery tasks to compare the performance of the new estimation method with that of techniques currently employed in the literature.

Let $\mathbf{m}_{io}$ be the vector constructed from the Cartesian coordinates of the matched, error-free (true) points $\mathbf{p}_{Aio}$ and $\mathbf{p}_{Bio}$. In general, $\mathbf{m}_{io} \in \mathcal{R}^k$ where $k = 2$ for the quadratic forms and $k = 4$ for the bilinear forms. Then we can always write the bilinear form (1) as a linear form

$$\mathbf{p}_{Aio}^{\top} \mathbf{F} \mathbf{p}_{Bio} = \alpha_o + \mathbf{z}(\mathbf{m}_{io})^{\top} \boldsymbol{\theta}_o = 0, \tag{2}$$

where the new vector $\mathbf{z}_{io} = \mathbf{z}(\mathbf{m}_{io}) \in \mathcal{R}^p$ is a *nonlinear* mapping of the elements of $\mathbf{m}_{io}$, and the parameter vector $\boldsymbol{\theta}_o$ is derived from the true matrix $\mathbf{F}$. Note that the intercept is treated separately. The relevance of this will become clear when the linearization is discussed in Section 2.2. Since (2) is determined up to a nonzero scaling factor we will normalize it by imposing

$$\|\boldsymbol{\theta}_o\|^2 = 1. \tag{3}$$

Thus, for example, the quadratic form of the conic fitting problem

$$\mathbf{p}_{io}^{\top} \begin{bmatrix} d & c/2 & a/2 \\ c/2 & e & b/2 \\ a/2 & b/2 & \alpha_o \end{bmatrix} \mathbf{p}_{io} = 0, \tag{4}$$

can be also regarded as a linear form with

$$\begin{aligned} \boldsymbol{\theta}_o &= [a, b, c, d, e]^{\top}, \\ \mathbf{z}_{io} &= \left[x_{io},\ y_{io},\ x_{io}y_{io},\ x_{io}^2,\ y_{io}^2\right]^{\top}. \end{aligned} \tag{5}$$

In practice, the available matched points $\mathbf{p}_{Ai}$ and $\mathbf{p}_{Bi}$ are error prone, and the matrix $\mathbf{F}$ has to be estimated. In the presence of measurement errors, the noisy vector $\mathbf{m}_i$ derived from the matched image points is observed. Assuming additive errors we have

$$\mathbf{m}_i = \mathbf{m}_{io} + \Delta\mathbf{m}_i, \quad \Delta\mathbf{m}_i \sim \mathrm{GI}(\mathbf{0}, \mathbf{C}_{\mathbf{m}}), \tag{6}$$

where '$\sim$GI($\mathbf{0}$, $\mathbf{C}_{\mathbf{m}}$)' means 'distributed independently with a general, symmetric p.d.f. having zero mean and positive semi-definite covariance matrix'. The error process (6) together with the constraint (2) represent a *nonlinear, errors-in-variables* (EIV) model with intercept (Van Huffel and Vandewalle, 1991, Chap. 8).

The common covariance matrix $\mathbf{C}_{\mathbf{m}} = \sigma_\nu^2 \mathbf{C}_{\mathbf{m}}^{\star}$ is assumed to be known up to a multiple σ_ν^2, the noise variance. Most often in applications the normalized covariance matrix $\mathbf{C}_{\mathbf{m}}^{\star}$ is taken to be the identity matrix $\mathbf{I}$. However, if a priori knowledge about the noise process is available (such as resolution differences between the two images) it should be incorporated into $\mathbf{C}_{\mathbf{m}}^{\star}$. In Section 4.3 we will show the importance of using non-identity $\mathbf{C}_{\mathbf{m}}^{\star}$ through an example. We further assume that no erroneous matches are present in the observed data. Such outliers can be eliminated before the estimation process by using robust matching methods as in Zhang et al. (1995).

In Section 2 it is shown that by approaching the estimation of $\mathbf{F}$ through the linear form, i.e., linearization of the bilinear problem, leads to a heteroscedastic regression. The currently proposed solutions for the estimation of the fundamental matrix and conic fitting, are discussed in this context. In Section 3 the new method for heteroscedastic regression is introduced and applied to estimation with a bilinear constraint. The performance of the new algorithm is compared with other techniques for ellipse fitting in Section 4, and for estimation of the fundamental matrix in Section 5. The connection to geometric distance minimization and other related issues are discussed in Section 6.

2. The Bilinear Problem

Any estimation process has two components, recovery of the model parameters (*parameter estimation*),

and recovery of the uncorrupted data points (*data correction*). The latter are also known as the nuisance parameters since the associated uncertainty does not depend on the model parameters. In linear estimation problems the two components are intertwined and all the parameter estimates can be obtained simultaneously. Parameter estimation with a bilinear constraint, however, is a nonlinear process and the estimation of $\hat{\mathbf{p}}_{Ai}$, $\hat{\mathbf{p}}_{Bi}$ and $\hat{\mathbf{F}}$ simultaneously is a complicated task. If the parameter estimation procedure provides a consistent estimator $\hat{\mathbf{F}}$, then subsequent estimation of $\hat{\mathbf{p}}_{Ai}$ and $\hat{\mathbf{p}}_{Bi}$ is usually satisfactory. Thus, first will focus on the estimation of the model parameters.

2.1. Linearized Model

To estimate the parameters of the nonlinear EIV model, the sum of squared Mahalanobis distances between the noisy points and the true points should be minimized in $\mathcal{R}^k$

$$[\hat{\alpha}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{m}}_i] = \arg \min_{\alpha_o, \boldsymbol{\theta}_o, \mathbf{m}_{io}} \sum_{i=1}^{n} (\mathbf{m}_i - \mathbf{m}_{io})^\top \mathbf{C}_{\mathbf{m}}^{-} (\mathbf{m}_i - \mathbf{m}_{io}), \quad (7)$$

$$\text{subject to} \quad \alpha_o + \mathbf{z}(\mathbf{m}_{io})^\top \boldsymbol{\theta}_o = 0 \ \text{ and } \ \|\boldsymbol{\theta}_o\|^2 = 1, \quad (8)$$

where $\mathbf{C}_{\mathbf{m}}^{-}$ is the pseudoinverse of $\mathbf{C}_{\mathbf{m}}$ which also accounts for the case when $\mathbf{C}_{\mathbf{m}}$ is singular. Note that the minimization (8) corresponds to the maximum likelihood solution for normal noise, under the functional model (Fuller, 1987, p. 124–139), in which the true values of the points $\mathbf{m}_{io}$ are considered as unknown constants. The solutions for $\hat{\alpha}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{m}}_i$ cannot be obtained in closed form and an iterative procedure is needed. Satisfactory convergence of the nonlinear optimization procedure requires a good initial estimate and it is computationally impractical without using approximations (Fuller, 1987, p. 230; Taubin, 1991).

However, if the bilinear constraint is written under its linear form (2) the estimation of $\boldsymbol{\theta}_o$ becomes the solution of a linear minimization problem in $\mathcal{R}^p$, the space of $\mathbf{z}$. This is indeed true, though with a very important caveat. In an optimal parameter fitting procedure for the linearized model, the error characteristics of the linearized model variables $\mathbf{z}_i$ must be taken into account. The mapped vectors $\mathbf{z}_i$, obtained from the observed vectors $\mathbf{m}_i$ can be approximated as

$$\mathbf{z}_i = \mathbf{z}_{io} + \Delta\mathbf{z}_i, \quad \Delta\mathbf{z}_i \sim \mathrm{GI}(\boldsymbol{\mu}, \mathbf{C}_i), \ i = 1 \ldots n. \quad (9)$$

The expressions of the mean and covariance matrix of $\Delta\mathbf{z}_i$ are obtained from the mapping function $\mathbf{z}(\mathbf{m})$. In our case the points $\mathbf{m}_i$ have a common covariance matrix $\mathbf{C}_{\mathbf{m}}$, and the linearized error vectors $\Delta\mathbf{z}_i$ have the same mean vector $\boldsymbol{\mu}$, but *data dependent* covariance matrices $\mathbf{C}_i$. Such data, with the distribution of each point depending on its true coordinates is called *heteroscedastic*.

For example, assuming $\Delta\mathbf{m}_i \sim \mathrm{NI}(\mathbf{0}, \sigma_\nu^2\mathbf{I})$, i.e., normal i.i.d. errors, and using (5) as the mapping function, we obtain

$$\boldsymbol{\mu} = \left[0, 0, 0, \sigma_\nu^2, \sigma_\nu^2\right]^\top, \quad (10)$$

$$\mathbf{C}_i = \sigma_\nu^2 \begin{bmatrix} 1 & 0 & y_{io} & 2x_{io} & 0 \\ 0 & 1 & x_{io} & 0 & 2y_{io} \\ y_{io} & x_{io} & x_{io}^2 + y_{io}^2 + \sigma_\nu^2 & 2x_{io}y_{io} & 2x_{io}y_{io} \\ 2x_{io} & 0 & 2x_{io}y_{io} & 4x_{io}^2 + 2\sigma_\nu^2 & 0 \\ 0 & 2y_{io} & 2x_{io}y_{io} & 0 & 4y_{io}^2 + 2\sigma_\nu^2 \end{bmatrix}$$
$$= \sigma_\nu^2 \mathbf{C}_i^\star. \quad (11)$$

Note the presence of the true (and thus unknown) coordinates of the data point in the expression of $\mathbf{C}_i^\star$.

In the estimation process an approximation for the normalized matrix $\mathbf{C}_i^\star$ must be available. From the mapping function it can be seen that in the general case the knowledge of the fourth order moments of the elements of $\Delta\mathbf{m}_i$ are required. The third order moments are zero since we assumed a symmetrical error distribution. The way to approximate $\mathbf{C}_i^\star$ will be discussed in Section 3.1.2. The estimation of the noise variance, σ_ν^2 is discussed in Section 3.1.

The parameters $[\alpha_o, \boldsymbol{\theta}_o]$ of the *linear, heteroscedastic* EIV model can be found by minimizing the squared Mahalanobis distances in $\mathcal{R}^p$

$$[\hat{\alpha}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{z}}_i] = \arg \min_{\alpha_o, \boldsymbol{\theta}_o, \mathbf{z}_{io}} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu} - \mathbf{z}_{io})^\top \times \mathbf{C}_i^{\star -} (\mathbf{z}_i - \boldsymbol{\mu} - \mathbf{z}_{io}), \quad (12)$$

$$\text{subject to} \quad \alpha_o + \mathbf{z}_{io}^\top \boldsymbol{\theta}_o = 0 \ \text{ and } \ \|\boldsymbol{\theta}_o\|^2 = 1, \quad (13)$$

with the linearized measurement errors defined in (9). Note that by subtracting the mean of the errors $\boldsymbol{\mu}$, the residuals used in the Mahalanobis distances have zero mean as required by the implicit assumption of the minimization. The price of moving from a nonlinear minimization problem to a linear one is the loss of

homoscedasticity (same distribution) of the errors. The estimated parameters are not necessarily the same as by solving the original, nonlinear problem (7), however, they are very close in practice. Special care is needed for the data correction procedure as will be discussed in Section 3.2.

2.2. Literature Review

For the moment we restrict ourselves to the algorithms proposed in the computer vision literature which estimate the matrix of the bilinear form making use of the linearization described in the preceding section and assume $\mathbf{C}_{\mathbf{m}}^{\star} = \mathbf{I}$. (Will return to the issue of the optimization criteria for image understanding problems in Section 6.) Two important misconceptions can be often noticed.

1. The effect of the nonlinear mapping of $\mathbf{m}_i$ into $\mathbf{z}(\mathbf{m}_i)$ on the errors $\Delta\mathbf{z}_i$ is ignored.
2. The intercept α_o is treated the same way as the other model parameters θ_{jo}, i.e., the elements of $\boldsymbol{\theta}_o$.

In the widely used *eight-point algorithm* for the estimation of the fundamental matrix, e.g, (Faugera, 1993, p. 274), (Hartley, 1997a) both misconceptions are present. For a critical discussion see Luong and Faugeras (1996), Torr and Murray (1997) and Zhang (1998a). A similar method is also popular for conic fitting, see Zhang (1997) for an evaluation.

The original eight-point algorithm assumes $\Delta\mathbf{z} \sim \mathrm{NI}(\mathbf{0}, \sigma_\nu^2\mathbf{I})$, and finds the parameter estimate $\hat{\mathbf{u}} = [\hat{\alpha},\ \hat{\boldsymbol{\theta}}^\top]^\top$ under the constraint $\|\hat{\mathbf{u}}\|^2 = 1$, as the eigenvector associated with the smallest eigenvalue of the matrix

$$\mathbf{S} = [\mathbf{1}_n, \mathbf{Z}]^\top [\mathbf{1}_n, \mathbf{Z}], \quad \mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n]^\top, \tag{14}$$

where $\mathbf{1}_n$ is an n-dimensional column vector of ones. This eigenvector spans the enforced null-space of the matrix $[\mathbf{1}_n, \mathbf{Z}]$ in $\mathcal{R}^{p+1}$. However, since in (13) the θ_j are associated with $\mathbf{z}_{i,j}$, while the intercept is associated with the *constant* 1, not the entire $\mathcal{R}^{p+1}$ should be searched for a valid solution. The least eigenvector algorithm fails to take this into account.

The poor performance of the least eigenvector algorithm is well know even beyond bilinear problems (e.g. see Faugeras (1993, p. 63) in the context of camera calibration), and often computationally expensive nonlinear techniques were recommended instead. Recently, Hartley (1997a) proposed the *normalized eight-point algorithm* in which before the estimation procedure the points in each image are independently transformed by translation and scaling to have their centroids in the origin and a scatter close to a unit circle or unit box. The nature of the errors $\Delta\mathbf{z}_i$ is not taken into account.

In the context of our model, the significant performance improvement of the normalized relative to the original eight-point algorithm is justified not only by the better condition number of the data matrix as Hartley motivates the need for the transformation. Due to centering the intercept is eliminated while the scaling reduces the influence of measurement errors. In Section 5.1 will show a quantitative example. Thus the least eigenvector algorithm is now applied in conditions closer to its validity, and the results improve. In general, however, using the incorrect model when estimating from data with significant measurement errors yields strongly biased parameter estimates, as will be shown in the experimental sections.

The correct, *total (orthogonal) least squares* (TLS) solution to solve (12) for $\boldsymbol{\mu} = \mathbf{0}$, and $\Delta\mathbf{z}_i \sim \mathrm{GI}(\mathbf{0}, \sigma_\nu^2\mathbf{I})$, i.e., zero mean, homoscedastic errors, is

$$\hat{\alpha} = -\bar{\mathbf{z}}^\top\hat{\boldsymbol{\theta}} \quad \text{where } \bar{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i, \tag{15}$$

and $\hat{\boldsymbol{\theta}}$ is the eigenvector associated with the smallest eigenvalue of the moment matrix of the *centered* data $\mathbf{Z}_c$,

$$\mathbf{M} = \mathbf{Z}_c^\top\mathbf{Z}_c = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^\top. \tag{16}$$

The proof is immediate following through Appendix A for this particular case. The least eigenvector algorithm can be applied only when the errors of the *augmented* data vectors $\mathbf{y} = [1\ \ \mathbf{z}^\top]$ have the *same, singular* covariance matrix

$$\mathbf{C} = \sigma_\nu^2\begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I} \end{bmatrix}. \tag{17}$$

The use of the singular covariance matrix in (12) automatically separates the treatment of the intercept from the rest of the parameters.

Taubin (1991) proposed the following approximation for the mean squared Euclidean distances ($\mathbf{C}_{\mathbf{m}}^{\star} = \mathbf{I}$)

in the minimization criterion (7)

$$\sum_{i=1}^{n}(\mathbf{m}_i - \mathbf{m}_{io})^\top(\mathbf{m}_i - \mathbf{m}_{io}) \approx \frac{\sum_{i=1}^{n}\left(\alpha_o + \mathbf{z}_i^\top \boldsymbol{\theta}_o\right)^2}{\boldsymbol{\theta}_o^\top \mathbf{N}\boldsymbol{\theta}_o}, \tag{18}$$

$$\text{where} \quad \mathbf{N} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \mathbf{z}_i}{\partial \mathbf{m}_i}\right)\left(\frac{\partial \mathbf{z}_i}{\partial \mathbf{m}_i}\right)^\top, \tag{19}$$

is a $p \times p$ positive semi-definite matrix. Note that $\frac{\partial \mathbf{z}_i}{\partial \mathbf{m}_i}$ is the Jacobian of the nonlinear transformation from $\mathbf{m}$ to $\mathbf{z}$ evaluated at $\mathbf{m}_i$, and every element of the sum can be regarded as the normalized covariance matrix of the vector $\mathbf{z}_i$ when neglecting the higher order moments. Thus, $\mathbf{N}$ has the meaning of the average normalized covariance matrix.

The solution through this approximation is the smallest generalized eigenvalue of the generalized eigenproblem

$$\mathbf{M}\hat{\boldsymbol{\theta}} - \lambda\mathbf{N}\hat{\boldsymbol{\theta}} = \mathbf{0} \quad \text{and} \quad \hat{\alpha} = -\bar{\mathbf{z}}^\top\hat{\boldsymbol{\theta}}. \tag{20}$$

The employed approximation (18) is known in the statistical literature as the generalized total least squares (GTLS) problem (Van Huffel and Vandewalle, 1989), where the noisy points in a linear model are assumed to have zero mean and a *common* covariance matrix $\sigma_\nu^2\mathbf{N}$. Although the result is improved relative to that of the eigenvector of $\mathbf{M}$ (16) alone, it is still not optimal. Taubin (1993) used this solution as the input into a nonlinear Levenberg-Marquardt algorithm to solve the weighted least squares problem based on a higher order approximation of the squared Euclidean distance. He also remarked that this initial solution was often unsatisfactory, and the optimization algorithm did not return a useful solution. Experimental results for ellipse fitting using Taubin's method, described in Fitzgibbon et al. (1999), have also shown problems with convergence.

Kanatani (1996, Chap. 9) assumed a small noise model and analyzed the statistical behavior of the errors under this assumption. He has shown that to obtain an unbiased solution to image understanding problems involving the bilinear constraint (up to a second order approximation), the criterion to be minimized has the form

$$\sum_{i=1}^{n} w_i\left(\alpha_o + \mathbf{z}_i^\top\boldsymbol{\theta}_o\right)^2, \tag{21}$$

where w_i are weights depending on the data *and* the unknown parameters $\boldsymbol{\theta}_o$. The errors $\Delta\mathbf{z}_i$ may have nonzero mean and different covariance matrices. The dependence of w_i on the sought parameters implies an iterative procedure, called *renormalization*. At each step the underlying heteroscedastic regression problem is transformed into a generalized eigenvalue problem whose solution is used to define the bias correction for the next step. Both a first order and a second order approximation for the bias correction were developed. The method was tested for conic fitting and motion analysis, see Kanatani (1993, 1994, 1996).

First order renormalization (Kanatani 1996, Sec. 9.4) is an elegant numerical way to compute the generalized eigenvector required for the parameter estimate while simultaneously updating the matrices toward bias removal. The first order renormalization is similar to the GTLS procedure suggested by Taubin (1991). The main difference is that to reduce the bias of the solution Taubin used a second, nonlinear estimation step with the Levenberg-Marquardt routine, while in the renormalization the successive iterations achieve that goal.

The second order renormalization algorithm (Kanatani, 1996, Sec. 9.6) takes into account all the noise related phenomena, up to fourth order moments, and provides a near optimal estimator under the small noise level assumption. However, the procedure starts with a biased initial solution and the correction matrices are evaluated at the noisy data points. These may cause the algorithm not to converge when the noise level becomes significant. Zhang (1997) recently introduced a new version of the second order renormalization where the effect of the fourth order moments is ignored. He has also shown that the original procedure implicitly assumes that all the error vectors have the same norm, an assumption which may not hold for real data. Similar conclusions were reached in Leedan (1997, p. 92).

There are important differences between the algorithm proposed in this paper and renormalization since the iterative minimization is approached differently. Renormalization starts with a biased estimation problem and at each step introduces a correction to reduce the bias. Our method iterates over an approximation of the underlying maximum likelihood (for normal noise) problem. See Chojnacki et al. (1999) for a thorough comparison of the two approaches. At each step renormalization solves the generalized eigenproblem through eigenvectors thus with a numerically sensitive technique, while we use a more robust method, the generalized singular value decomposition. For low to

moderate noise levels the performance of renormalization and HEIV is similar, the differences appear when the initial bias and/or the amount of noise become significant, as will be shown in the experimental sections.

3. Solution for Heteroscedastic Regression

To obtain a satisfactory solution for the linearized model (2), the error characteristics of the model variables $\mathbf{z}_i$ must be taken into account. The corresponding minimization criterion was defined in (12) and it is the approximation (under the mapping from $\mathbf{m}$ to $\mathbf{z}$) of the minimization problem (7). In Section 3.1 an iterative algorithm solving this minimization problem is presented. The method of obtaining the estimates for the data of the *original* problem, i.e., the data correction procedure, is described in Section 3.2.

3.1. Estimation of the Parameters

The model (2) is linear with respect to the parameters $[\alpha_o, \boldsymbol{\theta}_o]$ and the estimators $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$ can be found by solving the minimization problem (12). For the moment it is assumed that the normalized covariance matrices $\mathbf{C}_i^\star$ are known. How to approximate these matrices is discussed in Section 3.1.2.

The parameter estimates, and the corrected data points of the *linear model* are

$$\hat{\alpha} = -(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \hat{\boldsymbol{\theta}}, \quad \hat{\mathbf{z}}_i = \mathbf{z}_i - \boldsymbol{\mu} - \frac{[(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}]\mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}, \tag{22}$$

$$\text{with} \qquad \tilde{\mathbf{z}} = \frac{\sum_{i=1}^n = \dfrac{\mathbf{z}_i}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}}{\sum_{i=1}^n \dfrac{1}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}},$$

and $\hat{\boldsymbol{\theta}}$ is the solution to the equation

$$\mathbf{M}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\theta}} - \mathbf{C}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\theta}} = \mathbf{0} \tag{23}$$

$$\mathbf{M}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \frac{(\mathbf{z}_i - \tilde{\mathbf{z}})(\mathbf{z}_i - \tilde{\mathbf{z}})^\top}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}$$

$$\mathbf{C}(\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n \left[\frac{(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}\right]^2 \mathbf{C}_i^\star. \tag{24}$$

The *weighted* centered moment matrix of the data $\mathbf{M}(\hat{\boldsymbol{\theta}})$, and the *weighted* error covariance matrix $\mathbf{C}(\hat{\boldsymbol{\theta}})$ are both positive semi-definite. The proof is presented in Appendix A.

The minimization criterion (12) under the constraint (13) can be written as a function of $\hat{\boldsymbol{\theta}}$ by replacing the true vectors $\mathbf{z}_{io}$ with the estimated ones $\hat{\mathbf{z}}_i$ which yields

$$\begin{aligned}\hat{J} &= \sum_{i=1}^n \frac{[(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}]^2}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \\ &= \sum_{i=1}^n \frac{\hat{\boldsymbol{\theta}}^\top (\mathbf{z}_i - \tilde{\mathbf{z}})(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \\ &= \hat{\boldsymbol{\theta}}^\top \mathbf{M}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\theta}}.\end{aligned} \tag{25}$$

Solving (23) is the same as minimizing (25) with respect to $\hat{\boldsymbol{\theta}}$, since the left hand side of (23) is the gradient of (25) up to a scale factor. A similar derivation was obtained by Fuller (1987, pp. 217–218) for the regression model, and the solution $\hat{\boldsymbol{\theta}}$ was shown to be a consistent estimator for general error distributions with finite moments up to the fourth order (Fuller, 1987, Theorem 3.1.3). Since any continuous function of a consistent estimator is itself a consistent estimator (Mendel, 1995, p. 96), $\hat{\alpha}$ and $\hat{\mathbf{z}}_i$ are consistent estimators as well.

The noise variance estimate $\hat{\sigma}_\nu^2$ can be obtained from (25) by assuming that the estimation errors are uniformly distributed across the $n - p$ dimensions of the null space of the (full rank) data matrix $\mathbf{Z}^\top$ (Fuller, 1987, p. 243)

$$\hat{\sigma}_\nu^2 = \frac{\hat{\boldsymbol{\theta}}^\top \mathbf{M}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\theta}}}{n - p}. \tag{26}$$

There is no closed form solution to (23) and an iterative procedure must be used, which however, is much simpler than the one needed to solve the original problem (7).

3.1.1. The Iterative Procedure. We use an iterative procedure similar to that suggested by Fuller (1987, p. 218) for the regression model. Starting with a fixed $\mathbf{M}(\hat{\boldsymbol{\theta}})$ and $\mathbf{C}(\hat{\boldsymbol{\theta}})$ evaluated at an initial $\hat{\boldsymbol{\theta}}^{(o)}$, at each iteration the following generalized eigenproblem is solved

$$\mathbf{M}^{(j-1)}\hat{\boldsymbol{\theta}}^{(j)} - \lambda \mathbf{C}^{(j-1)}\hat{\boldsymbol{\theta}}^{(j)} = \mathbf{0}, \tag{27}$$

where $\mathbf{M}^{(j-1)} = \mathbf{M}(\hat{\boldsymbol{\theta}}^{(j-1)})$ and $\mathbf{C}^{(j-1)} = \mathbf{C}(\hat{\boldsymbol{\theta}}^{(j-1)})$. The left side of (27) is a symmetric positive semi-definite pencil (Golub and Van Loan, 1990, p. 471)

and the smallest generalized eigenvector of the matrix pair $[\mathbf{M}^{(j-1)}, \mathbf{C}^{(j-1)}]$ provides a decrease in the value of the minimization criterion (25).

The advantage of this approach is that at each iteration we deal with a symmetric, positive semi-definite eigenproblem that can be solved using techniques with superior numerical behavior. It can be shown that convergence $\hat{\boldsymbol{\theta}}^{(j)} = \hat{\boldsymbol{\theta}}^{(j-1)}$ implies $\lambda = 1$, and thus no data dependent threshold is required (Leedan, 1997, p. 63). Convergence is not guaranteed since (25) is not necessarily a convex function in $\hat{\boldsymbol{\theta}}$. Fuller (1987, p. 218) has shown that the convergence of the iterative procedure is similar to that of the Newton-Raphson method. In practice, even for significant noise levels, after only a few iterations a close to optimal solution is obtained as will be shown in the experimental sections.

Since by definition the matrices $\mathbf{M}^{(j-1)}, \mathbf{C}^{(j-1)}$ are positive semi-definite they can be written as

$$\mathbf{M}^{(j-1)} = \mathbf{Z}_c^{(j-1)\top}\mathbf{Z}_c^{(j-1)}, \quad \mathbf{C}^{(j-1)} = \mathbf{L}^{(j-1)\top}\mathbf{L}^{(j-1)}, \tag{28}$$

where from (22) and (24)

$$\mathbf{Z}_c^{(j-1)} = \left[\frac{\left[\mathbf{z}_1 - \tilde{\mathbf{z}}^{(j-1)}\right]^\top}{\sqrt{\hat{\boldsymbol{\theta}}^{(j-1)\top}\mathbf{C}_1^\star\hat{\boldsymbol{\theta}}^{(j-1)}}}, \ldots, \frac{\left[\mathbf{z}_n - \tilde{\mathbf{z}}^{(j-1)}\right]^\top}{\sqrt{\hat{\boldsymbol{\theta}}^{(j-1)\top}\mathbf{C}_n^\star\hat{\boldsymbol{\theta}}^{(j-1)}}}\right] \tag{29}$$

$$\tilde{\mathbf{z}}^{(j-1)} = \frac{\sum_{i=1}^n \dfrac{\mathbf{z}_i}{\hat{\boldsymbol{\theta}}^{(j-1)\top}\mathbf{C}_i^\star\hat{\boldsymbol{\theta}}^{(j-1)}}}{\sum_{i=1}^n \dfrac{1}{\hat{\boldsymbol{\theta}}^{(j-1)\top}\mathbf{C}_i^\star\hat{\boldsymbol{\theta}}^{(j-1)}}}$$

and $\mathbf{L}^{(j-1)}$ is the Cholesky decomposition (Golub and Van Loan, 1990, p. 146) of $\mathbf{C}^{(j-1)}$.

Thus the generalized eigenproblem (27) has a structure which allows the numerically more robust generalized singular vector (GSVD) decomposition to be used to obtain the solution $\hat{\boldsymbol{\theta}}^{(j)}$ in the jth iteration. See Appendix B for details. Note that the smallest generalized singular value of the matrix pair $[\mathbf{Z}_c^{(j-1)}, \mathbf{L}^{(j-1)}]$ is $\sigma_q^2 = \lambda$ and thus convergence means $\sigma_q^{(j)} = 1$.

The iterative procedure requires an appropriate initial solution $\hat{\boldsymbol{\theta}}^{(o)}$. Assume that the *known* normalized covariance matrices $\mathbf{C}_i^\star$ can be approximated as $\mathbf{C}_i^\star \approx \beta_i \bar{\mathbf{C}}^\star$ where $\bar{\mathbf{C}}^\star$ is an *unknown* covariance matrix and β_i are *unknown* positive factors. The values of β_i and $\bar{\mathbf{C}}^\star$ can be found by minimizing

$$[\beta_i, \bar{\mathbf{C}}^\star] = \arg\min_{\beta_i, \bar{\mathbf{C}}^\star} \sum_{i=1}^n \|\mathbf{C}_i^\star - \beta_i \bar{\mathbf{C}}^\star\|_F^2, \tag{30}$$

where $\|\cdot\|_F$ is the Frobenius norm. The solution of (30) is

$$\bar{\mathbf{C}}^\star = \frac{\sum_{i=1}^n \beta_i \mathbf{C}_i^\star}{\sum_{i=1}^n \beta_i^2}, \quad \beta_i = \frac{\mathrm{tr}\{\mathbf{C}_i^\star \bar{\mathbf{C}}^\star\}}{\mathrm{tr}\{\bar{\mathbf{C}}^{\star 2}\}}. \tag{31}$$

To obtain $\hat{\boldsymbol{\theta}}^{(o)}$, start with all $\beta_i = 1$, compute $\bar{\mathbf{C}}^\star$ and readjust β_i and $\bar{\mathbf{C}}^\star$. Further refinements do not yield improvements in the final result of the estimation. It should be emphasized that the accuracy of the initial solution is not crucial for satisfactory performance, and most often the same result is obtained by employing a random initialization. The above described procedure is needed when the noise variance (26) is to be determined prior to the iterations.

The generalized eigenproblem (23) after some manipulations is reduced to

$$\mathbf{Z}_c(\boldsymbol{\beta})^\top \mathbf{Z}_c(\boldsymbol{\beta})\hat{\boldsymbol{\theta}}^{(o)} - \lambda \bar{\mathbf{C}}^\star \hat{\boldsymbol{\theta}}^{(o)} = \mathbf{0}, \tag{32}$$

where

$$\mathbf{Z}_c(\boldsymbol{\beta}) = \left[\frac{[\mathbf{z}_1 - \tilde{\mathbf{z}}(\boldsymbol{\beta})]^\top}{\sqrt{\beta_1}}, \ldots, \frac{[\mathbf{z}_n - \tilde{\mathbf{z}}(\boldsymbol{\beta})]^\top}{\sqrt{\beta_n}}\right]^\top \tag{33}$$

$$\tilde{\mathbf{z}}(\boldsymbol{\beta}) = \frac{\sum_{i=1}^n \dfrac{\mathbf{z}_i}{\beta_i}}{\sum_{i=1}^n \dfrac{1}{\beta_i}}.$$

Thus, the initial solution $\hat{\boldsymbol{\theta}}^{(o)}$ is obtained from the GSVD of the matrix pair $[\mathbf{Z}_c(\boldsymbol{\beta}), \bar{\mathbf{L}}]$, where $\bar{\mathbf{L}}$ is the Cholesky decomposition of $\bar{\mathbf{C}}^\star$. Note that there are no iterations involved. The initial value of the intercept is

$$\hat{\alpha}^{(o)} = -[\tilde{\mathbf{z}}(\boldsymbol{\beta}) - \boldsymbol{\mu}]^\top \hat{\boldsymbol{\theta}}^{(o)}. \tag{34}$$

In the presence of outliers, high breakdown point estimators such as the LMedS technique (Meer et al., 1991) should be used to obtain the initial solution and reject the outliers from further processing. However, the computation of a model candidate with the elemental subset should employ the above procedure and not use the solution of the linear system of equations, since the latter produces a biased estimate. We return now to the issue of approximating the covariance matrices $\mathbf{C}_i^\star$.

3.1.2. The Approximate Distribution of $\Delta\mathbf{z}_i$. In Section 2.1 we have already shown that the nonlinear mapping between the observed data $\mathbf{m}_i$ (6) and the linearized model variable $\mathbf{z}_i$ makes the error relative to the true vector $\mathbf{z}_{io}$ (9), to have nonzero mean $\boldsymbol{\mu}$ and a covariance matrix $\mathbf{C}_i$ dependent on $\mathbf{C}_\mathbf{m}$ and $\mathbf{m}_{io}$. The example (11) for conic fitting was also given. Similar expressions will be obtained in Section 5 for the epipolar geometry problem for symmetrical, zero-mean distributions of $\Delta\mathbf{m}_i$. While the assumption of normal noise $\Delta\mathbf{m}_i$ is not necessary for the approximations developed in this section, it should be noted that it was used to obtain the analytical expressions of $\mathbf{C}_i^\star$ in the experiments. The general case is discussed in Leedan (1997, p. 67).

It is not possible to obtain $\boldsymbol{\mu}$ and $\mathbf{C}_i^\star$ exactly since $\mathbf{m}_{io}$ and σ_ν^2 have to be estimated from the data. However, the solution of the minimization problem (12) remains close to optimal if the approximations of $\boldsymbol{\mu}$ and $\mathbf{C}_i^\star$ are consistent (Fuller, 1987, Theorem 3.1.1). Two issues have to be addressed when approximating $\boldsymbol{\mu}$ and $\mathbf{C}_i$:

- How to reduce the influence of the fourth order moments of $\Delta\mathbf{m}_i$ on $\mathbf{C}_i^\star$.
- How to estimate $\mathbf{m}_{io}$ and σ_ν^2.

As will be shown in the sequel w.l.g. it can be assumed that $\mathbf{C}_\mathbf{m} = \sigma_\nu^2\mathbf{I}$. In this case, to reduce the effect of the fourth order moments it is required that

$$m_{(j)io}^2 \gg \sigma_\nu^2 \quad j = 1..k \quad i = 1..n, \tag{35}$$

where $m_{(j)io}$ is the jth element in the vector $\mathbf{m}_{io}$. A proper transformation of the data points $\mathbf{m}_i$ can satisfy these requirements without affecting the solution, while preserving the additional restrictions on the matrix $\mathbf{F}$ in (1).

A transformation prior to processing has often been used in computer vision, however, for different reasons. Usually the centroid of the data is translated to the origin of the image coordinates. Although centering the image does not affect the statistical properties of the image points, it can move some of the points too close to the origin. These points have a significantly decreased signal-to-noise-ratio and can degrade the overall performance of the estimation procedure. Often the image coordinates are also scaled down to reduce the condition number of the data matrix $\mathbf{Z}$ (Torr, 1995; Hartley, 1997a).

A general affine transformation of the data $\mathbf{m}_{io}$ is defined as

$$\tilde{\mathbf{m}}_{io} = s\mathbf{T}\mathbf{m}_{io} + \mathbf{a}, \tag{36}$$

where $\mathbf{a}$ is a translation vector, s is a nonzero scaling factor and $\mathbf{T}$ is a nonsingular transformation matrix in $\mathcal{R}^{k\times k}$. Let the estimate for $\mathbf{F}$ obtained from the original data $\mathbf{m}_i$ be $\hat{\mathbf{F}}$, and the one obtained from the transformed data $\tilde{\mathbf{m}}_i$ be $\tilde{\mathbf{F}}$.

It can be easily shown that the minimization criterion (7) is not affected (up to the scaling factor) by the affine transformation. The transformation also preserves the rank of $\mathbf{F}$, and in the case of quadratic forms the value class (e.g., positive definite, indefinite, etc.) and the sign of the determinant as well (Mirsky, 1990, p. 396). Moreover, it can be shown that the rank and the value class of the leading 2×2 submatrix of $\mathbf{F}$ are preserved, thus if $\tilde{\mathbf{F}}$ represents an ellipse so does $\hat{\tilde{\mathbf{F}}}$.

The transformed data points $\tilde{\mathbf{m}}_i$ have the normalized covariance matrix $\tilde{\mathbf{C}}_\mathbf{m}^\star = \mathbf{T}\mathbf{C}_\mathbf{m}^\star\mathbf{T}^\top$, and therefore $\mathbf{T}$ should be chosen to yield $\tilde{\mathbf{C}}_\mathbf{m}^\star = \mathbf{I}$, i.e., to whiten the errors in the image points. In order to find the translation vector $\mathbf{a}$, a trade-off is to be considered. After the isotropic scaling, the data is inside of a unit box in $\mathcal{R}^k$, and the variance of the noise is $\tilde{\sigma}_\nu^2 = s^2\sigma_\nu^2$. To satisfy (35), i.e., none of the transformed coordinates are very close to zero, the vector $\mathbf{a}$ should move the unit box away from the origin in $\mathcal{R}^k$. However, the condition number of the data matrix, constructed with the transformed vectors $\mathbf{z}(\tilde{\mathbf{m}}_i)$, increases with the magnitude of the translation from the origin. To satisfy the trade-off we set $\mathbf{a}$ to move the left bottom corner of the unit box at the point [1, 1].

Since we use the GSVD technique in the computations the sensitivity to the condition number of the data matrix is significantly reduced. Slight performance improvement due to the translation away from the origin was observed in both applications (ellipse fitting, fundamental matrix estimation) for very noisy data (Leedan, 1997, p. 117, 157). To simplify the notations, in the sequel we omit the tilde superscript but imply that all the processing is performed with the transformed data if not specified otherwise.

The second issue regarding the computation of $\mathbf{C}_i$ and $\boldsymbol{\mu}$ is their dependence on the true points $\mathbf{m}_{io}$ and the noise variance σ_ν^2. When random initialization is used, the coordinates of the noisy points and $\sigma_\nu^2 = 0$ should be substituted into the expressions of $\mathbf{C}_i^\star$ and $\boldsymbol{\mu}$, which are then updated after the first iteration.

It should be noted that under practical noise levels, neglecting the fourth order moments in $\mathbf{C}_i^\star$ does not seem to decrease performance. If the initial solution $\hat{\boldsymbol{\theta}}^{(o)}$ is obtained by solving the eigenproblem (32), the initial estimates $\hat{\mathbf{m}}_i^{(o)}$ and $\hat{\sigma}_\nu^2$ can be computed. Note that the covariance matrix $\mathbf{C}^\star$ is the weighted sum of the $\mathbf{C}_i^\star$ (31) and thus using the above substitutions has a lesser effect on the the $\hat{\boldsymbol{\theta}}^{(o)}$. The noise variance estimate $\hat{\sigma}_\nu^2$ is then computed from (26), and $\hat{\mathbf{m}}_i^{(o)}$ is obtained from the first k elements of $\hat{\mathbf{z}}_i^{(o)}$ (22). Using $\hat{\mathbf{m}}_i^{(o)}$ and $\hat{\sigma}_\nu^2$ the covariance matrix $\mathbf{C}_i^\star$ is updated. When this initialization is employed (as was in all our experiments) recomputing $\boldsymbol{\mu}$ and $\mathbf{C}_i^\star$ at each iteration step does not have a significant influence on the final result and the same values were used throughout the algorithm.

3.2. Data Correction

At convergence, the parameter estimation procedure provides the estimates $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$ defining a *plane* in $\mathcal{R}^p$. However, we must satisfy the constraints in the original problem (7), i.e., the points $\mathbf{m}_{io}$ must lie on the *surface* in $\mathcal{R}^k$ defined by $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$.

Substituting the parameter estimates into (22), the new estimate $\hat{\mathbf{m}}_i^{(t)}$ obtained from the first k elements (as we did for the initial solution) is not satisfactory since $\hat{\mathbf{m}}_i^{(t)}$ is not a consistent estimator for $\mathbf{m}_{io}$ and it will not necessarily be on the surface. The estimators $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$ are consistent under very mild conditions (See Section 3.1) and the final estimate for the data point $\hat{\mathbf{m}}_i$ should be obtained through the data correction procedure. Using a standard nonlinear optimization routine to solve the minimization problem (8) for $\hat{\mathbf{m}}_i$ with the constraint $\hat{\alpha} + \mathbf{z}(\hat{\mathbf{m}}_i)^\top \hat{\boldsymbol{\theta}} = 0$ is computationally expensive.

A simpler approach exploits the fact that an initial estimate $\hat{\mathbf{m}}_i^{(t)}$ is already available. Since the error $\Delta\mathbf{m}_i$ has identity normalized covariance matrix (recall the affine transformation of the data), the estimated error vector $\Delta\hat{\mathbf{m}}_i = \mathbf{m}_i - \hat{\mathbf{m}}_i$ has to be collinear with the normal to the surface in $\mathcal{R}^k$ evaluated at the unknown $\hat{\mathbf{m}}_i$. An approximation to the normal can be evaluated at $\hat{\mathbf{m}}_i^{(t)}$, and the obtained expression for $\hat{\mathbf{m}}_i$ is substituted into the constraint

$$\hat{\mathbf{m}}_i = \mathbf{m}_i - \delta \left(\frac{\partial \mathbf{z}_i}{\partial \mathbf{m}_i} \bigg|_{\hat{\mathbf{m}}_i^{(t)}} \right)^\top \hat{\boldsymbol{\theta}}, \quad \hat{\alpha} + \mathbf{z}(\hat{\mathbf{m}}_i)^\top \hat{\boldsymbol{\theta}} = 0. \tag{37}$$

For the bilinear problem the resulting quadratic equation in δ has either one, two or no real roots. The minimum magnitude δ is the desired solution. Experimental results have shown that this procedure yields good approximation even for significant noise levels. For the few nonregular points, where no real δ is obtained, $\hat{\mathbf{m}}_i$ can be computed using a standard nonlinear optimization routine with $\hat{\mathbf{m}}_i^{(t)}$ as the initial solution. This case, however, was never encountered in our experiments. In Taubin (1993) a similar approach was used to approximate the absolute value of δ, and solve the equation for the minimum $|\delta|$.

3.3. Algorithm for Estimation with Bilinear Constraints

We summarize now the processing steps described in the previous sections. The concatenated measurements $\mathbf{m}_i$ are assumed to have a common covariance matrix $\mathbf{C}_\mathbf{m} = \sigma_\nu^2 \mathbf{C}_\mathbf{m}^\star$.

Data Transformation

DT.1. Apply an affine transformation to the data $\mathbf{m}_i$. After the transformation the data $\tilde{\mathbf{m}}_i$ is, in each image, inside a unit box with the left bottom corner at $[1,\ 1]$ and has the normalized covariance $\tilde{\mathbf{C}}_\mathbf{m}^\star = \mathbf{I}$. The following steps refer to the transformed data and omit the tilde superscript for convenience.

Linearization of the Problem

LP.1. Define the nonlinear mapping of $\mathbf{m}$ into $\mathbf{z}(\mathbf{m})$.
LP.2. Compute the analytical expression of the covariance matrices $\mathbf{C}_i$.

Initial Estimates [Optional]

IE.1. Compute $\mathbf{C}_i^{(o)}$, β_i and $\bar{\mathbf{C}}^\star$.
IE.2. Find $\hat{\boldsymbol{\theta}}^{(o)}$ by applying GSVD to the matrix pair $[\mathbf{Z}_c(\beta),\ \bar{\mathbf{L}}]$.
IE.3. Compute $\hat{\sigma}_\nu^2$ and $\hat{\mathbf{m}}_i^{(o)}$.
IE.4. Compute $\boldsymbol{\mu}$ and $\mathbf{C}_i^\star$.

Parameter Estimation

PE.1. Start iterations. Set $j = 1$.
PE.2. Find $\hat{\boldsymbol{\theta}}^{(j)}$ by applying GSVD to the matrix pair $[\mathbf{Z}_c^{(j-1)}, \mathbf{L}^{(j-1)}]$ computed using the current solution $\hat{\boldsymbol{\theta}}^{(j-1)}$.
PE.3. Repeat Step PE.2 for $j = 2, 3 \ldots$ until $|\sigma_q^{(j)} - 1|$ is less than tolerance.
PE.4. Return $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(j)}$ and compute $\hat{\alpha}$.

Restriction Enforcement

RE.1. Apply additional restrictions on the parameters $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$ if required by the geometric constraints of the original problem.

Data Correction

DC.1. Find $\hat{\mathbf{z}}_i$, $i = 1 \ldots n$, and obtain $\hat{\mathbf{m}}_i^{(t)}$ from its first k elements.

DC.2. Compute $\hat{\mathbf{m}}_i$ by solving the quadratic equation in the distance δ. If no solution is available for a particular point, apply a standard nonlinear optimization routine.

Back Transformation

BT.1. Transform back the parameters $[\hat{\alpha}, \hat{\boldsymbol{\theta}}, \hat{\sigma}_\nu^2]$ and the corrected data points $\hat{\mathbf{m}}_i$ to the original image coordinate systems.

The order in which the last three procedures, restriction enforcement, data correction and back transformation are applied should be further investigated. Hartley (1997a) has found in the context of the eight-point algorithm that restriction enforcement preceding back transformation gave the best performance for the employed image. However, if data correction is also taken into consideration a more complex picture emerges in which the effect of back transformation can reduce the impact of data correction.

Some important issues should be noted. By processing the intercept $\hat{\alpha}$ separately from the parameter vector $\hat{\boldsymbol{\theta}}$ the covariance matrices $\mathbf{C}_i^\star$ and $\mathbf{C}(\boldsymbol{\theta})$ are nonsingular if $\mathbf{C}_\mathbf{m}^\star$ is nonsingular. Also, the parameter estimation problem is reduced by one dimension. It is evident from (22) that due to the nonlinearity of the mapping process, the centroid vector of the data matrix $\mathbf{Z}$ is *not* a point on the estimated hyperplane $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$ in $\mathcal{R}^p$. Thus, a simple centering of the data will not provide a good estimate for the intercept.

For high noise levels very rarely the algorithm may not converge and the value of the minimization criterion starts to increase. In such cases a standard optimization technique should be applied to reach the nearest local minimum. Experimental results have shown that this occurs when a degenerate solution is obtained which cannot account for all the data points. The problem of degeneracy is not addressed in this paper, however, was already analyzed in the context of the epipolar geometry (Torr, 1995; Luong and Faugeras, 1996). Another possible computational "trick" is to recompute the vectors $\hat{\mathbf{z}}_i^{(j)}$ and update the covariance matrices $\mathbf{C}_i$ whenever the difference between $\hat{\boldsymbol{\theta}}^{(j)}$ and $\hat{\boldsymbol{\theta}}^{(j-1)}$ becomes large.

When solving a computer vision problem, the estimated parameter matrix $\hat{\mathbf{F}}$ and hence $[\hat{\alpha}, \hat{\boldsymbol{\theta}}]$ must satisfy additional restrictions reflecting the underlying geometrical properties. The algorithm can be easily modified to incorporate linear restriction functions. See Leedan (1997, Sec. 3.4.4) for the discussion of several approaches. However, most restrictions are nonlinear and solving the minimization problem in (12) with additional nonlinear restrictions is not straightforward. We have adopted the strategy of obtaining an unrestricted $\hat{\mathbf{F}}$ and then readjust the parameters to obey the restrictions. In the remaining sections we discuss two applications, ellipse fitting and estimation of the fundamental matrix, in both of which additional restrictions have to be imposed.

The algorithm as described above is not limited to solving problems with bilinear constraints since it is a general approach to heteroscedastic EIV regression. Only the linearization and the restriction enforcement steps are problem specific. The algorithm was implemented in MATLAB and the code is available (for a more general case) at
`www.caip.rutgers.edu/riul/research/code.html`

4. Ellipse Fitting

The problem of fitting an ellipse to noisy data points is considered first. The problem was extensively analyzed in the computer vision literature, however, most of the proposed techniques provide biased solutions. Recent reviews can be found in Cabrera and Meer (1996), Zhang (1997) and Fitzgibbon et al. (1999). In Cabrera and Meer (1996) a bias removal technique of different nature was also proposed.

4.1. Background and Employed Methods

To have the quadratic form (4) represent an ellipse, the parameter matrix $\mathbf{F}$, hence the parameters $[\alpha, \boldsymbol{\theta}]$, must satisfy an additional, nonlinear restriction of having a negative value for the discriminant $c^2 - 4de$. The restriction can be written as

$$\hat{\boldsymbol{\theta}}^\top \mathbf{D} \hat{\boldsymbol{\theta}} = 1, \quad \text{where} \quad \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 & 0 \end{bmatrix}. \quad (38)$$

Note that **D** is an indefinite matrix with signature (1,2), i.e., one positive and two negative eigenvalues: 2, −1, −2. In Section 3.1.2 it was discussed that the affine transformation does not affect the imposed restrictions and the transformed parameters still represent an ellipse.

The simplest strategy is of obtaining an unrestricted $\hat{\boldsymbol{\theta}}$ from the linearized model and whenever the result is not an ellipse the restriction (38) is enforced on the solution yielding the new $\hat{\boldsymbol{\theta}}$. Since this case appears only very infrequently, we have used the direct least squares ellipse fitting method proposed in Fitzgibbon et al. (1999). This method being of the TLS class it is biased, as its developers also recognized it. We did center the data first, i.e., separated the intercept.

The performance of three different estimation techniques were compared for the task of ellipse fitting.

TLS – The total least squares solution obtained from the generalized eigenvector of **S** (14) and **D** (38), as it was implemented in Fitzgibbon et al. (1999). The noisy data is scaled down to a unit box and centered at the origin. Thus this method can be also seen as an implementation of Hartley's (1997a) proposal to improve on the performance of the eight-point algorithm.

KAN – The solution obtained by Kanatani's second order renormalization (Kanatani, 1994). The noisy data is scaled down to a unit box and centered at the origin.

EIV – The method proposed in Section 3.1. The noisy data is scaled down to a unit box with the lower left corner at [1, 1]. The ellipse restriction was imposed after transforming back the unrestricted estimator $\hat{\boldsymbol{\theta}}$. The expressions of $\boldsymbol{\mu}$ and $\mathbf{C}_i$ were given in (10) and (11).

All the methods were implemented in MATLAB. Experiments were performed with both synthetic and real data.

4.2. *Experiments with Synthetic Data*

The synthetic data was based on a canonical ellipse with

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 10000 \end{bmatrix}. \tag{39}$$

The ellipse is centered at the origin and the major and the minor radii are 100 and 50, respectively. To increase the bias present in the TLS method, the noisy points were always taken from a segment spanning only one quarter of the ellipse. The segment is represented by 40 points and its location on the ellipse was randomly chosen. The coordinates of the points were corrupted with zero-mean normal noise, $\mathbf{C_m} = \sigma_\nu^2 \mathbf{I}$, and tests were performed for different values of σ_ν. The number of points was chosen to be relative small in order to reproduce real situations. In images, after outlier rejection, the size of the remaining sample is often significantly reduced and the performance of the estimation process deteriorates due to the small sample effect.

The location of the data points on the ellipse is also an important factor affecting performance. When the points belong to a high curvature section (Fig. 1(a)) the estimation process is much more stable than when the points are located on a low curvature part (Fig. 1(b)).

For each experimental condition (the value of σ_ν and the employed estimation technique) 200 trials were run. Since these trials cover all possible locations on the

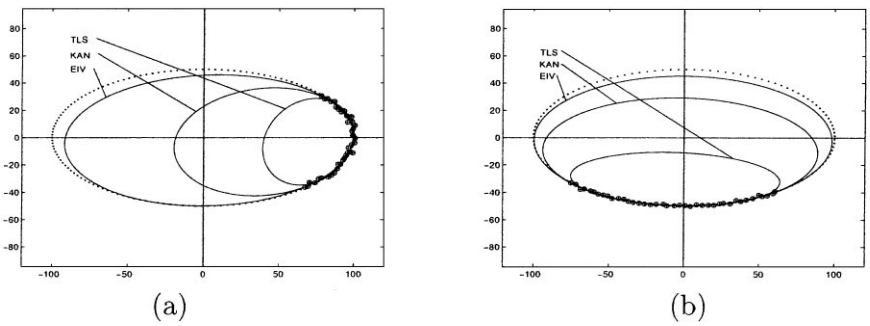

Figure 1. The two extreme cases for the data. (a) Best: 40 noisy points on the high curvature section. (b) Worst: 40 points on the low curvature section. The dotted line represents the true ellipse. The three different estimation techniques label the corresponding ellipses.

ellipse, the results can be regarded as the "expected" performance for data covering one fourth of the ellipse. The results can also be seen as a lower bound on the performance. In practice the data comes from longer segments (albeit not necessarily more points) which can only improve performance for similar operating conditions.

The results for $\sigma_\nu \leq 0.05$ were essentially the same for all three methods. For $0.05 \leq \sigma_\nu \leq 0.1$ the two bias correcting methods, KAN and EIV, produced similar results, both superior to the TLS method. Figure 2 shows the scatter plots of the major and minor radii of the estimated ellipses for larger noise values, $\sigma_\nu = 0.5$, 1, 1.5, and 2 with rounding of the corrupted value to the closest integer (data on discrete grid). The axes have log-log scale to show all the results. The points in the negative quadrant in Fig. 2 for KAN method indicate the cases when the solution converged

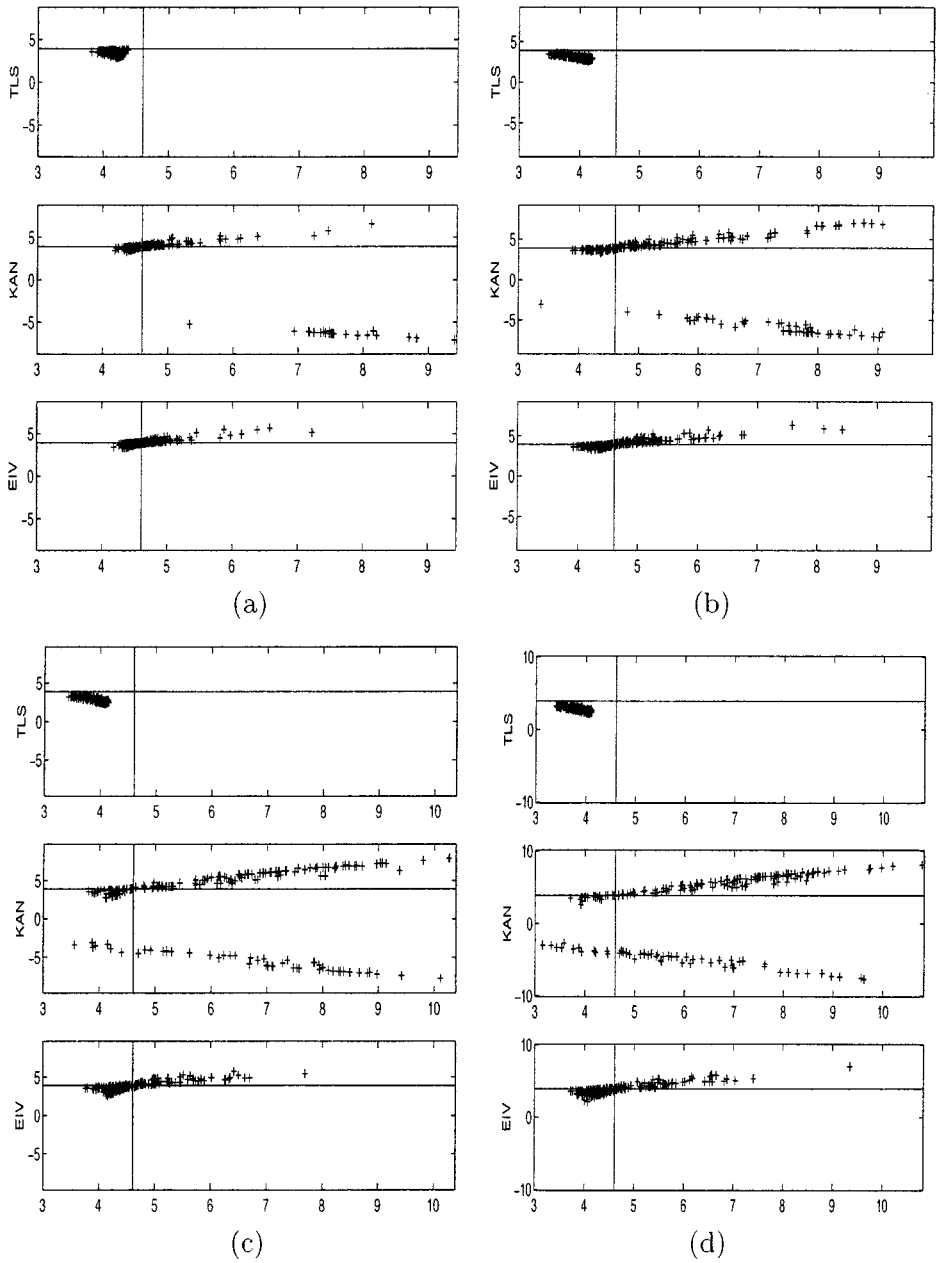

Figure 2. Ellipse fitting, synthetic data. The scatter (log-log) plots of the major and minor radii. (a) $\sigma_\nu = 0.5$. (b) $\sigma_\nu = 1$. (c) $\sigma_\nu = 1.5$. (d) $\sigma_\nu = 2$ with rounding. The true values are shown with the crossed lines.

to a hyperbola. The EIV produced much less hyperbolas (as will be shown) and they were converted into ellipses using the method mentioned at the beginning of this section. Note the strong bias of the TLS and the large spread of KAN for significant noise levels.

To obtain another quality measure for the results, the estimated data points $\hat{\mathbf{m}}_i$ were computed at each trial. In order to facilitate comparison across the three different estimation methods a standard nonlinear optimization tool (the constr function in the Optimization Toolbox of MATLAB) was used to compute $\hat{\mathbf{m}}_i$ after the parameters were estimated. However, when for the EIV method the estimates $\hat{\mathbf{m}}_i$ are computed with the data correction procedure described in Section 3.2 similar results were obtained at all noise levels. Figure 3 shows the histograms of the average squared distances between $\hat{\mathbf{m}}_i$ and the noisy points $\mathbf{m}_i$. The histograms for the distances between $\hat{\mathbf{m}}_i$ and the true points $\mathbf{m}_{io}$

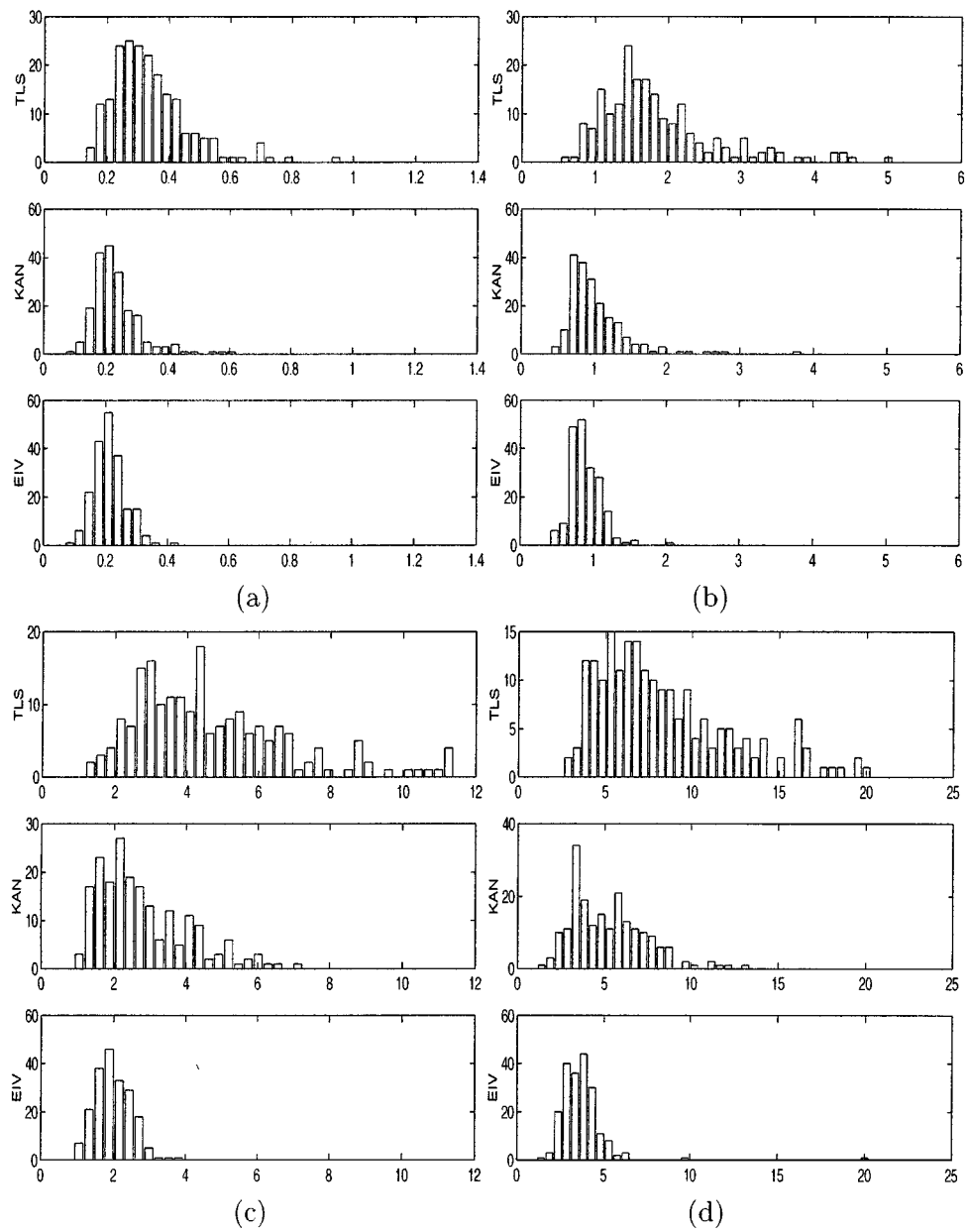

Figure 3. Ellipse fitting, synthetic data. The histograms of the average squared distances between the noisy points and the estimated ones. (a) $\sigma_\nu = 0.5$. (b) $\sigma_\nu = 1$. (c) $\sigma_\nu = 1.5$. (d) $\sigma_\nu = 2$ with rounding.

Table 1. Ellipse fitting, $\sigma_\nu = 1$. Median and (standard deviation) for 200 trials.

Method	Major	Minor	Angle	Center		Distance	Conv.	Steps	Hyp.
True	100	50	0	[0,0]					
TLS	54(10)	21(6)	1.5(24)	[−3,0]	([47,23])	1.65(0.8)	200	1	0
KAN	191(2800)	50(391)	0.1(37)	[−8,−1.5]	([1600,2900])	0.91(0.4)	39	7.3	47
EIV	104(420)	52(58)	0.4(13)	[−1,−1]	([396,175])	0.86(0.2)	197	4.1	2[a]

[a]See text.

were also generated (not shown) and were very similar to Fig. 3 indicating the adequacy of the minimization criterion (12).

The EIV method is the most stable, yielding the narrowest spread and the smallest average squared distances. The improvement relative to KAN method is due to more accurate modeling of the error and the better numerical behavior of the GSVD employed in the iterations. At high noise levels, a few trials yielded degenerate cases in which the data points were distributed across the two branches of a hyperbola. When such a hyperbola is converted into an ellipse very large average squared distance is obtained. See the right side of Fig. 2(d), EIV.

The median of the estimated ellipse parameters (the major and minor radii, the center and the orientation of the ellipse), as well as the median of the average squared distances for each estimation method are shown in Table 1 for the case $\sigma_\nu = 1$. The spread across the trials is measured by the standard deviations shown in parentheses. For KAN, only the cases yielding ellipses were considered.

Several parameters measuring the behavior of the iterative procedures are also shown in Table 1. The 'convergence' column indicates the number of trials in which the algorithm converged, and the 'steps' column has the average number of iterations till convergence. The processing time in MATLAB of an iteration for KAN and EIV are roughly the same. The 'hyperbola' column shows the number of trials where the result was a hyperbola.

The TLS method always provides strongly biased results. The estimated ellipses have parameters with small standard deviations but they provide the worst fit. Since the range of the data was the same (on a quarter segment of the ellipse) the TLS estimator produces similarly sized ellipses. Their size agrees with the prediction of the bias towards smaller eccentricities (Kanatani, 1994). The quality of the TLS results suggests that they are not adequate to be used as initial solutions.

The KAN algorithm did not converge in most of the cases when $\sigma_\nu \geq 1$. This value, in fact, exceeds the noise levels Kanatani used in his examples (Kanatani, 1994). However, the KAN method still produced reasonable results after we have modified it to abort whenever the value of the minimization criterion started to increase. In 47 trials the result was a hyperbola and the statistics in Table 1 thus corresponds to the 153 trials yielding ellipses. The histogram of the average squared distances (Fig. 3(b), KAN) remained practically the same when the hyperbola cases were excluded. This is not unexpected since the hyperbolas are valid, unrestricted solutions of the minimization problem.

The EIV estimation method produced a smaller number of hyperbolas and practically always converged. All the hyperbolas were converted into ellipses and are included in Table 1. Although converting a hyperbola into an ellipse yields in general a deterioration in the quality of the fit (since the hyperbola is the optimal fit for the given data points) there were no significant differences in the average squared distances.

The estimated fit minimizes the Euclidean distances ($\mathbf{C}^\star_\mathbf{m} = \mathbf{I}$) to the observed data. However, since the data comes from a small segment of the ellipse the fit may appear very different from the true ellipse, illustrating the ill-posed nature of such estimation problems. In Fig. 4(a) an example with a one quarter segment of an ellipse is shown ($\sigma_\nu = 0.5$). The ellipse estimated by the EIV method is shown in Fig. 4(b) with the dashed-dotted line. The averaged squared distance of the estimated ellipse is 0.2212, which is close to the median of the histogram in Fig. 3(a), EIV. The ellipse segment was used to generate 100 noisy ($\sigma_\nu = 0.5$) data sets, and for each the EIV method was applied. The two ellipses shown in Fig. 4(b) with solid lines have the same average squared distances 0.2203 and 0.2204, respectively. The results show the inadequacy of small segments when a valid fit is sought for the entire data space.

For a more detailed discussion of the experimental data see Leedan, (1997, Sec. 4.1.5).

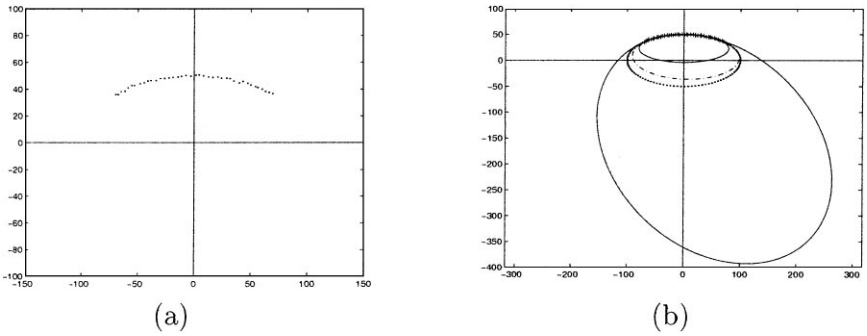

Figure 4. The ill-posed nature of fitting to small ellipse segments. (a) An example of a noisy quarter ellipse. (b) The true (dotted) the original EIV estimate (dash-dotted) and two extreme ellipses (solid).

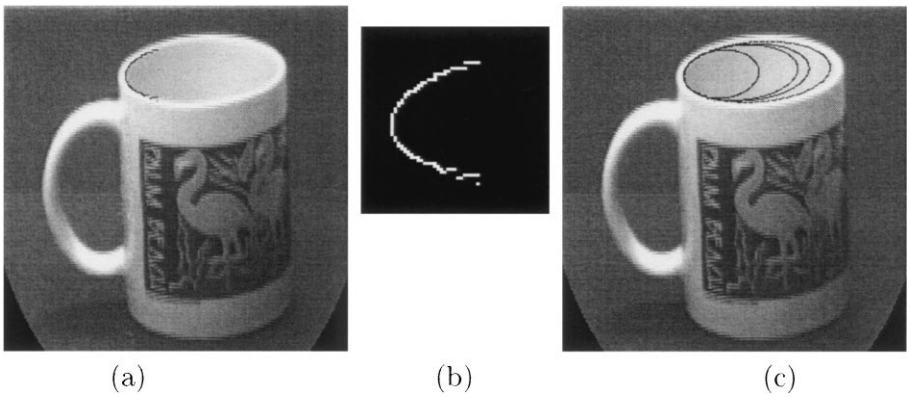

Figure 5. Ellipse fitting, real data. (a) The coffee mug image. The employed edge pixels are marked with black. (b) The input to the estimation algorithm. (c) Four estimated ellipses. From left to right: TLS, KAN, EIV and EIV-Cm.

4.3. Experiments with Real Data

As a real data example, the image of the coffee mug (Fig. 5(a)) was used. A small segment from the edge image was selected in order to estimate the inner elliptical rim of the mug. The 83 pixels of the employed data are shown in Fig. 5(b), and the results for each of the estimation methods are overlaid in Fig. 5(c). The TLS method produced the smallest ellipse, KAN the next, and the EIV method the third ellipse from the left.

The image of the coffee mug was part of a larger scene and was magnified after being "cut out". This introduced a significant aliasing clearly visible in the gray level image (Fig. 5(a)). The aliasing, together with the digital geometry of the edge image, introduced correlation between the two image coordinates of a data point. To take the correlation into account the covariance matrix of the image points was modified from $\mathbf{C}_{\mathbf{m}}^{\star} = \mathbf{I}$ to

$$\mathbf{C}_{\mathbf{m}}^{\star} = \begin{bmatrix} 1 & 0 \\ 0 & 0.25 \end{bmatrix}. \tag{40}$$

This change had a significant effect on the result. The new estimate (labeled EIV-Cm) is the rightmost ellipse in Fig. 5(c). Note that the employed covariance matrix $\mathbf{C}_{\mathbf{m}}^{\star}$ is only a coarse approximation of the error process. The data in Fig. 5 is typical for an edge image, which indicates the importance of using nonidentity covariance matrices for the pixels.

5. Estimation of the Fundamental Matrix

The problem of estimating the epipolar geometry of two uncalibrated cameras, i.e., the fundamental

matrix, is a central problem in 3D reconstruction (Faugeras, 1995). Various estimation methods were proposed, and excellent reviews can be found in Luong and Faugeras (1996), Torr and Murray (1997) and Zhang (1998a).

5.1. Background and Employed Methods

To have the bilinear form (1) represent the epipolar constraint between the matched points $\mathbf{p}_{Aio} = [x_{Aio}, y_{Aio}, 1]^\top$ and $\mathbf{p}_{Bio} = [x_{Bio}, y_{Bio}, 1]^\top$, in two images captured without any knowledge about the internal or external parameters of the two cameras, the matrix $\mathbf{F}$ must have rank two. Often the rank restriction is enforced by applying the matrix approximation theorem (Van Huffel and Vandewalle, 1991, p. 31). By this theorem, replacing with zero the smallest singular value of the matrix estimated from the data (which in general has the full rank 3) yields the optimal solution for a large class of matrix norms. It is important to be aware that the approximation theorem is based on the assumption that the elements of the estimated matrix are uncorrelated and have the same variance. This is not true in our case since the elements are derived from $\hat{\alpha}$ and $\hat{\boldsymbol{\theta}}$. It is possible to define a more general rank reduction procedure which takes into account the covariance matrix of the parameter estimates (Leedan, 1997, p. 132).

The fundamental matrix can be parametrized in several different ways (Luong and Faugeras, 1996). To compare the quality of different estimates we have used the two epipoles $\mathbf{e}_A$ and $\mathbf{e}_B$, defined as the null-spaces of the restricted, rank two matrix

$$\mathbf{F}\mathbf{e}_A = \mathbf{0}, \quad \mathbf{F}^\top \mathbf{e}_B = \mathbf{0}, \tag{41}$$

and the homography transformation defined by $\mathbf{H}$, the leading 2×2 submatrix of $\mathbf{F}$. From the singular value decomposition

$$\mathbf{H} = \mathbf{U}_H \begin{bmatrix} \sigma_{H_1} & \\ & \sigma_{H_2} \end{bmatrix} \mathbf{V}_H^\top. \tag{42}$$

The homography is parametrized by the rotation angles of the orthogonal matrices $\mathbf{U}_H, \mathbf{V}_H$ and the ratio of the two singular values $\frac{\sigma_{H_2}}{\sigma_{H_1}}$.

The mapping from the epipolar constraint to the linearized model (2) is defined by

$$\begin{aligned} \mathbf{z}_{io} = [&x_{Aio},\ y_{Aio},\ x_{Bio},\ y_{Bio},\ x_{Aio}x_{Bio}, \\ &x_{Aio}y_{Bio},\ x_{Bio}y_{Aio},\ y_{Aio}y_{Bio}]^\top \end{aligned} \tag{43}$$

and

$$\begin{aligned} \boldsymbol{\theta}_o &= [f_{31},\ f_{32},\ f_{13},\ f_{23},\ f_{11},\ f_{21},\ f_{12},\ f_{22}]^\top, \\ \alpha_o &= f_{33}. \end{aligned} \tag{44}$$

If normal errors are assumed for the image coordinates, i.e., $\Delta\mathbf{m}_i \sim \mathrm{NI}(\mathbf{0}, \sigma_\nu^2\mathbf{I})$, the mean $\boldsymbol{\mu} = \mathbf{0}$ and the normalized covariance matrix $\mathbf{C}_i^*$ of the error vector $\Delta\mathbf{z}_i$ is:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & x_{Bio} & y_{Bio} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & x_{Bio} & y_{Bio} \\ 0 & 0 & 1 & 0 & x_{Aio} & 0 & y_{Aio} & 0 \\ 0 & 0 & 0 & 1 & 0 & x_{Aio} & 0 & y_{Aio} \\ x_{Bio} & 0 & x_{Aio} & 0 & x_{Aio}^2 + x_{Bio}^2 + \sigma_\nu^2 & x_{Bio}y_{Bio} & x_{Aio}y_{Aio} & 0 \\ y_{Bio} & 0 & 0 & x_{Aio} & x_{Bio}y_{Bio} & x_{Aio}^2 + y_{Bio}^2 + \sigma_\nu^2 & 0 & x_{Aio}y_{Aio} \\ 0 & x_{Bio} & y_{Aio} & 0 & x_{Aio}y_{Aio} & 0 & x_{Bio}^2 + y_{Aio}^2 + \sigma_\nu^2 & x_{Bio}y_{Bio} \\ 0 & y_{Bio} & 0 & y_{Aio} & 0 & x_{Aio}y_{Aio} & x_{Bio}y_{Bio} & y_{Aio}^2 + y_{Bio}^2 + \sigma_\nu^2 \end{bmatrix} \tag{45}$$

The expression of $\mathbf{C}_i = \sigma_\nu^2 \mathbf{C}_i^*$ is also a good approximation for noise with symmetric p.d.f. (Leedan, 1997, p. 67).

The performance of five different estimation techniques were compared.

HAR – The normalized eight-point algorithm of Hartley (1997a). In both images the noisy data is isotropically scaled down to a unit box and centered at the origin. The solution is the eigenvector of $\mathbf{S}$ (14). The importance of scaling was investigated in a small pilot experiment. For an image the covariances (45) were averaged for 50 noisy points before and after normalization. The singular values of the average covariance

matrix for the original data were

$$10^5[2.508 \quad 1.355 \quad 1.159 \quad 0.006 \quad 0 \quad 0 \quad 0 \quad 0]$$

while after normalization they became

$$[1.32 \quad 1.18 \quad 1.17 \quad 1.03 \quad 1 \quad 1 \quad 1 \quad 1].$$

Thus normalization not only improved the condition number (the effective rank in the first case is only 3), but also yields matrices whose average is close to identity. The result supports the conjecture in Section 2.2 that the main effect of normalization is to make the linearized data more suitable for TLS estimation.

MOR – The virtual parallax solution suggested by Boufama and Mohr (1995) uses projective planar coordinates. The only transformation performed on the data is the parallax transformation which is based on four image points which should be located far from each other. The errors have now a very complicated structure. Zhang (1998a) discusses some of the theoretical issues related to the optimality of the estimation. A similar method, using 3D projective coordinates was proposed by Ponce and Genc (1996).

KAN – The solution obtained by Kanatani's second-order renormalization. The noisy data is isotropically scaled down to a unit box and centered at the origin.

NON – The nonlinear solution using an optimization routine (the fminu routine in MATLAB Optimization Toolbox) that minimizes the sum of squared distances between the noisy points and the epipolar lines in each image (Luong and Faugeras, 1996). The initial solution is the TLS solution. No transformation is performed on the data.

EIV – The method proposed in Section 3.1. The noisy data is scaled down to a unit box in each image with the same scaling factor. The bottom left corner of the unit box is then positioned at [1,1].

The techniques were implemented in MATLAB. The rank-two restriction is enforced for all the methods except MOR, after back transforming to the original coordinate system. We have used the general rank reduction procedure (Leedan, 1997, p. 132) for KAN and EIV. Experiments were performed with both synthetic and real data.

5.2. Experiments with Synthetic Data

The synthetic data was generated from 3-D points that are randomly scattered in a cube $[2500 \times 2500 \times 2500]$ with the lower left corner in the origin of the world coordinate system. Both cameras had focal length of 300. The first one was located at the origin of the coordinate system, the second one at $[500,\ 100,\ 200]$ and had a 3-D rotation $[5^\circ,\ 5^\circ,\ -30^\circ]$. The true fundamental matrix is

$$\mathbf{F} = \begin{bmatrix} 0.0025 & -0.0049 & 1 \\ 0.0066 & 0.0024 & -4.1059 \\ -2.8475 & 3.2885 & -134.11 \end{bmatrix}, \qquad (46)$$

and the corresponding epipoles are (in image coordinates)

$$\mathbf{e}_A = [460.79 \quad 439.78]^\top \quad \mathbf{e}_B = [750 \quad 150]^\top$$

the singular value ratio is $s_H = 0.75$, the rotation angles are $\mathrm{rot}_1 = 10.14^\circ$, $\mathrm{rot}_2 = 102.8^\circ$.

The coordinates of the points in each image were corrupted with zero-mean normal noise, $\mathbf{C_m} = \sigma_\nu^2 \mathbf{I}$, and tests were performed for different values of σ_ν. Forty points present in both images were chosen as matched pairs (Fig. 6). For each experimental condition (the value of σ_ν and the employed estimation technique) 200 trials were run. We have preferred this type of Monte-Carlo experiment, instead of randomizing the matched points configuration, in order to avoid the degenerate cases.

The results for $\sigma_\nu \le 0.05$ were essentially the same for all five methods. For $0.05 \le \sigma_\nu \le 0.5$ the bias correcting methods, KAN, NON and EIV, produced similar results, all of which were better than the HAR and MOR methods. Detailed investigation for noise levels $\sigma_\nu = 1$, 2 and 4 with rounding are described in Leedan (1997, Sec. 4.2.5). In Fig. 7 the results for the largest noise level are presented. Note the strong bias of the HAR and the large spread of MOR for the epipole estimates. This spread may be also due to the random selection of the bases of the parallax transformation since we did not implement the "good basis" search procedure suggested in Ponce and Genc (1996).

The median of the parameters measuring the behavior of the iterative procedures are shown in Table 2 together with the standard deviation. The processing time in MATLAB of an iteration for KAN and EIV are roughly the same. The NON method is measured by the number of function calls to the minimization

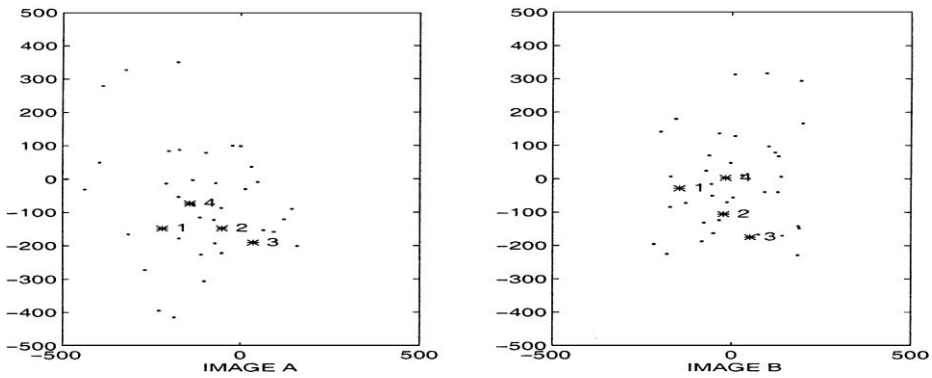

Figure 6. The two views of 40 points used in the experiments. Four matched points are labeled.

function where every 10 function calls are computationally equivalent to one EIV iteration. It is worth emphasizing that although the NON method uses a first-order approximation to the distances it still requires large amount of computation time.

The KAN method did not converge in most of the cases, and the results are obtained after we have modified the KAN algorithm to exit whenever the value of the minimization criterion starts to increase. Although the HAR method has a relative small standard deviation for the epipole estimates, these estimates are strongly biased. The EIV method has the results with the smallest average squared distance except the NON method.

The effect of transformation of the data was also investigated. We have found that it had a weak influence on the EIV results, but as expected a strong one on the HAR results. The KAN method performed slightly better without transformation probably because the influence of higher order moments is larger when the data is centered at the origin.

5.3. *Experiments with Real Data*

We have used the "standard" images for fundamental matrix estimation: *LIFIA-house, bridge, tribunal*. They were kindly provided, together with the ground truth for the last two, by Roger Mohr. These images have very low noise level and thus all the methods produced similar results. In Table 3 the results for the *bridge* image are shown.

The KAN method did not converge. Since the value of the minimization criterion increased immediately, the KAN method executed only one iteration thus yielding almost the same result as the HAR method (recall that the initial solution for the renormalization procedure is the TLS solution). In Fig. 8 the ground truth and the EIV estimates are used to draw the epipolar lines in the two images.

6. Discussion

The main concern raised against linearization is that the geometric nature of the original problem is not considered. The quantity

$$r_i = \mathbf{p}_{Ai}^\top \mathbf{F} \mathbf{p}_{Bi} = \alpha_o + \mathbf{z}(\mathbf{m}_i)^\top \boldsymbol{\theta}_o \tag{47}$$

is known as the *algebraic distance*, and minimizing $\sum_{i=1}^n r_i^2$ leads to the least eigenvector algorithm (see Section 2.2). Hartley (1997a, 1997b) has shown that by normalizing the data a performance comparable

Table 2. Epipolar geometry, $\sigma_\nu = 4$ with rounding. Median and (standard deviation) for 200 trials.

	Epipole 1		Epipole 2		H. Ratio	Distance	Conv.	Steps
True	[460.8, 439.8]		[750, 150]		0.75	0		
HAR	[399.5, 402.5]	([65.7, 43.1])	[678.0, 148.9]	([77.7, 13.3])	0.76 (0.03)	14.4 (4.4)	200	1
MOR	[503.3, 466.8]	([147.2, 115.7])	[782.9, 146.0]	([112.5, 16.7])	0.73 (0.06)	37 (82)	200	1
KAN	[430.8, 424.3]	([77.3, 49.6])	[714.5, 149.7]	([89.8, 13.7])	0.76 (0.03)	14.2 (4.5)	30	4.07
NON	[458.6, 437.1]	([59.6, 36.0])	[743.2, 149.8]	([77.6, 10.2])	0.75 (0.03)	12.7 (3.2)	200	1110
EIV	[462.7, 437.4]	([68.9, 45.4])	[748.1, 150.8]	([82.2, 14.2])	0.75 (0.03)	13.6 (4.1)	200	3.32

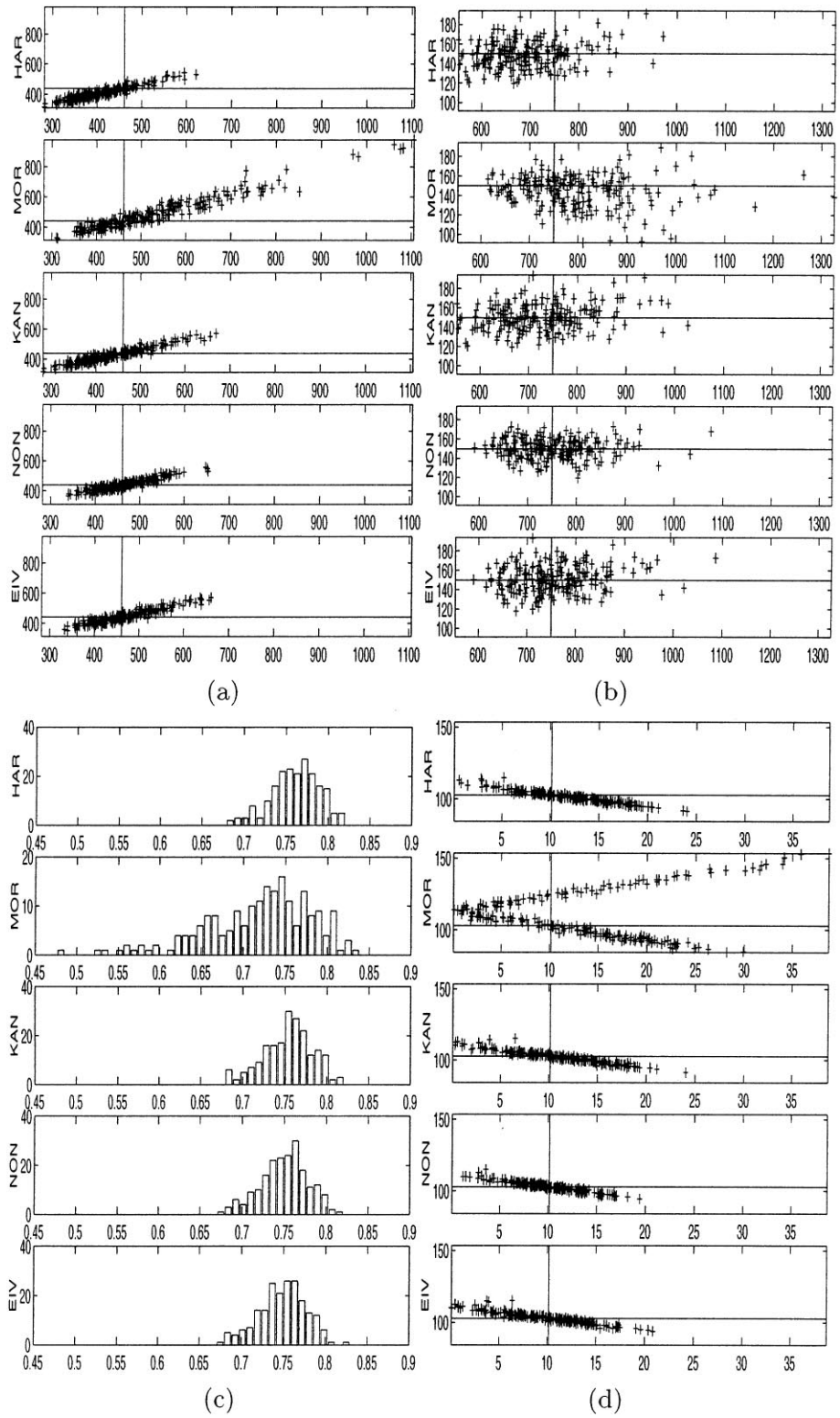

Figure 7. Epipolar geometry, synthetic data, $\sigma_\nu = 4$ with rounding. Scatter plots of the first (a) and second (b) epipole. (c) The histograms of the ratio of the singular values of the homography transformation. (d) Scatter plot of the two rotation angles.

Table 3. The fundamental matrix parameters estimated for the *bridge* image.

	Epipole 1	Epipole 2	Homography Ratio	Homography Rot1	Homography Rot2	Distance	Steps
True	[150.3, −1531.5]	[81.3, −1633.4]	0.89	54.90	34.43		
HAR	[219.2, −229.1]	[210.6, −233.9]	0.93	42.74	47.28	0.08	1
MOR	[−585.3, −6078.7]	[−1162.5, −7977.9]	0.77	86.99	6.35	0.21	1
KAN	[219.2, −229.2]	[210.6, −233.8]	0.93	42.57	47.45	0.08	1[a]
NON	[218.8, −352.8]	[205.5, −361.8]	0.93	45.92	44.1	0.07	1601
EIV	[122.8, −1275.1]	[55.0, −1385.5]	0.89	58.98	30.85	0.5	12

[a]See text.

with more computationally expensive nonlinear methods can be obtained. His examples, however, do not include ellipse fitting, for which our experiments (Section 4.2) has shown a significant difference. This is probably due to two factors. First, ellipse fitting has nonzero mean $\boldsymbol{\mu}$ for the linearized noise process, and neglecting it introduces more severe bias. Second, we used difficult conditions for estimation by taking data only from a small ellipse segment.

The nonlinear minimization criteria are related to minimizing Euclidean instead of algebraic distances. The Euclidean distance between the observed and the original data point is difficult to compute when the latter belongs to a surface. Sampson (1982) discussing the problem of ellipse fitting proposed to normalize the algebraic distances by their gradients, and to minimize

$$\sum_{i=1}^{n} \left[\frac{r_i}{\left\| \frac{\partial r_i}{\partial \mathbf{m}_i} \right\|} \right]^2, \tag{48}$$

where $\frac{\partial r_i}{\partial \mathbf{m}_i}$ is the gradient computed in the noisy points. Since up to the constant σ_ν^2 the denominator is an approximation of the variance of r_i for $\mathbf{C}_\mathbf{m}^\star = \mathbf{I}$, each term of the expression has now the same variance and traditional estimation methods are adequate. Note that to obtain the solution in the space of $\mathbf{m}$, $\mathcal{R}^k$, nonlinear minimization methods have to be used.

The *gradient criterion* (48) was often employed for ellipse fitting (Zhang, 1997) and fundamental matrix estimation (Luong and Faugeras, 1996; Torr and Murray, 1997; Zhang, 1998a). In the latter context it was compared with minimizing the distances to the epipolar lines, or between the observations and the reprojections of the reconstructed points. All these experiments generated the same result. The different nonlinear criteria yield very similar performance. Recently Zhang (1998b) gave a theoretical explanation of the phenomenon.

Let analyze now the minimization criterion (12) of the linearized model. It is easy to see using (A.7) that

$$\hat{J}_\mathbf{z} = \sum_{i=1}^{n} \frac{\left[\hat{\alpha} + \mathbf{z}_i^\top \hat{\boldsymbol{\theta}} - \boldsymbol{\mu}^\top \hat{\boldsymbol{\theta}}\right]^2}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} = \sum_{i=1}^{n} \frac{\left[r_i - \boldsymbol{\mu}^\top \hat{\boldsymbol{\theta}}\right]^2}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} . \tag{49}$$

To interpret the denominator, the variance of r_i should be approximated taking into account the mapping of $\mathbf{m}$ into the linearized variables $\mathbf{z}$. Making extensive use of Taylor series and the chain rule for vector variables we obtain

$$\mathrm{var}[r_i] = \sigma_\nu^2 \frac{\partial r_i^\top}{\partial \mathbf{m}_i} \mathbf{C}_\mathbf{m}^\star \frac{\partial r_i}{\partial \mathbf{m}_i} \tag{50}$$

$$= \sigma_\nu^2 \frac{\partial r_i^\top}{\partial \mathbf{z}_i} \frac{\partial \mathbf{z}_i}{\partial \mathbf{m}_i} \mathbf{C}_\mathbf{m}^\star \frac{\partial \mathbf{z}_i^\top}{\partial \mathbf{m}_i} \frac{\partial r_i}{\partial \mathbf{z}_i} \tag{51}$$

$$= \sigma_\nu^2 \frac{\partial r_i^\top}{\partial \mathbf{z}_i} \mathbf{C}_i^{\star(o)} \frac{\partial r_i}{\partial \mathbf{z}_i} = \sigma_\nu^2 \hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^{\star(o)} \hat{\boldsymbol{\theta}} \tag{52}$$

where $\mathbf{C}_i^{\star(o)}$ is the covariance matrix of $\mathbf{z}_i$ neglecting the higher order moments. Thus our minimization criterion is similar to the gradient criterion but also includes a correction for the nonzero mean of the mapped errors and uses a higher order approximation of the covariance matrix $\mathbf{C}_i^\star$ in the iterations. This equivalence explains the almost identical performance between EIV and NON for the fundamental matrix estimation (Table 2). However, to solve the minimization with the proposed method requires an order of

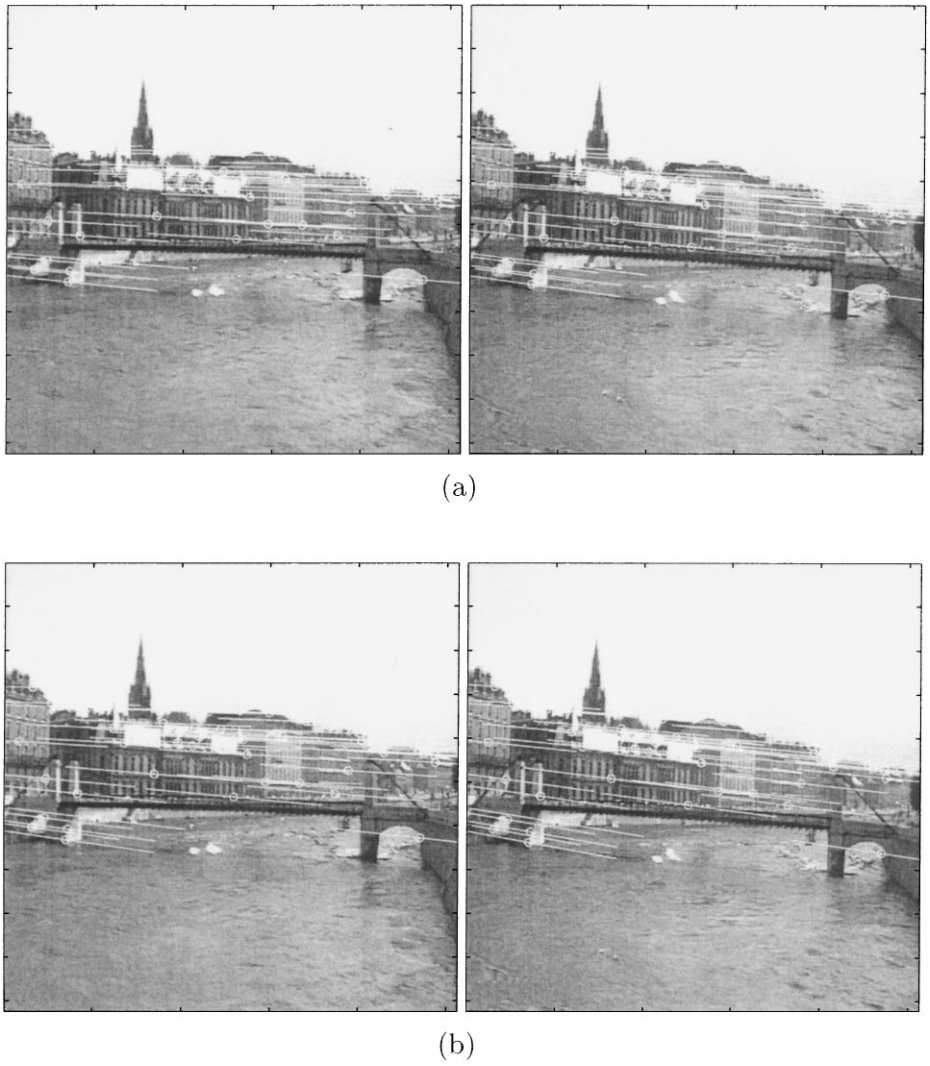

Figure 8. The *bridge* image. The matched data points are marked with circles together with the direction of the epipolar lines. (a) Ground truth. (b) Estimated with the EIV method.

magnitude less computations than the nonlinear optimization procedure.

A close to optimal solution for heteroscedastic regression is important for a much larger class of image understanding problems than the one discussed in this paper. The method described in this paper was recently generalized to vectorial constraints and was applied to the problem of 3D rigid motion of a stereo head (Matei and Meer, 1999). Other applications are currently under development.

To conclude, we have developed a new, numerically robust, quasi-optimal technique for parameter estimation and data correction under heteroscedastic noise. The technique provides a solution similar in quality with nonlinear optimization at much less computational cost. Heteroscedasticity arises often in vision problems and a reliable, relative simple approach to deal with it is of great importance.

Appendix A: Estimation of the Linearized EIV Model Parameters

In this Appendix the estimates of the parameters $[\alpha_o, \boldsymbol{\theta}_o]$ of the linear, heteroscedastic EIV model are found by solving the minimization problem (12)

repeated here for convenience.

$$[\hat{\alpha}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{z}}_i] = \arg \min_{\alpha_o, \boldsymbol{\theta}_o, \mathbf{z}_{io}} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu} - \mathbf{z}_{io})^\top \times \mathbf{C}_i^{\star -} (\mathbf{z}_i - \boldsymbol{\mu} - \mathbf{z}_{io}) \tag{A.1}$$

$$\text{subject to} \quad \alpha_o + \mathbf{z}_{io}^\top \boldsymbol{\theta}_o = 0 \quad \text{and} \quad \|\boldsymbol{\theta}_o\|^2 = 1. \tag{A.2}$$

Introducing the Lagrange multipliers η_i, and substituting the true parameters with the estimated ones, we define the function

$$L = \frac{1}{2} \sum_{i=1}^{n} (\mathbf{z}_i - \boldsymbol{\mu} - \hat{\mathbf{z}}_i)^\top \mathbf{C}_i^{\star -} (\mathbf{z}_i - \boldsymbol{\mu} - \hat{\mathbf{z}}_i) + \sum_{i=1}^{n} \eta_i \left(\hat{\alpha} + \hat{\mathbf{z}}_i^\top \hat{\boldsymbol{\theta}}\right). \tag{A.3}$$

Differentiating with respect to $\hat{\mathbf{z}}_i$ yields

$$-\mathbf{C}_i^{\star -} (\mathbf{z}_i - \boldsymbol{\mu} - \hat{\mathbf{z}}_i) + \eta_i \hat{\boldsymbol{\theta}} = 0 \tag{A.4}$$

from where

$$\hat{\mathbf{z}}_i = \mathbf{z}_i - \boldsymbol{\mu} - \eta_i \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}. \tag{A.5}$$

Note that the relation (A.5) is valid for singular $\mathbf{C}_i^\star$ since no data correction is required in the null-space of the matrix. Using the constraint (A.2) for the estimated values yields

$$\eta_i = \frac{\hat{\alpha} + (\mathbf{z}_i - \boldsymbol{\mu})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \tag{A.6}$$

from where the closed form of $\hat{\mathbf{z}}_i$ as a function of $\hat{\boldsymbol{\theta}}$ and $\hat{\alpha}$ is obtained

$$\hat{\mathbf{z}}_i = \mathbf{z}_i - \boldsymbol{\mu} - \frac{[\hat{\alpha} + (\mathbf{z}_i - \boldsymbol{\mu})^\top \hat{\boldsymbol{\theta}}] \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}. \tag{A.7}$$

Differentiating (A.3) with respect to $\hat{\alpha}$ yields

$$\sum_{i=1}^{n} \eta_i = 0, \tag{A.8}$$

and by (A.6)

$$\hat{\alpha} \sum_{i=1}^{n} \frac{1}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} + \sum_{i=1}^{n} \frac{(\mathbf{z}_i - \boldsymbol{\mu})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} = 0. \tag{A.9}$$

We define the weighted average

$$\tilde{\mathbf{z}} = \frac{\sum_{i=1}^{n} \dfrac{\mathbf{z}_i}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}}{\sum_{i=1}^{n} \hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}, \tag{A.10}$$

and thus the intercept estimate is

$$\hat{\alpha} = -(\tilde{\mathbf{z}} - \boldsymbol{\mu})^\top \hat{\boldsymbol{\theta}}. \tag{A.11}$$

Applying (A.11) to (A.7) we obtain the expression of the corrected data points of the *linear* model

$$\hat{\mathbf{z}}_i = \mathbf{z}_i - \boldsymbol{\mu} - \frac{[(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}] \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}. \tag{A.12}$$

Differentiating (A.3) with respect to $\hat{\boldsymbol{\theta}}$ yields

$$\sum_{i=1}^{n} \eta_i \hat{\mathbf{z}}_i = \mathbf{0}, \tag{A.13}$$

which using the previous results can be written as

$$\sum_{i=1}^{n} \left(\frac{(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \right) \left(\mathbf{z}_i - \frac{[(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}] \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \right) = \mathbf{0} \tag{A.14}$$

or

$$\left[\sum_{i=1}^{n} \frac{\mathbf{z}_i (\mathbf{z}_i - \tilde{\mathbf{z}})^\top}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \right] \hat{\boldsymbol{\theta}} - \left[\sum_{i=1}^{n} \left(\frac{(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \right)^2 \mathbf{C}_i^\star \right] \hat{\boldsymbol{\theta}} = \mathbf{0}. \tag{A.15}$$

From (A.10) we have the equality

$$\sum_{i=1}^{n} \frac{\mathbf{z}_i - \tilde{\mathbf{z}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} = \mathbf{0}. \tag{A.16}$$

and the expression (A.15) becomes

$$\left[\sum_{i=1}^{n} \frac{(\mathbf{z}_i - \tilde{\mathbf{z}})(\mathbf{z}_i - \tilde{\mathbf{z}})^\top}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \right] \hat{\boldsymbol{\theta}} - \left[\sum_{i=1}^{n} \left(\frac{(\mathbf{z}_i - \tilde{\mathbf{z}})^\top \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}^\top \mathbf{C}_i^\star \hat{\boldsymbol{\theta}}} \right)^2 \mathbf{C}_i^\star \right] \hat{\boldsymbol{\theta}} = \mathbf{0} \tag{A.17}$$

which taking into account the definitions (24) yields (23).

Appendix B: The Use of Generalized Singular Value Decomposition

The solution of the generalized eigenproblem is very unstable with ill-conditioned or singular matrices (Golub and Van Loan, 1990, p. 469). In Van Huffel and Vandewalle (1989) a stable numerical method was introduced based on the generalized singular value decomposition (GSVD), which simultaneously diagonalizes two matrices. In this Appendix we show how the GSVD technique can be exploited to solve generalized eigenproblems.

Let $\mathbf{A}$ and $\mathbf{B}$ be an arbitrary $n \times p$ and a positive semi-definite $p \times p$ matrix, respectively. The GSVD provides the decompositions (Golub and Van Loan, 1990, p. 471) and (Van Huffel and Vandewalle, 1989)

$$\mathbf{A} = \mathbf{U}_A \boldsymbol{\Sigma}_A \mathbf{X}^{-1}, \quad \mathbf{B} = \mathbf{U}_B \boldsymbol{\Sigma}_B \mathbf{X}^{-1}, \tag{B.1}$$

where $\mathbf{U}_A, \mathbf{U}_B$ are an $n \times n$ and a $p \times p$ orthogonal matrices, $\boldsymbol{\Sigma}_A, \boldsymbol{\Sigma}_B$ are an $n \times p$ and a $p \times p$ diagonal matrices, and $\mathbf{X}$ is a $p \times p$ nonsingular matrix. Then, the general eigenproblem

$$\mathbf{A}^\top \mathbf{A} \hat{\boldsymbol{\theta}} - \lambda \mathbf{B}^\top \mathbf{B} \hat{\boldsymbol{\theta}} = \mathbf{0} \tag{B.2}$$

is reduced to

$$\boldsymbol{\Sigma}^2 \mathbf{X}^{-1} \hat{\boldsymbol{\theta}} = \lambda \mathbf{X}^{-1} \hat{\boldsymbol{\theta}}, \quad \text{where } \boldsymbol{\Sigma}^2 = \left(\boldsymbol{\Sigma}_B^2\right)^- \boldsymbol{\Sigma}_A^\top \boldsymbol{\Sigma}_A. \tag{B.3}$$

The nonzero elements of the diagonal $p \times p$ matrix $\boldsymbol{\Sigma}^2$ are the squared generalized singular values of the matrix pair $[\mathbf{A}, \mathbf{B}]$, i.e.,

$$\sigma_i^2 = \left(\frac{\sigma_{Ai}}{\sigma_{Bi}}\right)^2 \quad i = 1 \,..\, p. \tag{B.4}$$

Note that if $\mathbf{A}$ is not of full rank some of the σ_i may be zero. The matrix $\mathbf{B}$ is semi-definite and some of the σ_{Bi} can be zero too. This is not a problem, however, since the corresponding values of σ_i will be very large and are of no interest.

Let σ_q be the smallest value (including zero) among the σ_i. Since $\mathbf{X}^{-1}\hat{\boldsymbol{\theta}}$ is the eigenvector of the symmetric, positive semi-definite matrix $\boldsymbol{\Sigma}^2$, we have

$$\mathbf{X}^{-1}\hat{\boldsymbol{\theta}} = \gamma \mathbf{e}_q \quad \hat{\boldsymbol{\theta}} = \gamma \mathbf{x}_q, \tag{B.5}$$

where $\mathbf{e}_q$ is the standard basis vector with one in the qth position, and thus $\mathbf{x}_q$ is the qth column of $\mathbf{X}$. The value of the scaling factor γ can be determined from additional constraints on $\hat{\boldsymbol{\theta}}$, such as unit norm.

The computation of the GSVD deals only with the matrices $\mathbf{A}$ and $\mathbf{B}$ whose condition numbers are the square root of that of $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{B}^\top\mathbf{B}$, the matrices defining the generalized eigenproblem. Also no matrix inversion is necessary during the computations. The technique can deal with rank deficient matrices which is not the case for the generalized eigenvalue decomposition. Thus, solving with GSVD general eigenproblems of the structure (B.2) has a better numerical behavior than using straightforward, eigenvector based techniques.

Acknowledgment

We thank Bogdan Matei and David Tyler for many helpful discussions and Roger Mohr for providing the images used in the fundamental matrix estimation experiment. The research was supported by the National Science Foundation under the grant IRI-9530546. A short version of this paper, with ellipse fitting as application was presented at the Sixth International Conference on Computer Vision, Bombay, India, January 1998.

References

Boufama, B. and Mohr, R. 1995. Epipole and fundamental matrix estimation using virtual parallax. In *Proc. 5th Int. Conf. on Computer Vision*, Boston, MA, pp. 1030–1036.

Cabrera, J. and Meer, P. 1996. Estimating ellipses by the bootstrap method for bias correction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18:752–756.

Chojnacki, W., Brooks, M.J., and Hengel, A. van den, 1999. Fitting surfaces to data with covariance information: Fundamental methods applicable to computer vision. Technical Report TR-99-03. Department of Computer Science, University of Adelaide, Australia.

Faugeras, O. 1993. *Three—Dimensional Computer Vision*. MIT Press: Cambridge, MA.

Faugeras, O. 1995. Stratification of three-dimensional vision: projective, affine and metric representations. *Journal of Optical Society of America*, 12:465–484.

Fitzgibbon, A.W., Pilu, M., and Fisher, R.B. 1999. Direct least squares fitting of ellipses. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21:476–480.

Fuller, W.A. 1987. *Measurement Error Models*. Wiley: New York.

Golub, G.H. and Van Loan, C.F. 1990 *Matrix Computations*, 2nd ed. Johns Hopkins University Press: Baltimore, MD.

Hartley, R.I. 1997a. In defense of the eight-point algorithm. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:580–593.

Hartley, R.I. 1997b. Minimizing algebraic error. In *Proc. DARPA 1997 Image Understanding Workshop*, New Orleans, LA, pp. 631–637.

Kanatani, K. 1993. Unbiased estimation and statistical analysis of 3-D rigid motion from two views. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:37–50.

Kanatani, K. 1994. Statistical bias of conic fitting and renormalization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:320–326.

Kanatani, K. 1996. *Statistical Optimization for Geometric Computation: Theory and Practice*. North-Holland: Amsterdam, The Netherlands.

Leedan, Y. 1997. Statistical analysis of quadratic problems in computer vision. Ph.D. Thesis. Department of Electrical and Computer Engineering. Rutgers University. Available at web-site `www.caip.rutgers/riul/research/theses.html`

Luong, Q.-T. and Faugeras, O. 1996. The fundamental matrix: Theory, algorithms and stability analysis. *Int. Journal on Computer Vision*, 17:43–76.

Matei, B. and Meer, P. 1999. Optimal rigid motion estimation and performance evaluation with bootstrap. In *Proc. 1999 IEEE Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, Vol. 1, pp. 339–345.

Meer, P., Mintz, D., Rosenfeld, A, and Kim, D.Y. 1991. Robust regression methods for computer vision: A review. *Int. J. Computer Vision*, 6:59–70.

Mendel, J.M. 1995. *Lessons in Estimation Theory for Signal Processing, Communication and Control*. Prentice Hall: Englewood Cliffs, NJ.

Mirsky, L. 1990. *An Introduction to Linear Algebra*. Dover: Mineola, NY.

Ponce, J. and Genc, Y. 1996. Epipolar geometry and linear subspace methods: A new approach to weak calibration. In *Proc. 1996 IEEE Conf. on Computer Vision and Pattern Recognition*, San Francisco, CA, pp. 776–781.

Sampson, P.D. 1982. Fitting conic sections to "very scattered" data: An iterative refinement of the Bookstein algorithm. *Computer Graphics Image Processing*, 18:97–108.

Taubin, G. 1991. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:1115–1138.

Taubin, G. 1993. An improved algorithm for algebraic curve and surface fitting. In *Proc. 4th Int. Conf. on Computer Vision*, Berlin, Germany, pp. 658–665.

Torr, P.H.S. 1995. Motion segmentation and outlier detection. Ph.D Thesis, OUEL 1987/93, Department of Engineering Science, University of Oxford.

Torr, P.H.S. and Murray, D.W. 1997. The development and comparison of robust methods for estimating the fundamental matrix. *Int. Journal on Computer Vision*, 24:271–300.

Van Huffel, S. and Vandewalle, J. 1989. Analysis and properties of GTLS in problem $AX \approx B$. *SIAM Journal on Matrix Analysis and Applications*, 10:294–315.

Van Huffel, S. and Vandewalle, J. 1991. *The Total Least Squares Problem*. SIAM: Philadelphia, PA.

Zhang, Z. 1997. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, 15:59–76.

Zhang, Z. 1998a. Determining the epipolar geometry and its uncertainty: A review. *Int. Journal on Computer Vision*, 27:161–195.

Zhang, Z. 1998b. On the optimization criteria used in two-view motion analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20:717–729.

Zhang, Z., Deriche, D., Faugeras, O., and Luong Q.T. 1995. A robust technique for matching two uncalibrated images through the recovery of unknown epipolar geometry. *Artificial Intelligence*, 78:87–119.