Efficient Approach for Computing the Discrimination Ratio-Based Variant of Information Entropy for Image Processing

PEICHAO GAO [1,2,3], **HONG ZHANG** [4], **DUO JIA** [2], **CHANGQING SONG** [2], **CHANGXIU CHENG** [1,2,5], **AND SHI SHEN** [2]

[1] State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China
[2] Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China
[3] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[4] Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu 611756, China
[5] National Tibetan Plateau Data Center, Beijing 100101, China

Corresponding author: Changxiu Cheng (chengcx@bnu.edu.cn)

ABSTRACT Information content is an important criterion for many image processing algorithms such as band selection and image fusion. Usually, information content is quantified by using information entropy (i.e., Shannon entropy); however, this is not a suitable measure because information entropy is independent of the spatial distribution of image pixels. Thus, improved information entropies and variants of information entropy have been developed. Among all the entropic measures, the discrimination ratio-based variant of information entropy (hereinafter DVIE) has recently been demonstrated to be the most effective. On the other hand, DVIE is the most inefficient measure in terms of computation time, which severely restricts its applications. To solve this problem, we present a three-strategy approach to efficiently compute the DVIE of an image. The first strategy is to use a simplified equation for DVIE. The second strategy is to selectively compute the two computationally intensive components of DVIE—intra-distance and extra-distance—based on the computational complexity. Only one distance was computed, and the other distance was derived based on the lookup table of average distances. The third strategy was to efficiently construct the lookup table based on geometric symmetry. We performed both validation and evaluation experiments to demonstrate that the proposed approach was not only valid for accurately computing DVIE, but it was also highly efficient. Our proposed approach saved more than 99% of the time taken for the original approach, without compromising the accuracy; therefore, DVIE was made applicable for processing images.

INDEX TERMS Band selection, discrimination ratio, image processing, information content, information entropy, information-theoretic, Shannon entropy.

I. INTRODUCTION

Image processing is usually the first step towards applying optical remote sensing data [1]. Its primary purpose is to improve the quality of an image and make the image suitable for certain applications [2]. However, one major problem faced during image processing is the objective assessment of the image quality [3], [4]. One simple and

The associate editor coordinating the review of this manuscript and approving it for publication was Francisco J. Garcia-Penalvo.

efficient method for this assessment is to employ information entropy (IE, also called Shannon entropy) [5]–[9], which is an information-theoretic metric that quantifies the information content of a dataset [10]. This method has been extensively used because of its theoretical elegance and practical simplicity, which helped create numerous image processing algorithms [11]–[13].

Despite its popularity, IE is not a perfect measure of the information content of an image. It characterizes only the statistical information of an image; it disregards the spatial

distribution of pixels [14]. However, this distribution is too important to be disregarded; therefore, an image can be called a spatial dataset. To solve this problem, efforts have been made to improve the IE [15]; these efforts have produced numerous improved information entropies and variants of information entropies [e.g.,] [16], [17]. Among them, discrimination ratio-based variant of IE (hereinafter DVIE) has been shown to be the most effective measure for characterizing the spatial distribution [15], [18].

However, DVIE is the most inefficient measure in terms of the computation time. According to Gao, *et al.* [15], the computation time of DVIE is 18,257 times that of the IE. The DVIE still takes 702 times as long as the second most inefficient measure. On a powerful desktop computer, it takes 37 s to evaluate an image of only 150 × 215 pixels [15]. Another serious concern is that the computation time of DVIE increases exponentially when the image size increases; this will be demonstrated in Section II. This efficiency problem has become a bottleneck in using DVIE to process images.

Therefore, the aim of this study is to remove this bottleneck by developing an efficient approach to compute DVIE. This approach will make it practically possible for DVIE to evaluate images on a personal computer environment.

The remainder of the article is organized as follows. Section II provides an overview of DVIE; it is compared with IE, and its efficiency problem is discussed. Section III explores the key idea of the proposed approach for efficiently computing DVIE; this section explains the three strategies proposed to reduce the amount of computation. Sections IV – VI further describe the three strategies, namely a simplified equation for efficiently computing DVIE, a complexity-based method for the selective computation of the intra- and extra-distance components, and a symmetry-based method for efficiently constructing the lookup table. Section VII presents the experiments for validating and evaluating the proposed approach. Section VIII illustrates the application of DVIE to process remote sensing image. Sections IX – X provide discussions on the importance of the proposed approach; we end with a few concluding remarks.

II. DVIE

A. DVIE VS. IE

IE was developed by Shannon [5], who is considered the father of information theory. IE is a measure of information or disorder (these two terms are often used interchangeably). The mathematical formula of IE is as follows:

$$IE = -\sum_{i=1}^{n} P(X = x_i) \cdot \log P(X = x_i) \qquad (1)$$

where X is a discrete random variable with the possible values $\{x_1, x_2, \cdots, x_n\}$, and i is an integer variable with possible values ranging from 1 to n. $P(X = x_i)$ is the probability of X taking the value of x_i; in other words, it is the proportion of all the instances of X with the value x_i.

Let us take the image shown in Fig. 1(a) as an example. In this case, X is the image pixel. It has three possible

4	5	4	4	IE = 1.4591
6	4	5	4	DVIE = 1.1853
6	4	5	5	(a)

4	4	4	4	IE = 1.4591
6	6	4	4	DVIE = 1.0516
5	5	5	5	(b)
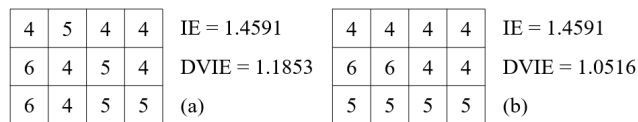
FIGURE 1. Two simple images and their information entropies (IEs) and their discrimination ratio-based variant of information entropies (DVIEs).

values, namely $x_1 = 4$, $x_2 = 5$, and $x_3 = 6$, with the probability of $P(X = x_1) = 6/12$, $P(X = x_2) = 4/12$, and $P(X = x_3) = 2/12$, respectively. As a result, the IE of this image is computed as 1.4591 using (1). Note that the IE remains unchanged even if the spatial distribution of pixels is altered, as shown in Fig. 1(b). This means that according to information theory [19], [20], the two images in Fig. 1 have exactly the same information content; in more general terms, any two images are regarded to have the same information content as long as they have the same pixel composition. However, an image is "a specific composition and configuration of pixels" [21]. Therefore, a perfect measure of the information content of an image should be capable of characterizing both the compositional and configurational information of an image. Clearly, IE fails to characterize the latter.

To address this problem, Claramunt [18] developed a variant of IE based on Tobler's first law of geography, which stated that "everything is related to everything else, but near things are more related than distant things" [22]–[24]. This law indicates that in a general case, similar entities are close in space. Claramunt [18] regarded such a case as an orderly state and assumed that as similar entities come close, the patterns of their entities would become more orderly. Similarly, as different entities come close, the pattern of their entities would become more disorderly. In dealing with an image, one can interpret the entities as pixels, and the pattern of entities as the whole image. To quantify the closeness of similar and different pixels, Claramunt [18] proposed the concepts of intra-distance and extra-distance, respectively. These two distances can be computed as follows:

$$d_{in}(X = x_i) = \begin{cases} \dfrac{1}{N_i \cdot (N_i - 1)} \displaystyle\sum_{j=1}^{N_i} \sum_{k=1}^{N_i} |A_{ij}A_{ik}| & N_i > 1 \\ \lambda & N_i = 1 \end{cases} \qquad (2)$$

$$d_{ex}(X = x_i) = \begin{cases} \dfrac{1}{N_i \cdot (N - N_i)} \displaystyle\sum_{j=1}^{N_i} \sum_{k=1}^{N - N_i} |A_{ij}B_{ik}| & N_i \neq N \\ \lambda & N_i = N \end{cases} \qquad (3)$$

where N is the total number of the pixels of an image, and λ is a pre-set parameter that takes a small value (e.g., 0.1 or 0.2). N_i, $d_{in}(X = x_i)$, and $d_{ex}(X = x_i)$ are the number, the intra-distance, and the extra-distance of the pixels at the ith gray level (i.e., pixels with a value of x_i), respectively. A_{ij} is

the jth pixel at the ith gray level, and B_{ik} is the kth pixel that does not belong to the ith gray level. $\left|A_{ij}B_{ik}\right|$ is the Euclidean distance between A_{ij} and B_{ik}. From (2) and (3), we can see that $d_{in}(X = x_i)$ is the average distance between the two pixels at the ith gray level, and $d_{ex}(X = x_i)$ is the average distance between a pixel at the ith gray level and a pixel at another gray level. In a sense, $d_{in}(X = x_i)$ and $d_{ex}(X = x_i)$ are analogous to the concepts of "within-class variance" and "between-class variance" [25].

Based on the intra-distance and extra-distance, the variant of IE is computed as follows:

$$DVIE = -\sum_{i=1}^{n} \frac{d_{in}(X = x_i)}{d_{ex}(X = x_i)} \cdot P(X = x_i) \cdot \log P(X = x_i) \tag{4}$$

In this variant, the intra-distance and extra-distance are combined as a ratio [26], which Leibovici, *et al.* [27] called a discriminant ratio. Therefore, this variant has been referred to as the DVIE in this paper.

To illustrate the computation of DVIE, let us take Fig. 1(a) again as an example. First, using (2), the intra-distances were computed to be $d_{in}(X = 4) = 1.9452$, $d_{in}(X = 5) = 1.6488$, and $d_{in}(X = 6) = 1$, according to (2). Then, using (3), the corresponding extra-distances were computed to be $d_{ex}(X = 4) = 1.8637$, $d_{ex}(X = 5) = 1.8611$, and $d_{ex}(X = 6) = 2.2054$. Finally, using (4), the DVIE of Fig. 1(a) was computed to be 1.1853 with the logarithmic base of 2. Similarly, the DVIE of Fig. 1(b) was also computed, and the result was 1.0516. We can see that DVIE is capable of distinguishing Fig. 1(b) from Fig. 1(a), namely it can characterize the configurational information of an image. Indeed, an experimental evaluation involving 100,000 images [15] reveals that DVIE is the most effective measure for characterizing configurational information.

B. EFFICIENCY PROBLEM OF DVIE
Although it is effective, DVIE is a highly computation-intensive measure. From (2) – (4), we can see that the basic computation involves calculating the distance between two pixels. The amount of such basic computation (A_C) can be expressed as follows using an image of N pixels:

$$A_C = \sum_{i=1}^{n}(N_i - 1) \cdot N_i + \sum_{i=1}^{n}(N - N_i) \cdot N_i$$
$$= \sum_{i=1}^{n}(N - 1) \cdot N_i \tag{5}$$

This involves an extremely large number of basic computational steps, even with a small image. For example, let us assume that the size of the image is 600×600 pixels, and these pixels are evenly distributed in five gray levels. The A_C of this image is 1.2960×10^{11}, which is very large.

III. IMPROVING THE COMPUTATIONAL EFFICIENCY OF DVIE
In general, there are two main ways to improve the computational efficiency of a measure; one method uses hardware, and the other uses software. The hardware method employs advanced computing devices such as cloud computing platforms, graphics processing units (GPUs), and tensor processing units (TPUs). This improves the computational efficiency by increasing the amount of computation per unit of time. By contrast, the software method develops new or improved methods for computing a measure. The computational efficiency is improved by either increasing the amount of computation per unit of time or reducing the amount of computation involved.

In this study, we focused on the software method because it is more cost-effective than the hardware method for most users. Our core idea involves three strategies. The first strategy was to simplify the equation for DVIE. The second strategy was to selectively compute the intra-distance and extra-distance based on the computational complexity. Only one distance was computed, and the other distance was derived based on a lookup table of average distances (LTAD). The third strategy was to efficiently construct the lookup table based on geometric symmetry.

IV. A SIMPLIFIED EQUATION FOR EFFICIENTLY COMPUTING DVIE
The DVIE equation can be simplified according to the property of distance. One basic property of the distance between the two points A_1 and A_2 on a Euclidean plane is commutativity, namely $|A_1A_2| = |A_2A_1|$. Using the commutative property, a simplified equation can be proposed for deriving the intra-distance for computing the DVIE:

$$d_{in}(X = x_i) = \begin{cases} \dfrac{2}{N_i \cdot (N_i - 1)} \sum_{j=1,}^{N_i} \sum_{k=j+1}^{j<k\leq N_i} \left|A_{ij}A_{ik}\right| & N_i > 1 \\ \lambda & N_i = 1 \end{cases} \tag{6}$$

By using (6), it is possible to avoid $0.5 \times \sum_{i=1}^{n}(N_i - 1) \cdot N_i$ basic computations. As a result, the number of basic computations is reduced to A'_C, as follows:

$$A'_C = 0.5 \times \sum_{i=1}^{n}(N_i - 1) \cdot N_i + \sum_{i=1}^{n}(N - N_i) \cdot N_i$$
$$= \sum_{i=1}^{n}(N - 0.5N_i - 0.5) \cdot N_i \tag{7}$$

For example, for a 600×600 image whose pixels are evenly distributed in five gray levels, the number of basic computations is reduced from 1.2960×10^{11} to 1.1664×10^{11}; that is, 1.2960×10^{10} basic computations are decreased.

V. COMPLEXITY-BASED METHOD FOR SELECTIVE COMPUTATION OF INTRA-DISTANCE AND EXTRA-DISTANCE

In the original method for computing DVIE, both the intra- and the extra-distance components needed to be computed for each gray level. In this study, we propose to selectively compute the intra- and the extra-distance components. Only one distance was selected for the computation, and the other distance was derived based on the first distance. This selective computation was performed for each gray level of an image in the following three steps.

A. COMPUTATIONAL COMPLEXITY-BASED SELECTION

This step selects one of the two distances having lower computational complexity. The computational complexities of the intra- and the extra-distance components were characterized using the number of basic computations involved in computing the two distances. As shown in (7), the number of basic computations for the intra-distance of the ith gray level is $0.5 \times (N_i - 1) \cdot N_i$, whereas that for the extra-distance of the ith gray level is $(N - N_i) \cdot N_i$. The selection was performed as follows. If $0.5 \times (N_i - 1) \cdot N_i$ is greater than $(N - N_i) \cdot N_i$, then the intra-distance is selected; otherwise, the extra-distance is selected.

B. NUMERICAL COMPUTATION OF THE SELECTED DISTANCE

In this step, the distance selected in the previous step was determined by performing all the basic computations required by this distance, as given by (3) and (6). For example, if the intra-distance was selected, then it was determined according to (6). This determination was numerical in nature; therefore, it could be called a numerical computation.

C. ANALYTICAL COMPUTATION OF THE UNSELECTED DISTANCE

The other distance (i.e., the unselected distance) was analytically computed according to the relationship between the intra- and extra-distance components. This relationship was derived as follows.

First, the following relationship holds for any pixel of an image:

$$(N - 1) \times d_{avg}(P_s) = \sum_r |P_s P_r| + \sum_t |P_s P_t| \quad (8)$$

where N is the total number of pixels of an image. P_s is the sth pixel of the image, and it belongs to the ith gray level; $d_{avg}(P_s)$ is the average distance between P_s and all the other pixels of the image. P_r and P_t are the other pixels that have and do not have the same gray value, respectively, as P_s. The function $|\cdot|$ returns the distance between the two pixels.

Second, the following relationship can be derived for any gray level based on (8):

$$(N - 1) \times \sum_s d_{avg}(P_s) = \sum_s \sum_r |P_s P_r| + \sum_s \sum_t |P_s P_t| \quad (9)$$

Third, (9) can be rewritten in terms of intra- and extra-distance components, as follows:

$$(N - 1) \sum_s d_{avg}(P_s) = N_i (N_i - 1) d_{in}(X = x_i)$$
$$+ N_i (N - N_i) d_{ex}(X = x_i) \quad (10)$$

Equation (10) implies that we can directly derive the intra-distance based on the extra-distance and vice versa, if we have a lookup table for $d_{avg}(P_s)$. A lookup table is an array or a matrix of pre-computed values for a time-consuming function; more details on this concept can be found in [28]. In the following section, we will show how to efficiently construct such a lookup table.

VI. SYMMETRY-BASED METHOD FOR EFFICIENTLY CONSTRUCTING A LOOKUP TABLE

As mentioned in Section V, we need an LTAD. The LTAD is a matrix whose size is the same as the image under consideration. LTAD (r, c) stores the average distance of the pixel on the rth row and cth column of the image from all other pixels.

To construct the LTAD of an image, the most direct approach is to determine the elements of LTAD one by one. However, this approach is inefficient because it involves a substantial amount of computation. For an image of $R \times C$ pixels, determining a single element of LTAD requires $R \times C - 1$ basic computations, and determining all the elements of LTAD requires $R^2 C^2 - RC$ (e.g., 1.2960×10^{11} for a 600×600 image) basic computations.

To reduce the amount of computation, we developed an improved method in this study by constructing the LTAD of an image based on geometric symmetry. The method involved three steps. The first step was to determine the upper-left quarter of the elements. Mathematically, if the image has $R \times C$ pixels, then the range of the row numbers of these elements and the range of the column numbers are $\left[0, \lfloor R/2 \rfloor\right]$ and $\left[0, \lfloor C/2 \rfloor\right]$, respectively, as shown in Fig. 2. The second step was performed based on the line symmetry to directly obtain the upper-right and the bottom-left quarters of the elements according to the mapping functions in (11), (12), as shown at the bottom of the next page. The last step was based on the point symmetry, namely to directly obtain the bottom-right quarters of the elements by using (13). The number of basic computations was $\lceil R/2 \rceil \times \lceil C/2 \rceil \times (R \times C - 1)$ with this improved method (e.g., 0.3240×10^{11} for a 600×600 image). Therefore, our proposed method reduced three-quarters of computations used for the direct method.

To further reduce the amount of computation, we developed a method for efficiently determining the upper-left quarter of the elements of a LTAD. This method was implemented on two other lookup tables, which were efficiently constructed based on the geometric symmetry.

The first lookup table (denoted by LT1) was constructed to store the distance between any two pixels. The size of this lookup table was the same as that of the image under

FIGURE 2. Upper-left quarter of elements (the gray area) of a lookup table of average distances (LTAD).

consideration, namely $R \times C$. An element of this lookup table, LT1 (r, c), denotes the distance between Pixels (x_1, y_1) and (x_2, y_2), where $\Delta x = |x_1 - x_2| = r$ and $\Delta y = |y_1 - y_2| = c$. The value of this element can be computed as $\sqrt{r^2 + c^2}$. Not all the values of the elements needed to be computed. According to the geometric symmetry, the following computations can be avoided to improve efficiency.

- If the image has more columns than rows (i.e., a wide image, $R \leq C$), then only the elements in the upper-right part (Fig. 3) of LT1 needed to be determined. The spatial extent of these elements can be mathematically expressed as $0 \leq r \leq R - 1$ and $r \leq c \leq C - 1$.
- If the image has fewer columns than rows, then only the elements in the bottom-left part (Fig. 4) of LT1 needed to be determined. Their spatial extent was $c \leq r \leq R - 1$ and $0 \leq c \leq C - 1$.

The second lookup table (denoted by LT2) was constructed to store the total distance between a pixel and a matrix of $r \times c$ pixels that were placed at the upper-left, upper-right, bottom-left, or bottom-right corners of the pixel (see Fig. 5). The size of LT2 was also the same as the image under consideration,

FIGURE 3. Elements involved in constructing LT1 with a wide image.

namely $R \times C$. The elements of LT2 were determined based on LT1. Similar to the LT1 case, not all the elements of LT2 needed to be determined. Only a few elements needed to be determined according to the geometric symmetry, as specified in the following cases.

- If the image had more columns than rows (i.e., a wide image, $R \leq C$), then only the upper-right part of the elements of LT2 needed to be determined; the elements in the first row were excluded. According to the definition of LT2, all the elements in the first row had zero value. The spatial extent of the elements to be determined is shown in Fig. 6; this is mathematically expressed as $1 \leq r \leq R - 1$ and $r \leq c \leq C - 1$. The determination of these elements is discussed in two subcases. The first subcase was for the elements at the diagonal of LT2. These elements were determined using (14) and (15), as shown at the bottom of this page. The second subcase

$$
\text{LTAD}(r, c) =
\begin{cases}
\text{LTAD}(r, C - c - 1) & r \in [0, \lfloor R/2 \rfloor], c \in [\lceil C/2 \rceil, C) & (11) \\
\text{LTAD}(R - r - 1, c) & r \in [\lceil R/2 \rceil, R), c \in [0, \lfloor C/2 \rfloor] & (12) \\
\text{LTAD}(R - r - 1, C - c - 1) & r \in [\lceil R/2 \rceil, R), c \in [\lceil C/2 \rceil, C) & (13)
\end{cases}
$$

$$
\text{LT2}(r, c) =
\begin{cases}
LT1(r, c) & r = c = 1 & (14) \\
LT2(r-1, r-1) + LT1(r, r) + 2 \sum_{i=1}^{i=r-1} LT1(i, r) & 2 \leq r = c < R & (15) \\
LT2(r, c-1) + \sum_{i=1}^{i=r} LT1(i, c) & 2 \leq r < R, r+1 \leq c < C & (16)
\end{cases}
$$

$$
\text{LT2}(r, c) =
\begin{cases}
LT1(r, c) & r = c = 1 & (17) \\
LT2(c-1, c-1) + LT1(c, c) + 2 \sum_{i=1}^{i=c-1} LT1(i, c) & 2 \leq c = r < C & (18) \\
LT2(r-1, c) + \sum_{i=1}^{i=c} LT1(r, i) & 2 \leq c < C, c+1 \leq r < R & (19)
\end{cases}
$$

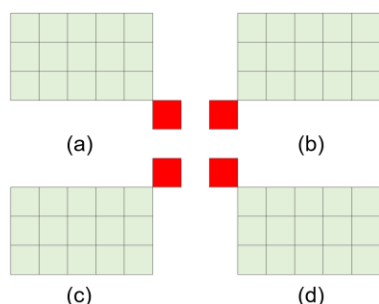**FIGURE 4.** Elements involved in constructing LT1 with a long image.

FIGURE 5. Matrix of pixels (displayed in green) that are placed at the (a) upper-left, (b) upper-right, (c) bottom-left, or (d) bottom-right corners of a single pixel (displayed in red).

FIGURE 6. Elements involved in constructing LT2 with a wide image.

FIGURE 7. Elements involved in constructing LT2 with a long image.

was for the other elements, which were determined using (16), as shown at bottom of the previous page.

- If the image had fewer columns than rows, only the bottom-left part of the element of LT2 needed to be determined. The elements in the first column were excluded because they had zero by the definition of LT2. The spatial extent of the elements to be determined is shown in Fig. 7; this is mathematically expressed as $1 \leq c \leq C - 1$ and $c \leq r \leq R - 1$. To determine their values, we also distinguished two subcases. The first subcase was for the elements at the diagonal of LT2, whose values were computed using (17) and (18), as shown at bottom of the previous page. The other subcase was for the remaining elements, which was determined using (19), as shown at bottom of the previous page.

VII. EXPERIMENT

A. DATA PREPARATION

We generated five sets of experimental data using the thermodynamics-based strategy developed by Gao, *et al.* [15] for testing spatial entropy. This strategy works with a user-supplied image (referred to as a seed image) to iteratively

simulate the mixing of ideal gases in a closed container. In an iteration, each pixel of the image is treated as a gas molecule, and half of the pixels switched their positions with randomly selected neighbors.

To generate the five sets of experimental data, we obtained and used five seed images. First, we obtained five remote sensing images from Google Earth, as shown in Fig. 8. The first two remote sensing images are of farmlands in Egypt. The third and fourth remote sensing images are Tsim Sha Tsui area and Kai Tak Airport of Hong Kong, respectively. The last remote sensing image was of the Qinghai Lake of China. All these remote sensing images had three bands, each of which had 300×300 pixels. Second, we converted these five remote sensing images to gray images. As shown in Fig. 9, these gray images were different in the histogram. Third, these gray images were used as seed images to simulate the iterative mixing of the ideal gases. The maximum number of iterations was set as 10,000. As a result, each of five sets of experimental data included a sequence of 10,000 gray images, which were increasingly disorderly in terms of thermodynamics (see Figs. 10 – 14).

B. EXPERIMENTAL VALIDATION

Validation involves testing the validity of the approach used in this study for computing DVIE; in other words,
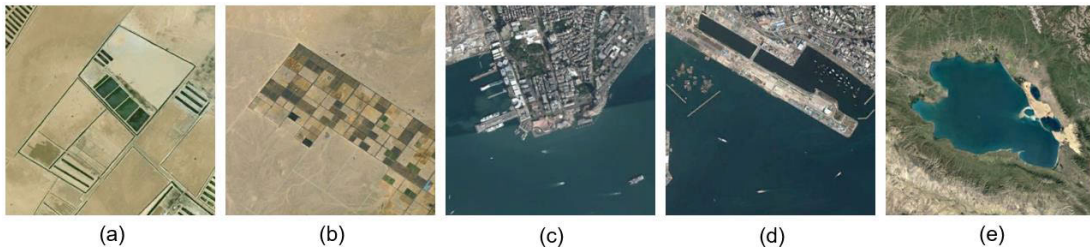
FIGURE 8. Remote sensing images for experiments: (a) Farmland 1 in Egypt, (b) Farmland 2 in Egypt, (c) Tsim Sha Tsui of Hong Kong, (d) Kai Tak Airport of Hong Kong, and (e) Qinghai Lake of China.
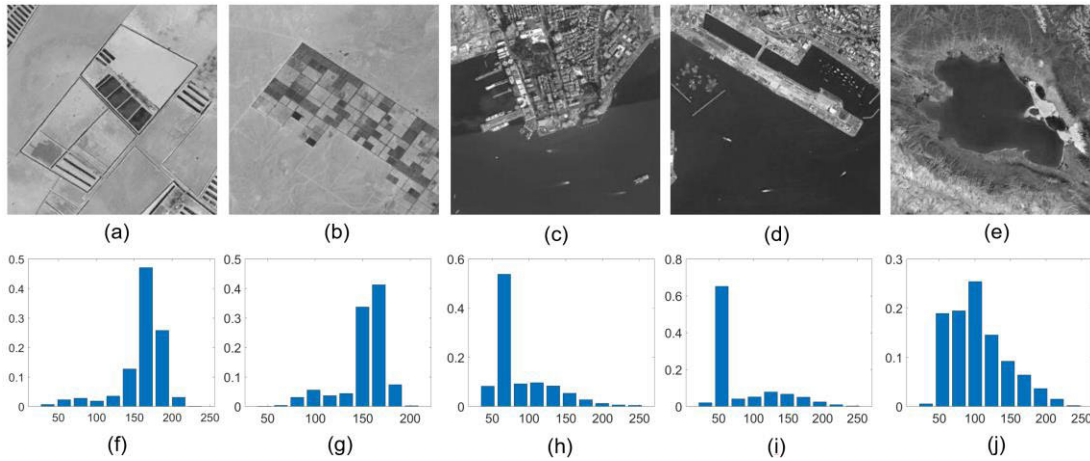
FIGURE 9. The gray images (a) – (e) correspond to the five remote sensing images in Fig. 7 and the histograms (f) – (j) of these gray images.
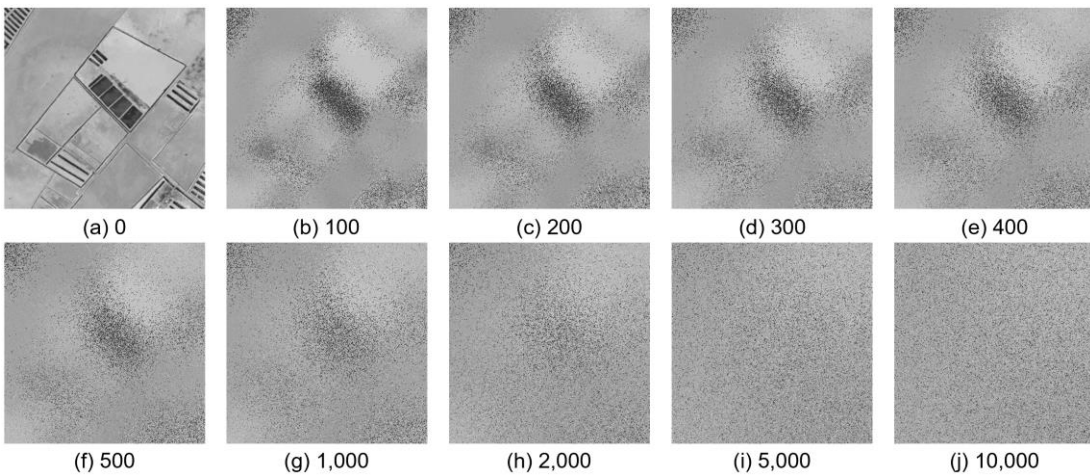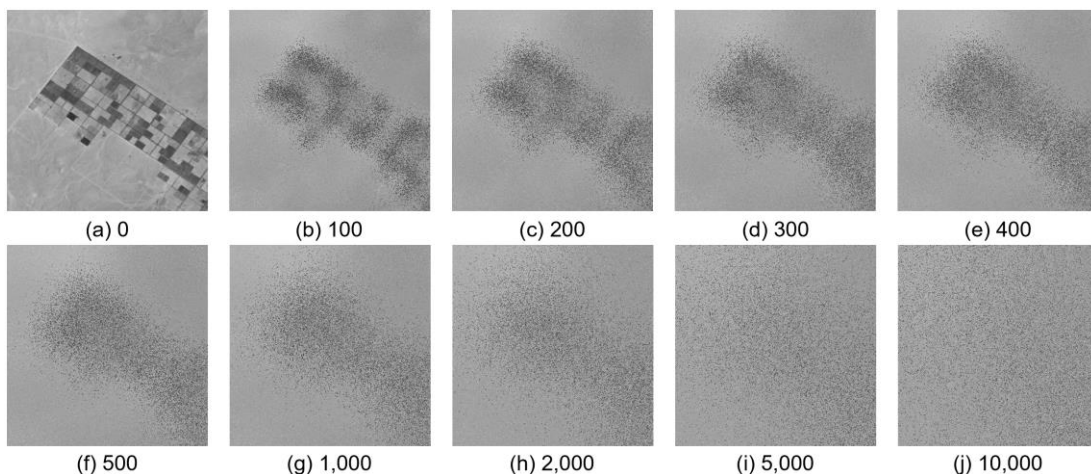

FIGURE 10. Images from the Experimental Dataset 1. These images are the outcomes of the iterations (a) 0, (b) 100, (c) 200, (d) 300, (e) 400, (f) 500, (g) 1,000, (h) 2,000, (i) 5,000, and (j) 10,000.

it involves testing whether the computational results of the proposed approach are the same as that of the original approach. If validity is not guaranteed, the efficiency results would be less meaningful. Hence, we performed validation before testing the efficiency.
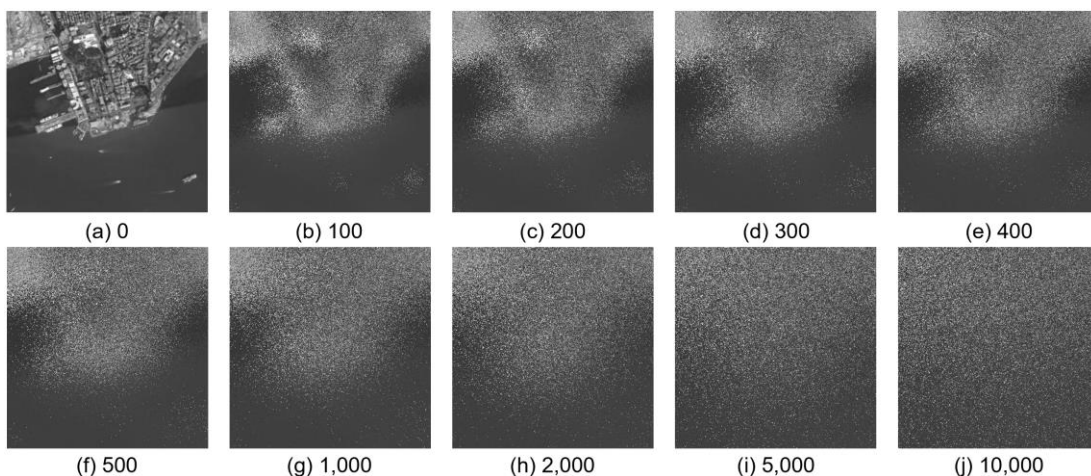
An ideal validation would compare the DVIEs of the experimental images (i.e., the 50,000 images of the five experimental datasets) computed using the proposed approach with those computed using the original approach. However, such

validation is not practical because it takes months to compute the DVIEs using the original approach. Hence, we test the validity of the proposed approach by using the following two experiments.

1) **Comparison with five samples of experimental images.** First, we generated a sample of 500 randomly selected images from each experimental dataset. Then, the DVIE of each sampled image was computed using both the proposed and original approaches. Finally,

FIGURE 11. Images from the Experimental Dataset 2. These images are the outcomes of the iterations (a) 0, (b) 100, (c) 200, (d) 300, (e) 400, (f) 500, (g) 1,000, (h) 2,000, (i) 5,000, and (j) 10,000.

FIGURE 12. Images from the Experimental Dataset 3. These images are the outcomes of the iterations (a) 0, (b) 100, (c) 200, (d) 300, (e) 400, (f) 500, (g) 1,000, (h) 2,000, (i) 5,000, and (j) 10,000.
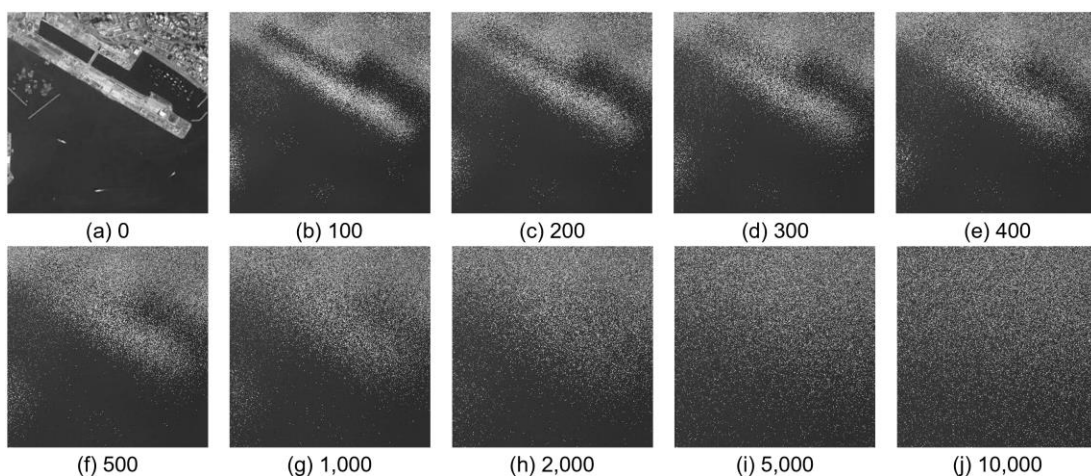
FIGURE 13. Images from the Experimental Dataset 4. These images are the outcomes of the iterations (a) 0, (b) 100, (c) 200, (d) 300, (e) 400, (f) 500, (g) 1,000, (h) 2,000, (i) 5,000, and (j) 10,000.

we compared the two DVIEs of each image to check whether they were the same.

2) **Trend derived from all experimental images**. The DVIEs of all the 50,000 images of the five experimental datasets were computed using the proposed approach. The changes in the DVIEs for varying iterations of mixing were graphically represented for each experimental dataset. An increasing trend should
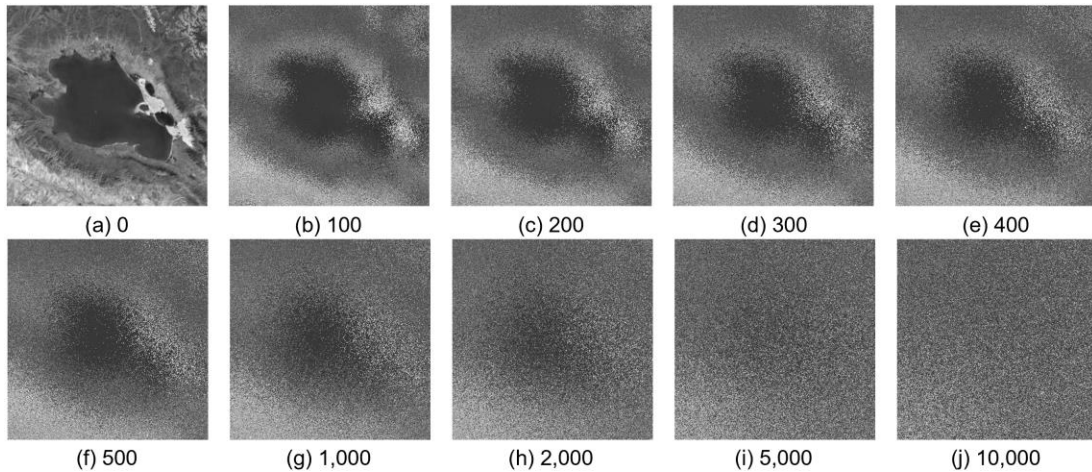
FIGURE 14. Images from the Experimental Dataset 5. These images are the outcomes of the iterations (a) 0, (b) 100, (c) 200, (d) 300, (e) 400, (f) 500, (g) 1,000, (h) 2,000, (i) 5,000, and (j) 10,000.
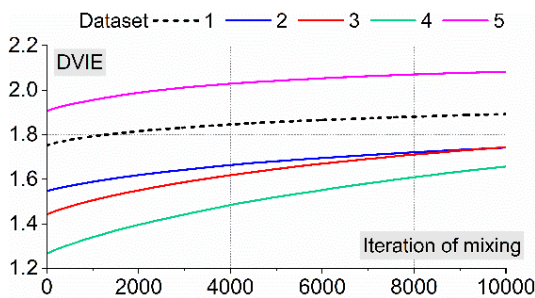

FIGURE 15. Changes in the DVIEs of the images of the experimental datasets along with the iterations of mixing.

FIGURE 16. Experimental hyperspectral remote sensing image (left) and its reference map (right) [43].

be seen unless the image pixels were fully mixed, as demonstrated by Gao, *et al.* [15].

The results of the first experiment demonstrated that the DVIEs computed using the two approaches for each sampled image had the same value. The results of the second experiment are shown in Fig. 15 (logarithms to base 10 were used). This figure shows that the DVIEs computed using the proposed approach for each of the five experimental datasets increased along with the iterative mixing. In summary, the results of the two experiments confirmed the validity of the proposed approach for computing DVIE.

C. EXPERIMENTAL EVALUATION

The evaluation aims to experimentally explore whether and how the proposed approach is more computationally efficient than the original approach. The measure of efficiency is the CPU time required by the proposed approach for computing the DVIE of an image from the experimental datasets. When the CPU time required by the proposed approach is less, this approach is more computationally efficient. The benchmark is the CPU time required by the original approach. It is not practical to compute the DVIEs of all the images using the original approach; therefore, we randomly selected 500 images from each exper-
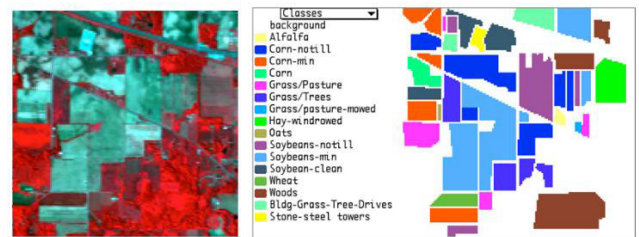
imental dataset. As a result, we had five subsets of the images.

The CPU time for dealing with an image (t) was determined as follows. First, the DVIEs of all images in a subset were computed, and the total CPU time (t_s) was recorded. The operating environment was an ordinary computer equipped with an Intel Core i5 processor, a 16-GB random access memory, and a Windows 10 system. Second, this computation was performed 10 times to derive the average total CPU time ($\bar{t}_s$). Third, the time t was calculated as $\bar{t}_s$ divided by 500. Fourth, this t was calculated using both the original and the proposed approaches for each of the five sub-datasets of images.

The results are shown in Table 1. Clearly, more than 99% time was saved by using the proposed approach to compute the DVIE of an image from any of the five sub-datasets. This fact demonstrates the outstanding efficiency of the proposed approach for dealing with varying images. Such efficiency makes it possible to compute the DVIEs of large datasets using ordinary computers. For example, Table 1 shows that it takes approximately 74 days to compute the DVIEs of the 10,000 images of the Experimental Dataset 5 using the original approach. By contrast, the proposed approach required only approximately 50 min.

We also evaluated the effect of the third strategy, namely the symmetry-based method for efficiently constructing LTAD. To this end, we reperformed the experiments in
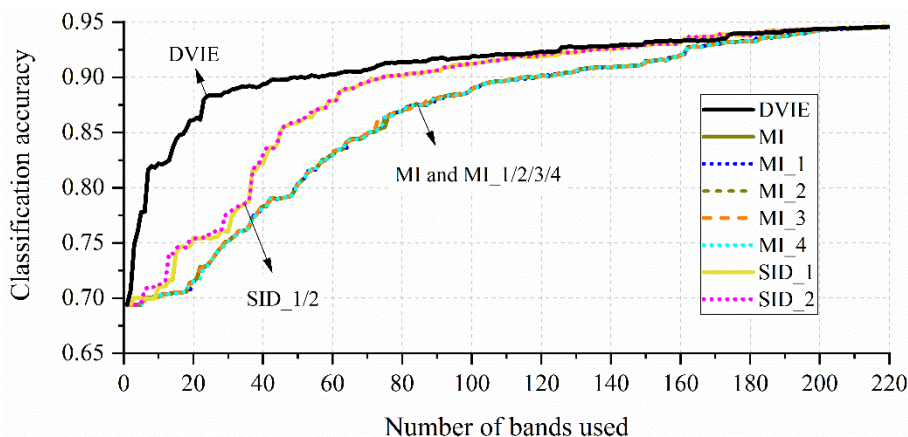
FIGURE 17. Change of classification accuracy with the number of bands selected based on different similarity metrics.

TABLE 1. CPU time required by the original and the proposed approaches to compute the DVIE of an image from sub-datasets 1 – 5.

Sub-dataset	Original (s)	Proposed (s)	Time saved
1	262.3	1.0	99.6%
2	256.9	1.1	99.6%
3	582.2	1.4	99.8%
4	305.5	1.4	99.6%
5	639.8	0.7	99.9%

Table 1 using the same sub-datasets but a different approach, which is an improved approach without the third strategy. The results for the five sub-datasets are 24.6, 24.1, 24.9, 25.1, and 24.5 seconds, respectively. A comparison between these results and that in Table 1 demonstrated that the third strategy indeed improved the computational efficiency.

VIII. APPLICATION

The proposed approach makes DVIE feasible for dealing with hyper- or ultra-spectral remote sensing images, where a single image may contain hundreds or even thousands of spectral bands [29]–[32]. In this section, we present an application of DVIE to the unsupervised band selection for hyperspectral image classification. In addition, we compare the performance of all this DVIE-based band selection with the band selections based on other entropies.

Band selection is one of the two methods for the dimensionality reduction of remote sensing images (the other method is feature extraction) [33]–[35]. The main idea behind band selection is to determine the "most distinctive and informative bands" [36]. In this process, there are two major aspects: an appropriate measurement criterion (i.e., a measure of band similarity) and an effective search strategy [33].

In this study, we proposed a DVIE-based method for unsupervised band selection for hyperspectral image classification. Specifically, we used the difference in DVIE between the two bands to measure similarities. The search strategy

was adopted from an existing study [21] for easy comparison of performances among different entropies and their variants. Also, for this reason, we used the same experimental hyperspectral remote sensing image (Fig. 16), performance analysis method, and benchmarks as the existing study [21]. (The performance analysis method included image classification for evaluating performance and a "band number versus classification accuracy" curve to present the results.) These benchmarks are seven IE-based measures, namely mutual information (MI), four variants of mutual information (MI_1, MI_2, MI_3, and MI_4) [37]–[40], and two variants of spectral information divergence (SID_1 and SID_2) [41], [42].

The experimental results are shown in Fig. 17. This figure clearly shows that the DVIE-based method for band selection significantly outperformed all the other methods based on IE. For example, the classification accuracy for the DVIE-based method was as high as 86% when only 20 bands were selected from the 220 original bands and used for image classification. By contrast, using the same number of bands, the highest classification accuracy using the IE-based methods was 75% (i.e., using SID_1 and SID_2). This fact suggests the vast potential of DVIE for image processing.

IX. DISCUSSION ON SIGNIFICANCE

The significance of the proposed method can be summarized as follows.

The method makes it possible to practically apply DVIE for processing remote sensing images. The previous study has shown that DVIE is the most effective measure for characterizing the spatial distribution of image pixels among all the IEs and their diverse variants. The proposed method can serve as a bridge between this most effective measure and multiple image processing applications, such as registration, classification, and change detection.

It can be used as the basis for further improvement of the computational efficiency of DVIE. As explained in Section III, the proposed method is designed to make DVIE

as universally applicable as possible. The proposed method is sequential; therefore, it can be used on both personal and high-performance computers. In the future, this method can be improved to perform in parallel with advanced computing devices such as clouds, GPUs, and TPUs.

It provides a possible way for addressing similar problems of practically determining a computation-intensive measure. An example of this problem is the determination of average linkage [44], which is a famous measure in geo-data clustering [45]. Average linkage is defined as the distance between two clusters of objects and is computed as the average distance between all the pairs of objects from the two clusters. This computation is intensive even for small clusters. However, the average linkage is comparable with the concept of extra-distance in computing DVIE; therefore, the idea for efficiently computing extra-distance can be borrowed.

X. CONCLUSIONS

In this study, we developed an efficient approach to compute the DVIE for processing remote sensing images. The approach was developed based on three strategies for reducing the amount of computation required by the original approach. The first strategy was to improve the equation of DVIE. The second strategy was a selective computation of the two key components of DVIE—intra-distance and extra-distance—based on the computational complexity. The distance with low computational complexity was computed by the original approach, whereas the distance with high computational complexity was directly derived using a proposed equation, which involved an LTAD. The third strategy was an efficient method to construct this lookup table. Experiments were performed to test both the validity (i.e., accuracy) and utility (i.e., efficiency) of the proposed approach for computing the DVIE of an image. We used five experimental datasets and 50,000 experimental images of 300×300 pixels. In addition, we described the applications of DVIE. The DVIE was used to select salient bands of hyperspectral remote sensing images. From the experimental results, we drew the following major conclusions about the approach proposed in this paper:

1) It is not only valid (i.e., its computational results are in line with those of the original approach), but it is also highly efficient because it saves more than 99% time as compared with the original approach.

2) It tends to be universally applicable because its algorithm is sequential in nature. It can be performed in either a personal or a high-performance computing environment.

3) Its performance can be further improved by employing a parallel computing strategy or more advanced computing devices such as clouds, GPU, and TPU.

4) It makes the application of DVIE possible, and the DVIE shows strong potential in applications for remote sensing processing. The DVIE-based band selection outperformed all the other band selection methods using IE-based measures.

The limitation of the proposed approach is that when applied for a multispectral remote sensing image, it constructs an LTAD for every band of the image. In fact, only one LTAD was necessary when dealing with multiple bands because all these bands had the same size.

Future research can cover the following three areas. First, more DVIE applications can be developed for processing remote sensing images, such as image registration, classification, coding, and change detection. Second, a systematic comparison, both theoretically and experimentally, can be made between the DVIE and the alternatives to IE for image processing, namely thermodynamic entropy (i.e., Boltzmann entropy) [46]–[48]. Third, we can determine whether there is a relationship between DVIE and another important measure of complexity, namely the fractal dimension [49]–[53]; determining the extent of this relationship will help better understand the nature of DVIE.

REFERENCES

[1] J. G. Liu and P. J. Mason, *Image Processing and GIS for Remote Sensing: Techniques and Applications*. Hoboken, NJ, USA: Wiley, 2016.

[2] J.-C. Pinoli, *Mathematical Foundations of Image Processing and Analysis*, no. 2. Hoboken, NJ, USA: Wiley, 2014.

[3] J. Liu, J. Huang, S. Liu, H. Li, Q. Zhou, and J. Liu, "Human visual system consistent quality assessment for remote sensing image fusion," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 79–90, Jul. 2015.

[4] X. Cao, X. Li, Z. Li, and L. Jiao, "Hyperspectral band selection with objective image quality assessment," *Int. J. Remote Sens.*, vol. 38, no. 12, pp. 3656–3668, Jun. 2017.

[5] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul./Oct. 1948.

[6] C. Claramunt, "Towards a spatio-temporal form of entropy," in *Proc. Int. Conf. Conceptual Modeling*. Florence, Italy: Springer, 2012, pp. 221–230.

[7] L. Hu, Z. He, J. Liu, and C. Zheng, "Method for measuring the information content of terrain from digital elevation models," *Entropy*, vol. 17, no. 12, pp. 7021–7051, 2015.

[8] J. Nowosad and T. F. Stepinski, "Information theory as a consistent framework for quantification and classification of landscape patterns," *Landscape Ecol.*, vol. 34, no. 9, pp. 2091–2101, Sep. 2019.

[9] W. Ji, J. Wu, M. Zhang, Z. Liu, G. Shi, and X. Xie, "Blind image quality assessment with joint entropy degradation," *IEEE Access*, vol. 7, pp. 30925–30936, 2019.

[10] J. X. Zhang, P. M. Atkinson, and M. F. Goodchild, *Scale in Spatial Information and Analysis*. Boca Raton, FL, USA: CRC Press, 2014.

[11] W. Sun, L. Zhang, B. Du, W. Li, and Y. Mark Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.

[12] H. Jia, K. Sun, W. Song, X. Peng, C. Lang, and Y. Li, "Multi-strategy emperor penguin optimizer for RGB histogram-based color satellite image segmentation using Masi entropy," *IEEE Access*, vol. 7, pp. 134448–134474, 2019.

[13] H. S. N. Alwerfali, M. Abd Elaziz, M. A. A. Al-Qaness, A. A. Abbasi, S. Lu, F. Liu, and L. Li, "A multilevel image thresholding based on hybrid salp swarm algorithm and fuzzy entropy," *IEEE Access*, vol. 7, pp. 181405–181422, 2019.

[14] S. A. Cushman, "Thermodynamics in landscape ecology: The importance of integrating measurement and modeling of landscape entropy," *Landscape Ecol.*, vol. 30, no. 1, pp. 7–10, Jan. 2015.

[15] P. Gao, Z. Li, and H. Zhang, "Thermodynamics-based evaluation of various improved Shannon entropies for configurational information of gray-level images," *Entropy*, vol. 20, no. 1, p. 19, 2018.

[16] S. Rakshit and A. Mishra, "Estimation of structural information content in images," in *Proc. Asian Conf. Comput. Vis.* Hyderabad, India: Springer, 2006, pp. 265–275.

[17] M. K. Quweider, "Spatial entropy-based cost function for gray and color Image segmentation with dynamic optimal partitioning," *Int. J. Comput. Sci. Netw. Secur.*, vol. 12, no. 4, pp. 64–75, 2012.

[18] C. Claramunt, "A spatial form of diversity," in *Spatial Information Theory*, vol. 3693, A. G. Cohn D. M. Mark, Eds. Berlin, Germany: Springer, 2005, pp. 218–231.

[19] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Urbana, IL, USA: Univ. Illinois Press, 1949.

[20] F. Q. Niu, D. H. Zhu, and C. X. Cheng, "Map information theories and adaptive visualization of electronic map in feature class-based zooming," *Proc. SPIE*, vol. 6420, Oct. 2006, Art. no. 64200L.

[21] P. Gao, J. Wang, H. Zhang, and Z. Li, "Boltzmann entropy-based unsupervised band selection for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 462–466, Mar. 2019.

[22] W. R. Tobler, "A computer movie simulating urban growth in the detroit region," *Econ. Geography*, vol. 46, pp. 234–240, Jun. 1970.

[23] D. Z. Sui, "Tobler's first law of geography: A big idea for a small world?" *Ann. Assoc. Amer. Geographers*, vol. 94, no. 2, pp. 269–277, Jun. 2004.

[24] W. Tobler, "On the first law of geography: A reply," *Ann. Assoc. Amer. Geographers*, vol. 94, no. 2, pp. 304–310, Jun. 2004.

[25] T. Warner, "Spatial autocorrelation analysis of hyperspectral imagery for feature selection," *Remote Sens. Environ.*, vol. 60, no. 1, pp. 58–70, Apr. 1997.

[26] X. Li and C. Claramunt, "A spatial entropy-based decision tree for classification of geographical information," *Trans. GIS*, vol. 10, no. 3, pp. 451–467, May 2006.

[27] D. G. Leibovici, C. Claramunt, D. Le Guyader, and D. Brosset, "Local and global spatio-temporal entropy indices based on distance-ratios and co-occurrences distributions," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 5, pp. 1061–1084, May 2014.

[28] M. Campbell-Kelly, M. Croarken, R. Flood, and E. Robson, *The History of Mathematical Tables: From Sumer to Spreadsheets*. New York, NY, USA: Oxford Univ. Press, 2003.

[29] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3185–3195, Jun. 2018.

[30] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[31] R. Yang and J. Kan, "An unsupervised hyperspectral band selection method based on shared nearest neighbor and correlation analysis," *IEEE Access*, vol. 7, pp. 185532–185542, 2019.

[32] Q. Chen, J. Sun, V. Palade, X. Shi, and L. Liu, "Hierarchical clustering based band selection algorithm for hyperspectral face recognition," *IEEE Access*, vol. 7, pp. 24333–24342, 2019.

[33] J. Feng, L. C. Jiao, X. Zhang, and T. Sun, "Hyperspectral band selection based on trivariate mutual information and clonal selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 4092–4105, Jul. 2014.

[34] W. Sun, M. Jiang, W. Li, and Y. Liu, "A symmetric sparse representation based band selection method for hyperspectral imagery classification," *Remote Sens.*, vol. 8, no. 3, p. 238, 2016.

[35] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.

[36] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

[37] Y. Horibe, "Entropy and correlation," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-15, no. 5, pp. 641–642, Sep./Oct. 1985.

[38] T. O. Kvalseth, "Entropy and correlation: Some comments," *IEEE Trans. Syst., Man, Cybern.*, vol. 17, no. 3, pp. 517–519, May 1987.

[39] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, no. 12, pp. 583–617, 2002.

[40] M. Hossny, S. Nahavandi, and D. Creighton, "Comments on 'Information measure for performance of image fusion,'" *Electron. Lett.*, vol. 44, no. 18, pp. 1066–1067, Aug. 2008.

[41] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[42] C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[43] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. (2015). *220 Band AVIRIS Hyperspectral Image Data Set: Jun. 12 1992 Indian Pine Test Site 3*. [Online]. Available: https://purr.purdue.edu/publications/1947/1

[44] X. Zhu and D. Guo, "Mapping large spatial flow data with hierarchical clustering," *Trans. GIS*, vol. 18, no. 3, pp. 421–435, Jun. 2014.

[45] Q. Liu, Z. Li, M. Deng, J. Tang, and X. Mei, "Modeling the effect of scale on clustering of spatial points," *Comput., Environ. Urban Syst.*, vol. 52, pp. 81–92, Jul. 2015.

[46] P. Gao and Z. Li, "Computation of the Boltzmann entropy of a landscape: A review and a generalization," *Landscape Ecol.*, vol. 34, no. 9, pp. 2183–2196, Sep. 2019.

[47] S. Cushman, "Calculation of configurational entropy in complex landscapes," *Entropy*, vol. 20, no. 4, p. 298, 2018.

[48] S. A. Cushman, "Calculating the configurational entropy of a landscape mosaic," *Landscape Ecol.*, vol. 31, no. 3, pp. 481–489, Mar. 2016.

[49] P. Gao, S. A. Cushman, G. Liu, S. Ye, S. Shen, and C. Cheng, "FracL: A tool for characterizing the fractality of landscape gradients from a new perspective," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 10, p. 466, 2019.

[50] T. Lan, Z. Li, and H. Zhang, "Urban allometric scaling beneath structural fractality of road networks," *Ann. Amer. Assoc. Geographers*, vol. 109, no. 3, pp. 943–957, May 2019.

[51] B. Mandelbrot, "How long is the coast of britain? Statistical self-similarity and fractional dimension," *Science*, vol. 156, no. 3775, pp. 636–638, May 1967.

[52] B. Jiang and J. Yin, "Ht-index for quantifying the fractal or scaling structure of geographic features," *Ann. Assoc. Amer. Geographers*, vol. 104, no. 3, pp. 530–540, May 2014.

[53] D. Ma and B. Jiang, "A smooth curve as a fractal under the third definition," *Cartographica: Int. J. Geographic Inf. Geovis.*, vol. 53, no. 3, pp. 203–210, Sep. 2018.

PEICHAO GAO received the B.S. degree from the Department of Surveying Engineering, Chengdu University of Technology, China, in 2012, the M.S. degree from the Department of Civil Engineering, Tsinghua University, China, in 2015, and the Ph.D. degree from the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, in 2018. He is currently an Assistant Professor at the Faculty of Geographical Science, Beijing Normal University.

HONG ZHANG received the B.S. degree in tourism management and the M.S. degree in human geography from Hubei University, China, in 2004 and 2007, respectively, and the Ph.D. degree in cartography and GIS from The Hong Kong Polytechnic University, in 2011. She is currently an Associate Professor at Southwest Jiaotong University, Chengdu, China. Her research interests include spatial complexity, urban geography, and multi-scale modeling.
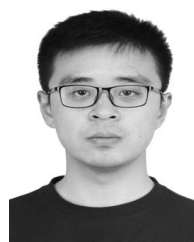
DUO JIA received the B.S. and M.S. degrees from the China University of Mining and Technology. He is currently pursuing the Ph.D. degree with the Faculty of Geographical Science, Beijing Normal University. His research interests include remote sensing image processing and geographical analysis.

CHANGQING SONG received the B.S. degree in geography from Northeast Normal University, China, in 1984, and the M.S. and Ph.D. degrees in geography from Peking University, China, in 1989 and 1992, respectively. He is currently a Full Professor of geography and the Dean of the Faculty of Geographical Science, Beijing Normal University, China. He serves as the Director of the Academic Working Committee of the Geographical Society of China, and an Associate Editor of *Chinese Geographical Science* and *Acta Geographica Sinica*.

CHANGXIU CHENG received the B.S. degree in computer science from the Beijing Institute of Technology, China, in 1997, and the Ph.D. degree in soil science from China Agricultural University, in 2001. She is currently a Full Professor and the Director of the Center for GeoData and Analysis, Faculty of Geographical Science, Beijing Normal University. She is the author of more than 90 research papers, three books (or chapters), and two national standards. She serves as a Council Member of the Eurasian System Science Research Association. She received the ''Yisheng Mao'' Science and Technology Award of Beijing, in 2017.

SHI SHEN received the B.S. degree in geographical information system from China Agricultural University, in 2011, and the Ph.D. degree in cartography and geographical information system from Beijing Normal University, in 2019. He is currently an Assistant Professor at the Faculty of Geographical Science, Beijing Normal University. His research interests include spatial complexity, land surface process analysis and modeling, and coupling human-earth systems.

• • •