

Shapelet-based orientation and defect identification method for nanostructured surface imaging

Matthew Peres Tino

Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario, Canada

Robert Suderman

Google Inc.

Nasser Mohieddin Abukhdeir

Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario, Canada

Department of Physics, University of Waterloo, Waterloo, Ontario, Canada

Waterloo Institute for Nanotechnology, University of Waterloo, Waterloo, Ontario, Canada

E-mail: nmabukhdeir@uwaterloo.ca

Abstract. Structure-property relations are of fundamental importance for continued progress in materials research. Determining these relationships for nanomaterials introduces additional challenges, especially when nanostructure is present, either through self-assembly or nano-lithographic processes. Recent advances have been made for quantification of nanostructured surfaces, for which many robust experimental imaging methods exist. One promising approach is based on the use of *shapelet functions* for image analysis, which may be used as a reduced-basis for surface pattern structure. These shapelet-based methods enable automated quantification of nanostructured images, guided by the user/researcher, providing pixel-level measures of local order without requiring detailed knowledge of order symmetries. In this work, enhancements to the existing shapelet-based *response distance method* are developed which enable further analysis of local order, including quantification of *local orientation* and identification of *topological defects*. The presented shapelet-based methods are applied to a representative set of images of self-assembled surfaces from experimental characterization techniques including scanning electron microscopy, atomic force microscopy and transmission electron microscopy. These methods are shown to be complementary in implementation and, importantly, provide researchers with a robust and generalized computational approach to comprehensively quantify nanostructure order, including local orientation and boundaries within well-aligned grains.

Keywords: nanostructure; surfaces; self-assembly; defects; shapelets; image processing; machine learning

Submitted to: *Nanotechnology*

1. Introduction

In contrast to relatively reliable and robust imaging-based characterization methods for nanomaterials [1], *post-processing* methods for this class of images are in a far less advanced state [2]. Post-processing methods enabling *quantification* of nanostructure are vital for scientific analysis and the development of structure/property relationships. These quantification tasks are relatively challenging due to the presence of complex nanostructures resulting from self-organization and self-assembly processes [2], in addition to the presence of measurement uncertainty and noise. However, analyzing ordered nanostructures and defects associated with this order is important to identify connections between nanomaterial synthesis, function, and application [3].

Recently there has been an increased focus on the presence of *defects* in nanomaterials and their resulting properties. Defects can be generally classified as orientational (disclinations) and translational (dislocations) [4]. Directly related to the presence of defects is variation of local nanostructure orientation, where interfaces between well-ordered regions (*e.g.* grains) are composed of defect structures (*e.g.* grain boundaries). These types of defects and defect structures, amongst others, are often inherent characteristics of self-assembly processes [5]. For example, they are intrinsic to liquid crystal phases and self-assembly materials such as colloids and block-copolymers [6].

Variation and manipulation of defects in nanostructured materials has been shown to enable the “tuning” of material properties (electrical, optical, magnetic, chemical) to obtain desired functionality [6, 7]. This approach, so-called *defect engineering*, has seen success in crystalline inorganic hard matter and organic hard materials [6]. However, while defect engineering in soft matter is less prevalent, recent work in this field has demonstrated significant promise. Defect engineering can be important for specific self-assembly applications, such as microelectronics, where material requirements may involve extremely low defect densities or specific defect placement within integrated circuits [5].

A recently developed method utilizing *shapelet functions* [8], originally developed for the analysis of images of galaxies, has proven useful for automated quantification of nanostructured surfaces [9, 10]. Suderman *et al* [9] showed that a subset of *polar shapelets* [11] can be used as a *local* reduced-basis to extract and compare a broad range of pattern orders observed in nanomaterial surfaces. This method is differentiated from past approaches, such as bond-orientational order analysis [2, 12], as it does not require image thresholding or segmentation and instead resolves pixel-scale detail of local order. This *response distance method* [9] is supervised, requiring the researcher or user to specify a reference subdomain in which there is well-defined local order, using this information to analyze the entire image. Results of applying the response distance method to simulated nanomaterial surface images are shown in Figure 1.

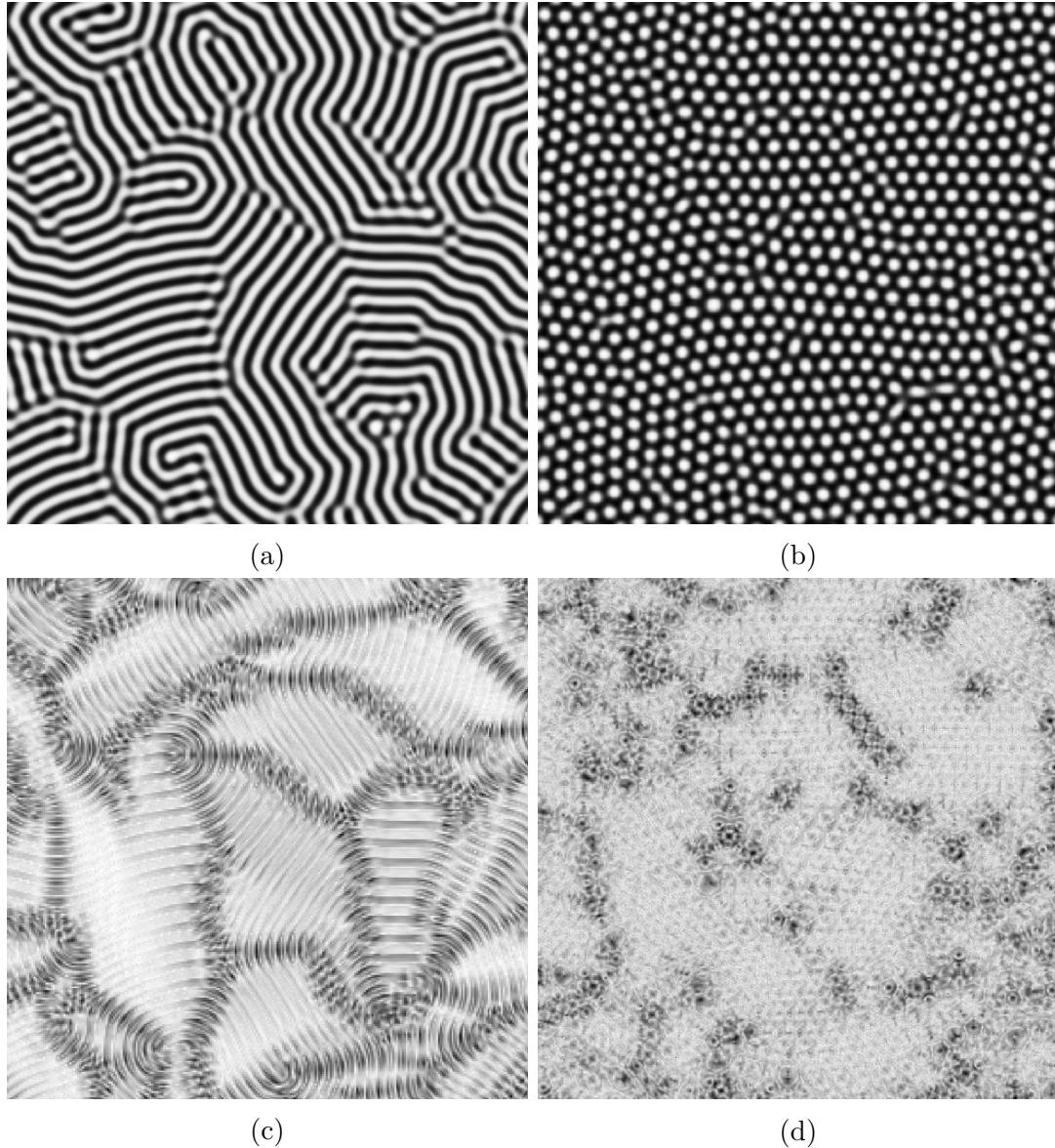


Figure 1: Simulated stripe and hexagonal self-assembly surface images and their response distance from [10]: (a) simulation result of a stripe self-assembly pattern. Reprinted with permission from [9]. (b) simulation result of a hexagonal self-assembly pattern. Reprinted with permission from [9]. (c) the response distance from [10] for the simulated stripe pattern in Figure 1a. Reprinted with permission from [10]. (d) the response distance from [10] for the simulated hexagonal pattern in Figure 1b. Reprinted with permission from [10].

The computational complexity of the original response distance method [9] was significantly reduced in later work [10], which also increased the robustness of the analysis in the presence of measurement uncertainty (noise). While these recently developed shapelet-based response distance methods have been applied to quantify the local degree of order, their utility to quantify local pattern orientation and identify

topological defects has not been adequately explored.

The overall objective of this work is to extend the shapelet-based response distance method [9] to include quantification of local pattern orientation and identification of topological defects. Given that the current approach is supervised by requiring user input to specify local pattern order, an emphasis on enhancements which reduce the possibility of incorrect user input is prioritized with specific objectives as follows:

- Analyze the effect of incorporating higher-order shapelets within the existing *response distance method* [9].
- Utilize shapelet orientation at maximum response, resulting from steerable filter theory [13], to determine local pattern orientation.
- Develop a defect identification method for self-assembly images which directly identifies defect surface patterns.

The paper is organized as follows: (i) Section 2 presents a background on the shapelet-based method for self-assembly imaging. (ii) Section 3 presents three sets of results: (a) incorporation of higher-order shapelets to reduce the noise associated with the *response distance method* [9], (b) quantification of local pattern order using shapelet orientation at maximum response, and (c) the development of a defect response distance method to identify topological defects with validation using stripe, square, and hexagonal pattern order. Finally, conclusions are presented in Section 4.

2. Shapelet-based pattern analysis

Shapelets are a class orthogonal functions originally developed for the decomposition and reconstruction of images of galaxies [8]. However, they have recently been shown to be useful to analyze patterns present in self-assembly imaging with a generalized method first presented by Suderman *et al* [9]. Several various shapelet formulations have been developed, with *polar shapelets* [11] found to be particularly suited for quantification of local pattern order.

Polar shapelets have several parameterizations, with parameters n and m corresponding to modulation in the radial and polar dimensions, respectively. Past work focused on self-assembly considers a subset of shapelets where $n = 0$ and $m \in \mathbb{N}$, which results in rapid radial decay. This subset of shapelets was further reformulated so that the radial scale is constant for increasing m and they are orthonormal, similar to an alternative formulation presented in eqn. 8 of ref. [11]. The constant scale orthonormal shapelet formulation used in this work is as follows [10],

$$S_m(r, \phi; \beta) = \frac{1}{\beta \sqrt{\pi m!}} \left(\frac{r}{\beta} \right)^m \exp \left(-\frac{r^2}{2\beta^2} - im\phi \right) \quad (1)$$

where the radial (constant) scale of the shapelet is λ and a rescaling of the standard polar shapelet is achieved through reformulation of the scale parameter,

$$\beta = \frac{f\lambda}{\sqrt{m}} \quad (2)$$

with f being a geometric factor mapped to specific m values; see ref. [10] for more details.

Figure 2 shows the real and imaginary components for $n = 0$ and $m \leq 8$ orthonormal polar shapelets with a length-to-image scale ratio of $\frac{\lambda}{N} = 0.25$. As is standard practice in image processing, shapelet functions are treated as *kernels* through projection onto a discrete domain, which then enables them to be used in convolution operations with target images. However, shapelet orientation has a strong effect on the resulting convolution, which was addressed by Suderman *et al* [9] who showed that shapelets are *steerable* [13]. Steerable kernels/filters may be decomposed into the following general form,

$$f^\theta(r, \phi) = \sum_{j=1}^M k_j(\theta) g_j(r, \phi) \quad (3)$$

where θ is the angle of rotation of the filter, $k_j(\theta)$ are rotation-dependent coefficients, and $g_j(r, \phi)$ are basis functions from a minimum set of M . The use of steerable formulations of filters enables computationally-efficient analysis of local orientation resulting in maximum (and minimum) response [13, Section V.A]. Setting the first derivative (w.r.t orientation θ) of the convolution of the steerable filter with an image function (I) to zero and solving the resulting expression can then be used to find both the maximum response of the steerable filter and the corresponding orientation at maximum response [14],

$$\frac{d}{d\theta}[I \otimes f^\theta(r, \phi)] = \sum_j^M \frac{d}{d\theta} k_j(\theta) [I \otimes g_j(r, \phi)] = 0 \quad (4)$$

where I is the target image (function) and \otimes represents the convolution operation.

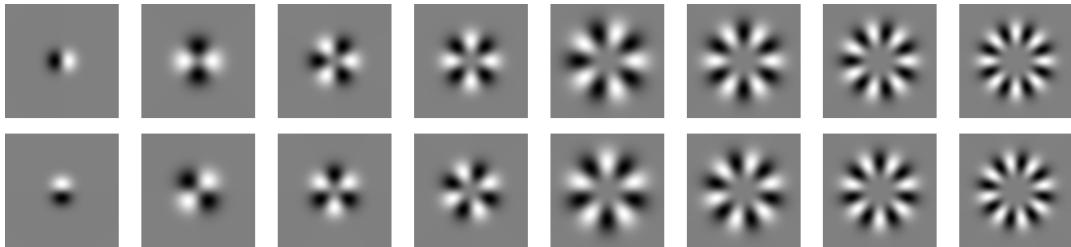


Figure 2: Real and imaginary basis shapelet components for $n = 0$ and $m \leq 8$ projected onto a discrete domain. All kernel windows are the same size and shapelet scale is fixed, highlighting the effect of the geometric factor f [10]. Length-to-image scale ratio is $\frac{\lambda}{N} = 0.25$.

Suderman *et al* [9] demonstrated that shapelets are steerable with $M = 2$,

$$S^\theta(r, \phi) = (\cos m\theta) \operatorname{Re}(S_m) + (\sin m\theta) \operatorname{Im}(S_m) \quad (5)$$

noting that this steerable formulation is real-valued. Based on refs. [13, 14] and solving eqn. 4 for the steerable shapelet in eqn. 5 (see Supplementary Material for more

details), the maximum shapelet response S_{max} (w.r.t orientation θ) and the orientation at maximum response (θ_{max}) for steerable shapelets are,

$$S_{max} = \sqrt{C_1^2 + C_2^2} \quad (6)$$

$$\theta_{max} = \frac{\arg(C_1, C_2)}{m} \quad (7)$$

where C_i is the result of convolving each steerable basis function with I , the target image (function).

The *response distance method* developed by Suderman *et al* [9] is a supervised image analysis method that quantifies local pattern order. The supervision required from the user involves the selection of a local reference region of the image in which the desired pattern is uniform (undeformed). A set of *response vectors* is then found for steerable shapelets centered at each pixel in the reference region. This reference set of response vectors is then used to compute the *response distance* for each pixel in the image, with the distance computed as the minimum of the 2-norm between the local response vector and all response vectors in the reference set,

$$d_{i,j} = \min \|\vec{R} - \vec{r}_{i,j}\|_2 \quad (8)$$

where $\vec{r}_{i,j}$ denotes the given response vector at pixel location i, j and \vec{R} is the reference set of response vectors.

Akdeniz *et al* [10] enhanced the response distance method by applying k -means clustering [15] on the user-specified reference set of response vectors to obtain a condensed representation with *cluster centroid vectors*. This reduces the set of reference response vectors from the number of pixels in the reference region (typically $10^2 - 10^3$) to a fixed set of centroid vectors (≈ 20). Applying k -means clustering [15] reduced the computational complexity of the method at least an order-of-magnitude, from $\propto 10^1$ to $\propto 10^0$ seconds) for the self-assembly images used both here and in past work [10].

3. Results and Discussion

The surface self-assembly imaging data used in this study is consistent with past work [9, 10], being a combination of simulation (Figures 1a-1b) and experimental images of nanostructured surfaces. Experimental imaging includes the use of a representative set of surface patterns (stripe, square, and hexagonal) and microscopy techniques including TEM (Figure 3a), AFM (Figure 3b), and SEM (Figure 3c) which include significant measurement noise.

The results presented in this work are divided into three different sections:

Higher-order shapelet basis – in this section the effect of including a higher-order ($m \geq 7$) shapelet basis is analyzed and compared to past response distance methods [9, 10], where a lower-order ($m \leq 6$) shapelet basis was used based on the maximum symmetry determined by the user.

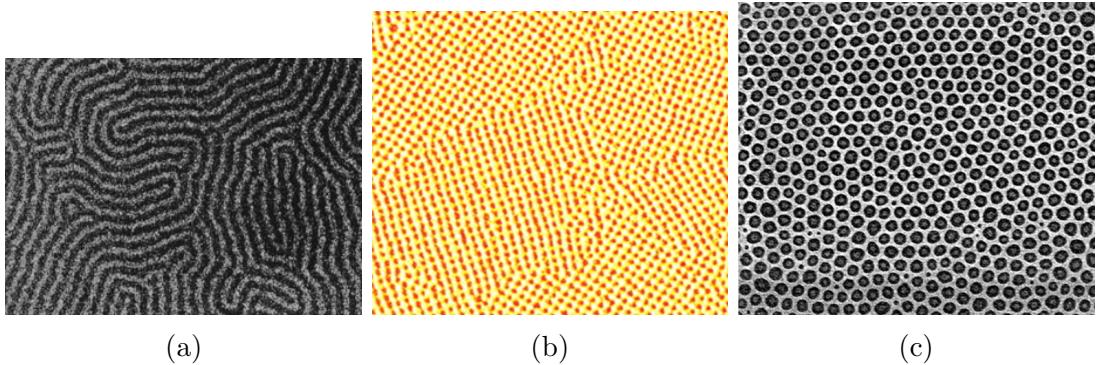


Figure 3: The set of simulated and experimental images of self-assembled surfaces used in this study: (Figure 1a) simulation results of a stripe self-assembled surface from [9], (Figure 1b) simulation results of heteroepitaxial surface self-assembly [9], (a) a TEM image of block copolymer surface self-assembly with stripe order [16], (b) an AFM image of a square patterned nanotemplate formed around surface self-assembled block copolymers [17], and (c) an SEM image of block copolymer surface self-assembly with hexagonal order [18]. From refs. [9, 16–18], reprinted with permission from APS, AAAS, AAAS, and Springer Nature, respectively.

Local pattern orientation – in this section a method is proposed and validated for the use of existing information from the response distance method, orientation at maximum response, to determine local pattern orientation without significant additional computational complexity. This is achieved through the use of steerable filter theory [13] and the “steerable” shapelet formulations presented in refs. [9, 10], as opposed to brute-force techniques, which involve explicitly computing a large number of shapelet (filter) orientations [19].

Defect identification method – in this section a defect response distance method is presented which identifies the presence of local pattern distortions, defects (dislocations, disclinations), and defect structures (grain boundaries [4]).

The analysis and methods presented are applied to a representative set of self-assembly images (Figure 3), however, it is important to note that all these images are both (i) inclusive of only one type of self-assembly pattern (stripe, square, or hexagonal) and (ii) do not include significant areas of disorder. An open-source software reference implementation provided in ref. [20] can be used to reproduce the presented results and apply these methods for the analysis of other self-assembly images.

3.1. Higher-order shapelets for pattern analysis

Previously developed shapelet response distance methods [9, 10] used lower-order ($m \leq 6$) shapelets based upon two considerations:

- from a qualitative perspective, dominant symmetries observed in surface self-assembly imaging range from 2-fold (stripe) to 6-fold (hexagonal),

- the magnitude of the response to shapelets of higher-order ($m \geq 7$) decreases rapidly with increasing order (m).

However, given the significantly reduced computational complexity involved in the shapelet-based response distance method presented in ref. [10], quantitative analysis of higher-order shapelet response can be considered. This is further motivated by the work of Massey *et al.* [11], which involved the use of higher-order shapelet expansions ($m \leq 20$) to optimize the tradeoff between galactic image reconstruction quality and computational complexity.

In order to quantitatively assess the overall contribution of higher-order shapelets, shapelet response vectors were computed (see Section 2) for the reference image set shown in Figure 3. The increase in computational complexity of the response distance method associated with the incorporation of higher-order shapelets was found to be negligible, $\approx 1 - 10\%$ of the lower-order shapelet computation time, depending on image size. For each image, the cumulative shapelet response for each individual shapelet component was compared, as shown in Figure 4. In agreement with past work, for all surface self-assembly images, lower-order shapelet responses are dominant in magnitude, however, the response for shapelets with $m = [7, 10]$ decays but is still significant. Additionally, for all experimental images, the rate of decay is much lower, with higher-order shapelet responses being $\geq 10\%$ of the maximum response up to $m = 29$. This is especially pronounced for experimental images of hexagonal surface self-assembly (Figure 3c).

These results show that higher-order shapelets do capture some information from self-assembly images, especially those with significant noise. This response behavior is similar to what is observed when applying the discrete Fourier transform (DFT) to images with considerable noise and/or spatial discontinuities, where coefficients for the DFT for higher frequency modes have increasing magnitude/moduli [21, Sec. 7.2].

To determine the local pattern features which are not captured using lower-order shapelets with the response distance method [10], analysis that included higher-order shapelets was performed and the response distance results were compared. Maximum shapelet order (m') was determined for each image by iteratively computing the cumulative shapelet response for increasing shapelet order, until the cumulative response was found to be below a normalized threshold value ($\epsilon = 0.1$) with respect to the maximum cumulative shapelet response. Referring back to Figure 4, the computed truncation index m' is highlighted in the caption for each of the five images in Figure 3.

The difference in response distance results is qualitatively (visually) similar and, consequently, these results are not shown here (see Supplementary Material for more details). After taking the difference in response distance using lower- versus including higher-order shapelet response and superimposing on the original images (Figure 5), the relation between higher-order shapelet response and local pattern features is more clear. Figure 5 shows the difference in response distance between using lower-order ($m \leq 6$) and the incorporation of higher-order ($m \leq m'$) shapelets overlayed on each of the self-assembly images. Focusing on this result for simulated self-assembly images of stripe

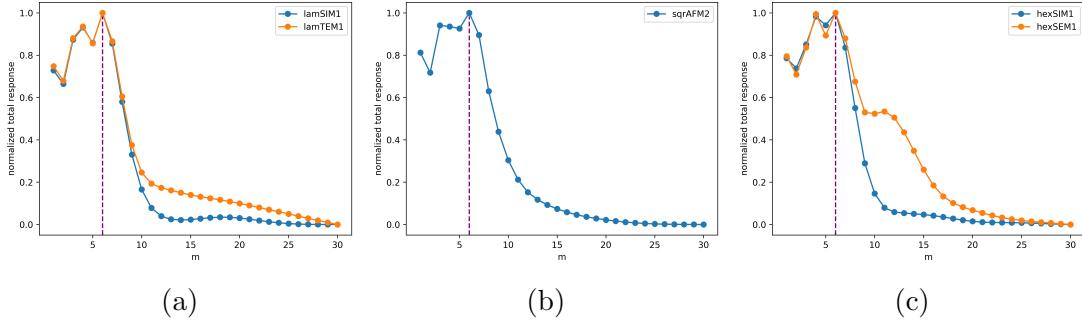


Figure 4: The normalized cumulative shapelet response for order $m \leq 30$ shapelets for the various self-assembly microscopy images in Figure 3. The cumulative shapelet response is the sum of the magnitude of responses over the entire domain. The dotted purple line at $m = 6$ highlights the upper bound for lower-order shapelet response in past work [9, 10]. The computed maximum shapelet order (m') is as follows: $m' = 10$ for both simulated stripe (Figure 1a) and hexagonal (Figure 1b) self-assembly images, and $m' = 29, 14$, and 23 for the experimental stripe (Figure 3a), square (Figure 3b), and hexagonal (Figure 3c) self-assembly images respectively.

(Figure 5a) and hexagonal (Figure 5b) patterns, higher-order shapelet-based response is observed to be significantly greater in areas with (i) pattern distortion (curvature or dilation) and (ii) topological defects. This is beneficial compared to the lower-order shapelet result in that, while regions with pattern distortion and defects have lower order than uniform regions, they still have order and are topologically different from regions *without* order (disorder). This is consistent with order parameter representations of local alignment in liquid crystalline phases [22], where order parameters are composed of two components, (order) magnitude and phase. Higher-order contributions to local order have also been shown to be important for quantification of alignment of nanostructures, such as nanorods and nanowires [23]. Topological defects, such as dislocations and disclinations, correspond to regions with non-zero order magnitude and *degenerate* phase [4]. This is most evident for the hexagonal simulation image (Figure 1b), where higher-order shapelet-based response directly corresponds to both disclinations (both plus and minus [4, Figure 9.2.19]) and dislocations, which are composed of adjacent pairs of oppositely charged disclinations for a hexagonal pattern.

Focusing on this result for experimental self-assembly images of stripe (Figure 5c), square (Figure 5d), and hexagonal (Figure 5e) patterns, higher-order shapelet-based response is observed to be significantly greater throughout the image. The difference in response between lower-order and higher-order shapelets can now be attributed to both (i) pattern distortion/defects and (ii) local uncertainty (noise). That is, in addition to significantly increased higher-order shapelet response in areas with distortion and defects, there is also difference in uniform regions where there is high-frequency noise. As with the simulated images, the difference in response distance for experimental images is most significant in areas directly corresponding to disclination defects.

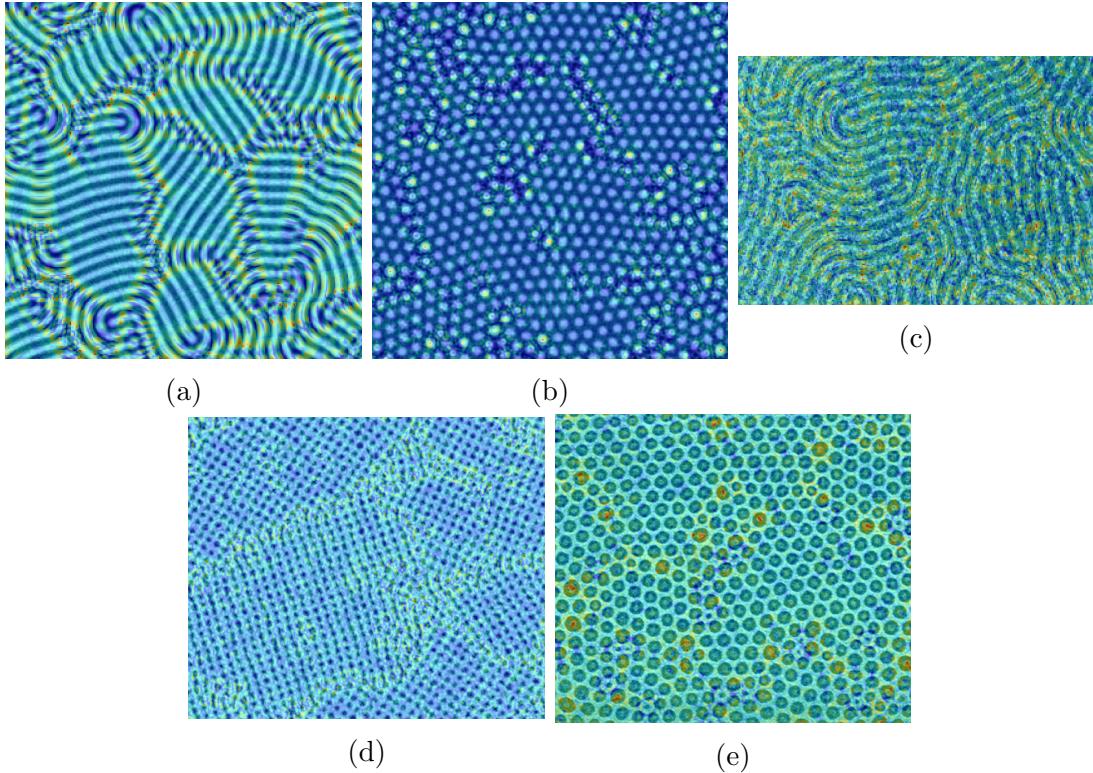


Figure 5: The normalized difference in the response distance using lower-order shapelets ($m \leq 6$) and higher-order shapelets ($m \leq m'$) overlayed onto the original image for the various simulated and experimental self-assembly microscopy images in Figure 3. Brighter regions correspond to larger differences in response distance.

3.2. Orientation

The shapelet response distance method also provides local orientation (at maximum response, see Section 2) for each shapelet function. This local orientation is a result of the use of *steerable* shapelet formulations [13, 14], which involves convolutions with two (basis) components of each shapelet function, instead of multiple rotations at a fixed rotation interval as with other methods [19]. Quantification of local pattern orientation and spatial variation is as important as quantifying the degree of order itself. However, this is not a straightforward task given that each shapelet of order m may have a different orientation at every pixel within an image. Furthermore, the orientation for each shapelet ranges from $[0, \frac{2\pi}{m}]$, further complicating the use of the local shapelet *orientation vector* in order to approximate local pattern orientation.

An approach to determine local pattern orientation was developed, inspired by a similar approach using bond-orientational order theory [12] but retains the inherent benefits of the shapelet-based approach where image segmentation is not required. Additionally, the shapelet method does not require different segmentation and pattern feature identification steps as with bond-orientational order theory. Instead of integrating all local orientational information provided by the shapelet orientation

vector, the dominant rotational symmetry (m'') of the pattern is determined and orientation from only this shapelet mode is used. For stripe, square, and hexagonal patterns, the dominant rotational symmetries are $m'' = 1, 4$, and 6 respectively. This results in a maximum orientation value of 2π , $\frac{\pi}{2}$, and $\frac{2\pi}{3}$ for stripe, square, and hexagonal patterns due to degeneracy in rotational symmetry.

However, unlike local pattern response, local orientation is not a feature-scale attribute which results in highly varying local orientation vectors. The proposed method uses local m'' -fold shapelet orientation only in regions with relatively high m'' -fold shapelet response, masking-out local orientation values where pattern intensity is low. Standard image-processing methods (masking, dilation, and blending) are then applied to interpolate local orientation in those regions with low pattern response. This approach is demonstrated for both simulation-based stripe (Figure 1a) and hexagonal (Figure 1b) self-assembly images and shown in Figure 6:

Masking: Figures 6a and 6b show the results of masking the m'' -fold shapelet orientation based on a threshold of scaled local response. This threshold, unique to each image, is found via an iterative scheme, beginning with the largest (and least strict) threshold and analyzing the resulting orientation plot after dilation and blending operations. If there exist undefined orientation regions after blending (i.e., either masked-out and/or not interpolated from blending), the masking threshold is reduced and the process repeated until the post-blending image contains an insignificant number of undefined orientations (< 1% of all pixels). This results in masking-out of regions of the pattern in which either (i) the pattern is relatively uniform with low image intensity (in between pattern features) or (ii) there are defects present (both orientational and translational).

Dilation: Figures 6c and 6d show the results of dilation of the masked orientation images in order to approximate the local orientation of areas with relatively low pattern response. Morphological greyscale dilation, conveniently available from the Scientific Python package (`scipy`) [24] multi-dimensional image processing module (`ndimage`), is used to “fill in” void space between neighbouring well-defined response regions. The dilation kernel size is chosen to be 2λ , where λ is the characteristic wavelength of the pattern and also the approximate distance between well-defined response regions. This scale was chosen to allow for adequate overlapping of neighbouring well-defined response regions which ultimately defines orientation both in void space and across orientational boundaries.

Blending: Figure 7 shows the result of blending the dilated orientation image for both simulated stripe (Figure 1a) and hexagonal (Figure 1b) self-assembly microscopy images superimposed on the original pattern. The blending operation is performed using a median filter kernel [24] with scale 4λ , so that the effective local orientation at each pixel is the median between two layers of surrounding well-defined features. This allows for effective transitions in orientation between well-defined response regions and void space/orientational boundaries.

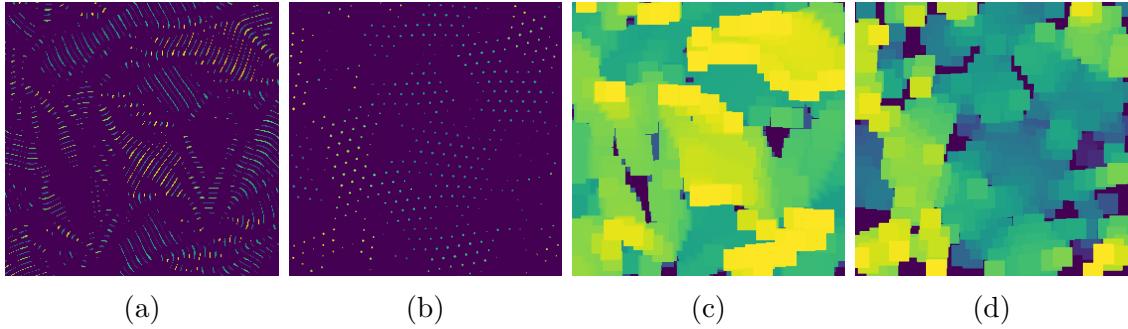


Figure 6: Masking and dilation operations applied to various simulated patterns: (a-b) masking operation applied to simulated stripe (Figure 1a) and hexagonal (Figure 1b) self-assembly images respectively, (c-d) dilation operation applied to simulated stripe (Figure 1a) and hexagonal (Figure 1b) self-assembly images respectively using kernel size 2λ .

For both simulated stripe (Figure 6a) and hexagonal (Figure 6b) masks, the resulting local orientation in uniform areas is qualitatively consistent and correct, with respect to manual (human) analysis. Several uniform regions in both images have gradual (defect-free) changes in local orientation corresponding to elastic-like pattern deformation. This is quantified well, with gradual local computed orientation changes corresponding to those observed manually. For sharp changes in orientation, which correspond to the presence of grain boundaries, there are corresponding sharp changes in computed orientation which validate the specific blending method used in that these orientational interfaces should be sharp and not diffuse. It is important to note that, even though the presented method estimates local orientation values in regions of orientational defects (disclinations), local orientation is degenerate in these areas, given that local m -fold order is present.

The shapelet-based local orientation method was then applied to experimental images of stripe (Figure 3a), square (Figure 3b), and hexagonal (Figure 3c) self-assembly, with results shown in Figures 7c, 7d, and 7e respectively. The presence of noise and non-ideal pattern features is not found to have a significant effect on the accuracy of the computed local orientation. Furthermore, like the shapelet-based response-distance method, the local orientation method is generalized and is able to analyze patterns with non-convex features (stripe).

3.3. Defect identification method

Changes in local orientation in self-assembly are mediated by the presence of orientation defects, disclinations, and defect structures such as arrays of dislocations and grain boundaries. Thus identification of defects is both complimentary to local orientation information and an important property for structure/property relations [5]. While crystalline and liquid crystalline phases have well-defined topological defects [4, 26]

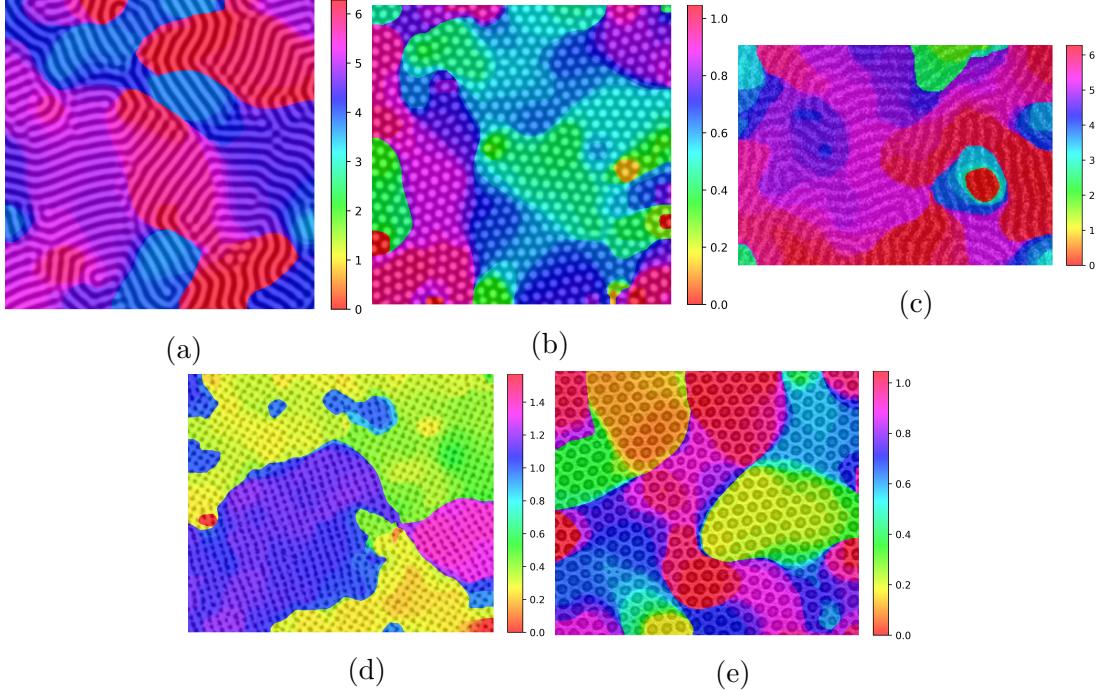


Figure 7: The result of applying the blending operation via median filter kernel [24] and superimposing this onto the original image for: (a) simulated stripe pattern (Figure 1a), (b) simulated hexagonal pattern (Figure 1b), (c) experimental stripe pattern (Figure 3a), (d) experimental square pattern (Figure 3b), and (e) experimental hexagonal pattern (Figure 3c). Cyclic colour map “hsv” from `matplotlib` [25] is used to account for degeneracy in shapelet rotational symmetry.

(orientational versus translational, dimensionality, charge), in self-assembly there can be significant order deformation which is not energetically possible for order on molecular and atomic scales. Thus the use of the shapelet-based response distance method [9, 10] with user supervision is challenging, in that the user would need to identify multiple subdomains with pattern defects of greatly varying character. This is opposite of the original approach where the user selects a subdomain that contains uniform pattern order.

To address this challenge, a complementary method for defect identification, the defect response distance method, is proposed which uses k -means clustering [15] of shapelet response vectors from the *whole* image instead of a pre-defined subdomain. The cluster associated with each pixel in the image is then visualized so that the user can select *multiple* clusters associated with one or more defects and/or defect structures. Figures 8a, 8b, and 8c show examples of this decomposition for simulated stripe (Figure 1a), experimental square (Figure 3b), and simulated hexagonal (Figure 1b) patterns. Stripe patterns required $k \geq 4$, where cluster centroids associated with ordered regions have higher response for specific subsets of m -fold shapelet symmetries and those associated with defects have relatively equal response. For square and hexagonal

patterns, additional clusters were needed ($k \geq 8$ and $k \geq 10$ respectively) with three or four centroids associated with defects, depending on both the amount of noise present in the image (which affects k -means clustering [15]) and user discretion for selecting defect clusters. Figures 8d, 8e, and 8f show visualizations of cluster centroid weights for shapelet order $m \leq 7$. Higher-order shapelets for each image (maximum shapelet order m' found in Figure 4) were included in this analysis, but their corresponding centroid weight values were omitted for brevity.

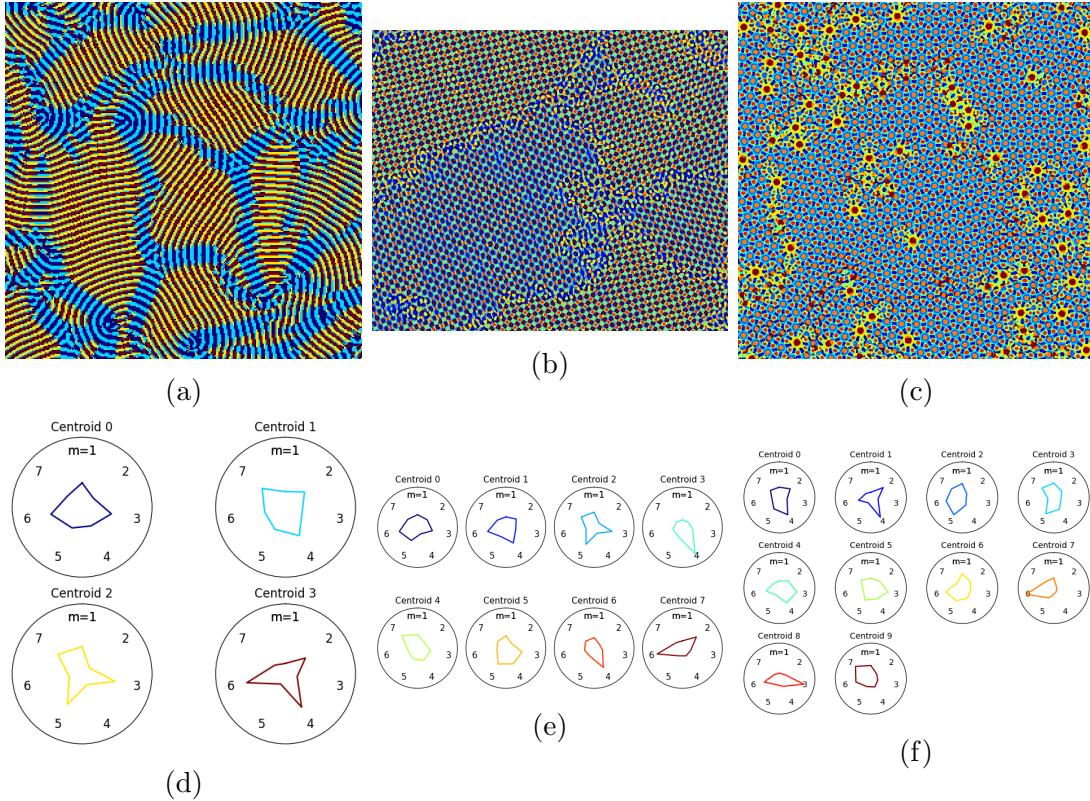


Figure 8: Decomposition of shapelet response into 4 (stripe), 8 (square), or 10 (hexagonal) centroid response vectors. Displaying the cluster associated with each pixel (response vector) (a-c) and the radar chart representation of the centroid vector weights (d-f) for order $m \leq 7$ shapelets for self-assembly images: (a/d) simulated stripe (Figure 1a) image, with centroids 0 and 1 selected as defect centroids, (b/e) experimental square (Figure 3b) image, with centroids 0, 1, and 4 selected as defect centroids, and (c/f) simulated hexagonal (Figure 1b) image, with centroids 2, 5, 6, and 9 selected as defect centroids.

User supervision is significantly more involved than the response distance method, as previously the user was required to choose a single reference subdomain, but now both the number of clusters k and the set of defect-related cluster centroids must be chosen. Given this information, the response distance method can then be applied by using the centroid vector of each defect cluster as the reference set, for each defect cluster, instead of a single reference set of response vectors associated with uniform

order (as in Section 3.1). This defect response distance method was performed for each of the simulated and experimental images from Figure 3 and the results are shown in Figure 9, with intermediate results for experimental images (similar to those in Figure 8) provided in the Supplementary Material. Additionally, a median filter [24] with kernel size $\frac{\lambda}{2}$, where λ is the characteristic length scale (wavelength) of the pattern, was used to smooth the defect response distance result used in the pattern overlays. This kernel size is significantly smaller than that used for smoothing (blending) of dilated orientation results (see Section 3.2) because of the very localized nature of defects and defect structures.

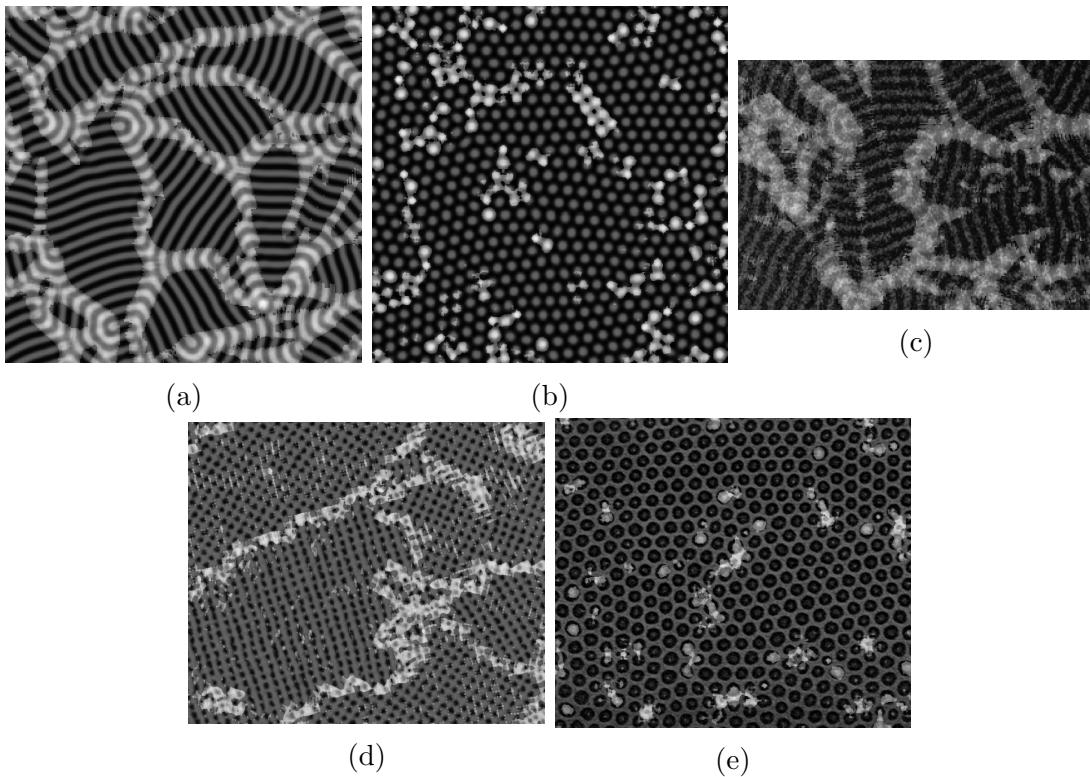


Figure 9: The defect response distance superimposed onto the original pattern image, after using a median filter [24] with kernel size $\frac{\lambda}{2}$ for smoothing which identifies defects for self-assembly images: (a) simulated stripe (Figure 1a) image, (b) simulated hexagonal (Figure 1b) image, (c) experimental stripe (Figure 3a) image, (d) experimental square (Figure 3b) image, and (e) experimental hexagonal (Figure 3c) image.

Referring to defect response distance results for both hexagonal self-assembly images (Figures 9b and 9e), the method identifies both positive and negative disclinations (local 5-fold and 7-fold symmetric regions). In Figure 9b, several grain boundaries and pairs of adjacent oppositely charged disclinations, corresponding to dislocations in a hexagonal pattern, are also correctly identified. In Figure 9e, defect identification results are similarly found to be correct, however, the presence of noise in this image (experimentally determined) is manifested in less strong response. This is

particularly true for positive disclinations, but each one is correctly identified to some extent.

Referring to defect response results for both stripe self-assembly images (Figures 9a and 9c), the method responds much more strongly due to the presence of different grain boundary structures. It is important to note that stripe patterns are not composed of convex pattern structures and, instead, they are non-convex and semi-continuous within the image. Thus instead of localized point-like defect structures as is present in the representative hexagonal patterns, defect structures in stripe patterns are found to be line-like, similar to lower dimensional orientational inversion wall defects as seen in liquid crystalline phases [26]. Both positive and negative disclinations are identified by the method, along with individual and arrays of dislocations. Finally, inversion wall-like defect lines are also correctly identified due to the discontinuity of local orientation that they represent.

Referring to the defect response results for the experimental square self-assembly image (Figure 9d), topological defects (disclinations, dislocations) and grain boundary defect structures are adequately identified.

While the defect response distance method is found to correctly identify *regions* which include defects, it is not currently suitable for defect *classification* or quantification of the number of defects present in a given pattern. However, the ratio of the image area to that in which defects are identified is a quantity that could be used for structure-property relationships.

4. Conclusions

Significant enhancements to the past shapelet-based response distance method [9, 10] were developed which enabled automated quantification of local pattern orientation and identification of topological defects via the defect response distance method within images of nanostructured surfaces. These enhanced methods were applied to a representative set of nanostructured surface images, including those with varying patterns (stripe, square, hexagonal) and using different microscopy techniques (SEM, TEM, and AFM).

The use of higher-order shapelets within the existing response distance method [9, 10] was shown to improve the performance of the method at identifying regions of the pattern with deformation and topological defects. Higher-order shapelet response vectors were especially important for experimental images with significant noise.

A method for determining local pattern orientation was presented and applied which uses information pre-computed by the response distance method, specifically, *steerable* [13] shapelet orientation at maximum response. This shapelet-based method was validated using the set of representative nanostructure images and was able to quantify local orientation for each pattern, including those composed of non-convex features (stripe). This additional data enables researchers to identify local grains and orientation, along with computing estimates of orientational correlation length [12].

Finally, a supervised method for identification of topological defects, the defect response distance method, was developed and validated. The defect response distance method allows the user or researcher to select response distance centroids which are locally present in dislocation and disclination defects, and then automates the identification of all similar defects throughout the image. The method is found to identify defects for both convex (square, hexagonal) and non-convex (stripe) patterns, with measurement noise present.

The shapelet-based methods presented in this work provide researchers in nanoscience and nanotechnology with comprehensive and robust computational techniques for local quantification of nanostructure order/disorder, orientation, and the presence of topological defects. Implementations of these methods are provided for the research community as an open-source software package [20].

Acknowledgments

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Digital Research Alliance of Canada.

References

- [1] Moudikoudis S, Pallares R M and Thanh N T 2018 *Nanoscale* **10** 12871–12934
- [2] Abukhdeir N M 2016 *Materials Research Express* **3** 082001 URL <http://stacks.iop.org/2053-1591/3/i=8/a=082001>
- [3] Gubicza J 2017 *Defect structure and properties of nanomaterials* (Woodhead Publishing)
- [4] Chaikin P M and Lubensky T C 1995 *Principles of condensed matter physics* (Cambridge Univ Press)
- [5] Li W and Müller M **6** 187–216
- [6] Jangizehi A, Schmid F, Besenius P, Kremer K and Seiffert S 2020 *Soft Matter* **16** 10809–10859
- [7] Rhodes D, Chae S H, Ribeiro-Palau R and Hone J 2019 *Nature materials* **18** 541–549
- [8] Refregier A 2003 *Monthly Notices of the Royal Astronomical Society* **338** 35–47 ISSN 1365-2966 URL <http://dx.doi.org/10.1046/j.1365-8711.2003.05901.x>
- [9] Suderman R, Lizotte D and Abukhdeir N 2015 *Phys. Rev. E*
- [10] Akdeniz T J, Lizotte D and Abukhdeir N M 2018 *Nanotechnology*
- [11] Massey R and Refregier A 2005 *Monthly Notices of the Royal Astronomical Society* **363** 197–210 ISSN 1365-2966 URL <http://dx.doi.org/10.1111/j.1365-2966.2005.09453.x>
- [12] Strandburg K J (ed) 1992 *Bond-orientational Order in Condensed Matter Systems* (Springer-Verlag)
- [13] Freeman W T and Adelson E H 1991 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13** 891–906
- [14] Suderman R 2014 *Regular Pattern Detection and Analysis Using Shapelets* Master's thesis University of Waterloo
- [15] Wu J 2012 *Advances in K-means clustering: a data mining thinking* (Springer Science & Business Media)
- [16] Park M, Harrison C, Chaikin P M, Register R A and Adamson D H 1997 *Science* **276**(5317) 1401–1404
- [17] Tang C, Lennon E M, Fredrickson G H, Kramer E J and Hawker C J 2008 *Science* **322**(5900) 429–432

- [18] Park W I, Kim Y, Jeong J W, Kim K, Yoo J, Hur Y H, Kim J M, Thomas E L, Katz A A and Jung Y S 2013 *Scientific Reports* **3**(3190)
- [19] Wang Z, Li J, Guo Y, Tang S and Wang J 2013 *Computer Physics Communications* **184** 2489–2493 ISSN 0010-4655 URL <http://www.sciencedirect.com/science/article/pii/S0010465513002129>
- [20] URL <https://github.com/uw-comphys/shapelets>
- [21] Briggs W L and Henson V E 1995 *The DFT: an owner's manual for the discrete Fourier transform* (SIAM)
- [22] de Gennes P and Prost J 1995 *The Physics of Liquid Crystals* 2nd ed (New York: Oxford University Press)
- [23] Dong J, Goldthorpe I A and Abukhdeir N M 2016 *Nanotechnology* **27** 235701 URL <http://stacks.iop.org/0957-4484/27/i=23/a=235701>
- [24] Virtanen P, Gommers R, Oliphant T E, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J *et al.* 2020 *Nature methods* **17** 261–272
- [25] Hunter J D 2007 *Computing in science & engineering* **9** 90–95
- [26] Kleman M 1989 *Reports on Progress in Physics* **52** 555–654 ISSN 0034-4885 URL <http://iopscience.iop.org/0034-4885/52/5/002>
<http://stacks.iop.org/0034-4885/52/i=5/a=002?key=crossref.0e27843bc55e70a3cab14a3b18413ed5>