Tina Memo No. 2005-006
Short version submitted to BMVC 2005 (rejected)

Curve Fitting and Image Potentials: A Unification within the Likelihood Framework.

N.A.Thacker.

Last updated
10 / 10 / 2005

Imaging Science and Biomedical Engineering Division,
Medical School, University of Manchester,
Stopford Building, Oxford Road,
Manchester, M13 9PT.

Curve Fitting and Image Potentials: A Unification within the Likelihood Framework.

N.A.Thacker. 10/10/2005

Abstract

This paper provides a purely theoretical discussion of the Likelihood basis for curve fitting. In particular it explains how to construct a valid likelihood for a point generated from an arbitrary curve in circumstances where data has errors in 2D, and where detected features are provided as connected strings. This likelihood is compared to closest point approaches and shown to generate additional terms which can be expected to be the source of parameter bias (previously quantified in other publications [5] and simulated here) if ignored. The Likelihood formulation is then also compared to approaches which make use of "image potentials" in the construction of cost functions. A better understanding of the relationship of fitting algorithms to conventional statistical techniques is needed in order to gain the benefits of a fully quantitative methodology, such as estimation of parameter covariances [2].

Introduction

Quantitative analysis of the location of objects based upon extracted features has to be seen as one of the prototypical problems in computer vision. Fitting of data to a theoretical curve is simple when the measurement errors are in one direction, such as a histogram or function fit. However, in image analysis we must fit curves to data with errors in 2 (or more) dimensions. There are several algorithmic approaches to this, one being where a simple distance measure taken between data points and some point on the curve (typically the closest) is minimised. This may be best described as iterated closest point. Another approach involves computing a cost function at each location in an image and summing this "image potential" along the curve in order to construct a measure of fit, this approach is generally taken by so-called "snakes". Ignoring for now the problem of identifying a global minima (or even being certain that the correct curve has been selected in the first place), both of these approaches are known to exhibit bias. In some cases, such as the fitting of ellipses using iterated closest point methods, the bias is well documented and empirical methods have been suggested to correct it [5]. In fact both of these approaches can be expected to have common origins in probability theory. If this common origin were identified it would give a better understanding of the problem and its solution. As the method of Maximum Likelihood is expected to provide a bias free estimate of parameters for any problem (when formulated in a data space of equal variance), it should be possible to identify both the theoretical basis of the methods and missing terms which would produce biased parameter estimates.

The aim of any likelihood function is to compute a surrogate for the probability of observing a particular data configuration. Likelihood density distributions must therefore agree with empirical observations of data density, this is particularly important if we wish to use these approaches quantitatively. The approach taken in this paper is to derive likelihood functions according to a specific (simplified but realistic) data generation process. These functions are valid for data which is generated in accordance with the assumed sampling process. We will show below how simple changes in the generation mechanism can produce systematic changes in density distributions. Our analysis of bias can then be done by comparing the likelihood functions we derive, with those currently in use.

This paper has not attempted to replicate empirically the biases demonstrated by other authors. Papers such as [5] already provide a good empirical demonstration of the effects of bias on single curves. However, it is our belief that the source of bias in the fit has in the past been incorrectly attributed to factors such as the linearisation performed in the Kalman filter construction. In contrast, the present document derives an equivalent correction for the bias from the definition of Likelihood itself. The distributions generated are confirmed here using Monte Carlo simulation.

Maximum Likelihood Curve Fitting

As mentioned above, construction of a likelihood method for curve fitting where errors on measurements are in only one direction is straight forward. The likelihood for each data point (x, y) can be constructed from knowledge of the expected distribution of measured values around the true solution. Taking a weighting of d^2 for example,

corresponds to a log likelihood assuming a Gaussian error distribution of unknown (but fixed) width σ.

$$p(x,y|d) \;=\; exp{-}(d^2/2\sigma^2) \qquad \rightarrow -\; ln(p(x,y|d)) \propto d^2$$

In 2D however, things are slightly more complicated. In what follows we will assume homogenous (ie: circular) Gaussian errors, though the result can be generalised later by the use of appropriate transforms or alternative error distributions, if required.

The common method of computing a cost from the distance to the closest point can be interpreted as assuming that the only way the observed measurement could have been produced is as a perturbation of data around that (nearest) location. In fact, for a 1 form curve it would be theoretically possible (with varying degrees of probability) to have generated the observed data point from anywhere along the curve. The total probability density would then be proportional to the sum of all possible generation mechanisms.

In concrete terms, we can define the total probability density for the construction of a given location as the path integral of the measurement distribution along a vector of data points $\mathbf{s}$ which comprise the curve. For the purposes of simplifying the mathematics we can work in a co-ordinate system centred on the measured point and oriented so that the point of closest approach is at a distance d in the y' direction (Figure 1(i)). The total probability density[1] is therefore;

$$p(x,y|\mathbf{s}) \;=\; \int_0^S exp[-(x'^2(s) \;+\; y'^2(s))/2\sigma^2]\; ds \qquad (1)$$

In order to construct the full likelihood, assuming independence, we can then write the optimisation function as a sum over all T of the (x_t, y_t) locations in the total data set $\mathbf{t}$.

$$L \;=\; -\; \sum_{t=0}^{T} ln(p(x_t,y_t|\mathbf{s})) \qquad (2)$$

Generally, $\mathbf{s}$ will be defined as a function of a limited number of fit parameters $\mathbf{s}(\mathbf{A})$, and equation 2 will be minimised in order to determine the best set. One of the likely requirements of such a fit would be to establish the scale of the fitted curve. This requires special consideration, as if we simply fix the scale parameters for likelihood normalisation (so that the same value is always computed for equivalent positions within the model regardless of model scale) then there is a trivial solution for an infinitely large curve positioned such that the single highest probability density is situated over all measured data points. This can be avoided by observing that the task we have defined falls into the category of probability density estimation, and the appropriate probability form is Extended Maximum Likelihood (EML, [8, 1]). This requires that equation 2 is minimised subject to the constraint that the integrated probability density is fixed (See Appendix A). This process "protects" the optimisation process from finding the trivial solution by making such a scaling improbable. If we wish to otherwise "fix" the likelihood normalisation, but do not want to solve the task of integrating the probability density, we can achieve the same aim by adding the term $ln(S)$ to equation 2 [2]. In what follows below, I will assume that the effects of curve scaling have been appropriately dealt with.

Fitting to a Circular Approximation

To proceed further we need to know the specific form of the curve s. However, if we assume that measurement errors are generally small in comparison to the shape of the curve it is reasonable to approximate the curve in the region of the measurement as a portion of a circle ($\mathbf{s}'$) with radius of curvature r (Figure 1 (ii)) [3]. Defining the point of closest approach as $s = 0$ for convenience, we can now write;

$$p(x,y|\mathbf{s}) \;\approx\; p(x,y|\mathbf{s}') \;=\; \int_{-S'/2}^{S'/2} exp{-}[r^2\; sin^2(s/r) \;+\; ((d+r) \;-\; r\; cos(s/r))^2]/2\sigma^2\; ds$$

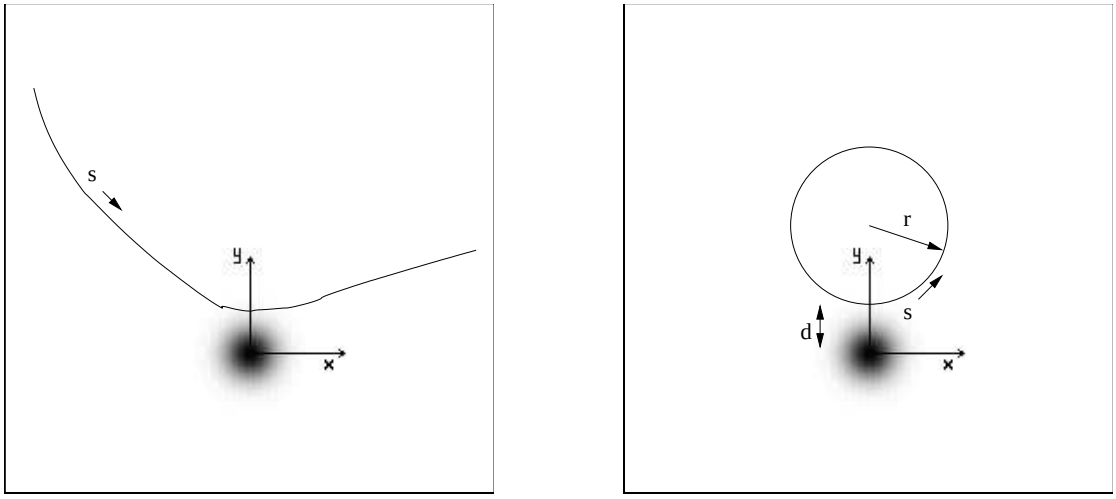
Though this integration is performed around the entire circle, we expect that contributions form the side away from the origin will be negligible[4].

[1] Notice that the use of a lower case p is used to denote the fact that this is not a true probability due to the arbitrary scaling of this value with our choice of s. As always there is an implied interval around the measurement of fixed size, which we may need to remember under some situations.

[2] As the total probability density is the integral of a (normalised) convolution of the curve, the normalised density is always proportional to the length of $\mathbf{s}$.

[3] Note we could also have chosen to perform this analysis with a low parameter polynomial, eg: a parabola, however, we would then have to decide how to scale steps along each part of the curve, the use of a circle here makes the arc length an appropriate and natural choice when working in a space of homogenous errors. See [7].

[4] In addition, note that approximating circles with negative radius of curvature ($-r$) correspond to the data point lying on the other side of the curve.



(i) (ii)

Figure 1: *(i) Path integral for likelihood of observed measurement from a curve, and approximation (ii) using a circle based upon the tangent to* $\mathbf{s}$ *as the point of closest approach and radius of curvature* $\mathbf{r}$*.*

This expression can be simplified to;

$$p(x, y|\mathbf{s}') \;=\; \int_{-S'/2}^{S'/2} exp{-}[d^2 + 2r(d+r)(1 \;-\; cos(s/r))]/2\sigma^2 \; ds \qquad (3)$$

It is worth examining this expression for the case of $r \to \infty$ before we proceed further. In this limit;

$$2r(d+r)(1 \;-\; cos(s/r)) \to s^2$$

In addition, as we have defined $s = 0$ at the point of closest approach, and knowing that in this limit the curve becomes a line parallel to the x' axis, then $s^2 \;=\; x'^2$ in this case. Thus for $(r \to \infty)$ we get the line $(\mathbf{s}'')$;

$$p(x, y|\mathbf{s}'') \;=\; exp[-d^2/2\sigma^2] \int_{-S''/2}^{S''/2} exp{-}[s^2/2\sigma^2] \; ds \;=\; \frac{1}{\sqrt{2\pi}\sigma} exp[-d^2/2\sigma^2]$$

ie:

$$-\; ln(p(x, y|\mathbf{s}'')) \propto d^2/2\sigma^2$$

Thus the use of the point of closest approach is appropriate for the case of fitting data with 2D homogenous errors to a straight line. This validates the common method of "orthogonal regression" described in statistical text books for the fitting of lines.

However, returning now to equation 3 , we can see that for anything other than a straight line use of the point of closest approach is not sufficient to describe the likelihood of the data. Notice that strictly the problem identified in the previous section regarding scale would still need to be addressed, but may be avoided by fixing σ (or ignoring it altogether as done in some fitting methods).

We can see that use of the distance to the closest point provides only the first term in the likelihood for the more general curve.

$$L \;=\; \sum_{t=0}^{T} d_t^2/2\sigma^2 \;+\; ln(\int_{-S'/2}^{S'/2} exp{-}[2r_t(d_t + r_t)(1 \;-\; cos(s/r_t))]/2\sigma^2 \; ds)$$

In this context the second term here would be considered as a second order "bias correction" for closest approach methods. In particular the probability estimates for points "outside" the curve will be less that that for a line, and probability estimates for points "inside" the curve will be more than that for a straight line. The scale of this bias will depend upon the size of the expected error (σ) in comparison to the radius of curvature (r).

This analysis also suggests that there will even be a bias on the fitted radii of circles when using closest point techniques, with a bias towards smaller values. Another way of understanding this problem is to consider the following. If we generate data on a perfect circle and then perturb each point with 2D homogenous Gaussian errors

a count of the proportion of data points outside the circle will give a value greater than 50 %. However, the effects of this process on local data density are also counteracted by the increased volume for data at increased radii. As the assumed Likelihood distributions must match the observed data distribution for an un-biased estimate of parameters, the consequent lack of symmetry has the consequence of bias for the (symmetrical) closest point estimate of parameters.

Analysis of Image Potential Approaches

We now consider approaches to curve alignment based upon the use of "image potentials". Though many researchers in this area would accept that there has to be a link between these techniques and quantitative methods for data analysis, we would argue that the majority view on these algorithms is that they are constructed in order to achieve particular behaviours, rather than from probability theory. This in turn may lead to the idea that it is valid to use a variety of alternative cost functions, with a consequently infinite set of possible behaviours. As with all such approaches, we believe that once the link to conventional methods has been identified it then becomes possible to interpret the cost function as a set of terms which must be justifiable by probability theory (see for example [9] and [10].

Generally these approaches comprise "internal" and "external" terms. However, the "internal" shape terms are clearly intended to define the class of shape behaviour and must therefore take the role of the assumed model in likelihood. We will concentrate here only on the "external" (data driven) term, and will return to the "internal" terms briefly later. In particular we want to consider the process of constructing a cost function from a sum over locations in a "potential image".

Imagine that we have a vector of discrete 2D measurements t, and that we represent these in the form of a binary image with some spatial quantisation $I(x, y)$, such that

$$I(x, y) \; = \; 1 \quad if \quad x, y \; = \; x_t, y_t$$

$$else \quad I(x, y) \; = \; 0$$

We can now convolve this binary image with a kernel (eg: Gaussian of width σ), in order to generate a new image which has real values that fall away gradually away from the data in t.

$$\mathbf{J} \; = \; \mathbf{I} \otimes G(\sigma)$$

At any location (x, y) in the new image J the value will be given by

$$J(x, y) \; = \; \sum_{0}^{T} exp{-}(x^2 \; + \; y^2)/2\sigma^2$$

Comparing this formulation with the curve likelihood (equation 1) we see that there is a lot of similarity. Except for the interchange of data $\mathbf{t}$ with theoretical curve $\mathbf{s}$ (and the approximation of an integral with a finite sum) these two constructs are essentially equivalent. In fact, using the same notation as previously, and for more general distributions (non-Gaussian), we can say that;

$$J(x, y) \; = \; p(x, y|\mathbf{t})$$

If we now locate a theoretical curve $\mathbf{s}$ within this image and construct the quantity

$$L \; = \; - \; \sum_{s=0}^{S} ln(p(x_s, y_s|\mathbf{t})) \qquad (4)$$

this would correspond closely to the kind of data driven terms normally found in "snake" based location algorithms. Thus, the "image potential" approach offers the possibility of working with a more flexible set of probability densities, than simply Gaussian (eg: the more robust bi-exponential). This can be done by substituting the appropriate convolution kernel.

Equation 4 corresponds to constructing the likelihood as a distribution of the model about the data. For snake based approaches this formulation is necessary so that the "potential image" does not change with changes in the parameterised curve. Although convention would lead us to expect the earlier approach to be the preferred option (a likelihood based upon a perturbation about the assumed model), it can also be seen as an interchange of definition between what constitutes the model and what the data. We therefore need to ask ourselves if this is a valid step; are there genuinely two alternative ways to interpret the same data?

Resolving this potential ambiguity requires us to consider where we wish to make the assumption of independence. While it is legitimate to believe that individual measured data are sampled independently from the curve, it would not make sense to suggest that theoretical curve points are sampled independently from the data distribution. The unfortunate conclusion from this is that this second approach must be interpreted as an approximation to equation 2 which becomes exact only for noise free data (and the correct model).

One consequence of this is that the likelihood definition (fixed normalisation for the probability density) no-longer "protects" the optimisation function from changes in scale to the fitted curve. This means that if scale changes are allowed during optimisation the best fit curve in all circumstances corresponds to the trivial solution of shrinking the curve onto the maximum probability (image potential) value. Correction for this problem can be derived by considering the underlying probabilities that the likelihood approach is attempting to capture, and (as with the alternative normalisation correction defined for EML described above) requires a correction term;

$$L \;=\; -\;\sum_{s=0}^{S}[ln(p(x_s, y_s|\mathbf{t}))\;+\;ln(S)] \qquad (4)$$

for an overall length change to the theoretical curve $\mathbf{s}$.

This analysis has been rather successful at linking conventional curve fitting and image potentials, within the Likelihood framework. Showing that an image potential generated by convolution with a fixed blurring function is equivalent to assuming a stochastic sampling of measured feature locations, with an error distribution related to the convolution kernel. Even so, this approach needs to go some way if we wish to generate realistic Likelihood estimates for the description of shape.

Returning now briefly to the issue of the "internal" terms, now that we know that the "external" term is approximating a likelihood, this tells us that the "internal" terms should strictly be describing other log probabilities, for example the expected geometrical relationships between adjacent points on the curve. Only by doing so could the combined cost function be composed of commensurate (like) terms. In addition, this also tells us that any terms we choose to use make specific assumptions regarding the expected distributions of data. If these distributions are sample dependant, so that they will change if the data (object set) is changed, then we must expect also to have to modify these terms to compensate. There can therefore be no single group of terms which are optimal for all sets of arbitrary objects. Alternatively "internal" terms must be conditional upon the object hypothesis. This logic leads naturally to the use of approaches such as active contour models (ACM's) as a principled extension of 'snakes'. In addition we can also see that for some tasks, such as 3D projection of rigid objects, all knowledge of object structure is already embodied in the object model and additional terms in snake formulation are unnecessary.

A Likelihood for Connected Features

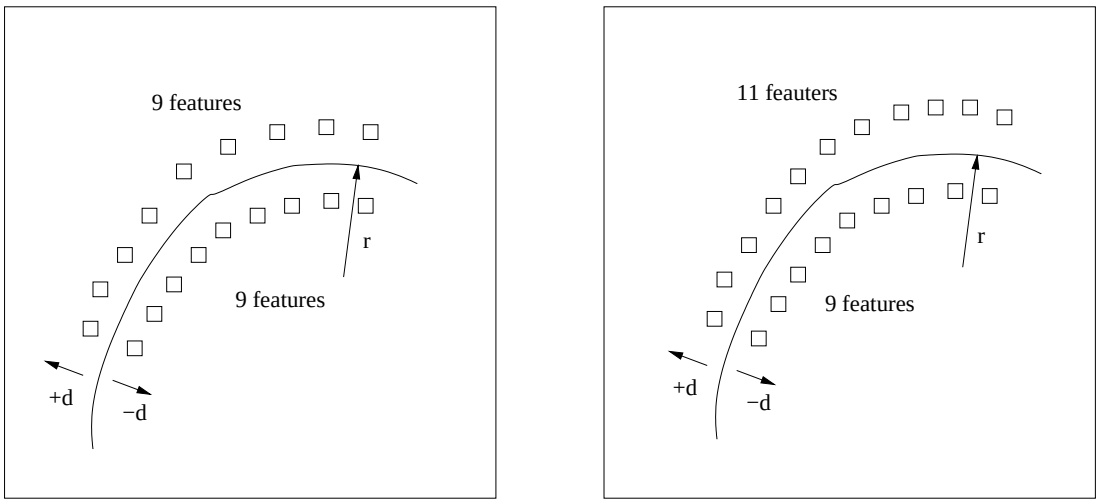
Although an assumed generation process which has a 2D perturbation about a point on a curve might be a reasonable model for the fitting of isolated features (such as corners), it is not appropriate for the matching of a curve to a set of edge features. In this case we know that the direction along the detected feature string conveys little or no information. Moreover, as algorithms such as edge detectors are designed to produce connected strings there is a fixed density of detected features along the string. We need to build this property into the definition of our likelihood model some way. Strictly we should start from the noise in the image and build up but we will assume here that feature locators have approximately fixed spatial accuracy for simplicity. In addition we will ignore the complication of quantization of an image into pixels, and assume that such effects average out in the statistics of large samples, particularly if an effort is made to work with sub-pixel feature locations.

If we start for example by assuming the 1 dimensional nearest point likelihood model this gives the following radial density model.

$$p(x, y|\mathbf{s}) \;=\; exp(-d^2/2\sigma^2)$$

This would imply that for every point along the curve we would have an equal probability of generating a point inside or outside of the curve at a distance d (Figure 2 (i)). The problem of trying to apply a 1D density model to 2D data is that it is easy to neglect the consequences of changing orientation for the sampling process. As we can see from the figure, a generation process with this property will appear to increase the separation between data moved by the noise process to the outside of the curve, relative to those on the inside. Therefore, it will not produce a fixed average interval edge string, and the change in interval will be a function of local radius of curvature (r), ie: a systematic effect of curve shape.

In order to get the density of edges to match on the inside and outside of the fitted curve we need to make a correction to the expected probability density in proportion to the expected increase in spacing as a function of



(i) (ii)

Figure 2: *(i) Edge density generation for the equivalent of a nearest point likelihood and (ii) The relative increase in edge density needed on the outside of the curve in order to generate a connected string.*

radius and distance from the curve. The density model for this case is;

$$p(x, y|\mathbf{s}) \;=\; \frac{r+d}{r} exp(-d^2/2\sigma^2) \qquad (5)$$

So that the corresponding likelihood is;

$$L \;=\; \sum_{t=0}^{T} d_t^2/2\sigma^2 \;-\; ln\,(\frac{r_t + d_t}{r_t})$$

Once again this demonstrates that simple closest point fitting routines, which do not include this second term, are likely to exhibit bias on fit parameters. This explains the bias observed when fitting ellipses to edge strings using point of closest approach methods and suggests (as is observed) that fitted ellipses will generally appear to be more elongated than the true curve [5]. The bias correction term suggested in this earlier work is identical in form to that derived here for connected features. However, the suggested source of this bias is not as discussed in this present document (Appendix B).

Results

A comparison of the theoretical predictions for isolated features and connected features (based upon a sampling process with homogenous errors), allows us to conclude that these two mechanisms (isolated features and connected features) are not equivalent. A Monte-Carlo simulation of the two processes generates the distributions shown in Figure 3. Results shown are for the rather extreme circumstance of a circle of radius 64 pixels and a feature location accuracy of 20 pixels so that the biases are visible at the level of a single measurement distribution. The theory explains how these biases scale with radii and inherent measurement accuracy, and the effect of these processes on practical curve fits have been detailed in previous work.

It can be seen that the enhancement in density is on opposite sides of the curve for the two approaches. Fitting of strings (edges) is therefore not equivalent to fitting of isolated points (corners) and both cannot be corrected with the same correction process. This observation also has implications for our analysis of "snake" potentials, which are more easily compared to the likelihood needed for isolated features than for connected curves.

Discussion and Conclusions

This document analyses methods for the location of curves in 2D image data from the standpoint that a genuine Likelihood based approach would deliver un-biased estimates of location and shape. I have explained how in the

(i) (ii)

Figure 3: *(i) Empirical radial distribution for uniform 2D errors and (ii) empirical distribution for connected feature model (arbitrary normalisation) showing skewing in opposite directions from theoretical central value (peak at 64).*

absence of definite correspondence between data and points on a curve, the likelihood for an observed measurement must be composed from probable contributions from any point on the curve. This leads to the exact probability density estimate defined in equation 1. An analysis of this expression shows that iterated closest point algorithms can be justified strictly only for line fits. For other curves a bias must be expected in estimated parameters. An alternative probability density estimate is provided for the case of circular distribution and it is argued that this might be used as the basis for a second order correction for arbitrary curves when fitting to isolated features (such as corners). In addition the approach taken to the construction of "image potentials" is related to this framework. Though this approach has some advantages, in that arbitrary curves are already treated in a way which might be expected to correctly account for local curvature, it is shown that here too this approach can only be an approximation to the required likelihood. The likelihood analysis has then been extended to cover the fitting of connected features, where the linear density of detected data is fixed. Once again the potential for a bias is demonstrated when fitting with closest point methods. While the 2D based likelihood techniques can be easily related to "potential image" fitting approaches, it may clearly be much more difficult to define a potential which corresponds to the required form for connected features. Certainly, the dependency of the likelihood upon local curvature would appear to make the construction of a valid potential from a convolution with a fixed kernel impossible. This departure from a clear-cut likelihood interpretation is enough to suggest that snake based approaches have the potential for both reduced accuracy and bias in comparison to what could have been done otherwise with the same data when fitting individual curves. This could yet lead to improved methods for locating objects in image data, and a suggestion is made in Appendix C for how this may be done. It must also be said however, that these bias processes may be considered negligible when simultaneously fitting multiple theoretical curves, such as in [4]. Here it was observed that even problems such as systematic effects of illumination could be dealt with, when there are many constraints (features) in the model and a robust kernel is used.

This discussion has been currently limited to purely theoretical considerations. These ideas should stand or fall on their own merits[5]. It is therefore important that previous work [5] has already demonstrated the form of the required bias terms. It is hoped that the reader will understand that this attempt to understand the theoretical basis of common algorithms is more than an excuse to invent another algorithm (the literature probably has far too many already). One of the first benefits of understanding the origins of these methods is that a direct linkage between Likelihood and an algorithm is a necessary precursor to the possibility of applying methods for the estimation of error covariances on parameter estimates [2, 3].

Simple curve fitting is no longer of significant interest in current literature. In particular, many recent techniques now take approaches which attempt to incorporate image based models directly into the feature location process. However, much of the emphasis of this work is based upon specific ways of representing the models, training parameters or subsequent search. Little has been said regarding a theoretical (probabilistic) justification for the cost functions used. The simple example shown here of the effects of scale change hint at the complications which may be found when dealing with, for example, models of deforming shape. It is therefore my belief that much can still be learned from considering the origins of shape based analysis using Likelihood. In particular, if biases are known to be present in ellipse fitting problems based upon iterated closest point approaches, it is difficult to believe that these problems will have gone away simply by extending these approaches to include image data.

[5]Indeed, no amount of supporting data would otherwise convince the sceptical that this particular analysis was valid, as (as for the Monte Carlo results presented here) we could always choose to use data which was consistent with the assumed data formation process.

Comments

My main motivation for writing this paper was to try to explain, how ideas such as 'snakes' might be justified from a statistical viewpoint, and specifically what we need to assume in order to derive them. Such a level of understanding is needed if they are to be used correctly. When this paper was submitted for publication it was rejected, largely on the basis that the contents were deemed already known by the community at large. However, even when reviewing for the same conference the following year I got papers describing 'snakes' which were written by researchers who seemed (as usual) oblivious to any link with probability. I seem to remember that my response was slightly more restrained than; "I've been told you should know better than this - REJECT!"

I did not persist in attempting to publish my work, I had already concluded what was really needed was an approach which matched the distributions of random variables seen in data, rather than just assuming spatial Gaussian forms. Though the work shows and interesting link between curve fitting and image potentials, the intricacies of the work in this document become rather pointless if the measured data does not have spatially isotropic measurement characteristics to begin with. Our following work, (see Tina memo 2006-007) addressed this issue directly by using a bootstrap approach to estimate both likelihoods and hypothesis probabilities based upon the statistical noise in an image and the algorithmic process of feature detection.

Appendix A: Likelihood Normalisation; Peak or Area

In this document we assert that the correct normalisation principle for EML is to normalise to the area under the probability density distributions. This is consistent with the definition of Likelihood found in many text books. In previous work we have asserted that Likelihood should be normalised so that the peaks of individual data distributions are set to 1. We know from statistical methods that this latter approach is what is actually needed to obtain a quantitative measure. So the question arises, why are there these two possibilities (which is "correct")?

The solution to this apparent contradiction requires us to remember the derivation of likelihood (and in particular EML) from probability theory. As explained in [7] the derivation of likelihood requires at some stage a move from discrete events to continuous variables. In EML this has been done assuming equal intervals and therefore equal variance. Clearly for fixed equal variance a normalisation to the area is equivalent (for practical purposes) to normalising to the peak. In other situations, such as bootstrap likelihoods, where there may be localised variations in distribution width (defined due to statistical precision of the measurement) we should strictly be normalising to the peak, as this is the one which correctly follows the underlying definition of the discrete event. EML is thus only valid because we do not expect the intrinsic measurement precision to vary across the space in which the data is measured.

This observation also therefore has consequences for the use of the EM algorithm for the fitting of data density distributions. The EM algorithm, and fixed normalisation of the total density distribution, is not appropriate when we believe that the source of the distribution widths is actually due to changes in measurement precision.

I have not seen before in the literature any indication that measurement precision and inherent sample distribution cannot be combined into one effective distribution width as though they could be treated equally under probability theory. Though clearly, there should be some difference, as measurement precision sets the limit of what can be considered statistically significant changes in the data, whereas sample distribution is related to the signal component of the measurement.

Appendix B: Bias Correction Ellipse Fitting

The paper by Porrill [5] provides a good empirical demonstration of the effects of bias on single curves. This correction process is still used for the fitting of edge strings in the Tina vision software. However, it is my belief that the source of bias in the fit is incorrectly attributed to the linearisation performed in the Kalman filter construction. In particular I believe there is an error in the interpretation of the use of the variational principle, which strictly only defined for Gaussian distributions, is used to derive a correction which changes the likelihood shape. In contrast, this document derives an equivalent correction for the bias from the definition of Likelihood itself.

Appendix C: A More Realistic Potential Image Approach for Connected Edge Features

In this section I would like to combine all of the result so far into one recommendation for a potential image approach. However, there are still a few additional issues which remain to be discussed before this can be done.

The first problem is that practical edge detectors can not really be considered as a stochastic sampling process. The results from an edge detector are actually far more stable than a mechanism for detection based upon random sampling from a single distribution would suggest. Secondly, in order to be able to recommend a theoretical approach for the construction of image potentials, we really need to address the issue introduced above of constructing the correct likelihood from $p(x, y|\mathbf{s})$ rather than $p(x, y|\mathbf{t})$.

The first of these observations can be addressed by observing that most edge detectors do not simply generate edge elements at quantised pixel locations. In fact the use of quite high signal to noise in images means that detected edges are not only quite reliable (much more so than a random sample) but can also support sub-voxel estimation of edge location. Thus we would like to be able to make appropriate use of this information in later stages (rather than producing a coarse binarisation as described previously). This suggests that what we may wish to work towards is the definition of an edge probability image as the summary of our measured data. As it happens, this idea also turns out to be useful in the solution of the second problem.

In order to simplify things, I will assume that the model and the data have been generated by an identical measurement process. Or equivalently, the modelling process is intended to generate data which is an instance of what could have been observed. This means that we can expect the Likelihood method we use to be symmetric under interchange of the data and model. The EML formulation is appropriate for the comparison of finite quantities of sample data to a theoretical distribution on the basis of joint probability. However, once we decided that we wish instead to compare probability densities (describing both the model and the data) then an approach which provides a prediction of the generalisation of the interpretation to unseen data becomes appropriate. This suggests using the Bhattacharyya measure. [6, 8].

$$L = \sum_{x,y} \sqrt{p(x, y|\mathbf{s})p(x, y|\mathbf{t})} \qquad (6)$$

This process requires integral normalisation of both probabilities, but as the data sample we wish to fit is fixed we need only worry about normalising the model probabilities $p(x, y|\mathbf{s})$. This normalisation is expected to solve the scale issues identified in previous sections.

Given this formulation, it only remains to construct appropriate estimates for $p(x, y|\mathbf{s})$ and $p(x, y|\mathbf{t})$. From the analysis in the previous section we know that the appropriate form for the probability distribution as we move away from the sub-pixel location of the a theoretical curve is given by;

$$p(x, y|\mathbf{s}) = \frac{r + d}{r}\Phi(d, \mathbf{s})$$

where $\Phi(d)$ is a kernel function which determines the density model for the measurement process and r is one again the local radius of curvature. By symmetry therefore we must also have;

$$p(x, y|\mathbf{t}) = \frac{r + d}{r}\Phi(d, \mathbf{t})$$

for the probability density around the measured data $\mathbf{t}$.

The process of locating a theoretical curve within measured feature data therefore becomes one of first computing a potential image, $\sqrt{p(x, y|\mathbf{t})}$, from a process of summing from spatially varying kernels along the ridges of detected edges. The possibilities of multiple radial contributions from different parts of the curve should be summed, as in the uniform error example. This is actually likely to be rather a complicated process, and we discuss this issue later below. Secondly a model potential image must also be constructed following changes in the theoretical parameterisation in the same way. Thirdly the likelihood can be computed from equation 6.

Clearly, we have attempted to address the issue of bias by correcting the probability densities considered. This is the main reason for the suggestion of these rather awkward distributions. However, this algorithm construction would not have the same problems with bias as standard curve fits even if we had not done this. This can be understood by performing a simple thought experiment; imagine a Monte-Carlo experiment in which we empirically observe the shift in mean parameter estimates from repeated independent data samples drawn from the expected distributions. Any bias observed for one labelling of model and data would need to change sign under interchange of labels. Yet a bias, by definition must be systematic and cannot change sign in an otherwise equivalent experiment. Thus, **any**

cost function which is symmetrical under interchange of (equivalent) model and data definitions, **cannot have parameter bias**. It may however, make sub-optimal use of the data (and therefore have reduced stability and increased error) in parameter estimates. As the probability density estimates cannot be exact, the issue of using the theoretical distributions then becomes one of believing that such a correction will make a significant improvement in probability and therefore parameter estimates. We may therefore have the freedom to use simplified density distributions, including those based upon uniform error distributions, without significant loss of theoretical rigour.

Notice that algorithmically this process does not have the simplicity of a standard "snake" approach, neither does it have the potential for closed form solution offered by "least-square" constructs. However, only having to attend to variation in the model potential (which can be constructed by consideration of the theoretical curve locations), should not be a heavy computation burden. It should also be of value that we are free within this framework to use a density distribution Φ which is in agreement with practical edge detector performance.

References

[1] R.J. Barlow. Statistics: A Guide to the use of Statistical Methods in the Physical Sciences. John Wiley and Sons, U.K., 1989.

[2] R.M. Haralick, *Performance Characterization in Computer Vision*, CVGIP-IE, 60, 1994, pp.245-249.

[3] P. Courtney and N.A. Thacker, *Performance Characterisation in Computer Vision: The Role of Statistics in Testing and Design*, "Imaging and Vision Systems: Theory, Assessment and Applications", Jacques Blanc-Talon and Dan Popescu (Eds.), NOVA Science Books, 2001, ISBN 1-59033-033-1.

[4] A.Lacey, N.A.Thacker, P.Courtney and S.Pollard, TINA 2001: The Closed Loop 3D Model Matcher. proc. BMVC, 2001, pp203-212.

[5] J.Porrill, 'Fitting Ellipses and Predicting Confidence Envelopes using a Bias Corrected Kalman Filter.' Proc. 5th. Alvey Vision Conference, 175-185, Sept. 1989.

[6] N.A.Thacker, F.Ahearne and P.I.Rockett, 'The Bhattacharyya Metric as an Absolute Similarity Measure for Frequency Coded Data.' Kybernetika, 34, 4, 363-368, 1997.

[7] N.A.Thacker, P.Bromiley, The Equal Variance Domain: Issues Surrounding the use of Probability Densities for Algorithm Construction. Tina memo, 2004-005, 2004.

[8] N.A.Thacker, Parameter Estimation for EM Mixture Modelling and its Relationship to Likelihood and EML., Tina memo, 2004-006, 2004.

[9] R.S. Stephens, A Probabilistic Approach to the Hough Transform, proc. BMVC90, pp 55-60, 1990.

[10] A. Roche, G.Malandain, N.Ayache and S.Prima. Towards a Better Comprehension of Similarity Measures Used in Medical Image Registration. MICCAI, 555-566, 1999.