Tina Memo No. 2006-010
Internal.

Models for Quantitative Testing of Edge Hypotheses.

Neil A. Thacker.

Last updated
25 / 11 / 2007

WWW.TINA-VISION.NET

Possible Models for Quantitative Testing of Edge Hypotheses.

N.A.Thacker

Abstract

Some thoughts on the calculation of quantitative probabilities of edge detection, for investigation during the 3D object recognition and localisation project. See also Tina memo 2006-007.

Introduction

Many applications require the detection of edges as a starting point for image analysis. We would like to localise a projected geometric feature by comparing the an IID feature saliency map (Fisher Information) with a predicted position. For example, a method for fitting curves to extended edge structure will require a model of location accuracy of an edge in order to construct a Likelihood based estimator.

Clearly, the Fisher Information measure is not a Likelihood and therefore cannot be directly combined with other such terms. One problem which needs addressing is therefore that of determining how to obtain a Likelihood from a significance map. If the dominant source of location error on the detected edge is due to the object being imaged then it is reasonable to construct such Likelihoods on the basis of image plane distances away from a local maxima. However, for accurately modelled objects, such as man made industrial components, the dominant source of error may be from the image measurement, ie: the grey level distribution. In this case we need to understand how the noise, illumination and local structure in the image affects the edge detection process. It is then obvious that such processes will **not** be correctly modelled purely in terms of spatial distances.

Quantitative use of edges also requires an understanding of the fundamental accuracy and stability of the detection process. We would like to be able to form valid hypothesis tests to confirm that edges are present in the locations expected. This becomes more complicated the greater the sophistication of the detection process. Our favourite algorithms may be so complex that it is difficult to provide any analytic understanding of the process. The aim of this document is therefore to construct a statistical model for a simple edge detector based upon ridge detection (a value being greater than some number of its nearest neighbours) and thresholding in a feature significance image.

Derivation of the Power Law Model

We might start by assuming that the dominant form of perturbation along an edge is purely the grey level noise process. In practice however, the appearance of an edge at a specific location will depend upon how the extended continuous structure generates the observed pixels. Under specific illumination, changes in position along this edge will also generate a form of data perturbation due to partial volume processes, which are perfectly allowable variations in feature apearance, but a simple edge model is in general unable to predict.

We can take a bootstrap approach to this problem. We define the probability that a particular pixel g (quantifying the local image gradient) would be greater than a single sample drawn from the distribution of N local pixels.

$$R \;=\; \frac{1}{N+1}(0.5 \;+\; \sum_i^N \int_{-\infty}^{x_i} p(g,y)\; dy)$$

We can regard this as the ''soft rank'' of the central pixel in comparison with its neighbours (see Tina memo 2009-008).

We can get closer to a feature detection process by defining the likelihood p_{max} that the central pixel would be larger than some number of its neighbours. Careful use of this strategy can even identify connected structures. On a 3x3 region labelling pixels as edges if they are greater than six neighbours and above the noise floor identifies connected (one pixel wide) edge ridges well.

If the probability that the central gradient value would be larger than its neighbours is R then the probability that it is greater than exactly six from eight randomly selected neighbours $P_{(6/8)}$ is obtained from the binomial distribution

$$P_{(6/8)} \;=\; 28R^6(1-R)^2$$

The probability that it is greater than 6 or more neighbours is then

$$p_{max} \;=\; P_{(6/8)} \;+\; P_{(7/8)} \;+\; P_{(8/8)} \;=\; 28R^6 \;-\; 48R^7 \;+\; 21R^8$$

The probability of a value being consistent with an edge must also take into account the threshold limit t for detection. We can combine this as an independent process, so that;

$$p_e \;=\; p_{max} p_t \;=\; p_{max} \int_t^{\infty} p(g,y)\; dy$$

Note that as we are working with gradient values and not image grey levels $p(g,y)$ is likely to Rician in shape for $g \approx 0$. This is the expression we would expect to use to define the location of a projected curve, ie: a continuous function of R or density, which can be taken to be a Likelihood. In order to make quantitative use of this probability we must confirm that the data will be distributed in accordance with the assumed bootstrap model. In particular we need to be able to construct a meaningful hypothesis test.

We can get from this density to a hypothesis test by observing that when, as in this case, there is a monotonic relationship between the two they are related by an integration (see Appendix)

$$H_e \;=\; \int_0^R p_e \; dR$$

integrating by parts we get

$$H_e \;=\; [p_t \int_0^R p_{max}\; dR]_0^R \;-\; \int_0^R (\int_0^R p_{max}\; dR) \frac{\partial p_t}{\partial R}\; dR$$

for which it is reasonable to assume that the second term is negligible for above threshold edges as $\frac{\partial p_t}{\partial R} \approx 0$ and $p_t \approx 1$. Then the corresponding hypothesis test H_e is approximated by the normalised integral

$$H_e \;\approx\; p_t(12R^7 \;-\; 18R^8 \;+\; 7R^9)$$

However, there are clearly many different ways that we can define a detection process.

- Within the same basic framework, we can have a different definition for an edge detector. We may choose to use a larger sample region than 8, or to say that a gradient greater than all neighbours is really a corner, and therefore not an edge.
- Several approximations, such as for H_e above, may be necessary in estimating a given value of P using the local rank (R). The estimated value of R might also be systematically affected by the processing we apply to images to reduce noise and estimate gradients.
- R may not be the best domain in which to compute p_{max}. Ideally, probability densities should be constructed in and "equal variance" domain in order to maintain quantitative validity. This is not an issue for p_t as we might expect the image data to have uniform independent random noise. However, if we believe that R approximates a Binomial sample then the equal variance space for Likelihood construction requires the "arcsin" transform [1]. Equivalently we can scale our density by multiplying with the appropriate interval , ie:

$$P_e \;=\; P_{max} P_t \;\propto\; \sqrt{R - R^2}\; p_{max} p_t$$

There is consequently some degree of flexibility involved with how we can choose to define and compute P_e, but crucially, under all of these circumstances, any specific form can always be expressed as a fixed polynomial in R and tested in data.

One approach to this problem is to seek edge density distribution models which are generated by the locations people would define as edges. However, as we have seen previously, our definition of maximal information may not correspond with a geometric definition of an object boundary. Our edge based models of shape therefore need to be capable of predicting where edges will be detectable on the basis of the applied statistical tests, ie: ideally our models should be learned from real image data rather than being predictions from naive CAD models.

It is therefore legitimate to select simple functions which approximate P_e and H_e well with a small number of parameters, such as

$$P_e \;\approx p_t R^M$$

Estimation of one parameter is not difficult even on small samples of data, but the next simplest polynomial

$$P_e \;\approx \alpha R^M \;+\; (1-\alpha) R^N$$

[1] This is non-standard for Likelihood construction, we may very well be the only group in the area who would advocate this.

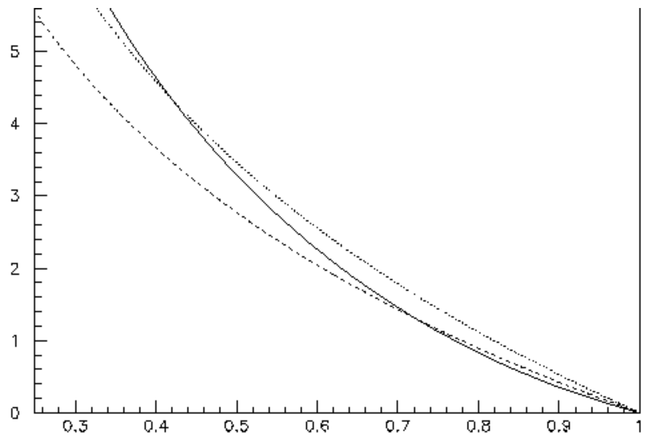


Figure 1: *The similarity of* $-log(4P)$ *(dotted curve) and* $-log(5P)$ *(dashed curve) to* $-log(H_e)$ *(solid curve). The minimum difference of* $|log(H_e) - Mlog(P)|$ *is less than 0.2 for* $H_e > 0.01$ *with* $M = 4.6$. *This value would therefore be expected to be the best matching curve (on the basis of a Kolmogorov-Smirnov test) in real data, if* H_e *is based upon a valid definition for an edge.*

Histogram	**All pixels**	**Edges**
R	constant	$R^{M'}$
$P = R^{M'}$	$P^{(1-M')/M'}$	$P^{1/M'}$
$H = R^{M'+1}$	$H^{-M'/(M'+1)}$	constant
$Y = -log(H)$	$exp(-Y/(M'+1))$	$exp(-Y)$

Table 1: *Theoretical data density distributions for a power law model. The distribution of rank values for all pixels is constant by definition for a 'hard' ranking process and should remain this way following the inclusion of the threshold probability term if it is independent of maxima selection. We further assume that true edges have a distribution proportional to* R^M, *all other results then follow on the basis of the specified non-linear transformations.*

has 3 free parameters. Estimation of parameters then becomes more difficult. It therefore makes sense, and is computationally more efficient, to stay with the simplest model if results are already acceptable. We can accommodate this process by adjusting the number of random comparison pixels M', with $P_{max} \approx R^{M'}$ in order to match the data. Figure 1 shows how $P_{max} \approx R^{M'}$ approximates H_e for $M' \approx 4.5$, we expect to see a value similar to this in real data.

The above argument is a theoretical prediction and overall justification for taking a 'power law' approach. We can test the validity of these conclusions by confirming that histograms of the derived variables (figure 2) have the distributions predicted for this model (table 1), computed in the standard way for non-linear transformations (Appendix). A cumulative distribution of the log hypothesis, which is suitable for direct use in a Kolmogorov-Smirnov test, will have an inverted exponential distribution, As long as this is found to be true in practice the actual argument for how it arises then becomes unimportant. The ultimate requirement is that the hypothesis distribution is uniform (fig 2(c)) (and behaves well around any hypothesis threshold we may use) for those locations we define in a model as true edges. Once this has been achieved the corresponding expressions for edge Likelihood can be identified.

Summary

This document has suggested an approach for the construction of quantitative probability models for edge detection. The calculation R^M defines an edge as a pixel with a gradient which is on average greater than M neighbours. Though this is a less sophisticated model of edge detection than P_e , it is not entirely unrealistic. It is perhaps less likely to identify a row of edges as a pixel wide connected structure, but will suppress responses from edges adjacent to stronger structures in the region used to define R. This model has the additional advantage that the log-Likelihood needed for object alignment differs from the log of the hypothesis by a fixed scaling factor (ie: it can be neglected for most purposes). Indeed, the logarithm of ranked pixel values multiplied by a suitable noise

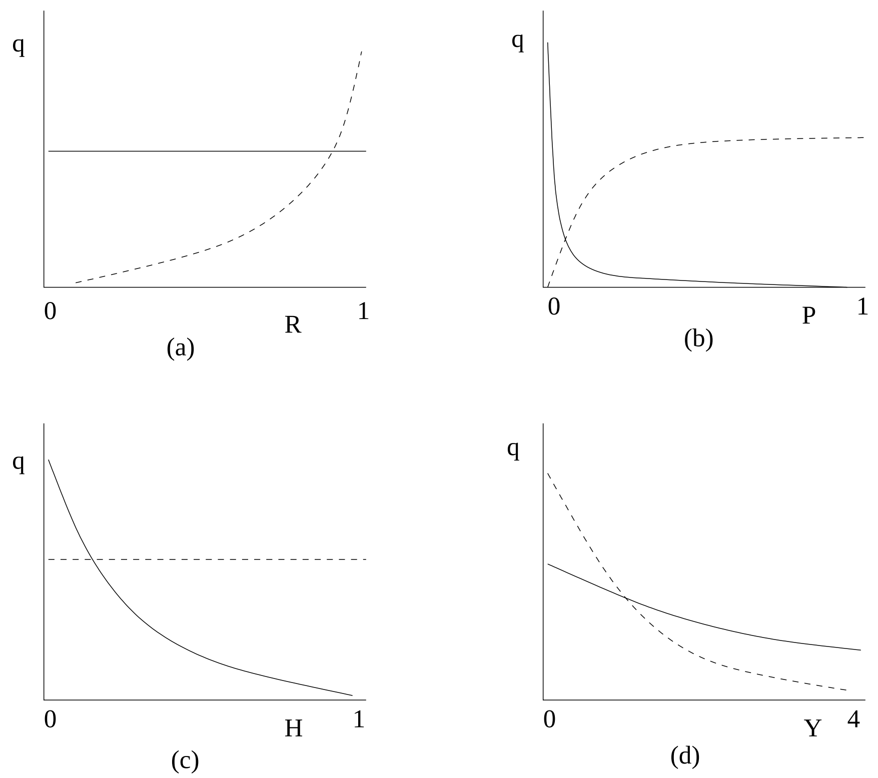


Figure 2: *Approximate data density distributions (histograms) for all image data (solid) and edge data (dotted) using a power law model of edge probability density ($M \approx 4$). Plotted as a function of (a) gradient rank R, (b) edge probability P, (c) edge hypothesis H, and (d)log of edge hypothesis according to the functions specified in table 1.*

threshold term, is an equivalent log-Likelihood for object aligment [2]. In addition, as the distribution of R for all image data is by definition uniform, from Bayes theorem the conditional probability $P(edge|data)$ of a gradient with local rank R corresponding to an edge is proportional to P_e. When we maximise the Likelihood for the position of an extended feature using the sum of logarithms of this quantity we are also maximising the joint probability of all locations being classifed as an edge.

The power law model has been adopted as the standard aproach for calculation of edge Likelihoods and Hypotheses, empirical studies are now required to determine if either has practical utility. Preliminary distributions, obtained while developing these ideas are shown in figure 3. More systematic evaluation of the power law model is given in 2006-007.

Appendix

General transformation of variables $y = f(x)$ results in a transformation of density distribution $q_x \rightarrow q_y$ according to

$$q_y = q_x / \frac{\partial y}{\partial x}$$

so that (for example) the histogram of all data in an image (uniform in R) for variable $P = R^M$ is given by

$$q_P \propto 1/\frac{\partial P}{\partial R} = 1/R^{M-1} = P^{(1-M)/M}$$

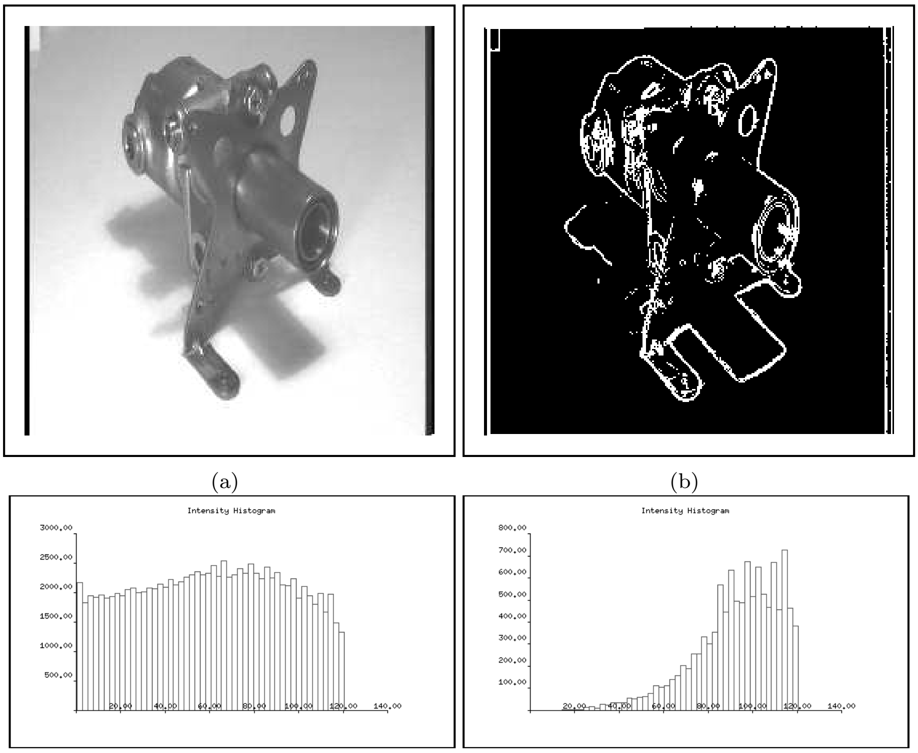
Note that for $q_x = f(x)$,if we define $y = \int f(x)dx$ then

$$q_y = f(x)/f(x) = constant$$

regardless of $f(x)$.

This probability integral transform is the basis of conventional hypothesis tests for monotonic $f(x)$.

[2]I have to point out the obvious analogy here with biology, where the retina is known to enhance edges, and subsequent neural processing is thought to support "winner-take-all" mechanisms for ultimate encoding of features as mean fequencies (or firing rates) in the brain. Micro-saccades could also be the biological analog of the bootrap resampling we used to motivate use of R.



(a) (b)

(c) (d)

Figure 3: The test shown (a) was smoothed and used for calculation of image gradients, these were thresholded and the mask used to select near edge pixels, histograms of ranked gradient (11x11 region) were then constructed for all (c) and thresholded data (d) c.w. figure 2 (a). The data follows the expected model quite well, for data below $R = 0.5$.