Analysing error of fit functions for ellipses

Paul L. Rosin

Department of Computer Science and Information Systems

Brunel University

Middlesex, UB8 3PH

email: Paul.Rosin@brunel.ac.uk

Abstract

We describe several established error of fit (EOF) functions for use in the least square fitting of ellipses, and introduce a further four new EOFs. Four measures are given for assessing the suitability of such EOFs, quantifying their linearity, curvature bias, asymmetry, and overall goodness. These measures enable a better understanding to be gained of the individual merits of the EOF functions.

1   Introduction

The fitting of ellipses to edge data is a common task in computer vision. In particular, this often arises in the context of industrial inspection since circular parts in the scene are projected into the image as ellipses. There are many algorithms for ellipse fitting, but in this paper we shall concentrate on minimisation techniques rather than others such as the Hough transform voting method. Despite its sensitivity to non-Gaussian noise, least squares (LS) fitting is probably the most widely used approach for estimating the ellipses' parameters, due to its computation efficiency and its high efficiency as an estimator. It operates by minimising the sum of squares of some error term e_j measured at each data point $\mathbf{X}_j = (x_j, y_j)$. Thus for N points the parameters $\mathbf{P}$ of the best fit ellipse are obtained by

$$\min_{\mathbf{P}} \sum_{j=1}^{N} e_j^2 .$$

Similarly, other estimation techniques such as the least median of squares depend on a suitable error term. However, there has been little comparison between the effects of the choice of error term (but see [3, 8, 9]). The Euclidean distance from $\mathbf{X}_j$ to the ellipse boundary would be a good choice for e_j, but requires solving a quartic equation which may have up to four solutions, requiring the one with the minimum distance to be determined [10]. To avoid the complexity of evaluating the true Euclidean distance it is usual practice to approximate it by some measure – the error of fit (EOF) function – that is simpler to calculate. Previously we described six approximations [9] (EOF_1 - EOF_6). In this paper we describe and analyse a further set of approximations (EOF_7 - EOF_{13}).

2 Distance Approximations

2.1 Algebraic distance

The simplest approximation is the algebraic distance [2] calculated from $\mathbf{X}_j$ to the ellipse

$$\text{EOF}_1 = Q(x_j, y_j) = Ax_j^2 + Bx_jy_j + Cy_j^2 + Dx_j + Ey_j + F, \tag{1}$$

where $Q(x, y) = 0$ is the general equation of a conic, describing ellipses when $B^2 - 4AC < 0$. An advantage of (1) is that a closed form LS solution is available whereas the other EOFs described below must be minimised by iterative procedures. Figure 1a shows the iso-value contours for EOF_1 (the ellipse boundary is drawn bold) from which we can make several observations. First, EOF_1 demonstrates the so called "high curvature bias" in that the spacing between the contours become wider at the pointed ends of the ellipse, i.e. near the regions of high curvature. Second, the contours get closer with increasing distance out from the ellipse. The effect of the curvature bias is to cause data near the ends of the ellipse to have less influence on the fit, often resulting in an overestimate of the eccentricity of the fitted ellipse. The second factor is the relationship between the EOF and the Euclidean distance as a function of the Euclidean distance. Ideally they should be linearly related; a constant scaling factor has no effect and can be ignored. The super-linear relationship shown by EOF_1 causes the outlying data to have a stronger influence on the fit than a linear or sub-linear relationship, and therefore increases the sensitivity of the fitting to noise.

2.2 Gradient weighted algebraic distance

The most commonly used improvement over the algebraic distance is to inversely weight $Q(x_j, y_j)$ by its gradient [1, 11, 13]

$$\text{EOF}_2 = \frac{Q(x_j, y_j)}{|\nabla Q(x_j, y_j)|}. \tag{2}$$

The plot of the iso-value contours (figure 1b) shows that the curvature bias is indeed reduced, although the magnitude of the gradient of the measure inside the ellipse is considerably larger than on the outside.[1] This asymmetry between the distance approximation inside and outside the ellipse means that the data inside the ellipse has more influence than the exterior data, which could cause the size of the fitted ellipse to be underestimated.

2.3 Second order approximation

Nalwa and Pauchon [6] used a second order Taylor series expansion[2] to approximate the distance d, expanding about $\mathbf{X}_j$. The distance is then obtained by

---

[1] Artifacts from the plotting process have caused some contours to be missed near the centre of the ellipse.

[2] Taubin [13] showed that EOF_2 is equivalent to a first order Taylor series expansion.

(a) EOF$_1$

(b) EOF$_2$

(c) EOF$_7$

(d) EOF$_8$

(e) EOF$_9$

(f) EOF$_{10}$

(g) EOF$_{11}$

(h) EOF$_{12a}$
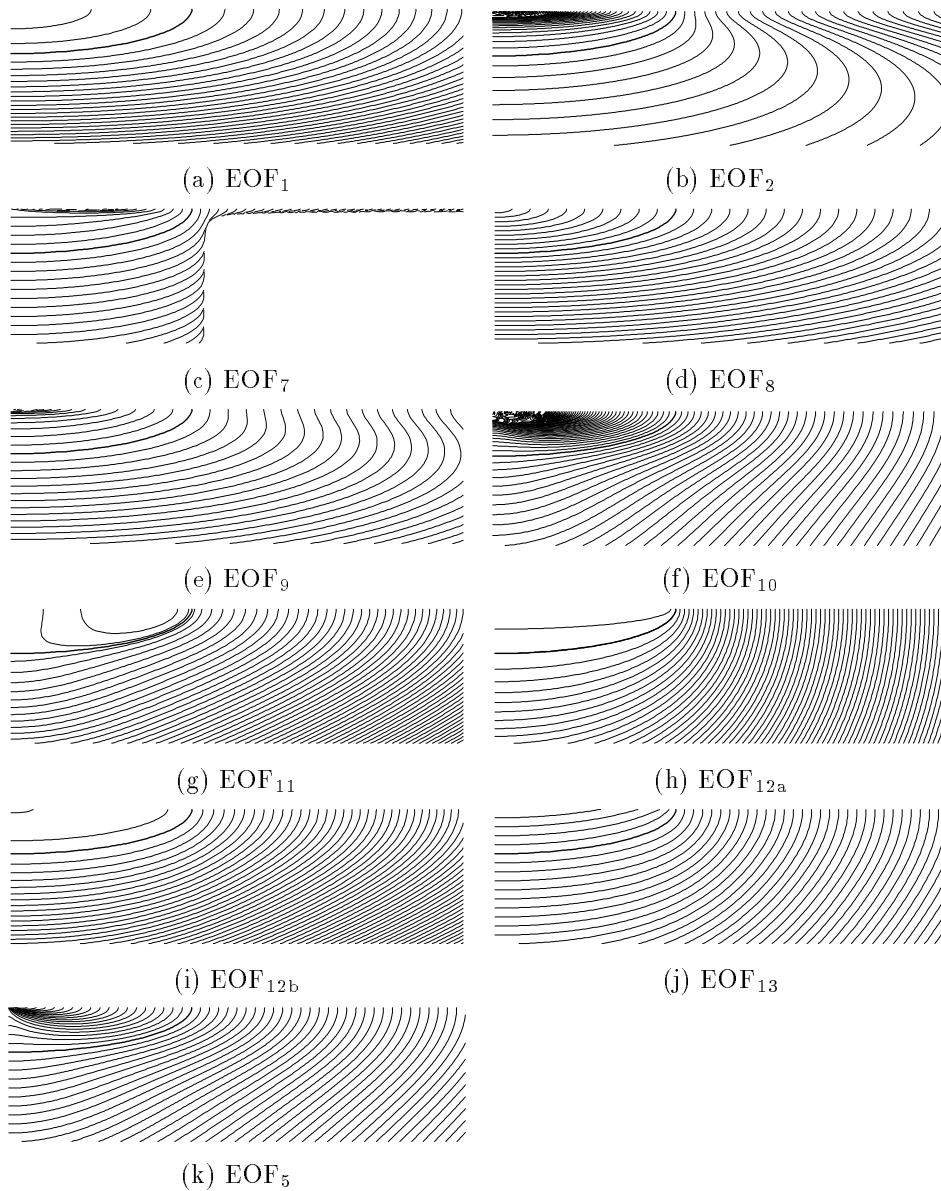
(i) EOF$_{12b}$

(j) EOF$_{13}$

(k) EOF$_5$

Figure 1: Iso-value contours

solving the resulting quadratic equation, obtaining

$$\text{EOF}_7 = d = \min \left| \frac{Q'(x_j, y_j) \pm \sqrt{Q'^2(x_j, y_j) - 2Q(x_j, y_j)Q''(x_j, y_j)}}{Q''(x_j, y_j)} \right|. \qquad (3)$$

where

$$
\begin{aligned}
Q'(x_j, y_j) &= |\nabla Q(x_j, y_j)| = \sqrt{Q_x^2 + Q_y^2} \\
Q''(x_j, y_j) &= \frac{Q_{xx}\, Q_x^2 + 2\, Q_{xy}\, Q_x\, Q_y + Q_{yy}\, Q_y^2}{|\nabla Q|^2}
\end{aligned}
$$

and all coefficients are evaluated at $\mathbf{X}_j$. This approximation does not hold when the roots of the quadratic are imaginary. In practise, we have found that for the high curvature sections of eccentric ellipses real roots can only be found close to the boundary. This is demonstrated in figure 1c where the blank right hand portion contains imaginary roots.

2.4 Pavlidis' approximation

Pavlidis [7] provides a distance measure which is an improvement over the basic algebraic distance, defined as

$$\text{EOF}_8 = \sqrt{Ax_j^2 + Bx_j y_j + Cy_j^2 + Dx_j + Ey_j + v} - \sqrt{F + v} \qquad (4)$$

where

$$v = \frac{1}{4} \begin{bmatrix} D & E \end{bmatrix} \begin{bmatrix} A & \frac{B}{2} \\ \frac{B}{2} & C \end{bmatrix}^{-1} \begin{bmatrix} D \\ E \end{bmatrix}.$$

The improvement of the algebraic distance is evident in the iso-value plot in figure 1d.

2.5 Reduced gradient weighted algebraic distance

We now describe four new error of fit functions which also provide some improvements over the algebraic distance. The first is based on the observation that the algebraic distance EOF_1 suffers from severe curvature bias while the weighted algebraic distance EOF_2 suffers from severe asymmetry. Both these factors degrade the effectiveness of the measures. Possibly some intermediate measure which gives the $|\nabla Q(x_j, y_j)|$ term less weighting might produce a superior compromise, especially since their gradients have opposite signs. This is simply achieved by raising the weighting term to some power $p = [0, 1]$.

$$\text{EOF}_9 = \frac{Q(x_j, y_j)}{|\nabla Q(x_j, y_j)|^p}. \qquad (5)$$

Thus EOF_9 reduces to EOF_1 and EOF_2 respectively when $p = 0$ and $p = 1$. Figure 1e shows the iso-value contours obtained when $p = 0.45$; it can be seen that both the curvature bias and asymmetry are moderate.

2.6   Directional derivative weighted algebraic distance

Although the gradient term in EOF_2 corrected much of the high curvature bias of EOF_1 it also introduced some other curvature bias, producing the wavy iso-value contours in figure 1b. These waves are avoided if we replace $|\nabla Q(x_j, y_j)|$ by the directional derivative of $Q(x_j, y_j)$ in the direction along the ray $\mathbf{r} = \langle \cos \theta, \sin \theta \rangle$ drawn between $\mathbf{X}_j$ and the centre of the ellipse $\mathbf{C}$, and we get

$$
\begin{aligned}
\text{EOF}_{10} &= \frac{Q(x_j, y_j)}{\mathbf{r} \cdot \nabla Q(x_j, y_j)} \\
&= \frac{Q(x_j, y_j)}{\cos \theta \, (2Ax_j + By_j + D) + \sin \theta \, (Bx_j + 2Cy_j + E)}.
\end{aligned}
\tag{6}
$$

The resulting iso-value contours (figure 1f) still have the asymmetry present in EOF_2, but the contours appear more regular.

2.7   Combined conic and circular distance

The iso-value contours for the algebraic distance show the effects of the curvature bias: at the pointed ends of the ellipse the distance is underestimated. If, instead of the algebraic distance of the ellipse, the algebraic distance of the inscribed circle (i.e. setting $B = 0$ and $A = C$) were used this would cause the opposite effect: the distance would be overestimated at the pointed ends of the ellipse. This suggests that the effects of curvature bias could be reduced by combining the two terms. We ignore constant values and just use the simplified symmetric term

$$
C(x_j, y_j) = (x_j - x_c)^2 + (y_j - y_c)^2
$$

and combine the terms as their geometric mean:

$$
\text{EOF}_{11} = \sqrt{|Q(x_j, y_j)| C(x_j, y_j)}
\tag{7}
$$

This produces iso-value contours (figure 1g) which have quite low curvature bias, although the asymmetry is substantial.

2.8   Concentric ellipse estimation

A property of ellipses is that for a point $\mathbf{P}$ on the ellipse with two foci $\mathbf{F}$ and $\mathbf{F}'$ then $\overline{\mathbf{PF}} + \overline{\mathbf{PF}'} = 2a$. We use this to make an approximation to the major axis of the ellipse with similar proportions through $\mathbf{X}_j$ as $\overline{\mathbf{X}_j \mathbf{F}} + \overline{\mathbf{X}_j \mathbf{F}'} = 2\tilde{a}$. We take the approximate distance as

$$
\text{EOF}_{12a} = \tilde{a} - a.
\tag{8}
$$

This is the same as the first part of Stricker's [12] two step method for estimating $\tilde{a}$. Plotting the iso-level contours in figure 1h reveals significant low curvature bias. We describe an alternative to Stricker's subsequent reestimation step. In a similar fashion to the combined conic and circular distance above we correct the curvature bias by combining the algebraic distance with EOF_{12a} to get

$$
\text{EOF}_{12b} = \sqrt{|Q(x_j, y_j)| (\tilde{a} - a)}.
\tag{9}
$$

The iso-value contours (figure 1i) are much improved, showing little curvature bias although there is significant asymmetry.

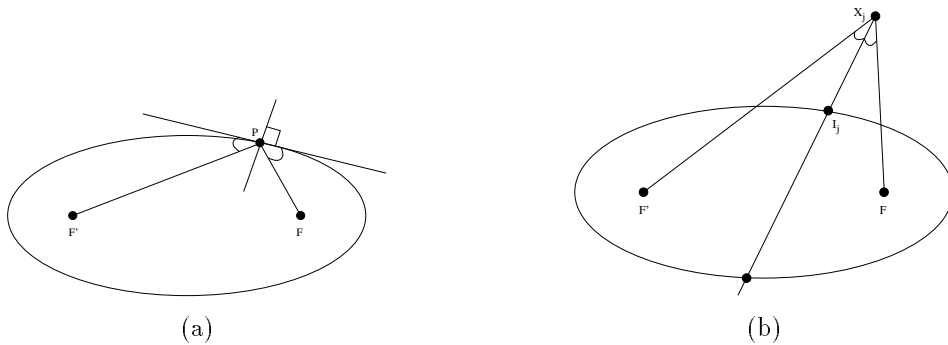## 2.9 Angular bisector of lines to foci

$$(a) \qquad\qquad (b)$$

Figure 2: Angular bisector of lines through $\mathbf{X}_j$ and the foci

The final distance approximation is based on the focal (reflection) property of ellipses: the lines $\mathbf{FP}$ and $\mathbf{F'P}$ make equal angles with the tangent to the curve at the point $\mathbf{P}$ on the ellipse (figure 2a). Thus the normal to the ellipse at $\mathbf{P}$ is the angular bisector of $\mathbf{FP}$ and $\mathbf{F'P}$. As shown in figure 2b we approximate the normal to the ellipse from $\mathbf{X}_j$ as the angular bisector of $\mathbf{X}_j \mathbf{F}$ and $\mathbf{X}_j \mathbf{F'}$. It would be expected to provide a good approximation for points close to the ellipse. We find the closest point of intersection $\mathbf{I}_j$ of the bisecting line with the ellipse, and take the distance between $\mathbf{X}_j$ and $\mathbf{I}_j$

$$\text{EOF}_{13} = \overline{\mathbf{X}_j\,\mathbf{I}_j}. \qquad (10)$$

As can be seen from the iso-value plot in figure 1j this appears to provide a very good approximate distance.

For comparison, the more easily determined ray through $\mathbf{X}_j$ and the center of the ellipse is also considered. The distance from $\mathbf{X}_j$ to the closest point of intersection between the ray and the ellipse was used by Nakagawa and Rosenfeld [5] as an EOF function. It can be seen from the resulting iso-value plot (figure 1k) that EOF_{13} is considerably better.

3 Assessment Measures

Despite the variety of approximations to the distance from a point to the boundary of an ellipse little comparative analysis has been carried out.[3] Although visualising the iso-value contours is a useful tool for the qualitative analysis of EOFs, a quantitative assessment would be valuable for objectively comparing them. We develop

---

[3]Gross and Boult [4] experimentally evaluated four EOFs for superquadrics using graphical plots comparing scaled summed EOF against the true RMS error, as well as showing one dimensional cross-sections of the errors of fit.

assessment measures based on the three factors of deviation from the Euclidean distance described above. One consideration is that the degree of deviation may not be constant with increasing distance from the ellipse. Therefore it may be necessary to make the measures a function of the distances. Second, although we show the measures applied to all values $\mathbf{X} = (x, y)$ in the $\mathcal{R}^2$ plane, in practice we must discreetly sample a subset of $\mathcal{R}^2$. This leads to the problem of which subset? One solution is to assume a noise model which, given an ellipse, specifies where the data is expected to occur. For instance, for noise with a Gaussian distribution $N(0, \sigma)$ we can weight the data in $\mathcal{R}^2$ by

$$w(d) = \frac{1}{\sigma\sqrt{2\pi}} e^{-d^2/2\sigma^2}$$

where d is the Euclidean distance to the ellipse boundary.

3.1   Linearity

To test for a linear relationship between the Euclidean distance values $d_{\mathbf{X}}$ and their approximations $e_{\mathbf{X}}$ it is natural to use the Pearson correlation coefficient

$$\rho = \frac{\sum_{\mathbf{X}}(e_{\mathbf{X}} - \bar{e}_{\mathbf{X}})(d_{\mathbf{X}} - \bar{d}_{\mathbf{X}})}{\sqrt{\sum_{\mathbf{X}}(e_{\mathbf{X}} - \bar{e}_{\mathbf{X}})^2 \sum_{\mathbf{X}}(d_{\mathbf{X}} - \bar{d}_{\mathbf{X}})^2}} \tag{11}$$

The value of ρ lies between ± 1, although in our context negative values are unlikely, and we can think of $\rho \in [0, 1]$ with increasing values meaning a better linear correlation.

A problem with (11) is that equidistant values further from the ellipse will have greater effect on the measure than close values since as iso-value contours become more distant from the ellipse they become longer. This can be compensated by weighting points in $\mathcal{R}^2$ by the length of the iso-value contour through the point. Since determining the length is not straightforward we take an alternative approach which is to sample a set of iso-value contours at regular intervals of the distance approximation. The iso-value contour of the distance approximation at some value[4] E_i is at $d_{\mathbf{X}}|e_{\mathbf{X}} = E_i$, and the mean Euclidean distance along each contour is

$$\mu_i = \mathrm{E}[d_{\mathbf{X}}|e_{\mathbf{X}} = E_i]. \tag{12}$$

The correlation coefficient is then calculated between the iso-value values and their Euclidean distance means

$$L = \frac{\sum_i w_i(E_i - \bar{E}_i)(\mu_i - \bar{\mu}_i)}{\sqrt{\sum_i w_i(E_i - \bar{E}_i)^2 \sum_i w_i(\mu_i - \bar{\mu}_i)^2}} \tag{13}$$

Thus each distance is now given equal weight in the assessment measure. In addition we have included the term w_i for each contour corresponding to the weighting factor associated with the noise model, where $w_i = w(\bar{\mu}_i)$.

---

[4] To avoid confusion we assume distances inside and outside the ellipse are signed, or can be made so.

3.2   Curvature Bias

To measure the departure of the iso-value contour from the desired constant Euclidean distance we use the variance of the underlying Euclidean distance

$$\sigma_i^2 = \text{Var}[d_{\mathbf{X}}|e_{\mathbf{X}} = E_i]. \tag{14}$$

Since this is a local measure it must be combined over contours to give a global measure

$$C = \sum_i w_i \sigma_i^2. \tag{15}$$

The ideal error of fit should have no curvature bias which is obtained when $C = 0$.

3.3   Asymmetry

Assuming signed distances, the mean Euclidean distances along corresponding iso-value contours inside and outside the ellipse are calculated

$$\begin{aligned} \mu_i^+ &= \text{E}[d_{\mathbf{X}}|e_{\mathbf{X}} = E_i] \\ \mu_i^- &= \text{E}[d_{\mathbf{X}}|e_{\mathbf{X}} = -E_i]. \end{aligned}$$

Asymmetry is calculated at each contour pair as the normalised difference in their mean Euclidean distances.

$$a_i = \frac{|\mu_i^+ - \mu_i^-|}{\mu_i^+ + \mu_i^-}.$$

Again, a weighted sum of the local measures over the contours is made to produce a global measure

$$A = \sum_i w_i a_i \tag{16}$$

which would equal zero for the ideal error of fit.

3.4   Overall Goodness

Rather than individually assess the three specific characteristics described above sometimes it may be more convenient to produce a single overall assessment of the distance approximations, ignoring the details. One approach would to produce a weighted sum of the three measures above. Instead we use an alternative which avoids the need for deciding on suitable weights and use the squared difference between the approximation and the true Euclidean distance

$$G = \sum_{\mathbf{X}} w_{\mathbf{X}}(d_{\mathbf{X}} - Se_{\mathbf{X}})^2. \tag{17}$$

Since we can ignore uniform scalings of the distances we allow the approximation to be scaled by some constant S, and choose S so as to minimise G. This is found when $\frac{\delta G}{\delta S} = 0$, yielding

$$S = \frac{\sum_{\mathbf{X}} w_{\mathbf{X}} e_{\mathbf{X}} d_{\mathbf{X}}}{\sum_{\mathbf{X}} w_{\mathbf{X}} e_{\mathbf{X}}^2}.$$

4 Results

We can now apply the assessment measures described in the previous section to the EOF functions described in section 2. A quadrant of a single ellipse with semi-major axis $a = 400$ and semi-minor axis $b = 100$ and centred at the origin was used. The Euclidean distance was found by plotting the ellipse into an image followed by performing a Euclidean distance transform. The distance was made signed by setting negative all non-zero distance values 4-way connected to the origin. The EOFs were generated and sampled at unit increments in the quadrant 1000×300. Iso-value contours crossing the Y axis at 10 pixel intervals were detected, and the average and variance of the Euclidean distances (12) and (14) along the contours were calculated.

To simplify calculation of the measures and ensure uniform sampling, for both the average and variance of the Euclidean distance values along the contours the EOF values were resampled at unit intervals using linear interpolation. It is then straightforward to find the corresponding points μ_i^+ and μ_i^- .

First we apply the assessment measures to EOF_9 to determine a suitable value for p. The effects of p on the measures is graphed in figure 3. Increasing p initially has little effect on the linearity before it degrades from $p \approx 0.4$. There is a roughly linear relationship between p and a reduction in curvature bias. Asymmetry degrades with increasing p in a non-linear fashion. Combining all the factors in the single measure of goodness G, the optimal value is found at $p \approx 0.45$. This is the value of p we have used when comparing the various techniques. If the severe asymmetry is ignored by using the G' measure then the optimal value is $p \approx 1$ – i.e. EOF_9 reduces to EOF_2.

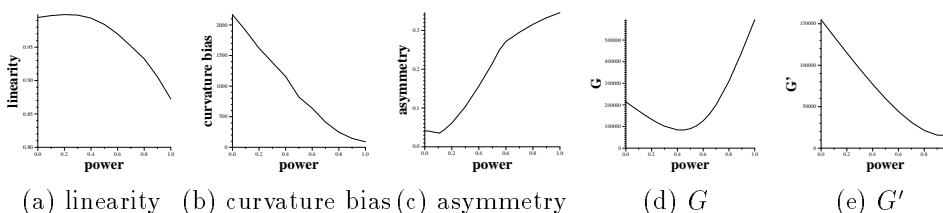(a) linearity (b) curvature bias (c) asymmetry   (d) G (e) G'

Figure 3: Effect of p on assessment measures

Table 1 gives the assessment values of the EOFs (with the exception of the incomplete EOF_7) for the Gaussian noise model with $\sigma = 64$. To facilitate comparison all values have been scaled wrt the corresponding EOF_1 value. Recall that we wish to maximise L and minimise the remaining assessment measures. Clearly, EOF_{13} is the best EOF since it has the best rating in all categories. With the exception of EOF_2 and EOF_{10} the EOFs have reasonable linearity. All the EOFs provide a lower curvature bias than the algebraic distance. However, despite its wavy iso-value contours, EOF_2 still has a lower curvature bias than EOF_8 – EOF_{11}. Apart from EOF_{13} the EOFs have rather poor asymmetry; as expected EOF_{10} is comparable with EOF_2, while EOF_{11} is worse yet. Overall the EOFs with the exception of EOF_2 and EOF_{10} gain better ratings by G than EOF_1. If asymmetry is factored out using G' then all the EOFs receive better scores.

Table 1: Normalised assessment results with $N(0, 64)$ noise model

EOF	L	C	A	G	G'
1	1.000	1.000	1.000	1.000	1.000
2	0.877	0.041	8.404	2.771	0.099
8	0.995	1.000	4.945	0.526	0.666
9	0.995	0.452	4.500	0.391	0.441
10	0.898	0.069	7.092	3.134	0.362
11	1.003	0.230	12.507	0.617	0.273
12b	0.990	0.035	3.508	0.425	0.163
13	1.006	0.002	0.747	0.007	0.009

References

[1] G.J. Agin. Fitting ellipses and general second-order curves. Technical Report CMU-RI-TR-81-5, The Robotics Institute, Carnegie-Mellon University, Pittsburgh, USA, 1981.

[2] A. Albano. Representation of digitised contours in terms of conic arcs and straight line segments. *CGIP*, 5:23–33, 1974.

[3] A.W. Fitzgibbon and R.B. Fisher. A buyer's guide to conic fitting. In *British Machine Vision Conf.*, pages 513–522, 1995.

[4] A.D. Gross and T.E. Boult. Error of fit for recovering parametric solids. In *ICCV*, pages 690–694, 1988.

[5] Y. Nakagawa and A. Rosenfeld. A note on polygonal and elliptical approximation of mechanical parts. *PR*, 11:133–142, 1979.

[6] V.S. Nalwa and E. Pauchon. Edgel aggregation and edge description. *CVGIP*, 40:79–94, 1987.

[7] T. Pavlidis. Curve fitting with conic splines. *ACM Trans. on Graphics*, 2(1):1–31, 1983.

[8] P.L. Rosin. A note on the least squares fitting of ellipses. *PRL*, 14:799–808, 1993.

[9] P.L. Rosin. Assessing error of fit functions for ellipses. Technical Report CSTR-95-13, Department of Computer Science & Information Systems, Brunel University, 1995.

[10] R. Safaee-Rad, I. Tchoukanov, B. Benhabib, and K.C. Smith. Accurate parameter estimation of quadratic curves from grey level images. *CVGIP: IU*, 54:259–274, 1991.

[11] P.D. Sampson. Fitting conic sections to very scattered data. An iterative refinement to the Bookstein algorithm. *CVGIP*, 18:97–108, 1982.

[12] M. Stricker. A new approach for robust ellipse fitting. In *Int. Conf. Automation, Robotics, and Computer Vision*, pages 940–945, 1994.

[13] G. Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Trans. PAMI*, 13(11):1115–1138, 1991.