Computer Vision and Image Understanding **78,** 69–83 (2000)
doi:10.1006/cviu.1999.0828, available online at http://www.idealibrary.com on IDEAL®

Heteroscedastic Hough Transform (HtHT): An Efficient Method for Robust Line Fitting in the ‘Errors in the Variables’ Problem

Nahum Kiryati

Department of Electrical Engineering—Systems, Faculty of Engineering, Tel Aviv University, Ramat Aviv 69978, Israel
E-mail: nk@eng.tau.ac.il

and

Alfred M. Bruckstein

Department of Computer Science, Technion—Israel Institute of Technology, Haifa 32000, Israel
E-mail: freddy@cs.technion.ac.il

Received February 23, 1999; accepted November 5, 1999

A versatile, systematic, and efficient line-fitting algorithm is presented, accommodating (1) errors in both coordinates (‘errors in the variables’), (2) correlation between the noise in the two coordinates (i.e., equal noise density ellipses that are not aligned with the coordinate axes), (3) heteroscedastic noise (different noise covariance matrices for different data points), and (4) outliers (achieving robustness by using finite support influence functions). The starting point for the analysis is the assumption of additive, zero mean, Gaussian measurement noise with point-dependent covariance matrix with crossterms. A maximum-likelihood approach is taken. The handling of outliers is inspired by robust M-estimation. Line fitting is viewed as a global optimization problem. It is shown that even in the rather general setup considered here, the objective function has a special structure in the normal parameters space, that allows efficient systematic computation. The suggested algorithm can be extended to deal with “repulsive” data points (from which the line should keep a distance) and with simultaneous fitting of several lines to the same data set. © 2000 Academic Press

Key Words: correlated errors in the variables; coarse-to-fine grid search; heteroscedastic noise; Hough transform; maximum likelihood regression; robust M-estimation.

1. INTRODUCTION

Consider the problem of fitting a straight line l to a planar set of points

$$S = \{(x_i, y_i), i = 1, \ldots, M\}.$$

In the simplest formulation of the problem, which is common in practically all the experimental sciences, it is assumed that the x_i's are error-free, while the y_i's are contaminated with additive random i.i.d. Gaussian noise with zero mean and variance σ_y^2. It is well known (and easy to show) that maximum likelihood (ML) estimation of the line parameters leads in this case to the classical method of least squares (LS), i.e., to the minimization problem

$$\hat{l} = \arg\min_l \sum_{i=1}^{M} (Y_i - y_i)^2,$$

subject to the constraint that the "adjusted" points $\{(X_i = x_i, Y_i) \mid i = 1, \ldots, M\}$ are on the line l. This problem has a well-known closed-form analytic solution.

In the 'errors in the variables' formulation of the line-fitting problem, it is no longer assumed that the x_i's are error-free. Generally, the 'errors in the variables' problem does not admit an analytic solution. Suppose, however, that the noise in the x_i's is also identically distributed Gaussian with zero mean and variance $\sigma_x^2 = \sigma_y^2$, and that for any given x_i it is independent of the noise in the other x_i's and is also independent of the noise in y_i. ML estimation of the line parameters leads, in this case, to the method of total least squares (TLS), i.e., to minimizing the sum of squared distances between the noisy data points and the fitted line, measured in the direction perpendicular to the line. The TLS line fitting admits an analytic solution. By scaling the coordinate system axes, the analytic solution can be extended to the case of $\sigma_x^2 \neq \sigma_y^2$.

In many applications, the assumptions behind the LS and TLS methods are acceptable. Suppose, however, that the variance of the noise is not identical for all the data points. This is known as the heteroscedastic case. Furthermore, assume that the noise in x_i is correlated to the noise in y_i. The latter situation leads to constant noise density ellipses that are not aligned with the coordinate system axes. Generally, there is no analytic solution for the cases of heteroscedastic and/or correlated noise.

Line-fitting problems with heteroscedastic and correlated noise are quite common. Whenever the points in the data set are obtained by fusion of several sensors, each the source of some of the points, the noise is generally heteroscedastic since each sensor has different noise characteristics. If, in addition, the fusion process involves coordinate system transformations, in particular rotation, the noise in the x and y coordinates will be correlated. Heteroscedastic regression problems in computer vision are studied in [13]. Efficient solutions to problems of this type are in high demand.

Line fitting based on ML estimation of the line parameters, in particular the LS and TLS methods, fails when the data set contains outliers that do not satisfy the assumed statistical noise model. Several approaches to robust regression, with various statistical properties, have been suggested; see, e.g., [1, 5, 12, 14, 20]. In early studies, the computational cost of robust regression received little attention. Later, significant efforts were aimed at developing computationally efficient robust regression methods; see, e.g., [3, 15].

The Hough transform [7, 11] is an effective way to fit a line to data points in the presence of (possibly many) outliers. In its standard form [2], each data point is transformed to a sinusoidal voting pattern in an accumulator array that represents the normal parameters plane. The Hough transform will fail if errors in the position of the genuine (nonoutlier) points are present, since the sinusoids in the parameter plane will not intersect, as required, in a point. The resulting spreading of "votes" implies that if the errors are large, the maximum in the accumulator array will no longer correspond to the parameters of the true line.

Shapiro [17] and Thrift and Dunn [19] extended the Hough transform to deal with errors in the position of the genuine data points. See also Princen *et al.* [16]. Kiryati and Bruckstein [8, 9] showed that the Hough transform can be tuned to perform M-estimation, to accommodate repulsive data points, and to allow for heteroscedastic noise, but only if the noise is not correlated, i.e., only if the main axes of the equal noise density ellipses are parallel to the coordinate system axes.

In this paper we present an efficient method for robust line fitting in the heteroscedastic 'errors in the variables' problem, with correlated noise. It is assumed that the covariance matrix associated with each data point is known. The method suggested is easy to implement and fast to compute and provides a systematic solution to this important practical problem. The organization of the paper is as follows. In Section 2 the problem is defined and formulated as a global optimization problem and the general approach to solving it is outlined. In Section 3 it is shown that the objective function can be simplified and has a special structure. It is further shown that this special structure leads to an elegant, efficient computational solution. Experimental results are presented in Section 4. In Section 5 an alternative definition of the problem is considered and limitations of the method are discussed.

2. PROBLEM STATEMENT

Let $S = \{(X_i, Y_i) \mid i = 1, \ldots, M\}$ be an unknown set of collinear points in the plane, but suppose that measurements of the coordinates of the points are available: $s = \{(x_i, y_i) \mid i = 1, \ldots, M\}$. Due to noise, $s \neq S$. The measurement noise is assumed to be additive and independent between points. Its probability density is modeled as a 2-D Gaussian with zero mean and a point-specific covariance matrix,

$$p(x_i, y_i \mid X_i, Y_i) = \frac{1}{2\pi |\boldsymbol{\Sigma}_i|^{\frac{1}{2}}} \exp\left(-\bar{\boldsymbol{\Delta}}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \bar{\boldsymbol{\Delta}}_i / 2\right), \tag{1}$$

where

$$\bar{\boldsymbol{\Delta}}_i \equiv \left(\Delta_{x_i}, \Delta_{y_i}\right)^{\mathrm{T}} \equiv (X_i - x_i, Y_i - y_i)^{\mathrm{T}},$$

$\boldsymbol{\Sigma}_i$ is the covariance matrix

$$\boldsymbol{\Sigma}_i \equiv \begin{bmatrix} \sigma_{x_i}^2 & \sigma_{xy_i} \\ \sigma_{xy_i} & \sigma_{y_i}^2 \end{bmatrix},$$

and $|\boldsymbol{\Sigma}_i|$ is the determinant of the covariance matrix,

$$|\boldsymbol{\Sigma}_i| \equiv \sigma_{x_i}^2 \sigma_{y_i}^2 - \sigma_{xy_i}^2.$$

This is, due to the dependence of the covariance matrix on i, a heteroscedastic Gaussian noise model, with correlated noise.

The line-fitting task may therefore be regarded as finding the maximum likelihood estimator for the set of collinear points S. This can be expressed as

$$\hat{S} = \arg\max_{S \in S^*} p\{s \mid S\}, \tag{2}$$

where S^* is the set of all collinear sets of M points in the plane. The independence of the noise between points leads to

$$\hat{S} = \arg\max_{S \in S^*} \prod_{i=1}^{M} p[(x_i, y_i) \mid (X_i, Y_i)]. \tag{3}$$

The logarithm function is monotonic, so

$$\hat{S} = \arg\max_{S \in S^*} \sum_{i=1}^{M} \log p[(x_i, y_i) \mid (X_i, Y_i)]. \tag{4}$$

Substituting Eq. (1) and discarding constants that are irrelevant for the minimization we obtain

$$\hat{S} = \arg\min_{S \in S^*} \sum_{i=1}^{M} \bar{\boldsymbol{\Delta}}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \bar{\boldsymbol{\Delta}}_i. \tag{5}$$

The points in S are collinear; thus all the points $(X_i, Y_i) \in S$ must lie on a line l in the plane. Given a line l, the fitting of the individual points can be carried out independently. Let C_i quantify the fitting error between the data point $(x_i, y_i) \in s$ and the matched point $(X_i, Y_i) \in l$,

$$C_i(l) = \min_{(X_i, Y_i) \in l} \bar{\boldsymbol{\Delta}}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \bar{\boldsymbol{\Delta}}_i. \tag{6}$$

Let l^* be the set of straight lines in the plane. The line-fitting problem is now that of finding the line l that leads to the smallest total fitting cost:

$$\hat{l} = \arg\min_{l \in l^*} C(l) = \arg\min_{l \in l^*} \sum_{i=1}^{M} C_i(l). \tag{7}$$

Note that the TLS method is a special case where $\forall i \;\; \sigma_{x_i}^2 = \sigma_{y_i}^2 = \sigma^2$ and $\sigma_{xy_i} = 0$. The LS method is the limiting case in which $\forall i \;\; \sigma_{y_i}^2 = \sigma^2$, $\sigma_{xy_i} = 0$, and $\sigma_{x_i}^2 = 0$ (singular covariance matrix).

We have so far assumed that there are no outliers in the set of data points s. In the presence of outliers, the above formulation of the line-fitting problem is inadequate. Since the contribution C_i of any data point to the total cost is unbounded, even a single outlier can throw the fitted line $\hat{l}$ arbitrarily far from its true position. To alleviate the problem, a robust approach similar to M-estimation can be taken, limiting the contribution of each point to the total cost. Formally, C_i, is replaced by $\varrho(C_i)$, where $\varrho(\cdot)$ is some saturation function such as

$$\varrho(C_i) = \min\{C_i, a_i^2\}. \tag{8}$$

If a descent-type optimization technique were to be used, a smoother differentiable saturation function $\varrho(\cdot)$ would be preferable.

The algebraic form of the constraint $(X_i, Y_i) \in l$, i.e., that the matched point is on the line, depends on the parameterization chosen for straight lines. The slope–intercept parameterization is common, but has several drawbacks, primarily the singularity of vertical lines. We use the normal parameterization (ρ, θ), which is generally well behaved and has the advantage that all lines passing through a bounded domain in the x–y plane are represented by normal parameters that belong to a bounded domain in the ρ–θ parameter plane. Robust line fitting with heteroscedastic correlated noise is thus the solution of the optimization problem

$$\arg\min_{(\rho,\theta)} C(\rho, \theta) = \arg\min_{(\rho,\theta)} \sum_{i=1}^{M} \varrho[C_i(\rho, \theta)], \tag{9}$$

where

$$C_i(\rho, \theta) = \min_{(X_i, Y_i)} \bar{\boldsymbol{\Delta}}_i^{\mathrm{T}} \boldsymbol{\Sigma}_i^{-1} \bar{\boldsymbol{\Delta}}_i \tag{10}$$

and the minimization in Eq. (10) is subject to the constraint that the "adjusted point" (X_i, Y_i) is on the line defined by (ρ, θ), i.e., that its coordinates satisfy

$$\rho = X_i \cos\theta + Y_i \sin\theta \quad i = 1, \ldots, M. \tag{11}$$

As discussed above, least squares (LS) and total least squares (TLS) line fitting admit analytic solutions. The correlated heteroscedastic line-fitting problem without outliers, i.e., where we can simply use $\varrho(C_i) = C_i$, does not generally admit an analytic solution. However, in this case the objective function $C(\rho, \theta)$ can be sufficiently well behaved to allow solution by a descent-type algorithm. When outliers have to be accounted for, the ϱ function has to be a saturation function (as in Eq. (8)). The robust, correlated heteroscedastic line-fitting problem, as defined by Eqs. (9)–(11), is generally a global optimization problem, often with many local minima. Then, descent-type optimization techniques usually fail unless a very good starting point is provided. Such good initial approximations are normally not available.

It is well known that the general global optimization problem, i.e., that of finding the global minimum of a general function above a continuous bounded domain, cannot be solved by a finite number of function evaluations. However, if the objective function is not too badly behaved, in particular if its variations are small within small neighborhoods, various global optimization approaches can yield good results [21].

In this paper an algorithm to solve the robust line-fitting problem with correlated heteroscedastic noise is developed. The general approach is coarse-to-fine grid search. In principle, a rectangular grid is placed in the domain of the objective function. The objective function is evaluated at all the grid points. Finer local grids are then positioned around significant minima locations. The process continues until the required precision is obtained. In the context of the standard Hough transform, coarse-to-fine search has been suggested in [6].

The viability of this approach depends on two issues. One is the level of certainty that the coarse-to-fine grid search converges to the true global minimum of the objective function, i.e., to the best line. Convergence analysis of coarse-to-fine grid search in a similar robust line-fitting problem has been studied in [18]. The other issue is computational. In particular, if it had been necessary to compute the objective function from scratch at every grid point, the overall computational cost would have been rather high. We show that efficient systematic evaluation of the objective function on a rectangular grid is possible, leading to rapid robust line fitting.

In a previous paper [9] we studied a special degenerate case of the line-fitting problem. There it was assumed that the correlation terms σ_{xy_i} in the covariance matrices $\mathbf{\Sigma}_i$ in Eq. (10) are all zero. In this paper an efficient systematic way to solve the general problem, with correlated noise, as defined by Eqs. (9)–(11) is presented and demonstrated.

3. EFFICIENT COMPUTATION ON A GRID

Consider a single term

$$C_i(\rho,\theta) = \min_{(X_i,Y_i)} \bar{\mathbf{\Delta}}_i^{\mathrm{T}} \mathbf{\Sigma}_i^{-1} \bar{\mathbf{\Delta}}_i \tag{12}$$

where the minimization is subject to the constraint (11). Inverting the covariance matrix and performing the vector–matrix multiplications, we obtain

$$C_i(\rho,\theta) = \min_{(X_i,Y_i)} \frac{1}{|\mathbf{\Sigma}_i|}\left(\Delta_{x_i}^2 \sigma_{y_i}^2 + \Delta_{y_i}^2 \sigma_{x_i}^2 - 2\Delta_{x_i}\Delta_{y_i}\sigma_{xy_i}\right). \tag{13}$$

In order to carry out the minimization subject to (11), we define the Lagrangian

$$\Phi_i = \frac{1}{|\mathbf{\Sigma}_i|}\left(\Delta_{x_i}^2 \sigma_{y_i}^2 + \Delta_{y_i}^2 \sigma_{x_i}^2 - 2\Delta_{x_i}\Delta_{y_i}\sigma_{xy_i}\right) + \lambda(\rho - X_i\cos\theta - Y_i\sin\theta) \tag{14}$$

and require that $\partial\Phi_i/\partial\lambda = 0$, $\partial\Phi_i/\partial X_i = 0$, and $\partial\Phi_i/\partial Y_i = 0$. The first requirement yields the constraint (11). The second and third requirements lead to

$$\Delta_{x_i}\sigma_{y_i}^2 - \Delta_{y_i}\sigma_{xy_i} = \frac{\lambda|\mathbf{\Sigma}_i|}{2}\cos\theta \tag{15}$$

$$\Delta_{y_i}\sigma_{x_i}^2 - \Delta_{x_i}\sigma_{xy_i} = \frac{\lambda|\mathbf{\Sigma}_i|}{2}\sin\theta \tag{16}$$

Equations (15) and (16) can be solved for $\Delta_{x_i} \equiv X_i - x_i$ and $\Delta_{y_i} \equiv Y_i - y_i$. We obtain

$$\Delta_{x_i} = \frac{\lambda}{2}\left(\sigma_{x_i}^2\cos\theta + \sigma_{xy_i}\sin\theta\right) \tag{17}$$

$$\Delta_{y_i} = \frac{\lambda}{2}\left(\sigma_{y_i}^2\sin\theta + \sigma_{xy_i}\cos\theta\right). \tag{18}$$

Now, Eq. (13) can be rewritten as

$$C_i(\rho,\theta) = \min_{(X_i,Y_i)}\left[\frac{1}{|\mathbf{\Sigma}_i|}\Delta_{x_i}\cdot\left(\Delta_{x_i}\sigma_{y_i}^2 - \Delta_{y_i}\sigma_{xy_i}\right) + \frac{1}{|\mathbf{\Sigma}_i|}\Delta_{y_i}\cdot\left(\Delta_{y_i}\sigma_{x_i}^2 - \Delta_{x_i}\sigma_{xy_i}\right)\right]. \tag{19}$$

By substituting Eqs. (15) and (16) it is seen that

$$C_i(\rho,\theta) = \frac{\lambda}{2}\left(\Delta_{x_i}\cos\theta + \Delta_{y_i}\sin\theta\right) \tag{20}$$

subject to (11). Substituting (11) in Eq. (20), we obtain

$$C_i(\rho,\theta) = \frac{\lambda}{2}(\rho - x_i\cos\theta - y_i\sin\theta). \tag{21}$$

Define

$$\rho_i(\theta) \equiv x_i\cos\theta + y_i\sin\theta$$

and

$$r_i(\rho,\theta) \equiv \rho - \rho_i(\theta).$$

We can now write

$$C_i(\rho,\theta) = \frac{\lambda}{2}[\rho - \rho_i(\theta)] = \frac{\lambda}{2}\, r_i(\rho,\theta). \tag{22}$$

Note that $\rho_i(\theta)$ is the sinusoid that corresponds to the data point (x_i, y_i) in the Duda and Hart formulation of the Hough transform [2] and that $|r_i(\rho,\theta)|$ is the distance between the data point (x_i, y_i) and the line parameterized by (ρ,θ) [8].

It is also possible to substitute Eqs. (17) and (18) into Eq. (20). We get

$$C_i(\rho,\theta) = \frac{\lambda^2}{4}\cdot v_i(\theta), \tag{23}$$

where

$$v_i(\theta) \equiv \sigma_{x_i}^2\cos^2\theta + \sigma_{y_i}^2\sin^2\theta + 2\sigma_{xy_i}\sin\theta\cos\theta.$$

By comparing the expressions for $C_i(\rho,\theta)$ in Eqs. (22) and (23), we obtain that

$$\frac{\lambda}{2} = \frac{r_i(\rho,\theta)}{v_i(\theta)},$$

so finally

$$C_i(\rho,\theta) = \frac{r_i^2(\rho,\theta)}{v_i(\theta)} = \frac{(\rho - x_i\cos\theta - y_i\sin\theta)^2}{\sigma_{x_i}^2\cos^2\theta + \sigma_{y_i}^2\sin^2\theta + 2\sigma_{xy_i}\sin\theta\cos\theta}. \tag{24}$$

Using this result, robust line-fitting with heteroscedastic correlated noise, as defined by Eqs. (9)–(11), takes the simpler form

$$\arg\min_{(\rho,\theta)} \sum_{i=1}^{M} \varrho\left[\frac{(\rho - x_i\cos\theta - y_i\sin\theta)^2}{\sigma_{x_i}^2\cos^2\theta + \sigma_{y_i}^2\sin^2\theta + 2\sigma_{xy_i}\sin\theta\cos\theta}\right]. \tag{25}$$

We proceed to show that coarse-to-fine grid search is a very efficient computational approach to solving this global minimization problem, assuming that the covariance matrix associated with each data point is known.

Without loss of generality, assume that all data points $\{(x_i, y_i)\}$ lie within a circle of radius R centered at the origin of the x–y plane. In practice, it is advisable to minimize R by translating the coordinate system origin to, say, $(\bar{x}, \bar{y})$, where

$$\bar{x} \equiv \left(\min_i x_i + \max_i x_i\right) \Big/ 2$$
$$\bar{y} \equiv \left(\min_i y_i + \max_i y_i\right) \Big/ 2.$$

Since all the data points lie within the circle, the fitted line must pass through the circle. Therefore, the line can be represented by normal parameters (ρ, θ) that are within the rectangular parameter space domain

$$A^{(0)} = \{(\rho, \theta) : \ -R < \rho < R \ , \ 0 \leq \theta < \pi\}$$

in the ρ–θ plane. A rectangular grid with $N_\rho \times N_\theta$ grid points is placed in this domain, and each grid point is represented by an accumulator in an accumulator array.

Naive brute-force evaluation of the objective function on all the grid points according to Eqs. (9)–(11) is computationally expensive. However, as expressed by Eq. (25), the objective function has a special structure that can be used to dramatically reduce the computing time. The general form of a fast algorithm, guided by Eq. (25), to compute $C(\rho, \theta)$ on a rectangular set of grid points is as follows.

```
Store a discrete approximation of the
  function r^2 in a vector D_0
Reset the accumulator array
For each data point (x_i, y_i)
    For each discrete value θ_j of θ
         Let D_1 be D_0 shifted by ρ_i(θ_j)
         Let D_2 be D_1/V_i(θ_j)
         Let D_3 be D_2 clipped by a_i^2 (Eq. 8)
         Add D_3 to the θ_j column of A^(0)
    Next θ
Next data point
```

The operations in the inner loop are very simple. Thus, while the time needed for computing $C(\rho, \theta)$ is asymptotically proportional to $M \cdot N_\theta \cdot N_\rho$, the proportionality constant is very small. Furthermore, the time critical operations (division and clipping) involve a vector and a scalar, hence any type of pipelining or parallel hardware can be fully utilized to obtain maximal speedup.

The next step is to find the minimum $[\rho^{(0)}, \theta^{(0)}]$ (or a few significant minima) in the accumulator array. These parameters are a discrete approximation of the parameters of the best line that can be fitted to the data. $[\rho^{(0)}, \theta^{(0)}]$ can be used as an excellent starting point for a descent-type line-fitting algorithm. Alternatively, a fine rectangular grid can be placed in a small domain $A^{(1)}$ centered at $[\rho^{(0)}, \theta^{(0)}]$. The objective function can be efficiently calculated in higher resolution in $A^{(1)}$, its minimum $[\rho^{(1)}, \theta^{(1)}]$ can be found by search and so on until convergence is achieved.

4. EXAMPLES

The suggested algorithm for heteroscedastic robust line-fitting with correlated errors in the variables has been implemented. The input is a list of data points, each represented by a record that includes the following items:

- The coordinates of the data point (x_i, y_i).
- The standard deviations σ_{x_i} and σ_{y_i} of the errors in the x and y coordinates.
- The correlation $\text{corr}_i \equiv \sigma_{xy_i}/\sigma_{x_i}\sigma_{y_i}$ between the errors in the x and y coordinates.
- The upper bound a_i^2 on the fitting cost (saturation level), see Eq. (8).

The output is the equation of the line fitted to the data.

The coordinates of the data points in the four parts of Fig. 1 are the same, but the standard deviations, the correlation and the saturation levels differ. In the top-left part, line-fitting according to the conventional least-squares criterion, with errors in the y coordinate only, is

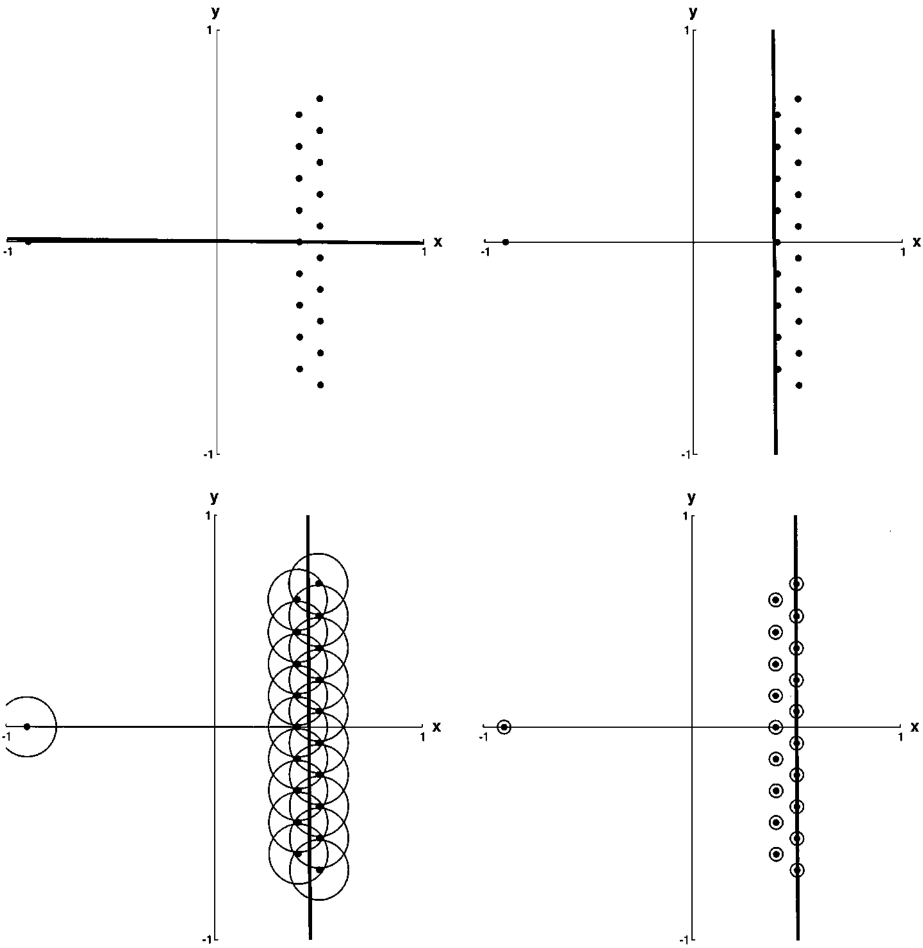

FIG. 1. (Top-left) Least-squares line fitting with errors in the y values only. The fitted line is horizontal. (Top-right) Total least squares (TLS) line fitting, as a special case of the suggested algorithm. (Bottom-left) Robust TLS. A data point has no influence beyond the circular contour. (Bottom-right) The standard Hough transform is obtained as a special case by setting the saturation level to a small value.

implemented using the suggested algorithm. This corresponds to $\sigma_{x_i} = 0$, $\sigma_{y_i} = \sigma$, $\mathrm{corr}_i = 0$ and $a_i^2 \to \infty$ (high saturation level, no clipping) for all points. The horizontal line fitted demonstrates that the conventional least-squares criterion is not suitable for the given data set. In the top-right part, for all points, $\sigma_{x_i} = \sigma_{y_i} = \sigma$, $\mathrm{corr}_i = 0$, and $a_i^2 \to \infty$. The algorithm in this case finds the total least squares (TLS) line. The damaging effect of the outlier is apparent. Robust line-fitting is made possible in the bottom-left part by setting $a_i^2 = 2$ for all points ($\sigma_{x_i} = \sigma_{y_i} = 0.1$, $\mathrm{corr}_i = 0$). The circular contour around each data point shows the locus of points (X_i, Y_i) for which

$$(X - x_i, Y - y_i)\begin{pmatrix} \sigma_{x_i}^2 & \sigma_{xy_i} \\ \sigma_{xy_i} & \sigma_{y_i}^2 \end{pmatrix}^{-1} \begin{pmatrix} X - x_i \\ Y - y_i \end{pmatrix} = a_i^2.$$

This means that the fitting cost is clamped to a_i^2 outside the circle, eliminating the damaging effect of the outliers. The bottom-right part of the figure shows that, by reducing a_i^2 to some small value (0.1) for all points, any point that is not exactly on the line is regarded as an outlier and the algorithm degenerates to the standard Hough Transform [2]. Note that in all parts of Fig. (1) there is no correlation between the errors in the x and y directions.

Robust line-fitting with correlated heteroscedastic noise is demonstrated in Fig. 2. The coordinates of the data points in both parts of Fig. 2 are the same and the standard deviations and saturation level for all the points are equal: $\sigma_{x_i} = \sigma_{y_i} = 0.1$, $a_i^2 = 2$. The correlation values differ, however. In the left part, the points in the larger collinear group are all with $\mathrm{corr}_i = -0.7$, the points in the smaller collinear group are with $\mathrm{corr}_i = 0.9$, and the noise in the remaining outlier is uncorrelated: $\mathrm{corr}_i = 0$. The elliptical (or circular) contour around each point again represents the influence region of that point, as explained above. In the right part of the figure, the noise correlation for the points in the large collinear group is taken as $+0.7$. The effect on the fitted line is apparent.

In the top-left part of Fig. 3, for all points, $\sigma_{x_i} = \sigma_{y_i} = 0.1$, $\mathrm{corr}_i = 0$, and $a_i^2 = 0.1$. This leads to line fitting as in the standard Hough transform. In the top-right part, with $a_i^2 = 2.0$ for

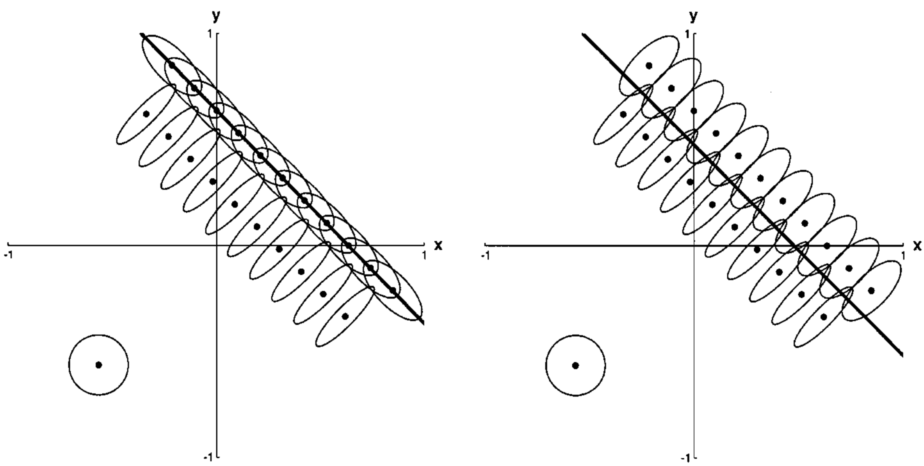

FIG. 2. Robust heteroscedastic line fitting with correlated noise. The differences in the records of the data points in these examples are only in the correlation values. (Left) The points in the smaller collinear group are regarded as outliers. (Right) The points in both collinear groups exert influence on the fitted line.

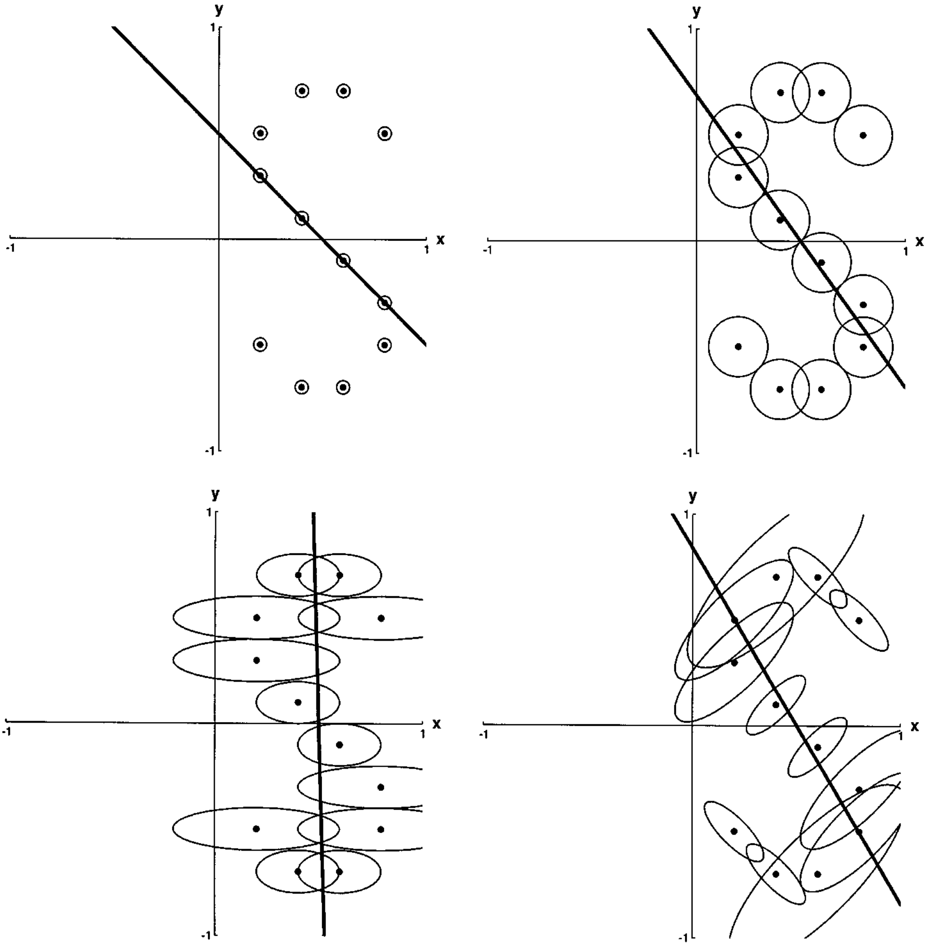

FIG. 3. (Top-left) Tuning the algorithm to compute the standard Hough transform. (Top-right) Robust TLS line-fitting. (Bottom-left) Robust heteroscedastic line fitting, without correlation. All points contribute to the fitted line, i.e., the saturation level is not reached in any point. (Bottom-right) Robust line fitting with heteroscedastic correlated noise.

all points, robust TLS line fitting is obtained. In the bottom-left part, for all points $\sigma_{y_i} = 0.1$, $\text{corr}_i = 0$ and $a_i^2 = 1.0$. Here, however, σ_{x_i} varies between the points. Observe that all the data points have influence on the fitted line in this case. Finally, in the bottom-right part, the standard deviations and the correlation values vary between the points. The saturation level is equal for all points, $a_i^2 = 2.0$, so the elliptical contour around each point, that represents the influence boundary, also illustrates the standard deviation and correlation.

Consider the gray-level image shown in Fig. 4. The top-left part of Fig. 5 is the corresponding edge image, with about 5000 edge points. The line fitted to this data set depends on the assumed standard deviations, correlations and saturation levels. The top-right part of Fig. 5 is obtained with $\sigma_{x_i} = \sigma_{y_i} = 0.01$, $\text{corr}_i = 0$ and $a_i^2 = 0.01$ for all points. As can be seen, due to the small errors assumed, the line is fitted to the smooth edge of the saw. When the standard deviations and saturation levels are increased, $\sigma_{x_i} = \sigma_{y_i} = 0.05$, $\text{corr}_i = 0$ and $a_i^2 = 0.05$ for all points, the line is fitted to the jagged edge of the saw, as seen in the bottom-left part of the figure. In the bottom-right part of Fig. 5, the saturation level a_i^2

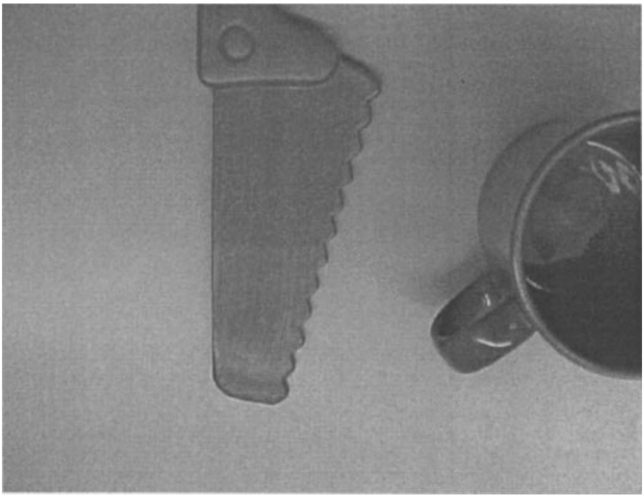

FIG. 4. A gray-level image.

associated with each edge point is made individually proportional to the absolute value of the sine of the gradient direction angle (provided by the edge detector). This has the effect of attenuating vertical and near-vertical edges, and, as seen, leads the algorithm to fit the line to the horizontal edge.

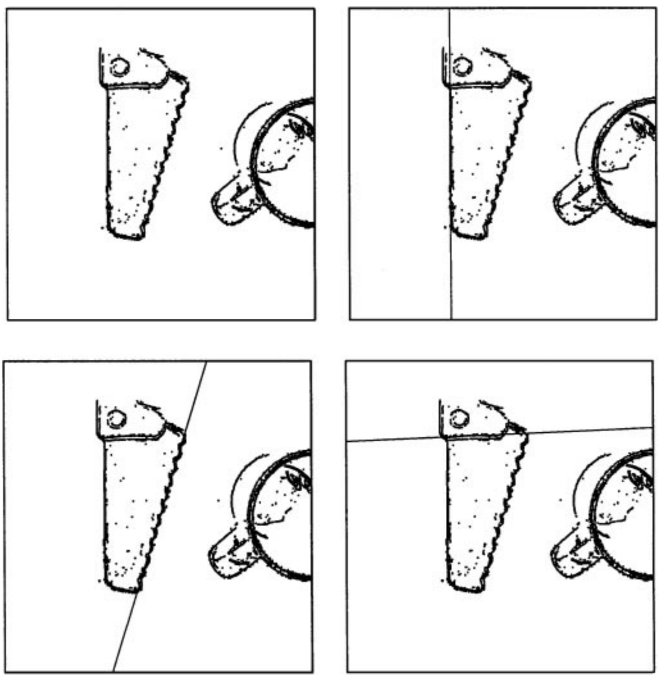

FIG. 5. (Top-left) An edge image obtained from the gray-level image shown in Fig. 4. (Top-right) With $\sigma_{x_i} = \sigma_{y_i} = 0.01$, $\text{corr}_i = 0$, and $a_i^2 = 0.01$ for all points, the line is fitted to the smooth edge of the saw. (Bottom-left) When larger errors are admitted, the longer (but jagged) edge is selected. (Bottom-right) By individually setting the saturation level associated with each point to be proportional to the absolute value of the sine of the gradient direction angle, the algorithm is made biased towards horizontal edges.

5. DISCUSSION

In this paper, we presented an efficient method for robust line-fitting in the heteroscedastic 'errors in the variables' problem, with correlated noise, based on the problem statement presented in Section 2. Suppose that the line fitting task were redefined as finding the maximum likelihood estimator for the line l on which the set of points S lies, i.e.,

$$\hat{l} = \arg\max_{l} p\{s \mid S \in l\}. \tag{26}$$

This is different from the definition in Eq. (2), since here all the possible positions of each point (X_i, Y_i) on the line l, rather than only the best one, are taken into account. From the independence of the noise between points and the monotonicity of the logarithm function it follows that

$$\hat{l} = \arg\max_{l} \sum_{i=1}^{M} \log p[(x_i, y_i) \mid (X_i, Y_i) \in l]. \tag{27}$$

Let $l(t) = [X(t), Y(t)]$ denote the arc-length parameterization of the line l. We get

$$p[(x_i, y_i) \mid (X_i, Y_i) \in l] = \int_{-\infty}^{\infty} p\{(x_i, y_i) \mid [X(t), Y(t)]\}\, dt. \tag{28}$$

It is now recognized that $p[(x_i, y_i) \mid (X_i, Y_i) \in l]$ is equal to $\mathcal{R}\{p[(x_i, y_i) \mid (X, Y)]\}$, the Radon transform of the noise probability density with respect to the spatial variables X and Y, where $p[(x_i, y_i) \mid (X, Y)]$ is as defined by Eq. (1). Using the normal parameterization of the line l, it can also be expressed as

$$\begin{aligned} p[(x_i, y_i) \mid (X_i, Y_i) \in l] &= \mathcal{R}\{p[(x_i, y_i) \mid (X, Y)](\rho, \theta)\} \\ &= \iint \delta(\rho - X\cos\theta - Y\sin\theta) p\{(x_i, y_i) \mid (X, Y)\}\, dX\, dY \end{aligned} \tag{29}$$

Using the rotation property of the Radon transform and additional algebraic steps, it can be shown that the line fitting problem, as defined by Eq. (26), is reduced to

$$\arg\min_{(\rho,\theta)} \sum_{i=1}^{M} C_i(\rho, \theta), \tag{30}$$

where $C_i(\rho, \theta)$ is exactly as given in Eq. (24). We conclude that this alternative formulation of the problem (Eq. 26) is equivalent to the one followed in the body of this paper (Eq. (2)).

Awareness of the following limitations of the suggested technique provides important insights and directions for future research.

- The line-fitting method presented here is suitable for planar data sets. Computationally efficient extensions to higher dimensions require further investigation [13].
- It is necessary to provide the noise covariance matrix and the saturation level (a_i^2) associated with each data point. If these are not available, the algorithm should be applied as a computational mechanism within a larger statistical estimation framework.

• The analysis is based on a (2-D heteroscedastic) Gaussian noise model for the inliers. The algorithm can be expected to provide useful (though not optimal) results in many other cases. Generally, extensions to non-Gaussian inlier noise models need to be developed.

• The method is based on solving a global optimization problem by coarse-to-fine grid search. Convergence is based on the well-behavedness of the objective function. Theoretical convergence properties in a related case were studied in [18]. Extension of their results is necessary.

The algorithm presented in this paper can be easily generalized to deal with "repulsive" data points, i.e., points from which the line should keep a distance [8]. Fitting several lines to the same data set can be accomplished using straightforward extensions of known Hough transform practices, e.g., [4, 10]. Beyond its applications in image analysis, the suggested algorithm is an example of a method rooted in computer vision theory that can be useful in various other scientific domains.

REFERENCES

1. M. Black and A. Rangarajan, On the unification of line processes, outlier rejection, and robust statistics with applications in early vision, *Int. J. Comput. Vision* **19**, 1996, 57–91.
2. R. O. Duda and P. E. Hart, Use of the Hough transform to detect lines and curves in pictures, *Comm. ACM* **15**, 1972, 11–15.
3. H. Edelsbrunner and D. L. Souvaine, Computing least median of squares regression and guided topological sweep, *J. Amer. Stat. Assoc.* **85**, 1990, 115–119.
4. G. Gerig, Linking image-space and accumulator-space: A new approach for object recognition, in *Proc. 1st Int. Conf. on Computer Vision (ICCV), London, 1987*, pp. 112–117.
5. C. Goodall, M-estimators of location: An outline of the theory, in *Understanding Robust and Exploratory Data Analysis* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, Eds.), Wiley, New York, 1983.
6. J. Illingworth and J. Kittler, The adaptive Hough transform, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 1987, 690–698.
7. J. Illingworth and J. Kittler, A survey of the Hough transform, *Comput. Vision Graphics Image Process.* **44**, 1988, 87–116.
8. N. Kiryati and A. M. Bruckstein, On navigating between friends and foes, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 1991, 602–606.
9. N. Kiryati and A. M. Bruckstein, What's in a set of points? *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 1992, 496–500.
10. V. F. Leavers, The dynamic generalized Hough transform: Its relationship to the probabilistic Hough transforms and an application to the concurrent detection of circles and ellipses, *CVGIP Image Understanding* **56**, 1992, 381–398.
11. V. F. Leavers, Which Hough transform? *CVGIP Image Understanding* **58**, 1993, 250–264.
12. G. Li, Robust regression, in *Exploring Data Tables, Trends and Shapes* (D. C. Hoaglin, F. Mosteller, and J. W. Tukey, Eds.), Wiley, New York, 1985.
13. Y. Leedan and P. Meer, Estimation with bilinear constraints in computer vision, in *Proc. 6th Int. Conf. on Computer Vision (ICCV), Bombay, 1998*, pp. 733–738; available at `www.caip.rutgers.edu/~meer/RIUL/projects.html`.
14. P. Meer, D. Mintz, D. Y. Kim, and A. Rosenfeld, Robust regression methods for computer vision: A review, *Int. J. Comput. Vision* **6**, 1991, 59–70.
15. D. M. Mount and N. S. Netanyahu, Computationally efficient algorithms for high-dimensional robust estimators, *Graph. Models Image Process.* **4**, 1994, 289–303.
16. J. Princen, J. Illingworth, and J. Kittler, A formal definition of the Hough transform: Properties and relationships, *J. Mathe. Imaging Vision* **1**, 1992, 153–168.

17. S. D. Shapiro, Generalization of the Hough transform for curve detection in noisy digital images, in *Proc. 4th Int. Joint Conf. on Patt. Rec., Kyoto, 1978*, pp. 710–714.

18. M. Soffer and N. Kiryati, Guaranteed convergence of the Hough transform, *Comput. Vision Image Understand.* **69**, 1998, 119–134.

19. P. R. Thrift and S. M. Dunn, Approximating point-set images by line segments using a variation of the Hough transform, *Comput. Vision Graphics Image Process.* **21**, 1983, 383–394.

20. C. V. Stewart, MINPRAN: A new robust estimator for computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **17**, 1995, 925–938.

21. A. Törn and A. Žilinskas, *Global Optimization*, Lecture Notes in Computer Science #350, Springer, 1989.