Pros and Cons of Euclidean Fitting*

Petko Faber and Bob Fisher
Division of Informatics, University of Edinburgh,
Edinburgh, EH1 2QL, UK
npf|rbf@dai.ed.ac.uk

Abstract

The purpose of this paper is to discuss pros and cons of fitting general curves and surfaces to 2D and 3D edge and range data using the Euclidean distance. In the past researchers have used approximate distance functions rather than the Euclidean distance. But the main disadvantage of the Euclidean fitting, computational cost, has become less important due to rising computing speed. Experiments with the real Euclidean distance show the limitations of suggested approximations like the Algebraic distance or Taubin's approximation. We compare the performance of various fitting algorithms in terms of efficiency, correctness, robustness and pose invariance.

1 Introduction

The ability to construct CAD or other object models from edge and range data has a fundamental meaning in building a recognition and positioning system. While the problem of model fitting has been successfully addressed, the problem of a high accuracy and stability of the fitting is still an open problem. On the one hand it is imperative to solve the problem of how curves and surfaces can be fitted to a given data set. But on the other hand it is obvious that accuracy and stability of the fitting has a substantial impact on the recognition performance especially in reverse engineering where we desire an accurate reconstruction of 3D geometric models of objects from range data. Thus it is very important to get good shape estimates from the data.

Implicit polynomial curves and surfaces are potentially among the most useful object or data representations for use in computer vision and image analysis. Their power appears by their ability to smooth noisy data, to interpolate through sparse or missing data, their compactness and their form being commonly used in numerous constructions. An implicit curve or surface is the zero set of a smooth function $f : \mathbb{R}^n \to \mathbb{R}^m$ of the n variables: $\mathcal{Z}(f) = \{\vec{x} : f(\vec{x}) = 0\}$. Let $f(\vec{x})$ be an *implicit polynomial* of degree d given by

$$f(\vec{x}) = \sum_{\substack{(i+j+k)\leq d \\ \{i,j,k\}\geq 0}} a_{ijk} \cdot x^i \cdot y^j \cdot z^k = 0 \ . \tag{1}$$

Then, we only have to determine the parameter set $\{a_{ijk}\}$ that describes the given data best.

2 Least squares fitting of general curves and surfaces

Parameter estimation, usually cast as an optimization problem, can be divided into three general techniques: least-squares fitting (e.g. [1, 3, 9, 10, 12]), Kalman filtering (e.g. [4, 5]), and robust techniques (e.g. [2, 6]). Given a finite set of data points $\mathcal{D} = \{\vec{x}_p\}$, $p \in [1, P]$, the problem of fitting a general curve and surface $\mathcal{Z}(f)$ to $\mathcal{D}$ by a least-squares method is to minimize a distance measure

$$\frac{1}{P}\sum_{p=1}^{P} \mathrm{dist}\left(\vec{x}_p, \mathcal{Z}(f)\right) \to \text{ Minimum} \tag{2}$$

from the data points $\vec{x}_p$ to the curve or surface $\mathcal{Z}(f)$, a function of the parameter set $\{a_{ijk}\}$. The distance from the point $\vec{x}_p$ to the zero set $\mathcal{Z}(f)$ is defined as the minimum of the distances from $\vec{x}_p$ to points $\vec{x}_t \in \mathcal{Z}(f)$:

$$\mathrm{dist}(\vec{x}_p, \mathcal{Z}(f)) = \min\left\{\| \vec{x}_p - \vec{x}_t \| : f(\vec{x}_t) = 0\right\} \ . \tag{3}$$

*The work was funded by the CAMERA (*CA*d *M*odelling of Built *E*nvironments from *R*ange *A*nalysis) project, an EC TMR network (ERB FMRX-CT97-0127).

In the past researchers have often replaced the real Euclidean distance by an approximation. But it is well known that a different performance function can produce biased results, and for a lot of primitive curves and surfaces a closed form expression exists for the Euclidean distance. In the following we summarize the Algebraic fitting, Taubin's fitting [13, 14] and an Euclidean fitting [7, 8].

Algebraic fitting ($\boldsymbol{AF}$) is based on the approximation of the Euclidean distance by the algebraic distance $\text{dist}_A(\vec{x}_p, \mathcal{Z}(f)) = f(\vec{x}_p)$. Given the Algebraic distance for each point, Eq.(2) can be formulated as an Eigenvector problem. To avoid the trivial solution $\{a_{ijk}\} = \vec{0}$ and any multiple of a solution, the parameter set $\{a_{ijk}\}$ may be constrained in some way. The *pros and cons* of using algebraic distances are the gain in computational efficiency, because closed form solutions can usually be obtained, but often the results are unsatisfactory.

Taubin's fitting ($\boldsymbol{TF}$) uses the first order approximation of Eq. (1) to estimate $\text{dist}_T(\vec{x}_p, \mathcal{Z}(f))$ [13].

$$\text{dist}_T(\vec{x}_p, \mathcal{Z}(f)) = \frac{|f(\vec{x}_p)|}{\|\nabla f(\vec{x}_p)\|} \tag{4}$$

The *pros and cons* of using Taubin's distance are no iterative procedures are required and it is a first order approximation to the exact distance, but the approximate distance is also biased in some sense. If, for instance, a data point $\vec{x}_p$ is close to a critical point of the polynomial, i.e., $\|\nabla f(\vec{x}_p)\| \approx 0$, but $f(\vec{x}_p) \neq 0$, the distance becomes large which is certainly a limitation. To minimize Eq.(2) the usage of the Levenberg-Marquardt (*LM*) algorithm is proposed.

Euclidean fitting ($\boldsymbol{EF}$) replaces the approximated distances again by the Euclidean distance, which is invariant to transformations in Euclidean space and not biased. For primitive curves and surfaces like straight lines, ellipses [15, 16], planes, cylinders, cones, and ellipsoids, a closed form expression exists for the Euclidean distance from a point to the zero set and we use these. However, as the general expression of the Euclidean distance is more complicated and there exists no known closed form expression, an iterative optimization procedure must be carried out. Given the Euclidean distance $\text{dist}_E(\vec{x}_p, \mathcal{Z}(f))$ for each point the following simple algorithm can be used:

1. The Euclidean fitting requires an initial estimate for the parameters $\{a_{ijk}\}$ and we have found that the result of Taubin's fitting method is more suitable than others. We get the initial parameter set $\{a_{ijk}\}^{[0]}$.
2. In the second step $\{a_{ijk}\}^{[s]}$,$s = 0, 1, \ldots$, is updated using the *LM* algorithm, minimizing the sum of Euclidean distances for all data points.
3. Finally, each $\{a_{ijk}\}^{[s+1]}$ is evaluated by a *M-estimator* $\mathcal{L}$ on the basis of $\text{dist}_E(\vec{x}_p, \mathcal{Z}(f))$. If $\mathcal{L}(\{a_{ijk}\}^{[s+1]}) < \mathcal{L}(\{a_{ijk}\}^{[s]})$, $\{a_{ijk}\}^{[s+1]}$ is accepted and the fitting will be continued with step 2. Otherwise the fitting is terminated and $\{a_{ijk}\}^{[s]}$ is the desired solution.

3 Evaluating the Euclidean fitting

To work out the pros and cons of Euclidean fitting we compare the performance of the *EF* method with the performance of *AF* and *TF* in terms of efficiency, correctness and robustness for both simulated and real data. In case of simulated data we have generated data sets which describe (elliptical) cylinders and cones. The 3D data were generated by adding isotropic Gaussian noise $\sigma = \{1\%, 5\%, 10\%, 20\%\}$. Additionally the surfaces were partially occluded. The visible surfaces were varied between 1/2 (maximal case) and 1/6 of the full 3D cylinder. To show that the *EF* works even for real data we have used several range data sets. For all experiments we include in all three fitting methods the same constraints which describe the expected surface type to enforce the fitting of a special surface type. Finally, we look to the pose invariance of the fitting methods.

3.1 Efficiency

A good fitting algorithm has to be efficient as possible in terms of run time and formal complexity. While the problem of computational cost is no longer a really hard problem because of the rapidly increasing machine speed, we should guarantee the fitting with acceptable computational cost as well as the algorithm with relatively low complexity. All algorithms have been implemented in C and the computation was performed on a Pentium III 466 MHz. The average computational costs for the *AF*, *TF* and *EF* are in Tab. 1. As expected the *AF* and *TF* supply the best performance. The *EF* algorithm

Table 1: Average computational costs in milliseconds per 1000 points.

	AF	*TF*	*EF*
Plane	0.958	1.042	2.417
Sphere	1.208	1.250	3.208
Circular cylinder	3.583	3.625	12.375
Elliptical cylinder	13.292	13.958	241.667
Circular cone	15.667	15.833	288.375
Elliptical cone	15.042	15.375	291.958
General quadric	18.208	18.458	351.083

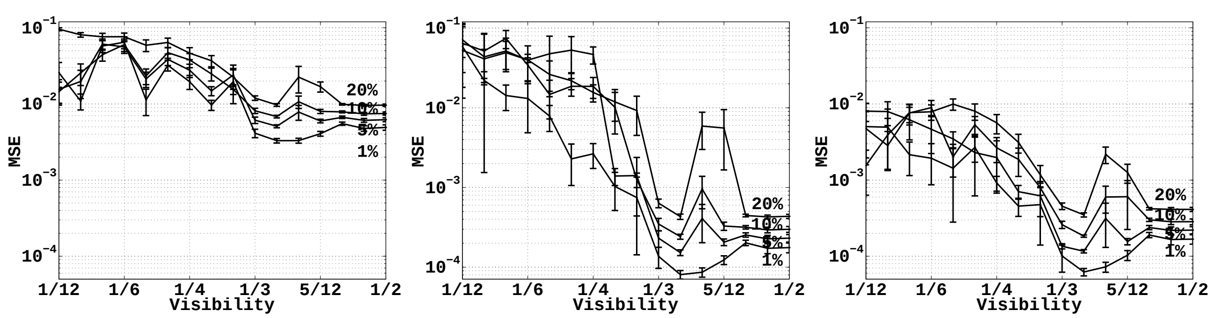
requires a repeated search for the point x_t closest to x_p and the calculation of the Euclidean distance. A quick review of the values in Tab.1 shows that the computational costs increase if we fit an elliptical cylinder, a circular or an elliptical cone respectively a general quadric, because the distance estimation is more complicated. In summary the efficiency is a *con* of *EF*, but is bounded by a factor of about 20 times the performance of the other algorithms and is still computationally reasonable for up to 10^6 data points if real-time performance is not needed.

3.2 Correctness

It is obvious that the fitting result should describe the data set by the correct curve or surface type. That means that it should not fit a false type to the data. To verify the correctness we tested if the fitting result of the (constrained) eigenvalue analysis corresponds to the general curve or surface invariants. If one solution satisfies the conditions for one curve or surface type, it is assumed that the fitting is correct in sense of an interpretable real curve or surface. Otherwise, the fitting will be defined as failure. In our experiments *AF* failed sometimes (up to 23 percent) respecting our expectations, especially with higher noise levels or a sparse data set (see Sec. 3.3). For *TF* and *EF* we had no failures in our experiments. In summary the correctness is a *pro* of *EF*.

3.3 Robustness

A fitting method must degrade gracefully with increasing noise in the data, with a decrease in the available relevant data, and with an increase in the irrelevant data. To evaluate the robustness of the proposed *EF*, we use synthetic generated data describing an elliptical cylinder by adding isotropic Gaussian noise $\sigma = \{1\%, 5\%, 10\%, 20\%\}$ and partially occlusion varied between 1/2 (maximal case) and 1/6 of the full 3D cylinder. In the first experiment the number of 3D points for the simulated cylinder was $n = 100$ and to measure the average fitting error each experiment runs 100 times. The reported error is the Euclidean geometric distance between the 3D data points and the estimated surfaces. The mean squares errors (*MSE*'s) and standard deviations of the different fittings are in Fig. 1. As expected, *TF* and *EF* yield

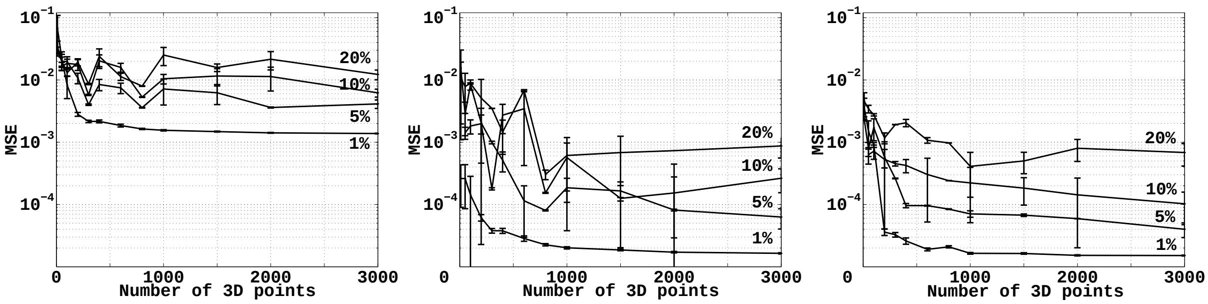


a. Algebraic fitting. b. Taubin's fitting. c. Euclidean fitting.

Figure 1: Average least squares error fitting a synthetic generated cylinder with added Gaussian noise $\sigma = \{1\%, 5\%, 10\%, 20\%\}$. The visible surfaces were varied between 1/2 (maximal case) and 1/6 of the full 3D cylinder. The number of trials was 500.

the best results respect with to the mean and standard deviation, and the mean for *EF* is always lower than for the other two algorithms. The results of *AF* are only partially acceptable, because of the mean and the standard deviation. In the direct comparison of *TF* with *EF* the results of *EF* are much better and the mean of *EF* is always lower than the mean of the other two algorithms. As mentioned in Sec.3.2, *AF* can give sometimes wrong results which means that the fitted curve or surface types does not come up with our expectations. We removed all failed fittings out of the considerations.

In the second experiment, the number of 3D points was stepwise decreased from $n = 1000$ down to $n = 10$ 3D data points to evaluate the behaviour of the several fitting methods. Each experiment runs 100 times. The mean squares errors (*MSE*'s) and standard deviations of the different fittings are in Fig. 2. As expected, *TF* and *EF* yield also the best results in this experiment. With decreased point



a. Algebraic fitting. b. Taubin's fitting. c. Euclidean fitting.

Figure 2: Average least squares error fitting a synthetic generated cylinder with added Gaussian noise $\sigma = \{1\%, 5\%, 10\%, 20\%\}$. The number of 3D points was stepwise decreased from 1000 up to 10. The visible surfaces was 5/12 of the full 3D cylinder. The number of trials was 500.

density especially the *AF* becomes more and more unstable which is reflected in the mean and standard deviation. Unexpectedly, the *EF* is very stable even with only $n = 10$ 3D data points. This underlines once more the outstanding performance of the *EF*. In summary the robustness is clearly a *pro* of *EF*.

3.4 Pose invariance

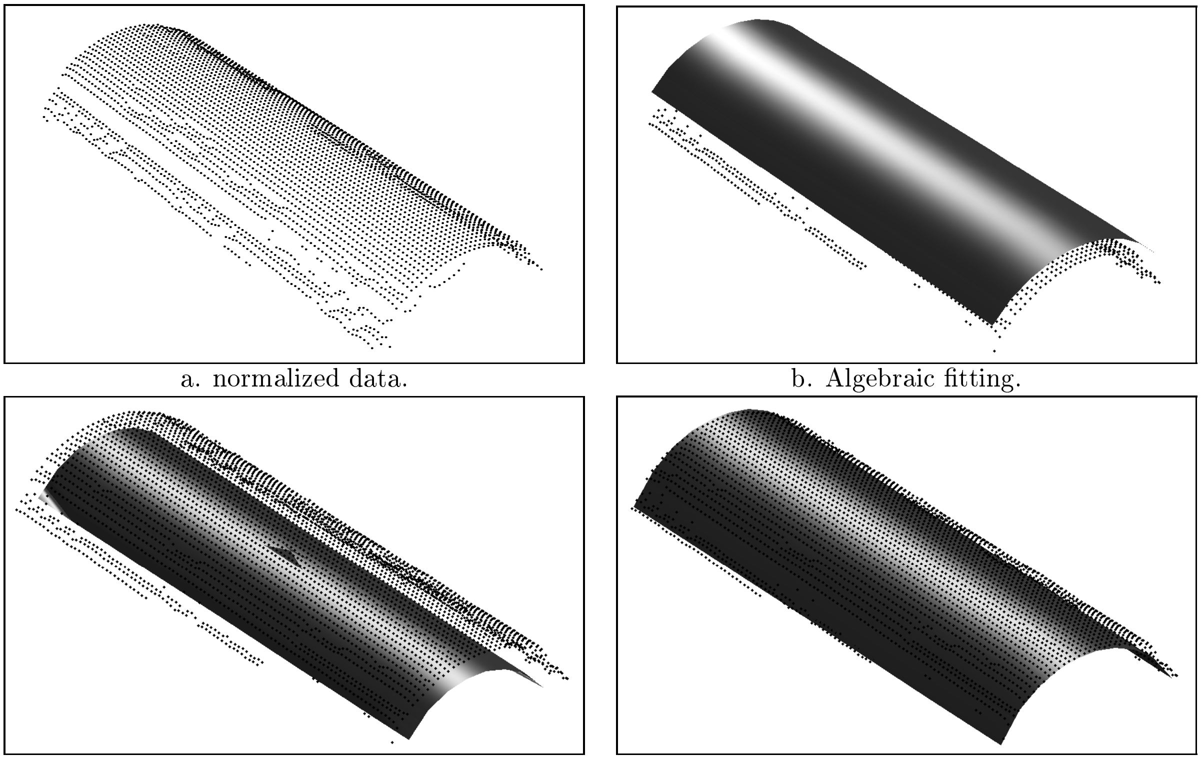
It is obvious that the fitting results should be pose invariant. But, it is well known that this reasonable and necessary requirement cannot be always guaranteed by all three fitting methods. To evaluate the pose invariance we use a real data set (see Fig. 3a.) describing an elliptical cylinder. The normalized data set was a) shifted, b) rotated, and c) both rotated and shifted. A quick review of the residuals (*MSE*) in Tab. 2 shows *AF* and *TF* are not pose invariant while the *EF* is pose invariant. To illustrate the pose dependency, the fitting results for position 3 are visualized in Fig. 3. In summary the pose invariance is clearly a *pro* of *EF*.

Table 2: Residuals fitting an elliptical cylinder (see Fig. 3). The normalized cylinder was shifted by $t = [0.3, 0.2, 0.1]$ (pos. 1), rotated by $\vartheta = \pi/12$ and $n = [0.5, 1.0, 0.5]$ (pos. 2), shifted and rotated (pos. 3).

		normal pos	position 1	position 2	position 3
AF	$[10^{-3}]$	0.5242	2.0181	2.6950	1.8271
TF	$[10^{-3}]$	0.5024	1.5143	2.0277	1.3817
EF	$[10^{-3}]$	0.4021	0.4152	0.8634	0.6088

4 Conclusion

The focus was on the pros and cons of Euclidean fitting compared with the commonly used Algebraic fitting and Taubin's fitting. Referring to our objective we can finally conclude that we have more pros than cons for the Euclidean fitting. While the main disadvantage of the Euclidean fitting, computational cost, has become less important due to rising computing speed, robustness and accuracy increases sufficiently compared to both other methods. Additionally, the Euclidean fitting is pose invariant.



a. normalized data. b. Algebraic fitting.

c. Taubin's fitting. d. Euclidean fitting.

Figure 3: Fitting results for a real range data (≈ 3300 points). The normalized data set was shifted by $t = [0.3, 0.2, 0.1]$ and rotated by $\vartheta = \pi/12$ and $n = [0.5, 1.0, 0.5]$.

References

[1] Allan, F. E. The general form of the orthogonal polynomial for simple series with proofs of their simple properties. In *Proc. Royal Soc. Edinburgh*, pp. 310–320, 1935.

[2] Besl, P. J. and R. C. Jain. Three-dimensional object recognition. *Computing Survey*, 17(1):75–145, 1985.

[3] Bookstein, F. L. Fitting conic sections to scattered data. *CGIP*, 9:56–71, 1979.

[4] Chui, C. K. and G. Chen. *Kalman filtering with real time applications.* Springer, Berlin-Heidelberg-New York, 1987.

[5] Dickmanns, E. D. and V. Graefe. *Dynamic monocular machine vision. MVA*, 1:223–240, 1988.

[6] Duda, R. O. and P. E. Hart. The use of Hough transform to detect lines and curves in pictures. *Comm. Assoc. Comp. Machine*, 15:11–15, 1972.

[7] Faber, P. and R. B. Fisher. Euclidean fitting revisited. In *4th IWVF*, pp. 165–175, 2001.

[8] Faber, P. and R. B. Fisher. Estimation of General Curves and Surfaces to Edge and Range Data by Euclidean Fitting. *submitted to the MVA.*

[9] Fitzgibbon, A. W. and R. B. Fisher. A buyer's guide to conic fitting. In *6th BMVC*, pp. 513–522, 1995.

[10] Kanatani, K. Renormalization for biased estimation. In *4th ICCV*, pp. 599–606, 1993.

[11] Ray, W. J. J. *Introduction to Robust and Quasi-Robust Statistical Methods.* Springer, Berlin-Heidelberg-New York, 1983.

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

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

[14] Taubin, G. An improved algorithm for algebraic curve and surface fitting. In *4th ICCV*, pp. 658–665, 1993.

[15] Voss, K. and H. Süße. *Adaptive Modelle und Invarianten für zweidimensionale Bilder*. Shaker, Aachen, 1995.

[16] Zhang, Z. Parameter estimation techniques: a tutorial with application to conic fitting. *IVC*, 15:59–76, 1997.