

**Question:** What are different measures of performance of data compression technique?

Similar questions Define the following terms:

- (i) Compression ratio
- (ii) Distortion
- (iii) Compression rate
- (iv) Fidelity and Quality
- (v) Self Information

1. **Compression ratio:** It is very logical way of measuring how well a compression algorithm compresses a given set of data is to look at the ratio of the number of bits required to represent the data before compression to the number of bits required to represent the data after compression. This ratio is called 'Compression ratio'. Ex. Suppose storing an image requires 65536 bytes, this image is compressed and the compressed version requires 16384 bytes. So the compression ratio is 4:1. It can be also represented in terms of reduction in the amount of data required as a percentage i.e 75%
2. **Distortion:** In order to determine the efficiency of a compression algorithm, we have to have the same way of quantifying the difference. The difference between the original and the reconstruction is called as 'Distortion'. Lossy techniques are generally used for the compression of data that originate as analog signals, such as speech and video. In compression of speech and video, the final arbiter of quality is human. Because human responses are difficult to model mathematically, many approximate measures of distortion are used to determine the quality of the reconstructed waveforms.
3. **Compression rate:** It is the average number of bits required to represent a single sample. Ex. In the case of the compressed image if we assume 8 bits per byte (or pixel) the average number of bits per pixel in the compressed

representation is 2. Thus we would say that the compression rate is 2 bits/pixel.

4. **Fidelity and Quality:** The difference between the reconstruction and the original are fidelity and quality. When we say that the fidelity or quality of a reconstruction is high, we mean that the difference between the reconstruction and the original is small. Whether the difference is a mathematical or a perceptual difference should be evident from the context.
5. **Self Information:** Shannon defined a quantity called Self – Information. Suppose we have an event A, which is a set of outcomes of some random experiment. If  $P(A)$  is the probability that event A will occur, then the self-information associated with A is given by:

$$i(A) = -\log_b P(A) \dots \dots \dots (1)$$

If the probability of an event is low, the amount of self-information associated with it is high. If the probability of an event is high, the information associated with it is low.

Ex. The barking of a dog during a burglary is a high probability event and therefore does not contain too much information theory. However if the dog did not bark during a burglary, this is a low-probability event and contains a lot of information. The information obtained from the occurrence of 2 independent events is the sum of the information obtained from the occurrence of individual events. Suppose A and B are 2 independent events. The self-information associated with the occurrence of both event A and event B is, by equation (1)

$$i(AB) = -\log_b$$

as A and B are independent.

$$P(AB) = P(A) \cdot P(B)$$

$$\text{And } i(AB) = -\log_b = -\log_b P(A) - \log_b P(B) = i(A) + i(B)$$

$$i(AB) = i(A) + i(B)$$