Proceedings of the 2018 IISE Annual Conference
K. Barker, D. Berry, C. Rainwater, eds.

SEM Based Image Data Mining for Quality Control of Nanofiber Reinforced Piezo-Electric Nanocomposites

Abstract ID: 2425

Tzu-Liang (Bill) Tseng[1,*], Zhonghua Hu[2], Jianguo Wu[3], Anabel Renteria-Marquez[4], Yuxin Wen[2], Chun-Che Huang[5]

[1] Department of Industrial, Manufacturing and Systems Engineering, University of Texas at El Paso, TX 79968, USA
[2] Department of Electrical and Computer Engineering, University of Texas at El Paso, El Paso, TX 79968, U.S.A.
[3] Department of Industrial Engineering and Management, Peking University, Beijing 100871, China.
[4] Department of Mechanical Engineering, University of Texas at El Paso, TX 79968, USA.
[5] Department of Information Management, National Chi Nan University, Pu-Li, Na-Tau, Taiwan 54561, ROC.

Abstract

Nowadays nanotechnology has been applied into various fields. In manufacturing area, nanofiber reinforced Piezo-Electric nanocomposites can provide significant enhanced to mechanical and other functional properties. The morphology (e.g., spatial distribution, fiber orientation) of nanofibers in the base material plays an important role in determining the enhancement of material properties. SEM (scanning electron microscope) images are widely used in quality control of the nanofiber products. The objective of this paper is to find the nanofibers' critical information data (length, width, angle and location) through SEM images automatically, then these data would be utilized in the section of analysis in quality control process. In this study, several new different image data mining approaches would be implemented and compared to help improving the performance.

Keywords

Nanofiber, quality control, image data mining, SEM

1. Introduction

The demand for electronics developments, especially on portable and flexible devices for applications has recently increased. These applications could be integrated with the human body for health monitors, artificial muscles, sensor networks and engineered tissue constructs. On the other hand, in this "Industrial 4.0" era, the concept "Internet of Things" describes the applications of portable and wireless devices, which needs new power sources beyond rechargeable batteries. Moreover, the power source should be scalable, since in these areas, flexible, rugged and lightweight constructions are critical requirements. With the help of nanotechnology, new nanofiber materials like piezoelectric material provides the ability to bend and stretch, they are also attractive for pressure sensors and mechanical energy harvests[1]. Piezoelectric effect is a unique property of certain crystal which generate an electric field or current if subjected to a physical stress. Among piezoelectric nanofibers, piezoelectric zinc oxide (ZnO) is one of the earliest used to build the Nano-generators. In the later years, PZT piezoelectric materials have been developed. PZT is a ceramic material which shows good piezoelectric performance and is commonly used in nanofibers-based energy harvesters. These several years, PVDF (Polyvinylidene fluoride)[2] nanofibers have been developed, because they have a unique combination of material properties, including good flexibility, lightweight, biocompatibility and availability in ultra-long lengths for various thicknesses and shapes. These excellent properties make PVDF nanofibers be a good candidate for organic Nano-generator and energy harvesting applications in wearable and implantable devices.

Piezoelectric nanofibers have been widely used as reinforcement to form various nanocomposites with significantly enhanced mechanical and other functional properties. The morphology (e.g., spatial distribution, fiber orientation) of nanofibers in the base material plays a critical role in determining the enhancement of material properties. In structural applications, uniform distribution of nanofibers in terms of both spatial location and orientation is desirable to achieve the best mechanical properties. However, in some other applications, alignment of nanofibers with the same orientation is more preferable. For example, in the dielectric nanocomposites manufacturing used in high performance capacitors, well-aligned nanofibers could increase both permittivity and breakdown strength, and thus increase the energy density of capacitors[3].

The standard quality inspection technique that is currently used is morphology analysis of nanofibers imbedded in the base material based on microscopic images, e.g., scanning electron microscope (SEM) images. The morphology analysis is often based on visual inspection of SEM images, which is subjective and therefore could lead to a lack of quantitative measurement of nanofibers. The objective of this research is to find several different methods to detect and collect important information such as length, width, location and orientation of the nanofibers through the SEM image automatically. Once this information is collected, it could be used in quality analysis. In this paper, we focus the research on the line shape nanofibers SEM images. In other words, we need to detect the lines and extract the metrics of them in the SEM image.

2. Literature Review

The image processing requires us to take huge amounts of low level pixel data and extract useful information. There are a lot of mechanisms of extracting lines exist today.

A line is a collection of edge points that are adjacent and have the same direction. Hough Transform[4] is proposed to detect straight lines in the image, and then it extend to arbitrary shapes like circle and ellipse.

The basic idea of Hough Transform is to transfer a point from the physical domain to a curve in its parameters' domain. A line in x-y plane can be uniquely defined by its distance r from the origin and the orientation angle θ as $x cos\,\theta + y\,sin\,\theta = r$, as shown in Figure 1. Therefore, this parameterization maps every line in x-y plane to a point (θ, r) in $\theta - r$ domain. For a point P_0 (x_0, y_0), all lines passing through it could be expressed as $x_0\,cos\,\theta + y_0\,sin\,\theta = r$, where r is the distance from the origin to the line, and θ is the angle. Consequently, the point P_0 is mapped from (x_0, y_0) to a curve in the $\theta - r$ domain. Based on this transform, for points lying on the same line in x-y plane, their corresponding sinusoidal curves in $\theta - r$ plane will pass through a common point, which is the line in the $\theta - r$ domain.

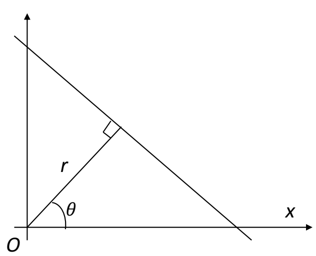

Figure 1: Illustration of Hough Transform

In the Hough transform algorithm, the θ-r parameter space is divided into accumulator cells to form a two-dimensional matrix. The value of the cell is the number of times a mapping line of points in x-y space intersects that cell. Cells receiving a minimum number of "votes" are assumed to correspond to lines in x-y space.

To extract nanofibers, we could use Hough Transform to detect lines in SEM images. However, due to the large width, a nanofiber could result in many lines in detection. To overcome this issue, "Skeleton" operation could be applied to get the morphological skeleton and then apply Hough Transform.

(a) (b)

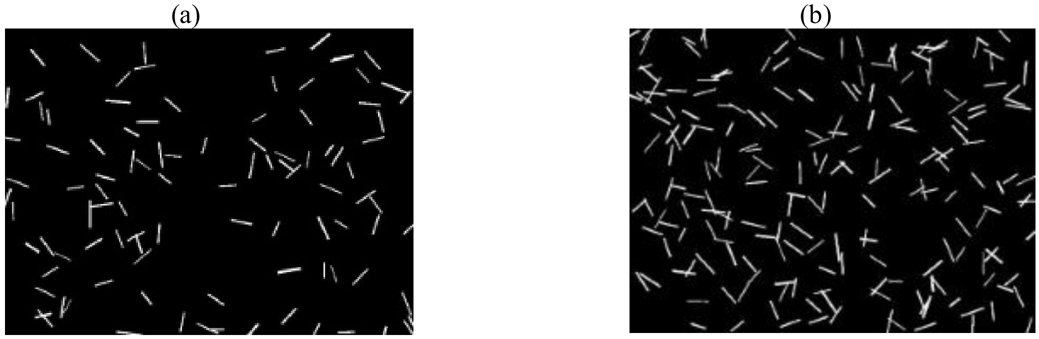

Figure 2: Illustration of simulated SEM images (a) 100 nanofibers (b) 200 nanofibers

In our project, we used Matlab to create two simulation image of SEM images, one contained 100 lines (nanofibers) and other contained 200 lines (nanofibers). The image resolution was set to 2400×1800 pixels. The orientation of

the nanofibers follows a uniform distribution. Also, the locations of nanofibers follow complete spatial randomness. The length and width of each nanofiber follow a normal distribution with a mean of 10 and 120 and standard deviation of 2 and 20 respectively. The simulation images display in Figure 2.
Detection results after applying the Hough Transform method are shown in Figure 3, the straight lines indicate the nanofibers. It could be seen that the accuracy drops when the density of nanofibers increases; the reason of that is because of the complexity. As the number of nanofibers increase in one image, there would be more overlapping. Then after the "Skeleton" operation, the shapes of the pattern would change dramatically.

(a) (b)

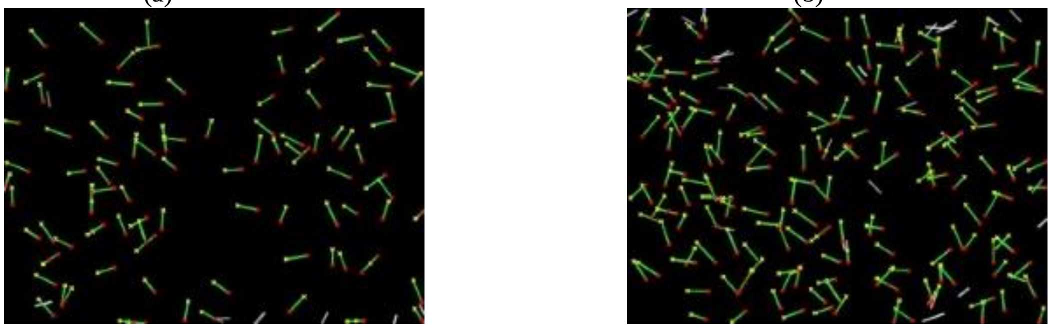

Figure 3: Hough transform detection results for (a) 100 nanofibers (b) 200 nanofibers

3. Methodology

The previous study and research of some general methods show that they could be used to find the lines (nanofibers) through the SEM image, but the result is not good enough. In this chapter, a novel method would be adapted to solve the issue/problem in the original methods and able to improve the performance. This method change the problem from a spatial image processing problem to a statistical problem. The detection process adapt a data mining approach to get the critical feature (change points), through these change points, we may extract the lines in some specific ways.
This statistical method included three major procedures.
First step: use "Moore Boundary Tracking Algorithm" to get the boundary data of each non-connected components in the SEM image. As a matter of fact, segmentation or morphology methods can also extract the boundary data, the reason using "Moore Boundary Tracking algorithm" is that the output will be an ordered sequence of points, which could be utilized in the following step. Once the boundary data has been generated, each edge point will contain an X and Y coordinates value.
Second step: adapt the change point method to detect changes of the X and Y value of the edge points. In statistical analysis, change detection or change point detection tries to identify the time when the probability distribution of a stochastic process or time series changes. In general the problem concerns both detecting whether or not a change has occurred, or whether several changes might have occurred, and identifying the times of any such changes.
Since in some complex SEM image, especially for some high-density Nanofiber SEM image, there could be some intersections or overlapped by several nanofibers; the change points in the boundary data could provide some critical information, for example the corners of one line or the intersection of two or more lines. Each change point indicates an abrupt steep trend change in this series of points. Our goal is to obtain these change points.
For m number of change points $\tau_{1:m} = (\tau_1, \dots, \tau_m)$ in an ordered sequence of data $y_{1:n} = (y_1, \dots, y_n)$, each change point's location is a positive integer between 1 and $n-1$. The change points' order is assume that $\tau_i < \tau_j$ if and only if $i < j$. Then, the m change points will split the data into $m+1$ segments, with the ith segment including $y_{(\tau_{i-1}+1):\tau_i}$. The common way to search multiple change points is to minimize the Equation (1), in this equation, $C\left(y_{(\tau_{i-1}+1):\tau_i}\right)$ is the cost function of segment $y_{(\tau_{i-1}+1):\tau_i}$ and β is the penalty.

$$\sum_{i=1}^{m+1} \left[C\left(y_{(\tau_{i-1}+1):\tau_i}\right)\right] + \beta f(m) \tag{1}$$

Based on this equation, here are two widely used methods which support multiple change points detection. One is "Binary Segmentation" (BS)[5] and another is "Optimal Partitioning" (OP)[6].
Binary Segmentation method is an extension application by the single change point detection method, it starts with applying one start change point for the entire data set and check if there is an existing τ that satisfies Equation (2).

$$C(y_{1:\tau}) + C\big(y_{(\tau+1):n}\big) + \beta < C(y_{1:n}) \tag{2}$$

If Equation (2) is false, this indicate that there is no more change point in the data segment $y_{1:n}$, and the method stops here. Otherwise, there should be a change point detected and the method will continue to be applied in these two segments $y_{1:\tau}$ and $y_{(\tau+1):n}$, which have been split by the change point τ. Then check the equation again in each segment, repeat the previous steps until the equation is not satisfied, which means no further change point are detected. The recursive procedure will get an $O(n \ln n)$ computation cost. Compared with the normal method of Equation (1), it is more efficient. However, this method will have an accuracy issue, the change points detected are the local change point in each separated data segment, which may cause that they are not the globe optimal change points found by in Equation (1).

Another Optimal Partitioning (OP) method utilize the Equation (1), and just let $f(m) = m$, then Equation (1) can be rewrote as following Equation (3)

$$\sum_{i=1}^{m+1}\left[C\big(y_{(\tau_{i-1}+1):\tau_i} + \beta\big)\right] \tag{3}$$

In the paper of "An algorithm for optimal partitioning of data on an interval"[7], Jackson use this OP method to decide the last change point at the beginning. After that, compute the optimal value of the cost function to the cost for the optimal partition of the data prior to the last change point plus the cost for the last segment, which is the rest of the segment starts with the last change point. He used $F(s)$ to denote as the minimization from Equation (3) for the data set $y_{1:s}$ and $T_s = \{\tau: 0 = \tau_0 < \tau_1 < \tau_2 \dots < \tau_m < \tau_{m+1} = s\}$ to be the set of change points in the data set. Also, let $F(0) = -\beta$ as the initial status.

Hence, the $F(s)$ can be derived as following Equation (4).

$$\begin{aligned} F(s) \quad &= \min_{\tau\in T_s}\left\{\sum_{i=1}^{m+1}\left[C\big(y_{(\tau_{i-1}+1):\tau_i}\big) + \beta\right]\right\} \\ &= \min_{t}\left\{\min_{\tau\in T_i}\sum_{i=1}^{m}\left[C\big(y_{(\tau_{i-1}+1):\tau_i}\big) + \beta\right] + C\big(y_{(t+1):n}\big) + \beta\right\} \\ &= \min_{t}\left\{F(t) + C\big(y_{(t+1):n}\big) + \beta\right\} \end{aligned} \tag{4}$$

From the Equation (4), actually it contains a recursion to get the minimal cost for data set $y_{1:s}$ in terms of the minimal cost for data set $y_{1:t}$ when $t < s$. The recursion could be solved in an order of $s = 1,2 \dots, n$. Then the overall computation time of the solution of $F(n)$ is $O(n^2)$.

These two popular methods are the basis of the PELT (Pruned Exact Linear Time)[8] algorithm. PELT algorithm takes advantage of "Binary Segmentation" to "pruning" the iteration procedure of Optimal Portioning method to improve the computing efficiency. It assumes that there exists a constant K and for all $t < s < T$. There could be Equation (5) which similar with Equation (2).

If

$$F(t) + C\big(y_{(t+1):s}\big) + K \ \geq F(s) \tag{5}$$

Then, for $t < s < T$, t cannot be the optimal last change point prior to T forever.

With this theorem the PELT method can reduce the number of points which need to be calculated on their cost function. Obviously, if K is bigger, the less change point will be detected. Note here, if $K = 0$, the pruning operation will only be related to the penalty β. After this "pruning" operation, the total computation cost could be reduced to $O(n)$.

Since we want to detect the lines (nanofibers) in the SEM image, we use linear regression model's mean square error (MSE) as shown in Equation (6)

$$MSE = \sum_{j=\tau_{i-1}+1}^{\tau_i} \varepsilon_j^2 \Big/ (\tau_i - (\tau_{i-1} + 1)) \tag{6}$$

Hence, the MSE is the cost of between $(\tau_{i-1} + 1)th$ point and $\tau_i th$ point.

However, there comes to several issues if we used simple linear regression model directly into our points data. First, it will be different from the normal regression data, out data is from the location of the points of computer display system. In computer's hardware, this point is called pixel. The pixels like a matrix on the display facilities. Hence, in 1965, Bresenham develop the line represent mechanism on these discrete pixel based system. In Bresenham's theory, we may find that different from the normal line connect two points, the shapes of these line pixels look like the steps, therefore, in some sections of the pixels, the slope of these pixels are 90 degree or $\frac{\pi}{2}$ grad. Especially the original line's slope is large than 45 degree, these pixels' section of 90 degree slope would be more and longer. This would cause to the slope parameter become close to infinite, then fail to get the correct regression model which means the residual would be wrong, also the incorrect cost function would lead to the inaccurate change point detection.

The second problem is, the linear regression model only contains the residual between y and *y-predicted*. It does not include the error of x. Actually, the data points including the x and y coordinate value are independent values which means that there is no relation between these two values, hence we cannot use the linear regression model here. To overcome this problem, we proposed a workaround solution, which we named it "Synthesized 2D" linear regression model. This solution may also resolve the problem of "infinite slope" which mentioned before.

Synthesized 2D residual method computes the linear regression twice, which will be one for X-axis coordinate value data set and another for Y-axis coordinate value data set. This method transfers the original 2D linear regression model (Deming Regression) into two simple 1D linear regression model. In detail, using the series number of each point in the boundary data set as the regressor and the X or Y axis coordinate value as the dependent variable to build two simple linear regression models. With these two separated regression models, we may get each MSE based on X and Y data set. This method can detect the residual on both X and Y side. Also, since in each point, there is only one X or Y value, which prevent the problem of "infinite slope". We just need to sum up the two MSE to get the synthesized cost which could be used to calculate the cost function $C\left(y_{(\tau_{i-1}+1):\tau_i}\right)$.

Using the synthesized 2D residual method, we may find the change points of the boundary file, notice here, different penalty choices would conduct to different results. We may tuning the penalty to get the higher accuracy. Figure 4 (a-d) shows the different results based on different penalty β choices (from $\beta = 0.01, 0.1, 1, 2$). After comparing and tuning, the penalty $\beta = 0.1$ would be a better choice.

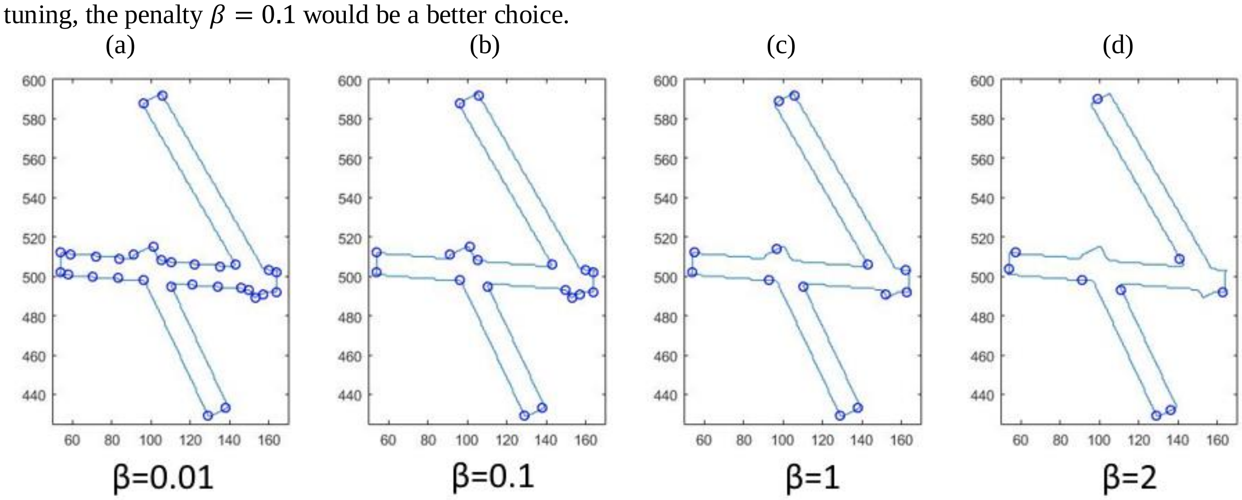

Figure 4: Change points detection when penalty (a) $\beta = 0.01$ (b) $\beta = 0.1$ (c) $\beta = 1$ (d) $\beta = 2$

Also, compared with OP method, the PELT can maintain the same detection result with much less computation time. Which means PELT method is much faster. Figure 5 illustrate the time consumption difference between OP and PELT methods with 5 different non-connected components' boundaries' points.

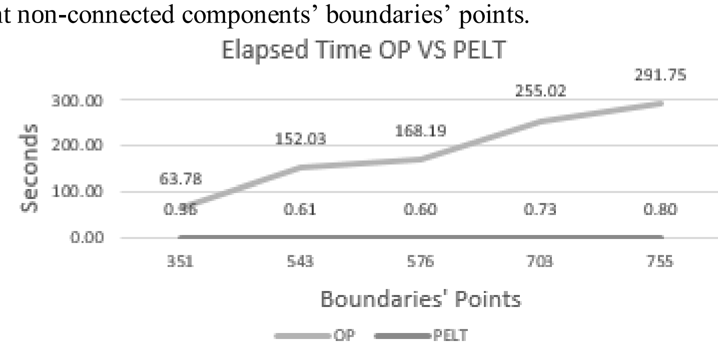

Figure 5: Time cost of OP method and PELT method

Thus, in the last step, we utilized some basic mathematics geometry method to extract the line through the change points detected in step 2. The 100 and 200 nanofibers simulation detection results are shown in Figure 6.

(a) (b)

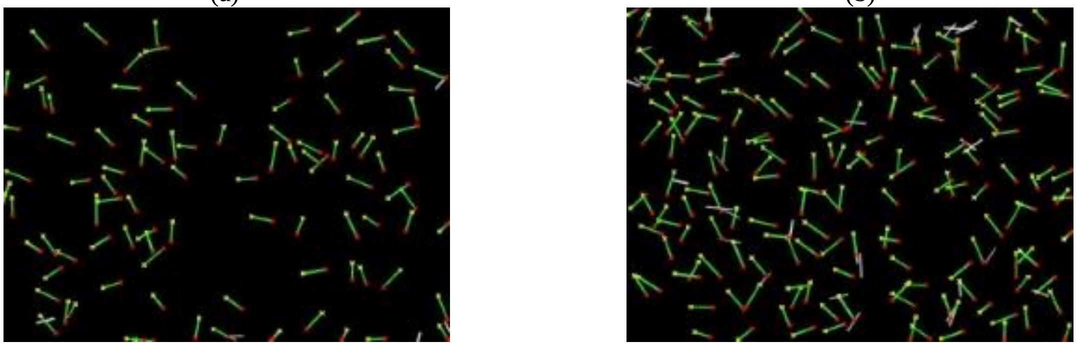

Figure 6: Statistical Method detection results for (a) 100 nanofibers (b) 200 nanofibers

Table 1 illustrates the performance results comparison between Hough Transform and Statistical Method. We find that the statistical method's accuracy is higher than the Hough Transform, especially for the high-density nanofibers simulation. Which prove that Statistical method is more robust.

Table 1: Performance of Hough Transform Method and Statistical Method in 100 and 200 nanofiber simulation

	Hough Transform Method detected	Accuracy of Hough Transform Method	Statistical Method Detected	Accuracy of Statistical Method
100 Nanofibers	90	90%	95	95%
200 Nanofibers	170	85%	184	92%

4. Conclusion

In conclusion, Hough Transform is very effective and efficient in the nanofiber segmentation, but in some complex condition, the performance may not be satisfied. The traditional image processing methods may not resolve this issue. Converting the pixel image to two ordered series data by extracting the boundaries of the components in the SEM image; data mining technology could be utilized to extract lines by changing point detection algorithm automatically. This method prevent the negative impact ("skeleton" operation) in Hough Transform method and can get better results. The improvement and application of this method to real SEM images will be left to our future work.

Acknowledgments

This work was partially supported by the National Science Foundation (DUE-TUES-1246050). The authors wish to express sincere gratitude for their financial support.

References

[1] A. Gheibi, R. Bagherzadeh, A. A. Merati *et al.*, "Electrical power generation from piezoelectric electrospun nanofibers membranes: electrospinning parameters optimization and effect of membranes thickness on output electrical voltage," *Journal of Polymer Research,* vol. 21, no. 11, pp. 1-14, 2014.

[2] L. Persano, C. Dagdeviren, Y. Su *et al.*, "High performance piezoelectric devices based on aligned arrays of nanofibers of poly (vinylidenefluoride-co-trifluoroethylene)," *Nature communications,* vol. 4, pp. 1633.

[3] H. Tang, Y. Lin, and H. A. Sodano, "Enhanced energy storage in nanocomposite capacitors through aligned PZT nanowires by uniaxial strain assembly," *Advanced Energy Materials,* vol. 2, no. 4, pp. 469-476, 2012.

[4] R. O. Duda, and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM,* vol. 15, no. 1, pp. 11-15, 1972.

[5] A. Sen, and M. S. Srivastava, "On tests for detecting change in mean," *The Annals of Statistics*, pp. 98-108, 1975.

[6] I. E. Auger, and C. E. Lawrence, "Algorithms for the optimal identification of segment neighborhoods," *Bulletin of mathematical biology,* vol. 51, no. 1, pp. 39-54, 1989.

[7] B. Jackson, J. D. Scargle, D. Barnes *et al.*, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Processing Letters,* vol. 12, no. 2, pp. 105-108, 2005.

[8] R. Killick, P. Fearnhead, and I. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association,* vol. 107, no. 500, pp. 1590-1598, 2012.