STATISTICAL TECHNIQUES FOR EDGE DETECTION IN HISTOLOGICAL IMAGES

D. Svoboda†‡, I. A. Williams†, N. Bowring†, E. Guest§

†*Department of Engineering and Technology, Manchester Metropolitan University Manchester M1 5GD*

‡*Centre for Biomedical Image Analysis, Faculty of Informatics,*
Masaryk University Botanická 68a, 602 00 BRNO, Czech Republic

§*School of Computing, Leeds Metropolitan University, Leeds LS1 3HE*

Keywords: Edge detection, Statistical, Histological images, Parametric and Non-Parametric tests.

Abstract: A review of the statistical techniques available for performing edge detection on histological images is presented. The tests under review include the Student's T Test, the Fisher test, the Chi Square test, the Kolmogorov Smirnov test, and the Mann Whitney U test. All utilize a novel two sample edge detector to compare the statistical properties of two image regions surrounding a central pixel. The performance of the statistical tests is compared using histological biomedical images on which traditional gradient based techniques are not as successful, therefore giving an overall review of the methods, and results. Comparisons are also made to the more traditional Canny and Sobel, edge detection filters. The results show that in the presence of noise and clutter in histological images both parametric and non-parametric statistical tests compare well robustly extracting edge information on a series images.

1 INTRODUCTION

Edge detection is a key process in many computer vision and image understanding applications. It is commonly understood how the edges in an image are vital to region segmentation and object recognition processes. Therefore any process subsequent to edge detection has its success directly dependent on the effectiveness and the accuracy of the edge detection stage. This has led to the development of many algorithms for edge and line detection, each with the unified goal of producing the optimum results for a range of applications and images.

Many of the early detectors developed are gradient and therefore derivative based, for example Roberts, Sobel, and Prewitt (Šonka et al., 1986). These original detectors were shown to perform effectively on synthetic images, or images with very little noise. Studies have since illustrated that when applied to images with significant noise, clutter, or texture as is common with histological images, their performance is found to diminish and significant spurious edges are detected. To partially overcome this problem of additive noise within images (Canny, 1986) introduced an analytical optimal step edge detector based on the first derivative of a Gaussian filter. This reduced the presence of image noise by providing Gaussian smoothing to the image prior to edge extraction, and therefore avoided to some extent the spurious effects of noise previously observed. The control of the strength of the Gaussian filter was dependent on the application and introduced a measure of user subjectivity, with the amount of smoothing acting as a user defined threshold on the image edges produced. Canny's filter is still widely used for the detection of edge information and is consequently seen by many as a benchmark for comparative tests of new edge detection methods. This said, the use of Gaussian smoothing prior to any edge detection can introduce edge localization problems due to the blurring nature of such smoothing filters. This effect can be avoided through the use of non-linear diffusion filtering (Perona and Malik, 1990) as a pre cursor to any gradient based edge detection technique.

As an alternative to these gradient based methods the problem of edge detection when applied to the noisy images like the mouse atlas (MA) (Brune et al., 1999) examples has been approached through the use of statistical tests. The comprehensive analysis of five such parametric and non-parametric statistical tests for the detection of edges and rib structures in X-Ray imagery was illustrated by (de Souza, 1983). Although the results presented in deSouza use only a single dimension of the image in their

Svoboda D., A. WilliamsâĂă I., BowringâĂă N. and Guest E. (2006).
STATISTICAL TECHNIQUES FOR EDGE DETECTION IN HISTOLOGICAL IMAGES.
In *Proceedings of the First International Conference on Computer Vision Theory and Applications*, pages 457-462
DOI: 10.5220/0001377904570462

analysis they are still significant to the application of two dimensional image data. The work of (Bovik et al., 1986) illustrated the theoretical use of many non-parametric tests within a two dimensional image corrupted with noise sampled at four different orientations. This work although largely theoretical and lacking in a comprehensive analysis of the results, did however evaluate the high computational requirements incurred by such ranking statistical tests. In spite of the computational cost being greater than the derivative based detectors, similar parametric tests were used effectively by (Beauchemin et al., 1998) to overcome the low signal to noise ratio that is evident in detecting edges in synthetic aperture radar (SAR) images, and also by (Huang and Tseng, 1988) to overcome the blurring effect found with gaussian smoothing filters.

To perform statistical tests effectively on two dimensional images and therefore allow comparisons to be made to the more traditional techniques, a novel edge detection algorithm was introduced by (Fesharaki and Hellestrand, 1994) that combined the use of a 5×5 pixel image mask and the popular two distribution Student's T test. This allowed them to effectively detect edges at eight different orientations in both noiseless and noisy images. Comparisons to the traditional gradient methods have illustrated a robust performance in the presence of noise (Kundu, 1990) and (Hou, 2003), who illustrated how statistics can outperform Canny on images corrupted with impulsive noise. Also found by (Lim and Jan, 2002) (Lim and Jan, 2006) was the possibility that a modified Student's T test could perform well on images with little noise, however was outperformed by the Kolmogorov Smirnov test in intense noise images.

All these methods, while removing the need for the smoothing parameters evident in Canny and other gradient based techniques, do not eliminate the need for a subjective user threshold. Through the use of a probability value of the test in question they perform a statistical confidence test using lookup tables. Work by (Bowring et al., 2004) has since indicated the possibility of producing images superior to both Canny and SUSAN (Smith and Brady, 1997) using novel statistical methods without the need for a confidence check simply by varying the size of the image mask and therefore the amount of data points used in the tests. Furthermore work by (Williams et al., 2005) illustrated how through the use of multiple masks of varying scales applied to the same image and artificial neural networks, it is possible to remove the need for any subjective threshold when producing superior statistical images, albeit at higher computational cost.

2 THE STATISTICAL EDGE DETECTION FILTER

For all of the results presented here, the same filter principle is used as that described by (Bowring et al., 2004). The reader is directed to that work for a more detailed full description of its operation. The statistical edge-detection filter principle is shown in simplified form in (Fig: 1). It details an edge section of a mouse atlas image (MA) (Brune et al., 1999) with a single square mask applied. Each mask used is divided in two equal areas surrounding a central pixel at various angles of $90°$, $60°$, $45°$ etc. If the mask lies entirely in a homogeneous region within the image, then there will be little or no difference in the computed statistical measures between both areas. The maximum difference will occur when the mask lies directly over the boundary between the two regions (as in Fig: 1), therefore generating greatly differing statistical measures for each of the regions. Using this technique, the likely edge direction is also determined and is used for later non-maximal suppression of the image when necessary.

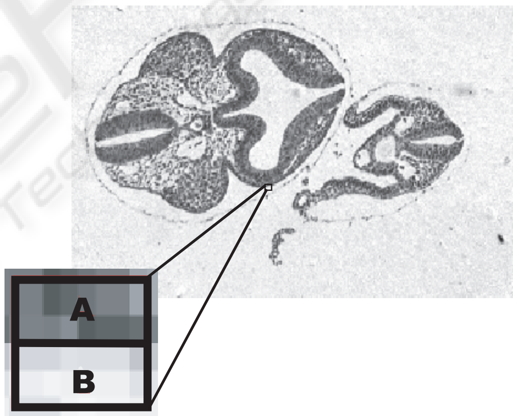


Figure 1: Illustrating a single statistical mask applied to an image region at an angle of $0°$. Each mask is divided into two equal sized regions *A* and *B* located around the central pixel of interest.

2.1 Implementing the Statistical Tests

For the analysis work, various statistical parametric and non-parametric tests have been used to compare two equal sized samples. Each of the tests used will give a high response if the two data sets *A* and *B* come from different regions of the image under evaluation, and likewise low values if they are from the same region.

The tests at use are:

Fisher Test

The Fisher Test (or the F Test as it is commonly referred) tests the hypothesis that two distributions will have the same variance. Fishers test is a non-parametric test, therefore making no assumption about the two data sets under evaluation.

The two-sample F-test is defined as follows:

$$F = \max\left(\frac{s_A}{s_B}, \frac{s_B}{s_A}\right) \quad (1)$$

Where s_A and s_B are the variances of the two *regions A* and *B* surrounding the central pixel.

Student's T-test

The Student's T-test is a parametric test based on the hypothesis that the two distributions will have the same or a similar mean value. The Student's T-test is generally used where it is expected that the two populations will have similar variances. However it has been shown that even with regions of greatly differing variances the test gives good results in practice (Bowring et al., 2004), (Williams et al., 2005), (Lim and Jan, 2002).

The T-test is given as:

$$T = \frac{|\bar{x}_A - \bar{x}_B|}{\sqrt{\frac{\alpha(|A|+|B|)}{|A \cup B|}}} \quad (2)$$

Where $\bar{x}_A$ is the mean and $|A|$ is the number of pixels from *region A*, and $\bar{x}_B$ and $|B|$ correspond to *region B*. α is defined as:

$$\alpha = \frac{|A|\bar{x}_A + |B|\bar{x}_B}{|A \cup B| - 2} \quad (3)$$

Kolmogorov Smirnov Test

The Kolmogorov Smirnov test (KS-Test) is a non-parametric test based on the empirical distribution function of ascending data points:

$$F_A(i) = n(i)/N \quad (4)$$

Where $n(i)$ is the number of data points less than the current data point in ranked set *A*, and N is the number of overall points contained in data set *A*.

The two sample KS tests checks for the maximum difference between the empirical and cumulative distribution functions for both data sets. From this it returns the value of D given in the equation:

$$D = \max_{i \epsilon \{1,\ldots,N\}} |F_A(i) - F_B(i)| \quad (5)$$

Where F_A is the empirical distribution function for data set *A*, and F_B is the empirical distribution for data set *B*

Chi Square Test

The Chi Square test uses checks for the independence of the two different data sets. The comparison is calculated by taking the difference at the same position for two binned datasets. Here the bins are defined by *region A* and *region B* of the mask.

The Two Sample Chi Square test is given as:

$$\chi^2 = \sum_i \frac{(R_i - S_i)^2}{R_i + S_i} \quad (6)$$

Where R_i is the number of values in *bin i* of *region A*, and S_i is the number of values in *bin i* of *region B.*

Mann Whitney U Test

The Mann Whitney U test checks the hypothesis that the two data sets under evaluation are taken from the same distribution. The statistical value U corresponds to a rank score which is calculated for both data sets.

$$R_A = \sum_{x \in A} \left(\sum_{y \in B; y < x} 1 \right) \quad (7)$$

$$R_B = \sum_{x \in B} \left(\sum_{y \in A; y < x} 1 \right) \quad (8)$$

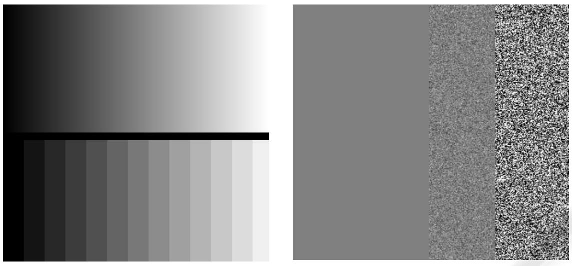
$$U = \min(R_A, R_B) \quad (9)$$

Where R_A refers to the data originating from *region A* of the mask and likewise R_B refers to *region B.* U is the overall statistical significance relating to the minimum value between R_A and R_B.

3 RESULTS AND ANALYSIS

3.1 Synthetic Images

To analyze the performance of the statistical tests it was important to test their function on synthetic images. The aim of these images was to exploit or hinder the specific characteristics of each test in question. The first of these images (Fig: 2(a)) featured a gradient image with continuous and stepped grayscale levels. This image specifically featured a gradual change in mean but a constant variance. The second of these test images (Fig: 2(b)) featured three step edges of varying levels of Gaussian noise. This image featured noise distributed with uniform mean although having a gradual change in variance



(a) (b)

Figure 2: Synthetic test images. a) Uniform variance and gradual mean change. b) Uniform mean and gradual variance change.

Figure of Merit:

$$R = \frac{1}{I_{sum}} \sum_{i=1}^{I_A} \frac{1}{1 + \beta d_i^2} \quad (10)$$

Where:
I_{sum} = $max(I, I_A)$.
I = The sum of the ideal edge points.
I_A = The sum of the detected edge points.
d_i = the distance of the i^{th} edge point from the ideal edge point.
β = A scaling constant (typically set to $\frac{1}{9}$).

Pratt's figure of merit (FOM) (Pratt, 1991) is adapted here to work with grayscale images. The performance value is calculated for each of the 256 grey levels in both the edge detected image, and the ideal gold standard image. The mean of these 256 Pratt's merit values is then assigned as the figure of merit for that particular image. Fig: 3 illustrates the FOM for the synthetic test images after non maximal suppression.

Table 1: Measured figure of merit values for the synthetic images at a range of mask sizes. "1" is the ideal result, "0" is a poor response. Values in boldface indicate the best response at the given resolution.

Image	Fig: 2(a)			Fig: 2(b)		
Mask Size	5×5	11×11	15×15	5×5	11×11	15×15
F	0.331	0.122	0.057	**0.405**	**0.567**	**0.567**
T	**0.905**	**0.602**	**0.469**	0.051	0.052	0.069
KS	0.158	0.136	0.123	0.116	0.369	0.437
χ^2	0.159	0.145	0.132	0.262	0.370	0.403
U	0.193	0.127	0.136	0.052	0.028	0.026
Variance	2	5	7	2	5	7
Canny	0.627	0.377	0.270	0.050	0.040	0.034
Sobel	0.880			0.053		

(a) (b)

Figure 3: Edge detection results for Fig: 2(a) and (b) images. a) T Test 5×5 b) F Test 11×11.

3.2 Histological Images

Histological images often tend to have poorly identifiable boundaries corrupted by noise, and generally have a very low contrast. The images used here for analysis are two typical histological images (Fig: 4). One is of a mouse atlas embryo (Brune et al., 1999), and the other is a section of human tissue colon courtesy of Faculty hospital Bohunice of Masaryk University (Brno, the Czech Republic). The FOM results are shown in (Table: 3). The performance is measured against gold standard images that have been pre-segmented by an expert in the field to include only their ideal edge points.

Table 2: Computational time of test for Fig: 2(b) (300×300 pixels). CPU specification: Pentium 4, 3.20GHz, 256 MB RAM.

Test	F	T	KS	χ^2	U
time (s)	2.59	2.48	10.25	5.46	6.42

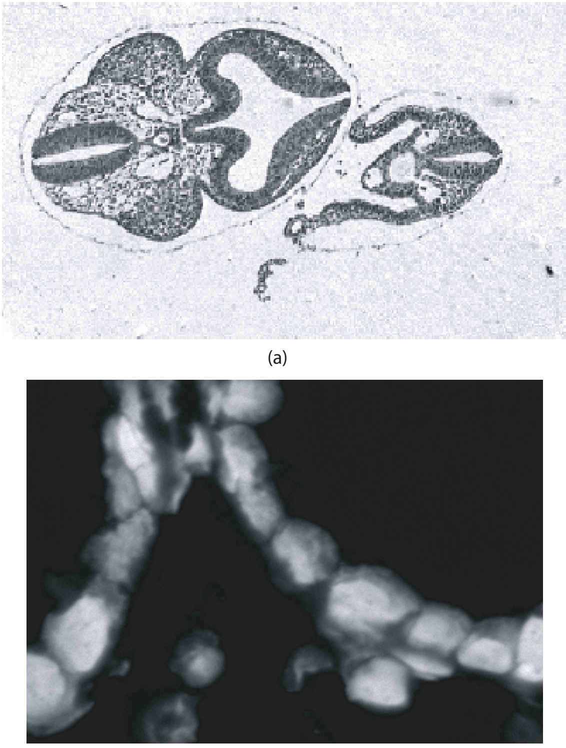


Figure 4: Histological images. a) Mouse Atlas Embryo, b) Human tissue colon.

4 CONCLUSION

An analysis of statistical tests available for detecting edges in noisy histological images has been presented here. The results (Table: 1, Fig: 3) have illustrated that with synthetic images the Student's T test and Sobel filters perform better where there is a known change in mean, however both are outperformed by the non parametric Fisher, KS and Chi square tests when the mean is constant and the variance changes. Further analysis of real image data illustrated that, when detecting edges in the histological MA images, the KS, Chi square and Student's T test performed the best overall. Even though the Canny filter outperformed the statistical tests at a fine resolution when compared to a mask of 5×5, it was outperformed as the number of data points and therefore the mask size increased. Also we illustrated how the Student's T test performed better with the tissue colon images, whereas the non-parametric KS and Chi square tests were poorer. We can assess this poorer performance for the non-parametric tests to be proportional to the size of the objects and edges within the image, thus allowing greater performance where the image contains larger objects. It also is commonly understood how the Canny produces edge images that, although

Table 3: Measured figure of merit (FOM) values for the histological image at a range of mask sizes. "1" is the ideal result, "0" is a poor response. Values in boldface indicate the best response at the given resolution.

Image	Mouse Atlas			Tissue Colon		
Mask Size	5×5	11×11	15×15	5×5	11×11	15×15
F	0.164	0.315	0.277	0.049	0.130	0.120
T	0.164	0.404	0.337	0.125	**0.392**	**0.368**
KS	0.130	0.391	0.405	0.036	0.098	0.104
χ^2	0.061	**0.411**	**0.432**	0.034	0.144	0.121
U	0.143	0.376	0.365	0.033	0.066	0.071
Variance	2	5	7	2	5	7
Canny	**0.346**	0.354	0.331	**0.304**	0.332	0.301
Sobel	0.360			0.311		

accurate in location, have edges that are not continuous. The statistical tests presented here incorporate an edge tracking process that is inherent in the algorithm. This tracking produces uninterrupted edges relative to the mask size. It is therefore a future goal to determine not only an FOM performance comparison of the detectors but also to assess edge continuity.
We have also illustrated that there is no specific statistical test suitable for all types of image. It is therefore a further goal to find the most successful combination of statistical tests which will perform well on all image data.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Education of the Czech Republic (Projects No.1K05021 and No. LC535) and by the Academy of Sciences of the Czech Republic (Grant No.A5004306). We would also like to thank (EMAP) the University of Edinburgh mouse atlas project for providing us with the MA images.

REFERENCES

Beauchemin, M., Thomson, K. P. B., and Edwards, G. (1998). On nonparametric edge detection in multilook sar images. *IEEE Transactions on GRS*, 36(5):1826–1829.

Bovik, A. C., Huang, T. S., and Jr, D. C. M. (1986). Non-parametric tests for edge detection in noise. *Pattern Recognition*, 19(3):209–219.

Bowring, N. J., Guest, E., Twigg, P., Fan, Y., and Gadsby, D. (2004). A new statistical method for edge detection on textured and cluttered images. In 4^{th} *IASTED VIIP Conf.*, pages 435–440.

Brune, M., Bard, J., Dubreuil, C., Guest, E., Hill, W., Kaufman, M., Stark, M., Davidson, D., and Baldock, R. (1999). A three-dimensional model of the mouse at embryonic day 9. 216(2):457–468.

Canny, J. (1986). A computational approach to edge detection. *IEEE T-PAMI*, 8:769–698.

de Souza, P. (1983). Edge detection using sliding statistical tests. *CVGIP*, 23(1):1–14.

Fesharaki, M. N. and Hellestrand, G. R. (1994). A new edge detection algorithm based on a statistical approach. In *ICSIPNN '94*, pages 21–24. IEEE.

Hou, Z. (2003). Robust edge detection. *Pattern Recognition*, 36(9):2083–2091.

Huang, J. S. and Tseng, D. H. (1988). Statistical theory of edge detection. *CVGIP*, 34(3):337–346.

Kundu, A. (1990). Robust edge detection. *Pattern Recognition*, 23(5):423–440.

Lim, D. H. and Jan, S. J. (2002). Comparison of two-sample tests for edge detection in noisy images. *Statistician*, 51(1):21–30.

Lim, D. H. and Jan, S. J. (2006). Robust edge detection in noisy images. *Computational Statistics and Data Analysis*, 50(3):803–812.

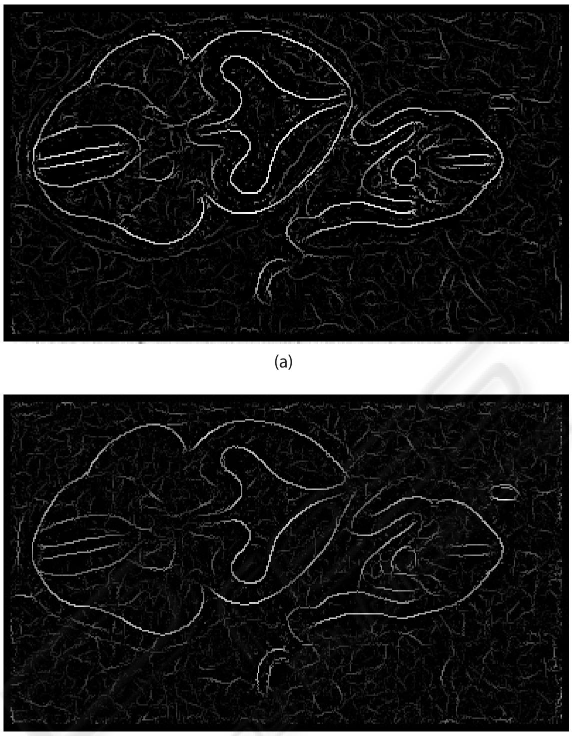
Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, 12(7):629–639.

Pratt, W. K. (1991). *Digital Image Processing*. Wiley. ISBN: 0-471-37407-5.

Smith, S. and Brady, J. (1997). Susan – a new approach to low level image processing. *IJCV*, 23(1):45–78.

Šonka, M., Hlaváč, V., and Boyle, R. (1986). *Image Processing Analysis and Machine Vision*. Chapman and Hall Publishing. London.

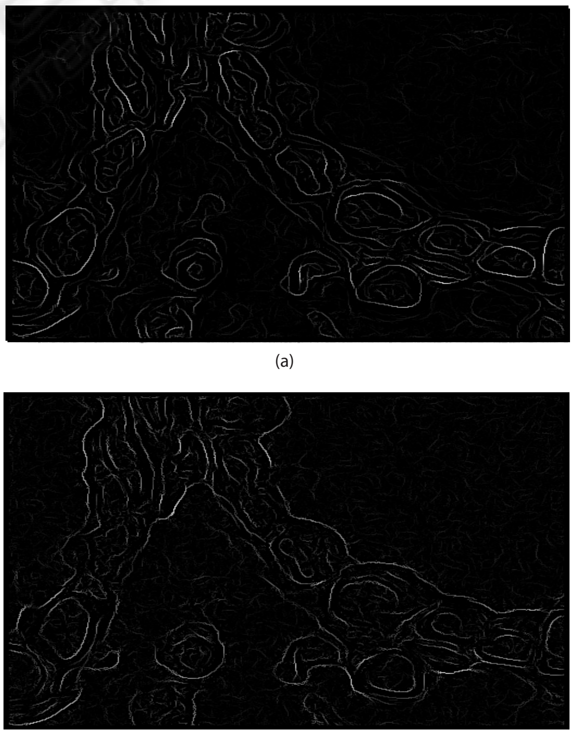
Williams, I., Bowring, N. J., Guest, E., Twigg, P., Fan, Y., and Gadsby, D. (2005). A combined statistical/neural network multi-scale edge detector. In 5^{th} *IASTED VIIP Conf.* ISBN: 0-88986-528-0, ref: 480-266.



(a)

(b)

Figure 5: Edge Detection results. a) T Test 11x11, b) Chi square Test 11x11.



(a)

(b)

Figure 6: Edge Detection results. a) T Test 11x11, b) F Test 11x11.