Statistical Edge Detection in CT Image by Kernel Density Estimation and Mean Square Error Distance

Xu Xu†, Student Member, Yi Cui††, and Shuxu Guo†††(a), Nonmembers

SUMMARY In this paper, we develop a novel two-sample test statistic for edge detection in CT image. This test statistic involves the non-parametric estimate of the samples’ probability density functions (PDF’s) based on the kernel density estimator and the calculation of the mean square error (MSE) distance of the estimated PDF’s. In order to extract single-pixel-wide edges, a generic detection scheme cooperated with the non-maximum suppression is also proposed. This new method is applied to a variety of noisy images, and the performance is quantitatively evaluated with edge strength images. The experiments show that the proposed method provides a more effective and robust way of detecting edges in CT image compared with other existing methods.

key words: edge detection, CT image, two-sample test statistic, kernel estimation, mean square error distance

1. Introduction

With the continuous development of medical imaging technologies, CT scan becomes increasingly important. With the help of automatic edge detection, it is possible for doctors to diagnose diseases more effectively, and the diagnostic accuracy can be greatly improved. A clearly detected edge of pathological and diseased tissues is helpful for doctors to observe and diagnose diseases. However, the resolution of CT images is not high enough, and the images are usually blurred due to tissue motions and noise interference. Therefore, improving the accuracy of tissue edge detection in CT image acquisition becomes an important practical subject [1].

By an intuitive definition, an edge is an abrupt spatial change of the gray level. Hence the first or second derivative of an image is mostly utilized in conventional edge detection methods, such as the Sobel, the Roberts, the Prewitt, and the Laplacian operators [2]. However, such methods are known to be vulnerable in the presence of noises, and false detections are likely to happen.

To address this problem, a common approach is to prefilter an image prior to the application of an edge detection operator. For example, a simple combination of the Gaussian filter and the Laplacian operator yields the well-known Laplacian of Gaussian (LoG) edge detector [2]. The first step of the Canny detector is also a Gaussian filter to improve the signal-to-noise ratio (SNR), and then gradient estimation is implemented [3]. Another recent example uses the Priestly-Chao kernel smoother which essentially serves as a pre-filter to smooth the image, and then the gradient is calculated in the analogue domain [4]. The local smoothing of the image followed by a statistical hypothesis testing on the gradient adopted by Abraham et al. [5] can also be treated as a noise filter. In order to achieve better filtering effect, some nonlinear filters are also proposed to deal with Gaussian or non-Gaussian noises [6]–[10]. Most recently, some methods have also been proposed to obtain simultaneous filtering and edge detection. For example, Laligant et al. proposed a non-linear derivative scheme in which both noise reduction and edge detection are accomplished in one-stage [11]. Wang modified the Laplacian operator to achieve a more robust result [12]. Of all the above methods, the essential idea of edge detection is still to make use of the derivatives. Therefore, we put them under the derivative framework.

Alternatively, both filtering and edge detection can take place at the same time under the statistical framework by making use of hypothesis testing [13]. According to different noise models, various parametric two-sample test statistics are available. For example, when the image is contaminated by Gaussian noise, the T-test can be used [14]. Unfortunately, parametric statistics are sensitive to the violation of the underlying distribution assumption. In order to give robust detection results, several non-parametric statistics, such as the WMW test statistic and the K-S test statistic, are introduced in [15]–[18]. Such a statistical framework is especially preferred in remote sensing applications in which noises are mostly non-Gaussian [19]–[21].

In this paper we follow the second framework by proposing a new two-sample test statistic for edge detection in CT image. The contributions of our work are:

1) We have derived an effective and yet robust edge detector for CT image contaminated by various noises. Unlike traditional edge detectors (e.g., the Canny detector) which implicitly assume Gaussian noises, our method is developed from non-parametric approach and shows robustness for general noise settings. We show that the proposed method performs comparably to the Canny detector when the noise is truly Gaussian. For the non-Gaussian cases which exist ubiquitously in practice, our method achieves a more stable performance over conventional detectors.

2) We have designed an edge detection algorithm that is...
able to extract single-pixel-wide edges. To our best knowledge, up to now, almost all of the two-sample test statistic based edge detectors report thick edges and very few edge extraction methods are systematically proposed for these statistical detectors. This becomes a bottleneck of their applicability when compared with derivative based detectors, such as the Canny detector which always produces single-pixel-wide edges. Also filling such a gap makes the comparison possible between the test statistic based detectors and the derivative based detectors.

3) We have compared the test statistic based edge detectors with derivative based detectors (the Canny detector, in particular) for different noisy CT image in a quantitative way. Although these two types of edge detectors have been extensively studied and successfully applied in their respective fields, no comparison has been made to reveal the advantages and disadvantages between them. This work is necessary for the choice of appropriate detector for different application purposes.

Accordingly, this paper is organized as follows. Section 2 introduces the new two-sample test statistic; Section 3 illustrates the edge detection scheme including two-dimensional implementation of the edge detector and the post-processing for single-pixel-wide edge extraction. Section 4 focuses on the performance evaluation. Last, Section 5 concludes the paper.

2. New Two-Sample Test Statistic

Under the statistical framework, edge detection can be modeled equivalently as a problem of a two-sample test. That is, we want to test the alternative hypothesis \( H_1 \) that the two samples are generated from different distributions and hence separated by an edge against the null hypothesis \( H_0 \) that they are generated from the same underlying distribution and covered in the same region. To this end, a test statistic is to be constructed and a tradeoff between significance level and the power of the test has to be achieved. In the context of detection problems, the terminology of detectors, false alarm and detection probability is often used and in this paper we will refer to them interchangeably.

Now we consider two sets of samples \( X = \{x_1, x_2, \ldots, x_{N_1}\} \) and \( Y = \{y_1, y_2, \ldots, y_{N_2}\} \), where \( N_1 \) and \( N_2 \) are the numbers of samples in each set. Then our purpose is to test the null hypothesis \( (H_0) \) that \( X \) and \( Y \) have the same distribution against the alternative hypothesis \( (H_1) \) that their distributions are different.

2.1 Kernel Density Estimation and MSE Distance

We first begin with the estimate of the underlying PDF's of \( X \) and \( Y \) with a kernel method as follows [14]:

\[
p_1(x) = \frac{1}{N_1 h} \sum_{i=1}^{N_1} K\left(\frac{x - x_i}{h}\right) \quad (1)
\]

\[
p_2(x) = \frac{1}{N_2 h} \sum_{i=1}^{N_2} K\left(\frac{x - y_i}{h}\right) \quad (2)
\]

where \( K(u) \) is the kernel function and \( h \) is the bandwidth of the estimator. If \( X \) and \( Y \) come from different regions, saying that they are separated by an edge, (1) and (2) will have significantly different estimated PDF's; otherwise, they will give close results.

Next we take the MSE distance of (1) and (2) as an indicator of their similarity, which is given by:

\[
\Lambda_{MSE} = \int_{-\infty}^{+\infty} |p_1(x) - p_2(x)|^2 \, dx
\]

(3)

In this paper, we choose the Gaussian kernel function for the PDF estimator in (1) and (2), i.e.

\[
K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}
\]

(4)

By substituting (4) into (3) and performing the integration, we can analytically derive the close form for (3) by

\[
\Lambda_{MSE} = d(X, X) - 2d(X, Y) + d(Y, Y)
\]

(5)

where

\[
d(X, X) = \frac{1}{N_1^2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} e^{-\frac{|x_i - x_j|^2}{2h^2}}
\]

\[
d(X, Y) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} e^{-\frac{|x_i - y_j|^2}{2h^2}}
\]

(6)

\[
d(Y, Y) = \frac{1}{N_2^2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_2} e^{-\frac{|y_i - y_j|^2}{2h^2}}
\]

Up to this point, we have essentially developed a new two-sample test statistic \( \Lambda_{MSE} \). This test statistic measures the estimated PDF difference between two sets of samples. It is interesting to note that \( \Lambda_{MSE} \) has a close relationship with the Kolmogorov-Smirnov two-sample test statistic by which two empirical cumulative distribution functions (CDF's) are compared instead.

On the other hand, \( \Lambda_{MSE} \) can be considered as a similarity metric between two real valued sets. It goes without saying that \( \Lambda_{MSE} \geq 0 \) for any \( X \) and \( Y \) and the equality holds if and only if \( X = Y \). An analogy can also be made between (5) and the Euclidean distance where \( d(X, X) \), \( d(X, Y) \), and \( d(Y, Y) \) only take on different forms. This analogy is mainly owing to (3) which is in fact a MSE distance as well.

As a result, given the MSE nature of \( \Lambda_{MSE} \), we will name it the MSE test statistic or MSE detector hereafter.

2.2 Choice of the Kernel Bandwidth

The bandwidth \( h \) of the kernel function should be carefully
chosen for the best detection. If \( X \) and \( Y \) are both normally distributed with the same variance, the expected value of \( \Lambda_{\text{MSE}} \) under the null hypothesis (\( H_0 \)) and the alternative hypothesis (\( H_1 \)) are respectively given by (Appendix A):

\[
E(\Lambda_{\text{MSE}}|H_0) = \frac{2}{N} \left(1 - \frac{h}{\sqrt{\sigma^2 + h^2}}\right)
\]

\[
E(\Lambda_{\text{MSE}}|H_1) = \frac{2}{N} \left(1 - \frac{h}{\sqrt{\sigma^2 + h^2}}\right) + \frac{2h}{\sqrt{\sigma^2 + h^2}} \left[1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4(h^2 + \sigma^2)}}\right]
\]

where \( \mu_1, \mu_2 \) and \( \sigma^2 \) are the means and variance of the samples in \( X \) and \( Y \). We also assume equal number of samples in each sample set for simplicity. It is interesting to note that \( E(\Lambda_{\text{MSE}}|H_1) \) is actually the sum of \( E(\Lambda_{\text{MSE}}|H_0) \) and a second term which is directly related to the input signal-to-noise ratio (SNR) \((\mu_1 - \mu_2)^2/\sigma^2\). In fact, it is this second term that contributes to the distinction between \( X \) and \( Y \). As a result, the idea is to choose \( h \) so as to maximize the relative difference \([E(\Lambda_{\text{MSE}}|H_1) - E(\Lambda_{\text{MSE}}|H_0)]/E(\Lambda_{\text{MSE}}|H_0)\), i.e., the following function according to (7) and (8):

\[
RD(h) = \frac{Nh}{\sqrt{\sigma^2 + h^2}} \left[1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4(h^2 + \sigma^2)}}\right]
\]

(9)

It can be proved that (9) is bounded with respect to \( h \) and it approaches its supremum of \( N(\mu_1 - \mu_2)^2/(2\sigma^2) \) when \( h \) goes to infinity (Appendix B). Hence we want \( h \) to be as large as possible. Figure 1 shows the ROC curves of \( \Lambda_{\text{MSE}} \) with different choices of bandwidth \( h \) for simulated data of which the SNR under \( H_1 \) is set to be 3 dB, i.e., \((\mu_1 - \mu_2)^2/\sigma^2 = 2\), and the number of samples \( N \) in each set is 12 (which is the case for a 5×5 scanning window, See Sect. 3). It can be seen from Fig. 1 that the performance of \( \Lambda_{\text{MSE}} \) indeed improves as the bandwidth \( h \) increases, as predicted by (9).

In practice, \( h \) doesn’t have to be arbitrarily large. Figure 1 also indicates that when \( h \geq 2\sigma \), \( \Lambda_{\text{MSE}} \) actually exhibits no significant performance improvement. Hence a choice of \( h = 2\sigma \) would suffice.

It is worthwhile to point out that the choice of the kernel bandwidth for the best detection is different from that purely for density estimation purpose. In the latter case, in order to prevent over or under smoothing, the optimal bandwidth is given by [22]

\[
h_{opt} = \left(\frac{c_2}{c_1 \cdot A \cdot N}\right)^{1/5}
\]

(10)

where \( c_1 = \int_{-\infty}^{\infty} u^2 K(u)du \), \( c_2 = \int_{-\infty}^{\infty} K^2(u)du \), \( A = \int_{-\infty}^{\infty} [p''(u)]^2 du \), \( N \) is the sample size and \( p(u) \) is the PDF to be estimated. If we stick to the Gaussian kernel function of (4) and the normal distribution for \( p(u) \), we have \( c_1 = 1 \), \( c_2 = 1/(2 \sqrt{\pi}) \), \( A = 3/(8 \sqrt{\pi} \sigma^2) \). Hence (10) becomes

\[
h_{opt} = \left(\frac{4}{3N}\right)^{1/5} \sigma
\]

(11)

While (11) is optimal in the sense of best smoothing the estimated density function, it becomes inferior to larger bandwidth in terms of the best detection according to (9). This assertion is also reflected in Fig. 1, where the curve for \( h = 0.6444\sigma \) corresponds to the bandwidth choice according to (11).

Last, the choice of the bandwidth requires the information about the noise level of the image, or more specifically, the variance \( \sigma^2 \) of the noise. If \( \sigma^2 \) is unknown, an estimation of the noise variance should be carried out [23], [24], either globally or locally. However, it should be pointed out that errors introduced in the estimation of the noise variance will not undermine the detection performance whatsoever as long as the bandwidth of the kernel function \( h \) is sufficiently large, i.e., \( h \gg \sigma \), as indicated by (9).

3. Edge Detection Scheme

3.1 Two-Dimensional Implementation

Scanning window is used as a common approach to apply the MSE test statistic or any other two-sample test statistic (e.g., the WMW statistic) for the detection of edges in an image. The size of the window is a compromise between noise resistance and edge detail preservation. Typical sizes are 5×5, 7×7, or 9×9. To detect edges in all possible directions, the window is partitioned into two sub-regions excluding the central pixel along different directions as shown in Fig. 2.

\( \Lambda_{\text{MSE}} \) are calculated in each of these different directions and the most significant (maximum) value is assigned to the central pixel as the edge strength. In this way, we are actually dealing with

\[
\Lambda_{\text{MSE}} = \max_{i=1,2,3,8} \Lambda_{\text{MSE}}^{(i)}
\]

(12)

We can also define the edge direction at the central pixel as the direction along which \( \Lambda_{\text{MSE}} \) is achieved, i.e.
Fig. 2 A set of 7 × 7 scanning windows partitioned along eight possible directions where the shadowed areas represent the samples of $X$ and the white areas represent the samples of $Y$.

\[
\varphi_{\text{max}} = \frac{\pi (i_{\text{max}} - 1)}{8}
\]

where $i_{\text{max}} = \max_{i=1,2,...,8} \{ \Lambda_{\text{MSE}}^{(i)} \}$.

3.2 Edge Extraction

Edges detected with a single threshold directly applied to the edge strength image are often thickened due to the use of the scanning windows with unignorable dimensions as shown in Fig. 4. This effect is observed in virtually all two-sample test statistic based edge detection methods. Hou et al. [18] tried to alleviate this problem by introducing contrast test. However, the result is still disappointing with multiple-pixel-wide edges. In order to extract thin edges, methods such as morphological processing [19] or the watershed processing [25] was proposed. Since the former method is applied to the post-detected binary image, it tends to give inaccurate edge locations. The latter one, on the other hand, is not only prone to give over-segmented results but also possible to lose important edges when they don’t form a complete basin.

In fact, the idea of detecting local maxima as potential edges proposed by Canny [3] can be adopted here. A quick review of Canny’s method is helpful. The image is first regularized or smoothed by a Gaussian low pass filter for noise reduction. Next the first derivative along the horizontal and vertical direction form the gradient $g = [g_x, g_y]^T$, from which the magnitude of the gradient $||g||$ and the direction of the gradient $\theta = \arctan(g_y/g_x)$ can be determined. Then the magnitude $||g||$ is inspected along $\theta$ to see whether it is a local maximum. If not, non-maximum suppression is initiated. Thus all potential edges are thinned to be single-pixel-wide. Last, either single thresholding or hysteresis thresholding is applied to the non-maxima suppressed gradient magnitude image.

In order to make use of such a non-maximum suppression technique in our method, an analogy is made to Canny’s algorithm. First, the edge strength $\Lambda_{\text{MSE}}^{\text{max}}$ of (12) is analogous to the magnitude of the gradient $||g||$; second, the edge direction $\varphi_{\text{max}}$ of (13) is analogous to the direction of the gradient $\theta$. It should be noted that $\varphi_{\text{max}}$ designates the tangent direction while $\theta$ denotes the normal direction.

The above analogy makes the implementation of the non-maximum suppression in our method straightforward. That is, the edge strength $\Lambda_{\text{MSE}}^{\text{max}}$ is inspected along the direction $\pi/2 + \varphi_{\text{max}}$ (remember $\varphi_{\text{max}}$ is the tangent direction), and it is suppressed if it is not a maximum. Besides, this analogy also makes the comparison between the proposed method and Canny’s method possible (See Sect. 4 for details).

To sum up, the entire edge detection scheme can be described as follows:

1) Initialization: select the size $r \times r$ and the possible number of edge directions $d$ for the scanning window;
2) Edge Strength Generation: for each pixel, calculate $\Lambda_{\text{MSE}}^{(i)}$ as the edge strength for the $i$th direction; save $\Lambda_{\text{MSE}}^{\text{max}}$ and $\varphi_{\text{max}}$ for that pixel according to (12) and (13), respectively.
3) Non-maximum Suppression: for each pixel, check whether its $\Lambda_{\text{MSE}}^{\text{max}}$ is a local maximum along the direction of $\pi/2 + \varphi_{\text{max}}$, if not, suppress it to zero.
4) Thresholding: Compare the non-maximum suppressed edge strength image with a given threshold $\gamma$ (either by manual or automatic selection) to obtain the binary edge map.

The above algorithm is generically applicable to all two-sample test statistic based detectors, only by substituting $\Lambda_{\text{MSE}}$ with other test statistics. In the next section, several two-sample test statistics will be incorporated into this same edge detection scheme for performance comparison.

4. Experiments

4.1 Procedure of the Performance Evaluation

Our main purpose is to use the newly developed two-sample test statistic $\Lambda_{\text{MSE}}$ for edge detection, so it should be evaluated as an edge detector.

As pointed out in Sect. 1, plethora of edge detectors has been proposed in the literature. From a conclusive point of view, all those detectors are preferably compared with our proposed method. However, studies have shown that the Canny detector still enjoys a general advantage in a variety of edge detection applications [26]. Hence, we take it for granted that the Canny detector will be a good representative derivative based method for comparison. In [17], the Wilcoxon–Mann–Whitney (WMW) test statistic $\Lambda_{\text{WMW}}$, and the Kolmogorov–Smirnov (K-S) test statistic $\Lambda_{\text{KS}}$ were compared and their utility for edge detection is analyzed. So we also include them as representative statistical methods for comparison.

In order to compare the test statistic based edge detector with the Canny detector in a meaningful way, recall (in Sect. 3) that the magnitude of the image gradient $||g||$ can also be considered as the edge strength as $\Lambda_{\text{max}}$ (which can
be any one of $\Lambda_{\text{MSE}}^\text{max}$, $\Lambda_{\text{WMW}}^\text{max}$, and $\Lambda_{\text{KS}}^\text{max}$ that we try to compare). Besides, non-maximum suppression can also be applied for both $\|g\|$ and $\Lambda_{\text{max}}$. Hence the non-maximum suppressed edge strength maps for different detectors are used for performance evaluation.

As for the evaluation criterion, several methods have been proposed in the literature [26]–[29]. In particular, Martin et al. [29] proposed to use the precision-recall (PR) curve for edge detection evaluation. “Precision” is defined as the fraction of detections that are true positives (TP’s) rather than false positives (FP’s), while “Recall” is the fraction of true positives that are detected rather than missed. Besides, they also used the F-measure as a summery evaluation parameter, which is given by:

$$F = \frac{P \cdot R}{\alpha P + (1 - \alpha) R}$$  \hspace{1cm} (14)

In this paper, we adopt this PR curve approach. The PR curve is drawn at each point calculated with a given threshold, and each detector is trained so that the maximum F-measure is obtained. In this way, we only compare detectors that are tuned to their best performance. To sum up, the evaluation procedure is as follows.

1) For each detector with a certain set of parameters, the non-maximum suppressed edge strength image is first generated.
2) The non-maximum suppressed edge strength image is compared with a set of thresholds to obtain a series of binary edge maps.
3) Each of the binary edge maps is compared with CT image (See Fig. 3) to calculate a point $(P, R)$ on the PR curve. In order to allow small error tolerance, for each truth edge point in CT image, the nearest detected edge point in the binary edge image within a certain tolerance of the truth edge point is counted as TP but only for once. In our experiment, we allow an edge displacement within a $3 \times 3$ neighborhood.
4) The detector is trained so that the best PR curve is obtained. The best PR curve is defined as the one with the maximum F-measure. In our experiment, we set $\alpha = 0.5$ in (14) so that “precision” and “recall” are weighted equally.
5) The best PR curves for each detector are compared in the same PR plane and the one that gives the maximum F-measure has the best performance.

4.2 Results with CT Image

Figure 3 shows example image (CT image) and its corresponding edge strength image.

The following three types of noises are added to the original image to verify the performances of different edge detectors in presence of different noises.

1) The image is contaminated by the additive white Gaussian noise (AWGN) with zero mean and a standard deviation of $\sigma = 15$.
2) The image is contaminated by the logarithm of the uniform noise (LUN) with zero mean and a standard deviation of $\sigma = 15$. The LUN is defined as: $u = \sigma \ln n$, where $u$ is a uniform random variable within the interval of $[0, e]$. It can be easily proved that $n$ has a zero mean and a standard deviation of $\sigma$. The use of this noise type is legitimated by the fact that when the image is contaminated by multiplicative noise with uniform distribution, a logarithmic transform is usually applied to produce additive noise, which is LUN.
3) The image is contaminated by a mixture of AWGN and the salt & pepper noise. The AWGN is same as 1) and 10% percent of the image pixels are corrupted by the salt & pepper noise. We will denote this noise as the mixture noise (MN) hereafter.

The last two cases represent typical non-Gaussian noises that may be commonly encountered. The first row of Fig. 5 shows the different noisy images.

In the training process, for the Canny detector, the only parameter that needs to be trained is the smoothing scale of the Gaussian low pass filter, which we now denote as $\sigma_C$. The threshold, on the other hand, becomes a parameter of the PR curve and thus need not be trained. For the proposed MSE edge detector, we only train on the size of the scanning window. Other parameters, such as the possible edge directions $d$ and the width of the kernel function $h$, are fixed. Let $d = 8$ and $h = 2\sigma$, where $\sigma$ is the image noise level. For the other test statistic based detectors, the size of the scanning window is also the only parameter that needs to be trained.

Table 1 illustrates the training process of the Canny detector for the noisy CT image. For the image contaminated by the AWGN, the best smoothing scale is achieved at $\sigma_C = 0.8$, with $F = 0.4979$; for the image contaminated by LUN, the best scale is achieved at $\sigma_C = 0.8$, 

![Fig. 3 CT image and edge strength image.](image)

| $\sigma$ | AWGN | LUN | MN | F-measure |
|---------|------|-----|----|-----------|
| 0.7     | 0.4916 | 0.4917 | 0.2853 |
| 0.8     | 0.4979 | 0.4994 | 0.3294 |
| 0.9     | 0.4954 | 0.4974 | 0.3647 |
| 1.0     | 0.4925 | 0.4923 | 0.3866 |
| 1.1     | 0.4893 | 0.4915 | 0.4018 |
| 1.2     | 0.4874 | 0.4895 | 0.4116 |
| 1.3     | 0.4835 | 0.4783 | 0.4178 |
| 1.4     | 0.4814 | 0.4737 | 0.4147 |

The first column is the smoothing scale of the Gaussian filter; the last column represents the F-measure of the Canny detector with different smoothing scales for the CT image contaminated by AWGN, LUN, and MN, respectively.
Table 2: Training process of the MSE detector.

| t x r  | AWGN | LUN | MN  |
|--------|------|-----|-----|
| 3x3    | 0.4648 | 0.4694 | 0.4643 |
| 5x5    | 0.5154 | 0.5158 | 0.5063 |
| 7x7    | 0.5048 | 0.4973 | 0.4922 |
| 9x9    | 0.4963 | 0.4788 | 0.4824 |

The first column is the size of the scanning window; the last three columns represent the F-measure of the MSE filter with different scanning windows for the CT image contaminated by AWGN, LUN, and MN, respectively.

Table 3: Training process of the WMW detector.

| t x r  | AWGN | LUN | MN  |
|--------|------|-----|-----|
| 3x3    | 0.1717 | 0.1629 | 0.1694 |
| 5x5    | 0.4592 | 0.4282 | 0.3586 |
| 7x7    | 0.4712 | 0.4447 | 0.3883 |
| 9x9    | 0.4791 | 0.4462 | 0.4058 |
| 11x11  | 0.4551 | 0.4263 | 0.3996 |
| 13x13  | 0.4223 | 0.4028 | 0.3873 |

The first column is the size of the scanning window; the last three columns represent the F-measure of the WMW filter with different scanning windows for the CT image contaminated by AWGN, LUN, and MN, respectively.

Table 4: Training process of the KS detector.

| t x r  | AWGN | LUN | MN  |
|--------|------|-----|-----|
| 3x3    | 0.1715 | 0.1631 | 0.1692 |
| 5x5    | 0.4592 | 0.4258 | 0.4042 |
| 7x7    | 0.4765 | 0.4485 | 0.4440 |
| 9x9    | 0.4742 | 0.4436 | 0.4432 |
| 11x11  | 0.4594 | 0.4294 | 0.4363 |
| 13x13  | 0.4373 | 0.4126 | 0.4148 |

The first column is the size of the scanning window; the last three columns represent the F-measure of the KS filter with different scanning windows for the CT image contaminated by AWGN, LUN, and MN, respectively.

with $F = 0.4994$; and for the image contaminated by MN, the best smoothing scale is achieved at $\sigma_C = 1.3$, with $F = 0.4178$.

Table 2 shows the training process of our MSE detector for the noisy CT image. It is found that a window size of $5 \times 5$ achieves the best performance for all kinds of noises. The training process for other two test statistic (i.e., $\Lambda_{WMW}$ and $\Lambda_{KS}$) based detectors reveals that for different noisy CT image, a $9 \times 9$ scanning window is the best choice for the WMW detector and a $7 \times 7$ scanning window is the best for the KS detector (Table 3 and Table 4).

Figure 4 shows the overlaid best PR curves of different detectors for different noisy CT image; Fig. 5 shows the non-maximum suppressed edge strength images (after normalization) of different edge detectors with respective best parameters for different noisy CT image.

From the Fig. 4, we see that when the image is contaminated by AWGN and LUN, the MSE detector and the Canny detector in general out-perform the other two detectors, i.e., the WMW detector and KS detector. What is noteworthy is that both the MSE detector and the Canny detector performs as good with the non-Gaussian noise of LUN as with AWGN, while the WMW detector and KS detector suffer obvious performance loss from the transition of AWGN to LUN. This, in fact, partially demonstrates that the Canny detector enjoys a rather good robustness, if the noise is not drastically non-Gaussian. As for the comparison between the MSE detector and the Canny detector, we can see that there is a performance transition occurring at the crossing point of their PR curves. At the lower recall regions, the MSE detector performs better than the Canny detector, while at the higher recall regions, the Canny detector has a better performance. In general, the MSE detector performs slightly better than the Canny detector. The superiority of our MSE detector can be clearly seen in the detection performance with image contaminated by MN, as shown in the third column of Fig. 4. Our method not only significantly out-performs the other three detectors, but also
undergoes very small performance loss compared to the result with AWGN and LUN. For example, the F-measure of the MSE detector is 0.7063 with AWGN and 0.6969 with MN. That is only a 1.33% performance loss. On the other hand, the Canny detector suffers as much as 8.04% performance loss from 0.7005 to 0.6442. The other two detectors, the WMW detector and KS detector, although used to be believed to perform robustly under such condition of MN, actually show no evidence that they can give better results than the Canny detector.

5. Conclusion

In this paper we have presented a new edge detection method based on the use of the MSE two-sample test statistic that we have developed. This test statistic is in essence non-parametric by estimating the samples’ PDF with kernel estimators and calculating the MSE distance between PDFs.

The difficulty of extracting a single-pixel-wide edge is resolved by incorporating the non-maximum suppression technique to our edge detection scheme, which is generically applicable to any two-sample test statistic based edge detectors.

The main advantages of our method include the following points. First, it has as good detection performance as the Canny detection when the noise is Gaussian; second, it suffers very small performance loss across different noise types; third, our method can be easily extended to edge detection in color or multi-dimensional images. The last one is worth some more elaboration here. When dealing with color images, the only modification is to use vector samples and Euclidean norm in (6) as

\[
d(X, X) = \frac{1}{N_1^2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} e^{-\frac{||x_i - x_j||^2}{4h^2}}
\]

\[
d(X, Y) = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} e^{-\frac{||x_i - y_j||^2}{4h^2}}
\]

\[
d(Y, Y) = \frac{1}{N_2^2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_2} e^{-\frac{||y_i - y_j||^2}{4h^2}}
\]

where \(x_i\) and \(y_i\) are d-dimensional vectors and \(||.||\) denotes the Euclidean norm. This treatment makes the multi-dimensional extension of our method much easier than that of other methods [30] in the literature. Hence the future work may focus on the performance analysis of our method in noisy color images.

References

[1] M. Gudmundsson, E.A. El-Kwae, and M.R. Kabuka, “Edge detection in medical images using a genetic algorithm,” IEEE Trans. Med. Imaging, vol.17, no.3, pp.469–474, 1998.
[2] R.C. Gonzalez and R.E. Woods, Digital Image Processing, 2nd ed., Prentice Hall, 2002.
[3] J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol.PAMI-8, no.6, pp.679–698, 1986.
[4] R.R. Rakesh, P. Chaudhuri, and C.A. Murthy, “Thresholding in edge detection: A statistical approach,” IEEE Trans. Image Process., vol.13, no.7, pp.927–936, 2004.
[5] I. Abraham, R. Abraham, A. Desolnix, and S. Li-Thiao-Te, “Significant edges in the case of non-stationary Gaussian noise,” Pattern Recognit., vol.40, no.11, pp.3277–3291, 2007.
[6] I. Pitas, and A. Venetsanopoulos, “Nonlinear mean filters in image processing,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-34, no.3, pp.573–584, 1986.
[7] A. Benazza-Benyahia, J.C. Pesquet, and H. Krim, “A nonlinear diffusion-based 3-band filter bank,” IEEE Signal Process. Lett., vol.10, no.12, pp.360–363, 2003.
[8] M.A. Schulze, “An edge-enhancing nonlinear filter for reducing multiplicative noise,” Nonlinear Image Processing VIII, vol.3026, pp.46–56, 1997.
[9] H. Hwang and R.A. Haddad, “Multilevel nonlinear filters for edge detection and noise suppression,” IEEE Trans. Signal Process., vol.42, no.2, pp.249–258, 1994.
Appendix A

This appendix is to derive the first moment of the MSE test statistic $\Lambda_{\text{MSE}}$. For simplicity, we assume that each sample set has an equal number of samples of $N$, i.e., $N_1 = N_2 = N$ in (6). We also assume that the samples $x_i$ in $X$ are independently and identically distributed and $x_i \sim N(\mu_1, \sigma_1^2)$; the samples $y_j$ in $Y$ are independently and identically distributed as well and $y_j \sim N(\mu_2, \sigma_2^2)$.

According to (5), the mean value (first order statistic) of $\Lambda_{\text{MSE}}$ is

$$E(\Lambda_{\text{MSE}}) = E[d(X, X)] - 2E[d(X, Y)] + E[d(Y, Y)] \quad (A\cdot1)$$

where $d(X, X)$, $d(X, Y)$ and $d(Y, Y)$ are given by (6). It is easy to notice that $E[d(X, X)] = E[d(Y, Y)]$, so (A·1) becomes

$$E(\Lambda_{\text{MSE}}) = 2E[d(X, X)] - 2E[d(X, Y)] \quad (A\cdot2)$$

According to (6), the expected value of $d(X, Y)$ can be written as

$$E[d(X, Y)] = \frac{1}{N^2} \left[ N - \sum_{i=1 \atop i \neq j}^N E \left( e^{-\frac{|x_i-y_j|^2}{4\sigma^2}} \right) \right] \quad (A\cdot3)$$

Notice that when $i \neq j$, $x_i$ and $x_j$ are independent and both normally distributed with mean $\mu_1$ and variance $\sigma_1^2$. Thus

$$E \left( e^{-\frac{|x_i-y_j|^2}{4\sigma^2}} \right) \approx \frac{1}{2\pi\sigma^2} \int e^{-\frac{|x_i-y_j|^2}{4\sigma^2}} e^{-\frac{|x_i-y_j|^2}{2\sigma_2^2}} e^{-\frac{|x_i-y_j|^2}{2\sigma_2^2}} dx_i dx_j$$

$$= \frac{h}{\sqrt{\sigma^2 + h^2}} \quad (A\cdot4)$$

By substituting (A·4) into (A·3), the expected value of $d(X, Y)$ is

$$E[d(X, Y)] = \frac{1}{N} \left[ 1 - (N-1) \frac{h}{\sqrt{\sigma^2 + h^2}} \right] \quad (A\cdot5)$$

Similarly, we have

$$E[d(X, Y)] = \frac{h}{\sqrt{\sigma^2 + h^2}} e^{-\frac{(\mu_1-\mu_2)^2}{2\sigma_2^2(\sigma^2+h^2)}} \quad (A\cdot6)$$

As a result, the expected value of $\Lambda_{\text{MSE}}$ under the alternative hypothesis $H_1$ that $\mu_1 \neq \mu_2$ is

$$E(\Lambda_{\text{MSE}} | H_1) = \frac{2}{N} \left( 1 - \frac{h}{\sqrt{\sigma^2 + h^2}} \right) + \frac{2 h}{\sqrt{\sigma^2 + h^2}} \left[ 1 - e^{-\frac{(\mu_1-\mu_2)^2}{2\sigma_2^2(\sigma^2+h^2)}} \right] \quad (A\cdot7)$$

Let $\mu_1 = \mu_2$ in (A·7), we have the expected value of $\Lambda_{\text{MSE}}$ under the null hypothesis $H_0$:

$$E(\Lambda_{\text{MSE}} | H_0) = \frac{2}{N} \left( 1 - \frac{h}{\sqrt{\sigma^2 + h^2}} \right) \quad (A\cdot8)$$
Appendix B

By a simple relation of $h < \sqrt{\sigma^2 + h^2}$, the following inequality holds for (9):

$$RD(h) = \frac{N}{\sigma^2} h (\sqrt{\sigma^2 + h^2} + h) \left[ 1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4\sigma^2 + 4h^2}} \right]$$

$$\leq \frac{2N}{\sigma^2} (\sigma^2 + h^2) \left[ 1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4\sigma^2 + 4h^2}} \right] \quad (A\cdot9)$$

If we let $t = \sigma^2 + h^2$, then we denote the right hand side of the above inequality by $g(t)$ which is

$$g(t) = \frac{2N}{\sigma^2} \left[ t - te^{-\frac{(\mu_1 - \mu_2)^2}{4}} \right], t \geq \sigma^2 \quad (A\cdot10)$$

Now we show that $g(t)$ is monotonically increasing with respect to $t$ when $t > 0$. In fact, the first derivative of $g(t)$ is given by:

$$g'(t) = \frac{2N}{\sigma^2} \left[ 1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4}} - \frac{(\mu_1 - \mu_2)^2}{4t} e^{-\frac{(\mu_1 - \mu_2)^2}{4}} \right]$$

$$\quad (A\cdot11)$$

We further take the second derivatives of $g(t)$ and get

$$g''(t) = -\frac{2N}{\sigma^2} \cdot \frac{(\mu_1 - \mu_2)^4}{16t^3} \cdot e^{-\frac{(\mu_1 - \mu_2)^2}{4t}} \quad (A\cdot12)$$

It is obvious that when $t > 0$, $g''(t) < 0$; so $g'(t)$ is monotonically decreasing for $t > 0$. Then according to (A·11), $g'(t) > g'(0) = 0$. As a result, $g(t)$ is a monotonically increasing function when $t \geq 0$ and by L’hopital’s rule we have:

$$g(t) < g(+\infty) = \lim_{t \to +\infty} \frac{2N}{\sigma^2} \left[ t - te^{-\frac{(\mu_1 - \mu_2)^2}{4}} \right] = \frac{N(\mu_1 - \mu_2)^2}{2\sigma^2} \quad (A\cdot13)$$

Hence the upper bound of $RD(h)$ is $N(\mu_1 - \mu_2)^2/(2\sigma^2)$. Furthermore, it is also the supremum of $RD(h)$ because by applying L’hopital’s rule we also have:

$$\lim_{h \to +\infty} RD(h) = \lim_{h \to +\infty} \frac{Nh}{\sqrt{\sigma^2 + h^2}} \left[ 1 - e^{-\frac{(\mu_1 - \mu_2)^2}{4\sigma^2 + 4h^2}} \right]$$

$$= \frac{N(\mu_1 - \mu_2)^2}{2\sigma^2} \quad (A\cdot14)$$