Direct Least Square Fitting of Ellipses

Andrew Fitzgibbon, Maurizio Pilu, and Robert B. Fisher

Abstract—This work presents a new efficient method for fitting ellipses to scattered data. Previous algorithms either fitted general conics or were computationally expensive. By minimizing the algebraic distance subject to the constraint $4ac - b^2 = 1$, the new method incorporates the ellipticity constraint into the normalization factor. The proposed method combines several advantages: It is ellipse-specific, so that even bad data will always return an ellipse. It can be solved naturally by a generalized eigensystem. It is extremely robust, efficient, and easy to implement.

Index Terms—Algebraic models, ellipse fitting, least squares fitting, constrained minimization, generalized eigenvalue problem.

1 Introduction

THE fitting of primitive models to image data is a basic task in pattern recognition and computer vision, allowing reduction and simplification of the data to the benefit of higher level processing stages. One of the most commonly used models is the ellipse which, being the perspective projection of the circle, is of great importance for many industrial applications. Despite its importance, however, there has been until now no computationally efficient ellipse-specific fitting algorithm [14], [5].

In this paper, we introduce a new method for fitting ellipses, rather than general conics, to segmented data. As we shall see in the next section, current methods are either computationally expensive iterative approaches, or perform ellipse fitting by least-squares fitting to a general conic and rejecting non-elliptical fits. These latter methods are cheap and perform well if the data belong to a precisely elliptical arc with little occlusion but suffer from the major shortcoming that under less ideal conditions—nonstrictly elliptical data, moderate occlusion or noise—they often yield unbounded fits to hyperbolae. In a situation where ellipses are specifically desired, such fits must be rejected as useless. A number of iterative refinement procedures [16], [8], [12] alleviate this problem, but do not eliminate it. In addition, these techniques often increase the computational burden unacceptably.

This paper introduces a new fitting method that combines the following advantages:

- ellipse-specificity, providing useful results under all noise and occlusion conditions;
- 2) invariance to affine transformation of the data;
- 3) high robustness to noise; and
- 4) high computational efficiency.

After a description of relevant previous ellipse fitting methods, in Section 3 we describe the method and provide a theoretical analysis of the uniqueness of the elliptical solution. Section 4 contains experimental results, notably to highlight behavior with

A. Fitzgibbon is with the Department of Engineering Science, University of Oxford, 19 Parks Road, Oxford, OX1 3BJ, England.
 E-mail: awf@robots.ox.ac.uk.

M. Pilu is with Hewlett-Packard Research Laboratories, Filton Road, Stoke Gifford, Bristol, BS12 6QZ, England. E-mail: mp@hplb.hpl.hp.com.

R.B. Fisher is with the Division of Informatics, University of Edinburgh, 5 Forrest Hill, Edinburgh, EH1 2QL, United Kingdom.
 E-mail: rbf@dai.edinburgh.ac.uk.

Manuscript received 4 Jan. 1999. Recommended for acceptance by R. Chin. For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 107704.

nonelliptical data, low-eccentricity bias, and noise resilience. We conclude by presenting some possible extensions.

2 Previous Methods and Their Limitations

The literature on ellipse fitting divides into two broad techniques: clustering (such as Hough-based methods [9], [19]) and least-squares fitting.

Least-squares techniques center on finding the set of parameters that minimize some distance measure between the data points and the ellipse. In this section, we briefly present the most cited works in ellipse fitting and its closely related problem, conic fitting. It will be shown that the direct specific least-square fitting of ellipses has, up to now, not been solved.

Before reviewing the literature on general conic fitting, we will introduce a statement of the problem that allows us to unify several approaches under the umbrella of constrained least squares. Let us represent a general conic by an implicit second order polynomial:

$$F(\mathbf{a}, \mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = ax^2 + bxy + cy^2 + dx + ey + f = 0,$$
 (1)

where $\mathbf{a} = [a\ b\ c\ d\ e\ f]^T$ and $\mathbf{x} = [x^2\ xy\ y^2\ x\ y\ 1]^T$. $F(\mathbf{a};\ \mathbf{x}_p)$ is called the "algebraic distance" of a point $(x,\ y)$ to the conic $F(\mathbf{a};\ \mathbf{x}) = 0$. The fitting of a general conic may be approached by minimizing the sum of squared algebraic distances

$$\mathcal{D}_{A}(\mathbf{a}) = \sum_{i=1}^{N} F(\mathbf{x}_{i})^{2}$$
 (2)

of the curve to the N data points \mathbf{x}_i [7]. In order to avoid the trivial solution $\mathbf{a} = \mathbf{0}_6$, and recognizing that any multiple of a solution \mathbf{a} represents the same conic, the parameter vector \mathbf{a} is constrained in some way. Many of the published algorithms differ only in the form of constraint applied to the parameters. For instance, many authors suggest $\|\mathbf{a}\|^2 = 1$, Rosin [14] and Gander [5] impose a + c = 1 while Rosin also investigates f = 1 [14]. Taubin's approximate square distance [17] may also be viewed as the quadratic constraint $\|N\mathbf{a}\|^2 = 1$ where N is the Jacobian $[\nabla F(\mathbf{a}; \mathbf{x}_i) \dots \nabla F(\mathbf{a}; \mathbf{x}_N)]^T$.

Note that these constraints are all either *linear*, of the form $\mathbf{c} \cdot \mathbf{a} = 1$ or *quadratic*, constraining $\mathbf{a}^T \mathbf{C} \mathbf{a} = 1$ where \mathbf{C} is a 6×6 constraint matrix.

In a seminal work, Bookstein [1] showed that if a quadratic constraint is set on the parameters (e.g., to avoid the trivial solution $\mathbf{a} = \mathbf{0}_{\theta}$) the minimization (2) can be solved by considering rank-deficient generalized eigenvalue system:

$$\mathbf{D}^{\mathsf{T}}\mathbf{D}\mathbf{a} = \lambda \mathbf{C}\mathbf{a},\tag{3}$$

where $\mathbf{D} = [\mathbf{x}_1 \ \mathbf{x}_2 \cdots \mathbf{x}_n]^T$ is called the *design matrix* and \mathbf{C} is the matrix that expresses the constraint.

A simple constraint is $\|\mathbf{a}\| = 1$ but Bookstein used the algebraic invariant constraint $a^2 + \frac{1}{2}b^2 + c^2 = 1$; Sampson [16] presented an *iterative* improvement to Bookstein method that replaces the algebraic distance (2) with a better approximation to the geometric distance, which was adapted by Taubin [17] to turn the problem again into a generalized eigensystem.

Despite the amount of work, direct *specific* ellipse fitting, however, was left unsolved. If ellipse fitting was needed, one had to rely either on generic conic fitting or on iterative methods to "nudge" the estimation towards ellipticity. For instance, Porrill [12], Ellis et al. [2], and Rosin [14] use conic fitting to initialize a Kalman filter that iteratively minimizes some error metric in order to gather new image evidence and to reject nonellipse fits by testing the discriminant $b^2 - 4ac < 0$ at each iteration. Another iterative algorithm is that of Haralick [7, Section 11.10.7], where the coefficients $\{a, b, c\}$ are transformed into $\{p^2, 2pq, q^2 + r^2\}$ so as to keep the conic discriminant always negative. A nonlinear minimization of the algebraic error over the space $\{p, q, r, d, e, f\}$ is performed.

In this journal, Rosin [15] reiterated this problem by stating that

ellipse-specific fitting is essentially a nonlinear problem and iterative methods must always be employed for this purpose. In the following section, we show that this is no longer true.

3 DIRECT ELLIPSE-SPECIFIC FITTING

In order to fit ellipses specifically while retaining the efficiency of solution of the linear least-squares problem (2), we would like to constrain the parameter vector \mathbf{a} so that the conic that it represents is forced to be an ellipse. The appropriate constraint is well known, namely, that the *discriminant* b^2-4ac be negative. However, this constrained problem is difficult to solve in general as the Kuhn-Tucker conditions [13] do not guarantee a solution. In fact, we have not been able to locate any reference regarding the minimization of a quadratic form subject to such a nonconvex inequality.

Although the imposition of this inequality constraint is difficult in general, in this case we have the freedom to arbitrarily scale the parameters so we may simply incorporate the scaling into the constraint and impose the *equality* constraint $4ac - b^2 = 1$ [4]. This is a quadratic constraint which may be expressed in the matrix form $\mathbf{a}^T \mathbf{C} \mathbf{a} = 1$ as

Now, following Bookstein [1], the constrained ellipse fitting problem reduces to

minimizing
$$E = \|\mathbf{Da}\|^2$$
 subject to the constraint $\mathbf{a}^T \mathbf{Ca} = 1$ (5)

where the design matrix D is defined as in the previous section.

Introducing the Lagrange multiplier λ and differentiating, we arrive at the system of simultaneous equations¹

$$2\mathbf{D}^{T}\mathbf{D}\mathbf{a} - 2\lambda\mathbf{C}\mathbf{a} = 0$$
$$\mathbf{a}^{T}\mathbf{C}\mathbf{a} = 1 \tag{6}$$

This may be rewritten as the system

$$Sa = \lambda Ca \tag{7}$$

$$\mathbf{a}^{T}\mathbf{C}\mathbf{a}=1\tag{8}$$

where **S** is the *scatter matrix* $\mathbf{D}^{\mathsf{T}}\mathbf{D}$. This system is readily solved by considering the generalized eigenvectors of (7). If $(\lambda_p, \mathbf{u}_i)$ solves (7), then so does $(\lambda_p, \mu \mathbf{u}_i)$ for any μ and from (8) we can find the value of μ_i as $\mu_i^2 \mathbf{u}_i^{\mathsf{T}} \mathbf{C} \mathbf{u}_i = 1$, giving

$$\mu_i = \sqrt{\frac{1}{\mathbf{u}_i^T \mathbf{C} \mathbf{u}_i}} = \sqrt{\frac{1}{\mathbf{u}_i^T \mathbf{S} \mathbf{u}_i}}.$$
 (9)

Finally, setting $\hat{\mathbf{a}}_i = \mu_i \mathbf{u}_i$ solves (6).

We note that the solution of the eigensystem (7) gives six eigenvalue-eigenvector pairs $(\lambda_p \ \mathbf{u}_p)$. Each of these pairs gives rise to a local minimum if the term under the square root of (9) is positive. In general, \mathbf{S} is positive definite, so the denominator $\mathbf{u}_i^T \mathbf{S} \mathbf{u}_i$ is positive for all \mathbf{u}_r . Therefore, the square root exists if $\lambda_i > 0$, so any solutions to (6) must have positive generalized eigenvalues.

Now we show that the minimization of $\|\mathbf{Da}\|^2$ subject to $4ac - b^2 = 1$ yields exactly one solution, which corresponds, by virtue of the constraint, to an ellipse [11]. For the demonstration, we will require Lemma 1.

1. Note that the method of Lagrange multipliers is not valid when the gradient of the constraint function becomes zero. In (5), this means Ca = 0, but then $\mathbf{a}^T C \mathbf{a} = 0$, so the constraint is violated and there is no solution.

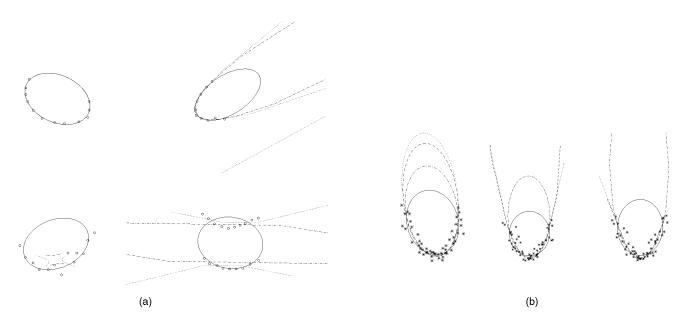


Fig. 1. (a) Fits to hand-input data to illustrate the ellipse specificity of the method. (b) Experiments with noisy parabolic data (after Sampson). Encoding is BOOK: dotted; GAND: dashed; TAUB: dash-dot; B2AC: solid.

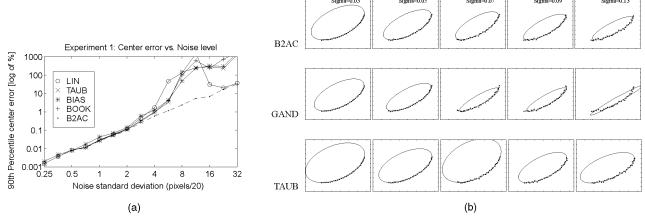


Fig. 2. (a) Variation of center position for increasing noise level when fitting to a whole ellipse. (b) Fits to arc of ellipse with increasing noise level. Notice how B2AC presents a much more graceful degradation with respect to noise.

LEMMA 1. The signs of the generalized eigenvalues of $\mathbf{Su} = \lambda \mathbf{Cu}$, where $\mathbf{S} \in \mathfrak{R}_{n \times n}$ is positive definite and $\mathbf{C} \in \mathfrak{R}_{n \times n}$ is symmetric, are the same as those of the constraint matrix \mathbf{C} , up to permutation of the indices.

PROOF. Let us define the *spectrum* $\sigma(S)$ as the set of eigenvalues of **S** and, analogously, $\sigma(S, C)$ the set of generalized eigenvalues of (7). Let the *inertia* i(S) be defined as the set of signs of $\sigma(S)$, and let i(S, C) analogously be the inertia of $\sigma(S, C)$. Then, the lemma is equivalent to proving that i(S, C) = i(C). As **S** is positive definite, it may be decomposed as \mathbf{Q}^2 for symmetric **Q**, allowing us to write $\mathbf{Su} = \lambda \mathbf{Cu}$ as $\mathbf{Q}^2\mathbf{u} = \lambda \mathbf{Cu}$. Now, substituting $\mathbf{v} = \mathbf{Qu}$ and premultiplying by \mathbf{Q}^{-1} gives $\mathbf{v} = \lambda \mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1}\mathbf{v}$ so that $\sigma(S, C) = \sigma(\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1})^{-1}$ and thus $i(S, C) = i(\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1})$. From Sylvester's Law of Inertia [18], we have that for any symmetric and nonsingular X, $i(S) = i(\mathbf{X}^T\mathbf{S}\mathbf{X})$. Therefore, substituting $\mathbf{X} = \mathbf{X}^T = \mathbf{Q}^{-1}$, we have $i(C) = i(\mathbf{Q}^{-1}\mathbf{C}\mathbf{Q}^{-1}) = i(S, C)$.

We can now state Theorem 1.

THEOREM 1. The solution of the conic fitting problem (5) subject to the constraint (4) admits exactly one elliptical solution corresponding to the single positive generalized eigenvalue of (7).

PROOF. Since the eigenvalues of \mathbf{C} are $\{-2, -1, 2, 0, 0, 0\}$, from Lemma 1 we have that (7) has *exactly one* positive eigenvalue $\lambda_i > 0$, giving the unique solution $\hat{\mathbf{a}} = \mu_i \mathbf{u}_i$ to (6). As $\mathbf{D}^T \mathbf{D}$ is positive semidefinite, the constrained problem has a minimum, which must satisfy (6), and we conclude that $\hat{\mathbf{a}}$ solves the constrained problem.

This unique solution has also some desirable properties in ellipse fitting:

- low eccentricity bias: An eigenvector of the eigensystem (7) is a local minimizer of the *Rayleigh quotient* $\frac{a^T S_a}{a^T C_a}$. In this case, the implicit normalization by $b^2 4ac$ turns singular for $b^2 4ac = 0$, which is a parabola. Since the minimization tends to "pull" the solution away from singularities [14], the unique elliptical solution tends to be biased towards low eccentricity.
- affine invariance: Let us represent the conic as $\mathbf{x}^{\top}A\mathbf{x} + \mathbf{x}^{\top}\mathbf{b} + \mathbf{c} = 0$. Under an affine transform H the leading form becomes $A' = H^{\top}AH$, so that $|A'| = |H'|^2|A|$. Being the Rayleigh quotient that we minimize $\frac{\mathbf{a}^T\mathbf{S}\mathbf{a}}{|A|}$, the new error measure is a

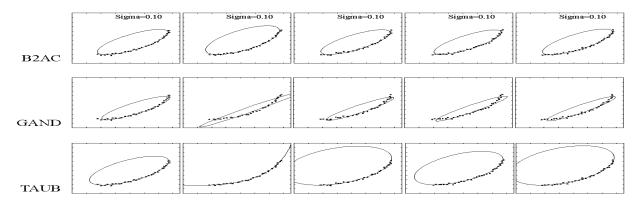


Fig. 3. Stability experiments for different runs with same noise variance (10% of data spread). The ellipse-specific method shows a remarkable stability.

scalar multiple of the original one and thus the new minimizer is transformed by H, which proved the affine invariance of the method.

4 EXPERIMENTAL RESULTS

This section describes some experiments that illustrate the interesting features of the new method and its noise performance compared to some of the least-squares fitting method reviewed in Section 2. In this short paper, we are not able to present a large body of results—which can be found in abundance in [3]—so we limited ourselves to those that are the most representative.

All experiments were conducted using the Matlab system [10]. Eigensystems are solved using the underlying EISPACK routines. We shall use the following abbreviations:

- LIN = linear method;
- BOOK = Bookstein method [1];
- TAUB = Taubin method [17];
- GAND = Gander method [5];
- BIAS = Kanatani bias-corrector method [8]; and finally
- B2AC = our new method.

4.1 Ellipse-Specificity

Despite the theoretical proof of the algorithm's ellipse specificity, it is instructive to observe its performance on some real data, of which Fig. 1a provides examples with hand-drawn datasets. The results of the method are superimposed on those of Bookstein and Gander. Dataset A is almost elliptical and indistinguishable fits were produced. The other sets exhibit varying degrees of nonellipticity and illustrate the potential of the method for coarse bounding of generic 2D data.

4.2 Low-Eccentricity Bias

Fig. 1b shows three experiments designed after Sampson [16] (following [6]) and basically consists of the same parabolic data but with different realizations of added isotropic Gaussian noise (σ =

```
% x,y are vectors of coordinates
function a=fit_ellipse(x,y)
% Build design matrix
  D = [ x.*x x.*y y.*y x y ones(size(x)) ];
% Build scatter matrix
  S = D'*D;
% Build 6x6 constraint matrix
  C(6,6)=0; C(1,3)=-2; C(2,2)=1; C(3,1)=-2;
% Solve generalised eigensystem
  [gevec, geval] = eig(S,C);
% Find the only negative eigenvalue
  [NegR, NegC] = find(geval<0 & ~isinf(geval));
% Get fitted parameters
  a = gevec(:,NegC);</pre>
```

Fig. 4. Simple six-line Matlab implementation of the ellipse fitting method.

7% of data spread). Sampson's iterative fit produced an ellipse with low eccentricity that was qualitatively similar to the one produced by our direct method (solid lines) but the *total* cost of our method is the same as that of acquiring his initial estimate. As anticipated in the previous section, the low eccentricity bias of our method is most evident in Fig. 1b when compared to Bookstein's, Taubin's, and Gander's results. It must be again remarked that this is not surprising, because those methods are not ellipse-specific, whereas ours is.

4.3 Noise Sensitivity

In this section, we describe some experiments concerning the noise performance of our method compared to others.

The first experiment is concerned with the stability of the estimated ellipse center with increasing noise levels. We consider a whole ellipse centered at the origin of semi-axis 1 and 0.5 and rotated by 40 degrees. The sampled ellipse was corrupted with noise (from 2^{-3} to 2^3) for 100 runs and the distance between the true ellipse center and the center of the conic returned by the fitting algorithm was recorded. Returned hyperbolae were included for the other algorithms. Fig. 2a shows the 90th percentile error in the centers as a function of noise level. At low noise levels (σ < 0.5), all algorithms can be seen to perform similarly, while at high levels, only the new (B2AC) algorithm degrades gracefully.

The good performance of the presented method is more evident when the data is occluded. In the second experiment, shown in Fig. 2b, increasing level of isotropic Gaussian was added to points on a given elliptical arc. The standard deviation of the noise varies from 3% in the leftmost column to 20% of data spread in the rightmost column; the noise has been set to a relatively high level because the performance of the three algorithms is substantially the same at low noise level of precise elliptical data. The top row shows the results for the method proposed here. Although, as expected, the fitted ellipses shrink with increasing levels of noise [8] (in the limit, the elliptical arc will look like a noisy line), it can be noticed that the ellipse dimension decreases smoothly with the noise level: This is an indication of well-behaved fitting. This shrinking phenomenon is evident also with the other methods but presents itself more erratically. Many more quantitative experiments on performance with occluded data can be found in [3].

The last experiment that we show here is perhaps the most interesting (although we have not seen it in related papers) and is concerned with assessing stability to different realizations of noise with the same variance. It is very desirable that an algorithm's performance be affected only by the noise level, and not by a particular realization of the noise. Fig. 3 shows five different runs for σ = 0.1, and the results of our method, Gander's method, and Taubin's method are given. This and similar experiments (see [11], [3]) show that our method has a greater stability to noise than the other methods.

5 Conclusion

In this paper, we have presented a least squares fitting method which is specific to ellipses and direct at the same time. Previous methods were either not ellipse-specific or were iterative.

We have theoretically demonstrated that our method uniquely yields elliptical solutions that, under the normalization $4ac-b^2=1$, minimize the sum of squared algebraic distances from the points to the ellipse.

Experimental results illustrate the advantages conferred by the ellipse-specificity in terms of occlusion and noise sensitivity. The stability properties widen the scope of application of the algorithm from ellipse fitting to cases where the data are not strictly elliptical but need to be minimally represented by an elliptical "blob."

In our view, the method presented here offers the best trade-off between speed and accuracy for ellipse fitting, and its uniqueness property makes it also extremely robust to noise and usable in many applications, especially in industrial vision. In cases where more accurate results are required, this algorithm provides an excellent initial estimate.

Its simplicity is demonstrated by the inclusion in Fig. 4 of a complete six-line implementation in Matlab. (An interactive Java demonstration is available at http://vision.dai.ed.ac.uk/maurizp/ElliFitDemo/demo.html.)

Future work includes the incorporation of the algorithm into a bias-correction algorithm based on that of Kanatani [8]. We note also that the algorithm can be trivially converted to a hyperbolaspecific fitter, and a variation may be used to fit parabolae.

REFERENCES

- [1] F.L. Bookstein, "Fitting Conic Sections to Scattered Data," *Computer Graphics and Image Processing*, no. 9, pp. 56-71, 1979.
- [2] T. Ellis, A. Abbood, and B. Brillault, "Ellipse Detection and Matching With Uncertainty," *Image and Vision Computing*, vol. 10, no. 2, pp. 271-276, 1992.
- [3] A.W. Fitzgibbon, "Stable Segmentation of 2D Curves," PhD thesis, Dept. of Artificial Intelligence, Univ. of Edinburgh, 1998.
- [4] A.W. Fitzgibbon and R.B. Fisher, "A Buyer's Guide to Conic Fitting," Proc. British Machine Vision Conf., Birmingham, England, 1995.
- [5] W. Gander, G.H. Golub, and R. Strebel, "Least-Square Fitting of Circles and Ellipses," *BIT*, no. 43, pp. 558-578, 1994.
- [6] R. Gnanadesikan, Methods for Statistical Data Analysis of Multivariate Observations. New York: Wiley, 1977.
- [7] R. Haralick and L. Shapiro, Computer and Robot Vision. Reading, Mass.: Addison-Wesley, 1992.
- [8] K. Kanatani, "Statistical Bias of Conic Fitting and Renormalization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 3, pp. 320-326, 1994.
- [9] V.F. Leavers, Shape Detection in Computer Vision Using the Hough Transform. New York: Springer-Verlag, 1992.
- [10] The MathWorks. The Matlab Package.
- [11] M. Pilu, "Part-Based Grouping and Recognition: A Model-Guided Approach," Dept. Artificial Intelligence, Univ. of Edinburgh, PhD thesis, Aug. 1996.
- [12] J. Porrill, "Fitting Ellipses and Predicting Confidence Envelopes Using a Bias Corrected Kalman Filter," *Image and Vision Computing*, vol. 8, no. 1, pp. 37-41, Feb. 1990.
- [13] S.S. Rao, Optimization: Theory and Applications, 2nd ed. New York: Wiley Estern, 1984.
- [14] P.L. Rosin, "A Note on the Least Squares Fitting of Ellipses," Pattern Recognition Letters, no. 14, pp. 799-808, Oct. 1993.
- [15] P.L. Rosin and G.A. West, "Nonparametric Segmentation of Curves Into Various Representations," *IEEE Trans. Pattern Analysis* and Machine Intelligence, vol. 17, no. 12, pp. 1,140-1,153, Dec. 1995.
- [16] P.D. Sampson, "Fitting Conic Sections to Very Scattered Data: An Iterative Refinement of the Bookstein Algorithm," Computer Graphics and Image Processing, no. 18, pp. 97-108, 1982.
- [17] G. Taubin, "Estimation of Planar Curves, Surfaces and Non-Planar Space Curves Defined by Implicit Equations, With Applications to Edge and Range Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 11, pp. 1,115-1,138, Nov. 1991.

- [18] J.H. Wilkinson, The Algebraic Eigenvalue Problem. Oxford, England: Clarendon Press, 1965.
- [19] H.K. Yuen, J. Illingworth, and J. Kittler, "Shape Using Volumetric Primitives," *Image and Vision Computing*, vol. 1, no. 7, pp. 31-37, 1989.