IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 September 2022, accepted 6 October 2022, date of publication 10 October 2022, date of current version 13 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3213280

RESEARCH ARTICLE

Characteristics of Peak and Cliff in Branch Length Similarity Entropy Profiles for Binary Time-Series and Their Application

SANG-HEE LEE AND CHEOL-MIN PARK
Division of Industrial Mathematics, National Institute for Mathematical Sciences, Daejeon 34047, South Korea

Corresponding author: Sang-Hee Lee (sunchaos.sanghee@gmail.com)

This work was supported by the National Institute for Mathematical Sciences.

ABSTRACT A binary time series can be transformed into a Branch Length Similarity (BLS) entropy profile by being mapped to a circumference called a *time-circle*. In this study, we explored how peaks and cliffs are formed and how they relate to time series. Peaks and cliffs are defined as spike shapes in their entropy profile and are called peaks (or cliffs) when their shape is symmetric (or asymmetric). We found that when signal bands with different signal densities are in the same time series, peaks or cliffs are formed on the side of the band with lower signal density. In addition, we found that when the signal density is moderately high, the distribution of peaks and cliffs appears as a global increase-decrease tendency of the entropy profile. The tendency appeared as a barrier in the entropy profile of the image. As an application of our findings, we successfully detected specific patterns in binary images using peaks, cliffs and the barriers.

INDEX TERMS Classification, data structure, discrete transforms, entropy, shape detection, signal analysis, signal processing algorithms, time-series analysis.

I. INTRODUCTION

In modern times, various digital devices connected to the internet have a great influence on social system stability. These devices produce a lot of information. In particular, a large portion of this information is occupied by the time series data. Therefore, collection and analysis of the time series data is essential in modern society [1], [2], [3]. A time series data is the information obtained in a chronological order. It has applications in various fields such as medical care, medical science, finance, economics, government, industry, environment, and socioeconomic [4], [5]. Several techniques for time series analysis have been developed to understand the systems in the field. The techniques include (1) finding similarities between time series [6], [7], (2) searching for subsequences in time series [8], (3) reducing dimensionality [9], [10] and subdivision [11]. In these techniques, the representation of the time series data is a fundamental problem that remains to be solved. Many literatures deal with various types of representation, and the purpose of the representation is to analyze the similarity between time series in various fields such as earthquake prediction [12], terrestrial ecosystem dynamics [13], stock price data, exchange rate analysis [14], medical pattern analysis [15], etc.

The similarity between time series data obtained in most fields is often measured using Euclidean Distance (ED) [16] or Dynamic Time Warping (DTW) algorithm [17]. ED is often used when comparing two time series of the same length. This measure computes the square root of the sum of squared differences between elements in the same time period. Although ED is intuitively clear, it has the disadvantage of being overly sensitive to outliers. In particular, this algorithm cannot compare time series of different lengths. To overcome this problem, the DTW algorithm was developed [18]. DTW performs nonlinear mapping by minimizing the total distance between the time series, followed by similarity search and detection [19]. DTW tends to cause singularity problems, which has the disadvantage of poor accuracy between time series and misalignment. For example, a single

The associate editor coordinating the review of this manuscript and approving it for publication was Hamed Azami.

point in one time series can be aligned with a large division in another. To overcome this problem, Keogh and Pazzani [20] estimated the derivative from the values of three temporally neighboring points in the time series. After obtaining the trend information on the original data based on derivatives, they proposed a differential DTW method to find new warping paths in outliers. Discrete Fourier Transform (DFT) and Discrete Wavelet Transform (DWT) algorithms are mainly applied to the time series of highly nonlinear patterns that do not have periodicity and simple trends. The DFT algorithm is a way to represent a time series as a set of sine and cosine functions in the frequency domain. The degree of similarity is measured by comparing the first few coefficient values of the function sets for the two time series. DWT provides multi-resolution signal decomposition capabilities. For this reason, DWT is used to accurately cluster homogeneous groups in time series with high similarity [21], [22], [23].

The abovementioned methods calculate the relative similarities of the two time series through comparison. Additionally, self-similarity, calculated from a time series itself, was newly introduced. Lee and Park [24] proposed a method to compute the branch length similarity (BLS) entropy profile by mapping the signal of a binary time series to a circumference called as time source. The authors defined the self-similarity of the time series from the entropy profile. The concept of self-similarity is completely different from the existing similarity in terms of information regarding the signal distribution structure of the time series itself. The self-similarity based on the BLS entropy profile has the strength to detect small structural changes in time series [25], [26]. To concretely demonstrate the practicality of the strength, they showed that there was a difference in self-similarity between the behavioral trajectories of *Caenorhabditis elegans* exposed to very low concentrations of toxic substances and the normal behavioral trajectories. In addition, the authors argued that there is a need to develop new indicators other than self-similarity for the characterization of entropy profiles. As an example, the authors mentioned a peak on the entropy profile. As a follow-up to the study, we explored peaks and cliffs formed on entropy profiles that reflect changes in signal distribution over time. Peaks and cliffs refer to the spike shape on the entropy profile. We classified the peaks and cliffs based on whether the shape is symmetrical or asymmetrical, respectively. We dealt with the problem of detecting anomalies in spatial distribution maps based on our findings on peak and cliff properties. This method is not only an algorithm for detecting a specific area of an image, but it can also be a method for detecting anomalies in a time series. In the discussion section, we briefly mentioned the idea of sophisticated detection of shapes consisting of closed curves.

II. PRELIMINARIES

A. BRANCH-LENGTH SIMILARITY ENTROPY

Branch length similarity (BLS) entropy and derivation statistics were proposed to provide a universally applicable means to characterize spatial network patterns [27]. The entropy was defined on a simple network consisting of one node and multiple branches (edges). The "length" of a branch can be an actual distance measure, but it can represent a unit measure such as time, area, probability, etc., or it can be a comparable character state such as a binary code, DNA sequence, character or behavioral state, etc. This indicates that the BLS entropy can be applied to a wide variety of problems. The ratio of the length of each branch to the sum of the lengths of all branches is defined as the probability of each branch, as follows:

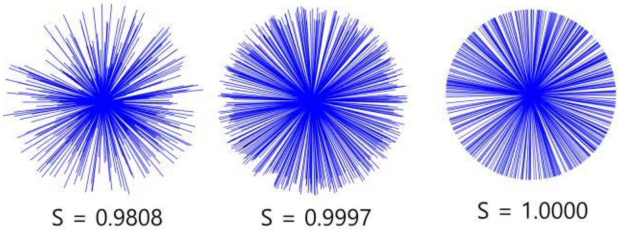


FIGURE 1. Examples of networks with different branch length similarity (BLS) entropy values.

$$p_j = L_j / \sum\nolimits_{k=1}^{n} L_k, \tag{1}$$

where n is the number of branches in the network, and L_k represents the length of the k^{th} branch ($k = 1, 2, 3, \ldots, n$), the BLS entropy can be mathematically represented as

$$S = -\sum\nolimits_{j=1}^{n} p_j \log\left(p_j\right) / \log\left(n\right). \tag{2}$$

In the BLS entropy definition, the more similar the lengths of all branches, the closer the entropy value is to 1.0, and the larger the length deviation, the closer the entropy value is to 0.0 [28]. To better visually understand the concept of BLS entropy, we compared three networks with one node with different branch lengths (Fig. 1). One of these networks had 300 branches with a length between 0.1 and 1.0. In this case, the S value was 0.9008. Another one had branch lengths between 0.9 and 1.0 and had an S value of 0.9997. In the other, the length of all branches was 1.0 and the value of S was 1.0. The higher the branch length similarity, the closer the entropy value was to 1.0.

B. STUDIES ON BRANCH LENGTH SIMILARITY ENTROPY

We encounter multi-node networks more frequently than single-node networks. For this reason, studies using entropy profiles, which are a set of entropy values, have been mainly conducted rather than studies using only a single BLS entropy value. Lee et al [27] obtained BLS entropy profiles from networks created by connecting each pixel on the battle tank shape outline with all other pixels on the edge of the shape. The network formation order was counterclockwise. The authors compared entropy profiles for different tank shapes using correlation coefficient values. They showed that the coefficient values successfully discriminate the shapes from each other, which indicates that the entropy profile effectively characterizes the shape of an object. Lee [28] showed

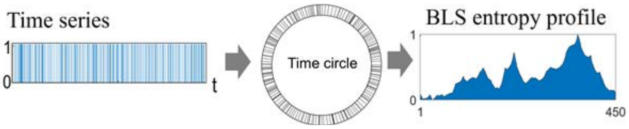


FIGURE 2. Process of constructing a time-circle from a binary time series and obtaining a BLS entropy profile.

that the entropy profile is also robust against noise. Kwon and Lee [29] used the BLS entropy profile to classify leaf shapes that have more complex shapes than battle tanks. The authors defined new descriptors based on entropy profiles: roundness, symmetry, and shape boundary roughness. The descriptors allowed us to successfully classify leaves of various species. Kang et al. [30] successfully classified different butterfly species using BLS entropy profiles. They obtained the entropy profile along the butterfly wing-shaped boundary line. The authors compared the results of applying the entropy profile to three machine learning techniques: Bayesian classifiers, multilayer perceptrons, and support vector machines with other well-known descriptors such as Fourier descriptors. The authors showed that the BLS entropy descriptor outperforms other well-known descriptors. Choi et al. [31] characterized the crawling behavior of *Caenorhabditis elegans* under (1) controlled conditions and (2) conditions treated with toxic substances (formaldehyde, toluene and benzene) using BLS entropy profiles. They showed that the entropy profile can be effectively used for state definition when constructing a hidden Markov model. In the model, the authors quantified the crawling behavior patterns as BLS entropy values, and showed that the patterns were classified into 5 groups using a self-organizing map.

Similarity is one of the main analysis methods in the field of shape classification. However, similarity is a relative quantity that indicates how similar two shapes are. If similarity could be defined based on the information of the shape itself rather than comparison, the existing method could be improved. Lee et al. [25] proposed a new measure (Γ) to quantify the degree of shape self-similarity using BLS entropy. The difference between Γ and other entropy application is that the Γ reflects the information of the entire entropy profile, whereas other entropy application studies previously used a specific value or a simple statistical value in the entropy profile. The authors calculated Γ values for groups of 70 individuals (20 shapes in each group) from the MPEG-7 shape database. Shapes belonging to each group had similar Γ values, and shapes belonging to groups that were geometrically (or topologically) similar also showed relatively similar Γ values.

C. TIME-CIRCLE FOR A BINARY TIME SERIES

We introduced the term, *time-circle*, to extend the BLS entropy concept defined for spatial networks to temporal data (time series). A binary time series is a sequence of signals indicated by signal "1" or "0" for the case in which an event occurs or not according to a discrete time flow. Figure 2 shows a binary time series consisting of 400 randomly distributed "1" signals and the normalized entropy profile in the [0, 1] range for the time series. To obtain the entropy profile, we mapped the binary time series to the circumference of a circle called the *time-circle* [24]. Then we connected each signal "1" with all other signals on the *time-circle* to form a network and calculated the entropy value for the network. That is, the distance between the two signals in the time series is converted into the distance between the two points on the *time-circle*. Except for the concept of the *time-circle*, it would difficult to find an appropriate way to calculate the BLS entropy profile for a time series. If we define the distance between signals in a time series as a branch length, the BLS entropy profile is dominated by some very distant signals regardless of the distribution of the entire signal. That is, information about the overall structure of the time series is diluted by some signals.

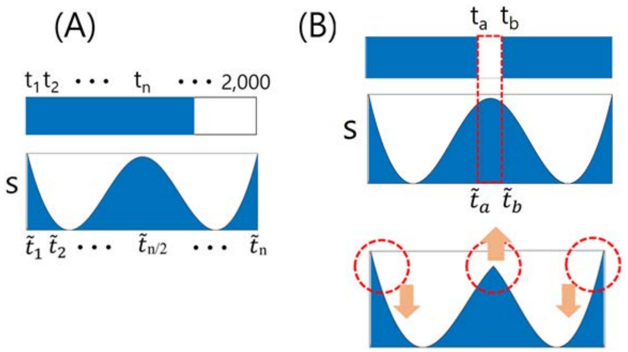

(A)
$t_1 t_2 \cdots t_n \cdots 2{,}000$
S
$\tilde{t}_1 \tilde{t}_2 \cdots \tilde{t}_{n/2} \cdots \tilde{t}_n$
(B)
t_a t_b
S
$\tilde{t}_a$ $\tilde{t}_b$

FIGURE 3. Process Peak generation mechanism. (A) A binary time series consisting of 200 consecutive "1" signals and their corresponding BLS entropy profile, (B) a time series in which all signals within the time interval [t_a, t_b] were removed and its BLS entropy profile.

III. OUR IDEAS AND FINDINGS

A. PEAK OCCURRENCE IN THE BLS ENTROPY PROFILE

The BLS entropy profile for a time series, $Q(t)$, consisting of 200 consecutive "1s" and 1800 "0s" has central axial symmetry (Fig. 3A). In this study, since the characteristics of the entropy profile were determined by the high and low levels rather than the actual entropy values, we normalized each profile to values in the range of [0, 1]. Here, the last signal of $Q(t)$ is changed from "0" to "1" for *time-circle* construction. t_1, t_2, ..., t_n on the time series correspond to $\tilde{t}_1, \tilde{t}_2, \ldots, \tilde{t}_n$ on the entropy profile. The profile was centrally convex. This is because there is only one signal band, and the center of the band has the highest entropy value due to left-right symmetry.

Next, we removed the time span [t_a, t_b] of equal length left and right from the center (Fig. 3B). Since the two endpoints t_a and t_b of the span are symmetrically equidistant from the center of the time series, the entropy values of $\tilde{t}_a$ and $\tilde{t}_b$ on the entropy profile are the same. It can be intuitively understood that the peak is formed by removing the part corresponding to the span from the entropy profile (see the red dotted line). Since the entropy profile has the property that partial strain affects the overall profile structure [25], the

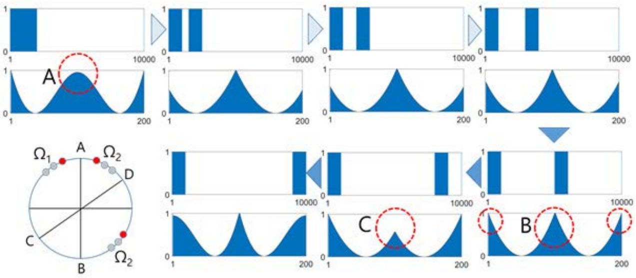


FIGURE 4. BLS entropy profile change as a function of distance *d* between two signal bands.

central peak height is increased and the height at both ends of the profile is relatively decreased (Fig. 3B). In summary, a peak is formed when the distribution of the left and right signals is the same for the time domain center where there is no signal. In addition, we can clearly infer that even if a band with a relatively low signal density replaces the removed domain, the peak is generated even though the height of the peak is slightly different compared to that for the removed domain.

B. PEAK HEIGHT CHANGE WITH DISTANCE BETWEEN SIGNAL BANDS

To understand the peak shape change on the BLS entropy profile, we investigated the entropy profiles for two signal bands Ω_1 ($w_1 = 1{,}000$ and $\delta_1 = 20$) and Ω_2 ($w_2 = 1{,}000$ and $\delta_2 = 20$) separated by a distance d. (Fig. 4). w_i and δ_i ($i = 1$, 2) represent the band length and inter-signal distance in the signal band, Ω_i, where the "1" signals are equally spaced. For this, we measured the peak height (H_peak) according to the change of d value. Here, the total length of the time series is L (=10,000). No peak was generated on the entropy profile when $d = 0$ (see red circle A). When $500 \leq d \leq 2{,}000$, H_peak was 1.0, and when $d = 5000$, three peaks with an H_peak value of 1.0 were formed (red circle B). In fact, the two peaks at both ends are combined on the *time-circle* to become one peak. The H_peak value of the central peak decreased sharply when $d = 8{,}000$ (red circle C). As the d value increases from 8,000 to 9,000, the H_peak value increases again. The peak height change can be understood by considering the signal distribution on the *time-circle* (bottom left in Fig. 4). Let us consider the case where Ω_1 is at arc AC and Ω_2 is at arc AD. Here, the entropy values of the red nodes within each band have the same entropy value because they are symmetric with respect to the line segment AB. In this case, as illustrated in Fig. 3, the entropy profile generated by the two bands has a sharp peak in the center. Even when Ω_2 is in arc BD, a peak occurs in the center of the entropy because Ω_1 and Ω2 are symmetric with respect to the line CD.

Figure 5 shows the change in H_peak for d and δ (=10, 20,…, 50). H_peak was significantly affected by d and relatively small in w. In order to show the effect of w in more detail, we investigated d^*, which is the d value that minimizes the H_peak value, and h^*, the minimum H_peak value (Fig. 6). As w increased, $d*$ decreased almost linearly and h^* decreased and increased. This indicates that the mutual influence between the bands is less complicated than what was anticipated.

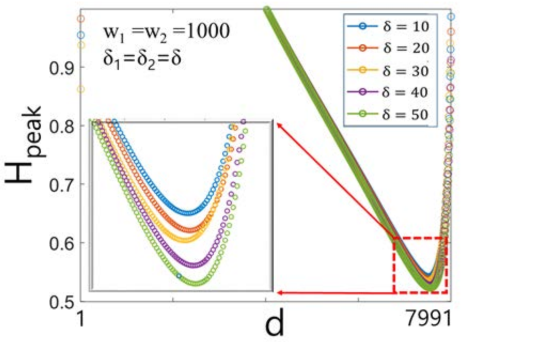


FIGURE 5. Change of peak height H_peak with separation distance d of two signal bands Ω_1 and Ω_2. Here, δ and w indicate the distance between the signals of Ω_1 and Ω_2 and the band length, respectively.

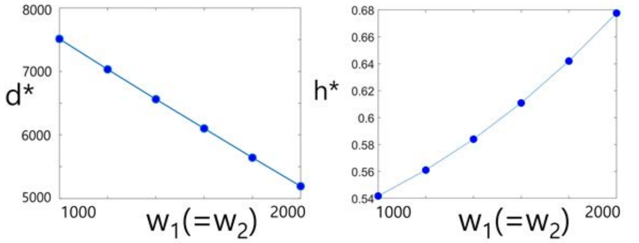


FIGURE 6. Distance d^* between the two signal bands that minimizes the peak height and the minimum height h^* of the peak.

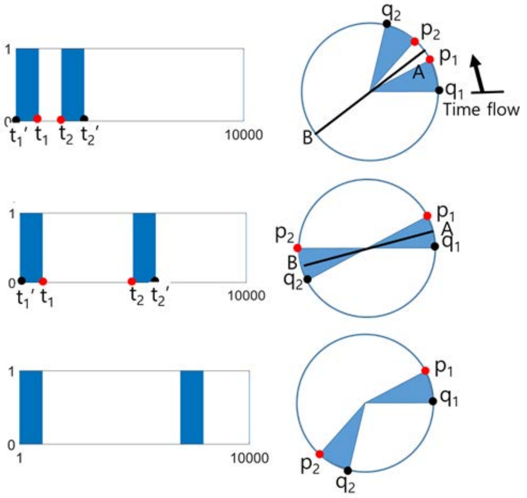


FIGURE 7. BLS entropy profile change as a function of distance *d* between two signal bands.

Figure 7 provides a qualitative understanding of the results of Fig. 5. t_1, t_2, t_1', and t_2' of the time series correspond to $p1$, $p2$, $q1$, $q2$ on the *time-circle*, respectively. When the two bands shown in Fig. 5 are close to each other ($d \leq 4{,}000$), a peak is formed between p_1 and p_2 (top of Fig. 7).

Here, since the entropy values at p_1 and p_2 are relatively higher than those at q_1 and q_2, a peak with an H_peak value of 1.0 occurs in the center of the entropy profile. When $d = 4{,}000$, $p1$, $p2$, $q1$, and $q2$ all have the same BLS entropy value. Therefore, in this case, two peaks with $H_\text{peak} = 1.0$,

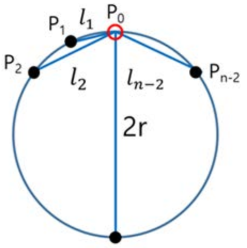


FIGURE 8. Case where signals within the time interval [101, 100+r] are removed from the time series (w = 1000, δ = 1) with L = 1000 and "1" signals; the difference between two entropy values S_1 (for t = 100) and S_2 (for t = 99), S_1-S_2, against r. Here, r = 1, 3, 5, ..., 599.

shown in circle B of Fig. 4, are generated (middle of Fig. 7). When $d > 4{,}000$, since the entropy values of p_1 and p_2 are smaller than the entropy values of q_1 and q_2, peaks with $H_{\text{peak}} = 1.0$ appear at both ends of the entropy profile, while peaks with $H_{\text{peak}} < 1.0$ are created in the center (bottom of Figure 7).

C. UPWARD PEAKS

All the peaks formed in the various binary time series we created were directed upwards. The shape of the peak contains information about the time series structure; thus, the direction of the peak can be used to understand the time series. As described earlier, a peak occurs when the "1" signals are removed for a time interval on a time series, and the *time-circle* for the time series makes the signal distribution symmetrical in a clockwise-counterclockwise direction with respect to the central axis of the interval. Thus, to prove that the peak is always upward, it must be mathematically demonstrated that the peak still points upward if consecutive "1" signals of any length are removed. The mathematical proof for removing one signal is included in Appendix A. However, it is not easy to prove when more than one signal is removed. Although this topic is very interesting for understanding peak characteristics, it is out of the scope of this study; thus, we leave it as a topic for future research. Instead of the mathematical proof, we numerically showed that the peak direction does not change even after the removal of consecutive signals (as shown in Fig. 8).

We created a binary time series with $L = 1000$ that included "1" signals at all times and "0" signals in the time interval [100, 100+r]. A peak was generated on the entropy profile for the time series. To verify whether the peak points upward, we calculated the entropy values S_1 and S_2 corresponding to $t = 100$ and 99. The difference value, S_1-S_2, was always positive, and as the r value increased, the difference value also increased. This means that the peak is upward irrespective of r.

D. CLIFF OCCURRENCE IN THE BLS ENTROPY PROFILE

For the signal distribution on the *time-circle*, if there were symmetry axes for the clockwise and counterclockwise directions, a peak was formed. Here we can ask the question what would happen to the entropy profile if there was no axis of symmetry.

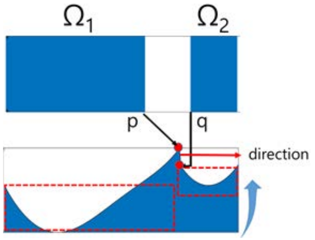


FIGURE 9. Mechanism by which a cliff is formed. The red arrow indicates the direction of the cliff.

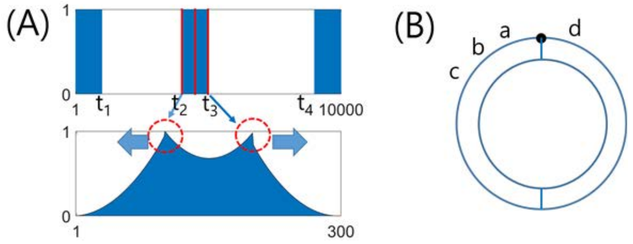


FIGURE 10. Mechanism for cliff direction determination. (A) Two cliffs created by two sections [t_1, t_2] and [t_3, t_4] between the three signal bands. The arrow points in the direction of the cliff. (B) Band positions on the time circle to understand the cliff direction. Here, t_1, t_2, t_3, and t_4 on the time series correspond to a, b, c, and d on the time-circle, respectively.

Figure 9 shows the time series with length $L = 10{,}000$, $Q(t)$, consisting of the two signal bands without the axis of symmetry Ω_1 ($w_1 = 6{,}000$, $\delta_1 = 10$) and Ω_2 ($w_2 = 6{,}000$, $\delta_2 = 10$). The range of $6{,}000 \leq t \leq 8{,}000$ has no signal. The entropy profile for $Q(t)$ has two concave parts. This is caused by the different entropy values of p and q for $t = 6{,}000$ and 8,000. We called this difference a cliff and defined the direction of the cliff as perpendicular to the direction of the height (see red arrow).

In our previous study [47], we mathematically proved that when two signal bands are on one time series, the entropy profile corresponding to the high signal density band is relatively more concave, whereas the entropy profile corresponding to the low signal density band is relatively less concave. Therefore, we can easily deduce that the entropy profile of Ω_1 is more concave downward compared to that of Ω_2.

E. CLIFF DIRECTION AND HEIGHT

Figure 10 shows a time series with $L = 10{,}000$ containing three signal bands. Each signal band has $w = 1{,}000$ and $\delta = 10$. When the time intervals in which the signal bands are separated from each other are [t_1, t_2] and [t_3, t_4], the interval does not have a symmetrical axis for the signal distribution on the *time-circle*. Thus, two cliffs are created. Here, we investigated the direction of the cliff by comparing the magnitudes of entropy values for a, b, c, and d corresponding to t_1, t_2, t_3, and t_4 in the time series (Fig. 10B). From comparing nodes a and b, we can easily infer that the entropy value of node a is higher than that of node b due to the fact that the position of node a is closer to the center of the time series than that of node b. Therefore, the cliff formed between nodes a and

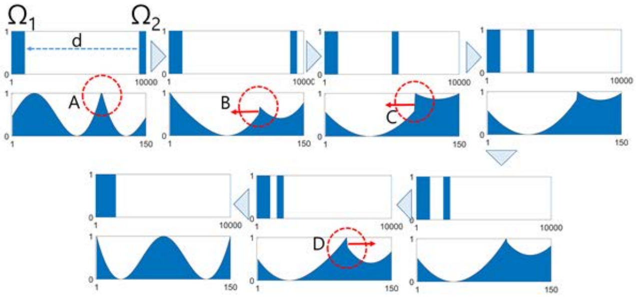


FIGURE 11. Changes in cliff height and direction for distance *d* between two signal bands of different lengths.

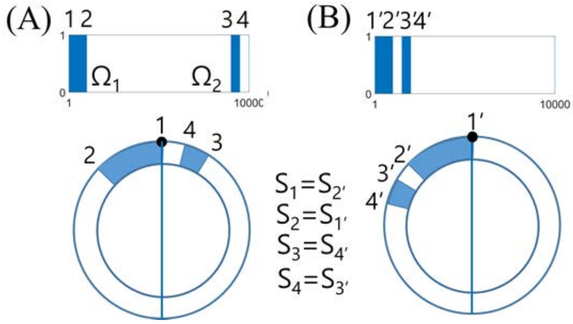


FIGURE 12. Cliff direction change on *time-circle* when the distance *d* between two signal bands is (A) relatively large and (B) small.

b points to the left. Comparing the entropy values at nodes c and d, node c has a relatively larger number of long branches while node d has shorter branches. In this case, based on the definition of entropy, Eq (2), it can be seen that the entropy value of node c is higher than that of d. That is, the cliff created between nodes c and d faces to the right.

To further understand the cliff orientation, we constructed a time series consisting of two signal bands, Ω_1 ($w_1 = 1{,}000$ and $\delta_1 = 10$) and Ω_2 ($w_2 = 500$ and $\delta_2 = 10$), separated by a distance d from each other (Fig. 11). We investigated the shape of the cliff by decreasing the d value from 8,500 to 0. For $d = 8{,}500$, no cliff was created because Ω_1 and Ω_2 are interconnected in the *time-circle* (see red circle A). When Ω_1 was fixed and Ω_2 shifted to the left ($d = 8{,}000$), a left-facing cliff was formed (see circle B). When $d = 4{,}000$, the direction of the cliff remained to the left and the height, h_c, increased. For $d = 500$, the cliff direction turned to the right and h_c decreased (see circle D). When the two bands made contact ($d = 0$), the cliff disappeared.

Figure 12 explains the cause of the change in direction of the cliff. For the case where the two bands are far apart, let the entropy values of the nodes on the *time-circle* corresponding to nodes 1, 2, 3, and 4 be S_1, S_2, S_3, and S_4, respectively (Fig. 12A). Let the entropy values for nodes 1', 2', 3', 4' be S_1', S_2', S_3', S_4', respectively, when the two bands are close (Fig. 12B). Comparing the entropy values of the two cases, we can see that $S_1 = S_2$', $S_2 = S_1$', $S_3 = S_4$', $S_4 = S_3$'. That is, the change in the direction of the cliff is a result caused by the relative change of the node position with respect to the flow direction of time.

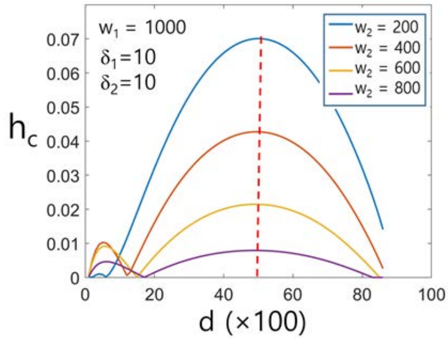


FIGURE 13. Plot of the cliff height, h_c, against the distance, d, between the two signal bands. Here, one band has length w_1 and inter-signal spacing, δ_1, and the other band has w_2 and δ_2.

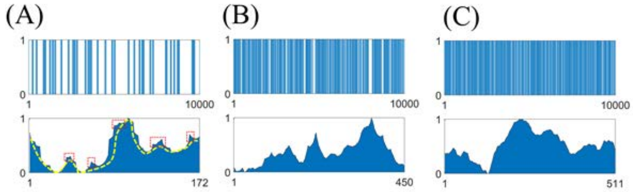


FIGURE 14. Peak, cliff, (see red boxes) and the global increase-decrease tendency (see yellow line) on BLS entropy profiles for time series with three different signal densities.

Figure 13 shows how the cliff height h_c varies with the spacing, d, between the two signal bands. One signal band Ω_1 has $w_1 = 1{,}000$, $\delta_1 = 10$ and the other band Ω_2 has $w_2 =$ 200, 400, 600, 800, $\delta_2 = 10$. For $d < 2{,}000$, h_c showed a small increase-decrease tendency. This is considered an effect caused by the nonlinear structure of the BLS entropy definition, Eq (2). When the value of d increased ($2{,}000 < d < 8{,}500$), h_c showed an increasing-decreasing trend regardless of the value of w_2. The symmetry structure for the tendency is due to the change in the relative positions of the nodes, as mentioned in Fig. 12.

F. PEAKS, CLIFFS, AND THE GLOBAL INCREASE-DECREASE TENDENCY ON BLS ENTROPY PROFILE

Figure 14 shows that changes in signal density are well captured through peaks, cliffs, and global increase-decrease tendencies of the entropy profile. We constructed a time series with $L = 10{,}000$ containing 40 randomly distributed signal bands ($w = 40$, $\delta = 1$) (Fig. 14A). Peaks and cliffs occurred in time intervals filled with "0" signals. Global information on the time series structure is reflected in the increase-decrease tendency (yellow dotted line) on the entropy profile, and local information is reflected in peaks and cliffs (red rectangle). When we increased the number of signal bands to 300 (Fig. 14B), many cliffs (or peaks) disappeared. This is because, as the distances between the bands decreased, the cliff (or peak) height decreased (see Figs. 5 and 13). On the other hand, the global increase-decrease tendency was more pronounced. When the number of signal bands was 1400 (Fig. 14C), most of the cliffs (or peaks) disappeared.

To better understand the global increase-decrease tendency, we investigated the entropy profiles for two simple

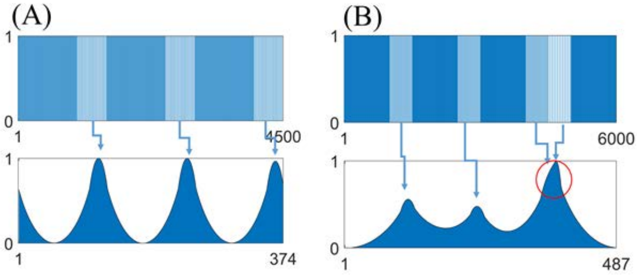


FIGURE 15. Time series consisting of connections of signal bands with different signal densities and their BLS entropy profile: (A) When two different signal bands are connected (B) When three different signal bands are connected. Here, each band has $w_1 = 1000$ and $\delta_1 = 10$, $w_2 = 500$ and $\delta_2 = 20$, and $w_3 = 500$ and $\delta_3 = 40$, respectively.

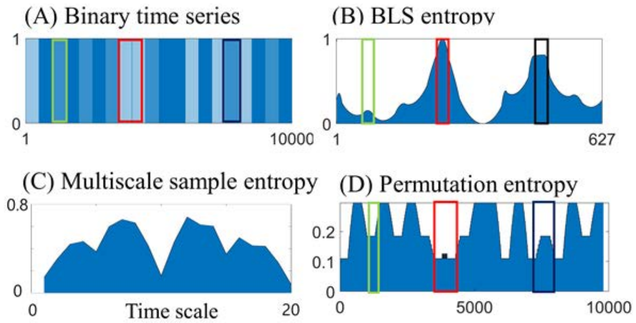


FIGURE 16. Three kinds of entropy for (A) a time series consisting of signal bands with different signal densities: (B) BLS entropy, (C) Multiscale sample entropy, and (D) Permutation entropy. Here, each band has $w_1 = 500$ and $\delta_1 = 10$, $w_2 = 500$ and $\delta_2 = 20$, and $w_3 = 500$ and $\delta_3 = 40$, respectively.

time series (Fig. 15). One time series consisted of two signal bands with different signal densities. One band had $w_1 = 1000$ and $\delta_1 = 10$ and the other band had $w_2 = 500$ and $\delta_2 = 20$. A relatively low signal density band formed a barrier (global increase-decrease tendency) in the entropy profile. If the low signal density band had a non-uniform signal, peaks and cliffs would be created on the barrier. To confirm this, we added one more signal band ($w_3 = 500$, $\delta_3 = 40$) (Fig. 15B). As expected, we could see that a right-biased structure was built above the third barrier (red circle). Due to the structure, the height of the other two barriers was relatively low.

The BLS entropy differ from existing entropies in that it contains both global and local properties for a given time series structure. To demonstrate this clearly, we calculated the BLS entropy profile, multiscale sample entropy [32] and permutation entropy [33] for a binary time series consisting of three different signal bands with different signal densities (Fig. 16). In the figure, the rectangles with the same color have the same time domain. In the time series (Fig. 16A), the red rectangle indicates the two bands with the lowest signal density, whereas the green and black rectangles indicate the band with the medium signal density. The BLS entropy profile showed that the profile values within the green and black rectangles are very different even though the two rectangles have the same signal density. This is due to the global information reflecting that the signal density around the green rectangle is higher than that around the black rectangle (Fig. 16B). On the other hand, multi-scale sample entropy is limited in capturing the dynamic characteristics of time series (Fig. 16C). Since the permutation entropy (delay = 1, order = 3, time window size = 200) contains only local information for the time series, the values within the green and black rectangles are the same (Fig. 16D).

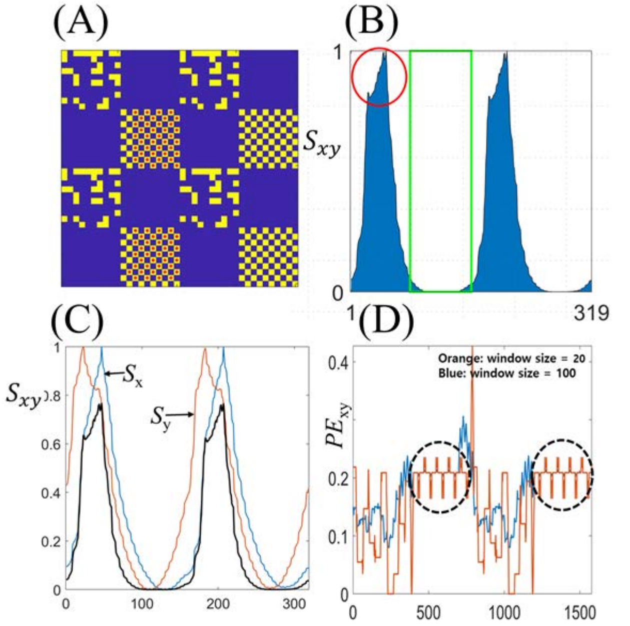


FIGURE 17. Detection of specific regions in binary images using BLS entropy profiles. (A) an image containing a check pattern and a random pattern, (B) BLS entropy profiles for two binary time series obtained by concatenation of images into rows and columns, S_{xy}, where the green rectangle indicates the selected section with a relatively low entropy profile value. (C) BLS entropy profiles obtained through row and column concatenation of images, S_x (by row) and S_y (by column), and S_{xy} (by S_x and S_y), (D) Permutation entropy (*PE*) profile obtained through row and column concatenation of images, PE_x (by row) and PE_y (by column), and PE_{xy} (by PE_x and PE_y). The degree and order values for PE_x and PE_y are 1 and 3, respectively.

IV. APPLICATION

A. DIRECTION OF A SPECIFIC REGION IN A SPATIAL DISTRIBUTION PATTERN

We applied our findings mentioned in the previous sections to the problem of detecting a specific region within a binary image. To this end, we generated a binary image with a grid size of 40 × 40 consisting of 4 random patterns and 4 checked patterns (Fig. 17A). The yellow and blue grids in the image represent 1's and 0's, respectively.

To detect a check pattern in an image, first we connect the second column of the image to the end of the first column and the third column to the end of the second column. We performed this process up to the 40th column to construct a binary vector (time series), Q_y, with $L = 1{,}600$. By applying the same method to the rows of the image, we obtained a binary time series, Q_x. Then, entropy profiles S_x and S_y were computed for the two time series Q_x and Q_y, respectively. Then we defined the BLS entropy profile for the image as below:

$$S_{xy} = S_x \times S_y \tag{3}$$

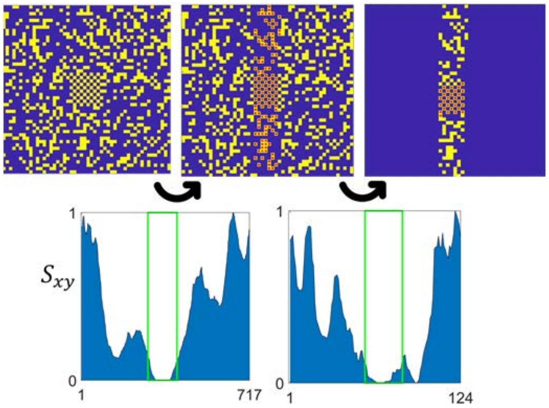


FIGURE 18. Check pattern detection in binary images' containing random noisy backgrounds. Here the yellow and blue grids represent "1" and "0" respectively.

where "×" is the element-wise product of two vectors of equal length S_x and S_y. The formation of two large periodicity barriers in S_{xy} indicates that the image has two periodic patterns (Fig. 17B).

From the fact that the check pattern has a higher signal density compared to the random pattern, we manually selected a region with a low entropy value between the two barriers (indicated by the green rectangle). The position on the image corresponding to the selected area is marked with a red dot. We successfully detected two of the four check patterns. The remaining two check patterns were detected by selecting a region with a low entropy value on the right side of the entropy profile. In this problem, we can see two advantages of the "element-wise product" of the S_{xy} definition.

One advantage is the containment of the phase difference information of S_x and S_y. In other words, as shown in Fig. 17C, if we only take either S_x or S_y, it becomes difficult to determine the boundary of the pattern we are looking for. The other advantage is that S_{xy} has a steeper barrier than that of S_x or S_y. This means that regions with different signal densities can be more easily distinguished. We compared our approach with the permutation entropy (*PE*) approach, which captures the dynamic properties of time-series. The comparison showed that the *PE* approach is limited in detecting the check pattern area (Fig. 17D). Here, the PE_{xy} was obtained by using the *PE* instead of *S* in Eq. (3). The PE_{xy} successfully captured regular regions (dotted circles) in the image, but showed difficulties in performing optimized thresholding to detect the check pattern regions. The difficulty is more likely to increase for more complex patterns. For the detection, it would be necessary to use a new thresholding algorithm or a formula different from Eq. (3).

To confirm that our method also works for more complex images, we created a random image of size 50 × 50 containing one check pattern of size 10 × 10 (see Fig. 18). We described the method in Appendix B. We first computed S_{xy} for this image (bottom left in Fig. 18). Since the signal density of the check pattern is relatively high compared to the other regions, we manually selected the region with the lowest entropy profile (green square). This selection successfully separated the vertical area containing the check pattern from the other areas. The detected area is marked with a red dot on the image (top center in Fig. 18). We computed S_{xy} once more for the selected region (bottom right of Fig. 18). The check pattern (upper right in Fig. 18) was successfully detected by reselecting the regions with relatively lower values in the entropy profile.

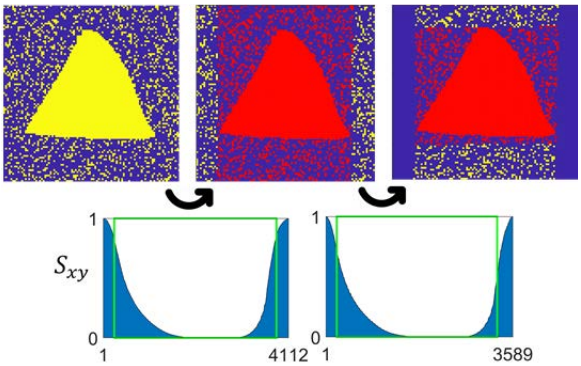


FIGURE 19. Triangle pattern detection in images with noisy backgrounds. Here the yellow and blue grids represent "1" and "0"respectively.

V. CONCLUSION AND FUTURE WORK

In this study, we explored the main features of the BLS entropy profile: peaks, cliffs, and global increase-decrease tendency. We found that peaks or cliffs are formed on the entropy profile when signal bands with different signal densities are adjacent to or separated from each other. Here, peaks and cliffs are created in the region of the entropy profile corresponding to the band of relatively low signal density. To understand more fundamentally, when a time series is mapped to a time-circle, a peak is formed if there are symmetrical axes in the clockwise and counterclockwise directions, otherwise a cliff is formed. When we consider both peaks and cliffs for the entire time series, we can see the global increase-decrease tendency on the entropy profile.

The tendency shown in the entropy profile of an image with spatially non-uniform signal density visualizes the density difference as a barrier (Fig. 17). Therefore, a barrier can be a useful indicator for determining a specific pattern boundary. However, as shown in Fig. 17, when we manually determined the inter-barrier interval, we missed a small part of the check pattern. Accurately detecting a specific pattern in an image is one of the most important issues in engineering and medical fields. Therefore, our method is novel and has advantages, but in order to increase its practicality, the problem of determining the interval between barriers needs to be explored more specifically. Through our preliminary study, it seemed that the distribution of signals "1" and "0" at the boundary of the pattern could affect the interval determination. Exploring this problem would be very interesting and worthwhile.

In Fig. 18, we successfully detected the check pattern through two entropy profile calculations. The two calculations were possible because the check pattern had a rectangular boundary. Furthermore, there is a need for a solution to detect a desired specific area, which is not a rectangle but an area surrounded by a closed curve. Figure 19 shows a noise image containing a triangular pattern. We performed two entropy profile calculations to detect a rectangular region (indicated by red dots) containing a triangular pattern. As a solution to this problem, we propose to divide the whole image into multiple images and detect the area of the triangular part in each image. Then, the triangular region can be extracted by summing all the partially detected regions. However, too many divisions may result in increased computational cost and reduced resolution of the signal density. Therefore, it is necessary to study division optimization.

This study is not only meaningful in that it provides an understanding of the entropy profile characteristics based on the concept of peaks and cliffs, in addition, it suggests a new method for detecting specific patterns in images. We believe that the BLS entropy profile has the potential to be applied to various problems in the field of pattern recognition by overcoming the above-mentioned problems.

APPENDIX A

Let P_0, P_1,..., P_n be the nodes on the *time-circle* corresponding to the "1s" in the time series in which the signal "1s" is distributed consecutively. Let l_i for $i=1,\ldots,n$ be the distance between P_0 and P_i. Let $L = l_2 + l_3 + \ldots l_n$ and $L'=l_1 + l_3 + \ldots l_n$ (See Fig. 19).

We consider the situation in which the first signal P_0 is removed. The entropy value of P_1 can be written as:

$$\mathrm{S}_1 = -\left[\frac{l_2}{L}log\left(\frac{l_2}{L}\right) + \frac{l_3}{L}log\left(\frac{l_3}{L}\right) + \ldots\frac{l_n}{L}log\left(\frac{l_n}{L}\right)\right] \times /\mathrm{log(n-1)} \quad (4)$$

Similarly, the entropy value at the P_2 is as follows:

$$\mathrm{S}_2 = -\left[\frac{l_1}{L'}log\left(\frac{l_1}{L'}\right) + \frac{l_3}{L'}log\left(\frac{l_3}{L'}\right) + \ldots\frac{l_n}{L'}log\left(\frac{l_n}{L'}\right)\right] \times /\mathrm{log(n-1)} \quad (5)$$

Therefore, the difference between S_1 and S_2 can be written as follows:

$$\begin{aligned}(\mathrm{S}_1 - \mathrm{S}_2)log(\mathrm{n}-1) &= -\left[\frac{l_2}{L}log\left(\frac{l_2}{L}\right) - \frac{l_1}{L'}log\left(\frac{l_1}{L'}\right)\right] \\ &\quad - \left[\frac{l_3}{L}log\left(\frac{l_3}{L}\right) - \frac{l_3}{L'}log\left(\frac{l_3}{L'}\right)\right] \\ &\quad - \left[\frac{l_4}{L}log\left(\frac{l_4}{L}\right) - \frac{l_4}{L'}log\left(\frac{l_4}{L'}\right)\right] \\ &\quad - \cdots - \left[\frac{l_n}{L}log\left(\frac{l_n}{L}\right) - \frac{l_n}{L'}log\left(\frac{l_n}{L'}\right)\right] \quad (6)\end{aligned}$$

If n is sufficiently large, it can be said that $L \approx L'$. Therefore, all terms after the second term can be all 0. Also, L_2 is always greater than L_1, $l_1 < l_2$. Therefore, if the following inequality is satisfied, the peak must always point upward.

$$\frac{l_2}{L}log\left(\frac{l_2}{L}\right) < \frac{l_1}{L}log\left(\frac{l_1}{L}\right) \quad (7)$$

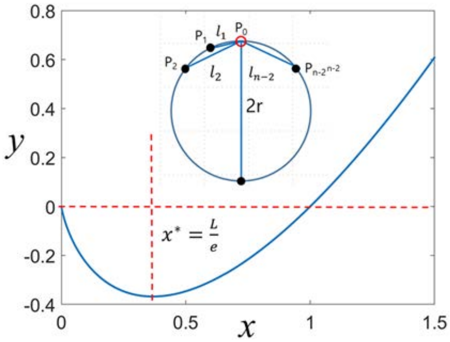


FIGURE 20. Illustration of peak pointing upwards. $p_1, p_2\ldots, p_\mathrm{n}$ denote nodes located on the time-circle with uniform intervals, and the distance between p_i and $p_\mathrm{i+1}$ is $l_\mathrm{i+1}$. In addition, p_0 is the removed node. x denotes l_1 and l_2, and y represents the function of $(x/L)\log(x/L)$.

The above inequality condition is satisfied as long as the function below increases.

$$y = \frac{x}{L}log\left(\frac{x}{L}\right)(L > 0), \quad (8)$$

The function for x is shown in Fig. 20. The range over which the function always increases must be greater than the point (x^*) where x is zero in the first derivative.

$$\begin{aligned} y' &= \frac{1}{L}log\left(\frac{x}{L}\right) + \frac{1}{L} = 0 \quad (9) \\ x^* &= \frac{L}{e}\end{aligned}$$

Finally, we get the following condition

$$l_2 < \frac{L}{e} \quad (10)$$

Since $L = l_2 + \ldots 2r + \ldots + l_{n-2} + l_{n-1} + l_n$, $2r$ is greater than n_2, and $L_{\mathrm{n}-1}$ is the same as L_2. Therefore, $el_2 < 3l_2 < L$ always holds. As a result, we proved that the peak generated when one signal is removed from a time series of evenly distributed signal "1s" always points upward.

APPENDIX B

Pseudo-code for an algorithm to find a specific "1" signal density region in a binary image

```
DATA: Binary image consisting of "0" and "1".
RESULT: A rectangular area for a specific density distribution of a "1" signal in a binary image.
1: $U$(i, j) ← binary image
2: Concatenate each row (column) in $U$(i, j) to form a binary time series $Q_\mathrm{x}$ ($Q_\mathrm{y}$).
3: Create time circles for $Q_\mathrm{x}$ and $Q_\mathrm{y}$ respectively (Fig. 2)
4: Generation of entropy profiles, $S_\mathrm{x}$ and $S_\mathrm{y}$ from two time circles (Fig. 2)
5: Calculate $S_\mathrm{xy} = S_\mathrm{x} \times S_\mathrm{y}$ (Eq. (3))
6: Select a specific section∗ in $S_\mathrm{xy}$
7: Extract the image area Ω(i, j) corresponding to the section from step 6 (Fig. 17)
8: Set $U$(i, j ∉ Ω) = 0 and then obtain $U\prime$ = rotate($U$, 90 degrees).
9: Repeat step 1 ∼ step 8 for $U\prime$.
```

10: Output the intersection of the two domains Ω and Ω′ obtained in step 7 and step 9.

* Specific section: (Example) When finding the region corresponding to the maximum signal density in the image, it is better to select a section close to 0 in S_{xy}, whereas in the case of the minimum signal density, select a section in which S_{xy} is close to 1.

CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this paper.

ACKNOWLEDGMENT

The authors would like to thank U. Jin Choi for encouraging us to develop ideas on the issue of applying BLS entropy to time series.

REFERENCES

[1] C. H. Fontes and O. Pereira, "Pattern recognition in multivariate time series—A case study applied to fault detection in a gas turbine," *Eng. Appl. Artif. Intell.*, vol. 49, pp. 10–18, Mar. 2016.

[2] X. Huang, Y. Ye, L. Xiong, R. Y. K. Lau, N. Jiang, and S. Wang, "Time series k-means: A new k-means type smooth subspace clustering for time series data," *Inf. Sci.*, vol. 1, pp. 367–368, 2016.

[3] H. Izakian, W. Pedrycz, and I. Jamal, "Fuzzy clustering of time series data using dynamic time warping distance," *Eng. Appl. Artif. Intell.*, vol. 39, pp. 235–244, Mar. 2015.

[4] Y. Sadahiro and T. Kobayashi, "Exploratory analysis of time series data: Detection of partial similarities, clustering, and visualization," *Comput., Environ. Urban Syst.*, vol. 45, pp. 24–33, May 2014.

[5] K. S. Tuncel and M. G. Baydogan, "Autoregressive forests for multivariate time series modeling," *Pattern Recognit.*, vol. 73, pp. 202–215, Jan. 2018.

[6] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," in *Proc. 4th Int. Conf. Found. Data Org. Algorithms*, 1993, pp. 69–84.

[7] C. Bettini, X. S. Wang, S. Jajodia, and J.-L. Lin, "Discovering frequent event patterns with multiple granularities in time sequences," *IEEE Trans. Knowl. Data Eng.*, vol. 10, no. 2, pp. 222–237, Mar. 1998.

[8] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast subsequence matching in time-series databases," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 1994, pp. 419–429.

[9] E. J. Keogh and M. J. Pazzani, "Relevance feedback retrieval of time series data," in *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 1999, pp. 183–190.

[10] E. Keogh and M. Pazzani, "A Simple dimensionality reduction technique for fast similarity search in large time series databases," in *Proc. 4th Pacific–Asia Conf. Knowl. Discovery Data Mining*, 2000, pp. 122–133.

[11] J. Abonyi, B. Feil, S. Nemeth, and P. Arva, "Modified Gath–Geva clustering for fuzzy segmentation of multivariate time-series," *Fuzzy Sets Syst.*, vol. 149, no. 1, pp. 39–56, Jan. 2005.

[12] Z. Han, Y. Li, Y. Du, W. Wang, and G. Chen, "Noncontact detection of earthquake-induced landslides by an enhanced image binarization method incorporating with monte-carlo simulation," *Geomatics, Natural Hazards Risk*, vol. 10, no. 1, pp. 219–241, Jan. 2019.

[13] M. D. Abramoff, P. J. Magalhaes, and S. J. Ram, "Image process with ImageJ," *Biophoton. Int.*, vol. 11, no. 7, pp. 36–42, 2004.

[14] R. Bruni, "Stock market index data and indicators for day trading as a binary classification problem," *Data Brief*, vol. 10, pp. 569–575, Feb. 2017.

[15] K. Yang and C. Shahabi, "A PCA-based similarity measure for multivariate time series," in *Proc. 2nd ACM Int. Workshop Multimedia Databases (MMDB)*, 2004, pp. 65–74.

[16] A. Amirteimoori and S. Kordrostami, "A Euclidean distance-based measure of efficiency in data envelopment analysis," *Optimization*, vol. 59, no. 7, pp. 985–996, 2010.

[17] H. Li and C. Wang, "Similarity measure based on incremental warping window for time series data mining," *IEEE Access*, vol. 7, pp. 3909–3917, 2019.

[18] A. Khaleghi, D. Ryabko, J. Mary, and P. Preux, "Consistent algorithms for clustering time series," *J. Mach. Learn. Res.*, vol. 17, no. 3, pp. 1–32, 2016.

[19] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. KDD Workshop*. Seattle, WA, USA, Apr. 1994, pp. 359–370.

[20] E. J. Keogh and M. J. Pazzani, "Derivative dynamic time warping," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2001, pp. 1–11.

[21] M. Bhaduri and J. Zhan, "Using empirical recurrence rates ratio for time series data similarity," *IEEE Access*, vol. 6, pp. 30855–30864, 2018.

[22] M. M. Zhang and D. Pi, "A new time series representation model and corresponding similarity measure for fast and accurate similarity detection," *IEEE Access*, vol. 5, pp. 24503–24519, 2017.

[23] F. K. P. Chan, A. W. C. Fu, and C. Yu, "Haar wavelets for efficient similarity search of time-series: With and without time warping," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 3, pp. 686–705, May 2003.

[24] S.-H. Lee and C.-M. Park, "A new measure to characterize the self-similarity of binary time series and its application," *IEEE Access*, vol. 9, pp. 73799–73807, 2021.

[25] S.-H. Lee, C.-M. Park, and U. Choi, "A new measure to characterize the degree of self-similarity of a shape and its applicability," *Entropy*, vol. 22, no. 9, p. 1061, Sep. 2020.

[26] O. Kwon and S.-H. Lee, "Properties of branch length similarity entropy on the network in R^k," *Entropy*, vol. 16, no. 1, pp. 557–566, Jan. 2014.

[27] S.-H. Lee, P. Bardunias, and N.-Y. Su, "A novel approach to shape recognition using shape outline," *J. Korean Phys. Soc.*, vol. 56, no. 3, pp. 1016–1019, Mar. 2010.

[28] S.-H. Lee, "Robustness of branch length similarity entropy approach for noise-added shape recognition," *J. Korean Phys. Soc.*, vol. 57, no. 3, pp. 501–505, Sep. 2010.

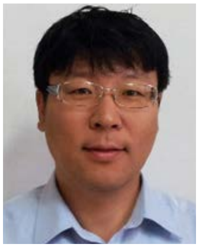
[29] O. Kwon and S.-H. Lee, "Branch length similarity entropy-based descriptors for shape representation," *J. Korean Phys. Soc.*, vol. 71, no. 10, pp. 727–732, Nov. 2017.

[30] S.-H. Kang, J.-H. Cho, and S.-H. Lee, "Identification of butterfly based on their shapes when viewed from different angles using an artificial neural network," *J. Asia–Pacific Entomol.*, vol. 17, no. 2, pp. 143–149, Jun. 2014.

[31] Y. Choi, W. Jeon, S.-H. Kang, S.-H. Lee, and T.-S. Chon, "Characterizing temporal patterns in the swimming activity of caenorhabditis elegans," *J. Korean Phys. Soc.*, vol. 60, no. 11, pp. 1840–1844, Jun. 2012.

[32] Y. Wang, Z. Liang, L. J. Voss, J. W. Sleigh, and X. Li, "Multi-scale sample entropy of electroencephalography during sevoflurane anesthesia," *J. Clin. Monitor. Comput.*, vol. 28, no. 4, pp. 409–417, Aug. 2014.

[33] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, no. 17, Apr. 2002, Art. no. 174102.



SANG-HEE LEE received the Ph.D. degree from Pusan National University, in 2005, for research on nonlinear dynamics related to biological systems. He is currently a Postdoctoral Researcher, conducted research on termite social behavior and control at the Entomological Institute, University of Florida. He is also a Senior Researcher at the National Institute for Mathematical Sciences, working to solve various industrial problems. He holds two patents and has published more than 120 journal articles covering the fields of animal behavior, ecological dynamics, and infectious disease control.

CHEOL-MIN PARK received the Ph.D. degree in mathematics from Seoul National University, Seoul, South Korea, in 2006. He has been a Researcher at the National Institute for Mathematical Sciences, Daejeon, South Korea, since 2011. His research interests include algorithms, number theory, and cryptography.