Predicting the Lossless Compression Ratio of Remote Sensing Images With Configurational Entropy

Xinghua Cheng, *Student Member, IEEE*, and Zhilin Li

***Abstract*—Compression of remote sensing images is beneficial to both storage and transmission. For lossless compression, the upper and lower limits of compression ratio are defined by Shannon's source coding theorem with Shannon entropy as the metric, which measures the statistical information of a dataset. However, the calculation of the actual Shannon entropy of a large image is not an easy task, which limits the practicality of predicting the lossless compression ratio with Shannon entropy. On the other hand, most recently developed compression techniques take into consideration the configurational information of images to achieve a high compression ratio. This leads us to hypothesize that a metric capturing configurational information can be employed to build mathematical models for predicting compression ratios. To test this hypothesis, a two-step investigation was carried out, i.e., to find the most suitable metric through extensive experimental tests and to build a model upon this metric. A total of 1850 8-b images with 15 compression techniques were used to form the experimental dataset. First, 29 metrics were analyzed in terms of correlation magnitude, distinctiveness, and model contribution. As a result, the configurational entropy outperformed the rest. Second, six configurational entropy-based prediction models for predicting the compression ratio were established and tested. Results illustrated that these models work well. The PolyRatio model with 9.0 as a numerator, which was in a similar form to Shannon's theorem, performed best and was thus recommended. This article provides a new direction for building a theoretical prediction model with configurational entropy.**

***Index Terms*—Compression ratio, configurational information, empirical model for predicting compression ratios, image coding, Shannon's source coding theorem.**

I. Introduction

HIGH spatial resolution remote sensing images are widely available, leading to the dramatic increase in the data volume size. To reduce the burden on image storage and transmission, data compression is one of the efficient solutions. As a result, lots of compression techniques have been developed over the past seven decades (see [1]–[15]). It is well known that compression can be either lossless or lossy. Lossy compression chooses to discard some information for achieving a very high compression ratio. On the other hand, lossless compression means that the original image can be completely recovered after decompression, which is very essential for some tasks (e.g., geological surveying, image quality assessment). Therefore, it is very desired and hence selected as the research topic of this article.

Fig. 1. Three images with the same Shannon entropy but different configurations.

The prediction of lossless compression ratio can be traced back to Shannon's paper entitled "A Mathematical Theory of Communication" [1]. In that paper, Shannon's source coding theorem established the upper and lower limits to lossless compression through coding and demonstrated the operational definition of Shannon entropy. Some techniques (e.g., Huffman encoding [2] and arithmetic encoding [4]) that encode discrete data into variable-length codewords based on the statistical information of images (i.e., the proportion of pixels at different gray levels) cannot perform better than Shannon's theorem. Since the 1980s, such techniques are usually combined with other techniques that utilize the configurational information (spatial structures) of images to implement compression. Indeed, some algorithms, e.g., LZMA [3], [8], CALIC [5], JPEG-LS [6], [7], FLIF [9], and HEIC [10], [11], have achieved a compression ratio much higher (see Section II for more details) than the upper limits defined by Shannon's theorem [16]–[18]. This can be easily reasoned that Shannon entropy, which is widely used in the image processing field [19]–[21], captures only statistical information. For instance, the three images shown in Fig. 1 are the same in terms of Shannon entropy although the configurations are quite different [22]–[25].

In fact, different order Shannon entropies with length n, which measures the correlation among pixels, can be used to improve the prediction performance of the compression ratio. However, the nth-order Shannon entropy when n approaches

Manuscript received August 3, 2021; revised September 29, 2021; accepted October 21, 2021. Date of publication October 29, 2021; date of current version December 3, 2021. This work was supported in part by the Research Grant Council of HKSAR under Grant 15221918, and in part by the National Natural Science Foundation of China under Grant 41930104. (*Corresponding author: Zhilin Li*.)

Xinghua Cheng is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China (e-mail: cxh9791156936@gmail.com).

Zhilin Li is with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong, China, and also with the State-Province Joint Engineering Laboratory of Spatial Information Technology for High-speed Railway Safety, Southwest Jiaotong University, Chengdu 611576, China (e-mail: dean.ge@swjt.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3123650

infinity (referred to as the actual Shannon entropy) is not easy to calculate [26] in practice and cannot capture the configurational information of images from the theoretical perspective, which will be introduced in the following sections.

From the literature, it can be found that some alternative techniques for predicting the lossless compression ratio based on the entropy of the "residual data" generated at preprocessing stage have been developed, such as the context tree weighting algorithm [27], prediction by partial matching [28] and the probabilistic suffix tree prediction algorithm [29]. However, such residual data is not available to (compression software) users in practice.

This means that there is an urgent need to develop new mathematical models for the reliable prediction of the lossless compression ratio for any given image in hand. It is hypothesized here that such a model is possible when a metric capable of characterizing the configurational information of the image is employed. Indeed, some preliminary experiments [16]–[18] have supported such a hypothesis. Therefore, this article is dedicated to establishing new models for this purpose.

Apart from the introduction, the remainder of this article is structured as follows. First, Shannon's source coding theorem concerning the prediction of the lossless image compression ratio is introduced and described. Then, a novel solution for building new models is outlined in general, i.e., the detection of the best metric candidate that captures the configurational information of images and the construction of mathematical prediction models. Second, the solution to identifying the best metric candidate is explained and implemented in detail; in particular, configurational entropy (Boltzmann entropy) is identified as the best metric candidate. Third, empirical models with configurational entropy are constructed and tested. Finally, a discussion and some concluding remarks are presented.

II. Shannon's Source Coding Theorem in Regard to the Prediction of the Image Compression Ratio: Analysis and a Novel Solution

A. Basic Introduction to Shannon's Source Coding Theorem

Compression can be thought of as a kind of coding and should follow the source coding theorem first proposed by Shannon in 1948 [1]. This theorem theoretically defines the upper and lower limits of the average codeword length (referred to as L_n) for a zero-memory source while suffering no information loss. This theorem was built with *n*th order block codes and can be written as follows:

$$H_n \le L_n < H_n + \tfrac{1}{n} \tag{1}$$

where H_n denotes the *n*th-order Shannon entropy of the zero-memory source and is calculated as follows:

$$H_n = \tfrac{1}{n} \sum_{i=1}^{n} P(X = x_i) \log P(X = x_i) \tag{2}$$

where X is a random variable that represents the symbol blocks $\{x_1, x_2, x_3, x_4 \cdots x_n\}$ and $P(X = x_i)$ denotes the occurrence probability of the ith symbol block. Note that the higher the n is, the lower the H_n. In Equation (1), L_n is the sum of the product of $l(x_i)$, the codeword length of x_i, and $P(X = x_i)$ in the nth order coding. It can be computed as follows:

$$L_n = \tfrac{1}{n} \sum_{i=1}^{n} P(X = x_i) \times l(X = x_i) \tag{3}$$

Note that Shannon's theorem holds when n tends to infinity. That is, L_n approaches the upper and lower limits shown in Equation (1). Fig. 2 shows an example in which each symbol is independent of calculating Shannon entropy (i.e., $n = 1$).

TABLE I
First-Order Block Coding for the Image Shown in Fig. 3

Gray level block	Count	Occurrence probability	Codeword	Codeword length
128	15	$15/36$	1	1
120	6	$6/36$	010	3
255	12	$12/36$	01	2
64	3	$3/36$	011	3

TABLE II
Second-Order Block Coding for the Image Shown in Fig. 3

Gray level block	Count	Occurrence probability	Codeword	Codeword length
(128, 128)	8	$8/35$	10	2
(128, 120)	3	$3/35$	010	3
(120, 255)	6	$6/35$	111	3
(255, 128)	6	$6/35$	00	2
(128, 255)	3	$3/35$	1100	4
(255, 64)	3	$3/35$	1101	4
(64, 255)	3	$3/35$	0110	4
(255, 120)	3	$3/35$	0111	4

B. Shannon's Source Coding Theorem for Predicting the Compression Ratio of Images: Analysis with Theoretical and Practical Perspectives

Shannon's source coding theorem was originally used to define the theoretical compression limits of a zero-memory source without any information loss. Regarding image compression, when an image is under the assumption that it was derived from an image source sending statistically independent pixels, according to a probability law [26], Shannon's theorem can be applied to determine the theoretical upper and lower limits of the compression ratio. This section analyzes such a case from both theoretical and practical perspectives. For the image shown in Fig. 3, regarding first-order block coding (see Table I where Huffman encoding [5] is used), based on Equations (2) and (3), H_1 and L_1 are 1.78 and 1.83 bits per pixel, respectively.

The fundamental premise of using Shannon's theorem is that the process of generating a source will be modeled as a probabilistic model. Then, by assuming that the image is connected row by row from beginning to end, the image can be mapped to a linear sequence. Subsequently, second-order block coding can be performed, which is given in Table II. H_2 and L_2 are 1.44 and 1.47 bits per pixel, respectively. Clearly, L_2 is smaller than L_1, and H_2 is smaller than H_1.

In practical compression applications, the compression ratio (C_R) [26] is one of the primary concerns for developers and

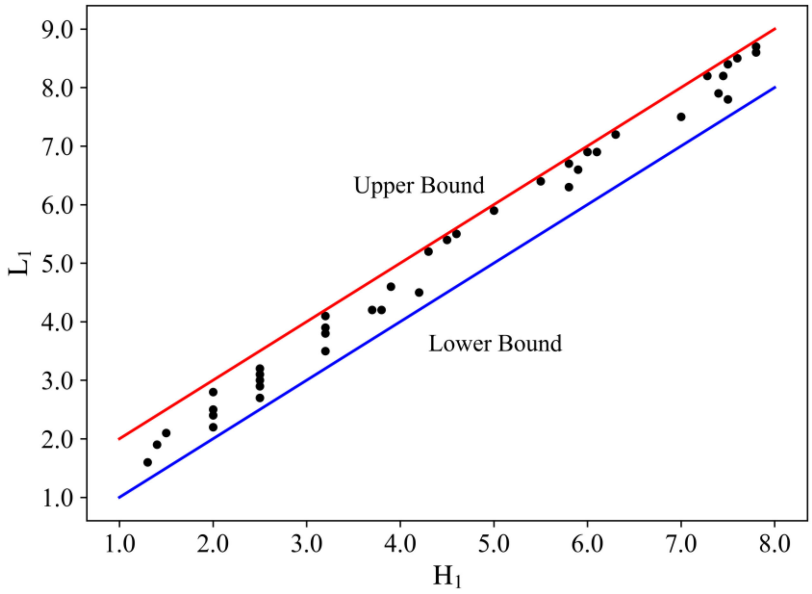

Fig. 2. Example of Shannon's source coding theorem under first-order block coding.

128	**128**	**128**	**120**	**255**	**128**
128	**128**	**128**	**120**	**255**	**128**
128	**128**	**128**	**120**	**255**	**128**
255	**64**	**255**	**120**	**255**	**128**
255	**64**	**255**	**120**	**255**	**128**
255	**64**	**255**	**120**	**255**	**128**

Fig. 3. Image of size 6×6.

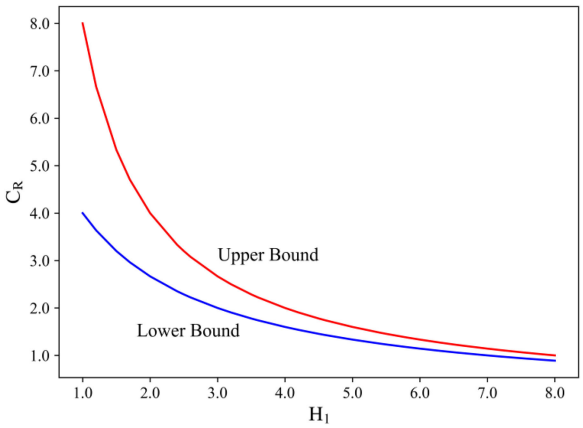

Fig. 4. Upper and lower bounds of the compression ratio defined by Shannon's theorem with H_1.

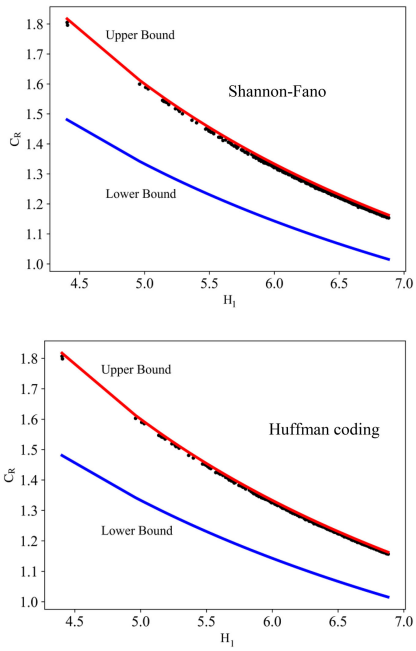

Fig. 5. Plots of Shannon's theorem against the compression ratio by two techniques that consider statistical information only.

users. The upper and lower limits defined by Shannon's theorem for 8-b images with first-order entropy can be described in Fig. 4, which provides perfect guidelines for traditional coding techniques (see Fig. 5).

However, for modern compression techniques which take into consideration configurational information, the C_R is much higher (see Fig. 6) than predicted by Fig. 5. Even with higher order Shannon entropy is employed (see Fig. 6), the modern compression algorithms still perform better than predicted by Fig. 5.

The above examples tell us that high-order Shannon entropy can improve the prediction. Then one may wonder whether very high-order Shannon entropy or actual Shannon entropy should be good enough for the prediction of compression ratios. However, the following problems cannot be solved.

1) Very High-Order Shannon Entropies are Computationally Expensive for Ordinary Users: When calculating the 4-th order Shannon entropy of an 8-b image, relative occurrence probabilities of $(2^8)^4$ pixel pairs should be computed. Unfortunately,

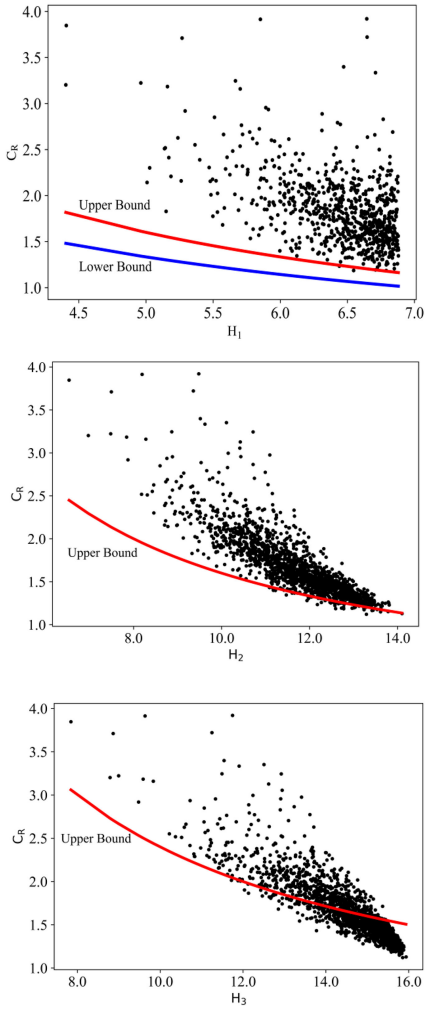

Fig. 6. Plots of Shannon's theorem with Shannon entropy of different orders against the compression ratio derived from JPEG-LS.

when sixth-order Shannon entropy is considered, we shall first compute the occurrence probabilities of $(2^8)^6$ possible tuples [26], indicating a high memory cost that is unendurable. The higher order the Shannon entropy, the higher computation cost (i.e., computation time and memory) we need to take.

2) Calculation of the Actual Shannon Entropy is Impossible and Unreliable: Originated from the field of telecommunication, the actual Shannon entropy is available for a memoryless source. Unfortunately, images are not a kind of memoryless source.

C. A Novel Solution for Building Models for Predicting the Compression Ratio

By analyzing Shannon's theorem for predicting the lossless compression ratio, one might imagine that image metrics, which are structural measures, could be used to build new prediction models. Therefore, we carry out investigations to build the relationship between the compression ratio and metrics measuring the configurations of images. As shown in Fig. 7, the objective is to select the best metric, then build mathematical models for the upper and lower bounds of the compression ratio.

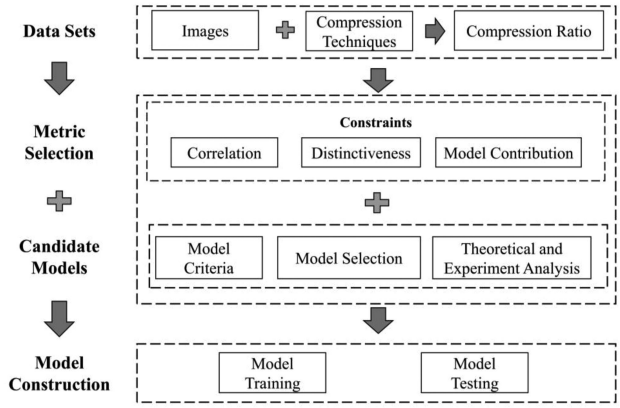

Fig. 7. Flowchart of building models for predicting the compression ratio.

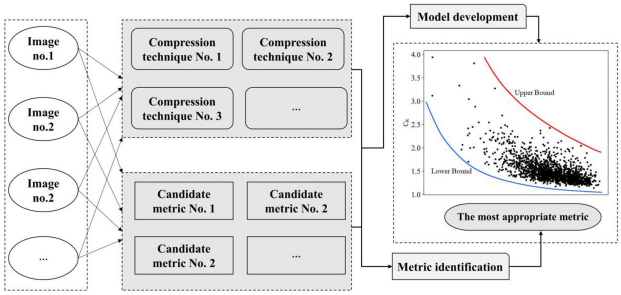

Fig. 8. A schematic diagram of building prediction models.

To realize such an objective, the following solution is proposed.

1) Typical techniques are used to compress a large number of images with different complexities to form experimental datasets.
2) The most appropriate image metric for measuring configurational information is identified to build models.
3) Mathematical models of the upper and lower bounds of the lossless compression ratio are built with the most appropriate metric.

The general process of implementing such a solution is shown as a diagram in Fig. 8, where three main stages can be distinguished, i.e., the formation of experimental datasets, metric selection, and model construction. For the first stage, from the perspective of statistical modeling, the larger the volume of the training dataset is, the more reliable and general the constructed models will be. Hence, a large number of images with different complexities and many typical compression techniques will be employed. Notably, data in which the compression ratio is larger than Shannon's upper bound are reserved. The reason is that for any given image, compression techniques cannot always achieve a compression ratio larger than the upper bound given by Shannon's theorem. After the construction of the experimental datasets, we can build new mathematical models for predicting the compression ratio, which will be introduced in Sections III and IV. Such mathematical models can be obtained by theoretical analysis (i.e., theoretical models) or be trained by experience (i.e., empirical models). In this article, only empirical models are considered.

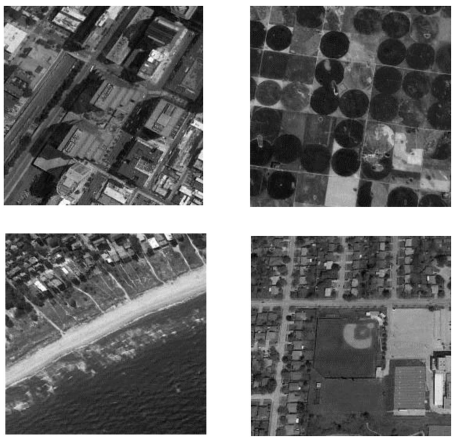

Fig. 9. Four remote sensing images (red band) from "NWPU-RSCI45" dataset [31].

III. Identification of the Most Appropriate Metric for Model Development

A. Metric Identification: Basic Ideas and Procedures for Implementation

As mentioned earlier, we need to identify the most appropriate metric for building models. Many metrics for quantifying the configurational information of images have already been developed. Hence, we should identify the most appropriate metric considering the three constraints shown in Fig. 7. To realize such an objective, a multilayered solution is proposed based on these three constraints from multivariate mathematical statistics, as follows.

1) *Correlation*: In statistical modeling, a correlation is any relationship between two variables, indicating the degree to which they change in coordination with one other. Therefore, metrics holding a high statistical correlation with the compression ratio are first selected.
2) *Distinctiveness*: A correlation may exist between metrics. In this respect, the selected variables should be highly distinctive.
3) *Model Contribution*: Different independent variables have different contributions to predicting the value of a dependent variable. Thus, the independent variable making the highest contribution should be selected.

More detailed descriptions about the employment of these constraints to identify the most appropriate metric for measuring configurational information will be introduced in Section III-C–E.

The image dataset called "NWPU-RSCI45" (some examples can be seen in Fig. 9) [31], where 1650 images (first 150 images for each of the first 11 classes) are used with nine compression methods given in Table III.

B. Selection of Candidate Metrics for Measuring Configurational Information

Many image metrics that measure configurational information have been developed for image processing. From a theoretical standpoint, these metrics are classified into four categories, as given in Table IV. The first is based on the gray level matrix of an image. The second is based on the results of the operations on the gray level matrix (i.e., the gray level co-occurrence matrix, neighboring gray level dependence matrix, neighboring gray tone difference matrix, gray level run length matrix, and gray level size zone matrix). The third category is based on the Sobel gradient, which detects edges to measure the configurational information of images. The final category involves entropy. Since different improved Shannon entropies have been examined and none of them can measure the configurational information of images [19], only metrics derived from configurational entropy are investigated here. Table IV gives the 29 metrics used in this article. To distinguish the metrics that belong to the second category, their abbreviations are combined with operation directions

TABLE III
Compression Techniques Used to Compress Images

Compression techniques	Source
LZW Encoding	[3], [8]
Deflate Encoding	[30]
LZMA Encoding	[32]
Bzip2	[33],[34]
Zip	[35], [36]
JPEG-LS	[6], [7]
1D Linear Prediction Encoding & Huffman Encoding	[37]
Lifting Wavelet Transformation & Huffman Encoding	[38]
Bit-plane Encoding & Run Length Encoding & Huffman Encoding	[39]

TABLE IV
Metrics Capturing Configurational Information

Category	Source	Derived Metrics (Abbreviation)
Configurational Entropy	[40-42]	*Configurational Entropy* by Resampling (S_R) *Configurational Entropy* by Aggregation (S_A)
Gray Level Matrix (*GLM*)	[43], [44]	*Gray Level Uniformity* (*GLU*) *Spatial Frequency* (*SF*)
Gray Level Cooccurrence Matrix (*GLCM*)	[45-49]	*Contrast* (*CON*) *Inverse Moment* (*IM*) *Sum of Squares* (*SUS*) *Correlation* (*COR*) *Homogeneity 1* (*HOM1*) *Homogeneity 2* (*HOM2*) *Dissimilarity* (*DIS*)
Neighboring Gray Level Dependence Matrix (*NGLDM*)	[50]	*Number Nonuniformity* (*NN*) *Second Moment* (*SM*) *Shannon Entropy* (*SE*)
Neighboring Gray Tone Difference Matrix (*NGTDM*)	[51]	*Coarseness* (*COA*) *Contrast* (*CON*) *Busyness* (*BUS*)
Gray Level Run Length Matrix (*GLRLM*)	[52]	*Gray Level Nonuniformity* (*GLN*) *Run Length Nonuniformity* (*RLN*) *Run Percentage* (*RP*) *Run Shannon Entropy* (*RSE*)
Gray Level Size Zone Matrix (*GLSZM*)	[53]	*Gray Level Nonuniformity* (*GLN*) *Size Zone Nonuniformity* (*SZN*) *Gray Level Variance* (*GLV*) *Zone Variance* (*ZV*) *Zone Shannon Entropy* (*ZSE*)
Sobel Gradient	[54], [55]	Mean of an Image by Sobel Kernels (SI_{mean}) Root Mean Square by Sobel Kernels (SI_{rms}) Standard Deviation by Sobel Kernels (SI_{stdev})

(i.e., right (R), right-down (RD), down (D), left-down (LD), up (U)). For example, the dissimilarity metric based on the gray level co-occurrence matrix using operation in the right direction is referred to as *GLCM-DIS-R*. Other metrics will be used with their abbreviations.

C. Identification of Image Metrics With a Strong Correlation Magnitude

The correlation between the independent variable (i.e., the compression ratio) and the dependent variable (image metrics measuring the configurational information) is an essential reference used to identify the most appropriate metric for building models. To quantitatively measure the first constraint (i.e., the strength of the relationship between two variables), Spearman's rank correlation coefficient [56], [57] (referred to as R_S) is employed. In this article, metrics having strong relationships with the compression ratio are first reserved. As a rule of thumb, the strength of the relationship between the variables is considered strong when the absolute R_s value at the confidence interval is larger than 0.70 [59]. The R_S values of 29 metrics are illustrated in Table V, where all p values [60] are smaller than 0.01, indicating that these metrics are statistically significantly correlated with compression ratio. However, only 18 metrics are reserved for the next identification process. The scatter plots of those metrics against the compression ratio are shown in Fig. 10.

D. Identification of Image Metrics With High Distinctiveness

In Section III-C, 18 metrics having strong correlations with compression ratio have been identified. However, many of them are correlated with each other. Thus, we need to identify metrics with the second constraint (i.e., high distinctiveness). To this end, two multivariate statistical analysis approaches are used, namely, factor analysis (FA) [61] and hierarchical cluster analysis (HCA).

1) Factor Analysis: FA is a statistical method for describing the variance among variables and reducing them into many factors [62]. To assess the appropriateness of conducting FA on the metrics, this article conducts two tests, namely, the Kaiser–Meyer–Olkin (KMO) measurement [63] and Bartlett's test of sphericity [64]. In this article, when the KMO measure of sampling adequacy is larger than 0.5 and Bartlett's test of sphericity significance is smaller than 0.05 [65], the data are adequate for conducting FA. Table VI gives the test results, which indicate that FA is appropriate.

Principal component analysis is used in the FA to form components. Fig. 11 shows the scatter plots of components, where the eigenvalue represents the measure of significance. Eigenvalues of 1.0 or higher are considered significant [66].

Two principal components with eigenvalues larger than 1 are significant. The loading matrix for all metrics is given in Table VII, where the loading values are the correlation coefficients for the metric and components. The rule is that the higher the loading value for the component is, the more important it is for this component. Then, we classify 18 metrics into two categories, as given in Table VIII.

TABLE V
METRICS AND THEIR CORRESPONDING R_s VALUES

Metric	R_s	Reserved
S_R	-0.78	Yes
S_A	-0.78	Yes
GLM-GLU	0.01	No
GLM-SF	-0.67	No
GLCM-CON-R	-0.60	No
GLCM-CON-RD	-0.61	No
GLCM-CON-U	-0.61	No
GLCM-CON-LD	-0.61	No
GLCM-IM-R	0.68	No
GLCM-IM-RD	0.68	No
GLCM-IM-U	0.69	No
GLCM-IM-LD	0.68	No
GLCM-SUS-R	-0.21	No
GLCM-SUS- RD	-0.21	No
GLCM-SUS-U	-0.21	No
GLCM-SUS-LD	-0.21	No
GLCM-COR-R	-0.46	No
GLCM-COR-RD	-0.46	No
GLCM-COR-U	-0.46	No
GLCM-COR-LD	-0.46	No
GLCM-HOM1-R	0.68	No
GLCM-HOM1-RD	0.67	No
GLCM-HOM1-U	0.68	No
GLCM-HOM1-LD	0.67	No
GLCM-HOM2-R	-0.74	Yes
GLCM-HOM2-RD	-0.74	Yes
GLCM-HOM2-U	-0.75	Yes
GLCM-HOM2-LD	-0.74	Yes
GLCM-DIS-R	-0.74	Yes
GLCM-DIS- RD	-0.75	Yes
GLCM-DIS-U	-0.75	Yes
GLCM-DIS-LD	-0.74	Yes
NGLDM-NN	0.52	No
NGLDM-SM	0.61	No
NGLDM-SE	-0.60	No
NGTDM-COA	0.49	No
NGTDM-CON	-0.55	No
NGTDM-BUS	-0.41	No
GLRLM-GLN-R	0.48	No
GLRLM-GLN-RD	0.50	No
GLRLM-GLN-U	0.48	No
GLRLM-GLN-LD	0.50	No
GLRLM-RLN-R	-0.75	Yes
GLRLM-RLN-RD	-0.75	Yes
GLRLM-RLN-U	-0.76	Yes
GLRLM-RLN-LD	-0.75	Yes
GLRLM-RSE-R	-0.15	No
GLRLM-RSE-RD	-0.25	No
GLRLM-RSE-U	-0.15	No
GLRLM-RSE-LD	-0.25	No
GLRLM-RP-R	-0.73	Yes
GLRLM-RP-RD	-0.73	Yes
GLRLM-RP-U	-0.73	Yes
GLRLM-RP-LD	-0.73	Yes
GLSZM-GLN	-0.35	No
GLSZM-SZN	-0.10	No
GLSZM-GLV	0.31	No
GLSZM-ZV	0.37	No
SI_{mean}	0.06	No
SI_{rms}	0.05	No
SI_{stdev}	-0.16	No

TABLE VI
RESULTS OF THE KMO AND BARTLETT TESTS

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.90
Bartlett's Test of Sphericity	Approximate Chi-Square	970987.37
	Degrees of Freedom	13205
	Significance	0.00

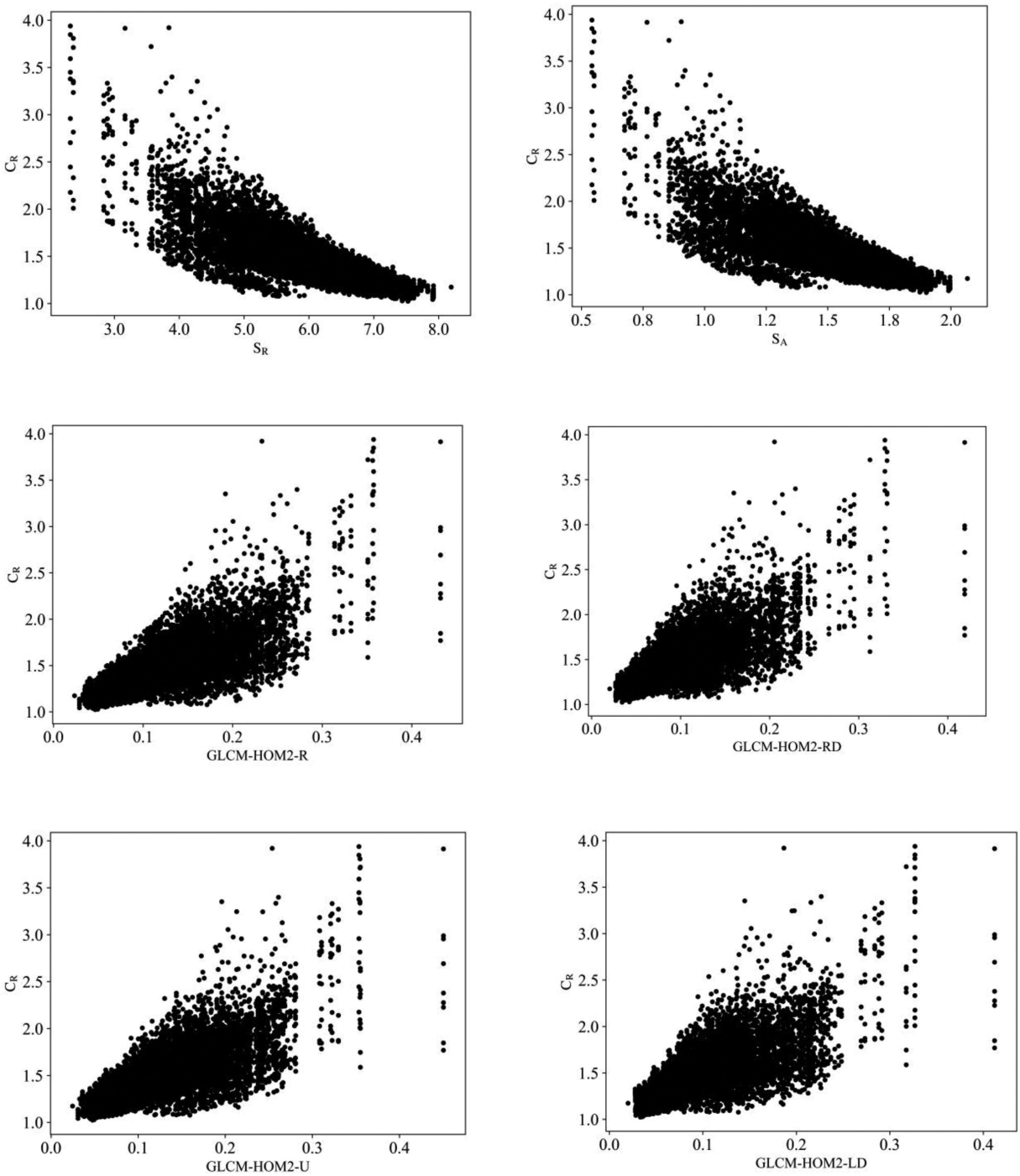

Fig. 10. Scatter plots of the metrics against C_R.

To further identify the metrics with high distinctiveness, HCA on the metrics is conducted. HCA is a method that reveals the underlying structures of objects by clustering or dividing objects one by one in a hierarchical process [67]. In this article, the squared Euclidean distance is used as the criterion for evaluating the similarity between two objects; the average linkage method is used as the linkage criterion. The HCA results are graphically shown in Fig. 12. Clearly, S_R and S_A are the most distinctive metrics in category no. 1. However, the metrics in category no. 2 are difficult to distinguish. Thus, S_R, *GLCM-DIS-R*, *GLCM-DIS-RD*, *GLCM-DIS-U*, and *GLCM-DIS-LD* are reserved and will be further investigated for their appropriateness for building models predicting the compression ratio.

E. Final Identification of Configurational Entropy (S_R) as the Recommended Metric With Model Contribution

In Section III-D, six metrics have been selected out of the 18 metrics selected in Section III-C. The final stage is to identify the most appropriate metric following the last constraint (i.e., the contribution to the construction of prediction models). To this end, the least absolute shrinkage and selection operator

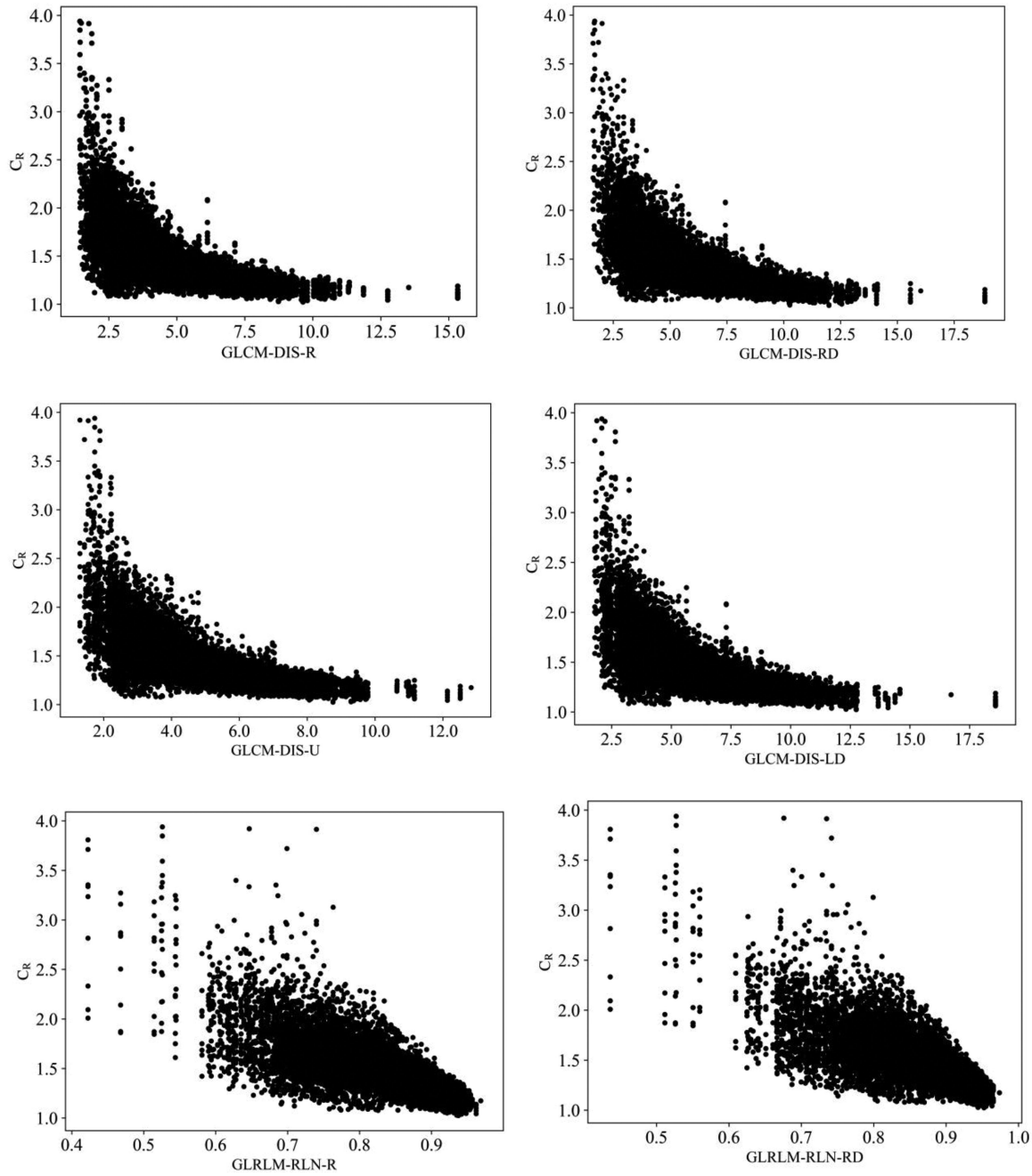

Fig. 10. (*Continued.*)

(LASSO) is used. The LASSO sets some coefficients of variables to 0, thus generating an interpretable model, and it is believed to select certain important variables from irrelevant features [68]–[70]. Imagine that we have a matrix X of features and an n-dimensional response γ. A linear model describing the relationship between y and X is as follows:

$$\gamma = X\beta + z \tag{4}$$

where z is a noise term. The LASSO is the family of solutions to the following problem:

$$\hat{\beta}\ (\lambda) = \operatorname{argmin}\frac{1}{2}\|\boldsymbol{y} - xb\|^2 + \lambda\|b\|_1 \tag{5}$$

where $b \in \mathbb{R}^p$, p is the number of variables and λ is the regularizing parameter. $\hat{\beta}(\lambda)$ is referred to as the LASSO path, which reflects the importance of a variable in building models to predict the response variable value.

As a result, the LASSO path is generated for the five metrics reserved in the aforementioned Section III-D. The results are illustrated in Fig. 13, where each line represents a metric. Clearly, only the coefficient values of S_R have been nonzero numbers in the whole regularization process, which means that S_R is more important than the remaining metrics. Therefore, S_R is finally considered to be the most appropriate metric for building prediction models.

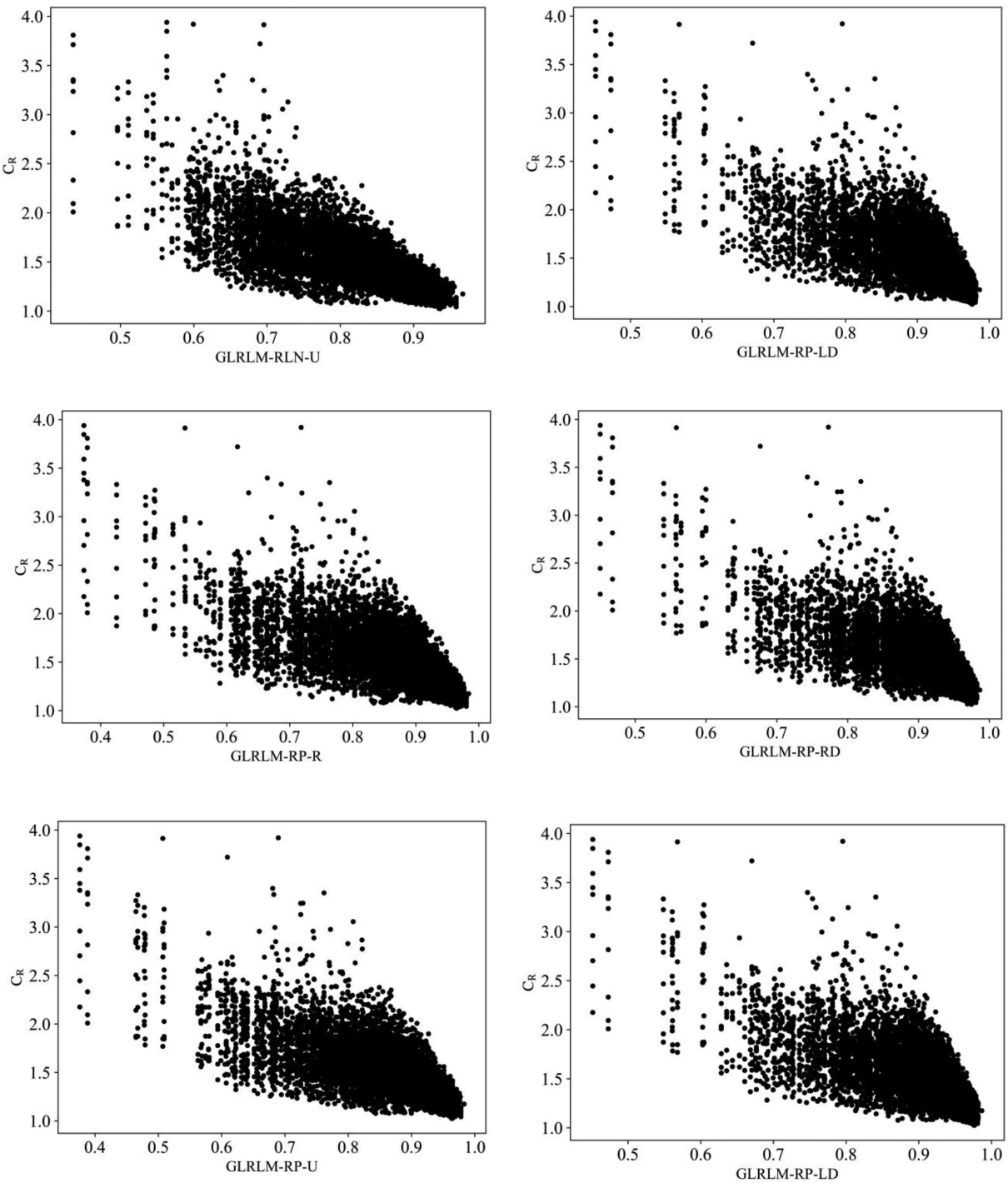

Fig. 10. (*Continued.*)

F. Verification of Configurational Entropy (S_R) as the Best Choice

In order to avoid running the risk of introducing significant statistic bias. Another six compression techniques (i.e., two-dimensional (2-D) linear prediction encoding [71], FLIF [9], HEIC [10], [11], TIFF [72], PNG [73], and JPEG 2000 [74]) were employed apart from-9 compression methods already used. Another three image datasets given in Table IX have been utilized apart from the "NWPU-RESISC45" image datasets. With 15 compression methods and 13 combinations (e.g., "NWPU-RESISC45" and "UCID") of image datasets, we used the proposed method to search for the best metric. Fortunately, the experimental results showed that the configurational entropy was always the best one.

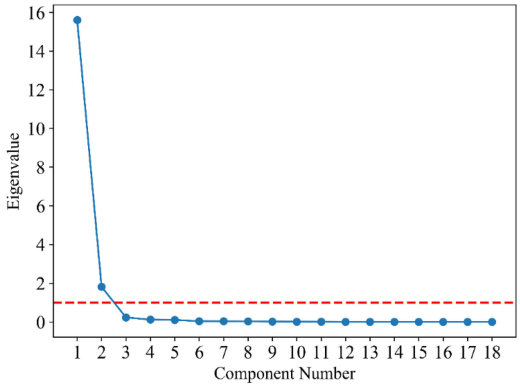

Fig. 11. Scatter plot of the components.

TABLE VII
LOADING MATRIX FOR ORTHOGONALLY ROTATED COMPONENTS

	Component	
Metric	1	2
S_R	0.77	0.63
S_A	0.78	0.62
GLCM-DIS-R	0.36	0.91
GLCM-DIS-RD	0.36	0.92
GLCM-DIS-U	0.33	0.92
GLCM-DIS-LD	0.35	0.93
GLCM-HOM2-R	-0.87	-0.45
GLCM-HOM2-RD	-0.90	-0.40
GLCM-HOM2-U	-0.86	-0.46
GLCM-HOM2-LD	-0.90	-0.40
GLRLM-RLN-R	0.83	0.48
GLRLM-RLN-RD	0.89	0.42
GLRLM-RLN-U	0.83	0.49
GLRLM-RLN-LD	0.89	0.42
GLRLM-RP-R	0.94	0.29
GLRLM-RP-RD	0.95	0.25
GLRLM-RP-U	0.94	0.31
GLRLM-RP-LD	0.95	0.25

TABLE VIII
TWO CATEGORIES OF METRICS

Category No.	Metrics
1	S_R, S_A, *GLCM-HOM2-R, GLCM-HOM2-RD, GLCM-HOM2-U, GLCM-HOM2-LD, GLRLM-RLN-R, GLRLM-RLN-RD, GLRLM-RLN-U, GLRLM-RLN-LD, GLRLM-RP-R, GLRLM-RP-RD, GLRLM-RP-U, GLRLM-RP-LD*
2	*GLCM-DIS-R, GLCM-DIS-RD, GLCM-DIS-U, GLCM-DIS-LD*

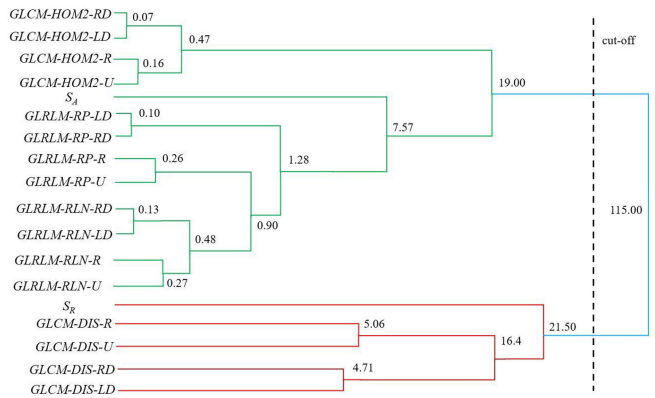

Fig. 12. Dendrogram of the results by HCA.

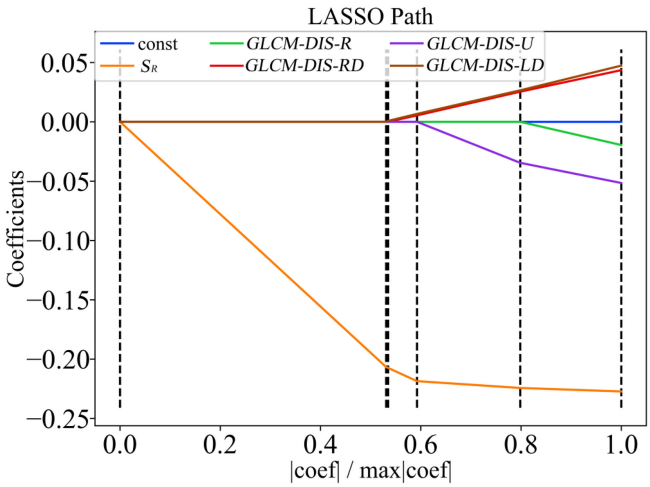

Fig. 13. LASSO path for metrics.

TABLE IX
EXPERIMENTAL IMAGE DATASETS

Name	Number of images	Source
CSIQ	30	[75], [76]
UCID	110	[77]
USC-SIPI-Textures	60	[78]

IV. BUILDING EMPIRICAL MODELS FOR PREDICTING THE LOSSLESS IMAGE COMPRESSION RATIO WITH CONFIGURATIONAL ENTROPY

A. Model Development: Basic Ideas and a Solution

After the most appropriate metric is identified, the next step is to build an appropriate empirical model. In doing so, three aspects should be considered as follows.

1) The data sample size is essential for regression modeling. Theoretically, the size of the data used to train models should be as large as possible.
2) Candidate mathematical models for predicting the lossless compression ratio should be selected from mathematical model sets and analyzed both theoretically and experimentally.
3) An approach will be developed to determine the coefficients of candidate models for fitting the data boundary. The best trend line for the training dataset can be determined and then utilized to determine the upper and lower bounds of the compression ratio.

By following this line of thought, a solution for building models can be proposed. Its implementation procedures are illustrated in Fig. 14. This article aims to develop mathematical models for the upper and lower bounds of the compression ratio. To obtain a reliable model with high accuracy, all images and compression methods used in section III are employed again here to develop mathematical models.

1) Criteria for Candidate Model Selection: In this article, mathematical models must be selected based on specified criteria. In fact, [79] has proposed criteria, which are given in Table X. These criteria are employed here to select models.

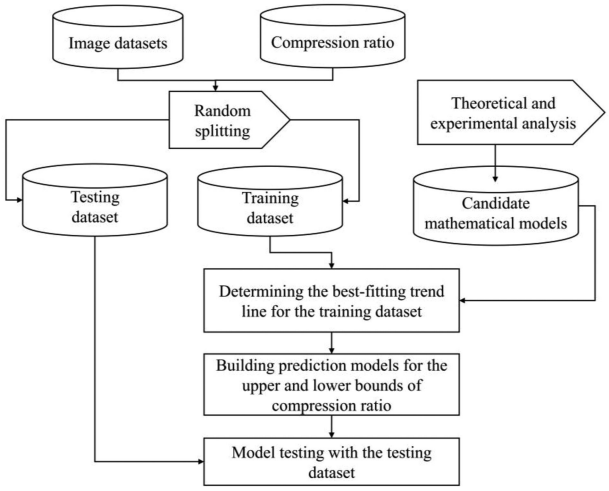

Fig. 14. General procedures for developing models.

TABLE X
CRITERIA FOR EVALUATING MATHEMATICAL MODELS [78]

Criteria	Description
accuracy	the dependent variable values obtained by a model are either correct or nearly correct
descriptive realism	the model should be based on correct assumptions
precision	the prediction results of the model are definite numbers, functions, or geometric figures
robustness	the model is able to tolerate errors in the independent variable data
generality	the model should be useful in various situations
fruitfulness	the model indicates inspiration for building other, better models (e.g., it gives hints for building theoretical models)
parsimony	the model should require the smallest possible number of parameters that adequately represent data obtained from the real world

In this article, for fruitfulness, descriptive realism and accuracy, candidate models should be capable of describing a monotonic relationship, and from the theoretical perspective, the dependent variable values obtained by models should be infinitely large when the independent variable value is 0.

2) Selection of the Most Appropriate Type of Mathematical models: Theoretical and Experimental Analysis: After determining the criteria for model selection, mathematical models are formulated. First, the number of candidate models must be kept small, but it must be large enough to avoid omitting a good prior model [80]. Second, the selection of mathematical models should be conducted through both theoretical and experimental analysis. Based on these two points, theoretically, six mathematical models from CurveExpert [81] are selected; they are given in Table XI. Notably, in terms of form, the equations of the first two models are the same as the theoretical upper and lower bounds defined by Shannon's theorem for the compression ratio. The remaining models are selected from different families.

The candidate mathematical models have been outlined briefly from the theoretical standpoint. Experimental analysis is now conducted to investigate whether in practice these candidate models are reliable for fitting a dataset consisting of the compression ratio and S_R. Fig. 15 shows the plots of the fitted models. Clearly, all models are appropriate for predicting the lossless compression ratio in accordance with the criteria given in Table X. However, they have different performances when S_R is located at different intervals. The increase in the C_R value becomes increasingly larger with the decrease in S_R, as shown in Fig. 15(c), (e) and (f).

TABLE XI
CANDIDATE MATHEMATICAL MODELS

Family	Name	Mathematical equation
PolyRatio	PolyRatio (0,1) Model	$y = \frac{a}{x}$
	Modified PolyRatio (0, 1) Model	$y = \frac{a}{x+b}$
Growth	Michaelis-Menten Model	$y = \frac{ax}{x+b}$
Power	Power Fit Model	$y = ax^b$
Exponential	Modified Exponential Model	$y = ae^{\frac{b}{x}}$
Yield-Density	Harris Model	$y = \frac{1}{a+bx^c}$

B. Establishment of Empirical Models

1) Determination of the Trend Line Based on the Best-Fitting Approach: Here, we use the regression approach to fit all candidate models to the data points. To that end, we need to first determine the weighting of each data point shown in Fig. 16(a). Kernel density estimation (KDE) [82] is employed here to check the distribution of the compression ratio.

As illustrated in Fig. 16(b), the compression ratio values are obviously not consistent with the normal distribution. Moreover, the data are clearly distributed unevenly when S_R falls in the intervals [8.0, 9.0] and [2.0, 3.5], and we do not know how the compression ratio varies as the S_R value varies. To lessen the effect of an insufficient number of data points and to robustly estimate the values of the model coefficients, a hybrid of the bootstrap method [83] and the weighted least squares [84] method is utilized to complete the regression task. Furthermore, considering that many data points have the same S_R value, the following regression is used for minimization:

$$Sum - of - Square = \sum \left(\frac{\mathrm{Y_{data}} - \mathrm{Y_{model}}}{\mathrm{SD}}\right)^2 \quad (6)$$

where Y_{data} is the observed C_R value; Y_{model} is the predicted value; and *SD* is the standard deviation of the observed C_R values.

By using the regression, all fitted models are generated and are plotted in Fig. 17. To determine which model has the best trend line, the Akaike information criterion (*AIC*) [85] is used. The lower the *AIC* value of the fitted model is, the greater the reliability of the model. From the information-theoretic perspective, the *AIC* is used to measure the extent to which the model approaches the "true model" for the data, and it is calculated as follows:

$$AIC = N \times \ln\left(\frac{\mathrm{SS}}{N}\right) + 2K \quad (7)$$

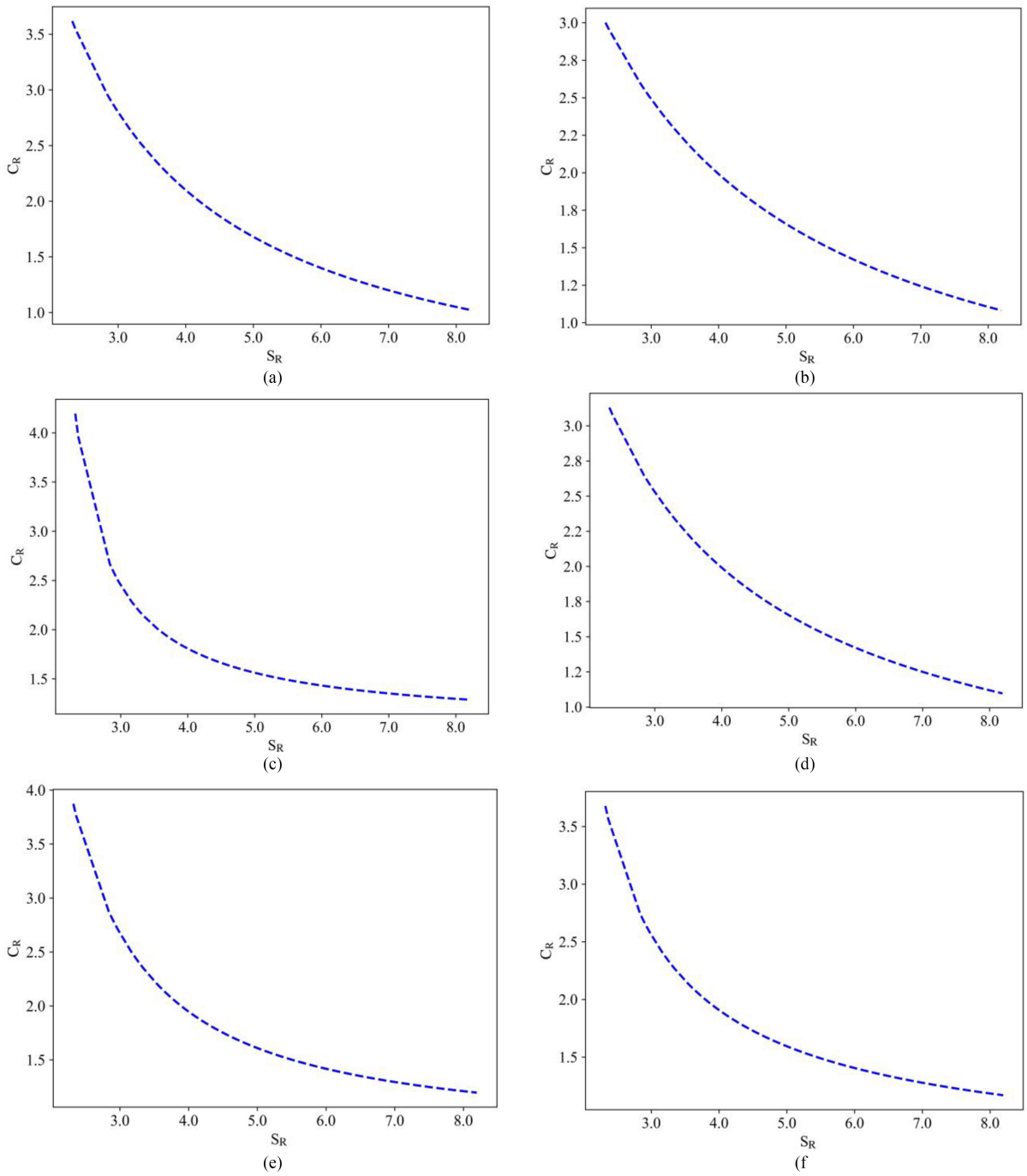

Fig. 15. Plots of models fitted to practical data points. (a) PolyRatio (01) model. (b) Modified PolyRatio (0, 1) model. (c) Michaelis–Menten model. (d) Power fit model. (e) Modified exponential model. (f) Harris model.

where *SS* is the sum of square shown in Equation (6), N is the number of data points used to estimate the model coefficients, and K is the number of parameters plus 1.

Since the calculation of *AIC* is greatly affected by the size of the data samples and model selection uncertainty is involved, *AIC* is used as a reference, not as the sole criterion for selecting the best trend line. Thus, considering the implications of Shannon's coding theorem, the selection priority of the best trend lines is given to the PolyRatio (0, 1) model and the modified PolyRatio (0, 1) model. Fig. 17 clearly shows that there is no significant difference in the *AIC* values of the fitted models. Therefore, both the PolyRatio (0, 1) model and the modified PolyRatio (0, 1) model are considered to have the best trend lines. Their coefficients and the standard errors of the coefficients are given in Table XII. The numerator of the second model is close to 9.0, which is the bit depth of the experimental images plus one.

2) Building Empirical Models: Determination of the Upper and Lower Bounds: Based on the best trend lines, their coefficient values can be adjusted to acquire the upper and lower

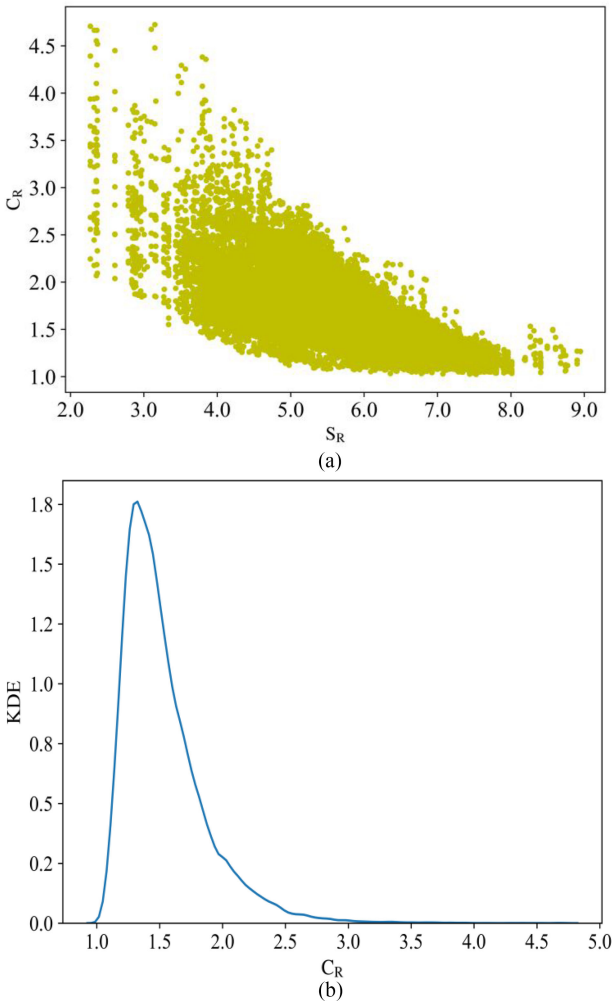

Fig. 16. Visual analysis of the dataset. (a) Scatter plot of S_R against C_R. (b) KDE plot of C_R.

TABLE XII
TWO BEST TREND LINES FOR THE TRAINING DATASET

Model equation	Fitted model	Coefficient standard error (a, b)
$y = \frac{a}{x+b}$	$C_R = \frac{9.7}{S_R + 0.6}$	0.03, 0.02
$y = \frac{a}{x}$	$C_R = \frac{8.8}{S_R}$	0.01

bounds of data. This article aims to build a series of mathematical models for these bounds. To meet this objective, the following strategy is proposed.

1) Determine the general model form for the bounds of the compression ratio based on trend lines.
2) Select the coefficients of the trend lines to be adjusted to generate the upper and lower bounds.

The generation of bounds is aided by a generalized least squares approach [86]. To form models with high consistency, the numerator of the models should be set to be constant, which we can do. Based on the trend lines given in Table XII and the empirical knowledge learned from Shannon's source coding theorem, the general form of the upper and lower bounds is

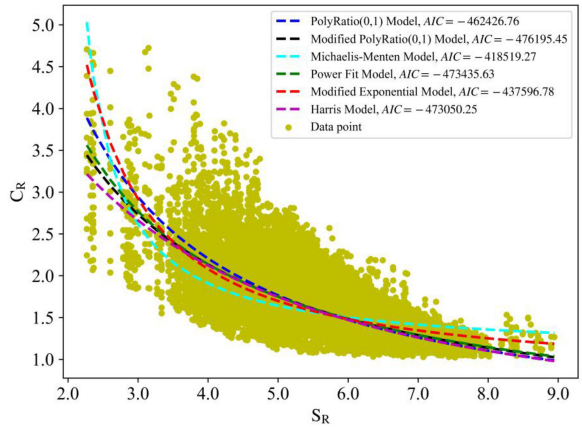

Fig. 17. Scatter plots of S_R against C_R and plots of the fitted models.

TABLE XIII
GENERAL MODEL FORMS FOR THE UPPER AND LOWER BOUNDS

General model form No.	Equation
1	$C_R = \frac{10.0}{e \times S_R + f}$
2	$C_R = \frac{9.0}{e \times S_R + f}$

TABLE XIV
ALL MODELS FOR UPPER AND LOWER BOUNDS

Trend line	Adjusted coefficient	Upper bound	Lower bound
$C_R = \frac{10.0}{e \times S_R + f}$	e, f	$\frac{10.0}{0.7 \times S_R - 0.5}$	$\frac{10.0}{1.5 \times S_R + 1.4}$
	f	$\frac{10.0}{S_R - 1.9}$	$\frac{10.0}{S_R + 3.9}$
$C_R = \frac{9.0}{e \times S_R + f}$	e	$\frac{10.0}{0.4 \times S_R + 0.6}$	$\frac{10.0}{1.8 \times S_R + 0.6}$
	e, f	$\frac{9.0}{0.7 \times S_R - 0.4}$	$\frac{9.0}{1.4 \times S_R + 1.3}$

expressed as follows:

$$C_R = \frac{C}{S_R \times e + f} \tag{8}$$

where C is a constant. Notably, the reason behind the existence of e is that the constant k of configurational entropy was temporarily set to 1[87]. With the two best trend lines for the training dataset, the numerators, i.e., 9.7 and 8.8, are rounded up to the next integer values, i.e., 10.0 and 9.0, respectively. Then, two general forms are given in Table XIII.

Two strategies are employed to fine-tune the coefficients (i.e., e and f) to determine the upper and lower bounds. The first strategy is to adjust the two coefficients simultaneously. The second is to adjust only one coefficient, with the other coefficient remaining unchanged. Table XIV gives all available models for the upper and lower bounds by these two approaches.

These models are graphically shown in Fig. 19. Clearly, in Fig. 18(a) and (d), the model seems to be satisfactory for fitting the boundary. However, in Fig. 18(b), the difference between the upper bound and the lower bound is increasingly higher when the S_R value decreases, whereas in Fig. 18(c), this difference increases smoothly.

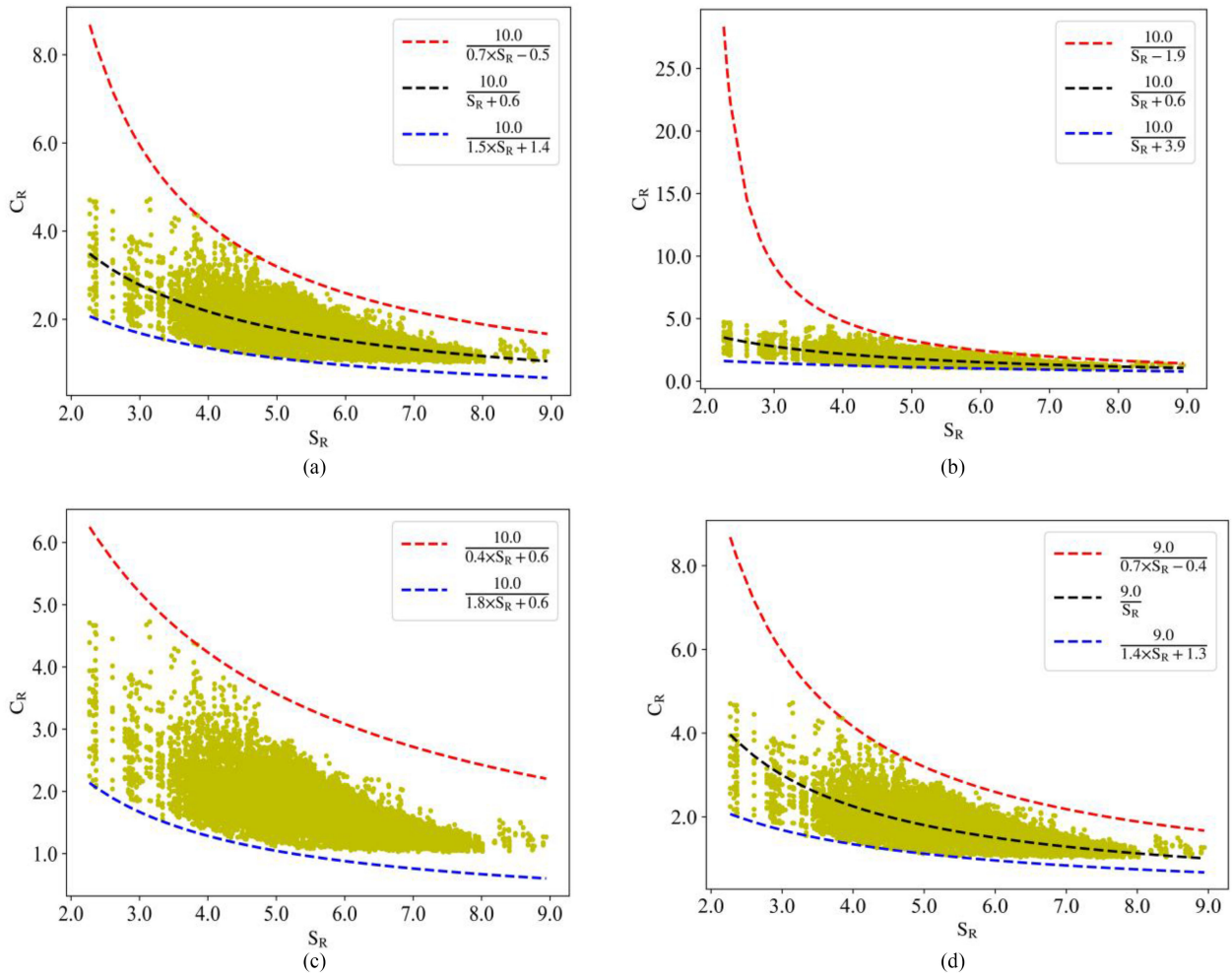

Fig. 18. Plots of the models for the upper and lower bounds based on the general forms. (a)–(c) Based on general model form no. 1. (d) Based on general model form no. 2.

C. Model Testing and Recommendation

Once the empirical models for the upper and lower bounds of the compression ratio have been established, their usefulness should be assessed by the testing dataset. Fig. 19 shows the plots of the models and the testing dataset, indicating that most data points fall within the range between the upper and lower bounds. This means that the established models are effective in predicting the lossless compression ratio in most cases. Nevertheless, some data points are slightly higher than the upper bound. The prediction error of the models is measured by the absolute difference between the C_R value and the predicted value for a given grayscale image regarding data points that fall out of the range formed by the upper and lower bounds. As a result, the maximum prediction errors in Fig. 19(a)–(d) are 0.54, 0.72, 0.03, and 0.54, respectively.

Regarding the recommendation of models, two aspects should be considered. The first aspect is the performance of the models in fitting the data boundary and predicting the compression ratio. The second is that, from the theoretical perspective, the models that are beneficial for developing theoretical models deserve to be recommended. Based on these aspects, the following model is the final recommendation:

$$C_R = \begin{cases} \frac{9.0}{0.7\times S_R - 0.4} & \text{Upper bound} \\ \frac{9.0}{1.4\times S_R + 1.3} & \text{Lower bound} \end{cases}$$

Compared to the other models, this model fits the data boundary well when S_R falls in the whole interval [2.0, 9.0]. Meanwhile, its maximum prediction error (i.e., 0.54) is relatively small and is acceptable. Moreover, the coefficient values in the denominator of this model are very close to each other, indicating high consistency and fruitfulness.

Specifically, the numerator value is 9.0, which is the maximum configurational entropy value of the 8-b experimental grayscale images used in this article and can be regarded as the maximum Shannon entropy (i.e., 8.0) plus one. Therefore, in theoretical terms, it is beneficial to use this model to connect Shannon's theorem and to then build a more comprehensive family of models. Regarding the prediction ability of this model, two points are intuitive. First, when S_R is in the interval [0.57, 9.0], we can use this model to predict the compression ratio of a given image. Second, the C_R predicted by this model is effective when it is not significantly different (e.g., one time, two times) from the true C_R.

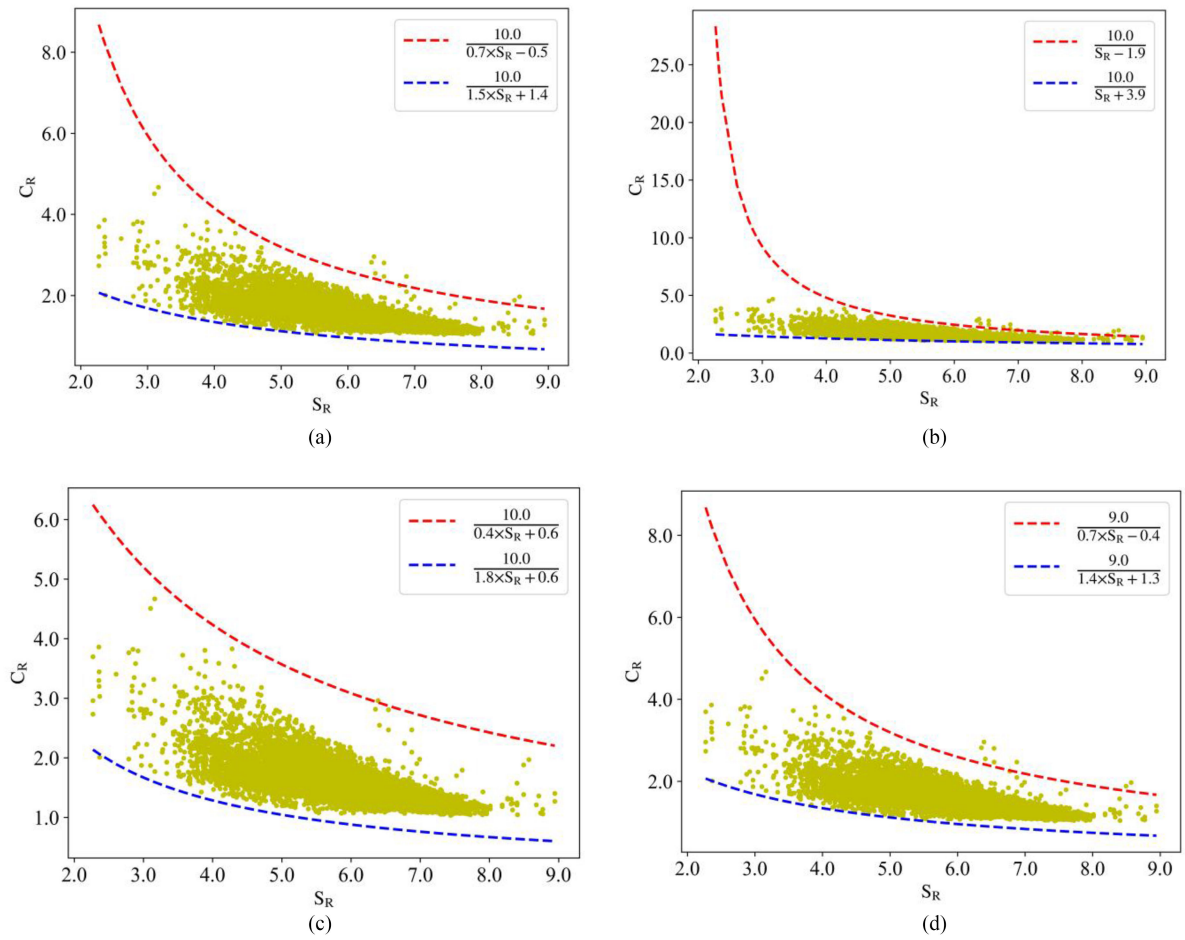

Fig. 19. Plots of the validation dataset and the trained models.

V. Discussion: Shannon's Source Coding Theorem Versus Configurational Entropy-Based Models

This section explores two issues. First, it compares configurational entropy-based models and Shannon's source coding theorem regarding the prediction of the compression ratio. As mentioned earlier, the recommended empirical configurational entropy-based model is very similar to Shannon's theorem. Thus, one may wonder whether there is a definite relationship between Shannon entropy and configurational entropy. We performed a preliminary exploration, and the results, as illustrated in Fig. 20, indicate a correlation between these two kinds of entropy in general. Some points (e.g., A in Fig. 20) have high Shannon entropy but very low configurational entropy. On the other hand, other points (e.g., B in Fig. 20) may have low Shannon entropy, but very high configurational entropy. This also implies that it is not possible to adjust the coefficients of models based on Shannon's theorem to predict the compression ratio by state-of-the-art compression techniques that consider the configurational information of images.

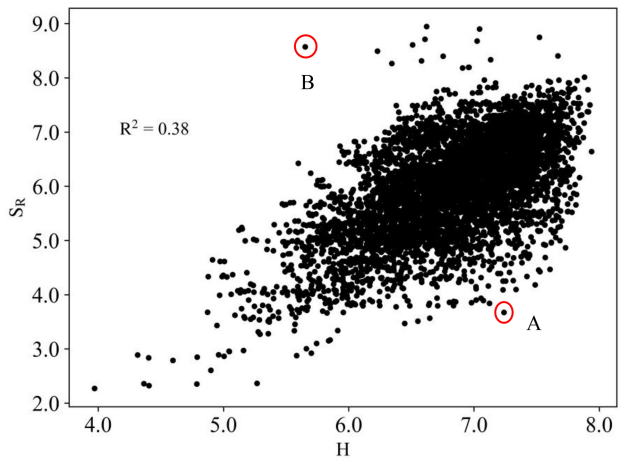

Fig. 20. Scatter plot of Shannon entropy versus configurational entropy.

Second, we observe the performance of models in which the numerator of the models is set to 8.0. By using the approach employed in Section IV, two models are generated and shown in Fig. 21. Clearly, the data boundaries are fitted well by these models, and the model coefficients in the denominator are closer to each other. The range between the upper and lower bounds in Fig. 21(b) is narrower than Fig. 21(a) when S_R is in the interval [2.0, 4.0]. This result means that the model shown in Fig. 21(b) performs better than the other model. Nevertheless, the upper bounds of the compression ratio by these two models

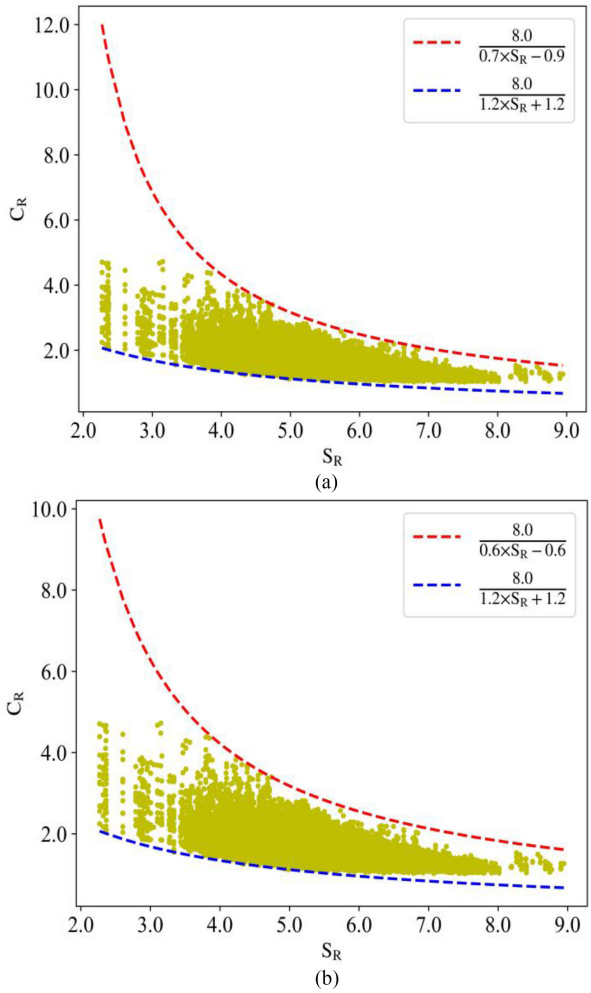

Fig. 21. Models of the upper and lower bounds where the numerator is set to 8.0.

are higher than those of the models illustrated in Fig. 18(a)–(d). Thus, recommending them is not appropriate.

Note that the computation cost for the configurational entropy is far smaller than that for higher order Shannon entropy. For instance, regarding an 8-b 1024 × 1024 image, the computation time for configurational entropy is around 0.4 s which is smaller than that (i.e., around 1.3 s) for third-order Shannon entropy in the operating environment (Intel Core i7–4790 CPU @ 3.60 GHz, 8.00 GB RAM, and 64-b Windows 10). Compared with higher-order Shannon entropy, the memory cost for computing the configurational entropy is always small and acceptable for ordinary users (see [42] for more details).

VI. Conclusion

The compression ratio is a critical factor in the development of compression techniques. Shannon's source coding theorem defines the upper and lower limits of the compression ratio with Shannon entropy, guiding data compression and determining the applicability of various techniques for different data. However, in practice, it is not convenient to use Shannon's theorem to predict the image compression ratio by techniques that consider the configurational information of image. The reason is that the calculation of Shannon entropy is limited by the measurement scale and Shannon entropy measures the statistical information of pixels only, not the actual configurational information of images.

Therefore, it is very desirable to construct new models for predicting the lossless compression ratio without the need to remove interpixel redundancy beforehand. In doing so, a two-step investigation was conducted in this article. The first step was to identify the most appropriate metric to use for building models. The second step was to build empirical models based on this metric. In doing so, a total of 29 metrics that capture the configurational information of pixels were selected, and a multilayered solution was developed to identify the most appropriate metric for building models. Then, a weighted least squares approach was developed to build models. From the experiments, the following conclusions can be drawn as follows.

1) Configurational entropy is the most appropriate for building models.
2) Configurational entropy-based models are effective in predicting the lossless compression ratio. Specifically, a model having high similarity to Shannon's source coding theorem with regard to the compression ratio is finally recommended.

This model provides directions for further building theoretical models. Three aspects are considered for future research. First, theoretically, the maximum configurational entropy for a given grayscale image must be determined. Second, for a single band of a remote sensing image, we can use configurational entropy to mathematically represent its quantification, transformation and context modeling stages in a complex compression scheme and then construct theoretical models. Third, considering the prediction of the lossless compression ratio of multispectral image, the configurational entropy is hopefully conceptualized and then calculated by a feasible solution.

References

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
[2] D. A. Huffman, "A method for the construction of minimum-redundancy codes," *Proc. IRE*, vol. 40, no. 9, pp. 1098–1101, Sep. 1952.
[3] T. Welch, "A technique for high-performance data-compression," *Computer*, vol. 17, no. 6, pp. 8–19, Jun. 1984.
[4] J. Rissanen and G. Langdon, "Arithmetic coding," *IBM J. Res. Develop.*, vol. 23, no. 2, pp. 149–162, Mar. 1979.
[5] X. Wu and N. Memon, "Context-based, adaptive, lossless image coding," *IEEE Trans. Commun.*, vol. 45, no. 4, pp. 437–444, Apr. 1997, doi: 10.1109/26.585919.
[6] JPEG-LS standard, ISO/IEC-14495-1 /ITU-T Rec. T.87, 2002.
[7] M. J. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1309–1324, Aug. 2000, doi: 10.1109/83.855427.
[8] G. Martín, "Range encoding: An algorithm for removing redundancy from a digitised message," in *Proc. Video Data Recording Conf.*, 1979, pp. 24–27.
[9] J. Sneyers and P. Wuille, "FLIF: Free lossless image format based on MANIAC compression," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 66–70.
[10] V. Sze, M. Budagavi, and G. J. Sullivan, "High efficiency video coding (HEVC)," in *Integrated Circuit and Systems, Algorithms and Architectures*, vol. 39, 1st ed., Berlin, Germany: Springer, 2014, pp. 49–90.

[11] High efficiency image file (HEIF), ISO/IEC international standard 23008-12, 2015.

[12] I. Blanes, E. Magli, and J. Serra-Sagrista, "A tutorial on image compression for optical space imaging systems," *IEEE Geosci. Remote Sens. Mag.*, vol. 2, no. 3, pp. 8–26, Sep. 2014, doi: 10.1109/MGRS.2014.2352465.

[13] E. Magli, G. Olmo, and E. Quacchio, "Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 1, pp. 21–25, Jan. 2004, doi: 10.1109/LGRS.2003.822312.

[14] M. I. Afjal, P. Uddin, A. Mamun, and A. Marjan, "An efficient lossless compression technique for remote sensing images using segmentation based band reordering heuristics," *Int. J. Remote Sens.*, vol. 42, no. 2, pp. 756–781, Nov. 2020.

[15] C. Shi, J. Zhang, and Y. Zhang, "Content-based onboard compression for remote sensing images," *Neurocomputing*, vol. 191, pp. 330–340, Feb. 2016.

[16] X. Cheng, "Entropy-based models for predicting lossless compression ratio of grayscale images: from Shannon back to Boltzmann," M.S. thesis, Dept. Land Surverying and Geo-Informatics, The Hong Kong Polytechnic Univ., Hong Kong SAR, China, 2020.

[17] N. Tavakoli, "Entropy and image compression," *J. Vis. Commun. Image Represent.*, vol. 4, no. 3, pp. 271–278, Sep. 1993.

[18] X. Cheng and Z. Li, "How does Shannon's Source Coding Theorem fare in prediction of image compression ratio with current algorithms?," in *Proc. Internatioanl. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2020, pp. 1313–1319.

[19] 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, pp. 19, Jan. 2018.

[20] J. W. Roberts, J. A. van Aardt, and F. B. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, p. 023522, May 2008.

[21] M. Feixas, A. Bardera, J. Rigau, Q. Xu, and M. Sbert, *Information Theory Tools For Image Processing*, San Rafael, CA, USA: Morgan & Claypool, 2014, pp. 1–164.

[22] J. L. Starck and F. Murtagh, "Astronomical image and signal processing: Looking at noise, information and scale," *IEEE Signal Process. Mag.*, vol. 18, no. 2, pp. 30–40, Mar. 2011, doi: 10.1109/79.916319.

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

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

[25] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, Jan. 1996.

[26] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Upper Saddle River, NJ, USA: Prentice Hall, 2007, pp. 409–514.

[27] F. M. Willems, Y. M. Shtarkov, and T. J. Tjalkens, "The context-tree weighting method: Basic properties," *IEEE Trans. Inf. Theory*, vol. 4, no. 3, pp. 653–664, May 1995, 10.1109/18.382012.

[28] R. Begleiter, R. El-Yaniv, and G. Yona, "On prediction using variable order Markov models," *J. Artif. Intell. Res.*, vol. 22, no. 1, pp. 385–421, Dec. 2004.

[29] J. Cleary and I. Witten, "Data compression using adaptive coding and partial string matching," *IEEE Trans. Commun.*, vol. 32, no. 4, pp. 396–402, Apr. 1948, doi: 10.1109/TCOM.1984.1096090.

[30] P. Deutsch, "RFC1951: DEFLATE compressed data format specification version 1.3," pp. 5–14, Accessed: Dec. 5, 2020. [Online]. Available: https://dl.acm.org/doi/pdf/10.17487/RFC1951

[31] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017, doi: 10.1109/JPROC.2017.2675998.

[32] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. 24, no. 5, pp. 530–536, Sep. 1978, doi: 10.1109/TIT.1978.1055934.

[33] M. Burrows and D. Wheeler, "A block sorting lossless data compression algorithm," Digital Equip. Corp., Maynard, MA, USA, Tech. Rep. SRC-RR-124, May 10, 1994. [Online]. Available: https://www.hpl.hp.com/techreports/Compaq-DEC/SRC-RR-124.pdf

[34] J. Seward, Bzip2 file compression program, Sep. 10, 2019. [Online]. Available: https://sourceforge.net/projects/bzip2/

[35] J. A. Storer and T. G. Szymanski, "Data compression via textual substitution," *J. ACM*, vol. 29, no. 4, pp. 928–951, Oct. 1982.

[36] Zip file format Specification, PKWARE Inc., Milwaukee, WI, USA, 2004. Accessed: Jul. 15, 2020, [Online]. Available: https://web.archive.org/web/20040819182806/http://www.pkware.com/company/standards/appnote/

[37] D. O'Shaughnessy, "Linear predictive coding," *IEEE Potentials*, vol. 7, no. 1, pp. 29–32, Feb. 1988, doi: 10.1109/45.1890.

[38] S. Dewitte and J. Cornelis, "Lossless integer wavelet transform," *IEEE Signal Process. Lett.*, vol. 4, no. 6, pp. 158–160, Jun. 1997, doi: 10.1109/97.586035.

[39] J. W. Schwartz and R. C. Barker, "Bit-plane encoding: A technique for source encoding," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-2, no. 4, pp. 385–392, Jul. 1996, doi: 10.1109/TAES.1966.4501787.

[40] P. Gao, H. Zhang, and Z. Li. "A hierarchy-based solution to calculate the configurational entropy of landscape gradients," *Landscape Ecol.*, vol. 32, no. 6, pp. 1133–1146, Apr. 2017.

[41] P. Gao and Z. Li, "Aggregation-based method for computing absolute configurational entropy of landscape gradient with full thermodynamic consistency," *Landscape Ecol.*, vol. 34, no. 8, pp. 1837–1847, Jul. 2019.

[42] P. Gao, H. Zhang, and Z. Li. "An efficient analytical method for computing the boltzmann entropy of a landscape gradient," *Trans. GIS*, vol. 22, no. 5, pp. 1046–1063, Jan. 2018.

[43] B. Bhanu, A. S. Politopoulos, and B. A. Parvin, "Intelligent autocueing of tactical targets," in *Proc. Architectures Algorithms Digit. Image Process*, 1983, pp. 90–98.

[44] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995, doi: 10.1109/26.477498.

[45] R. M. Haralick and K. Shanmugam, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, doi: 10.1109/TSMC.1973.4309314.

[46] F. Albregtsen, Statistical texture measures computed from gray level coocurrence matrices. pp. 1–12, Accessed: Dec. 5, 2020. [Online]. Available: https://www.researchgate.net/profile/Fritz_Albregtsen/publication/255652481_Statistical_Texture_Measures_Computed_from_Gray_Level_Coocurrence_Matrices/links/5460a0aa0cf27487b451351e/Statistical-Texture-Measures-Computed-from-Gray-Level-Coocurrence-Matrices

[47] J. J. Van Griethuysen *et al.*, "Computational radiomics system to decode the radiographic phenotype," *Cancer Res.*, vol. 77, no. 21, pp. e104–e107, Nov. 2017.

[48] F. T. Ulaby, F. Kouyate, B. Brisco, and T. L. Williams, "Textural information in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 2, pp. 235–245, Mar. 1986, doi: 10.1109/TGRS.1986.289643.

[49] C. Parmar *et al.*, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PloS One*, vol. 9, no. 7. Jul. 2014, Art. no. e102107.

[50] C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Comput. Vis. Graph. Image Process.*, vol. 23, no. 3, pp. 341–352, Sep. 1983, doi: 10.1016/0734-189X(83)90032-4.

[51] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Trans. Syst., Man, Cybern.*, vol. 19, no. 5, pp. 1264–1274, Sep./Oct. 1989, doi: 10.1109/21.44046.

[52] M. M. Galloway, "Texture classification using gray level run length," *Comput. Graph. Image Process.*, vol. 4, no. 2, pp. 172–179, Jun. 1975.

[53] G. Thibault, "Shape and texture indexes application to cell nuclei classification," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 27, no. 1, pp. 1357002, Mar. 2013.

[54] I. Sobel and G. Feldman, "A 3x3 isotropic gradient operator for image processing," presented at the *Stanford Artificial Intelligence Project (SAIL)*, 1968. [Online]. Available: https://www.researchgate.net/publication/281104656_An_Isotropic_3x3_Image_Gradient_Operator?channel=doi&linkId=55d5876408ae43dd17de57a4&showFulltext=true

[55] H. Yu and S. Winkler, "Image complexity and spatial information," in *Proc. 5th Int. Workshop Qual. Multimedia Experience*, 2013, pp. 12–17.

[56] J. Hauke and T. Kossowski, "Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data," *Quaestiones Geographicae*, vol. 30, no. 2, pp. 87–93, Jun. 2011.

[57] T. D. Gautheir,"Detecting trends using Spearman's rank correlation coefficient," *Environ. Forensics*, vol. 2, no. 4, pp. 359–362, Dec. 2010.

[58] T. R. Vetter, "Fundamentals of research data and variables: The devil is in the details," *Anesthesia Analgesia*, vol. 125, no. 4, pp. 1375–1380, Oct. 2017.

[59] P. Schober, C. Boer, and L. A. Schwarte,. "Correlation coefficients: Appropriate use and interpretation," *Anesthesia Analgesia*, vol. 126, no. 5, pp. 1763–1768, May 2018.

[60] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, vol. 50, no. 302, pp. 157–175. Jul. 1900.
[61] D. N. Lawley and A. E. Maxwell, "Factor analysis as a statistical method," *J. Roy. Stat. Soc. Ser. D (Statistician)*, vol. 12, no. 3, pp. 209–229. 1962.
[62] T. Asparouhov and B. Muthén, "Exploratory structural equation modeling, Structural equation modeling: A multidisciplinary journal," vol. 16, no. 3, pp. 397–438. Jul. 2009.
[63] H. F. Kaiser and J. Rice, "Little Jiffy, Mark iV," *Educ. Psychol. Meas.*, vol. 34, no. 1, pp. 111–117, Feb. 1974.
[64] G. W. Snedecor and W. G. Cochran, *Statistical Methods*, 8th ed., Ames, IA, USA: Iowa State Univ. Press, 1989.
[65] R. MacCallum, "A comparison of factor analysis programs in SPSS, BMDP, and SAS," *Psychometrika*, vol. 48, no. 2, pp. 223–231, Jun. 1983.
[66] J. O. Kim, O. Ahtola,, P. E. Spector, and, and C. W. Mueller. *Introduction to Factor Analysis: What it is and How to Do it*, Newbury Park, CA, USA: Sage Univ. Press, 1987, pp. 7–50.
[67] J. A. S. Almeida, L. M. S. Barbosa, A. A. C. C. Pais, and S. J. Formosinho, "Improving hierarchical cluster analysis: A new method with outlier detection and automatic clustering," *Chemometrics Intell. Lab. Syst.*, vol. 87, no. 2, pp. 208–217, Jun. 2007.
[68] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc., Ser. B (Methodol.)*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.
[69] W. Su,, M. Bogdan, and E. Candes, "False discoveries occur early on the Lasso path," *Ann. Statist.*, vol. 45, no. 5, pp. 2133–2150, Oct. 2017.
[70] P. Bühlmann and S. Van De Geer, *Statistics For High-Dimensional Data: Methods, Theory and Applications*. New York, NY, USA: Springer, 2011.
[71] D. B. Kidner and D. H. Smith, "Advances in the data compression of digital elevation models," *Comput. Geosci.*, vol. 29, no. 8, pp. 985–1002, Oct. 2003.
[72] TIFF 6.0 Specification, Adobe Syst., Adobe Systems Incorporated, Mountain View, CA, USA, 1992. Accessed: May 20, 2019. [Online]. Available: https://www.adobe.io/content/dam/udp/en/open/standards/tiff/TIFF6.pdf
[73] Portable Network Graphics (PNG) specification, (Second Edition), RFC 2083, ISO/IEC 15948:2003, World Wide Web Consortium, Cambridge, MA, USA, 2003. Accessed: Dec. 20, 2020. [Online]. Available: https://www.w3.org/TR/2003/REC-PNG-20031110/
[74] A. Skodras, C. Christopoulos and T. Ebrahimi, "The JPEG 2000 still image compression standard," *IEEE Signal Process. Mag.*, vol. 18, no. 5, pp. 36–58, Sep. 2001, doi: 10.1109/79.952804.
[75] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electron. Imag.*, vol. 19, no. 1, May 2010.
[76] E. C. Larson and D. M. Chandler, "Categorical Image Quality (CSIQ) Database," 2010. Accessed: May 20, 2019. [Online]. Available: http://vision.okstate.edu/csiq
[77] G. Schaefer and M. Stich. "UCID: An uncompressed color image database," in *Proc. Storage Retrieval Methods Appl. Multimedia, Int. Soc. Opt. Photon.*, 2004, pp. 472–481.
[78] "The USC-SIPI image database" Signal and Image Processing Institute, USC, USA,, 1997. Accessed: Dec. 10, 2020. [Online]. Available: http://sipi.usc.edu/services/database/data-base.html
[79] W. J. Meyer, *Concepts of Mathematical Modelling*, Mineola, New York, NY, USA: Dover Publ., 2012. pp. 139–405.
[80] D. Anderson and K. Burnham, *Model Selection and Multi-Model Inference*. 2nd ed. New York, NY, USA: Springer, 2004, pp. 352–434.
[81] D.G. Hyams, "CurveExpert professional documentation," Release 2.7.3, Oct. 17, 2020. [Online]. Available: https://www.curveexpert.net/docs/curveexpert/pro/pdf/CurveExpertProfessional.pdf
[82] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Ann. Statist.*, vol. 38, no. 5, pp. 2916–2957. Aug. 2010.
[83] J. Nevitt and G. R. Hancock, "Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling," *Struct. Equation Model.*, vol. 8, no. 3, pp. 353–377, Nov. 2009, doi: 10.1207/S15328007SEM0803_2.
[84] T. Strutz, "Data fitting and uncertainty: A practical introduction to weighted least squares and beyond," Wiesbaden, Germany: Springer, 2010, pp. 25–103.
[85] H. Akaike, "Information theory and an extension of the maximum-likelihood principle," in *Selected Papers of Hirotugu Akaike*, New York, NY, USA: Springer, 1998, pp. 199–213.
[86] N. Cardiel, "Data boundary fitting using a generalized least-squares method," *Monthly Notices Roy. Astron. Soc.*, vol. 396, no. 2, pp. 680–695, doi: 10.1111/j.1365-2966.2009.14749.x.
[87] S. A. Cushman, "Calculating the configurational entropy of a landscape mosaic," *Landscape Ecol.*, vol. 31, no. 3, pp. 481–489, Nov. 2016.

Xinghua Cheng (Student Member, IEEE) received the B.S. degree in GIS from Nanjing Normal University, Nanjing, China, in 2018 and the M.S. degree in geomatics from The Hong Kong Polytechnic University, Hong Kong, in 2020.

He is currently a Research Associate with the Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University. His research interests include Boltzmann entropy (thermodynamic entropy), spatial information theory, statistical thermodynamics of remote sensing imagery, and machine learning.

Zhilin Li received the B.Sc. degree in photogrammetry and remote sensing from Southwestern Jiaotong University, Chengdu, China, in 1982 and his Ph.D. degree from the University of Glasgow, Glasgow, U.K., in 1990.

He is a Full Professor in geo-informatics. He was an Assistant Professor with the Hong Kong Polytechnic University in 1996. He was an Associate Professor in 1998 and a Professor in 2003. He has a more than 200 journal papers and three research monographs. His research interests include digital cartography, spatial data modeling, feature extraction from remote sensing images.