
Edges: Saliency measures and automatic thresholding

Paul L. Rosin

Institute for Remote Sensing Applications
Joint Research Centre
I-21020 Ispra (VA), Italy
paul.rosin@jrc.it

Technical Note No. I.95.58
May 1995

Abstract

Edges are useful features for structural image analysis, but the output of standard edge detectors must be thresholded to remove the many spurious edges. This paper describes experiments with both new and old techniques for: 1) automatically determining appropriate edge threshold values, and 2) determining edge saliency (as alternatives to gradient magnitude).

keywords: edge threshold saliency assessment

1 Introduction

One of the most common image features used in machine vision are edges, and there is a substantial body of research on various techniques for performing edge detection. A drawback with using edges is that not only do edge detectors extract meaningful and useful edges, but also many other spurious ones which arise from noise and minor changes in intensity values. If all such edges are kept then the resulting clutter is hard for subsequent processing stages to analyse, while the large number of edge points can seriously degrade computational performance. The alternative is to select a subset of edges for further analysis, and disregard the remainder. This is generally done by a threshold on the gradient magnitude of pixels.

Surprisingly little attention has been paid to the problem of automating the thresholding of edges. Most standard image intensity thresholding algorithms cannot be applied since they assume bimodal (or multi-modal) intensity histograms while gradient magnitude histograms are more likely to be unimodal. Moreover, other inappropriate assumptions are often made such as modelling the two populations by Gaussian distributions, and not expecting the sizes of the populations to be too dissimilar. Unfortunately, in practice edge thresholding is often done in an *ad-hoc* manner, frequently requiring user tuning of parameters. To enable the building of robust machine vision systems it would preferable to automate the edge thresholding process so that the systems can adapt to different scenes and imaging conditions without manual intervention.

This paper describes two facets concerning edge thresholding. The first is the determination of appropriate edge saliency measures. Although the gradient magnitude is generally used other possibilities such as wigginess, lifetime, etc. also exist. Second, two methods for performing automatic edge thresholding are described. These follow on from Voorhees and Poggio (Voorhees and Poggio 1987) work on modelling the gradient magnitudes arising from noise, and our previous work (Venkatesh and Rosin 1995) that analysed complete edge curves rather than independently thresholding each pixel. Both synthetic and real image data with associated ground truth are used to assess the edge saliency measures and automatic thresholding techniques. The effectiveness of the saliency measures are analysed by considering the trade-offs between the amount of true and false positives for all the different threshold values, as shown by their receiver operating characteristic (ROC) curves. The correctness of the automatic thresholding is determined by a quality metric defined as the product of the proportions of true positives and true negatives (Venkatesh and Rosin 1995). Finally, it is shown how a combination of certain saliency measures can provide an improvement over their individual scores.

2 Saliency Measures

Although gradient magnitude is the measure traditionally used to discriminate between salient and non-salient edges various other possibilities have also been suggested. They could be used in place of gradient magnitude or in addition to it. Some are briefly described below.

2.1 Lifetime

This involves blurring the image over a range of scales, detecting edges at each scale, and tracking them across scale. The longer an edge persists before annihilation (i.e. its lifetime) the more likely it is to be significant. This approach follows from Marr's (Marr 1982) assertion that spatial coincidence of edges over scale correlates with real objects in the scene.

Tracking can be performed in a fine-to-coarse or coarse-to-fine manner. Bischof and Caelli (Bischof and Caelli 1988) measure an edge pixel's "stability" by tracking from fine through to coarse scales with a fixed window centered at the finest scale. Alternatively, in his edge focussing scheme Bergholm (Bergholm 1987) tracks from coarse to fine scales, and follows edge pixels as they spatially drift over scale. The spatial drift is limited by sampling scale-space sufficiently finely. Bergholm's scheme used the unthresholded coarse scale edge map to determine appropriate edges at the finest scale. Thus it involves tracking only those edges that exist at the coarsest scale. The weakness of this approach is that the resulting edge map depends on the coarsest scale of analysis,

which was manually chosen. In contrast, we wish to assign all edge pixels at the finest scale a lifetime value, and their relative values should be insensitive to the coarsest scale. Therefore, when tracking coarse-to-fine we also track edges that are created at all intermediate scales.

A slight modification can be made to these two approaches to combine edge lifetime and gradient magnitude. Rather than simply counting the number of scales an edge exists over, the gradient magnitudes at each scale are summed during the tracking process.

An alternative method for combining lifetime and gradient magnitude suggested by Griffin *et al.* (Griffin, Colchester, Robinson, and Hawkes 1992) is the product of the gradient magnitude $\|\nabla I\|$ with the angle of the iso-luminance curve through scale space $\arctan(\|\frac{\nabla I}{\nabla^2 I}\|)$. However, apart from an initial sharpening effect we have found that this measure is essentially the same as the unmodified gradient after non-maximal suppression. This is not unexpected since $\nabla^2 I$ equals zero close to edges (e.g. Marr and Hildreth's zero-crossing edge detector), and so $\arctan(\|\frac{\nabla I}{\nabla^2 I}\|)$ simply evaluates to $\frac{\pi}{2}$.

2.2 Wiggliness

The expectation is that noisy edges are less likely to be locally straight or smooth than significant edges (Rogowitz and Voss 1990). There are several possibilities for measuring edge wiggliness. We have experimented with k-regularity (Vasselle and Giraudon 1994) which is defined at each point p_i on the curve as:

$$r_{s,k}(i) = \frac{\|p_{i+ks} - p_i\|}{\sum_{j=1}^k \|p_{i+js} - p_{i+(j-1)s}\|}$$

and averaged over the curve, and the circular variance of orientation as defined in Gregson (Gregson 1993) or Mardia (Mardia 1972). Other possibilities include orientation coherence (Larré and Montseny 1994), complexity (Dubuc and Zucker 1994), and difference in curvature (Rosin and West 1995).

A drawback with such approaches is that they generally require some parameters specifying the scale of analysis, e.g. s and k for k-regularity, and the window size for orientation dispersion.

2.3 Width and Clutter

The spatial width of an edge may also be useful to discriminate between different edge types. As mentioned by Prager (Prager 1980), a combination of gradient magnitude and edge width can be utilised. Non-maximal suppression is performed on the raw edge map I , producing the image I_{nms} . The strength of each edge in I_{nms} is determined by integrating the two sets of monotonically decreasing gradient magnitudes in I perpendicular on both sides to the edge.

Hodgetts and Fallside (Hodgetts and Fallside 1992) describe a method for thresholding straight edges that incorporates width and length. Edgels with similar orientation are grouped, and edge support regions are then thresholded based on their size rather than gradient magnitude.

Rather than considering the width of the intensity discontinuity, the spatial extent on either side of the edge can be used. The larger these areas the further the closest adjacent edges are. The degree of isolation of an edge may correspond with its saliency since noisy edges will often be embedded in a surrounding clutter of other noisy edges.

This approach is implemented by Qian and Huang (Qian and Huang 1994) who suppress edges in noisy neighbourhoods. The “noisiness” measure is constructed at each edgel by counting the number of edgels within a square neighbourhood (whose size is a function of the scale of edge detection). For each connected edge curve the measure is then averaged over all its edgels.

We propose a different clutter measure that was previously applied to augment the distance transform (DT) with saliency measures (Rosin and West 1995). Its advantage is that it avoids the need for a parameter specifying the neighbourhood. The Voronoi diagrams of the edges are constructed using the standard DT. The neighbourhood of each edge curve is then specified by its Voronoi cell. The clutter measure is given by the Voronoi cell size divided by the length of the curve, providing a scale invariant measure approximately proportional to the average distance from each edgel to the closest edge from a different curve.

2.4 Projection onto Edge-Subspace

Frei and Chen (Frei and Chen 1977) suggested that the local 3×3 image window should be projected onto a set of nine orthogonal feature basis functions which can be divided into edge E (gradient and ripple) and non-edge (line, spot, and average) categories. Thresholding is then based on the angle of the window's projection onto the edge

subspace formed by four of the bases, which is proportional to the normalised dot product of the set of edge basis functions $B_{i \in E}$ and the 3×3 set of image values I ,

$$\frac{\sum_{i \in E} (I \cdot B_i)^2}{\sum_{k=1}^9 i_k^2}$$

where i_k is the k 'th image value in the window. This approach enables edges with low gradient magnitudes to be retained if their projection onto edge subspace is relatively high compared to their projection onto non-edge subspace. Conversely, edges with large gradient magnitudes can be rejected if their projection onto edge subspace is relatively low compared to their projection onto non-edge subspace.

2.5 Local Intensity and Contrast

Whereas edge thresholding is often performed globally over the image, it has been suggested that local thresholding may be appropriate. This can be achieved by locally normalising the image gradients, which can then be globally thresholded as before. For instance, Johnson considered normalising the gradient by dividing by the local average of the image intensity (Johnson 1990):

$$\frac{\|\nabla I\|}{\bar{I} + d} \quad (1)$$

where the constant d was included to reduce the effects of noise. The opposite approach was taken by Qian and Huang (Qian and Huang 1994) who *inhibit* edges in low intensity regions rather than boost their contrast.

In a similar manner, Osorio *et al.* (Osorio, Snyder, and Srinivasan 1987) effectively normalised the gradient, but divided by the local average of the gradient:

$$\frac{\|\nabla I\|}{\|\bar{\nabla I}\|}. \quad (2)$$

2.6 Phase Congruency

Morrone *et al.* (Morrone, Ross, Burr, and Owens 1986) have proposed that edges occur at points of maximum phase congruency. This can be calculated in a more convenient manner using the local energy function (Venkatesh and Owens 1990), which for a 1D signal is defined as $E(x) = \sqrt{F^2(x) + H^2(x)}$ where $F(x)$ and $H(x)$ are the Fourier and Hilbert transforms of the signal. Kovesi (Kovesi 1991) gave initial results using phase congruency as an edge saliency measure. These have been recently improved (Kovesi 1995) by: using wavelets to perform multi-scale analysis, better extension of the 1D technique to 2D, and adaptive noise estimation and suppression.

3 Automatic Thresholding

3.1 Gradient Histogram Analysis

Voorhees and Poggio (Voorhees and Poggio 1987) showed that if the image noise consists of additive Gaussian noise then the magnitude of the gradient of the image

$$\|\nabla I\| = \sqrt{I_x^2 + I_y^2}$$

has a Rayleigh distribution

$$R(\|\nabla I\|) = \frac{\|\nabla I\|}{\sigma^2} e^{-\frac{\|\nabla I\|^2}{2\sigma^2}}.$$

For an acceptable proportion of false edges (P_F) this allows a suitable threshold to be selected:

$$\tau = \sigma \sqrt{-2 \ln(P_F)}.$$

However, in practice the edge distribution in a real image is a combination of different sources of noise and significant features, complicating the identification of the non-noise components in the edge gradient magnitude histogram. Voorhees and Poggio's approach was to smooth the histogram and attempt to locate the peak. Unfortunately this requires various heuristically set parameters which diminishes the robustness of the method. Alternatively, Amodaj and Popovic (Amodaj and Popovic 1990) iteratively fit a Rayleigh function to the lower portion of the histogram. However, the fitting may not be robust since it depends on the portion of the histogram used as well as on the initial estimate.

Another approach to estimate the peak uses the relationship between the mode (σ) and mean (μ) of the Rayleigh function

$$\sigma = \sqrt{\frac{2}{\pi}}\mu \quad (3)$$

to estimate the parameters of the noise. Since the gradient magnitude histogram is contaminated by contributions from edge pixels Hancock and Kittler (Hancock and Kittler 1991) used an iterative method in which all gradient magnitudes greater than 5σ were rejected. Although they also mention preliminary applications of a robust estimation method, no details are given.

A simpler procedure is to model the noise by a Gaussian distribution. Canny (Canny 1986) estimates the mean from the bottom 80% of the histogram. Alternatively, Qian and Huang (Qian and Huang 1994) histogram equalise the gradient image. A histogram is then generated from the average gradient magnitudes calculated over each connected edge curve, from which the parameters of the Gaussian distribution are estimated. In a slightly different context, in order to diminish the effect of the tail, Kovesi (Kovesi 1995) suggests taking the exponential of the mean of the log-mapped histogram:

$$\hat{\mu} = \exp \left(\sum_{\forall i} H(i) \ln(i) \right)$$

where $H(i)$ is the histogram count of gradient i .

We propose to use the least median of squares (LMedS) from robust statistics (Rousseeuw and Leroy 1987):

$$\hat{\sigma} = \min_j \text{med}_i (H(i) - j)^2.$$

The LMedS of an uncontaminated Rayleigh function is close to the mode ($\approx 1.033\sigma$, see Appendix A), and we have found it to give good estimates even with substantial contamination.

For instance, the edges of the 256×256 image in Fig. 1a are shown in Fig. 1b, and their gradient magnitude histogram is given in Fig. 1c. The various estimates of the mode are shown in Fig. 1d. Particularly in small images the histogram may be noisy, which poses problems for peak finding procedures. The mode estimate based on the mean normalised as in (3) is substantially biased by the edge contribution. Although the iterative version of the mean provides an improvement, it is still outperformed by the normalised log-mapped mean and simple median. The best results are obtained using the LMedS. An iterative version of the LMedS estimator was implemented (repeatedly rejecting elements in the histogram $> 5\hat{\sigma}$), but it was found to give almost identical results to the non-iterative LMedS estimator. It can be seen that when thresholding with $P_F = 0.01$ using $\hat{\sigma}$ obtained from the iterative mean the edges are overthresholded (Fig. 1e). Thresholding using the LMedS estimate gives better results (Fig. 1f).

3.2 Connected Curve Analysis

Rather than threshold edge pixels independently some contextual information can be utilised. Koffi *et al.* (Koffi, Solaiman, and Mouchot 1994) consider the 3×3 window centred at each edgel. Canny (Canny 1986) popularised hysteresis which employs two thresholds T_U and T_L , where it was experimentally determined that $T_L \approx \frac{T_U}{2}$. Two sets of edgels E_U and E_L are determined, whose gradient magnitudes are greater than T_U and T_L respectively. The final edge map contains all edgels E_U and those edgels in E_L that are connected to an edgel from E_U by a path of E_L edgels. In the 1960's Sakai *et al.* (Sakai, Nagao, and Fujibayashi 1969) presented a similar approach involving three thresholds and several linking evaluation functions based on edge direction and gradient magnitude. In addition to Canny's hysteresis method, Ramesh *et al.* (Ramesh, Haralick, Zhang, Nadadur, and Thornton 1994) employ a curve length threshold.

Alternatively, Lowe (Lowe 1987) thresholded complete connected edge lists by their average gradient magnitude. To include a preference for long edges this was adapted by Rosin and West (Rosin and West 1989) to the summed gradient magnitude, so that weak edges would be retained if they were long enough. Also based on complete curves, Otte and Nagel (Otte and Nagel 1992) incorporate mean gradient magnitude, length l_k and orientation standard deviation s_k , and define the following thresholding criterion:

$$\frac{S_k^3}{l_k^2 \tau_1^3 \tau_2} \geq \frac{s_k}{\tau_3}$$

where S_k is the sum gradient magnitude of curve k , and τ_1 , τ_2 , and τ_3 are thresholds on the mean gradient magnitude, length, and orientation deviation respectively.

3.2.1 CLT Method

Whereas in the schemes of Lowe, Rosin and West, and Otte and Nagel the thresholds had to be set manually, an automatic threshold selection method was provided by Venkatesh and Rosin (Venkatesh and Rosin 1995). If it assumed that most of the edges arise from the noise population whose gradient magnitude is distributed with mean μ_{noise} and variance σ_{noise}^2 , then the edge lists of length n can be considered as n -sized samples of the noise population. The means of the samples will be the same as for the noise population μ_{noise} , but with variance $\frac{\sigma_{noise}^2}{n}$. From the Central Limit Theorem (CLT), for each n , the distribution of the means of the samples approaches a Gaussian with increasing n . Thus, the majority of noisy edge curves of length n will lie within $\mu_{noise} \pm 3\frac{\sigma_{noise}}{\sqrt{n}}$, which provides a suitable threshold. In practise, Venkatesh and Rosin (Venkatesh and Rosin 1995) devised a heuristically based scheme to select a linear decision rule in edge curve “length”-“average gradient magnitude” feature space. To ensure reliability in presence of contamination by significant (non-noise) edges this was based on the robust quantities of median and median absolute deviations (MAD) of the global set of edge curve lengths and average gradient magnitudes as shown in Fig. 2.

We now analyse more closely how this selected decision rule relates to feature space. It is not possible to determine the exact density of the points in feature space corresponding to noise since this depends on the distribution of lengths of noisy curves, which depends on the edge detector and edge linker. However, the linear decision rule was motivated by the observation that the cluster in feature space corresponding to noise is approximately triangular with uniform density. Under this assumption the median and MAD can be used to define the triangle boundaries. The median and MAD of the (e.g. X) co-ordinates of a right angled triangle of length x_T (reflected for convenience as shown in Fig. 17) are (see Appendix B):

$$x_{MED} = \frac{1}{\sqrt{2}}x_T \text{ and } x_{MAD} = \frac{1}{4\sqrt{2}}x_T.$$

Therefore, the diagonal triangle boundary is at $x_{MED} - \alpha x_{MAD} = 0$ which occurs at $\alpha = 4$. In practice it was found that better results were achieved with $\alpha > 5$, and the slope of the line was also halved (FIT2 in Fig. 2). This reflects the fact that the assumed uniform triangular density of noise is only an approximation.

Venkatesh and Rosin’s method is now extended to avoid the heuristic criterion. Instead of approximating the $3\frac{\sigma_{noise}}{\sqrt{n}}$ curve by a linear decision rule, σ_{noise} is estimated so that the actual curve is used directly instead. Each set of gradient magnitude means $\mathcal{G}^n = \{G_i^n, i = 1 \dots |\mathcal{G}^n|\}$ derived from the edge curves of length n has an approximately Gaussian distribution. Therefore, the standard deviation of the noise can be estimated as $\hat{\sigma}_{noise} = \sqrt{\text{Var}(\mathcal{G}^n)n}$. To improve robustness, the MAD is used in place of standard deviation. Moreover, we can combine the estimates provided by each set of gradient magnitude means \mathcal{G}^n . The standard error of $\hat{\sigma}_{noise}^2$ estimated from \mathcal{G}^n is $\sqrt{\frac{2}{n-1}\hat{\sigma}_{noise}^2}$ since the data is normally distributed. Therefore we weight each estimate by $\sqrt{|\mathcal{G}^n| - 1}$, getting:

$$\hat{\sigma}_{noise} = \frac{\sum_{\forall n} f \text{MAD}(\mathcal{G}^n) \sqrt{n} \sqrt{|\mathcal{G}^n| - 1}}{\sum_{\forall n} \sqrt{|\mathcal{G}^n| - 1}}$$

where

$$f = 1.4826 \left(1 + \frac{5}{|\mathcal{G}^n| - 1} \right)$$

is a Gaussian normalisation and finite-sample correction factor (Rousseeuw and Leroy 1987).

Fig. 3 gives an example of a feature space plot derived from the image in Fig. 1a. The gray line shows the median average gradient magnitude, and the linear and non-linear decision boundaries are marked in black. Each point corresponds to a curve, and the local estimates of $3\frac{\sigma_{noise}}{\sqrt{n}}$ are marked by crosses. Although both decision rules separate out the dense triangular wedge of noisy curves it can be seen that the non-linear rule eliminates more of the very short curves.

3.3 Other Thresholding Methods

Various other methods have been suggested for determining edge thresholds. Some are reviewed in Otte and Nagel (Otte and Nagel 1992) and Venkatesh and Rosin (Venkatesh and Rosin 1995), while further examples are given here. Many approaches use decision criteria explicitly based on the noise level (assumed to follow a zero mean Gaussian distribution) in the image (e.g. (Abdou and Pratt 1979; Brügelmann and Förstner 1992; Haddon 1988; Kittler 1983; Zuniga and Haralick 1988)), and use various schemes for noise estimation.

Kitchen and Rosenfeld (Kitchen and Rosenfeld 1981) defined a means for assessing different edge detectors using a combination of good continuation and thinness measures which are calculated within 3×3 windows. This

was extended to larger neighbourhoods by Haralick and Lee (Haralick and Lee 1990). Kitchen and Rosenfeld suggested that the value maximising the combined measure could be a suitable edge strength threshold. However, while the method worked on simple synthetic images, we found that it performed poorly on real images despite various attempts to remedy its shortcomings (Venkatesh and Rosin 1995).

Although the global gradient magnitude histogram is generally unimodal, Ibrahim and Aggoun (Ibrahim and Aggoun 1992) suggested analysing local windows in which the histogram would appear bimodal if a significant edge exists. Standard techniques can then be used to threshold the window. In areas containing few (or no) edges Ibrahim and Aggoun interpolate a threshold from nearby windows. However, the problem arises that it is necessary to robustly determine whether the window contains edges or not.¹ No details are given by Ibrahim and Aggoun of how to tackle this problem.

A different approach to thresholding is to perform relaxation labelling such that local edge magnitude characteristics are reinforced by the properties of adjacent pixels (Zucker, Hummel, and Rosenfeld 1977). An alternative labelling scheme is presented by Prager (Prager 1980) which avoids the difficulties of setting the many compatibility coefficients. Four different configurations of adjacent pixels labelled as edges or non-edges either increase, decrease, or leave unchanged the probability of the central pixel being an edge. Thus, during relaxation labelling, much of the initial gradient magnitude image converges towards a binary edge map. The remainder can be thresholded at the pixel probability $p = 0.5$ of being an edge.

4 Threshold Assessment

When discussing thresholding measures and methods some metric is necessary to assess their relative performance. Several related measures have been proposed for evaluating different edge detectors (Abdou and Pratt 1979). However, they are not totally appropriate here since they are concerned with factors such as edge displacement and multiple responses that are not relevant to threshold evaluation. Instead, we shall consider the edge thresholding process as a classification task: it must classify the output from the edge detector into true and false edges. Since perfect performance will not always be possible the thresholding should attempt to minimise the number of misclassifications (i.e. the true edges which are eliminated and the false edges which are retained by thresholding).

An aid for analysing the results of classification is to first construct a confusion matrix which contains the number of items assigned to each category relative to the actual category (Congalton 1991). The columns represent the reference data while the rows give the category assigned by the classifier. Thus element c_{ij} of the confusion matrix is the number of elements in category i that have been classified as category j . In our case there are just two categories: true edges, and false edges.

There are many methods that analyse the confusion matrix to quantify the accuracy of classification by a single figure of merit (*et al.* 1975). One common example for 2×2 tables is the cross product ratio

$$\alpha = \frac{c_{11}c_{22}}{c_{12}c_{21}}$$

Another wide-spread measure, for arbitrary sized matrices, is the Kappa coefficient of agreement (Cohen 1960) which is calculated as

$$\hat{\kappa} = \frac{\theta_1 - \theta_2}{1 - \theta_2}$$

where

$$\theta_1 = \frac{1}{N} \sum_i c_{ii}$$

specifies the proportion of correct classified elements, and

$$\theta_2 = \frac{1}{N^2} \sum_i \left(\sum_j c_{ij} \sum_j c_{ji} \right)$$

specifies the number of elements correctly classified by chance.

In this paper we use the following measure (Venkatesh and Rosin 1995)

$$\phi = \prod_i \frac{c_{ii}}{\sum_j c_{ji}}$$

¹This was also one of the problems that arose with a local window version (Venkatesh and Rosin 1995) of the Kitchen and Rosenfeld approach.

which is the product of the proportions of correctly classified real edges and false edges. Unlike the Kappa coefficient of agreement this measure has the desirable property that it is invariant under scaling of the column values. This avoids the problem that otherwise arises when some categories contain more items than others. Without multiplicative invariance the larger categories tend to dominate the measure unless the confusion matrix is normalised in some way (Congalton 1991). This is relevant in the context of edge thresholding since edge detection with little image smoothing typically produces many more false than true edges. Therefore, just using the percentage total classification accuracy for instance, would enable good scores to be achieved by simply setting such a high threshold that all edges are removed, so that the majority of potential edge pixels (i.e. non-edges) were correctly classified. Obviously this is not appropriate in the context of computer vision.

A second desirable property of ϕ is that at either extremes of threshold values the measure produces the value of zero since all true or false edges will have been incorrectly classified. Otherwise, the value reaches a maximum of one for the ideal case where all true and false edges have been correctly classified. In contrast, the cross product ratio does not have an upper bound since it will equal ∞ if any category is perfectly classified.

A limitation of all the above measures is that they are pixel based. By only analysing local information they ignore any structure properties of the data. Previously we found that this biased the measure against the complete curve analysis methods (Venkatesh and Rosin 1995) since they tend to occasionally include a spurious curve. Retaining such a complete curve involves misclassifying many edgels, substantially reducing the overall assessment score. However, the complete curve analysis methods produced perceptually better results, which are also better suited to further analysis, since curves enable more symbolic processing than individual pixels.

As well as generating a single figure of merit to assess a particular threshold it is also useful to analyse the trade-offs between the amount of true and false positives for all the different threshold values. This is done by generating ROC (receiver operating characteristic) curves (Egan 1975). In addition, the area under the ROC curve corresponds to the probability of a correct decision in a two-interval, forced choice task, and can be used as another assessment measure (Egan 1975; Hanley and McNeil 1982).

A prerequisite for generating the confusion matrix is *a priori* knowledge of the correct classification of the data. The most ready solution is to use synthetic images, which have the advantage that all the true edges are known exactly. The disadvantage is that the edge and noise characteristics are unlikely to be entirely representative of real images. On the other hand, ground truth edge maps of real images are not readily available, and if generated manually might be biased by the operator's preconceptions and higher-level knowledge.

Even given ground truth edges there is the problem of labelling the edge detector output since the edges will be imperfect as edge detectors are prone to mislocalisation, drop-out, and multiple responses. This problem could be circumvented for some saliency measures by calculating them at the ideal edge locations known from the ground truth. For other measures however this is not possible. For instance, calculating lifetime requires tracking edges over scale, and so edge detector output must be used. To cater for edge mislocalisation edgels within ± 1 pixel of a ground truth edgel are labelled as edges (true positives), all others being non-edges (false positives). Unfortunately, errors due to drop-out and multiple responses are still present, but appear to be negligible relative to the full edge set.

We use both types of data. The synthetic image is the one previously used in Venkatesh and Rosin (Venkatesh and Rosin 1995). It was created by randomly dropping onto the image a set of randomly generated polygons each filled with a random intensity; Gaussian noise (zero mean, $\sigma^2 = 200$) is then added to the image. For the real images, use is made of a set of reference edges recently generated for the RADIUS model-board imagery (Thornton, Nadadur, Ramesh, Liu, Zhang, Bedekar, and Haralick 1994).

5 Experiments

Fig. 4a shows the synthetic image pear. All the edges found by the Marr-Hildreth detector ($\sigma = 1.68$) are shown in Fig. 4b; true edges are coloured white, false edges are coloured black. The Marr-Hildreth edges were used to indicate the location of potential edges, and, in addition to gradient magnitude, the following edge salience measures were calculated at these points: summed gradient magnitude (Prager 1980), phase congruency (Kovesi 1995), edge projection (Frei and Chen 1977), orientation variance (Gregson 1993), k-regularity (Vasselle and Giraudon 1994), clutter (Rosin and West 1995), lifetime (fine \rightarrow coarse (Bischof and Caelli 1988) and coarse \rightarrow fine (Bergholm 1987)), and summed gradient lifetimes. The ROC curves are shown in Fig. 5. It can be seen that the gradient magnitude, summed gradient magnitude, gradient lifetime, and phase congruency measures perform very well, enabling most true edges to be detected whilst misclassifying few false edges. The edge projection method also performs well, but misclassifies more false edges, as does orientation variance. Due to the discretisation of their values (integers between 0 and 255) the k-regularity and clutter ROC curves are discontinuous. Neither are able to correctly classify more than 80% true edges without also misclassifying most false edges. The two

lifetime approaches produced more continuous results (105 images were used, evenly spaced in $\log(\sigma)$ space with $\sigma = [1.68, 16]$) but without incorporating gradient magnitude they are poor compared to the first five measures.

In the column headed ALONE, table 1 provides the optimal assessment values of the various saliency measures. These are derived by considering all possible thresholds (in the integer range [0, 255], and selecting the one maximising ϕ . This is intended to demonstrate the discrimination abilities of the measures without addressing the problem of automatic thresholding. Measures producing better ϕ values than gradient magnitude are printed bold. Both the summed gradient magnitude and fine \rightarrow coarse gradient lifetime methods outperform the gradient magnitude, by about 8% and 4% respectively,

In addition, the gradient normalisation methods given by (1) (Johnson 1990) and (2) (Osorio, Snyder, and Srinivasan 1987) were evaluated. These required choosing a neighbourhood size for averaging the image intensities or gradient magnitudes. As the averaging window size increases the normalised gradient approaches the unnormalised gradient. It was found that both methods gave consistently poorer results than the gradient magnitude, approaching its performance with increasing window size.

For comparison with the ϕ assessment, the area under the ROC curve is also given in the column headed A(ROC). Although this roughly follows the ϕ assessment, there are some discrepancies. In particular, the lifetime measures (not incorporating gradient magnitude) are given a relatively higher rating.

If the various saliency measures are not entirely correlated then it should be possible to combine them so as to produce a composite measure that outperforms any individual measure. The problem is how to combine the measures since they are produced by very different methods, have different ranges and distributions of values, and are consequently incommensurate. To test the possibility of improvement, again without considering the problem of threshold determination, we have experimented with several combination techniques. The first simply combines the gradient magnitude with another measure by forming the product of the pair. As shown in table 1 this enables the summed gradient magnitude, phase congruency, orientation variance, and fine \rightarrow coarse gradient lifetime to improve on gradient magnitude alone, but they are still inferior to summed gradient magnitude alone.

On examination of some of the significance measures it appeared that high values reliably indicated edges although other edges had low values easily confused with non-edges. This suggested the second approach for combining the same measure pairs as above. All possible thresholds for each of the same measure pairs were considered (i.e. up to 256^2 combinations) and a (i) MAX or (ii) MIN operation was performed, so that at each pixel an edge is output if either (i) both or (ii) at least one thresholded measure contain an edge respectively. Table 1 shows that this enables a further slight improvement over summed gradient magnitude alone by performing MAX(gradiant magnitude, summed gradient magnitude). Also, MAX or MIN combinations with phase congruency, orientation variance, and lifetime can improve gradient magnitude alone.

A different approach to measure combination is to use a training set and perform supervised learning. For instance, the table based combination of multiple experts technique by Huang and Suen (Huang and Suen 1995) would be appropriate except that for a set of continuous input values it has large memory requirements. Instead, we have used a multi-layer perceptron neural network employing the backpropagation learning algorithm to combine five diverse measures: gradient magnitude, edge projection, lifetime, k-regularity, and orientation variance. The network had five inputs, one hidden layer with twenty nodes, and two outputs, corresponding to the edge and non-edge classes. During classification the output node with the highest activation was chosen, thereby enabling the network to perform automatic thresholding as well as measure combination. A training set of 24000 training samples was made up from the synthetic and real images using edges from three scales ($\sigma = \{1, 1.68, 2\}$). Two approaches were taken to sample selection. Either the training samples could reflect the distribution of edge/non-edge pixels in the images, or they can be stratified to give each an equal number of samples to both classes. The two methods can have quite different results since there are so many more non-edges than edges in typical images. Unstratified sampling incorporates a bias to prefer non-edge labels, reflecting the underlying likelihoods. However, although this can improve overall classification accuracy, it is generally achieved by having a high threshold at the expense of the true edges, since this enables most non-edges to be correctly classified while the smaller number of edges are poorly classified. Since the ϕ measure operates on the proportions of correctly classified labels, it gives equal weighting to both edges and non-edges. Table 2 shows that stratified sampling gives better results than unstratified sampling. An improvement is made over gradient magnitude alone, and the results are comparable with the best of the MAX/MIN combinations. Moreover, whereas the results in table 1 were exhaustively optimised for all threshold combinations for the specific data set, the neural net was trained on a wider range of data.

Koffi *et al.* (Koffi, Solaiman, and Mouchot 1994) described a neural network approach to learning appropriate thresholds from gradient magnitude images. The 3×3 window centred at each edgel was input to the network, enabling some contextual information to be incorporated. We also experimented with including neighbouring pixels, but avoided the disadvantages of Koffi *et al.*'s scheme which requires increasing the quantity of the input data nine-fold, and necessitates the network separately learning edge fragments with different orientations and shapes. Since most edgelets will only have two significant adjacent edgelets, only the two edgelets with the strongest

responses from the set of 8 in the 3×3 neighbourhood are input in addition to the central pixel itself. Thus, only a three-fold increase in input data is required. Using the same five saliency measures as above, a second network with fifteen inputs, one hidden layer with fifty nodes, and two outputs was trained on 36000 training samples taken from the same images as before. Table 2 shows that incorporating this additional contextual information enables a further improvement.

Table 2 also shows the results of the other automatic thresholding methods, which are sectioned into local, semi-local, and global thresholding methods, applied to the gradient magnitude. Using the LMedS estimator of the Rayleigh distribution mode and thresholding at $P_F = 0.01$ produces a good ϕ value which is superior to using the iterative mean method for estimating the mode. Relaxation labelling with Prager's formulation (Prager 1980) also performs well. The methods that operate on complete curves give lower ϕ values. As noted before, this is partly a deficiency of point-wise assessment measures since structural information is not incorporated. For instance, visual inspection of the results produced by the Rayleigh method (Fig. 6a) compared to the non-linear CLT method (Fig. 6b) suggests that the curve based approach generates a more useful set of curves for subsequent analysis even though they receive a lower rating. Applying hysteresis after thresholding with the Rayleigh method was not successful since it caused the noisy edge fragments evident in Fig. 6a to grow, creating much unwanted clutter. It was found that the optimal result for the Rayleigh method with hysteresis was obtained with a much higher threshold found at $P_F = 10^{-5}$, and this produced a similar result to the method without hysteresis that was shown in Fig. 6a. Since the iterative mean Rayleigh method overthresholded the image, applying hysteresis afterwards does not have such a bad effect, resulting in a better score than the LMed Rayleigh method.

In addition, for comparison, two common image thresholding techniques are included: Kittler and Illingworth's minimum error thresholding (Kittler and Illingworth 1986) and Tsai's moment preserving thresholding (Tsai 1985). It can be seen that the minimum error thresholding performed very poorly probably due to its modeling of the populations by Gaussian distributions. The moment preserving thresholding was much better, but is still outperformed by all the other pixel based edge thresholding methods.

The optimal assessment values (found by exhaustive search) for thresholding global average and total curve gradient magnitudes are given for reference. Compared with them, the Central Limit Theorem methods achieve near optimal results. Experimenting with the non-linear CLT method, it was found that thresholding at $2\frac{\sigma_{noise}}{\sqrt{n}}$ gave slightly better results than at $3\frac{\sigma_{noise}}{\sqrt{n}}$.

Since table 1 shows that the summed gradient magnitude can potentially outperform gradient magnitude the automatic thresholding was also performed on the summed gradient magnitude. Table 3 shows that the thresholded results are improved except for the linear CLT method. The problem with the latter technique is that due to the more concentrated distribution of values the MAD of the summed gradient magnitude yielded zero, resulting in a poor decision rule.

Fig. 7a shows the 1314×1044 image J25 of a model board from the RADIUS project. This provides a more realistic image than the synthetic pear, and also has ground truth available (Fig. 7b). A portion of the output of the Marr-Hildreth edge detector at $\sigma = 1$ with the edges labelled as true or false edges is shown in Fig. 7c. The inaccuracies due to edge mislocation can be seen, represented mainly by missing true edge labels. Due to computational limitations the phase congruency was not applied to this larger image. The ROCs of the remaining saliency measures are shown in Fig. 8. It can be seen that for a real image, as compared to the previous synthetic example, the overall standard of discrimination is much poorer. Table 4 shows that summed gradient magnitude, edge projection, and fine \rightarrow coarse gradient lifetime perform better than gradient magnitude, and that with the exception of clutter all the measures can be improved by combining them with the gradient magnitude using the MAX operation.

The same analysis was performed for edges extracted by the Marr-Hildreth edge detector at a coarser scale $\sigma = 2$, which removes some noise and also reduces edge fragmentation. Again, on their own, edge projection produces the best ϕ , and fine \rightarrow coarse gradient lifetime also outperforms gradient magnitude, and most measures can be improved by combining them with the gradient magnitude using the MAX operation.

The results of the automatic thresholding methods are shown in table 6. Of the local thresholding techniques the neural network with stratified training data and incorporating neighbourhood information performs best. The Rayleigh approach does reasonably well. Again the minimum error thresholding performed very poorly, and the moment preserving thresholding better. However, this time the latter results are relatively poor compared to the other local methods. The performance levels of the semi-local methods are on par with the pear example. The non-linear CLT approach does rather poorly on the $\sigma = 1$ example although the linear method does adequately. Both perform better at $\sigma = 2$ due to the improved connectivity. Examining the feature space plot in Fig. 10 produced for $\sigma = 1$ it can be seen that the non-linear decision rule produces a higher threshold than the linear decision rule. When analysing the longer curves tend the noise tends to be overestimated. Despite the lower weighting given to these high estimates due to the smaller number of long curves, the final noise estimate is still too high.

To demonstrate some of the thresholding methods, Fig. 9 shows the results of applying the neural network using neighbourhood, and the linear CLT method. Although the former local method receives a better rating, the latter method appears superior since the edges are not fragmented – the neural network produces 4056 edge curves whereas the CLT method only produces 839. However, the CLT method has retained some undesirable noise (e.g. within the roofs of some of the buildings at the bottom of the image).

Fig. 11a shows the 1st principal component of a 256×256 portion of a Landsat TM image taken in Portugal. The Canny (Canny 1986) edge detector ($\sigma = 1$) was independently applied to the six non-thermal bands, the results ORED together, and nonmaximal suppression applied. The other saliency measures were also calculated by ORing the results obtained from the individual bands. The most significant edges were determined manually to provide an approximate reference set (Fig. 11b).

The ROCs of the saliency measures are shown in Fig. 12. Again, compared to the synthetic example, the standard of discrimination is poor. Orientation variance, k-regularity, and clutter perform particularly badly. From table 8 it can be seen that on its own only fine \rightarrow coarse gradient lifetime outperforms gradient magnitude. Combination with the gradient magnitude using the product or MAX operations enables some further improvements.

This image is more demanding than the previous ones since many field boundaries have weak edge responses, although they are still clearly visible on viewing a false colour composite made up from bands 3, 4, and 5. Whereas in the earlier examples the Rayleigh distribution approach produced results whose assessment values were comparable with the other methods, here they are much poorer, as seen in table 9. A lower threshold must be used to retain the real edges since many of them are so weak. Without applying hysteresis this requires setting P_F as high as 0.4 to achieve optimal results (i.e. allowing 40% false positives). When using hysteresis the optimal setting is $P_F = 0.05$ (Fig. 13a). In contrast to the previous examples, using hysteresis enables improvements on the pointwise Rayleigh approach.

Again, the moment preserving thresholding produces acceptable results but is outperformed by all the other pixel based edge thresholding methods. Relaxation labelling and the CLT thresholding schemes perform relatively poorly. The feature space plot in Fig. 14 shows that the non-linear CLT method overestimates the noise level, causing the edge map to be overthresholded. The linear CLT method (Fig. 13b) does better than the non-linear CLT method but still misses many of the true edges.

Since table 8 shows that the potentially best saliency measure was fine \rightarrow coarse gradient lifetime the thresholding methods were applied to it, giving in some cases substantial improvements (table 10). This demonstrated in Fig. 15 which shows the results of thresholding the gradient lifetime using the Rayleigh and linear CLT methods.

A final qualitative example is given to demonstrate more realistically the effects of noise. A simple indoor scene was photographed with different aperture settings and the same shutter speeds. Increasing the F stop effectively reduces the signal to noise ratio. Fig. 16a shows the 512×512 image with F/1.3. All the edges obtained with the Marr-Hildreth detector at $\sigma = 2$ for aperture settings F/1.3 and F/6 are shown in Figs. 16b&c. The non-linear CLT method gave slightly better results than the linear CLT method, and the results of thresholding are shown in Figs. 16d&e. Despite the substantial amount of noise evident at F/6 the curve based thresholding has done well. In comparison, although the Rayleigh-based approach successfully locates the mode from the histogram, the output is fragmented without applying hysteresis (Fig. 16f), and contains spurious clutter when hysteresis is applied (Fig. 16g).

6 Conclusions

A number of edge saliency measures have been described which offer alternative means of discriminating between the true and false edge responses produced by edge detectors. We have tested and compared the following measures: summed gradient magnitude (Prager 1980), phase congruency (Kovesi 1995), edge projection (Frei and Chen 1977), orientation variance (Gregson 1993), k-regularity (Vasselle and Giraudon 1994), clutter (Rosin and West 1995), lifetime (fine \rightarrow coarse (Bischof and Caelli 1988) and coarse \rightarrow fine (Bergholm 1987)), and summed gradient lifetimes. Several different image types were used; each had associated ground truth, which allowed a quantitative assessment of the relative merits of the measures. The traditional gradient magnitude measure performed relatively well on all the images, although on some images the summed gradient magnitude, edge projection, and summed gradient lifetime measures produced better results. There was no single measure that always gave the best results, but the fine \rightarrow coarse summed gradient lifetime was consistently better than the gradient magnitude.

We experimented with several ways of combining the saliency measures, and found that improvements over the individual measures can often be obtained by forming the product or by applying a MAX operation. Alternatively, we showed that a neural network could be used to automatically learn combination rules as well as edge thresholds. (The concept of learning suitable thresholds from training data has also been considered by Koffi *et al.* (Koffi,

Solaiman, and Mouchot 1994) and Ramesh *et al.* (Ramesh, Haralick, Zhang, Nadadur, and Thornton 1994).) Although good results were obtained for the set of test images considered here, the value of the approach depends on how representative the training data is of unseen images as well as the generalisation ability of the network.

Several thresholding schemes were considered, and can be classified as: local (pixel based), semi-local (restricted neighbourhood), and global (complete curve). In the local category the Rayleigh distribution method was considered, and the LMedS mode finding method was proposed. It was shown that for an uncontaminated Rayleigh distribution it produces a close estimate of the mode. In practise we have found it more reliable than the iterative mean estimation technique, producing better results. Applying hysteresis after the Rayleigh method resulted in underthresholding the images. Since the iterative mean method tended to produce higher thresholds than the LMedS method, its results were less degraded by hysteresis. Tsai's (Tsai 1985) moment preserving image thresholding technique was also tested, and found to give acceptable results, but was outperformed by the Rayleigh method.

The global thresholding method of Venkatesh and Rosin (Venkatesh and Rosin 1995) was analysed to explain the rationale behind the heuristic decision boundary estimation method. A method for determining a more accurate non-linear decision boundary was suggested. It gave better results than the linear decision boundary for several of the images but did not perform so well on others. This occurs because the linear approach is more robust since it uses global statistics of the distribution of curves in feature space, whereas the non-linear approach uses more local statistics. Examining the feature space plots suggested that the non-linear method tended to overestimate the noise level on "difficult" images, and therefore overthresholded the edge maps. Possibly the noise estimation procedure could be improved by performing a weighted median of the individual noise estimates rather than a weighted average, although this would require arbitrarily quantizing the weighting factors. A limitation of the global curve thresholding methods is their dependence on reasonable connectivity which is required so that edge curves can be extracted from the image. To this effect it would be beneficial to use some of the recently developed more sophisticated edge linkers (e.g. (Fischler 1994; Rothwell, Mundy, Hoffman, and Nguyen 1994; Wang and Binford 1994)).

Acknowledgments

I would like to thank V. Ramesh, D. Nadadur, and Prof. R. Haralick for providing the RADIUS test imagery and ground truth, Yannis Kanellopoulos for providing the neural network, and Peter Kovesi for providing the phase congruency code.

References

- Abdou, I. and W. Pratt (1979). Qualitative design and evaluation of enhancement/thresholding edge detectors. *Proc. IEEE* 67, 753–763.
- Amodaj, N. and M. Popovic (1990). Adaptive procedure for threshold selection in directional derivative edge detectors. In *Proc. SPIE Applications of Digital Image Processing XIII* Vol. 1349, pp. 128–138.
- Bergholm, F. (1987). Edge focusing. *IEEE Trans. PAMI* 9, 726–741.
- Bischof, W. and T. Caelli (1988). Parsing scale-space and spatial stability analysis. *Computer Vision, Graphics and Image Processing* 42, 192–205.
- Brügelmann, R. and W. Förstner (1992). Noise estimation for color edge extraction. In W. Förstner and S. Winter (Eds.), *Robust Computer Vision*, pp. 90–106. Wichmann.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. PAMI* 8, 679–698.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37–40.
- Congalton, R. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of the Environment* 37, 35–46.
- Dubuc, B. and S. Zucker (1994). Curve-like sets, normal complexity, and representation. In *Int. Conf. Pattern Recognition*, pp. 216–221.
- Egan, J. (1975). *Signal Detection Theory and ROC Analysis*. Academic Press.
- et al., Y. B. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT.
- Fischler, M. (1994). The perception of linear structure: a generic linker. In *ARPA Image Understanding Workshop*.

- Frei, W. and C. Chen (1977). Fast boundary detection: a generalization and a new algorithm. *IEEE Comp.* 26, 988–999.
- Gregson, P. (1993). Using angular dispersion of gradient direction for detecting edge ribbons. *IEEE Trans. PAMI* 15, 682–696.
- Griffin, L., A. Colchester, G. Robinson, and D. Hawkes (1992). Structure-sensitive scale and the hierarchical segmentation of grey-level images. In *Proc. SPIE Visualization in Biomedical Computing*, Vol. 1808.
- Haddon, J. (1988). Generalized threshold selection for edge detection. *Pattern Recognition* 21, 195–203.
- Hancock, E. and J. Kittler (1991). Adaptive estimation of hysteresis thresholds. In *Proc. CVPR*, pp. 196–201.
- Hanley, J. and B. McNeil (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Diagnostic Radiology* 143, 29–36.
- Haralick, R. and J. Lee (1990). Context dependent edge detection and evaluation. *Pattern Recognition* 23, 1–19.
- Hodgetts, M. and F. Fallside (1992). Extracting straight edges in images without conventional thresholding. In *4th Int. Conf. on Image Processing and Applic.*, pp. 307–310.
- Huang, T. and C. Suen (1995). Combination of multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Trans. PAMI* 17, 90–94.
- Ibrahim, M. and A. Aggoun (1992). DPCM edge detection using local histogram. In *4th Int. Conf. on Image Processing and Applic.*, pp. 542–545.
- Johnson, R. (1990). Contrast based edge detection. *Pattern Recognition* 23, 311–318.
- Kitchen, L. and A. Rosenfeld (1981). Edge evaluation using local edge coherence. *IEEE Trans. on Sys., Man, and Cybernetics* 11, 597–605.
- Kittler, J. (1983). Statistical image processing. In O. Braddick (Ed.), *Physical and Biological Processing of Images*, pp. 232–242. Springer-Verlag.
- Kittler, J. and J. Illingworth (1986). Minimum error thresholding. *Pattern Recognition* 19, 41–47.
- Koffi, R., B. Solaiman, and M. Mouchot (1994). Remote sensing images segmentation by Deriche's filter and neural network. In *Proc. SPIE Image and Signal Processing for Remote Sensing*, Vol. 2315.
- Kovesi, P. (1991). A dimensionless measure of edge significance. In *Proc. Digital Image Computing: Techniques and Applications*, pp. 281–288.
- Kovesi, P. (1995). Image features from phase congruency. Technical Report 95/4, Dept. Comp. Sci., Univ. Western Australia.
- Larré, A. and E. Montseny (1994). A step edge detector algorithm based on symbolic analysis. In *Int. Conf. Pattern Recognition*, Volume A, pp. 6–10.
- Lowe, D. (1987). Three-dimensional object recognition from single two-dimensional views. *AI* 31, 355–395.
- Mardia, K. (1972). *Statistics of Directional Data*. Academic Press.
- Marr, D. (1982). *Computer Vision*. Freeman.
- Morrone, M., J. Ross, D. Burr, and R. Owens (1986). Mach bands are phase dependent. *Nature* 324(6094), 250–253.
- Osorio, D., A. Snyder, and M. Srinivasan (1987). Bi-partitioning and boundary detection in natural scenes. *Spatial Vision* 2, 191–198.
- Otte, M. and H. Nagel (1992). Extraction of line drawings from gray value images by non-local analysis of edge element structures. In *Europ. Conf. Computer Vision*, pp. 687–695.
- Prager, J. (1980). Extracting and labelling boundary segments in natural scenes. *IEEE Trans. PAMI* 2, 16–27.
- Qian, R. and T. Huang (1994). A two-dimensional edge detection scheme for general visual processing. In *ARPA Image Understanding Workshop*.
- Ramesh, V., R. Haralick, X. Zhang, D. Nadadur, and K. Thornton (1994). Automatic selection of tuning parameters for feature extraction sequences. In *Proc. CVPR*, pp. 672–677.
- Rogowitz, B. and R. Voss (1990). Shape perception and low-dimension fractal boundary contours. In *Proc. SPIE Human Vision and Electronic Imaging: Models, Methods, and Applications*, Vol. 1249, pp. 387–394.
- Rosin, P. and G. West (1989). Segmentation of edges into lines and arcs. *Image and Vision Computing* 7, 109–114.

- Rosin, P. and G. West (1995). Salience distance transforms. *Graphical Models and Image Processing* 57, 483–521.
- Rothwell, C., J. Mundy, B. Hoffman, and V. Nguyen (1994). Driving vision by topology. Technical Report 244, INRIA.
- Rousseeuw, P. and A. Leroy (1987). *Robust Regression and Outlier Detection*. Wiley.
- Sakai, T., M. Nagao, and S. Fujibayashi (1969). Line extraction and pattern detection in a photograph. *Pattern Recognition* 1, 233–248.
- Thornton, K., D. Nadadur, V. Ramesh, X. Liu, X. Zhang, A. Bedekar, and R. Haralick (1994). Groundtruthing the RADIUS model-board imagery. In *ARPA Image Understanding Workshop*.
- Tsai, W. (1985). Moment-preserving thresholding: a new approach. *Computer Vision, Graphics and Image Processing* 29, 377–393.
- Vasselle, B. and G. Giraudon (1994). A multiscale regularity measure as a geometric criterion for image segmentation. *Machine Vision and Applications* 7, 229–236.
- Venkatesh, S. and R. Owens (1990). On the classification of image features. *Pattern Recognition Letters* 11, 339–349.
- Venkatesh, S. and P. Rosin (1995). Dynamic threshold determination by local and global edge evaluation. *Computer Vision, Graphics and Image Processing* 57(2), 146–160.
- Voorhees, H. and T. Poggio (1987). Detecting textons and texture boundaries in natural images. In *Proc. ICCV*, pp. 250–258.
- Wang, S. and T. Binford (1994). Model-based edgel aggregation. In *ARPA Image Understanding Workshop*.
- Zucker, S., R. Hummel, and A. Rosenfeld (1977). An application of relaxation labelling to line and curve enhancement. *IEEE Comp.* 26, 394–403.
- Zuniga, O. and R. Haralick (1988). Gradient threshold selection using the facet model. *Pattern Recognition* 21, 493–503.

Appendix A - Least Median of Squares Estimate of a Rayleigh Distribution

To determine the Least Median of Squares (LMedS) estimate of a Rayleigh distribution we use the formulation (Rousseeuw and Leroy 1987) that for a 1D signal $X = \{x_1 \dots x_n\}$ which has been sorted so that $x_i \leq x_j; \forall i < j$ the LMedS estimate is equivalent to finding i to minimise $x_{i+n/2} - x_i$, and the estimate is given as $\frac{x_{i+n/2} + x_i}{2}$. Thus, for the Rayleigh distribution $R(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ we need to find the range $x = [a, b]$ containing half of the density function, i.e.

$$\int_a^b R(x) dx = e^{\frac{-a^2}{2\sigma^2}} - e^{\frac{-b^2}{2\sigma^2}} = \frac{1}{2} \int_0^\infty R(x) dx = \frac{1}{2}. \quad (4)$$

Solving (4) for b gives

$$b = \sigma \left[-2 \ln \left(e^{\frac{-a^2}{2\sigma^2}} - \frac{1}{2} \right) \right]^{\frac{1}{2}}. \quad (5)$$

Since we want to minimise the range of x we differentiate $b - a$ after first substituting (5), giving

$$\frac{d(b-a)}{dx} = \frac{ae^{\frac{-a^2}{2\sigma^2}}}{\sigma \left[-2 \ln \left(e^{\frac{-a^2}{2\sigma^2}} - \frac{1}{2} \right) \right]^{\frac{1}{2}} \left[e^{\frac{-a^2}{2\sigma^2}} - \frac{1}{2} \right]} - 1. \quad (6)$$

Setting (6) to zero, and substituting

$$A = \frac{a}{\sigma} \quad (7)$$

we need to solve the following equation for A

$$\frac{Ae^{\frac{-A^2}{2}}}{\left[-2 \ln \left(e^{\frac{-A^2}{2}} - \frac{1}{2} \right) \right]^{\frac{1}{2}} \left[e^{\frac{-A^2}{2}} - \frac{1}{2} \right]} = 1.$$

Since this is difficult to perform analytically it was solved numerically giving

$$A = 0.5941278. \quad (8)$$

Substituting (8) and (7) into (5) we find that the LMedS estimate equals

$$\frac{a+b}{2} = 1.0333066\sigma.$$

Appendix B - Median and Median Absolute Deviation of a Right Angled Triangle

For convenience, we will consider the triangle to be translated and reflected as shown in Fig. 17. The bounding diagonal line can now be expressed simply as $y = rx$. The area of the section of the triangle from the origin to x_p is $A_p = \frac{r}{2}x_p^2$. If the total length of the triangle is x_T then the median of the X axis values is the point x_{MED} such that $A_T = 2A_{MED}$, which is given at $x_{MED} = \frac{1}{\sqrt{2}}x_T$. For a reflected triangle with the horizontal side along the Y axis (such as those generated by the feature space plots in section 3.2.1)

$$x_{MED} = \left(1 - \frac{1}{\sqrt{2}}\right)x_T.$$

To determine the median absolute deviation we consider the area of the triangle distant from x_{MED} . Again we use the translated and reflected triangle as shown in Fig. 18. Points on either side of x_{MED} need to be considered, this can be visualised as reflecting the area to the right of x_{MED} about x_{MED} (so that x_T goes to x_b , and then analysing only the summed areas from 0 to x_{MED}):

$$S_p = \begin{cases} \frac{r}{2}x_p^2 & \text{if } x_p < x_b \\ \frac{r}{2}x_b^2 + 2y_a(x_p - x_b) & \text{otherwise} \end{cases} \quad (9)$$

$$(10)$$

where $x_b = (\sqrt{2} - 1)x_T$ and $y_a = \frac{r}{\sqrt{2}}x_T$. We wish to find the point x_M at which

$$S_T = 2S_M. \quad (11)$$

We solve (11) for x_M by trying both expressions for S_M given in (9) & (10) and checking their consistency with the conditions. Using (9) gives $x_M = \frac{1}{\sqrt{2}}x_T$ while (10) gives $x_M = \frac{3}{4\sqrt{2}}x_T$. Since $\frac{1}{\sqrt{2}}x_T < x_b$ and $\frac{3}{4\sqrt{2}}x_T \geq x_b$, only the second expression holds, and $x_M = \frac{3}{4\sqrt{2}}x_T$. Finally, the median absolute deviation is given by

$$x_{MAD} = x_{MED} - x_M = \frac{1}{4\sqrt{2}}x_T.$$

Note that both the median and median absolute deviation are independent on the slope of the triangle. The same reasoning holds to find the median and median absolute deviation of the Y values of the triangle.

Table 1: Optimal assessment values for saliency measures on `pear` image

SALIENCY MEASURE	OPTIMAL ASSESSMENT ϕ				A(ROC)
	ALONE	PRODUCT	MAX	MIN	
gradient magnitude	0.823	—	—	—	0.951
summed gradient magnitude	0.888	0.881	0.890	0.823	0.973
phase congruency	0.821	0.838	0.840	0.823	0.949
edge projection	0.748	0.822	0.823	0.823	0.937
orientation variance	0.732	0.858	0.867	0.823	0.928
k-regularity	0.614	0.678	0.823	0.823	0.777
clutter	0.545	0.671	0.823	0.823	0.734
lifetime (fine \rightarrow coarse)	0.547	0.812	0.826	0.826	0.810
lifetime (coarse \rightarrow fine)	0.537	0.745	0.824	0.823	0.795
gradient lifetime (fine \rightarrow coarse)	0.858	0.867	0.870	0.823	0.956
gradient lifetime (coarse \rightarrow fine)	0.769	0.810	0.839	0.823	0.919

Table 2: Assessment values for thresholding methods on gradient magnitude of `pear` image

THRESHOLDING METHOD	ASSESSMENT ϕ
minimum error	0.006
moment preserving	0.710
neural network (unstratified)	0.785
(stratified)	0.889
using neighbourhood (stratified)	0.892
Rayleigh model (LMedS)	0.817
(mean)	0.783
(LMedS) with hysteresis	0.635
(mean) with hysteresis	0.747
relaxation labelling	0.863
<i>average curve magnitude</i>	0.694
<i>total curve magnitude</i>	0.673
CLT approach – linear	0.686
non-linear ($f = 2$)	0.695
non-linear ($f = 3$)	0.689

Table 3: Assessment values for thresholding methods on summed gradient magnitude of `pear` image

THRESHOLDING METHOD	ASSESSMENT ϕ
Rayleigh model (LMedS)	0.872
(mean)	0.888
(LMedS) with hysteresis	0.740
(mean) with hysteresis	0.742
CLT approach – linear	0.296
non-linear ($f = 2$)	0.713

Table 4: Optimal assessment values for saliency measures on J25 image; $\sigma = 1$

SALIENCY MEASURE	OPTIMAL ASSESSMENT ϕ			A(ROC)
	ALONE	MAX	MIN	
gradient magnitude	0.418	—	—	0.690
summed gradient magnitude	0.435	0.442	0.418	0.708
edge projection	0.442	0.451	0.418	0.712
orientation variance	0.338	0.435	0.418	0.609
k-regularity	0.403	0.437	0.418	0.649
clutter	0.337	0.418	0.418	0.593
lifetime (fine \rightarrow coarse)	0.325	0.422	0.418	0.604
lifetime (coarse \rightarrow fine)	0.346	0.423	0.418	0.624
gradient lifetime (fine \rightarrow coarse)	0.432	0.443	0.418	0.701
gradient lifetime (coarse \rightarrow fine)	0.402	0.429	0.418	0.670

Table 5: Optimal assessment values for saliency measures on J25 image; $\sigma = 2$

SALIENCY MEASURE	OPTIMAL ASSESSMENT ϕ			A(ROC)
	ALONE	MAX	MIN	
gradient magnitude	0.475	—	—	0.746
summed gradient magnitude	0.464	0.479	0.475	0.741
edge projection	0.480	0.491	0.475	0.754
orientation variance	0.370	0.479	0.475	0.650
k-regularity	0.459	0.503	0.475	0.701
clutter	0.329	0.475	0.475	0.598
lifetime (fine \rightarrow coarse)	0.345	0.475	0.475	0.605
lifetime (coarse \rightarrow fine)	0.332	0.476	0.475	0.595
gradient lifetime (fine \rightarrow coarse)	0.476	0.484	0.475	0.747
gradient lifetime (coarse \rightarrow fine)	0.454	0.481	0.475	0.726

Table 6: Assessment values for thresholding methods on gradient magnitude of J25 image

THRESHOLDING METHOD	ASSESSMENT ϕ	
	$\sigma = 1$	$\sigma = 2$
minimum error	0.026	0.025
moment preserving	0.296	0.398
neural network (unstratified)	0.230	0.242
(stratified)	0.455	0.508
using neighbourhood (stratified)	0.467	0.534
Rayleigh model (LMedS)	0.413	0.467
(mean)	0.303	0.277
(LMedS) with hysteresis	0.351	0.430
(mean) with hysteresis	0.417	0.459
relaxation labelling	0.419	0.369
<i>average curve magnitude</i>	0.399	0.471
<i>total curve magnitude</i>	0.406	0.460
CLT approach – linear	0.401	0.468
non-linear ($f = 2$)	0.296	0.461
non-linear ($f = 3$)	0.251	0.424

Table 7: Assessment values for thresholding methods on edge projection of J25 image

THRESHOLDING METHOD	ASSESSMENT ϕ	
	$\sigma = 1$	$\sigma = 2$
Rayleigh model (LMedS)	0.439	0.466
(mean)	0.327	0.273
(LMedS) with hysteresis	0.393	0.441
(mean) with hysteresis	0.439	0.466
CLT approach – linear	0.407	0.462
non-linear ($f = 2$)	0.278	0.413

Table 8: Optimal assessment values for saliency measures on sub2 image

SALIENCY MEASURE	OPTIMAL ASSESSMENT ϕ				A(ROC)
	ALONE	PRODUCT	MAX	MIN	
gradient magnitude	0.607	—	—	—	0.858
summed gradient magnitude	0.536	0.584	0.608	0.607	0.807
phase congruency	0.577	0.607	0.607	0.607	0.835
edge projection	0.522	0.592	0.607	0.607	0.794
orientation variance	0.338	0.608	0.607	0.607	0.606
k-regularity	0.434	0.519	0.607	0.607	0.631
clutter	0.248	0.460	0.607	0.607	0.452
lifetime (fine \rightarrow coarse)	0.459	0.586	0.608	0.607	0.739
lifetime (coarse \rightarrow fine)	0.450	0.580	0.608	0.607	0.730
gradient lifetime (fine \rightarrow coarse)	0.629	0.634	0.629	0.607	0.878
gradient lifetime (coarse \rightarrow fine)	0.581	0.604	0.612	0.607	0.843

Table 9: Assessment values for thresholding methods on gradient magnitude of sub2 image

THRESHOLDING METHOD	ASSESSMENT ϕ
minimum error	0.003
moment preserving	0.388
neural network (unstratified)	0.451
(stratified)	0.456
using neighbourhood (stratified)	0.479
Rayleigh model (LMedS) $P_F = 0.01$	0.265
(LMedS) $P_F = 0.05$	0.378
(LMedS) $P_F = 0.4$	0.607
(mean) $P_F = 0.01$	0.125
(mean) $P_F = 0.05$	0.195
(mean) $P_F = 0.4$	0.490
(LMedS) with hysteresis $P_F = 0.01$	0.589
(LMedS) with hysteresis $P_F = 0.05$	0.612
(LMedS) with hysteresis $P_F = 0.4$	0.441
(mean) with hysteresis $P_F = 0.01$	0.437
(mean) with hysteresis $P_F = 0.05$	0.523
(mean) with hysteresis $P_F = 0.4$	0.583
relaxation labelling	0.324
average curve magnitude	0.508
total curve magnitude	0.586
CLT approach – linear	0.397
non-linear ($f = 2$)	0.200
non-linear ($f = 3$)	0.111

Table 10: Assessment values for thresholding methods on fine → coarse gradient lifetime of sub2 image

THRESHOLDING METHOD	ASSESSMENT ϕ
Rayleigh model (LMedS) $P_F = 0.05$	0.540
(mean) $P_F = 0.05$	0.326
(LMedS) with hysteresis $P_F = 0.05$	0.631
(mean) with hysteresis $P_F = 0.05$	0.570
CLT approach – linear	0.495
non-linear ($f = 2$)	0.268

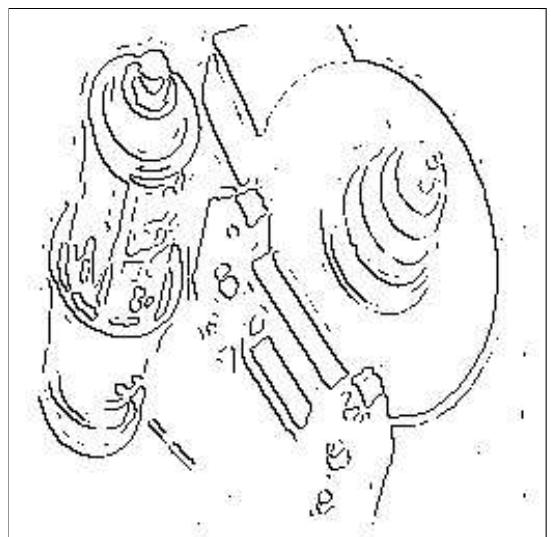
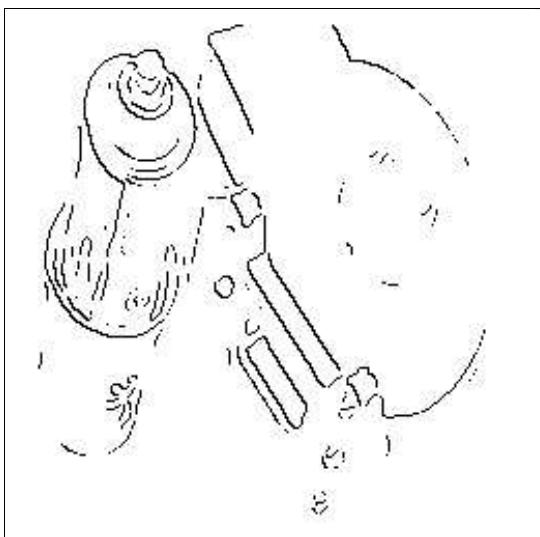
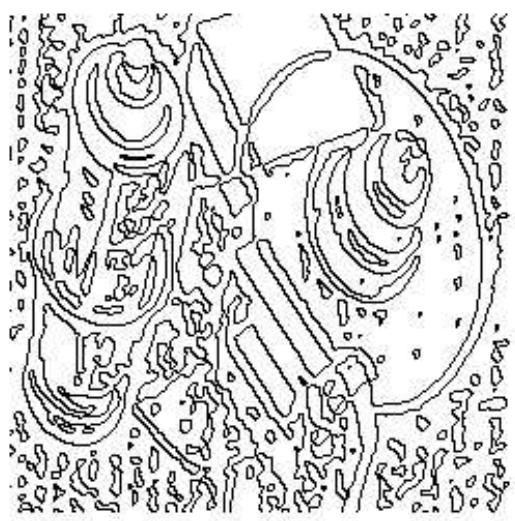


Figure 1: Thresholding using the gradient histogram; (a) can image; (b) all edges; (c) gradient magnitude histogram; (d) portion of histogram with estimated modes; (e) thresholded edges using iterative mean method; (f) thresholded edges using LMedS method

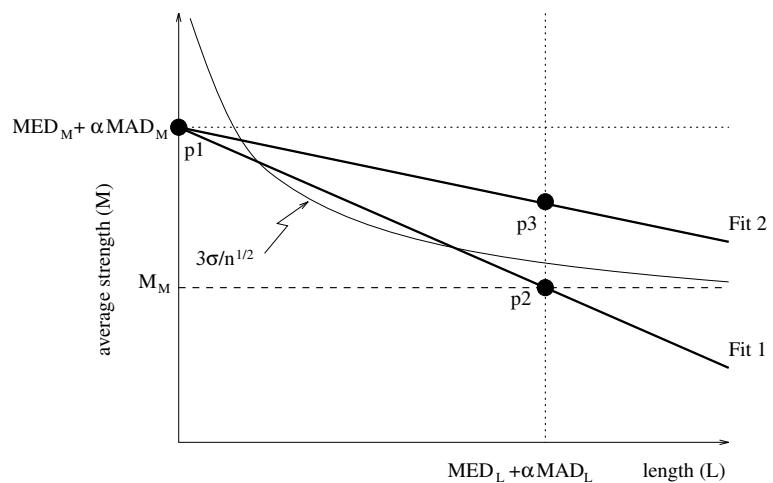


Figure 2: Linear approximation of noise boundary in feature space

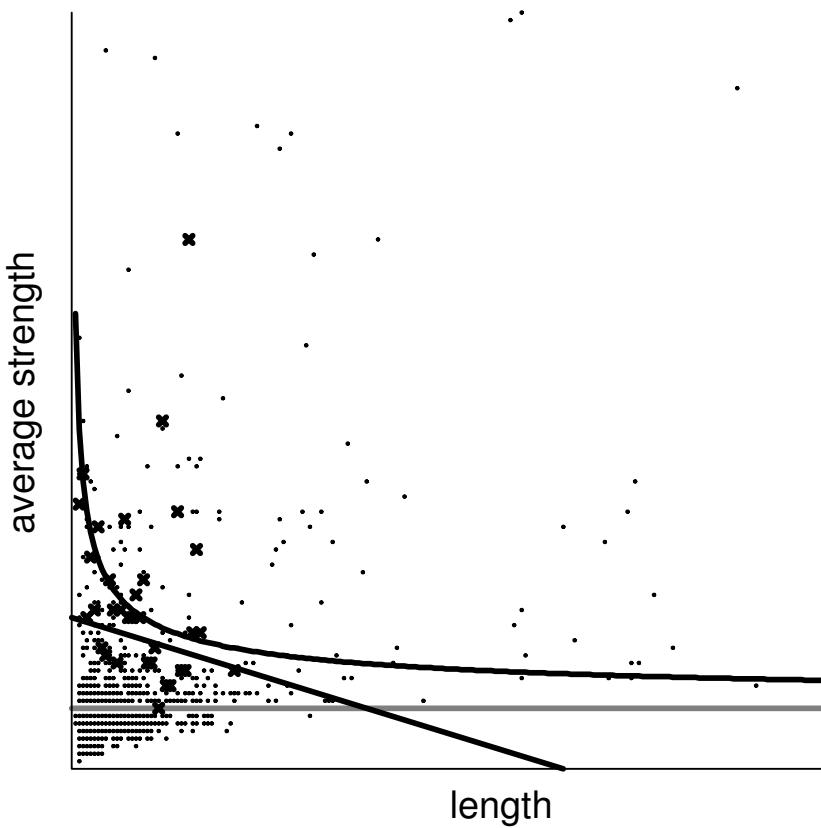


Figure 3: Decision rules in feature space for the can image

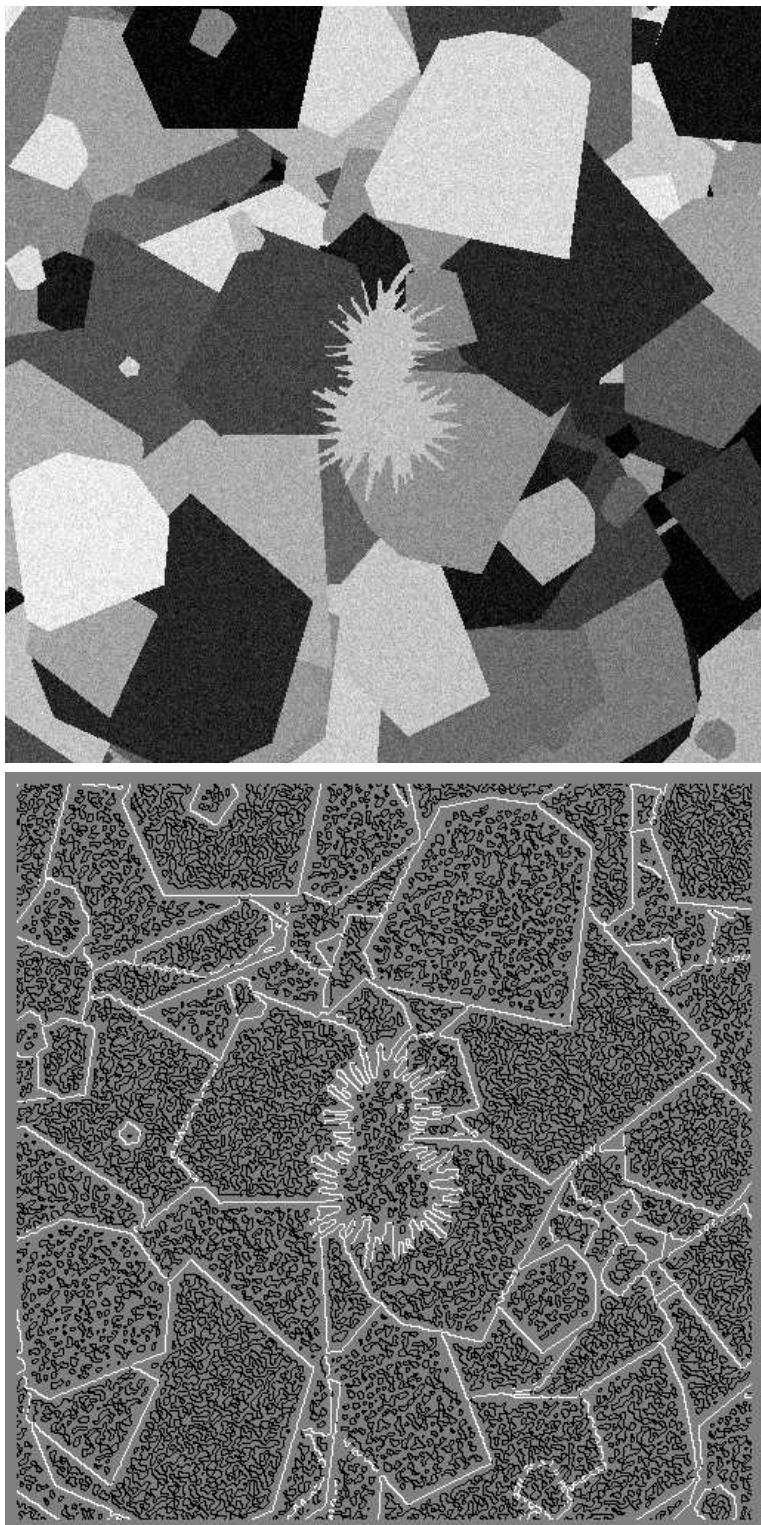
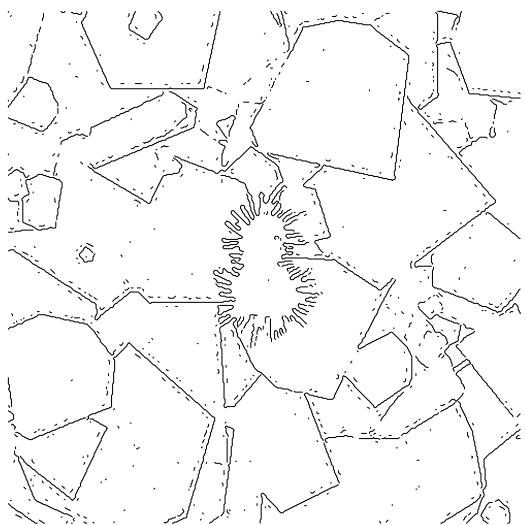


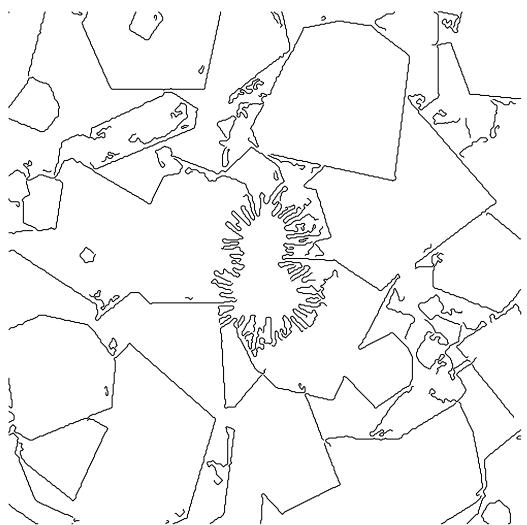
Figure 4: Synthetic pear image and detected edges



Figure 5: ROCs of saliency measures applied to pear image

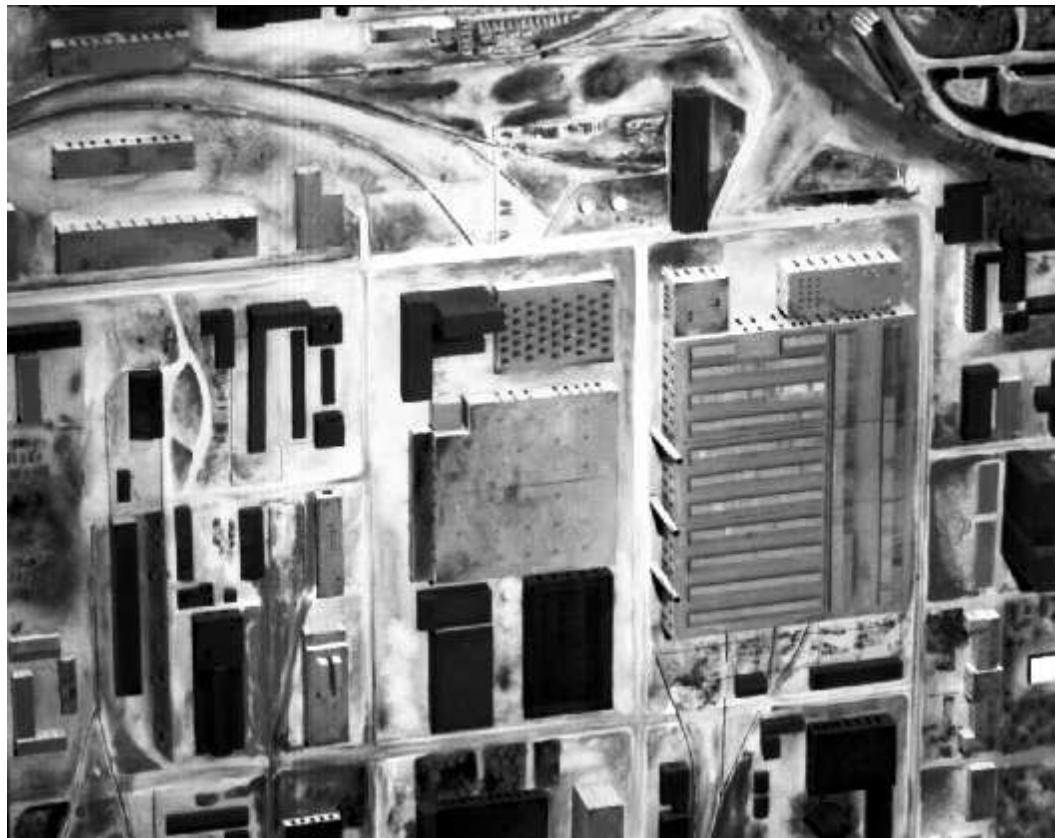


(a) Rayleigh method

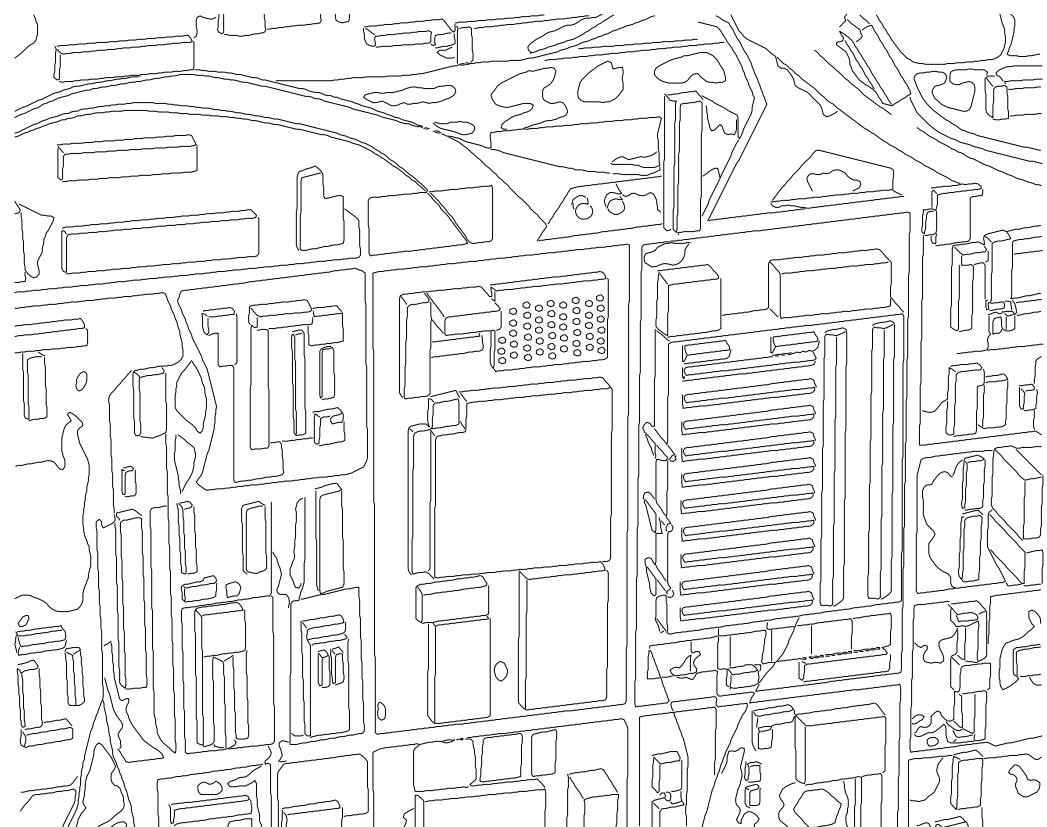


(a) non-linear CLT method

Figure 6: Thresholded edges of pear image



(a) intensity image



(b) ground truth edges

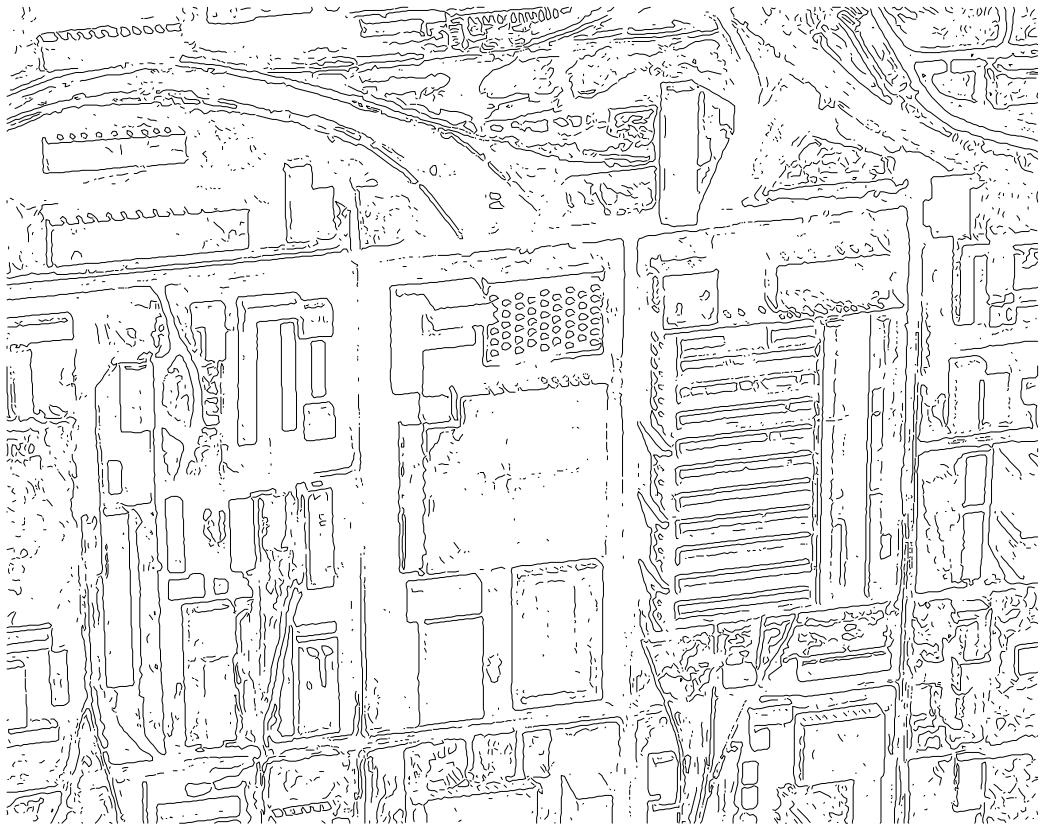


(c) 512×512 portion of labelled edge detector output

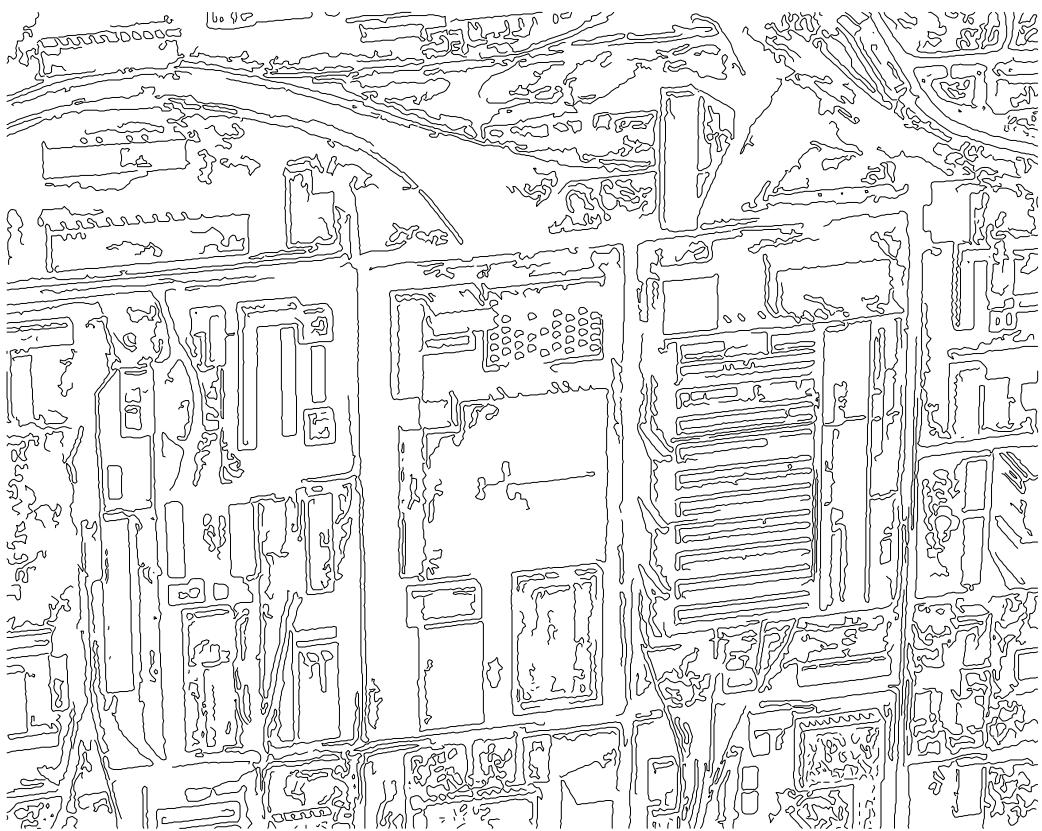
Figure 7: J25 image and detected edges



Figure 8: ROCs of saliency measures applied to j25 image; $\sigma = 1$



(a) neural network using neighbourhood



(b) linear CLT method

Figure 9: Thresholded edges of J25 image

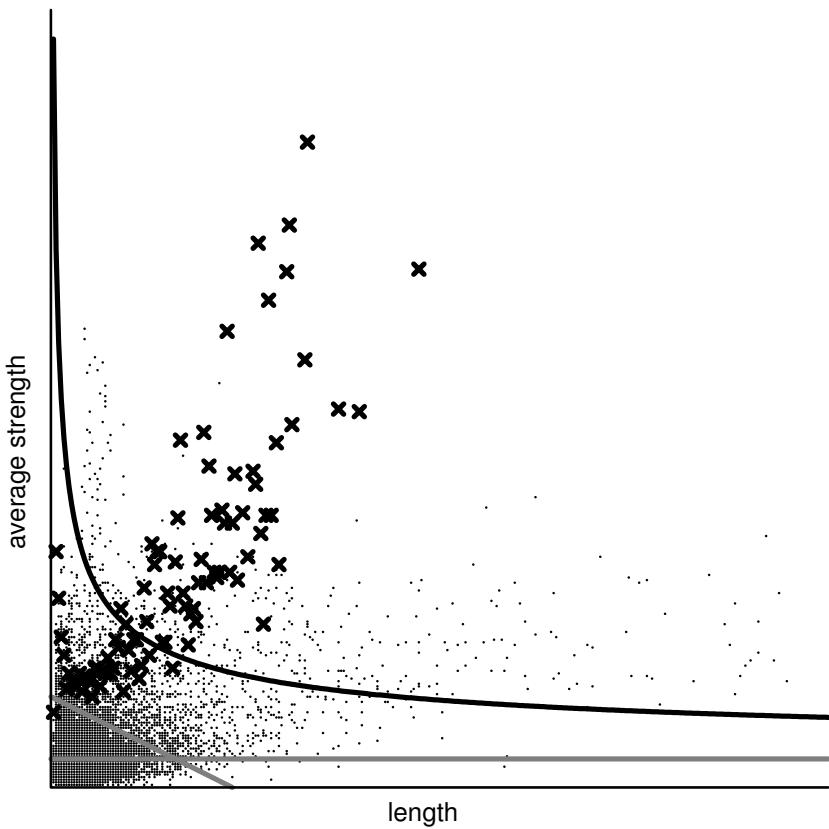


Figure 10: Decision rules in feature space for J25 image; $\sigma = 1$



Figure 11: sub2 image; a) 1st principal component, b) reference edges



Figure 12: ROCs of saliency measures applied to `sub2` image



(a) Rayleigh method with $P_F = 0.05$ using hysteresis



(b) linear CLT method

Figure 13: Thresholded edges of sub2 image using gradient magnitude

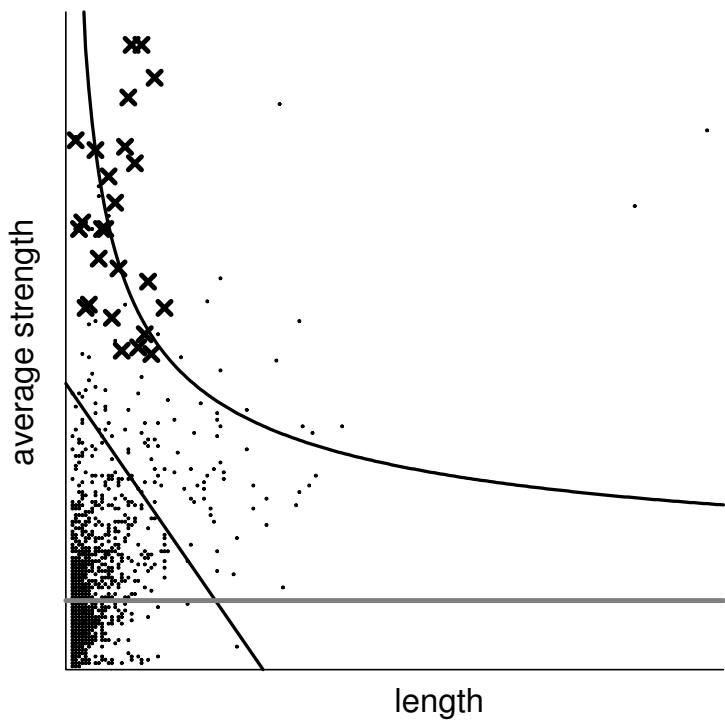


Figure 14: Decision rules in feature space for sub2 image

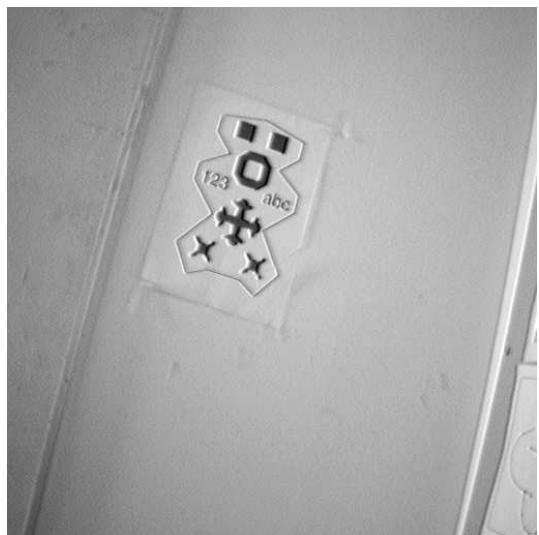


(a) Rayleigh method with $P_F = 0.05$ using hysteresis

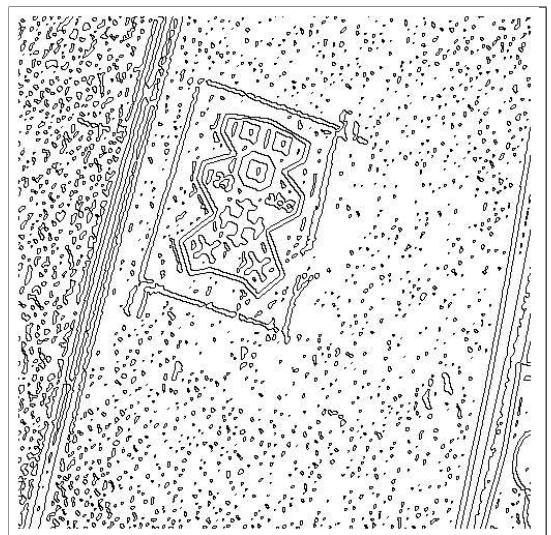


(b) linear CLT method

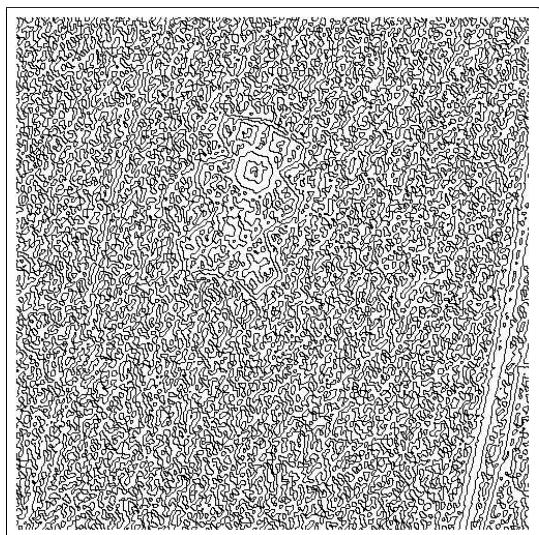
Figure 15: Thresholded edges of `sub2` image using fine → coarse gradient lifetime



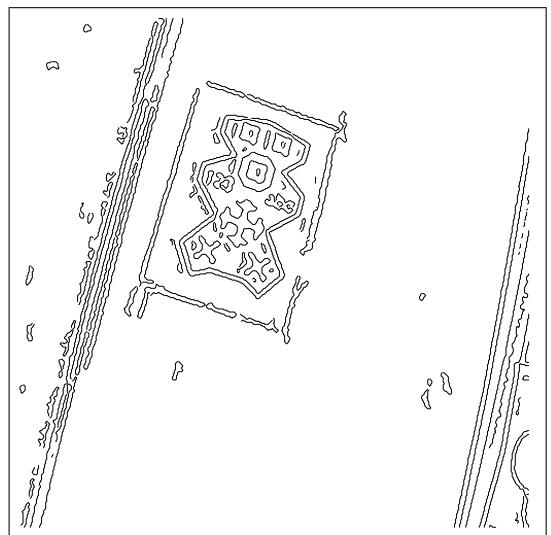
(a)



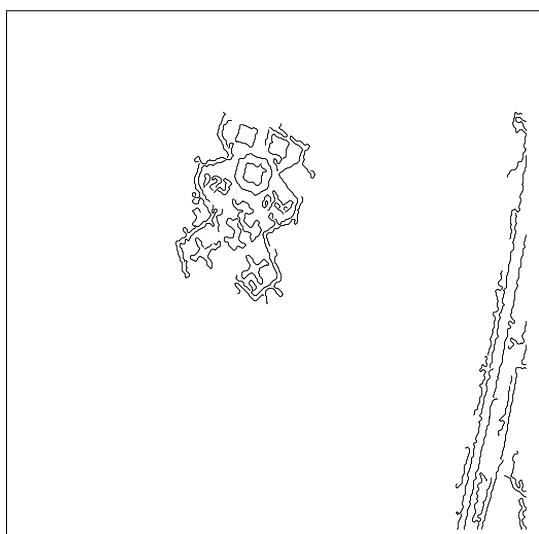
(b)



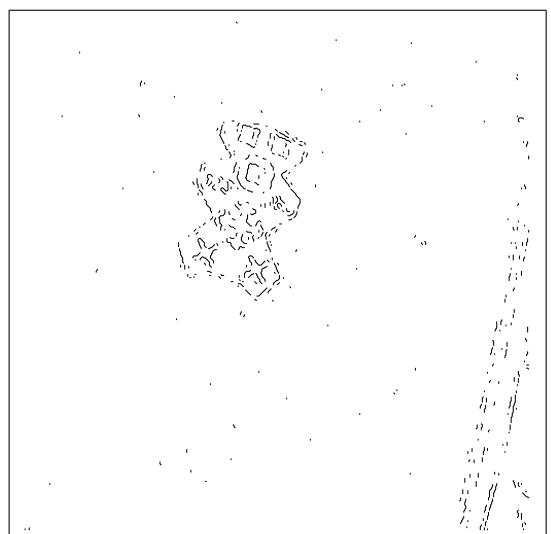
(c)



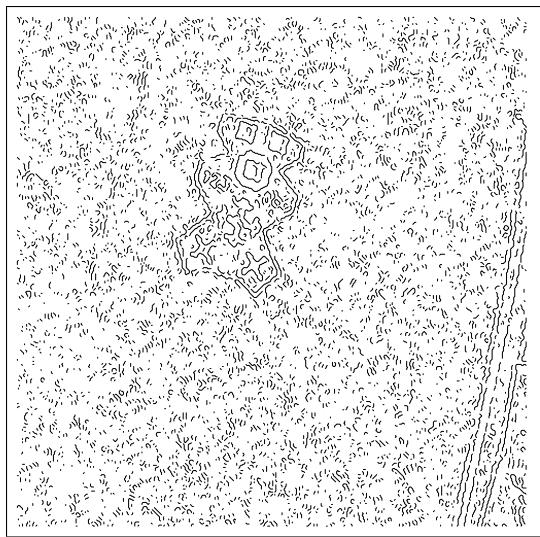
(d)



(e)



(f)



(g)

Figure 16: a) image at F/1.3, b) all edges at F/1.3, c) all edges at F/6, d) thresholded edges at F/1.3 using non-linear CLT method, e) thresholded edges at F/6 using non-linear CLT method, f) thresholded edges at F/6 using Rayleigh method, g) thresholded edges at F/6 using Rayleigh method with hysteresis

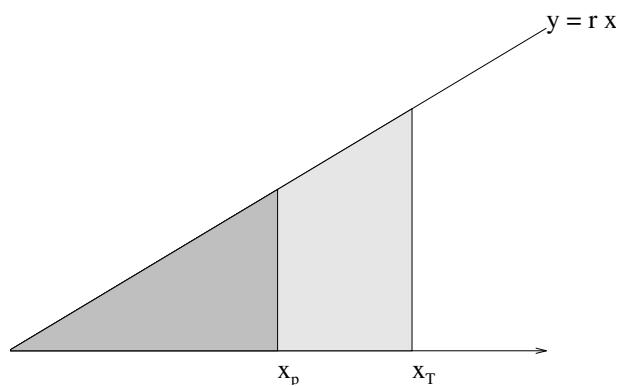


Figure 17: Determining median of triangle

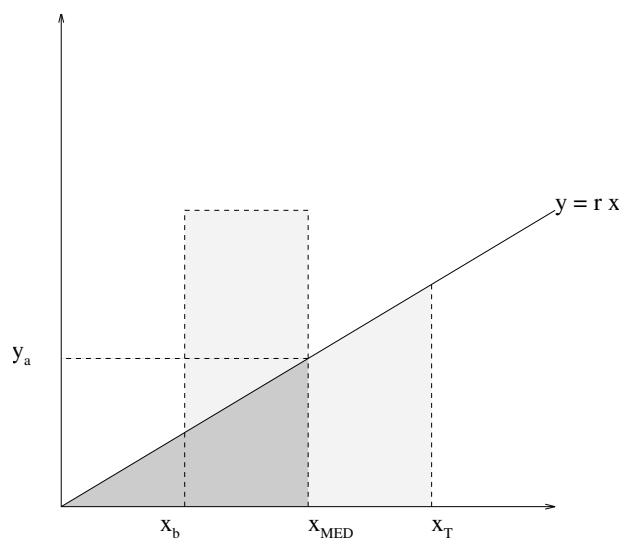


Figure 18: Determining median absolute deviation of triangle