CIRP Journal of Manufacturing Science and Technology 33 (2021) 176–187

Contents lists available at ScienceDirect

CIRP Journal of Manufacturing Science and Technology

journal homepage: www.elsevier.com/locate/cirpj

# Automatic morphological extraction of fibers from SEM images for quality control of short fiber-reinforced composites manufacturing

Md. Fashiar Rahman[a], Jianguo Wu[b,*], Tzu Liang (Bill) Tseng[c]

[a] Computational Science Program, University of Texas at El Paso, TX 79968, USA
[b] Industrial Engineering and Management, College of Engineering, Peking University, Beijing, China
[c] Department of Industrial, Manufacturing and Systems Engineering, University of Texas at El Paso, TX 79968, USA

ARTICLE INFO

Article history:
Available online xxx

*Keywords:*
Fiber-reinforced composites
Fiber extraction
Hough transform
Opening
Morphological analysis

ABSTRACT

The properties of fiber-reinforced composite materials greatly depend on the morphology of reinforcing fibers within the base materials, i.e., spatial uniformity, orientation, and length distribution. Accurately extracting this information in an automated manner from SEM images is essential for quality assessment, quality control, and process optimization. However, due to overlapping or cross-linking issue, morphological fiber extraction is very challenging and has not been well addressed in the existing literature. This paper takes into account this research gap and proposes five different methods, namely, the opening method, simple Hough transform (HT), partitioning HT, gradient-based HT, and break-merge method to automatically extract the straight fibers from SEM images to facilitate the morphological analysis. The robustness and accuracy of the proposed methods are evaluated and compared through both simulation and real case studies.

© 2021 CIRP.

## Introduction

Micro/nanofibers are noble reinforcements to increase the mechanical properties of composite materials. Fiber-reinforced composites exhibit promising advantages, such as high strength, high stiffness, and lightweight in comparison with conventional materials. These outstanding properties have led to their desirability and exploration in a wide range of applications. The spatial homogeneity, length, and alignment orientation of fibers in the base material play a decisive role in determining the final properties of composites [1–3]. For instance, with the increase of mean fiber length, the tensile strength increases significantly [4]. Besides, the composites have stronger mechanical properties in the direction of fiber alignment than other directions [5]. The desired orientation and length distribution largely depend on the application of composites. In structural applications, uniform distribution of fibers in terms of both spatial location and orientation is desirable to achieve the best isotropic mechanical properties. However, in other applications, the alignment of fibers in one direction may be preferable. Literature reports that the alignment of dielectric fillers in the direction of the applied electric field can significantly enhance the dielectric properties of the base material, especially the dielectric permittivity and breakdown strength [6,7]. Besides the alignment, the spatial distribution is also crucial for material properties. Almost in all applications, the homogeneous spatial distribution of fibers in the specimen is required to achieve optimal performance. Nevertheless, similar to particle-reinforced nanocomposites [8–10], the clustering or aggregation of fibers often exists due to imperfectly controlled processing, which may significantly reduce the material properties [11]. Therefore, it is highly desirable to extract the fiber morphology for quality evaluation and process control.

Nowadays, the conventional quality inspection approach is to perform the morphology analysis of fibers through visual checking of the scanning electron microscope (SEM) images of the composite material. However, this approach is subjective and often very time-consuming. Besides, it may not be realistic to collect all the quantitative information of interest manually. Therefore, an automated fiber extraction method for morphological analysis is desirable. Image processing techniques have been successfully used for the morphology analysis of nanoparticles in the biomedical and material science field. For example, ImageJ, a popular freeware tool provided by NIH, is used for the extraction and quantification of fluorescence particles [12]. A Machine learning approach has also been introduced to locate the particles' position in the images based on Haar features [13]. Ellipse fitting techniques and watershed algorithm are used to segment uniformly distributed particles from the background [14,15]. The

* Corresponding author.
*E-mail address:* j.wu@pku.edu.cn (J. Wu).

https://doi.org/10.1016/j.cirpj.2021.03.010
1755-5817/© 2021 CIRP.

problem of segmenting overlapped particles is also well tackled [16]. However, there are very limited morphological extraction methodologies for fibers in the existing literature. Kimura *et al.* [17] proposed an algorithm to measure the root length of plants through skeleton or thinning operation. Kawabata et al. [18] developed an image processing technique to detect and count asbestos fibers using both color and shape information. Peng et al. [19] presented a set of image processing techniques to estimate the length, position, and orientation of nanowires in the fluidic workspace from optical section microscopy images. However, all of these methods are not applicable to fiber extraction of SEM images, where fibers are often overlapped or cross-linked with each other. Jeon et al. [20] developed a method to characterize the nanowire alignment in a microchannel using ridge detection, texton analysis, and autocorrelation function (ACF) calculation. The textons of different orientation angles are convoluted with the autocorrelation field to detect the distribution of wire alignment angles. This method can approximately estimate the orientation distribution. Nevertheless, it is not capable of identifying the fiber locations, which are essential for spatial homogeneity assessment.

To address the aforementioned issues, this paper develops and compares five different methods to automatically identify the fibers from SEM images. Of the five methods, one is based on the morphological opening operation, three are based on the Hough transform (HT) algorithm, namely, the simple Hough transform approach, partitioning Hough transform and gradient-based Hough transform, and the last one is based on the identification of cross points. Hough transform (HT) algorithm is a very efficient tool to identify a certain class of shapes, such as lines, circles, and ellipses, by a voting procedure [21]. The classic HT was used to detect straight lines in the image. Intuitively, the straight fibers can be identified by detecting the long boundaries or by detecting the skeleton after thinning process through the simple HT method. However, various issues make the simple HT not effective, especially when the fiber density is high. To solve these issues, this paper proposes two improved approaches, the partitioning Hough transform and gradient-based Hough transform. Besides, another innovative approach named break-merge (BM) method has also been proposed. In this method, the partitioning and morphological thinning operations are performed first, and then the DBSCAN clustering algorithm [22] is used to classify the skeleton into cross points and straight line points. After that, the cross points are removed to break connected fibers into shorter segments, and then the DBSCAN is used again to identify these segments as clusters. Finally, these short segments are matched or merged based on their distance and orientation to form complete fibers for morphological information extraction. The contribution of this paper is twofold. First, it develops four tailored techniques by nontrivially customizing the natural morphological operations, and also develops a brand new technique (BM) to fill the research gap of fiber extraction. Second, the shortcomings of these methods are discussed, the performance is evaluated and compared to provide insight and guidance for practitioners.

The rest of the paper is organized as follows. In Section "Methodologies for automatic fiber extraction", the technical details of the proposed methods are presented. In Section "Simulation study for performance evaluation and comparison", the proposed methods are evaluated and compared in terms of identification accuracy through simulation studies. The real case studies are given in Section "Application to real images". The conclusion and discussion are provided in Section "Conclusion and discussion".

## Methodologies for automatic fiber extraction

In this paper, we only focus on short fibers or fillers that could keep straight within the base materials. Therefore, these fibers could be approximately treated as line segments with width. Due to their "line" shape, the morphological operations for line detection could be intuitively employed, such as morphological opening operation and Hough transform. However, since the fibers have width and are not simple lines, directly employing these methods may not work. In this section, a customized opening method and improved Hough transform approaches will be introduced. Besides, the break-merge method will also be proposed as a better alternative.

### *Opening based fiber extraction*

This method is based on the morphological "opening" [23] operation, an important operator derived from the fundamental operations of erosion and dilation. It is a basic operation to remove small objects compared with the structuring element from the foreground and can be used to find specific shapes, such as edges, corners, and lines in an image. Mathematically, the opening of image $A$ by a structuring element (SE) $B$ (denoted by $A \circ B$) can be expressed as

$$A \circ B = (A \ominus B) \oplus B, \tag{1}$$

where $\ominus$ and $\oplus$ denote the "erosion" and "dilation" operation, respectively. Dilation is a morphological transformation that combines two sets using vector addition of set elements. If $A$ is a binary image in a Euclidian space ($E$) and $B$ is a structuring element, the dilation of $A$ by $B$, i.e., $A \oplus B$, is defined as

$$A \oplus B = \underset{b \in B}{\cup} A_b, \tag{2}$$

where $A_b$ is the translation of $A$ by $b$, i.e., $A_b = \{a + b | a \in A\}, \forall_b \in E$. Conversely, erosion is the morphological transformation that combines two sets using the vector subtraction of set elements. The erosion of $A$ by $B$ is defined as

$$A \ominus B = \underset{b \in B}{\cap} A_{-b}, \tag{3}$$

where $A_{-b}$ is the translation of $A$ by $-b$. The structuring element $B$ can be considered as a shape detector.

The opening based method is illustrated in Fig. 1. The structuring element B sweeps over the image $A$ and eliminates the objects that are smaller than SE using erosion operation. Later, the shape of the resulting objects is restored by deploying the dilation operation. However, restoring accuracy highly depends on the type of structuring element and the shape of restoring objects. In the proposed method, we employ line structuring elements of various orientation angles with certain lengths to extract fibers with the same orientation angles. The orientation of SE is varied from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ with the angle difference $\Delta\theta$.

The fibers which are shorter than the SE in the orientation of $\theta$ are removed. Since the length and width of fibers are often larger than those of SE, a single fiber may be detected multiple times by different SE's of close orientations. As illustrated in Fig. 2(a), each fiber has been detected multiple times (marked by boundaries of different colors). Hence, we need to merge these fibers or remove the duplicated segmentation. It is obvious that, if a fiber is opened multiple times, their centroids and the orientations will be very close. Based on this fact, we can check the closeness of the opened

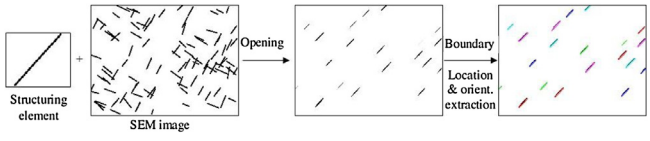



**Fig. 1.** The procedure for opening based method.

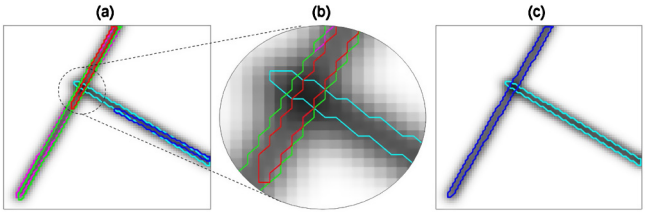



**Fig. 2.** Illustration of extraction duplication and fiber merging: (a) duplicated extraction using the opening method; (b) zoomed area showing duplicated extraction; (c) merging fibers with close centroid and orientation.

fibers and keep the desired one. The procedure is illustrated in Fig. 2 and the algorithm is provided in Table 1.

*Simple Hough transform based fiber extraction*

Hough transform is a mapping of a line from the spatial domain to another parameter space. It was first introduced by Paul Hough in 1962 [24]. Later it was extended to identify the arbitrary shapes, e.g., circles and ellipse, and it was named "generalized Hough transform" [25]. In Hough transform, a straight line is represented by

$$\rho = x\cos\theta + y\sin\theta \tag{4}$$

where $\rho$ is the length of the normal vector from the origin to the straight line, and $\theta$ is the orientation angle of the normal vector with respect to the $x$-axis. Based on this parameterization, each image point $(x, y)$ generates sinusoidal curves in $(\theta, \rho)$ space. The points on a particular line are mapped to the $(\theta, \rho)$ space to form many sinusoidal curves and these curves intersect at a common point $(\theta, \rho)$ that represents the line in $(\theta, \rho)$ space. The $\rho$–$\theta$ parameter space is subdivided into small accumulator cells to form a two-dimensional matrix, which is known as the Hough Transform bin. The parameter $\theta$ and $\rho$ are usually limited to $\pm\pi 2$ and $\sqrt{M^2 + N^2}$ respectively, where $(M, N)$ is the image size. This accumulator cell counts the number of sinusoidal curves that cross the corresponding point $(\theta, \rho)$. Thus, the resulting peaks in the accumulator array represent strong evidence that a corresponding straight line exists in the image. As illustrated in Fig. 3, three points {(19,1), (15,25), and (1,29)} generate three sinusoidal curves. Since these three points belong to the same line, they intersect at a common point $d$, which gives the parameter of $\rho = 20.5$ and $\theta = 45^\circ$ for the corresponding line.

To extract the fibers, we could use Hough transform to detect straight lines in SEM images. However, there are two major issues with the direct use of Hough transform. The first issue is that the fibers are not simply lines, but with width or thickness. Multiple

**Table 1**
Opening based fiber segmentation.

1. Convert the SEM image into a binary image
2. For different orientation angle $\theta$ ranging from $-\frac{\pi}{2}$ to $\frac{\pi}{2}$
   a. Perform opening operation with line structuring element of angle $\theta$ to extract fibers
   with orientation angles around $\theta$
   b. Extract the boundary of each fiber within the opened image
   c. Calculate the centroid and orientation of each extracted boundary
   End
3. For all the orientation angle $\theta$
   a. Check the pairwise distance of detected fibers based on their centroid and orientations
   b. If the distance of the fibers is below a certain threshold
   i. Treat them as a single fiber
   End

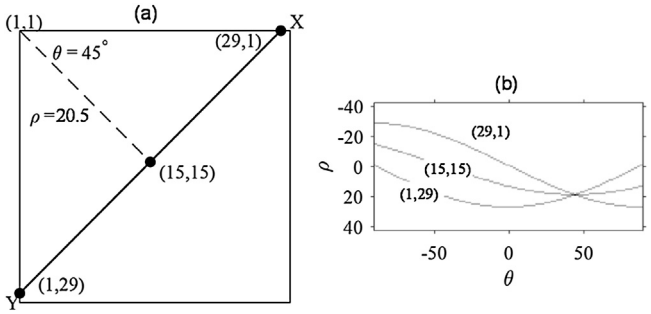



**Fig. 3.** Illustration of Hough transform: (a) three points on the same line; (b) three sinusoidal curves in the $(\theta, \rho)$ space mapped from the three points.

lines with different orientations and locations may be detected on the same fiber. Secondly, the accumulator cell can gather points that are on the same line but are actually from different fibers. To address the first issue, the "skeleton" operation [26] could be used to get the skeletonized image. The original structure is skeletonized by the use of a structuring element convoluted over the image. This method can successively erode pixels from the boundary until no more thinning (pixels removal) is possible. The orientation and location of the fibers are well preserved after the skeleton operation. Given a point set $A$, the skeleton operation $S(A)$ is defined as

$$S(A) = \bigcup_{k=0}^{M} S_k(A), \tag{5}$$

$$S_k(A) = (A \ominus kB) - [(A \ominus kB) \circ B], \tag{6}$$

where $M$ is the maximum iterative steps before $A$ erodes to an empty set, which is mathematically defined by Eq. (7), $k$ indicates how many times $A$ is eroded ($\ominus$) with the structuring element $B$, and ($\circ$) represents the opening operation:

$$M = \max\{k | (A \ominus kB) \neq \emptyset\}. \tag{7}$$

To overcome the second issue, the continuity of the points on the same line is evaluated by computing the distance between two line segments. Distances between two line segments associated with the same Hough transform bin are specified as a positive number. When the distance between the line segments is less than the value specified, the line segments are merged into one; otherwise, they are considered as separated lines. In our case, we set this threshold as 30$px$ through a trial-and-error process. This parameter needs to be adjusted while applying the HT algorithm depending on the images/applications.

*Partitioning Hough transform*

In the simple Hough transform approach, the "skeleton" operation is employed to alleviate the detection of multiple lines on a single fiber. However, there is still an issue that needs to be considered. The skeleton of other fibers still could contribute to the accumulator cell values and may significantly influence the detection accuracy, especially when the density of fiber is very large. To overcome this problem, we propose to add a partitioning step before applying the Hough transform algorithm.

The fibers are partitioned using the connected component labeling algorithm [27]. In this algorithm, the subsets of connected components (cross-linked fibers) are uniquely labeled based on the graph traversal method. Once the first pixel of a connected component is found, all the connected components are labeled before going onto the next pixel in the image. This results in linked

lists of the indices of the pixels that are connected to each other. Based on these lists of pixel indices, multiple images can be created, where each image contains a connected component. After that, the partitioned fibers are skeletonized by the morphological skeleton operation and then the simple Hough transform approach is applied to extract the fibers. The partitioning Hough transform approach is illustrated in Table 2.

*Gradient-based Hough transform*

As the density of fibers increases, there would be more fibers cross-linked with each other. The fiber crossing will significantly influence the shape of the skeleton, e.g., changing straight lines to curved lines, as shown in Fig. 4. Therefore, it would reduce the detection accuracy of the Hough transform method.

To solve this issue, another effective method, the gradient-based Hough transform [28] is employed. This method is an extension of the standard Hough transform. Instead of mapping each pixel of the fiber skeleton to a sinusoidal curve in the $(\theta, \rho)$ space, this approach maps each pixel of the fiber boundaries to a single cell based on its gradient. The rationale is that the gradient direction of the fiber boundary is approximately equal to the line parameter $\theta$ in Eq. (4). We only increase the corresponding accumulator cell with $\theta$ derived from the gradient, which can reduce the number of useless votes.

In this method, there are three steps. The first step is to extract the boundaries of all fibers or connected components with cross-linking. The second step is to calculate the gradient of each pixel of the boundaries. The third step is to map the boundary pixels based on the calculated gradient to the $(\theta, \rho)$ space. Boundaries can be easily extracted through various existing methods, such as "Freeman Chain Code" [29], "Minimum-Perimeter Polygon (MPP)" [30], and "Moore Boundary Tracing Algorithm" [31]. Let $g_{x_i}$ and $g_{y_i}$ be components of the gradient at $(x_i, y_i)$. As the boundary data set $B = \{(x_i, y_i)\}$ is a collection of sequential pairs of $x$ and $y$ coordinates along the boundary, we can easily measure the $g_{x_i}$ and $g_{y_i}$ by taking the difference between two points:

$$\left(g_{x_i}, g_{y_i}\right) = (x_{i+l} - x_i, y_{i+l} - y_i) \tag{8}$$

where $l$ is the step size. Then the gradient direction $\theta$ for pixel $(x_i, y_i)$ can be calculated by

$$\theta_i = \tan^{-1}\frac{g_y}{g_{x_i}} \tag{9}$$

After the gradient is obtained, the gradient-based Hough transform is applied to the boundaries to obtain the Hough matrix. On the boundary of each fiber, there are two long straight segments, which will result in two peaks with almost the same $\theta$ and slightly different $\rho$ in the accumulator matrix. We match these two peaks to identify the corresponding fiber. The detailed algorithm is illustrated in Table 3.

**Table 2**
Partitioning Hough transform for fiber extraction.

1. Convert the SEM image into a binary image
2. Partition the binary image into $n$ SEM images with each having one connected component.
3. For $i = 1{:}n$
   (a) Extract the morphological skeleton of image $i$
   (b) Perform Hough transform on image $i$ to get the Hough matrix
   (c) Identify the peaks of the Hough matrix
   (d) Detect Hough lines based on peaks and Hough matrix
   End

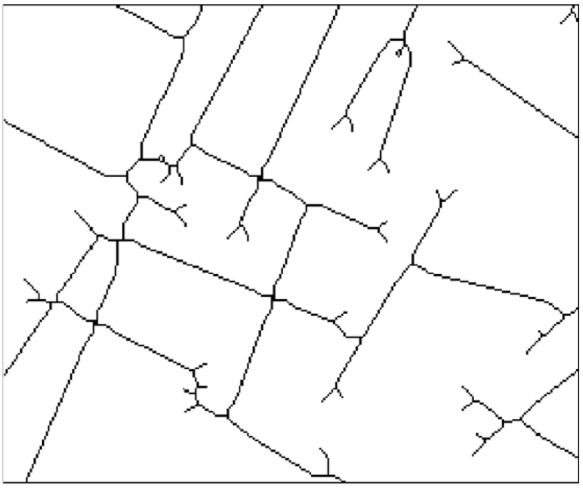


**Fig. 4.** The skeleton of multiple cross-linked fibers.

**Table 3**
Gradient-based Hough transform algorithm for fiber extraction.

1. Convert the SEM image into a binary image
2. Extract the fiber boundaries
3. For each of the boundary pixel $(x_i, y_i)$
   a. Calculate the gradient $\theta_i$
   b. Calculate the parameter $\rho_i = x_i\cos\theta_i + y_i\sin\theta_i$
   c. Increase the accumulator $A(\theta_i, \rho_i) = A(\theta_i, \rho_i) + 1$
   End
4. Detect the fibers based on the Hough matrix $A$

*Break-merge method for fiber extraction*

The key idea of the break-merge method is to first identify the cross points, and then break the cross-linked fibers at the cross points into multiple shorter segments, and eventually match the segments that are from the same fiber to obtain the orientation and location information. The main challenge lies in the identification of the cross points. To overcome this challenge, we propose to perform the partitioning and morphological thinning operation first to obtain the skeleton of each connected component, and then apply the DBSCAN clustering algorithm to classify the skeleton pixels into cross points and straight line points. DBSCAN was first introduced in 1996 by Ester et al. [22]. It requires two parameters, namely the radius of neighborhoods around a data point $p$, denoted by *eps*, and the minimum number of data points in a neighborhood to define a cluster, denoted by *minPts*. Using these two parameters, this algorithm categorizes the data points into three classes, i.e., core points (the $\varepsilon$-neighborhood contains at least *minPts* points), border points (has fewer than *minPts* points within its $\varepsilon$-neighborhood, but is in the neighborhood of a core point) and outliers (not a core nor a border point).

We set different values of the parameters based on the objectives at the break and merge step of this proposed method. Setting up the value of DBSCAN parameters is very critical, and they need to be selected through proper justification. In our case, the thinning operation makes this task very straightforward. Notice that, after thinning operation, the fibers become a single-pixel line except in the crossing and ending points. In the first step, we break the fibers at the crossing point by identifying crossing point clusters using the DBSCAN algorithm. From Fig. 5(a), it is obvious that any *line point* must have two neighbors, as fibers are nothing but a single-pixel line at this stage. However, the crossing point must have at least 3 neighbors. Therefore, for any point, if

there are at least 4 points within its neighborhood of diameter 3 pixels (including itself), then it belongs to a crossing-point cluster. Thus, we set the value of eps = 1.5 (radius of a neighborhood) and minPts = 4 (number of minimum points in the neighborhood) at the first step. Based on these parameters, the detected clusters (core and border points) are classified as cross points while the rest points (noise points) are labeled as the straight line points, as shown in Fig. 5. Note that in some rare cases, the endpoints may be classified as cross points. The reason is that the morphological skeleton around the endpoints may have two very short branches. However, treating the endpoints as cross points does not affect the fiber extraction much. The only influence is that the length of the extracted fiber may be slightly shorter than the original one.

At the next step, the crossing pixels are removed to break the connected components into multiple separated segments. Again, the DBSCAN algorithm is employed to label the segments as clusters, as shown in Fig. 6. Note that the clusters have been separated at the cross points region. Similar to the first step, for any point, if there are exactly three (3) points within its neighborhood of diameter 3 pixels (including itself), then it belongs to a line-points cluster as shown in Fig. 6(a) with different colors. For this reason, the DBSCAN parameters are set to *eps*=1.5 and *minPts*=3 at this step. Since a single fiber may be separated into two or even more clusters, we need to merge those clusters to accurately extract the fiber attributes. In this method, we test the difference among the segment orientations and center-to-center orientation (orientation of line segment formed by connecting the two centroids) of any two segments. If the differences between the orientation $\theta_i$ for segment $i$, $\theta_j$ for segment $j$ and center-to-center orientation $\theta_{ij}$ are below certain threshold, segment $i$ and segment $j$ are treated as from the same fiber. Note that it is possible that two segments from two different fibers may have exactly the same orientation and are also on the same line. In that case, these two segments may be mistakenly merged to form one fiber. However, due to the usage of the partitioning step, this case is very unlikely to happen in practice. The detailed algorithm is illustrated in Table 4.

The algorithms described in Section “Opening based fiber extraction” to “Break-merge method for fiber extraction” were implemented in MATLAB®. Notice that, converting the SEM images into binary images is the common step for all of these algorithms. This is an important step, as binary conversion greatly affects the task of fiber detection, especially for noisy images. Converting SEM images to binary images replaces all values above globally determined thresholds with ‘ls’ and setting all other values to ‘0s’. The threshold is determined using the Otsu's method [32], which chooses the threshold value to minimize the intraclass variance of the thresholded black and white pixels. The Otsu's method is one of the most successful and robust methods for image thresholding [33]. It can automatically find an optimal threshold based on the distribution of pixel values, and is very effective in distinguishing between the foreground and background. It uses a 256-bin image histogram to compute the threshold. The threshold could be different from image-to-image based on the pixel intensity. For a more detailed description, the reader is referred to [32].

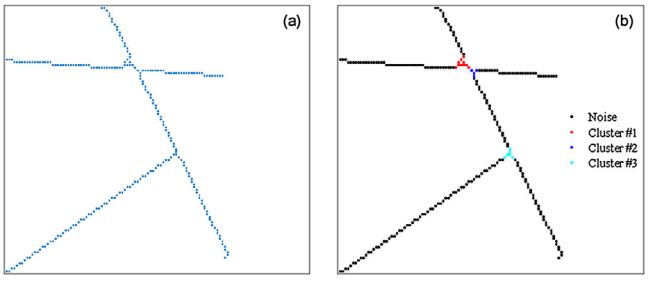



**Fig. 5.** Identifying the crossing points: (a) original skeleton and (b) crossing points identified as clusters.

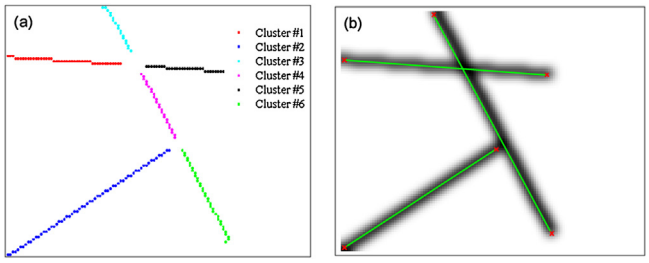



**Fig. 6.** Merging the line segments: (a) line points are clustered into different segments; (b) extracted fibers by merging the clusters.

## Simulation study for performance evaluation and comparison

This section is to evaluate and compare the five proposed methods through simulations. In the simulation, artificial SEM images with different fiber densities are randomly generated. The image resolution is set to 2400 × 1800 pixels. The fiber orientations are uniformly distributed between $-\pi/2$ to $\pi/2$. The length of fibers follows a normal distribution with a mean of 100 pixels and a standard deviation of 20 pixels. The fiber width remains fixed at 4 pixels. The centers of these fibers are uniformly distributed in the sample, which follows the complete spatial randomness (CSR) assumption widely used in spatial point analysis [34] The intensity of the gray level image at each pixel follows a truncated normal distribution within the range of 0–255 with $192 \pm 32$ (mean ± deviation). After the raw SEM image is generated, a 2-D Gaussian filter with a standard deviation of 2 is applied to smooth the fibers. Fig. 7 depicts four artificially generated SEM images with 50, 100, 200, and 400 fibers respectively. Note that in Fig. 7, the brightness of each pixel is reversed for illustration purposes.

The opening method, simple HT method, partitioning HT method, gradient-based HT method and break-merge method are applied to each artificial image. For the opening method, we tried different $\Delta\theta$s to see how this parameter influences the accuracy and execution time. Through our experiment, it is observed that the $\Delta\theta$ of 2 opens the fiber with reasonable accuracy and reduced execution time, as demonstrated in Fig. 8. With smaller $\Delta\theta$, though there is lower possibility to miss-detect the fibers, the accuracy drops due to the increasing error of merging fibers with close centroid and orientation. On the other hand, the larger $\Delta\theta$ fails to open some fibers, resulting in higher miss-detection rate. The execution time goes down linearly with the increase of $\Delta\theta$, which is expected since larger $\Delta\theta$ will result in less opened images. Therefore, in the study we choose $\Delta\theta = 2$.

Figs. 9–11 show the extraction results of the opening method, partitioning HT method, and break-merge method on the randomly generated artificial SEM images of four levels of fiber density. Due to space limitation, the extraction results for the other two methods are not shown here. The green line segments marked on the fibers represent the detected fibers where the red crosses specify the starting and the ending pixels of the detected fibers. For the undetected fibers, there are no green segments on the top. We can see that the break-merge method has identified almost all the fibers, while the other two methods missed several fibers when the density is high. To evaluate and compare these methods, three performance metrics are proposed: the miss-detection rate (*MDR*),

**Table 4**
Break-merge algorithm for fiber extraction.

1. Convert the SEM image into a binary image
2. Skeletonization by morphological thinning
3. Partition the skeletonized image into multiple images with each having one connected component
4. For each of the separated image
   (a) Apply DBSCAN to identify the cross points
   (b) Remove the cross points
   (c) Apply DBSCAN again to get separated clusters
   (d) Test the difference of segment orientations and center-to-center orientation of any two segments
   • If the orientation is below a certain threshold, merge them to form a single fiber
   End

i.e., the percentage of fibers missed, the mean-squared-error of the extracted positions ($MSE_l$), and the mean-squared-error of the extracted orientations ($MSE_o$). $MSE_l$ is defined as

$$MSE_l = \frac{1}{N}\sum_{i=1}^{N}\left[(\hat{x}_i - x_i)^2 + (\hat{y}_i - y_i)^2\right], \tag{10}$$

where $N$ is the number of detected fibers, $(\hat{x}_i, \hat{y}_i)$ and $(x_i, y_i)$ are the centroids of the $i$-th extracted and the original fiber respectively. $MSE_o$ is calculated as

$$MSE_o = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\theta}_i - \theta_i\right)^2, \tag{11}$$

where $\hat{\theta}_i$ and $\theta_i$ are the orientations of the $i$-th extracted and original fiber respectively.

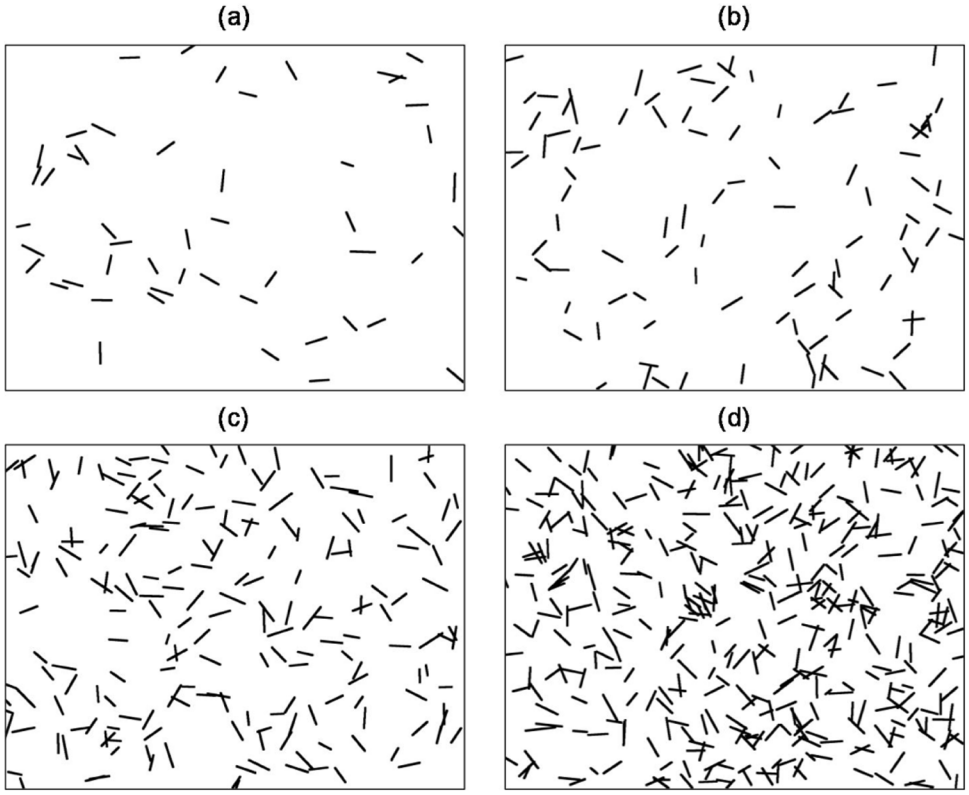



**Fig. 7.** Artificially generated images: (a) 50 fibers; (b) 100 fibers; (c) 200 fibers; (d) 400 fibers.

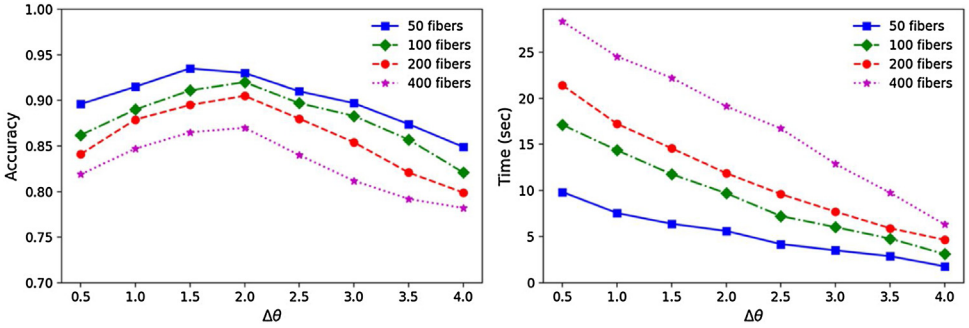



**Fig. 8.** The performance of the opening method at different $\Delta\theta$: (a) accuracy vs $\Delta\theta$; (b) execution time vs $\Delta\theta$.

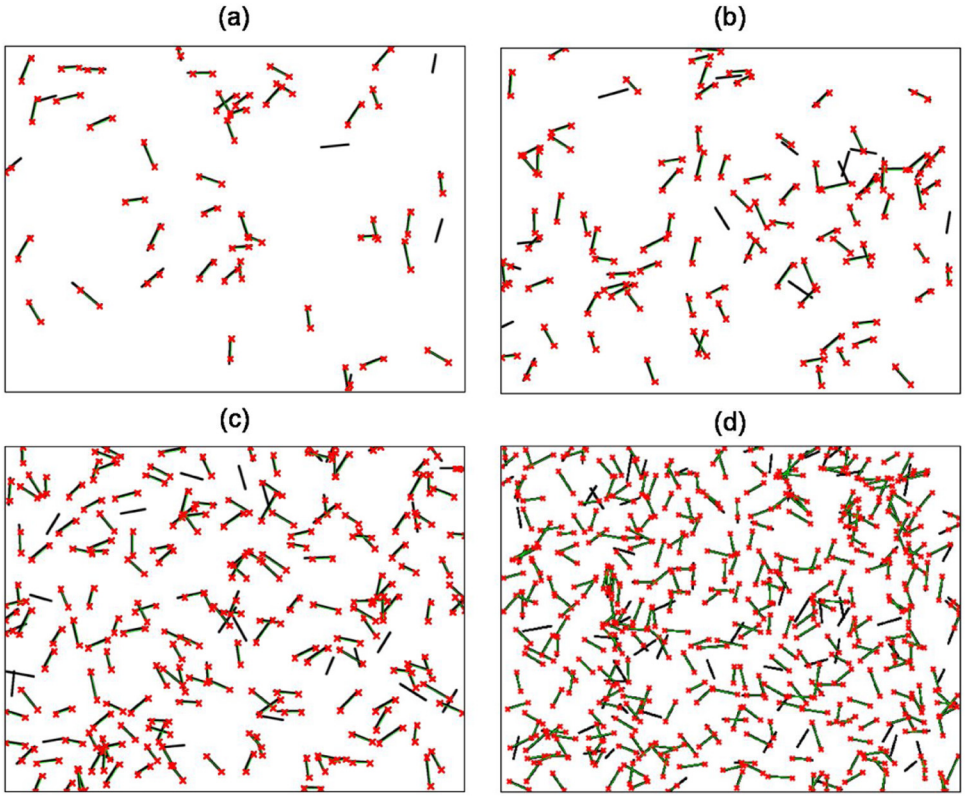



**Fig. 9.** Extraction using opening method: (a) 50 fibers, (b) 100 fibers, (c) 200 fibers, (d) 400 fibers.

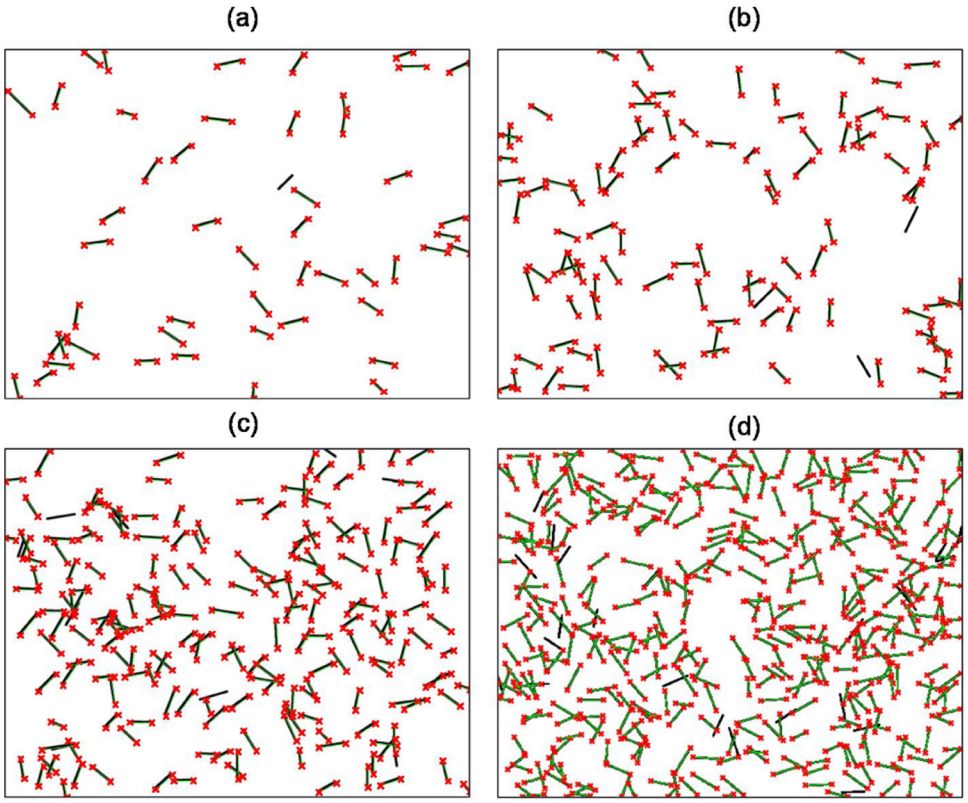



**Fig. 10.** Extraction using partitioning HT method: (a) 50 fibers, (b) 100 fibers, (c) 200 fibers, (d) 400 fibers.

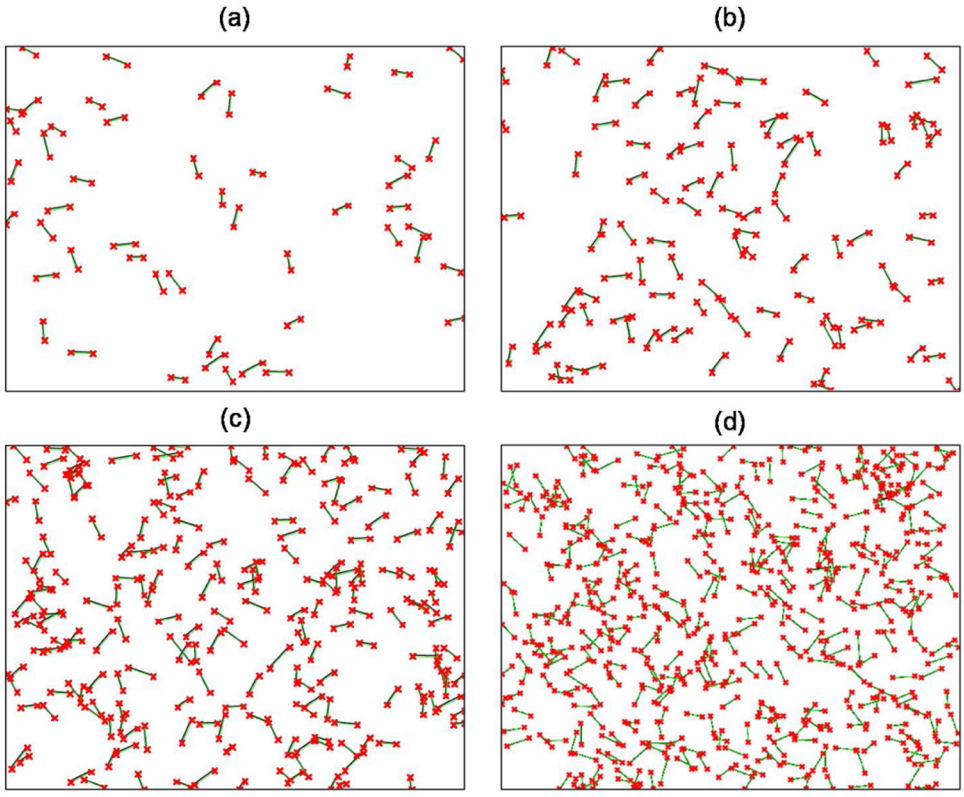



**Fig. 11.** Exaction using break-merge method: (a) 50 fibers, (b) 100 fibers, (c) 200 fibers, (d) 400 fibers.

The simulation is replicated 50 times for each of the four fiber densities (50, 100, 200, and 400 fibers). Fig. 12 shows the comparison of the five proposed methods in terms of the three evaluation metrics: *MDR*, $MSE_I$ and $MSE_o$. The detailed results are provided in Tables 5–7 respectively.

Clearly, the performance of all four methods decreases as the number of fibers increases. This is expected since increasing the number density will cause more over-lapping or crossing, and thus makes the fibers more difficult to segment. Of the five methods, the break-merge method outperforms all the other methods, while the simple HT method has the lowest extraction accuracy in almost all the cases. When the fiber density is low, the performance of the simple HT method is comparable to the opening method. However, as the fiber number density increases, the accuracy of the simple HT method decreases significantly. The reason is that increasing the fibers will increase the number of useless votes in the accumulator cells, i.e., pixels on the line but from fibers of different orientations

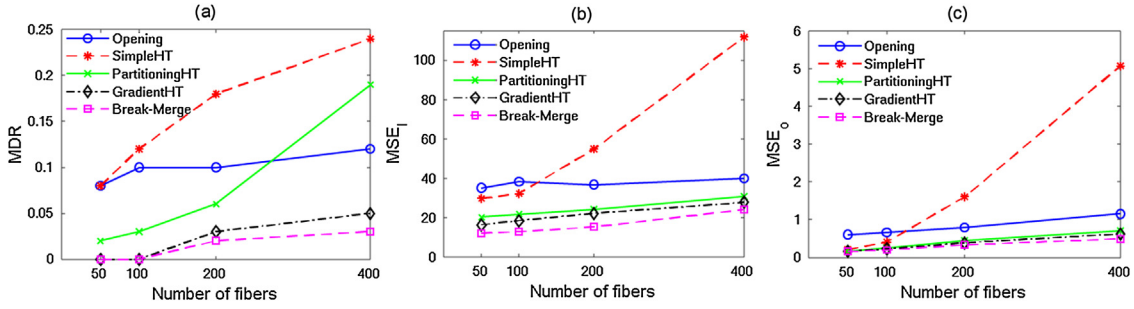



**Fig. 12.** Comparison of the five proposed methods in terms of *MDR*, $MSE_I$ and $MSE_o$.

**Table 5**
The MDR of the five proposed methods.

| No. of fibers | *MDR* | | | | |
|---|---|---|---|---|---|
| | Opening | Simple HT | Partitioning HT | Gradient-based HT | Break-merge |
| 50 | 0.08 | 0.08 | 0.02 | 0 | 0 |
| 100 | 0.10 | 0.12 | 0.03 | 0 | 0 |
| 200 | 0.10 | 0.18 | 0.06 | 0.03 | 0.02 |
| 400 | 0.12 | 0.24 | 0.19 | 0.05 | 0.03 |

**Table 6**
The MSE of extracted positions of the five proposed methods.

| No. of fibers | $MSE_l$ | | | | |
|---|---|---|---|---|---|
| | Opening | Simple HT | Partitioning HT | Gradient-based HT | Break-merge |
| 50 | 35.16 | 29.58 | 20.24 | 16.42 | 12.19 |
| 100 | 38.34 | 32.20 | 21.79 | 18.53 | 12.92 |
| 200 | 36.83 | 54.91 | 24.28 | 22.21 | 15.28 |
| 400 | 39.97 | 111.79 | 30.89 | 27.76 | 24.11 |

**Table 7**
The MSE of extracted orientations of the five proposed methods.

| No. of fibers | $MSE_o$ | | | | |
|---|---|---|---|---|---|
| | Opening | Simple HT | Partitioning HT | Gradient-based HT | Break-merge |
| 50 | 0.61 | 0.21 | 0.16 | 0.16 | 0.16 |
| 100 | 0.67 | 0.41 | 0.25 | 0.22 | 0.19 |
| 200 | 0.79 | 1.61 | 0.45 | 0.39 | 0.33 |
| 400 | 1.15 | 5.07 | 0.70 | 0.62 | 0.49 |

**Table 8**
Comparison of the execution time of the five proposed methods.

| No. of fibers | Execution time (Sec) | | | | |
|---|---|---|---|---|---|
| | Opening | Simple HT | Partitioning HT | Gradient-based HT | Break-merge |
| 50 | 5.42 | 2.60 | 8.95 | 10.32 | 3.82 |
| 100 | 9.84 | 4.36 | 16.76 | 18.56 | 7.14 |
| 200 | 12.74 | 7.93 | 30.61 | 33.98 | 14.09 |
| 400 | 18.78 | 14.58 | 53.67 | 59.62 | 26.72 |

contributing to the corresponding accumulator cell as well, which will reduce the visibility of the peaks corresponding to real fibers in the image. In comparison, a significant improvement is observed in partitioning based HT. Obviously, the partitioning step has effectively reduced the useless votes for the selected density levels. It is worth noting that if the density is extremely high, the partitioning step may not help, as most of the fibers may be connected to each other to form a large component. The gradient-based HT method performs much better than the simple and partitioning based HT. The reason is that incorporating the gradient information in the mapping process can more effectively eliminate the useless votes of other fibers of different orientations. The break-merge method has the highest fiber segmentation accuracy. It can effectively handle the overlapping or cross-linking issue by removing the crossing pixels and later merging the broken segments. The opening method outperforms the simple HT method at high fiber density levels, while it has lower extraction accuracy than all the other three methods. It is relatively more stable than the other methods. It is due to the probing nature of the structuring element in the opening operation.

The average computational cost of the extraction of fiber morphology is sown in Table 8. The execution time includes the image simulation and fiber morphology extraction. The above algorithms were implemented using MATLAB on a regular desktop with the configuration having Intel Core (TM) i5-6500 CPU @ 3.20 GHz and 8 GB RAM. It is observed that the simple HT method is less computationally expensive. As expected, the partitioning step over the HT method increases the computation time which is observed both for partitioning HT and gradient-based HT. Remember that, in gradient-based HT we calculate the gradient following the fiber partitioning step. The break-Merge method has a moderate computation efficiency with respect to opening, partitioning HT, and Gradient-based HT.

## Application to real images

In this section, the proposed break-merge method is applied to extract fibers from two real SEM images for morphological analysis. These two images are shown in Fig. 13(a) and (b), which are from Refs. [35,36], respectively. In Ref. [35], discontinuous pitch-based carbon fiber reinforced aluminum matrix (Al-CF)

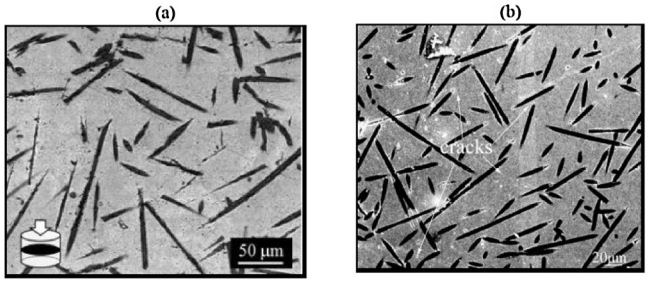



**Fig. 13.** Two real SEM images: (a) *AL–CF* composite (b) $C_f$/*SiC* composite.

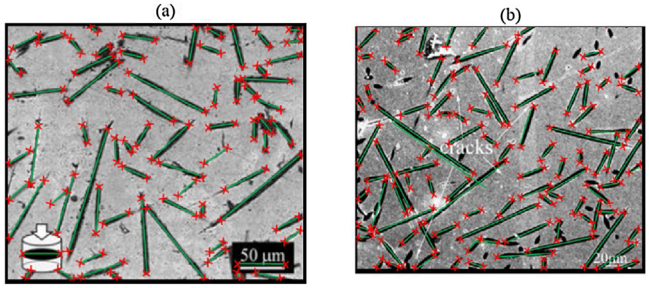



**Fig. 14.** Fiber extraction result for (a) *AL–CF* composite (b) $C_f$/*SiC* composite.

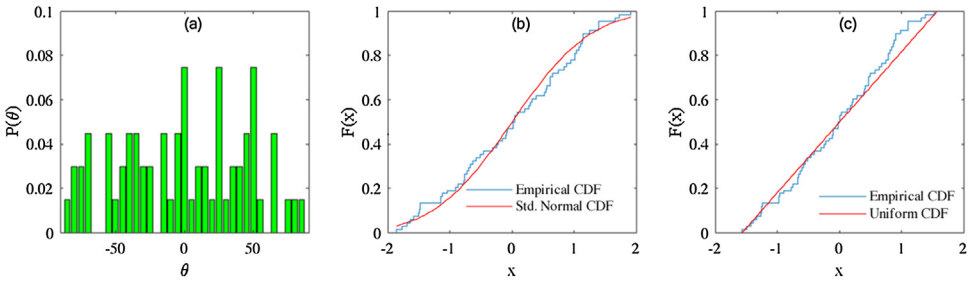



**Fig. 15.** (a) Histogram of the extracted orientations for the AL-CF composite; (b) empirical CDF and CDF of the standard normal distribution; and (c) empirical CDF and CDF of uniform distribution.

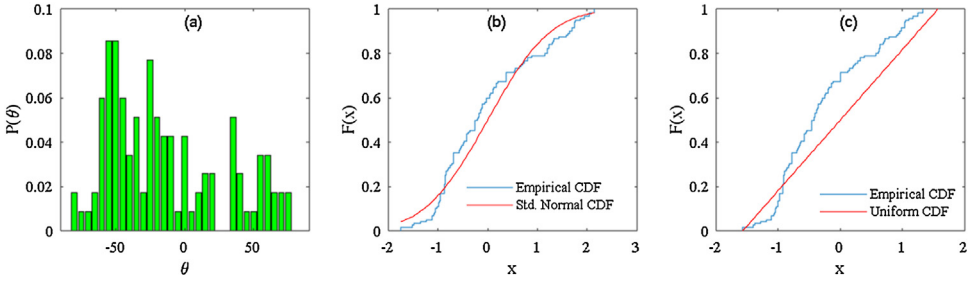



**Fig. 16.** (a) Histogram of the extracted orientations for the $C_f/SiC$ composite; (b) empirical CDF and CDF of the standard normal distribution; and (c) empirical CDF and CDF of uniform distribution.

composites with aluminium–silicon alloy (Al–Si) were fabricated through hot processing. The short carbon fiber and matrix powder were mixed for 5 min in air. Later, the mixed composite powder was hot-processed for 30 min at 600 °C under uniaxial compressive stress of 60 MPa. After the processing of the Al-CF composites, the microstructure is observed through scanning electron microscopy (SEM, Tescan, VEGA©). In Ref. [36], short carbon fibers were used as the reinforcement to fabricate $C_f/SiC$ composites by spark plasma sintering (SPS). The carbon fibers were firstly ultrasonically dispersed into the slurry to any possible fiber damage during the sintering process. After drying, the SPS samples were cut into 4 mm × 1.8 mm × 20 mm and polished to observe the surface by SEM. As the true locations and orientations are not available, we do not compare the five proposed methods here. We only use the break-merge method, whose superiority has been demonstrated in simulation study, to extract morphological attributes for subsequent analysis, e.g., testing the spatial and orientation uniformity.

The extracted fibers are shown in Fig. 14. There are 70 fibers detected on the first image ($MDR \approx 0$) and 91 fibers detected on the second image ($MDR \approx 2\%$), which are very accurate.

Based on detected fibers, we extract the centroids and orientation distribution for both images. Figs. 15(a) and 16(a) display the orientation distribution for AL-CF and $C_f/SiC$ composite respectively. We use the Kolmogorov–Smirnov (KS) test [37] to test whether the orientation follows a uniform or normal distribution. In the KS test, the alternative hypothesis indicates that $F_n(x)$ and $F(x)$ follow a different continuous distribution. The KS statistic for a given cumulative distribution function $F(x)$ is defined as

$$D_n = \sup_x |F_n(x) - F(x)|, \tag{12}$$

where $\sup_x$ is the supremum of the set of distances and $F_n(x)$ is the empirical distribution function. Clearly, $D_n$ is the maximum vertical difference between these two distributions function. If $D_n$ is above a certain threshold, the null hypothesis will be rejected. Here, we use the uniform and normal distributions as the candidate models to determine the extracted orientation distribution. To test the normal distribution, the observations are standardized first and then the empirical CDF is compared with that of the standard normal distribution. For the uniform distribution, the null distribution is set to be $U[-\pi/2, \pi/2]$.

Figs. 15(b) and (c) and 16(b) and (c) show the KS test results for the two SEM images respectively. The $p$-values of the test in Fig. 15(b) and (c) are 0.69 and 0.23 respectively. If we specify a 5% significance level, then both null distributions cannot be rejected. However, compared with uniform distribution, the normal distribution has a much larger $p$-value, indicating that the normal distribution is more appropriate to model the orientation distribution. For the second image, the $p$-values are 0.112 and $2.8 \times 10^{-4}$ respectively, indicating that the orientation distribution is not uniform if a 5% significance level is chosen. However, we cannot reject the assumption of fiber orientation being normally distributed.

The extracted locations of centroids are shown in Fig. 17. The spatial homogeneity of the centroids is examined, as it plays a decisive role in determining the fabrication quality in many applications. To justify whether the fibers follow complete spatial randomness (CSR), Ripley's $K$ function [38] is used, which is defined as

$$K(t) = \lambda^{-1} E\,(\text{number of centroids within distance } t \text{ of a randomly chosen centroid})$$

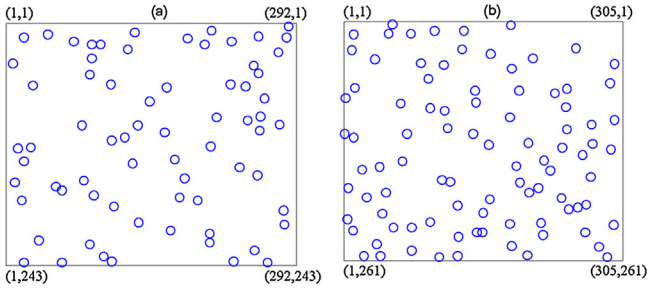



**Fig. 17.** Extracted fiber centroids for (a) AL-CF composite and (b) $C_f/SiC$ composite.

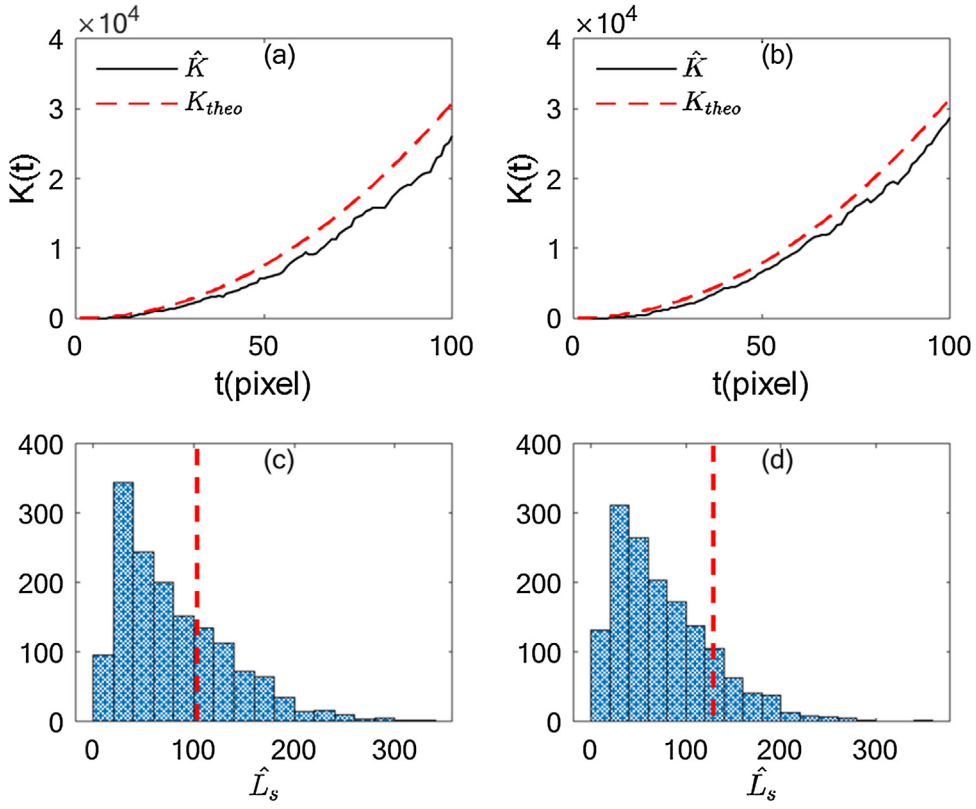



**Fig. 18.** CSR test using Ripley's *K* function. (a) and (b): the estimated and theoretical *K* functions for low and high number density image respectively; (c) and (d): the histograms of the simulated $L_s$ and observed $\hat{L}_s$ statistics denoted by vertical dashed lines for low and high number density images respectively.

Under CSR, it can be theoretically shown that $K(t)=\pi t^2$. Therefore, the difference between the theoretical $K(t)$ and the estimated $\hat{K}$ based on the observations can be used to test whether the fibers follow a homogeneous Poisson process. Alternatively, the variant $L(t)=[K(t)/\pi]^{1/2}$ is often used as $var\left[\hat{L}(t)\right]$ is approximately constant [39]. Under CSR, $L(t)=t$. To test whether $L(t)=t$, the test statistics $\hat{L}_s=\sum_t\left|\hat{L}(t)-t\right|$ is used. The distribution of $L_s$ under CSR is obtained through Monte Carlo simulations. Fig. 18 shows the theoretical *K* function under CSR, the estimated $\hat{K}$, the histogram of the simulated $\hat{L}_s$, and the observed test statistics for the extracted centroid location from the two images.

From Fig. 18(a) and (b), we can see that the estimated $\hat{K}(t)$ is very close to the theoretical $K(t)$. For the first image with low fiber density, the *p*-value of the hypothesis testing based on $L(t)$ is 0.647, which is not significant. Therefore, it is reasonable to consider that all the centroids are uniformly distributed with complete spatial randomness. For the second image with high fiber density, the *p*-value is 0.563, which is not significant as well. Consequently, we could not reject the null hypothesis that these fibers are uniformly distributed. We can justify this statement by looking into Fig. 17, i.e., the centroids are uniformly distributed.

## Conclusion and discussion

In this paper, we developed several automatic morphological extraction procedures to extract fibers from SEM images for quality assessment of fiber-reinforced composites manufacturing. Among them, one is based on the morphological opening operation, three are based on Hough transform, and the last one is the break-merge method, which first breaks the crossing fibers into segments from the crossing points, and then matches these short segments to form complete fibers. The performance of these methods has been evaluated and compared by using artificially generated SEM images of different fiber number densities in the simulation study. The results show that the simple Hough transform method has the lowest segmentation accuracy. The extraction accuracy increases after adding the partitioning step in the partitioning based HT method. However, its miss-detection rate continues to increase significantly as the fiber density increases. This problem is well tackled in the gradient-based HT method, which shows the best performance among the HT based methods. The break-merge method outperforms all the other methods in terms of all the three evaluation metrics, i.e., *MDR*, $MSE_l$ and $MSE_o$. The opening method is better than the simple HT method, while worse than all other methods. However, its performance is relatively stable with respect to the increase of fiber densities. In the real case study, the break-merge method is applied to two real SEM images. The extracted morphological information is further utilized for spatial and orientation uniformity test.

There are some open issues for future research in the fiber extraction from SEM images. First of all, in the current work, we only deal with short and straight fibers. However, in many fiber-reinforced composites, long and curved fibers of different widths and lengths may be cross-linked with each other. Secondly, when the fiber density is very high, two fibers may be aligned perfectly on a line, and be mistakenly merged into one fiber by the break-merge method. Finally, the resolution and contrast issue may pose an additional challenge to extract fibers from the background of the SEM image. How to accurately segment fibers and extract morphological information in such settings needs to be investigated.

## Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was partially supported by the Natural Science Foundation of China under Grant 51875003 and Grant 71932006, National Science Foundation (ECR-PEER-1935454), (ERC-ASPIRE-1941524) and Department of Education (Award # P120A180101).

## References

[1] Doshi, S., Charrier, J.M., 1989, A simple illustration of structure-properties relationships for short fiber-reinforced thermoplastics. Polym Compos, 10/1: 28–38.
[2] Yu, Z., Brisson, J., Ait-Kadi, A., 1994, Prediction of mechanical properties of short kevlar fiber-nylon-6, 6 composites. Polym Compos, 15/1: 64–73.
[3] Vaxman, A., Narkis, M., Siegmann, A., Kenig, S., 1989, Short-fiber-reinforced thermoplastics. Part III: effect of fiber length on rheological properties and fiber orientation. Polym Compos, 10/6: 454–462.
[4] Fu, S.-Y., Lauke, B., 1996, Effects of fiber length and fiber orientation distributions on the tensile strength of short-fiber-reinforced polymers. Compos Sci Technol, 56/10: 1179–1190.
[5] Frangopol, D.M., Recek, S., 2003, Reliability of fiber-reinforced composite laminate plates. Prob Eng Mech, 18/2: 119–137.
[6] Bowen, C., Newnham, R., Randall, C., 1998, Dielectric properties of dielectrophoretically assembled particulate-polymer composites. J Mater Res, 13/1: 205–210.
[7] Tang, H., Lin, Y., Sodano, H.A., 2012, Enhanced energy storage in nanocomposite capacitors through aligned PZT nanowires by uniaxial strain assembly. Adv Energy Mater, 2/4: 469–476.
[8] Wu, J., Zhou, S., Li, X., 2013, Acoustic emission monitoring for ultrasonic cavitation based dispersion process. J Manuf Sci Eng, 135/3: 031015.
[9] Wu, J., Zhou, S., Li, X., 2015, Ultrasonic attenuation based inspection method for scale-up production of A206-Al2O3 metal matrix nanocomposites. J Manuf Sci Eng, 137/1: 011013.
[10] Liu, Y., Wu, J., Zhou, S., Li, X., 2016, Microstructure modeling and ultrasonic wave propagation simulation of A206-Al2O3 metal matrix nanocomposites for quality inspection. J Manuf Sci Eng, 138/3: 031008.
[11] Bednarcyk, B.A., Aboudi, J., Arnold, S.M., 2015, Analysis of fiber clustering in composite materials using high-fidelity multiscale micromechanics. Int J Solids Struct, 69:311–327.
[12] Sage, D., Neumann, F.R., Hediger, F., Gasser, S.M., Unser, M., 2005, Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. IEEE Trans Image Process, 14/9: 1372–1383.
[13] Jiang, S., Zhou, X., Kirchhausen, T., Wong, S.T., 2007, Detection of molecular particles in live cells via machine learning. Cytometry Part A: J Int Soc Anal Cytol, 71/8: 563–575.
[14] Kothari, S., Chaudry, Q., Wang, M.D., 2009, Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. IEEE international symposium on biomedical imaging: from nano to macro 2009 ISBI'09, 795–798.
[15] Jung, M.-R., Shim, J.-H., Ko, B., Nam, J.-Y., 2008, Automatic cell segmentation and classification using morphological features and Bayesian networks. Image processing: machine vision applications vol 6813, 68130G.
[16] Park, C., et al, 2012, A multistage, semi-automated procedure for analyzing the morphology of nanoparticles. IIE Trans, 44/7: 507–522.
[17] Kimura, K., Kikuchi, S., Yamasaki, S.-i., 1999, Accurate root length measurement by image analysis. Plant Soil, 216/1–2: 117–127.
[18] Kawabata, K., Komori, Y., Mishima, T., Asama, H., 2009, An asbestos fiber detection technique utilizing image processing based on dispersion color. Part Sci Technol, 27/2: 177–192.
[19] Peng, T., Balijepalli, A., Gupta, S.K., LeBrun, T.W., 2009, Algorithms for extraction of nanowire lengths and positions from optical section microscopy image sequence. J Comput Inform Sci Eng, 9/4: 041007.
[20] Jeon, Y.J., Kang, H.W., Ko, S.H., Sung, H.J., 2013, Pattern analysis of aligned nanowires in a microchannel. Meas Sci Technol, 24/3: 035303.
[21] Illingworth, J., Kittler, J., 1988, A survey of the Hough transform. Comput Vision Graph Image Process, 44/1: 87–116.
[22] Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996, A density-based algorithm for discovering clusters in large spatial databases with noise. KDD vol 96, 226–231.
[23] Haralick, R.M., Sternberg, S.R., Zhuang, X., 1987, Image analysis using mathematical morphology. IEEE Trans Pattern Anal Mach Intell, 4:532–550.
[24] Hough PV. Method and means for recognizing complex patterns. Google Patents (1962).
[25] Ballard, D.H., 1981, Generalizing the Hough transform to detect arbitrary shapes. Pattern Recogn, 13/2: 111–122.
[26] Maragos, P., Schafer, R., 1986, Morphological skeleton representation and coding of binary images. IEEE Trans Acoust Speech Signal Process, 34/5: 1228–1244.
[27] Devijver, P.A., Ronse, C., 1984, Connected components in binary images: the detection problem. John Wiley & Sons, Inc.
[28] Petković, T., Lončarić, S., 2015, An extension to Hough transform based on gradient orientation.. a**rXiv:1510.04863.
[29] Saghri, J.A., Freeman, H., 1981, Analysis of the precision of generalized chain codes for the representation of planar curves. IEEE Trans Pattern Anal Mach Intell, 5:533–539.
[30] Sklansky, J., Chazin, R.L., Hansen, B.J., 1972, Minimum-perimeter polygons of digitized silhouettes. IEEE Trans Comput, 100/3: 260–268.
[31] Pradhan, R., Kumar, S., Agarwal, R., Pradhan, M.P., Ghose, M., 2010, Contour line tracing algorithm for digital topographic maps. Int J Image Process (IJIP), 4/2: 156–163.
[32] Otsu, N., 1979, A threshold selection method from gray-level histograms. IEEE Trans Syst Man Cybern, 9/1: 62–66.
[33] Sezgin, M., Sankur, B., 2004, Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging, 13/1: 146–165.
[34] Diggle, P.J., 2013, Statistical analysis of spatial and spatio-temporal point patterns. CRC Press.
[35] Kurita, H., Feuillet, E., Guillemet, T., Heintz, J.-M., Kawasaki, A., Silvain, J.-F., 2014, Simple fabrication and characterization of discontinuous carbon fiber reinforced aluminum matrix composite for lightweight heat sink applications. Acta Metall Sin (English Letters), 27/4: 714–722.
[36] Ding, Y., Dong, S., Huang, Z., Jiang, D., 2007, Fabrication of short C fiber-reinforced SiC composites by spark plasma sintering. Ceram Int, 33/1: 101–105.
[37] Massey Jr F.J.. 1951, The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc, 46/253: 68–78.
[38] Dixon, P.M., 2014, Ripley's K function. Wiley StatsRef: Statistics Reference Online.
[39] Wu, J., Yuan, Y., Gong, H., Tseng, T.-L., 2018, Inferring 3D ellipsoids based on cross-sectional images with applications to porosity control of additive manufacturing. IISE Trans, 50/7: 570–583.