Recurrence network modeling and analysis of spatial data
Cheng-Bang Chen, Hui Yang, and Soundar Kumara

Citation: Chaos 28, 085714 (2018); doi: 10.1063/1.5024917
View online: https://doi.org/10.1063/1.5024917
View Table of Contents: http://aip.scitation.org/toc/cha/28/8
Published by the American Institute of Physics

Articles you may be interested in
Multivariate weighted recurrence network analysis of EEG signals from ERP-based smart home system

Is Bach’s brain a Markov chain? Recurrence quantification to assess Markov order for short, symbolic, musical compositions

Phase space reconstruction for non-uniformly sampled noisy time series

Recurrence quantification analysis for the identification of burst phase synchronisation

Recurrence threshold selection for obtaining robust recurrence characteristics in different embedding dimensions

Deriving inferential statistics from recurrence plots: A recurrence-based test of differences between sample distributions and its comparison to the two-sample Kolmogorov-Smirnov test
Recurrence network modeling and analysis of spatial data

Cheng-Bang Chen, Hui Yang, and Soundar Kumara
Harold and Inge Marcus Department of Industrial and Manufacturing Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, USA

(Received 5 February 2018; accepted 22 May 2018; published online 27 August 2018)

Nonlinear dynamical systems exhibit complex recurrence behaviors. Recurrence plot is widely used to graphically represent the patterns of recurrence dynamics and further facilitates the quantification of recurrence patterns, namely, recurrence quantification analysis. However, traditional recurrence methods tend to be limited in their ability to handle spatial data due to high dimensionality and geometric characteristics. Prior efforts have been made to generalize the recurrence plot to a four-dimensional space for spatial data analysis, but this framework can only provide graphical visualization of recurrence patterns in the projected reduced-dimension space (i.e., two- or three-dimensions). In this paper, we propose a new weighted recurrence network approach for spatial data analysis. A weighted network model is introduced to represent the recurrence patterns in spatial data, which account for both pixel intensities and spatial distance simultaneously. Note that each network node represents a location in the high-dimensional spatial data. Network edges and weights preserve complex spatial structures and recurrence patterns. Network representation is shown to be an effective means to provide a complete picture of recurrence patterns in the spatial data. Furthermore, we leverage network statistics to characterize and quantify recurrence properties and features in the spatial data. Experimental results in both simulation and real-world case studies show that the generalized recurrence network approach yields superior performance in the visualization of recurrence patterns in spatial data and in the extraction of salient features to characterize recurrence dynamics in spatial systems. Published by AIP Publishing. https://doi.org/10.1063/1.5024917

This paper is motivated by the need for the representation and visualization of recurrence dynamics in the spatial data. Due to the high dimensionality and geometric characteristics, spatial data pose significant challenges on traditional recurrence methods for representation and characterization. The generalized recurrence plot involves the four-dimensional space for spatial data analysis, thereby providing limited visualization power in the reduced-dimension space. Thus, we propose a new weighted recurrence network approach that accounts for both the similarity level of pixel intensities and spatial closeness for spatial data analysis. Furthermore, we evaluate and validate the proposed methodology with both simulation and real-world case studies for the predictive modeling of quality variables in the manufacturing systems.

I. INTRODUCTION

Recurrence behaviors are common in real-world and natural systems. For instance, an electrocardiogram (ECG) shows near-periodical recurrences of heartbeats. The time series of rainfall and temperature indicate the recurrence cycle of the climate system. The vibration signals reflect the performance of recurrent productions in the manufacturing processes. The Poincaré recurrence theorem shows that the trajectory of a dynamic system reappears at the neighborhood of former states after a sufficiently long but finite time.

Recurrence plot (RP) is widely used to graphically represent the recurrence dynamics in the time series data which characterizes the proximity of two states in the time-delay reconstructed state space of a dynamical system. RP has found general applications in different disciplines. For example, Marwan et al. studied the climate change in the northwestern Argentine Andes and the interrelationship between commodity and stock indices movement. They also studied cardiac arrhythmias through the recurrence patterns in time series of heart rate variability. Yang et al. studied multiscale and heterogeneous recurrences in human heartbeats and manufacturing systems for process monitoring and control. RP provides an effective means to visualize recurrent dynamics in the time series data. Furthermore, recurrence quantification analysis (RQA) provides statistical measures to quantify the patterns and features in RP.

However, traditional RP tends to be limited in its ability to handle spatial data. Note that spatial data structure is high-dimensional with geometric correlation patterns, which is different from time series data. Traditional RP uses the state index or time index, to characterize the proximity of states in the space. Nonetheless, spatial data contain more than one dimension of geometric indices (e.g., two axes for one image) that pose challenges to the traditional RP. The recurrence may occur along each geometric index and/or in any spatial direction in the spatial data. Prior research attempts to resolve these issues by increasing the dimensionality of the RP to represent recurrence patterns of spatial data. For instance, Marwan et al. generalize the RP for recurrence analysis of two-dimensional imaging data. Nonetheless, the
generalized RP is extended to the four-dimensional space for spatial data analysis, but this framework can only provide graphical visualization of recurrence patterns in the projected reduced-dimension space (i.e., two- or three- dimensions). To provide a complete picture of recurrence patterns in the spatial data, new recurrence methods and tools are urgently needed.

Thus, we propose a generalized framework based on complex networks for recurrence analysis of spatial data. The advantageous features of network theory are that each network node can go beyond the geometric index and represents each pixel in the imaging data. Network edges and weights preserve complex spatial structures and recurrence patterns. The proposed approach accounts for geometric features of the spatial data while constructing the recurrence network. Network weights are computed with both pixel similarities and geometric distances to characterize the recurrences in the spatial data. The proposed method circumvents the shortcoming of generalized RP that visualizes the recurrence patterns in the reduced-dimension space and further provides a complete picture of recurrence dynamics using network representation. We also propose network statistics to measure and quantify recurrence properties and features of the spatial data. Experimental results in both simulation and real-world case studies show that the generalized recurrent network approach yields superior performance in the visualization of recurrence dynamics of spatial data, as well as in the prediction of surface roughness of workpieces from ultra-precision machining (UPM).

The rest of this paper is organized as follows: Sec. II presents the review of relevant works in recurrence analysis, Sec. III details the proposed methodology of recurrence network for spatial data analysis, Sec. IV evaluates and validates the proposed methodology in both simulation and real-world examples, and Sec. V includes the conclusions arising from this investigation.

II. RESEARCH BACKGROUND

In the literature, most of the previous efforts focus on recurrence analysis of time series data. As shown in Fig. 1, the RP provides a graphical representation of state proximity in the state space that is reconstructed from the time series. Mutual information is commonly used to select the optimal delay in time. The embedding dimension is determined by the false nearest neighbor method. Mathematically, the RP is represented as a symmetric matrix $R$:

$$R_{i,j} = \Theta(\xi - ||\vec{s}_i - \vec{s}_j||), \quad \vec{s}_i \in \mathbb{R}^m, \quad (1)$$

where $\vec{s}_i$ is the state at time $i$, $\Theta(\cdot)$ is the Heaviside function, and $\xi$ is the threshold. As shown in Fig. 1, if the state $\vec{s}_i$ is close to $\vec{s}_j$ and the distance measure is below a certain threshold $\xi$, then $R_{i,j}$ is equal to 1 and is represented as a black dot in the RP. Otherwise, $R_{i,j}$ is equal to 0 and is represented as a white dot in RP. Figure 1(c) shows the RP of the lag-reconstructed state space of the Lorenz time series. Both axes of RP are the state indices. Black dots in the RP denote when the state $i$ recurs in the neighborhood of state $j$. RP contains intriguing patterns of recurrence dynamics. The discontinuity in the lines parallel to the diagonal shows the level of system nonlinearity. Based on the line structures in RP, Webber and Marwan\textsuperscript{10} proposed the recurrence quantification analysis (RQA) that provides statistical measures of recurrence patterns in the RP. Also, Yang and Chen investigated various types of recurrences in the state space, namely, heterogeneous recurrence analysis.\textsuperscript{11} The new heterogeneous recurrence methods yield superior performance in the identification of dynamic transients in the manufacturing processes\textsuperscript{12,13} as well as the detection of sleep apnea events in healthcare.\textsuperscript{14}

However, traditional RP tends to be limited in the ability to handle spatial data due to the high dimensionality and special geometric data structure. Hence, Marwan et al.\textsuperscript{9,15} propose to generalize the RP framework for two-dimensional spatial data as

$$R_{\vec{p},\vec{q}} = \Theta(\xi - s_{\vec{p}} - s_{\vec{q}}), \quad s \in \mathbb{R} \text{ and } \vec{p}, \vec{q} \in \mathbb{N}^2, \quad (2)$$

where $s_{\vec{p}}$ represents the pixel intensity value of a pixel at the location $\vec{p}$, where $\vec{p}$ is the two-dimensional vector of spatial coordinates. Note that $\Theta(\cdot)$ is still the Heaviside function and $\xi$ is the threshold. In this framework, if the intensity difference between two pixels is below the threshold $\xi$, then there are spatial recurrences. Note that this method extends the two-dimensional $R_{i,j}$ in traditional RP to the four-dimensional $R_{\vec{p},\vec{q}}$ because $\vec{p}$ and $\vec{q}$ are two-dimensional vectors of spatial coordinates, i.e., $R_{\vec{p},\vec{q}} = R_{(x_p, y_p), (x_q, y_q)}$.

Figure 2 illustrates the extended recurrence plot of imaging data. Figure 2(a) shows the two locations of two pixels in the image, i.e., $\vec{p} = (x_p, y_p)$ and $\vec{q} = (x_q, y_q)$. As $R_{(x_p, y_p), (x_q, y_q)}$ is
the four-dimensional data, spatial recurrence patterns can only be visualized in the reduced-dimension space. For example, Fig. 2(b) shows the 3D visualization with three dimensions \( x_p, y_q, z_q \) that are arbitrarily selected from the four-dimensional space. It is difficult to get a complete picture of the extended recurrence plot \( R_{\tilde{p}, \tilde{q}} \). Only partial recurrence information can be visualized due to arbitrary selection of dimensions and conditional restriction. Also, it will be even more difficult to create the extended RP for three-dimensional imaging data because \( \tilde{p} \) and \( \tilde{q} \) will be three-dimensional vectors of spatial coordinates, i.e., \((x_p, y_p, z_p)\) and \((x_q, y_q, z_q)\). As a result, \( R_{\tilde{p}, \tilde{q}} \) will involve a total of six dimensions. The extended RP method in Eq. (2) is influenced by the curse of dimensionality and cannot adequately provide effective visualization of spatial data. There is a definite need to develop new recurrence methods and tools for recurrence analysis of spatial data that can provide effective characterization and quantification of recurrence patterns, as well as provide better visualization capability.

Network theory opens a new area to investigate the recurrence patterns of spatial data. Note that social networks, computer networks, and neural networks have received increasing interests in the past decade.\(^\text{16-18}\) The complex interactions between components within a system can be readily represented as the node-to-node interactions in the network. Although the network theory has been widely used for the analysis of nonlinear and nonstationary systems, it is difficult to get a complete picture of the extended recurrence plot \( R_{\tilde{p}, \tilde{q}} \). Only partial recurrence information can be visualized due to arbitrary selection of dimensions and conditional restriction. Also, it will be even more difficult to create the extended RP for three-dimensional imaging data because \( \tilde{p} \) and \( \tilde{q} \) will be three-dimensional vectors of spatial coordinates, i.e., \((x_p, y_p, z_p)\) and \((x_q, y_q, z_q)\). As a result, \( R_{\tilde{p}, \tilde{q}} \) will involve a total of six dimensions. The extended RP method in Eq. (2) is influenced by the curse of dimensionality and cannot adequately provide effective visualization of spatial data. There is a definite need to develop new recurrence methods and tools for recurrence analysis of spatial data that can provide effective characterization and quantification of recurrence patterns, as well as provide better visualization capability.

Network theory opens a new area to investigate the recurrence patterns of spatial data. Note that social networks, computer networks, and neural networks have received increasing interests in the past decade.\(^\text{16-18}\) The complex interactions between components within a system can be readily represented as the node-to-node interactions in the network. Although the network theory has been widely used for the analysis of nonlinear and nonstationary systems,\(^\text{11,19,20}\) little has been done for recurrence analysis of spatial data through network theory. A network \( G = \{V, E\} \) is constructed by a set of nodes, \( V \), and a set of edges, \( E \). Donner \textit{et al.}\(^\text{20}\) introduced the recurrence network for nonlinear time series analysis. The recurrence matrix, \( R_{ij} \), is used to derive the adjacent matrix, \( A_{ij} \), of the recurrence network as

\[
A_{ij}(\xi) = R_{ij}(\xi) - \Delta_{ij} = \Theta(\xi - x_i - x_j) - \Delta_{ij},
\]

where \( \Delta_{ij} \) is the Kronecker delta used to avoid artificial self-loop in the network. The recurrence network is an undirected and unweighted network for nonlinear time series analysis. Yang \textit{et al.}\(^\text{21}\) derived the self-organizing network structures of networks from the recurrence adjacency matrix \( A_{ij} \), and further leveraged self-organizing network for variable clustering and predictive modeling.\(^\text{22-24}\)

An attractive feature of network representation is that each network node is flexible to represent the \( x_i \) in the time series, \( \tilde{s}_i \) in the state space, or each pixel in the imaging data. Network edges and weights preserve complex spatial structures and recurrence patterns. The other attractive feature is that network theory provides a rich set of statistical measures of network characteristics, which are also conducive to the characterization and quantification of system dynamics.\(^\text{25}\) Prior research efforts have also shown the relationship between these network statistics and traditional \( \text{RQA} \) measures and how they complement each other to characterize the nonlinear and nonstationary systems.\(^\text{11,19,20}\) However, very little has been done to develop the recurrence network approach for spatial data analysis. Therefore, we propose to investigate the application of weighted network representation of spatial data, and further leverage network statistics to build a predictive model that is aimed at predicting the surface finish from the UPM images.

### III. RESEARCH METHODOLOGY

Recurrence analysis focuses on the proximity of states in the state-space trajectory. The “position” and “values” of the states are essential to recurrence analysis although system dynamics evolves over time. For the nonlinear time series, the timestamp gives the “position” of the states in the lag-reconstructed state space. The state vector provides the “status” of system dynamics at a specific position. However, both “positions” and “values” of spatial data possess special structures that are different from time series. In other words, both spatial coordinates and pixel values should be considered in the recurrence analysis of spatial data. As shown in Fig. 3, if we only take the pixel values into account and overlook the spatial correlation as in traditional recurrence network methods [Eq. (3)], then two different images will generate isomorphic networks if these two images follow the same intensity distribution.

Spatial data are different from time series because of the involvement of both pixel intensity and spatial location. However, traditional recurrence network theory is defined based on the adjacency matrix of data with a single index (e.g., time series with the time index). In the case of spatial data,
If we only consider the pixel intensity, then pertinent information about spatial recurrences will be missing. Therefore, traditional methods of recurrence network will generate two isomorphic networks for two images with different graphic complexity but the same distribution of intensities. In summary, it is critical to consider both pixel intensity and spatial location so as to capture the recurrence behaviors in the spatial data.

Therefore, we propose a weighted network method for recurrence analysis of spatial data that will account for both the pixel similarity and the spatial closeness simultaneously. In this section, we first introduce the proposed weighted network representation of spatial data, and then we discuss the use of network statistics to characterize recurrence patterns for predicting the surface finish from the UPM images.

A. Weighted network representation of spatial data

Spatial data contain important contextual information around the spatial position or location of an object. As opposed to the one-dimensional index or timestamp in the time series, the dimensionality of spatial coordinates can be more than one. For instance, the dimensionality of spatial coordinates in a line, a 2D image, and a 3D object is one, two, and three, respectively. Also, the pixel values provide a great deal of information on the spatial patterns. In the grayscale, the pixel value is scalar in the range of [0, 255]. In the RGB scale, the pixel value is the combination of different levels of red, blue, and green. Thus, we propose to include both pieces of essential information, spatial location and pixel value, to build the weighted network representation of spatial data.

\[ D_{p,r} = \frac{\phi(||\vec{p} - \vec{r}||)}{\phi(0)} \]

\[ D_{p,q} = \frac{\phi(||\vec{p} - \vec{q}||)}{\phi(0)} \]

FIG. 3. Two images with different spatial complexity but the same intensity distribution lead to the isomorphic network via the traditional framework of recurrence network.

FIG. 4. The spatial correlation index, $D_{p,r}$, using Gaussian function $\phi(\cdot)$ in 2D space, where $\phi(||\vec{p} - \vec{q}||)$ represents the geometric utility from $\vec{p}$ to arbitrary location $\vec{q}$, and $\phi(0)$ is the normalization constant. If $\vec{p} - \vec{q} < \vec{p} - \vec{r}$, then $\phi(||\vec{p} - \vec{q}||) > \phi(||\vec{p} - \vec{r}||)$, which leads to $D_{p,q} > D_{p,r}$. 

A. Weighted network representation of spatial data

Spatial data contain important contextual information around the spatial position or location of an object. As opposed to the one-dimensional index or timestamp in the time series, the dimensionality of spatial coordinates can be more than one. For instance, the dimensionality of spatial coordinates in a line, a 2D image, and a 3D object is one, two, and three, respectively. Also, the pixel values provide a great deal of information on the spatial patterns. In the grayscale, the pixel value is scalar in the range of [0, 255]. In the RGB scale, the pixel value is the combination of different levels of red, blue, and green. Thus, we propose to include both pieces of essential information, spatial location and pixel value, to build the weighted network representation of spatial data.
describes a relative intensity similarity index of two pixels, posed weighted network framework, if any two data points (i.e., the spatial correlation index) as in Eq. (4). The function of intensity similarity is constructed for the spatial data as shown in the equations below:

\[ w_{p,q} = I_{p,q} \cdot D_{p,q}, \]  
\[ I_{p,q} = 1 - \frac{\tilde{S}_p - \tilde{S}_q}{\max \|\tilde{S}\| - \min \|\tilde{S}\|}. \]  
\[ D_{p,q} = \frac{\phi(\tilde{p} - \tilde{q})}{\phi(|0|)}. \]  

The edge weight \( w_{p,q} \) is the product of \( I_{p,q} \) (i.e., the similarity level of pixel intensities) and \( D_{p,q} \) (i.e., the spatial correlation index) as in Eq. (4). The function of intensity similarity \( I_{p,q} \) describes a relative intensity similarity index of two pixels, as in Eq. (5), which gives a normalized similarity level of two observed pixels. If \( I_{p,q} = 1 \), two pixels have the same intensity level. If \( I_{p,q} = 0 \), then two pixels have opposite intensities. The index \( D_{p,q} \) in Eq. (6) describes the spatial correlation levels.

As shown in Fig. 4, we illustrate the use of utility function \( \phi(\cdot) \) to describe the correlation levels of spatial closeness. In general, the utility function \( \phi(\cdot) \) should have the following properties: (1) locally supported, (2) nonnegative, and (3) monotonically decreasing function with respect to the spatial distance \( \delta = \tilde{p} - \tilde{q} \). Without loss of generality, we choose the Gaussian utility function, \( \phi(x|\Sigma) = (2\pi|\Sigma|)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(x^T \Sigma x)\right] \), because it is compact and locally supported. Therefore, it is conducive to capture the geometric correlation patterns. In other words, if two pixels are located closer to each other, then they tend to have a higher correlation. The spatial correlation index \( D_{p,q} \) is in the range of \([0,1]\) because of the use of Gaussian function. This makes two closer pixels tend to have a higher correlation. Figure 4 illustrates that if \( \tilde{p} - \tilde{q} < \tilde{p} - \tilde{r} \), then \( D_{p,q} > D_{p,r} \).

The edge weight \( w_{p,q} \) takes into account both the similarity level of pixel intensities and spatial distances diminishing the marginal effects.

With a given recurrence threshold \( \xi \), we can build an adjacency matrix \( A_{p,q} \) of the recurrence network for spatial

FIG. 5. The variation of the edge weight with respect to the \( I_{p,q} \) and \( \tilde{p} - \tilde{q} \).

FIG. 6. The structural visualization results of three weighted recurrence networks for three different images: (a) uniformly distributed white noise, (b) auto-correlated process, and (c) periodical images.
data as

\[ A_{pq} = \Theta(w_{pq} \geq \xi) - \Delta_{pq}, \]  

where \( \Theta \) is the Heaviside function, and \( \Delta_{pq} \) is the Kronecker delta used to avoid artificial self-loop in the network.

### B. Recurrence measures from the complex network theory

Note that the edge weight, \( w_{pq} \), is the product of the similarity level of pixel intensities, \( I_{pq} \), and the spatial correlation index, \( D_{pq} \). As shown in Fig. 5, for two pixels with the same intensity similarity, the edge weight is bigger if their spatial distance \( \vec{p} - \vec{q} \) is smaller (i.e., a bigger \( D_{pq} \)). Conversely, the edge weight is smaller if two pixels share the same intensity value but they are far from each other. As a result, the proposed network method uses the measures of both intensity similarity and spatial closeness level to characterize recurrence patterns in the spatial data. For different recurrence patterns, their corresponding network structures will also be different.

Note that the network statistics are established measures of the characteristics of network topology. Network structures provide rich information that is invaluable to predictive modeling and statistical inference.\(^ {26,27} \) Hence, we further leverage network statistics to quantify recurrence patterns of the spatial data. Table I summarizes the network statistics and mathematical equations.

#### 1. Degree

The degree of a network node refers to the number of edges connected to that node. For any node \( i \) of an undirected network, the degree \( k_i = \sum_{j=1}^{N} A_{ij} \). In the proposed network, the degree \( k_p \) refers the frequency of the recurrence relative to the pixel \( p \). The nodes connected to a node correspond to the nodes which have both the image and spatial similarity. Moreover, the degree distribution \( \sim k_p \) provides the information of recurrence distribution of an image.

#### 2. Path

A path in a network is a route across the network that a node takes to reach another node along the edges. The path length is the number of edges within this path. The path characterizes how the recurrence connects pixels in the spatial data. The longer a path is, the more frequent the corresponding recurrence pattern is.

#### 3. Distance

The distance between two nodes is the shortest path between them. In the proposed network, the distance between two nodes implies the geometric distance between these two nodes. Two nodes linked together means the corresponding pixels have a similar intensity and are close in the distance. Therefore, if the node distance is larger than one, then two pixels are not close to each other. A larger distance implies a pattern recurs in the longer distance.

#### 4. Average path length

The average path length of a network is the mean shortest path length of all pairs of nodes. This statistic provides the information that the average distance the spatial patterns recurred.

#### 5. Betweenness centrality

The betweenness centrality is a measure of centrality from the shortest paths and quantifies the number of shortest paths that pass through one node.

#### 6. Transitivity

The transitivity, also known as clustering coefficient, measures the probability that the adjacent nodes of a node are connected, which provides information on the number of triples centered at the node. The transitivity measures the extent recurrent nodes tend to cluster together.
IV. EXPERIMENTAL DESIGN AND RESULTS

To evaluate and validate the proposed recurrence network, we performed computer experiments with both simulated imaging datasets and microscopic images from ultra-precision machining processes. First, we derive the visualization results of recurrence networks for three images, namely, random noises, two-dimensional auto-correlated process of 2nd order (2D-AR2), and periodic recurrences. Second, we illustrate the recurrence network structures for three simulated images that share the same distribution of pixel intensities but different spatial patterns. Note that traditional recurrence network methods tend to be limited in their ability to capture the spatial closeness level (also see Fig. 3). However, the proposed weighted network method accounts for both intensity similarity and spatial closeness in the spatial data, thereby providing a better picture of recurrence dynamics. Third, the proposed weighted network method is evaluated with microscopic images from ultra-precision-machining processes. Experimental results demonstrate that the proposed network statistics yield superior models for predicting surface roughness of UPM workpieces.

A. Simulation study

As shown in Fig. 6, we generated three simulated images with different recurrence levels, namely, uniformly distributed random noises, second-order autoregressive process, and 2D periodic data. All three images are of resolution 200 × 200 pixels in size and 256 gray scales in pixel intensity. The visualization results of recurrence networks are shown in the bottom of Fig. 6, which reveals the nonlinear and nonstationary recurrences in the corresponding images. The stronger the recurrence patterns in the images, the more organized the structural patterns in the corresponding networks. It is worth mentioning that the proposed network approach provides a complete picture for recurrence visualization through the network structure, while prior method has to project to the reduced-dimension space to visualize the recurrence dynamics of spatial data.

Furthermore, Fig. 7 shows three simulated images (i.e., randomly distributed pattern, stripe pattern, and the checkerboard pattern) with the same distribution of pixel intensities but different spatial patterns. Each image is of 80 × 80 pixels in resolution, the pixel intensity is in the eight levels from black to white, and the pixel intensity follows a uniform distribution, which means that each intensity level has 800 pixels. Note that traditional recurrence networks will lead to isomorphic network structures. Figure 8 shows the structures of proposed recurrence networks corresponding to each of the images at Fig. 7. It may be noted that the network for the random noise image is shown to yield a random structural pattern. However, strip and checkboard images result in the distinct patterns in the corresponding network structures. The network structure for the strip image shows that the nodes corresponding to the pixels having the same intensity (in the same strip) are all connected to form clusters, and each cluster has a strong connection to its neighbor clusters (the strips aside of the original strip) but has weak or no connection to
other clusters. Similarly, for the checkboard image, the nodes representing the pixels in each block are connected as a cluster, but only a few blocks having a similar intensity level and close to each other make the corresponding clusters linked to each other. In this example, network structures have distinct patterns to different image patterns, which illustrate that the proposed method is effective to capture nonlinear and nonstationary recurrence patterns in the images.

Moreover, we extract network statistics for the three recurrence networks. Figure 9 shows the differences in the distribution of node degrees for the three simulated images (i.e., randomly distributed pattern, stripe pattern, and the checkerboard pattern) with the same distribution of pixel intensities but different spatial patterns. The recurrence network for the randomly distributed image has bi-mode bell-shaped distribution. The network for the strip pattern image has two significant patterns on node degree: the distribution before degree 450 is a low-frequency pattern that is close to uniform distribution and the distribution after degree 450 is a high-frequency pattern. The network for the checkboard image has an extremely high peak at degree 400.

We also extract the betweenness centrality of the recurrence networks. The distributions for the corresponding networks are shown in Fig. 10. Similar to the node degree distribution, the distributions of betweenness centrality for these three networks exhibit different patterns. The distribution for the randomly distributed image forms an approximately exponential distribution; the distribution for the strip image still looks like an exponential distribution but has several peaks and the distribution for the checkboard image has an extremely high peak around 1. Figures 9 and 10 show that the proposed weighted recurrence networks effectively capture the critical information pertinent to recurrence patterns in the spatial data, and network statistics can effectively characterize image patterns for the predictive modeling and statistical inference.

**B. Real-world case study**

Furthermore, we evaluate and validate the proposed generalized recurrence network with the real-world images of surface finish from the ultra-precision machining (UPM) process. The UPM is equipped with air-bearing spindles and diamond tools to generate surface finish. The commonly used measure for surface roughness in UPM is arithmetical mean deviation (Ra). For a higher Ra, the quality of UPM finishes is not satisfactory. It is required that UPM surface finish should be within the specification <30 nm to achieve the satisfactory quality of production. Figure 11 illustrates two different levels of surface roughness (Ra = 43.95 nm vs. Ra = 585.93 nm). There are significantly different patterns in the UPM microscopic images. The surface characteristics of UPM images are relevant to the setup and conditions of cutting tools. In this experiment, we collected a total of 73 images with a variety of surface roughness.

Without loss of generality, we set the images 256 gray scales in color and 100 x 100 pixels in size. We first normalize the geometric distance with 100 pixels and set the \( \phi \sim N\left(0, \left(\frac{\xi}{4} \right)^2\right) \), with \( \xi = 0.95 \), we then construct complex recurrence networks and calculate the network statistics for each image. Results reported in Table II show that network statistics have strong correlations with the surface roughness Ra. The results of univariate feature analysis show that the average degree, average path length, and network density have a strong correlation with the surface roughness (p-value < 0.05).

Furthermore, we perform multivariate regression analysis of all the network statistics extracted from 73 UPM images. Table III shows the estimation of regression model parameters with standard errors and p-values. It may be noted that network statistics that are sensitive (i.e., p-value < 0.05) to the variations of surface roughness Ra include min(degree), mean(degree), average path length, Q3(local clustering coefficient), 1/degree^2, and 1/mean(degree). Fig. 12 shows the correlation between actual Ra values and model predictions, which yields the adj-R^2 value of 80.19%. Experimental results show that the proposed weighted network method effectively

<table>
<thead>
<tr>
<th>Network Statistics</th>
<th>( b_1 )</th>
<th>p-value for ( b_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean(degree)</td>
<td>-0.14079</td>
<td>3.97 x 10^{-7}</td>
</tr>
<tr>
<td>SD(degree)</td>
<td>-0.31864</td>
<td>1.78 x 10^{-7}</td>
</tr>
<tr>
<td>Average path length</td>
<td>145.46</td>
<td>1.25 x 10^{-8}</td>
</tr>
<tr>
<td>Density</td>
<td>-1407.71</td>
<td>3.97 x 10^{-7}</td>
</tr>
</tbody>
</table>

FIG. 11. Two UPM images with different surface roughness.

FIG. 12. The predictive performance of the regression model with network statistics (the adj-R^2: 80.19%).
extracts the recurrence patterns from UPM images and further provides an effective prediction model of the surface roughness of UPM workpieces.

V. CONCLUSIONS

This paper presents a generalized network approach for recurrence analysis of spatial data. A key idea is to study the recurrence patterns with the diminishing effects of spatial correlations. As such, we develop an undirected weighted network that takes into account both intensity similarities and the spatial distance to characterize recurrence dynamics of the spatial data. The computational speed is more dependent on the image size than the graphic complexity. For example, Fig. 7 shows three images with different complexity but the same size. The computational speed is approximately the same (i.e., 1.627 s) for these three images. However, if we double the image size, then the computational time will increase from 1.627 s to 3.408 s. Furthermore, we leverage the network metrics to statistically quantify the spatial recurrences. We evaluate the proposed network method with simulated images. Experimental results show that network representation provides an effective means to visualize recurrence dynamics hidden in image data. Furthermore, we evaluate and validate the proposed method with real-world imaging data of workpieces from the ultra-precision-machining processes. Experimental results show that network statistics capture salient features of the surface quality of workpieces, thereby providing a robust model for the prediction of surface roughness.

An attractive feature of the proposed network approach is that network node can go beyond the geometric indices and represents each pixel in the high-dimensional imaging data (e.g., three-dimensional or higher dimensional image). Network edges and weights preserve complex spatial structures and recurrence patterns. Network representation provides an effective means to visualize recurrence characteristics in high-dimensional spatial data. In future work, more research can be done to go beyond Gaussian diminishing effects of spatial correlation and investigate the various types of spatial effects on the spatial recurrence analysis.

ACKNOWLEDGMENTS

The authors would like to thank the National Science Foundation (No. CMMI-1617148) for the support of this research. The authors also thank Harold and Inge Marcus Career Professorship (H.Y.) and Allen E. Pearce and Allen M. Pearce Professorship (S.K.) for additional financial support.


| Coefficients | Estimate | Std. error | t value | Pr(>|t|) |
|--------------|----------|------------|---------|---------|
| (Intercept)  | 4.09 × 10^3 | 1.11 × 10^3 | 3.694 | 0.000455 |
| Min(degree)  | 2.49 | 4.01 × 10^-1 | 6.206 | 4.27 × 10^-8 |
| Mean(degree) | −1.39 | 2.38 × 10^-1 | −5.82 | 1.97 × 10^-7 |
| Max(degree)  | 6.51 × 10^-2 | 4.47 × 10^-2 | 1.458 | 0.149627 |
| Average path length | −3.00 × 10^2 | 5.69 × 10^-1 | −5.278 | 1.61 × 10^-6 |
| Q3(local clustering coefficient) | 3.64 × 10^3 | 4.65 × 10^-2 | 7.833 | 5.78 × 10^-11 |
| 1/degree^2 | 5.64 × 10^1 | 9.01 | 6.259 | 3.46 × 10^-3 |
| 1/mean(degree) | −9.62 × 10^6 | 1.76 × 10^-6 | −5.463 | 7.91 × 10^-7 |

TABLE III. Regression parameters for Ra prediction using network statistics. The adj-R^2= 0.8019.


