Image Processing and Analysis
Lecture 10、Image Compression

Weiqiang Wang
School of Computer Science and Technology, UCAS

December 20, 2023

Outline

1. Basis of Image Compression
2. Image Compression Model
3. Elements of Information Theory

Basis Concepts

- The problem solved by image compression is to minimize the amount of data required to represent a digital image. The basic principle of reducing data volume is to remove redundant data.
- The term "data compression" refers to reducing the amount of data required to represent a given amount of information. A clear distinction must be made between data and information.
- Data redundancy is the main issue in digital image compression. It is an entity that can be quantified mathematically.

Data Redundancy

- If n_1 and n_2 represent the number of information units carried in two data sets representing the same information, then the Relative Data Redundancy R_D for the first data set (represented by n_1) can be defined as

$$R_D = 1 - \frac{1}{C_R}$$

$$C_R = \frac{n_1}{n_2}$$

- C_R is referred to as the Compression Ratio. When $n_1 = n_2, C_R = 1$, $R_D = 0$, the first expression of information relative to the second data set contains no redundant data. When $n_2 \ll n_1$, it implies significant compression and high redundancy. When $n_2 \gg n_1$, it indicates that the amount of data contained in the second set greatly exceeds the data volume of the original expression.

Data Redundancy

- Coding Redundancy
- Interpixel Redundancy
- Psychovisual Redundancy

Data Redundancy -- Coding Redundancy

- Here, once again, we assume that a discrete random variable r_k within the interval [0,1] represents the grayscale level of an image and the probability of occurrence for each grayscale level is:

$$p_r(r_k) = \frac{n_k}{n}, k = 0,1,\ldots,L-1$$

- If $l(r_k)$ represents the number of bits used to express each r_k, then the average number of bits required to represent each pixel is:

$$L_{avg} = \sum_{k=0}^{L-1} l(r_k) p_r(r_k)$$

- The number of bits required to encode an image of size M×N is MNL_{avg}

Data Redundancy -- Coding Redundancy

r_k	$p_r(r_k)$	Code 1	$l_1(r_k)$	Code 2	$l_2(r_k)$
$r_0 = 0$	0.19	000	3	11	2
$r_1 = 1/7$	0.25	001	3	01	2
$r_2 = 2/7$	0.21	010	3	10	2
$r_3 = 3/7$	0.16	011	3	001	3
$r_4 = 4/7$	0.08	100	3	0001	4
$r_5 = 5/7$	0.06	101	3	00001	5
$r_6 = 6/7$	0.03	110	3	000001	6
$r_7 = 1$	0.02	111	3	000000	6

TABLE 8.1 Example of variable-length coding.

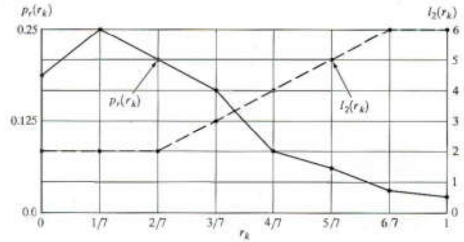
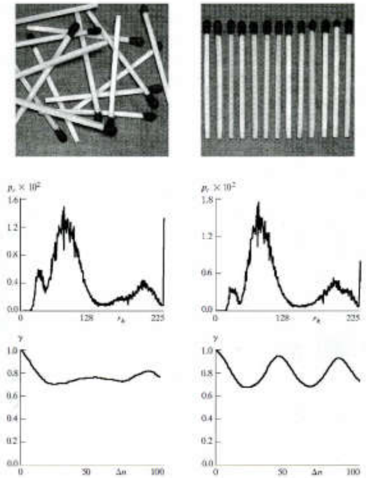


FIGURE 8.1 Graphic representation of the fundamental basis of data compression through variable-length coding.

Data Redundancy -- Interpixel Redundancy



a b
c d
e f

FIGURE 8.2 Two images and their gray-level histograms and normalized autocorrelation coefficients along one line.

Data Redundancy -- Interpixel Redundancy

- Figures 8.2(e) and (f) show the respective autocorrelation coefficients computed along one line of each image. These coefficients were computed using a normalized version of Eq. (4.6-30) in which

$$\gamma(\Delta n) = \frac{A(\Delta n)}{A(0)}$$

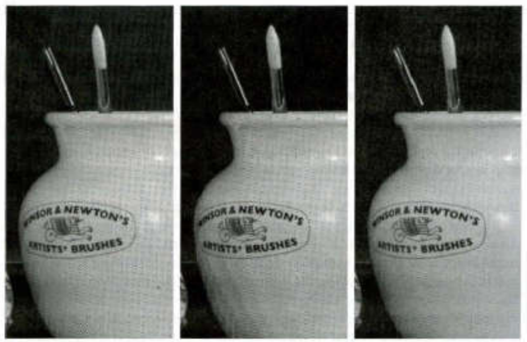
where

$$A(\Delta n) = \frac{1}{N-\Delta n}\sum_{y=0}^{N-1-\Delta n} f(x,y)f(x,y+\Delta n)$$

- A variety of names, including spatial redundancy, geometric redundancy, and interframe redundancy, have been coined to refer to these interpixel dependencies. We use the term interpixel redundancy to encompass them all.

Data Redundancy -- Psychovisual Redundancy

a b c
FIGURE 8.4 (a) Original image. (b) Uniform quantization to 16 levels. (c) IGS quantization to 16 levels.



Fidelity Criteria

- Removal of psychovisually redundant data results in a loss of real or quantitative visual information.
- Two general classes of criteria are used as the basis for such an assessment:
  (1) objective fidelity criteria
  (2) subjective fidelity criteria

Fidelity Criteria

- An example of objective fidelity criteria:

For any value of x and y, the error $e(x,y)$ between $f(x,y)$ and $\hat{f}(x,y)$ can be defined as

$$e(x,y) = \hat{f}(x,y) - f(x,y)$$

so that the total error between the two images is

$$\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]$$

The root-mean-square error e_{rms} between $f(x,y)$ and $\hat{f}(x,y)$ then is the square root of the squared error

$$e_{rms} = \left[\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[\hat{f}(x,y) - f(x,y)]^2\right]^{1/2}$$

Outline

Image Compression Model

- A compression system consists of two distinct structural blocks: an encoder and a decoder.
- In general, $\hat{f}(x, y)$ may or may not be an exact replica of $f(x, y)$. If it is, the system is error free or information preserving; if not, some level of distortion is present in the reconstructed image.
- Both the encoder and decoder shown in Fig. 8.5 consist of two relatively independent functions or subblocks: a source/channel encoder/decoder.

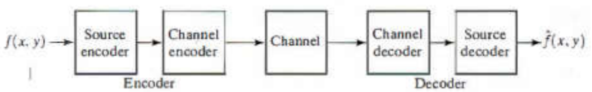
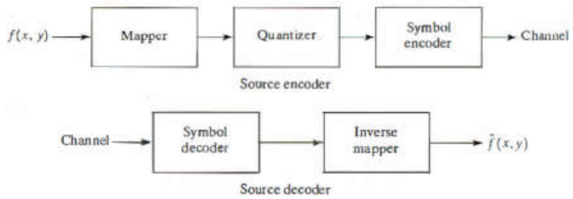


FIGURE 8.5 A general compression system model.

The Source Encoder and Decoder

- The source encoder is responsible for reducing or eliminating any coding, interpixel, or psychovisual redundancies in the input image.
- Normally, the source encoder can be modeled by a series of three independent operations: mapper, quantizer and symbol encoder.
- The source decoder contains only two components: symbol decoder and an inverse mapper.



The Channel Encoder and Decoder

- The channel encoder and decoder are designed to reduce the impact of channel noise by inserting a controlled form of redundancy into the source encoded data.
- As the output of the source encoder contains little redundancy, it would be highly sensitive to transmission noise without the addition of this "controlled redundancy."

The Channel Encoder and Decoder

- Hamming encoder and decoder

  It is based on appending enough bits to the data being encoded to ensure that some minimum number of bits must change between valid code words.

- 7-bit Hamming (7,4) code word

$$h_1 = b_3 \oplus b_2 \oplus b_0 \qquad h_3 = b_3$$
$$h_2 = b_3 \oplus b_1 \oplus b_0 \qquad h_5 = b_2$$
$$h_4 = b_2 \oplus b_1 \oplus b_0 \qquad h_6 = b_1$$
$$h_7 = b_0$$

- nonzero parity word

$$c_1 = h_1 \oplus h_3 \oplus h_5 \oplus h_7$$
$$c_2 = h_2 \oplus h_3 \oplus h_6 \oplus h_7$$
$$c_4 = h_4 \oplus h_5 \oplus h_6 \oplus h_7$$

Measuring Information

- The fundamental premise of information theory is that the generation of information can be modeled as a probabilistic process that can be measured in a manner that agrees with intuition.
- A random event E that occurs with probability $P(E)$ is said to contain $I(E)$ units of information.

$$I(E) = \log \frac{1}{P(E)} = -\log P(E)$$

The quantity $I(E)$ often is called the self-information of E.

Outline

Measuring Information(CONT.)

- The base of the logarithm in Eq. (1) determines the unit used to measure information. If the base m logarithm is used, the measurement is said to be in $m-ary$ units.

$$I(E) = \log \frac{1}{P(E)} = -\log P(E) \qquad (1)$$

- If the base 2 is selected, the resulting unit of information is called a bit.

The Information Channel

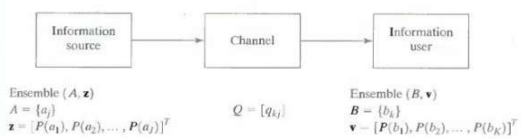


FIGURE 8.7 A simple information system.

- Assume that the information source in Fig generates a random sequence of symbols from a finite or countably infinite set of possible symbols. The set of source symbols $\{a_1, a_2, \ldots, a_J\}$ is referred to as the source alphabet A, and the elements of the set are called symbols or letters.

The Information Channel(CONT.)

- The probability of the event that the source will produce symbol a_j is $P\,(a_j)$

$$\sum_{j=1}^{J} P(a_j) = 1$$

- A $J \times 1$ vector $z = [P\,(a_1), P\,(a_2), \ldots, P\,(a_J)]^T$ customarily is used to represent the set of all source symbol probabilities $\{P\,(a_1), P\,(a_2), \ldots, P\,(a_J)\}$. The finite ensemble (A, z) describes the information source completely.

The Information Channel(CONT.)

- The self-information generated by the production of a single source symbol is $I(a_j) = -\log P\,(a_j)$.
- If k source symbols are generated, for a sufficiently large value of k, symbol a_j will (on average) be output $kP\,(a_j)$ times. Thus the average self-information obtained from k outputs is

$$-kP(a_1)\log P(a_1) - kP(a_2)\log P\,(a_2) - \cdots - kP(a_J)\log P(a_J)$$

- or

$$-k\sum_{j=1}^{J} P(a_j) log P(a_j)$$

The Information Channel(CONT.)

- The average information per source output, denoted $H(z)$, is

$$H(z) = -k\sum_{j=1}^{J} P(a_j) log P(a_j)$$

- This quantity is called the uncertainty or entropy of the source. It defines the average amount of information obtained by observing a single source output. As its magnitude increases, more uncertainty and thus more information is associated with the source. If the source symbols are equally probable, the entropy or uncertainty is maximized and the source provides the greatest possible average information per source symbol.

The Information Channel(CONT.)

- Because we modeled the input to the channel in $Fig.8.7$ as a discrete random variable, the information transferred to the output of the channel is also a discrete random variable(B : $\{b_1, b_2, \ldots, b_K\}$).
- The probability of the event that symbol b_k is presented to the information user is $P(b_k)$. The finite ensemble (B, v), where $v = [P(b_1), P(b_2), \ldots, P(b_K)]^T$, describes the channel output completely and thus the information received by the user.
- The probability $P(b_k)$ of a given channel output and the probability distribution of the source z are related by the expression

$$P(b_k) = \sum_{j=1}^{J} P(b_k|a_j) logP(a_j)$$

The Information Channel(CONT.)

- Each b_k has one conditional entropy function, this conditional entropy function, denoted $H(z|b_k)$, can be written as

$$H(z|b_k) = -\sum_{j=1}^{J} P(a_j|b_k) logP(a_j|b_k)$$

- $P(a_j|b_k)$ is the probability that symbol a_j was transmitted by the source, given that the user received b_k. The expected (average) value of this expression over all b_k is

$$H(z|v) = \sum_{k=1}^{K} H(z|b_k)P(b_k)$$

The Information Channel(CONT.)

- $P(b_k|a_j)$is the conditional probability that output symbol b_k is received, given that source symbol a_j was generated.
- The conditional probabilities are arranged in a matrix $K \times J$ matrix Q

$$Q = \begin{pmatrix} P(b_1|a_1) & P(b_1|a_2) & \cdots & P(b_1|a_J) \\ P(b_2|a_1) & P(b_2|a_2) & \cdots & P(b_2|a_J) \\ \vdots & \vdots & \ddots & \vdots \\ P(b_K|a_1) & P(b_K|a_2) & \cdots & P(b_K|a_J) \end{pmatrix}$$

- The probability distribution of the complete output alphabet can be computed from

$$v = Qz$$

- Matrix Q, with elements $q_{kj} = P(b_k|a_j)$, is referred to as the forward channel transition matrix or by the abbreviated term channel matrix.

The Information Channel(CONT.)

- After substitution and some minor rearrangement $H(z|v)$ can be written as

$$H(z|v) = -\sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j, b_k) logP(a_j|b_k)$$

- $P(a_j, b_k)$ is the joint probability of a_j and b_k.That is, $P(a_j, b_k)$ is the probability that a_j is transmitted an b_k is received.
- The term $H(z|v)$ is called the equivocation of z with respect to v. It represents the average information of one source symbol, assuming observation of the output symbol that resulted from its generation.

The Information Channel(CONT.)

- Because $H(z)$ is the average information of one source symbol, assuming no knowledge of the resulting output symbol, the difference between $H(z)$ and $H(z|v)$ is the average information received upon observing a single output symbol. This difference, denoted $I(z,v)$ and called the mutual information of z and v, is

$$I(z,v) = H(z) - H(z|v)$$

- Substituting for $H(z)$ and $H(z|v)$, and recalling that $P(a_j) = P(a_j,b_1) + P(a_j,b_2) + \cdots + P(a_j,b_K)$, yields

$$I(z,v) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j,b_k) log \frac{P(a_j,b_k)}{P(a_j)P(b_k)}$$

- After further manipulation, can be written as

$$I(z,v) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(a_j) q_{kj} log \frac{q_{kj}}{\sum_{i=1}^{J} P(a_i) q_{kj}}$$

The Information Channel(CONT.)

- The average information received upon observing a single output of the information channel is a function of the input or source symbol probability vector z and channel matrix Q.
- The minimum possible value of $I(z,v)$ is zero and occurs when the input and output symbols are statistically independent, in which case $P(a_j,b_k) = P(a_j)P(b_k)$.
- The maximum value of $I(z,v)$ over all possible choices of source probabilities in vector z is the capacity, C, of the channel described by channel matrix Q.That is,

$$C = \max_{z}[I(z,v)]$$

- The capacity of the channel defines the maximum rate at which information can be transmitted reliably through the channel. Moreover, the capacity of a channel does not depend on the input probabilities of the source but is a function of the conditional probabilities defining the channel alone.