Direct Least Squares Fitting of Ellipses

Andrew W. Fitzgibbon Maurizio Pilu Robert B. Fisher
Department of Artificial Intelligence
The University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL
SCOTLAND
andrewfg@ed.ac.uk

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 new method combines several advantages: (i) It is ellipse-specific so that even bad data will always return an ellipse; (ii) It can be solved naturally by a generalized eigensystem and (iii) it is extremely robust, efficient and easy to implement. We compare the proposed method to other approaches and show its robustness on several examples in which other non-ellipse-specific approaches would fail or require computationally expensive iterative refinements. Source code for the algorithm is supplied and a demonstration is available on* `http://vision.dai.ed.ac.uk/maurizp/ElliFitDemo/demo.html`

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 [15, 6].

In this paper we introduce a new method of fitting ellipses, rather than general conics, to segmented data. As we shall see in the next section, current methods are either computationally expensive Hough transform-based 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 — non-strictly 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 [18, 8, 13] 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 advantages of: *Ellipse-specificity* yielding useful results under all noise and occlusion conditions; *Affine invariance*; and *Robustness* to noise and occlusion.

After a description of previous algebraic 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 noise resilience, invariance properties and behaviour for non-elliptical data. We conclude by presenting some possible extensions.

2 Previous Methods and their Limitations

The literature on ellipse fitting divides into two general techniques: clustering and least-squares fitting.

Clustering methods are based on mapping sets of points to the parameter space, such as the Hough transform [10, 22] and accumulation methods [17]. These Hough-like techniques have some great advantages, notably high robustness to occlusion and no requirement for pre-segmentation, but they suffer from the great shortcomings of high computational complexity and non-uniqueness of solutions, which can render them unsuitable for real applications. Particularly when curves have been pre-segmented, their computational cost is significant.

1015-4651/96 $5.00 © 1996 IEEE
Proceedings of ICPR '96

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.

2.1 Problem statement

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:

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

where $\mathbf{a} = [a\ b\ c\ d\ e\ f]^T$ and $\mathbf{x} = [x^2\ xy\ y^2\ x\ y\ 1]^T$. $F(\mathbf{a}; \mathbf{x}_i)$ 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 [7] by minimizing the sum of squared algebraic distances

$$\sum_{i=1}^{N} F(\mathbf{x}_i)^2 \quad (2)$$

of the curve to the N data points $\mathbf{x}_i$. 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:

- Many authors [11] suggest $||\mathbf{a}||^2 = 1$.
- Bolles and Fischler [2] suggest $a = 1$.
- Rosin [15] and Gander [6] impose $a + c = 1$.
- Rosin [15] also investigates $f = 1$.
- Bookstein [3] proposes $a^2 + \frac{1}{2}b^2 + c^2 = 1$.
- Agin [1] and Taubin [20] use the data-dependent quadratic constraint $||N\mathbf{a}||^2 = 1$ where N is the Jacobian $[\nabla F(\mathbf{a}; \mathbf{x}_1) \ldots \nabla F(\mathbf{a}; \mathbf{x}_N)]^T$.

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

A number of papers have concerned themselves with the specific problem of recovering ellipses rather than general conics. The above methods do not restrict the fitting to be an ellipse, in the sense that they can return an hyperbola or a parabola even given elliptical data.

Porrill [13] and Ellis *et al.* [4] use Bookstein's method to initialize a Kalman filter. The Kalman filter iteratively minimizes the gradient distance in order to gather new image evidence and to reject non-ellipse fits by testing the discriminant $b^2 - 4ac < 0$ at each iteration. Porrill also gives nice examples of the confidence envelopes of the fittings. Although these methods transform the disadvantage of having a non-specific ellipse fitting method into an asset by using the ellipse constraint to check whether new data has to be included or to assess the quality of the fit, the methods require many iterations in the presence of very bad data, and may fail to converge in extreme cases.

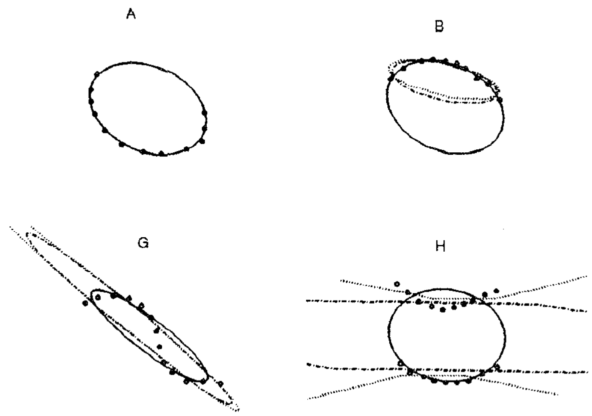

Figure 1. Hand-drawn data. The linetype / algorithm correspondences are Bookstein: dotted; Taubin: dash-dot; Ellipse-specific: solid.

Rosin [15] initializes a Kalman Filter with a circle, and tightens the corresponding initial covariance until the returned conic is an ellipse. In [16] he restates that ellipse-specific fitting is a non-linear problem and that iterative methods must be employed. He also [15] analyses the pro and cons of two commonly used normalizations, $f = 1$ and $a + c = 1$ and shows that the former biases the fitting to have smaller eccentricity, therefore increasing the probability of returning an ellipse, at the cost of losing transformational invariance.

Haralick [7, §11.10.7] takes a different approach. Effectively, he guarantees that the conic is an ellipse by replacing the coefficients $\{a, b, c\}$ with new expressions $\{p^2, 2pq, q^2 + r^2\}$ so that the discriminant $b^2 - 4ac$ becomes $-4p^2r^2$ which is guaranteed negative. Minimization over the space $\{p, q, r, d, e, f\}$ then yields an ellipse. His algorithm is again iterative, and an initial estimate is provided by a method of moments. Keren *et al.* [9] apply a similar technique to Haralick's in the context of the fitting of bounded quartic curves. Again, their algorithm is iterative.

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 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 [14] 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 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 $4ac - b^2 = 1$.

This is a quadratic constraint which may be expressed in the matrix form $\mathbf{a}^T \mathbf{C} \mathbf{a} = 1$ as

$$\mathbf{a}^T \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{a} = 1 \tag{3}$$

3.1 Solution of the quadratically constrained minimization

Following Bookstein [3], the constrained fitting problem is:

$$\text{Minimize } E = \|\mathbf{D}\mathbf{a}\|^2, \text{ subject to } \mathbf{a}^T \mathbf{C} \mathbf{a} = 1 \tag{4}$$

where the *design matrix* $\mathbf{D}$ is defined as the $n \times 6$ matrix $[\mathbf{x}_1\ \mathbf{x}_2 \cdots\ \mathbf{x}_n]^T$. Introducing the Lagrange multiplier λ and differentiating we arrive at the system of simultaneous equations[1]

$$\boxed{\begin{aligned} 2\mathbf{D}^T\mathbf{D}\mathbf{a} - 2\lambda \mathbf{C}\mathbf{a} &= 0 \\ \mathbf{a}^T \mathbf{C}\mathbf{a} &= 1 \end{aligned}} \tag{5}$$

This may be rewritten as the system

$$\mathbf{S}\mathbf{a} = \lambda \mathbf{C}\mathbf{a} \tag{6}$$

$$\mathbf{a}^T \mathbf{C}\mathbf{a} = 1 \tag{7}$$

where $\mathbf{S}$ is the *scatter matrix* $\mathbf{D}^T\mathbf{D}$. This system is readily solved by considering the generalized eigenvectors of (6). If $(\lambda_i, \mathbf{u}_i)$ solves (6) then so does $(\lambda_i, \mu\mathbf{u}_i)$ for any μ and from (7) we can find the value of μ_i as $\mu_i^2 \mathbf{u}_i^T \mathbf{C}\mathbf{u}_i = 1$ giving

$$\mu_i = \sqrt{\frac{1}{\mathbf{u}_i^T \mathbf{C}\mathbf{u}_i}} = \sqrt{\frac{\lambda_i}{\mathbf{u}_i^T \mathbf{S}\mathbf{u}_i}} \tag{8}$$

[1] Note that the method of Lagrange multipliers is not valid when the gradient of the constraint function becomes zero. In (4) this means $\mathbf{C}\mathbf{a} = 0$, but then $\mathbf{a}^T \mathbf{C}\mathbf{a} = 0$ so the constraint is violated and there is no solution.

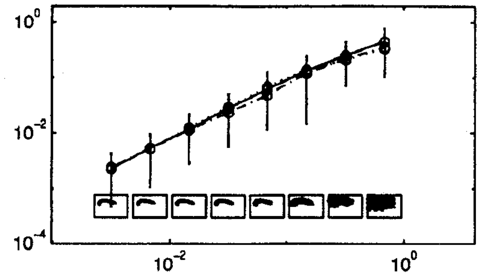

Figure 2. Average geometric distance error as a function of increasing noise level. The errorbars are at $\pm 1\sigma$. The pictures along the noise axis indicate visually the corresponding noise level. Encoding is Bookstein: dotted; Taubin: dash-dot; New: solid.

Finally, setting $\hat{\mathbf{a}}_i = \mu_i \mathbf{u}_i$ solves (5). As in general there may be up to 6 real solutions, the solution is chosen that yields the lowest residual $\hat{\mathbf{a}}_i^T S \hat{\mathbf{a}}_i = \lambda_i$.

We note that the solution of the eigensystem (6) gives 6 eigenvalue-eigenvector pairs $(\lambda_i, \mathbf{u}_i)$. Each of these pairs gives rise to a local minimum if the term under the square root in (8) 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 $\mathbf{u}_i$. Therefore the square root exists if $\lambda_i > 0$, so any solutions to (5) must have positive generalized eigenvalues. A Cholesky decomposition $S = Q^2$ reveals that the signs of the eigenvalues of (6) are the same as those of $Q^{-1}CQ^{-1}$, and therefore, by Sylvester's law of Inertia [21] the same as those of C. The constraint matrix (3) for the ellipse-specific problem has only one positive eigenvalue, and therefore the elliptical solution is unique [12].

3.2 Observation

Solving (4) is equivalent to minimizing the *Rayleigh quotient* $\frac{\mathbf{a}^T\mathbf{D}\mathbf{a}}{\mathbf{a}^T\mathbf{C}\mathbf{a}}$, which in this case is the cost function

$$\frac{1}{4ac - b^2} \sum_{i=1}^{N} F(\mathbf{x}_i)^2$$

Now, as the discriminant is inversely proportional to the product of the radii, this is analogous to multiplying the cost function by the area of the ellipse. Interestingly, this is the technique adopted by Solina and Bajcsy for fitting superquadrics to range data [19], in order to bias the fit towards smaller volume. This provides an intuitive feeling for the low-curvature bias of the method.

4 Experimental Results

In this section we present experimental results that compare the ellipse-specific solution to previous methods in terms of quality and robustness. We include both quantitative and qualitative results in order to allow other researchers to evaluate the utility of the ellipse-specific algorithm with respect to the others cited. Fitzgibbon [5] provides further theoretical and quantitative results for a wide range of conic-fitting algorithms.

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 Figure 1 provides some examples with hand-drawn datasets. The results of our 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 non-ellipticity, and illustrate the potential use of ellipse-specificity for coarse data bounding of nonelliptical data.

4.2 Noise sensitivity

The first noise experiment measures the average geometric distance error for each of the algorithms over 100 runs. In order to verify that the ellipses returned by the new algorithm are reasonable approximations to the minimum geometric distance ellipse, non-elliptical fits returned by the Bookstein and Taubin algorithms were ignored. It can be seen that our algorithm produces a closer ellipse on average than Bookstein's for medium noise, but that Taubin's—when it returns an ellipse—produces the smallest geometric distance error. We note however that all results are within each other's 1σ error bars over the 100 runs, meaning that the variations within runs are greater than the difference between the algorithms across runs.

The second experiment, illustrated in Figure 3, is perhaps more important (although we have not seen it in related papers) and is concerned with assessing the stability of the fitting with respect to different realizations of noise with the *same variance*. It is very desirable that the algorithm performance be affected only by the noise level, and not by a particular realization of the noise. Figure 3 shows ten different runs in which a different noise population with same variance ($\sigma = 0.1$) was generated and results for each of the three methods is displayed. In this and similar experiments we found that the stability of the method is noteworthy. Gander's algorithm shows a greater variation in results and Taubin's, while improving on Gander's, remains less stable than the proposed algorithm.

```
% x,y are lists 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 eigensystem
[gevec, geval] = eig(inv(S)*C);
% Find the positive eigenvalue
[PosR, PosC] = find(geval>0 & ~isinf(geval));
% Extract eigenvector corresponding to positive eigenvalue
a = gevec(:,PosC);
```

Figure 4. Complete 6-line Matlab implementation of the proposed algorithm.

4.3 Affine invariance

The constraint $4ac - b^2 = 1$ not only constrains the fitted conics to be ellipses but it is also covariant with affine transformations of the data points. In an experimental determination, we applied random nonsingular affine transforms to the points of a data set, applied the fitting algorithm, inverse-transformed the ellipse parameters by the original transform and compared the recovered parameters to the expected ones. The difference was zero up to machine precision.

5 Conclusions

This paper has presented a new method for direct least square fitting of ellipses. We believe this to be the first noniterative ellipse-specific algorithm. Previous conic fitting methods rely (when applied to ellipse fitting) either on the presence of good data or on computationally expensive iterative updates of the parameters.

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 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 tradeoff between speed and accuracy for ellipse fitting—its simplicity is demonstrated by the inclusion in Figure 4 of a complete 6-line implementation in MATLAB. In cases where more accurate results are required, this algorithm provides an excellent initial estimate. We note also that the algorithm

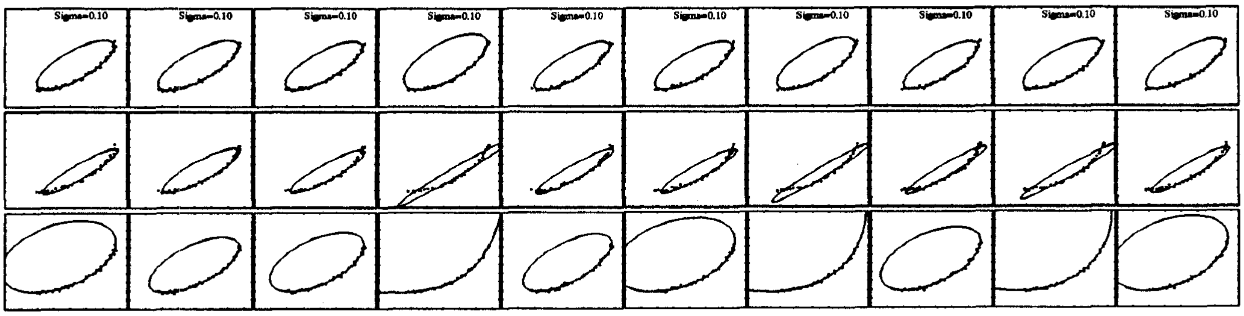

Figure 3. Stability experiments for different runs with same noise variance. Top row: proposed method; Mid Row: Gander's Method; Bottom Row: Taubin's method

can be trivially converted to a hyperbola-specific fitter, and a variation may be used to fit parabolae.

The algorithm is however biased towards ellipses of low eccentricity, and future work includes the incorporation of the algorithm into a bias-correction algorithm based on that of Kanatani [8]. In a similar vein, a theoretical analysis of the noise performance of the methods using eigensystem perturbation theory is under investigation.

6 Acknowledgements

The second author is partially sponsored by SGS-THOMSON Microelectronics UK. This work was partially funded by UK EPSRC Grant GR/H/86905.

References

[1] G. J. Agin. *Representation and description of curved objects.* PhD thesis, AIM-173, Stanford AI Lab, 1972.

[2] R. C. Bolles and M. A. Fischler. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In *Proceedings, International Joint Conference on AI*, pages 637–643, 1981.

[3] F. Bookstein. Fitting conic sections to scattered data. *Computer Graphics and Image Processing*, (9):56–71, 1979.

[4] T. Ellis, A. Abbood, and B. Brillault. Ellipse detection and matching with uncertainty. *Image and Vision Computing*, 10(2):271–276, 1992.

[5] A. Fitzgibbon and R. Fisher. A buyer's guide to conic fitting. In *Proceedings of British Machine Vision Conference*, Birmingam, 1995.

[6] W. Gander, G. Golub, and R. Strebel. Least-square fitting of circles and ellipses. *BIT*, (43):558–578, 1994.

[7] R. M. Haralick and L. G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley, 1993.

[8] K. Kanatani. Statistical bias of conic fitting and renormalization. *IEEE T-PAMI*, 16(3):320–326, 1994.

[9] D. Keren, D. Cooper, and J. Subrahmonia. Describing complicated objects by implicit polynomials. *IEEE T-PAMI*, 16(1):38–53, 1994.

[10] V. Leavers. *Shape Detection in Computer Vision Using the Hough Transform*. Springer-Verlag, 1992.

[11] K. A. Paton. Conic sections in automatic chromosome analysis. In B. Meltzer and D. Michie, editors, *Machine Intelligence 5*, pages 411–434. Edinburgh University Press, Edinburgh, 1969.

[12] M. Pilu. *Part-based Grouping and Recogntion: A Model-Guided Approach.* PhD Thesis, Department of Artificial Intelligence, University of Edinburgh, Scotland, 1996. Jun, 1996.

[13] J. Porrill. Fitting ellipses and predicting comfidence envelopes using a bias corrected Kalman filter. *Image and Vision Computing*, 8(1), Feb. 1990.

[14] S. S. Rao. *Optimization: Theory and Applications.* Wiley Eastern, 2 edition, 1984.

[15] P. Rosin. A note on the least square fitting of ellipses. *Pattern Recognition Letters*, (14):799–808, Oct. 1993.

[16] P. Rosin and G. West. Segmenting curves into lines and arcs. In *Proceedings of the Third International Conference on Computer Vision*, pages 74–78, Osaka, Japan, Dec. 1990.

[17] P. L. Rosin. Ellipse fitting by accumulating five-point fits. *Pattern Recognition Letters*, (14):661–699, Aug. 1993.

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

[19] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case of superquadrics with global deformations. *IEEE PAMI*, 12(2), Feb. 1990.

[20] 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 PAMI*, 13(11):1115–1138, Nov. 1991.

[21] J. H. Wilkinson. *The algebraic eigenvalue problem*. Claredon Press, Oxford, England, 1965.

[22] H. K. Yuen, J. Illingworth, and J. Kittler. Detecting partially occluded ellipses using the Hough transform. *Image and Vision Computing*, 7(1):31–37, 1989.