Optimal parameter selection for derivative estimation from range images

A. J. Stoddart, J. Illingworth, T. Windeatt
Dept. of Electronic and Electrical Engineering
University of Surrey, Guildford, GU2 5XH, U.K.
email: a.stoddart@ee.surrey.ac.uk

Abstract

Range images may be used for a variety of applications in object recognition, inspection and reverse engineering. In many of these applications it is important to obtain good estimates of the local surface curvature. Good curvature estimates require good derivative estimates, but the estimation of derivatives from sampled data is highly susceptible to noise. In this paper we introduce a new way of characterizing range data by a single parameter. From this characterization we show how to make an optimal choice of whatever parameters there are in a particular derivative estimation method, and obtain an estimate of the error one might expect. Finally we demonstrate how this analysis can be applied to measuring the curvature of a cylinder.

1 Introduction

Range images may be used for a variety of applications in object recognition, inspection and reverse engineering. In many of these applications it is important to obtain good estimates of the local differential properties. Properties such as normals or slopes depend on first derivatives, and various curvature properties such as the principal curvatures, mean and Gaussian curvature all depend on both first and second derivatives.

Derivative estimates [1, 2, 3] from noisy sampled data are difficult to compute for two main reasons. Firstly, the data may include discontinuities in value (C^0 discontinuities), slope (C^1 discontinuities), or higher derivatives. Secondly, because derivatives are computed by local differencing operations they are sensitive to noise. To solve the second problem it is common to smooth the data, but this runs the risk of concealing small discontinuities, and smoothing out features that we wish to observe.

In this paper we address the problem of estimating derivatives from noisy data obtained from a smooth surface. In particular we consider the problem of estimating derivatives from a dense range image of height measurements on a regular grid. In this paper we consider only one derivative estimation method, the weighted facet model. However our analysis applies to any filter based derivative estimation method.

We will ignore the problem of discontinuity detection, and assume that our image contains no discontinuities. Even if there are discontinuities present in a range image, a method of obtaining good derivative estimates in interior regions

BMVC 1994 doi:10.5244/C.8.16

is still required. Once the discontinuities have been detected the method can be applied to all the data using a method such as normalized convolution [4].

Almost all methods of derivative estimation involve one or more parameters. These parameters relate to filter sizes or the smoothing window sizes and possibly the order of the method, e.g. order of polynomial fit. (We consider smoothing to be part of derivative estimation.) It is possible to spend considerable time in choosing an appropriate parameter, typically by trial and error. If the sample or sensor characteristics change, then the previous choices of parameters may not be relevant.

In this paper we introduce a new way of characterizing range data by a single parameter which we call the *variation length*. This enables comparison between data sets obtained under widely varying conditions. From this characterization we show how to make an optimal choice of whatever parameters there are in a particular derivative estimation method, and obtain an estimate of the error one might expect. We do not directly address the question of curvature, but simply note that it is usually computed in a standard way from derivative data [3]. Finally we look at a common example, the cylinder, and apply the analysis.

What do we mean by an optimal parameter choice? Consider the two sources of error in our estimate of a derivative. The measurements will be corrupted by noise, and as a result the estimate will be corrupted by a measurement error. There will also be some systematic error in our estimate which is intrinsic to the method.

Suppose the parameter we must choose is a length (neighborhood) over which we must smooth. For zero smoothing our estimate will be very sensitive to measurement noise, but as we increase the smoothing length we will reduce the measurement noise and obtain a better estimate. However the data will have some significant scale of variation. When we smooth too much the systematic error will grow, for example we will start to smooth out real features. We must choose the smoothing length to balance out these two competing effects. It will then be an optimal choice, and the estimate will be the best possible estimate given the data and that particular method.

Before continuing we discuss briefly some previous work. Flynn and Jain [1] have conducted an empirical study of five different curvature estimation techniques. More recently Lee and Haralick [2] made a detailed performance characterization of the estimation of curvature from noisy sampled data. They study planar curve fitting to a ordered sequence of points, and provide a framework for performance evaluation. We base our strategy for optimal parameter choice on this kind of performance characterization.

2 Derivative Estimation

We assume that the function we wish to measure is $f(x)$. The sensor operation is modelled as sampling the function at a regularly spaced set of points $x_i = i\Delta$. Each measurement is assumed to be corrupted by Gaussian noise n_i of variance σ, so the data obtained from the sensor is $f(x_i) + n_i$. An estimate of $f(0)$ obtained from noise free data will be denoted $\hat{f}(0)$ and from noisy data the estimate will be denoted $\hat{f}_\sigma(0)$

In order to estimate $f(x)$ and its derivatives at $x = 0$ we make a least squares polynomial fit of order k in a neighborhood of $x = 0$. This is called the facet model. To specify the size of the neighborhood we introduce a Gaussian weighting function $w(x) = \exp\left\{-x^2/\alpha^2\right\}$. We call the parameter α the 'smoothing length'. The polynomial fit is obtained by minimizing the weighted least mean squares cost C given by

$$C = \sum_i w(x_i)\left[f(x_i) - \sum_{p=0}^{k} a_p x_i^p\right]^2 \tag{1}$$

with respect to the polynomial coefficients a_p. The polynomial fit may now be used to compute an estimate of $f(0)$ or higher derivatives. This kind of least squares problem is easily solved analytically using orthonomal polynomials [3], [for Gaussian weighting use discrete Hermite polynomials]. The process of obtaining these estimates may be reduced to a simple filter operation where we denote the filter coefficients by c_i, d_i and e_i. For convenience a factor of Δ is included to make the filter coefficients dimensionless,

$$\hat{f}(0) = \sum_i c_i f(x_i), \qquad \hat{f}'(0) = \sum_i \frac{d_i}{\Delta} f(x_i), \qquad \hat{f}''(0) = \sum_i \frac{e_i}{\Delta^2} f(x_i) \tag{2}$$

The filters all fall off sharply at about $|x| \sim \alpha$ due to the exponential weighting term. This means that smoothing length α is a measure of the size of the neighborhood over which we will sample to obtain our estimate.

Both α and Δ have dimensions of (horizontal) length and we may define a dimensionless measure of neighborhood size $\alpha_D = \alpha/\Delta$. Note that the coefficients c_i, d_i and e_i depend only on α_D. α_D is a measure of how many samples to either side of $x = 0$ are considered when making the polynomial fit ($w(x_i) = \exp\left\{-i^2/\alpha_D^2\right\}$).

3 Error Estimates

The estimate for $f(0)$ in the presence of noise is

$$\hat{f}_\sigma(0) = \sum_i c_i[f(x_i) + n_i] \tag{3}$$

We denote the expected value by $\langle\rangle$. We define the *measurement* error σ_n, as

$$\sigma_n^2 = \left\langle \left[\hat{f}_\sigma(0) - \hat{f}(0)\right]^2 \right\rangle = \sigma^2 \sum c_i^2 \tag{4}$$

The measurement error is not the only reason why our estimate will differ from the true value. In the absence of noise there is usually a systematic error due to the fact that values of $f(x_i)$ other than $f(0)$ are used in the estimate. For example if the function is smoothed then peaks will be flattened and troughs filled in. This degrades the estimate, but is not due to noise, it is a systematic effect of the method we choose. We call this systematic error the bias B, and it is given in terms of the coefficients c_i by

$$B = \hat{f}(0) - f(0) = \left[\sum_i c_i f(x_i)\right] - f(0) \tag{5}$$

Clearly the two sources of error behave differently. More smoothing reduces the random noise, but increases the bias. If we can estimate the respective errors caused then we can choose an amount of smoothing that is a good compromise between the two effects. For this we need an estimate of the bias, which we can obtain by using a Taylor expansion of the function f(x) to estimate $\hat{f}(0)$,

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f'''(0) + \cdots \tag{6}$$

The filter must obviously satisfy $\sum_i c_i = 1$ and if it is symmetric about $i = 0$ it will also satisfy $\sum_i ic_i = 0$. We cannot in general assume anything about the higher moments of the filter, although sooner or later some moment, say the nth will be non-zero, $\sum i^n c_i \neq 0$. Substituting equation (6) into equation (5) and truncating the Taylor series at n we may approximate the bias by,

$$B \simeq \left[\sum_i i^n c_i\right] \frac{\Delta^n}{n!} f^{(n)}(0) \tag{7}$$

It will be useful to define a quantity χ_n which has units of (horizontal) length. It is defined in terms of the noise variance and the nth derivative,

$$(\chi_n)^n = \frac{\sigma}{f^{(n)}(0)/n!} \tag{8}$$

The Taylor series can now be recast in an illuminating way,

$$f(x) = f(0) + \sigma\left\{\left(\frac{x}{\chi_1}\right) + \left(\frac{x}{\chi_2}\right)^2 + \left(\frac{x}{\chi_3}\right)^3 + \cdots\right\} \tag{9}$$

What is the significance of χ_n? For example χ_1 is the horizontal distance that you must travel such that σ is equal to the change in height of the function due to the first derivative, i.e. $f'(0)\chi_1 = \sigma$. For $n > 1$ χ_n is a natural extension of that concept. It is a measure of the distance you must travel before the nth derivative starts to influence the height by an amount equal to σ. It is the significant scale of horizontal variation for a given problem. We call it the *variation length*. Normally the order n is obvious from the context so we will not usually show the subscript n of χ_n explicitly.

As in the case of α it proves useful to consider a dimensionless version of χ, namely $\chi_D = \chi/\Delta$. We note that the larger χ_D is, the better the signal has been sampled. χ_D is the ratio of two horizontal lengths, the variation length and the sampling interval. It is a dimensionless measure of sampling 'quality', and we call it the *sampling ratio*.

The bias can be written in terms of the sampling ratio χ_D and the noise σ as

$$B \simeq \left[\sum_i i^n c_i\right] \frac{\sigma}{\chi_D^n} \tag{10}$$

It is now time to consider the choice of the parameter α_D, the dimensionless smoothing length or neighborhood size. Intuitively we expect a longer smoothing length will reduce the measurement error but too much smoothing will smooth out genuine features and increase the bias B. For typical filters c_i this is indeed the case. [See figure 1.] These effects both degrade the estimate but are different in nature. The measurement error is a zero mean random effect, whereas the bias B is systematic and depends on the shape of the object. These effects may have different significance to the user. However, in order to proceed we will assume that the two errors are of equal significance, in other words we wish to minimize the error as measured by

$$\sigma_{TOT}^2 = \left\langle \left[\hat{f}_\sigma(0) - f(0)\right]^2 \right\rangle = \sigma_n^2 + B^2 \tag{11}$$

If we define the *normalized total error* as $\gamma = \sigma_{TOT}/\sigma$ then in our approximation γ is given in terms of the filter coefficients c_i by

$$\gamma^2 = \frac{\sigma_{TOT}^2}{\sigma^2} = \frac{\sigma_n^2}{\sigma^2} + \frac{B^2}{\sigma^2} \simeq \left[\sum_i c_i^2\right] + \frac{1}{\chi_D^{2n}} \left[\sum_i i^n c_i\right]^2 \tag{12}$$

We note that γ depends on only two parameters, namely the sampling ratio χ_D and, through the coefficients c_i, the dimensionless smoothing length α_D. This suggests a way of choosing the optimal parameter α_D. It must be chosen to minimize γ and is a function of χ_D. We also obtain an estimate of the lowest possible total error $\sigma_{TOT} = \gamma\sigma$ for that particular method.

In summary the procedure is as follows. Choose a method and find out the order n of the lowest non-zero moment of the filter coefficients c_i. Using $f^{(n)}(0)$ and σ compute the variation length χ. Divide this by the sampling interval Δ to get the sampling ratio χ_D. Then by minimizing the total error γ with respect to α_D the optimal smoothing length is obtained. An example of this is given in the results section.

Presumably the user would fix the variation length to its maximum value in an image. The error at a given position in the image will then always be bounded by the (lower) measurement error and the (upper) total error for that χ_D and α_D.

3.1 Estimating derivatives

We are not primarily interested in estimates of $f(0)$, our main concern is the first and second derivatives. The above treatment may easily be extended to these cases, For the first derivative the measurement error σ_n is given as

$$\sigma_n^2 = \left\langle \left[\hat{f}'_\sigma(0) - \hat{f}'(0)\right]^2 \right\rangle = \sigma^2 \sum_i \frac{d_i^2}{\Delta^2} \tag{13}$$

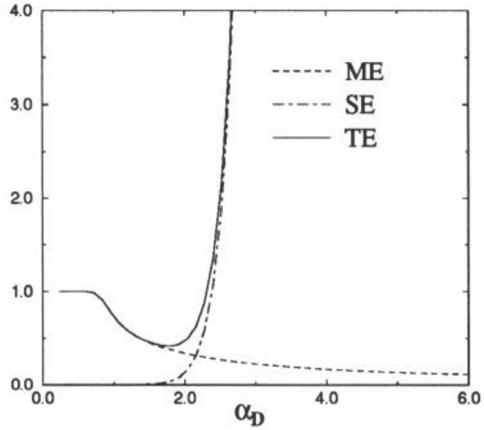

Figure 1: The zeroth derivative: Measurement (ME), systematic (SE) and total (TE) errors as a function of smoothing length for $\chi_D = 3$

The filter coefficients will satisfy $\sum_i d_i = 0$, $\sum_i id_i = 1$ and if the filter is antisymmetric also $\sum_i i^2 d_i = 0$. The bias B is defined as $B = \hat{f}'(0) - f'(0)$ and, assuming that the lowest non-zero moment is order n, it is approximated by

$$B \simeq \left[\sum_i i^n \frac{d_i}{\Delta}\right] \frac{\Delta^n}{n!} f^{(n)}(0) = \left[\sum_i i^n d_i\right] \frac{\sigma}{\Delta \chi_D^n} \tag{14}$$

The total error σ_{TOT} and normalized total error are

$$\sigma_{TOT}^2 = \left\langle \left[\hat{f}'_\sigma(0) - \hat{f}'(0)\right]^2 \right\rangle = \sigma_n^2 + B^2 \tag{15}$$

$$(\gamma')^2 = \left(\frac{\sigma_{TOT}}{\sigma/\chi}\right)^2 \simeq \chi_D^2 \left\{ \left[\sum_i d_i^2\right] + \frac{1}{\chi_D^{2n}} \left[\sum_i i^n d_i\right]^2 \right\} \tag{16}$$

All this generalizes quite straightforwardly to the second derivative, and we list the results below

$$\sigma_n^2 = \sigma^2 \sum_i \frac{e_i^2}{\Delta^4}, \quad \sum_i e_i = 0, \quad \sum_i ie_i = 0, \quad \sum_i i^2 e_i = 2 \tag{17}$$

$$B \simeq \left[\sum_i i^n e_i\right] \frac{\sigma}{\Delta^2 \chi_D^n} \tag{18}$$

$$\sigma_{TOT}^2 = \left\langle \left[\hat{f}''_\sigma(0) - f''(0)\right]^2 \right\rangle = \sigma_n^2 + B^2 \tag{19}$$

$$(\gamma'')^2 = \left[\frac{\sigma_{TOT}}{\sigma/\chi^2}\right]^2 \simeq \chi_D^4 \left\{ \left[\sum_i e_i^2\right] + \frac{1}{\chi_D^{2n}} \left[\sum_i i^n e_i\right]^2 \right\} \tag{20}$$

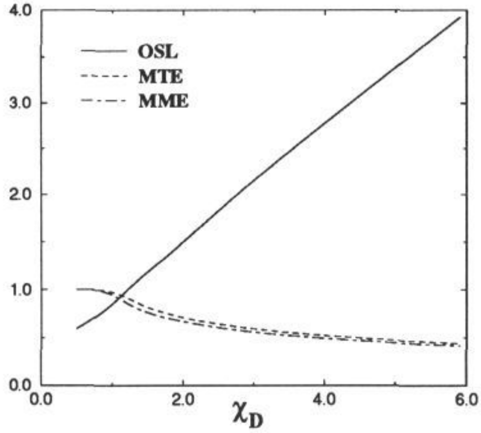

Figure 2: The zeroth derivative: Minimum total error (MTE), minimum measurement error (MME) and optimal smoothing length (OSL) as a function of sampling ratio

4 Results

We firstly consider as an example estimating $\hat{f}_\sigma(0)$ using a second order ($k = 2$) weighted facet model. For this case the lowest non-vanishing moment is order $n = 4$, so we need the order 4 variation length χ_4. In figure 1 we show the normalized measurement error σ_n^2/σ^2 and the normalized systematic error B^2/σ^2 as a function of the smoothing length α_D. In this figure we fixed the sampling ratio $\chi_D = 3$. For small smoothing length we see that $\sigma_n^2/\sigma^2 \to 1$ as would be expected. As we increase the smoothing length the measurement error decreases. However as the smoothing length increases beyond about 2 the bias picks up strongly. The normalized total error γ has a minimum at about $\alpha_D = 2$ where γ^2 is a little less than 0.5. This is how we choose the optimal smoothing length α_D. In figure 2 we plot the optimal smoothing length α_D and minimum normalized total error γ as a function of the variation length χ_D. The smoothing length is almost exactly proportional to χ_D. This is a very useful relation. It suggests that knowledge of the sampling ratio χ_D can very much simplify the choice of the optimal smoothing length α_D. *The linearity also confirms that the definitions of α_D and χ_D have been chosen in a physically meaningful way.* From the graph we see that to a good approximation we should always choose $\alpha_D = 0.6\chi_D$. We also see that as the sampling ratio improves (i.e. χ_D increases) the total error γ decreases and we obtain better estimates. However it doesn't fall off very fast and we don't really do much better than $\sigma_{TOT} = \gamma\sigma \sim 0.5\sigma$ for $\chi_D \sim 6$. In other words we can reduce the variance of the estimate by about half for moderate sampling ratio and a sensible choice of smoothing length.

In figure 2 we also show the measurement error $(\sigma_n/\sigma)^2$ for optimal smoothing length. It is possible that we may have overestimated the variation length and the

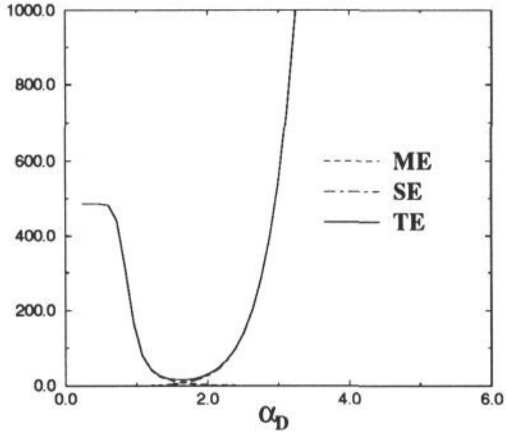

Figure 3: The second derivative: Measurement (ME), systematic (SE) and total (TE) errors as a function of smoothing length for $\chi_D = 3$

bias may be smaller than our approximation. However the total error γ cannot fall below the measurement error σ_n/σ once the choice of smoothing length has been made.

The next case we consider is obtaining an estimate of the second derivative (the first derivative follows similarly) $\hat{f}''_\sigma(0)$, again by the second order facet model. In figure 3 we again show the normalized measurement error $\frac{\sigma_n^2}{(\sigma/\chi^2)^2}$ and the normalized systematic error $\frac{B^2}{(\sigma/\chi^2)^2}$ as a function of the smoothing length α_D for fixed sampling ratio $\chi_D = 3$. We note that as the smoothing length $\alpha_D \to 0$ we see that $\gamma'' \to 486$ as expected since $\sum_i e_i^2 = 6$. We also note that the measurement error is much more sensitive to smoothing than the corresponding quantity in figure 1. This is not surprising as local differencing operations are much more sensitive to smoothing. As α_D starts to approach $\chi_D = 3$ the bias grows rapidly, and the minimum in $(\gamma'')^2$ occurs for about $\alpha_D \sim 2$.

In figure 4 we plot the optimal smoothing length α_D and minimum normalized total error γ'' as a function of the sampling ratio χ_D. The smoothing length is almost exactly proportional to χ_D, and from the graph we see that to a good approximation we should always choose $\alpha_D \sim 0.6\chi_D$.

There is some abrupt behaviour as χ_D decreases below about 1.4. The optimal smoothing falls suddenly to zero. This occurs because the bias has begun to dominate even at small α_D and, as can be seen in figure 3, $\frac{\sigma_n^2}{(\sigma/\chi^2)^2}$ is very flat for small α_D. For small χ_D, γ'' shoots up very quickly. (In the figure the dashed line becomes unreliable here.) What this tells us is that if you wish to get a reasonable estimate of a second derivative you must sample reasonably finely. If the samples are too far apart relative to the variation length i.e. if the sampling ratio $\chi_D \leq 1.4$ then *a meaningful estimate of the second derivative cannot be extracted.*

Apart from the region $\chi_D < 2$, we see that as the sampling quality improves

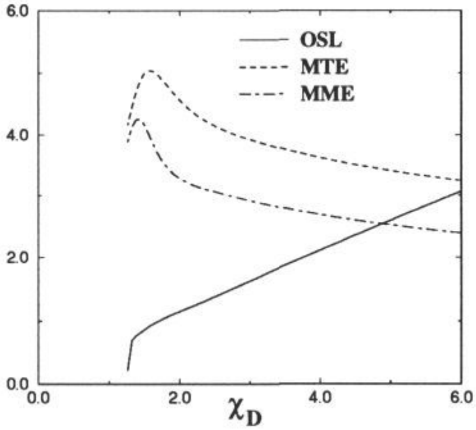

Figure 4: The second derivative: Minimum total error (MTE), minimum measurement error (MME) and optimal smoothing length (OSL) as a function of sampling ratio

γ'' is roughly constant. Recall that $\sigma_{TOT} = \gamma''\sigma/\chi^2$. What this means is that for fixed sample interval Δ, changes in χ or σ affect σ_{TOT} significantly. However when χ and σ are fixed, making Δ smaller won't improve the estimate very much. However as soon as the sampling interval Δ approaches the variation length χ in size you get a sharp increase in the total error.

In summary: sample too coarsely and you get bad results, but you only get limited gains by sampling extra finely!

5 Discussion

How useful is our analysis? If the user is unable to estimate the variation length χ then the analysis is of no use. It may indeed be difficult to estimate χ_4, especially if one does not have the luxury of knowing the function a priori. However one should realize that this an important quantity in the experiment.

Consider the case of measuring the zeroth derivative. Before choosing a sampling interval Δ the user will ascertain the variation length of the sample. Then Δ will be chosen to sample this between say 1 and 10 times. In our terminology this translates to $\chi_D \sim 1 - 10$. $\chi_1 = \sigma/f'$ is a reasonable way to estimate the variation length. Of course the precise choice must depend on how the estimate $\hat{f}(0)$ will be obtained. The relevant smoothing length may well be χ_2.

The point is this: In the zeroth derivative case it is considered routine to have an estimate of the relevant scale of horizontal variation before choosing the sampling interval. By analogy it would not be unreasonable to require an estimate of the relevant scale of horizontal variation when measuring higher order quantities. Obviously these will relate to the rate of change of derivative quantities, and the required knowledge will be some high derivative.

σ	χ_4	1st derivative $\sigma_{TOT} \simeq \sigma/\chi$	2nd derivative $\sigma_{TOT} \simeq 3\sigma/\chi^2$
0.001	0.124	0.008	0.195
0.005	0.185	0.027	0.438
0.01	0.220	0.045	0.619
0.05	0.329	0.152	1.386
0.1	0.392	0.255	1.952
0.2	0.466	0.429	2.763

Table 1: The order 4 sampling ratio for various noise values at $x = 0.70711$

We now apply our technique to a practical problem. In range images of manufactured objects one of the most common surfaces will be cylindrical. Applying the analysis outlined in this paper we can make a number of *purely theoretical* predictions which will be useful when using the facet model to estimate f''.

We consider the problem of recovering a second derivative from a curve of the form $f(x) = \sqrt{1-x^2}$. Clearly the surface has a C^1 discontinuity at $x = \pm 1$, and we do not attempt to estimate f'' near $x = 1$.

Because we know the precise analytic form of the function we can test the validity of the bias approximation [truncating the Taylor series]. We have confirmed that it is reasonable, and does not affect the optimal smoothing length by more than about 10%.

In summary if we wish to estimate the first or second derivative over a cylinder the following rough guide may be useful. This is for the case of the order 2 weighted facet model. Suppose we use $x = 1/\sqrt{2}$ for our estimate of χ_4. Table 1 lists values for χ_4. Then choose some reasonable sampling length, e.g. if $\chi_D = 4$ then $\Delta = \chi/4$. Next read off α_D and γ'' from figure 4, and then the total error for the first derivative and second derivatives are given by $\sigma_{TOT} = \gamma'\sigma/\chi$, and $\sigma_{TOT} = \gamma''\sigma/\chi^2$ respectively. If we take the normalized total errors $\gamma' \sim 1$, and $\gamma'' \sim 3$ then an estimate of the total errors for the first and second derivatives is shown in table 1. Since $f'(0.707) = -1$ and $f''(x)$ varies between about -1 and -3 for $|x| < 0.7$ most of the values for the total error in the second derivative are not very pleasing. The difficulty of accurate curvature measurements becomes plain.

References

[1] P. J. Flynn, A. K. Jain, Surface Classification: Hypothesis testing and parameter estimation, Proc. CVPR Ann Arbor, (1988) pp 261-267. P. J. Flynn, A. K. Jain, On reliable curvature estimation, Proc. CVPR (1989) pp 110-116.

[2] C.-K. Lee, R. M. Haralick, K. Deguchi, Estimation of Curvature from sampled noisy data, Proc. CVPR New York (1993) pp 536-541.

[3] P. J. Besl, Surfaces in range image understanding, Springer, NY, (1988).

[4] H. Knutsson, C.-F. Westin, Normalized Convolution: A technique for filtering incomplete and uncertain data, Proc. 8th SCIA, Tromso, Norway, (1993).