

# HOW DOES SHANNON'S SOURCE CODING THEOREM FARE IN PREDICTION OF IMAGE COMPRESSION RATIO WITH CURRENT ALGORITHMS?

Xinghua Cheng<sup>1</sup>, Zhilin. Li<sup>1,2,\*</sup>

<sup>1</sup> Dept. of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, Hong Kong - xingh.cheng@connect.polyu.hk

<sup>2</sup> Faculty of Geosciences and Environmental Engineering & State-Province Joint Engineering Laboratory in Spatial Information Technology for High-Speed Railway Safety, Southwest Jiaotong University, Chengdu, China - lszlli@polyu.edu.hk

**KEY WORDS:** Coding theorem, Image quality, Information theory, Shannon entropy, Compression ratio

## ABSTRACT:

Images with large volumes are generated daily with the advent of advanced sensors and platforms (e.g., satellite, unmanned autonomous vehicle) of data acquisition. This incurs issues on the storage, processing, and transmission of images. To address such issues, image compression is essential and can be achieved by lossy and/or lossless approaches. With lossy compression, a high compression ratio can usually be achieved but the original data can never be completely recovered. On the other hand, with lossless compression, the original information is well reserved. Lossless compression is very desirable in many applications such as remote sensing, geological surveying. Shannon's source coding theorem has defined the theoretical limits of compression ratio. However, some researchers have discovered that some compression techniques have achieved a compression ratio that is higher than the theoretical limits. Then, two questions naturally arise, i.e., "When this happens?" and "Why this happens?". This study is dedicated to giving answers to these two questions. Six algorithms are used to compress 1650 images with different complexities. The experimental results show that the generally acknowledged Shannon's coding theorem is still good enough for predicting compression ratio by the algorithms with consideration of statistical information only, but not capable of predicting compression ratio by the algorithms with consideration of configurational information of pixels. Overall, this study indicates that new empirical (or theoretical) models for predicting lossless compression ratio can be built with metrics capturing configurational information.

## 1. INTRODUCTION

Images with large volumes are generated daily with the advent of advanced sensors (e.g., high spatial and spectral resolutions) and platforms (e.g., satellite, unmanned autonomous vehicle, and mobile devices) of data acquisition. This incurs a big headache on the storage, processing, and transmission of images. To solve such a problem, image compression is essential and thus has become an essential research topic in the remote sensing community. In General, image compression can be achieved by lossy and/or lossless approaches. Lossy compression usually achieves a high compression ratio, while the original data can never be completely reconstructed. On the other hand, with lossless compression, information is completely reserved though a lower compression ratio is achieved. Compared with lossy compression, lossless compression is very desirable in many applications such as remote sensing, geological surveying, cartography, and medical imaging.

Owing to the remarkable efforts devoted by researchers from different fields, lots of lossless compression techniques have been developed. Those techniques are various in terms of performances (e.g., compression ratio, compression cost, compression time). It is interesting to point out that the compression ratio has attracted much attention from investigators of compression techniques. Shannon's source coding theorem (Shannon 1948) originated in the field of telecommunication has already defined the upper and lower limits of compression ratio. From the theoretical point of view, Shannon's coding theorem works well for some early compression techniques based on statistical coding principles. That is, the statistical information

(i.e., the proportion & values) of image pixels is utilized to compress an image. However, some researchers (Tavakoli, 1993, Larkin, 2016) have discovered that some compression techniques have achieved a compression ratio that is higher than the theoretical limit defined by Shannon's source coding theorem. In this respect, two questions naturally arise, i.e., "When this happens?" and "Why this happens?". As a result, this study is dedicated to giving answers to these two questions and indicating the implausible research topics in the future.

## 2. SHANNON'S CODING THEOREM AND IMAGE COMPRESSION

Traditionally, Shannon source coding theorem (Shannon, 1948) clearly states that the average code length (refers to  $L_{av}$ ) we can best achieve is as follows:

$$H \leq L_{av} < H + 1 \quad (1)$$

where  $H$  represents the Shannon entropy of an image.

Shannon entropy measures the uncertainty in a random variable and is calculated with the occurrence probability of individual gray level.

$$H = - \sum_{i=1}^n P(r_i) \log P(r_i) \quad (2)$$

Where  $r_i$  denotes the gray level;  $P(r_i)$  is its occurrence probability within an image;  $\log(\cdot)$  denotes the logarithm to base (e.g., 2, 10).

\* Corresponding author

In practice, the units of information are decimal units when the logarithm to bases 10 is used; that of information become bits when the logarithm to base 2 is used. For a binary image, the logarithm base to 2 shall be employed.

Regarding the average codeword length, its calculation formula is as follows:

$$L_{av} = - \sum_{i=1}^n P(r_i) \times l(r_i) \quad (3)$$

Where  $l(r_i)$  is the length of the codes for gray level  $r_i$ .

To explain Shannon entropy and average codeword length, an example shown in Table 1 is employed, in which  $H$  is 2.2 bits per pixel and  $L_{av}$  is 2.4 bits per pixel. This case complies with Shannon's coding theorem.

$r_i$	$P(r_i)$	Code	$l(r_i)$	$P(r_i) \times l(r_i)$
12	0.36	11	2	0.72
15	0.24	01	2	0.48
16	0.15	10	2	0.30
17	0.13	111	3	0.39
18	0.12	1110	4	0.48

Table 1 An example of average codeword length

Shannon's coding theorem is graphically shown in Figure 1. We can easily find that the range between the upper bound and the lower bound is unchanged along with the increase in the Shannon entropy value.

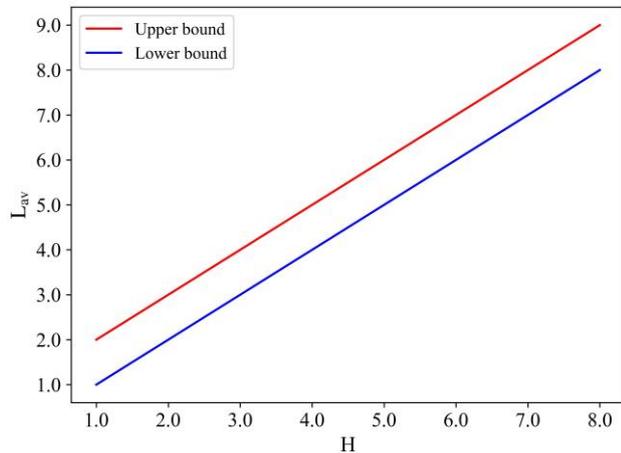


Figure 1. Shannon's source coding theorem on the average codeword length

In practical compression projects, it is not always able to calculate the average codeword length. However, the compression ratio (refers to  $C_R$ ) can be easily calculated. Therefore, when Shannon's coding theorem is applied to image compression, supposing each pixel of the original image is encoded with a byte (8 bits), it can be converted into as follows:

$$\frac{8}{H+1} < C_R \leq \frac{8}{H} \quad (2)$$

where  $C_R$  is the ratio of bytes for storing the original data and that for storing the compressed data.

At this point, the plots shown in Figure 1 are thus converted into plots shown in Figure 2. It is obvious that the theoretical upper and lower limits are inversely proportional to Shannon entropy of an image. This is dependent on the amount of information contained in an image. Moreover, we can easily find that the

range between the upper bound and lower bound of  $C_R$  become wider and wider along with the reduction in the Shannon entropy value. According to Shannon's coding theorem, when the Shannon entropy value approaches 0, the theoretical maximum compression ratio is infinitely large.

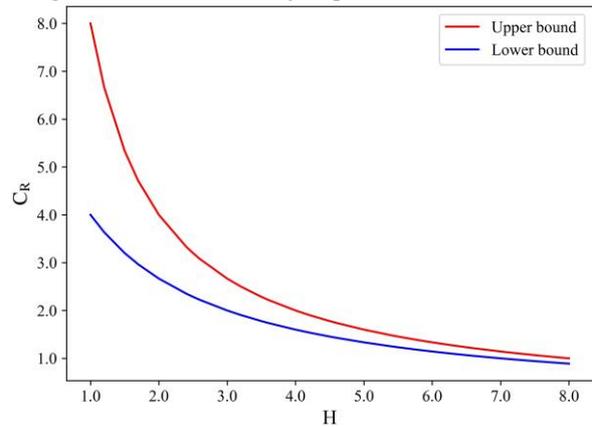


Figure 2. Shannon's source coding theorem on compression ratio

### 3. METHODS AND DATA

In this study, a preliminary investigation is reported into the validity of Shannon's source coding theorem for predicting lossless compression ratio by different lossless image compression techniques. To conduct such evaluation experiments, typical compression techniques are selected in accordance with two categories, i.e., (a) with consideration of statistical information only and (b) with consideration of configurational information, based on our assumption that Shannon's source coding theorem will become invalid in the prediction of compression ratio when an algorithm takes into account not only the statistical information but also the configurational information of image pixels in that the Shannon entropy is able to capture only the statistical information of image pixels but not the configurational information (Gao et al. 2018). At this sense, we selected three algorithms, i.e., Shannon coding (Shannon, 1948), Huffman coding (Huffman 1952), and Arithmetic coding (Rissanen, Langdon, 1979) from category (a), and three algorithms, i.e., LZMA (Ziv, Lempel, 1977), JPEG-LS (Richter, Ogawa, 1999 Weinberger et al. 2000) and Deflate (Ziv, Lempel, 1977, Deutsch, 1996) from category (b).

Figure 3 shows the general processes for generating data sets by those algorithms. More specifically, a total of 1650 images with

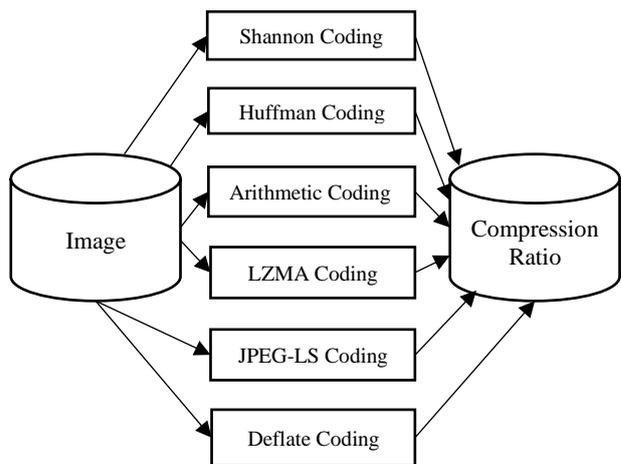


Figure 3 The flowchart of generating data sets

different complexities were extracted from the open accessed image dataset called “NWPU-RESISC45” (Cheng et al., 2017). Concretely, the first 150 images were extracted in accordance with each of the first 11 scene classes in this dataset. Only grayscale images were involved in this study. Thus, red bands of those images were used and compressed to generate the compression ratio datasets which were then compared with the theoretical upper bound defined by Shannon’s coding theorem.

#### 4. RESULTS AND ANALYSIS

This section shows the experimental results by all selected techniques. Concretely, Figure 4, Figure 5 and Figure 6 respectively show the scatter plots of Shannon entropy against compression ratio obtained by Shannon coding, Huffman coding and Arithmetic coding.

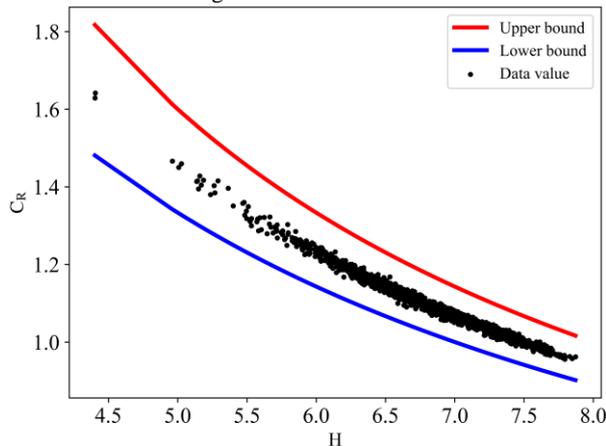


Figure 4. Theoretical bounds of Shannon’s theorem and the scatter plot of H against  $C_R$  obtained by Shannon coding

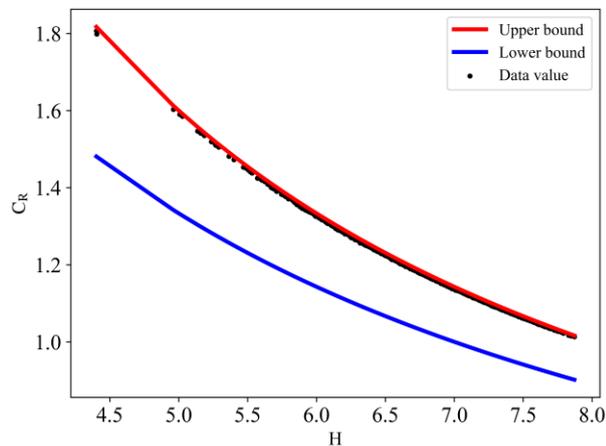


Figure 5. Theoretical bounds of Shannon’s theorem and the scatter plot of H against  $C_R$  values obtained by Huffman coding

We can easily find that Shannon’s source coding theorem is good enough for predicting such lossless compression ratio. Concretely, a high correlation between H and those compression ratios can be easily discovered. The compression ratios by Shannon coding fall within the upper bound and the lower bound defined by Shannon entropy. Note that Huffman coding achieves compression ratios that are very closed to the upper bound. It is worth noting that Arithmetic coding is the most powerful in comparison with the other three algorithms from category (a) in that its resulted achieved compression ratios have reached Shannon’s coding bound. As far, all experimental results comply with Shannon’s coding theorem.

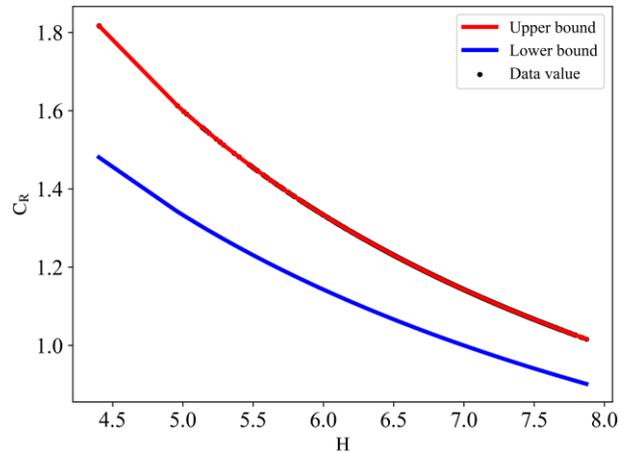


Figure 6. Theoretical bounds of Shannon’s theorem and the scatter plot of H against  $C_R$  values obtained by Arithmetic coding

However, when algorithms (i.e., LZMA, Deflate and JPEG-LS) considering structural information of pixels are utilized to compress images, the resulting compression ratio values significantly surpass the upper bound given by Shannon’s coding theorem. This can be seen in Figure 7, Figure 8, and Figure 9. Moreover, some compression ratio values are much higher than the upper bound as LZMA, JPEG-LS and Deflate utilize the structures of image pixels to achieve compression. This means that Shannon’s coding theorem is invalid when contexts and structures of image pixels are considered to compress an image. Meanwhile, we can discover that no obvious relationships between H and  $C_R$  by JPEG-LS can be discerned. It is noticed that JPEG-LS and Deflate are not able to completely exceed the bounds by Shannon’s coding theorem since some compression ratio values fall within the upper and the lower bound. This is true as the same compression technique perform differently upon different images.

As mentioned above, Shannon’s coding theorem is ineffective for predicting lossless compression ratio by those techniques that consider the structural information of pixels. In fact, those techniques utilize the inter-pixel redundancy to compress data, achieving a high compression ratio that is higher than Shannon’s coding bound.

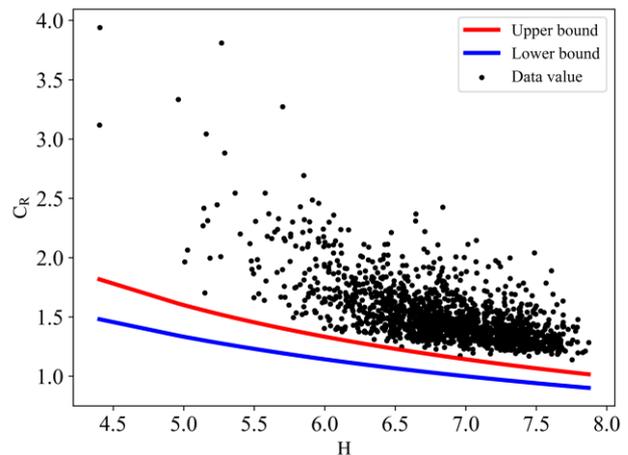


Figure 7. Theoretical bounds of Shannon’s theorem and the scatter plot of H against  $C_R$  obtained by LZMA

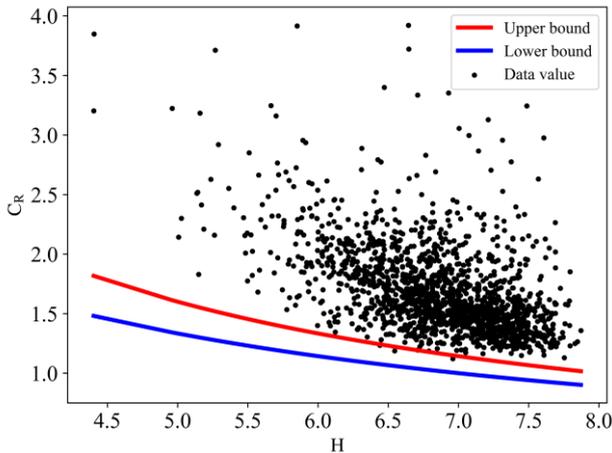


Figure 8. Theoretical bounds of Shannon's theorem and the scatter plot of H against  $C_R$  obtained by JPEG-LS

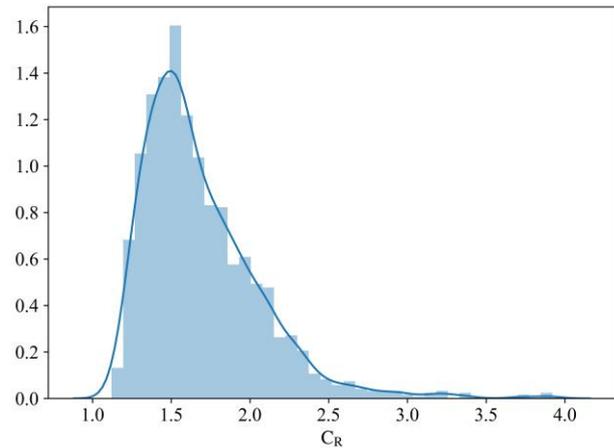


Figure 11. Distribution plot of  $C_R$  obtained by JPEG-LS

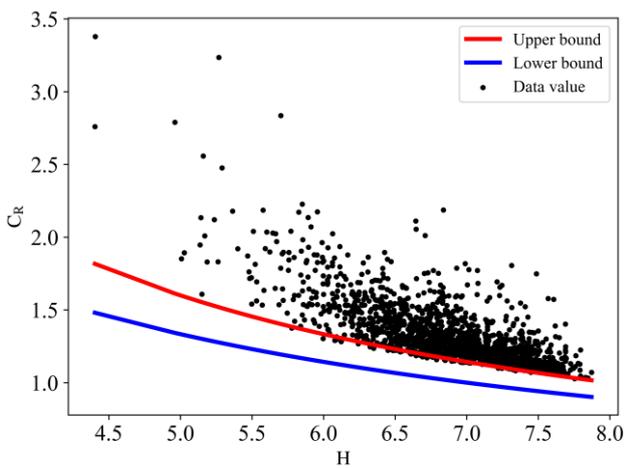


Figure 9. Theoretical bounds of Shannon's theorem and the scatter plot of H against  $C_R$  obtained by Deflate

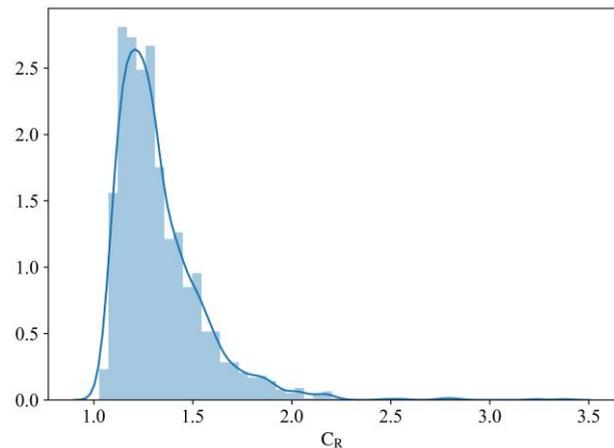


Figure 12. Distribution plot of  $C_R$  obtained by Deflate.

To further decipher the distribution pattern of compression ratio obtained by LZMA, JPEG-LS and Deflate, Kernel Density Estimation (KDE) (Terrell, Scott, 1992, Botev et al. 2010) is employed here. Three such plots are shown in Figure 10, Figure 11 and Figure 12. Indeed, most compression ratio values are located within an interval, i.e., 1.0 to 2.5. This is due to the range

of Shannon entropy values of images is not enough wide.

The generally acknowledged Shannon entropy is also called first-order Shannon entropy since it is calculated with the occurrence probability of individual grey level. In fact, Shannon (1948) has already proposed the definition of different order Shannon entropies. The definitions of first- and second-order Shannon entropies are as follows:

*First-order approximation has independent symbols with different probabilities.*

*Second-order approximation has symbol pairs with known probabilities.*

For the image shown in Figure 13, its second-order Shannon entropy is calculated as 2.9 bits per pixel according to the gray-level pair probabilities tabulated in Table 2.

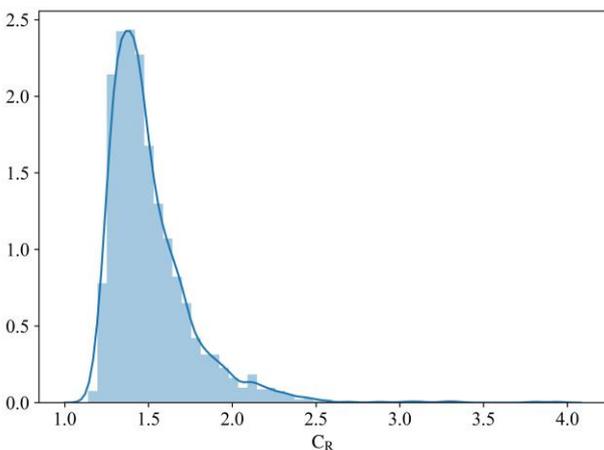


Figure 10. Distribution plot of  $C_R$  obtained by LZMA

11	11	11	10	10	11
10	11	10	11	11	10
10	11	11	11	11	11
254	254	109	109	254	254
244	244	244	244	244	244
254	254	11	11	10	11

Figure 13 A  $6 \times 6$  grayscale image

Gray-level pair	Count	Probability
(11, 11)	8	8/30
(11, 10)	4	4/30
(10, 10)	1	1/30
(10, 11)	5	5/30
(254, 254)	3	3/30
(254, 109)	1	1/30
(109, 109)	1	1/30
(109, 254)	1	1/30
(244, 244)	5	5/30
(254, 11)	1	1/30

Table 2 Occurrence probabilities of gray-level pairs for the image shown in Figure 13

With the definition of higher-order Shannon entropies, corresponding upper bounds are shown in Table 3 where NA means “unavailable”. Higher-order Shannon entropies consider a few structures of pixels. Thus, they are more powerful than first-order Shannon entropy.

Order	Symbol	Upper bound	Lower bound
1	H	$\frac{8}{H}$	$\frac{8}{H + 1}$
2	H <sub>2</sub>	$\frac{16}{H_2}$	NA
3	H <sub>3</sub>	$\frac{24}{H_3}$	NA
n	H <sub>n</sub>	$\frac{8 \times n}{H_n}$	NA

Table 3 Higher-order Shannon entropies and bounds for an 8-bits image

When the second-order Shannon entropy is utilized to evaluate Shannon’s theorem in the prediction of lossless compression

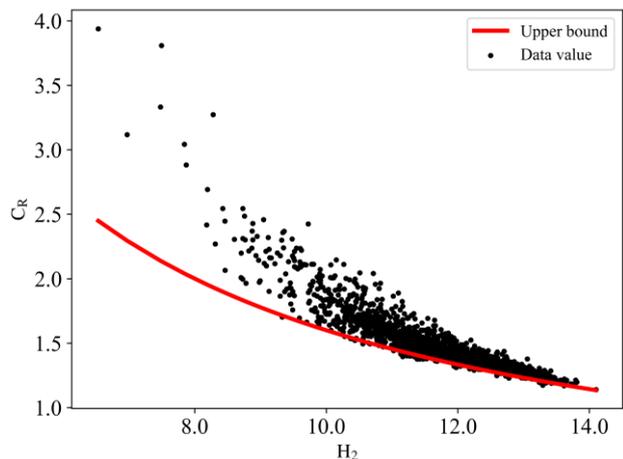


Figure 14 The upper bound by H<sub>2</sub> and scatter plot of H<sub>2</sub> against C<sub>R</sub> obtained by LZMA

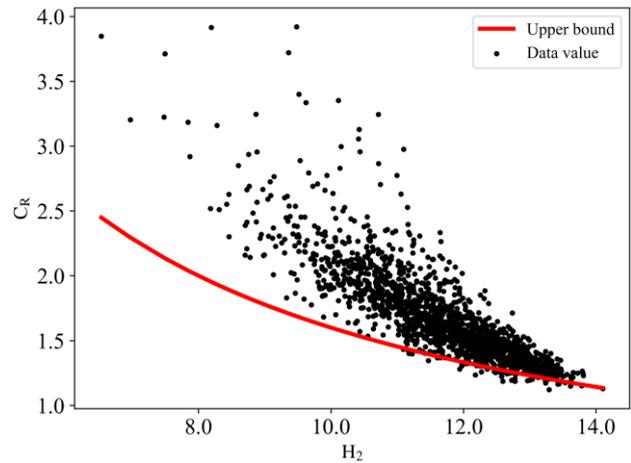


Figure 15 The upper bound by H<sub>2</sub> and scatter plot of H<sub>2</sub> against C<sub>R</sub> obtained by JPEG-LS

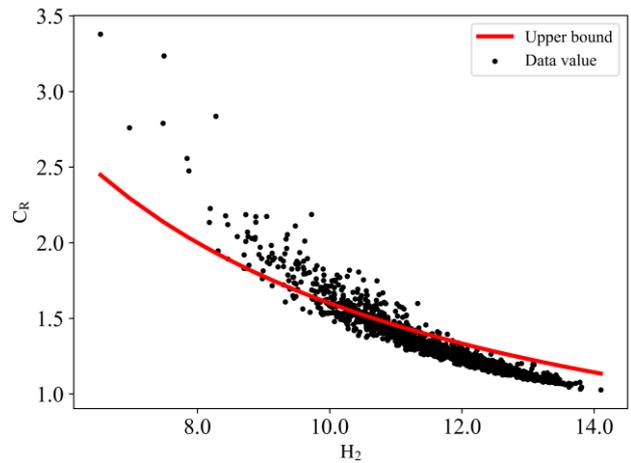


Figure 16 The upper bound by H<sub>2</sub> and scatter plot of H<sub>2</sub> against C<sub>R</sub> obtained by Deflate

ratio, the results are graphically described in Figure 14, Figure 15 and Figure 16. Obviously, second-order Shannon entropy is more powerful than first-order Shannon entropy as more compression ratio values are below the upper bound, as shown in Figure 16. Nevertheless, second-order Shannon entropy is still invalid for predicting lossless compression ratio. From the theoretical point of view, higher-order Shannon entropies are not able to completely capture structure information of pixels as they only consider the 1D structure of pixels, not the 2D structures and contexts. Thus, they are destined to be ineffective for predicting compression ratio by techniques considering structures or contexts of pixels

## 5. DISCUSSION

Shannon’s coding theorem is based on Shannon entropy. As a kind of entropy capturing the statistical information of pixels, Shannon entropy is not useful for describing structural information. Four images shown in Figure 17 are with the same composition. Thus, the same Shannon entropy values (i.e., 7.6 bits per pixel) are calculated for four images shown in Figure 17. In this respect, we can discover that Shannon entropy is questionable for distinguishing images. This can be used to explain why Shannon’s source coding theorem is ineffective for

predicting lossless compression ratio by techniques that consider structure information of pixels.

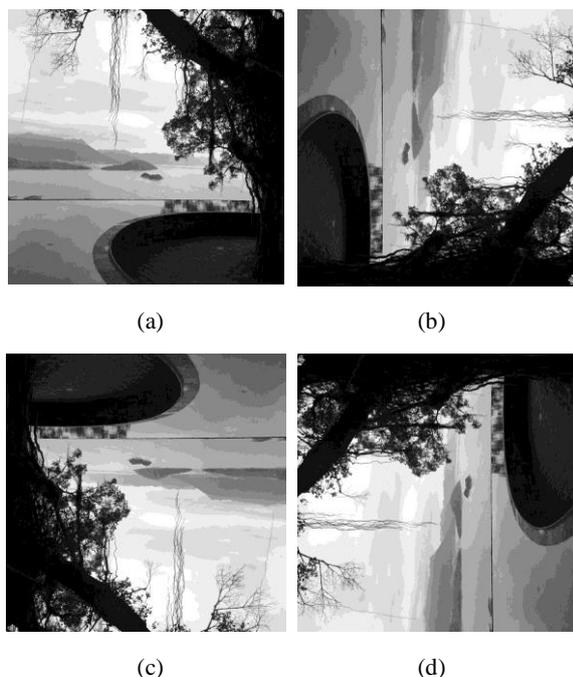


Figure 17 grayscale images with the same composition information (i.e., same Shannon entropy values), but different configurational information.

However, it is noted that Shannon's source coding theorem is always effective when it applies under the circumstance that redundancies between pixels have been well removed or modelled. Nevertheless, we are not likely to completely remove redundancies before using techniques (e.g., Huffman coding, Shannon coding) that consider compositional information to fulfil the requirement of applying Shannon's theorem. In practical projects, the bandwidth and storage sources are limited. In this respect, we need to know the theoretical maximum compression ratio for a specified image in order to allocate feasible storage sources. Based on the results and analysis described in Section 4, we can infer that image metrics capturing configurational information of image pixels, e.g., Boltzmann entropy (Gao et al. 2017) can be investigated upon whether they can be used to predict lossless compression ratio.

In this study, mosaic images and black-white images are not involved. It is noted that gray-level values are homogeneous within a region of such images. Figure 18 (a) shows an image composed of blocks of pixels. In each block, gray-level values are the same. When we use techniques to compress those images, the calculation of compression ratio is an issue. Moreover, the information content of such images should be quantitatively measured. Of course, Shannon entropy is not suitable for such a task. We need to employ other metrics to measure the compositional and configurational information contained in such images. Regarding the image shown in Figure 18 (b), the structures of pixels are not regular. Issues of measuring the information content of such images are also needed to be addressed. In the meantime, theoretical models for predicting lossless compression ratios of such images are also required.

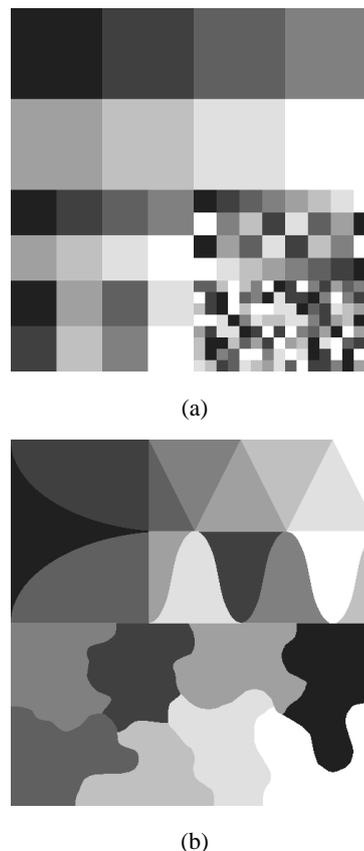


Figure 18 Two images from USC-SIPI-Textures data set (Weber 2004)

## 6. CONCLUSION

This study first introduces the Shannon's source coding theorem and image compression. Thereafter, methodologies and data used for evaluating the performance of Shannon's theorem are presented. Experimental results and analysis show that Shannon's coding theorem is not effective anymore for predicting lossless compression ratio obtained by techniques considering configurational information of pixels. This is true in the context where inter-pixel redundancies are not been removed or well modelled. Shannon entropy is calculated with only consideration of the occurrence frequency of individual gray-level. This means that Shannon entropy is blind to the contexts and structures of image pixels, whereas they are very important to help achieve a high compression ratio. At this sense, new empirical (or theoretical) models based on image metrics capturing configurational information can be built to guide the development of compression techniques and to help users choose suitable techniques for compressing images in order to save the storage space.

## ACKNOWLEDGEMENTS

This research was financially supported by a grant (#15221918) from the Research Grant Council of Hong Kong Special Administrative Region, China. Heartfelt thanks are given to anonymous reviewers.

## REFERENCES

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *The annals of Statistics*, 38(5), 2916-2957.

Cheng, G., Han, J., & Lu, X., 2017: Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE*, 105(10), 1865-1883.

Deutsch, P., 1996: DEFLATE compressed data format specification version 1.3.

Gao, P., Zhang, H., & Li, Z., 2017: A hierarchy-based solution to calculate the configurational entropy of landscape gradients. *Landscape ecology*, 32(6), 1133-1146.

Gao, P., Li, Z., & Zhang, H., 2018: Thermodynamics-based evaluation of various improved Shannon entropies for configurational information of gray-level images. *Entropy*, 20(1), 19.

Huffman, D. A., 1952: A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9), 1098-1101.

Larkin, K. G., 2016: Reflections on shannon information: In search of a natural information-entropy for images. *arXiv preprint arXiv:1609.01117*.

Rissanen, J., & Langdon, G. G., 1979: Arithmetic coding. *IBM Journal of research and development*, 23(2), 149-162.

Richter, T., & Ogawa, S., 1999: Information technology - Scalable compression and coding of continuous-tone still images, lossless and near-lossless coding. *International Organization for Standardization-ISO/IEC*, 18477-8.

Shannon, C. E., 1948. A mathematical theory of communication. *Bell system technical journal*, 27(3), 379-423.

Terrell, G. R., & Scott, D. W., 1992: Variable kernel density estimation. *The Annals of Statistics*, 1236-1265.

Tavakoli, N., 1993: Entropy and image compression. *Journal of Visual Communication and Image Representation*, 4(3), 271-278.

Weber, A. G., 2004: The use texture mosaic images. *University of Southern California Viterbi School of Engineering*, <http://sipi.usc.edu/database/>.

Weinberger, M. J., Seroussi, G., & Sapiro, G., 2000: The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS. *IEEE Transactions on Image processing*, 9(8), 1309-1324.

Ziv, J., & Lempel, A., 1977: A universal algorithm for sequential data compression. *IEEE Transactions on information theory*, 23(3), 337-343.