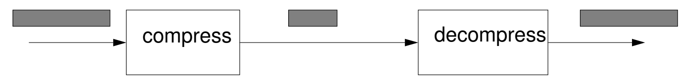
CSE and EE 461
Lecture 25

Autumn Quarter 1997

Data Compression

- Compression - reducing the amount of data for
  - faster data transmission
  - reduced storage space
  - faster computation
- Compression ratio = compressed length / uncompressed length
- Lossless compression - no data is lost, the decompression method completely recovers the data
- Lossy compression - data is lost, but the remaining data is still coherent
  - compression ratio vs. distortion

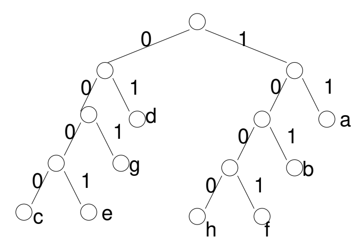


Lossless Compression

- Sources
  - text
  - images
- Morse code and Braille 1800's
- Huffman Coding
- Dictionary Techniques - Lempel-Ziv
  - unix compress, zip
  - GIF
- Arithmetic Coding, sometimes called entropy coding
- Run-length Coding
- Predictive Coding
  - lossless JPEG

Huffman Coding

a	.25	11
b	.2	101
c	.05	0000
d	.2	01
e	.05	0001
f	.1	1001
g	.1	001
h	.05	1000

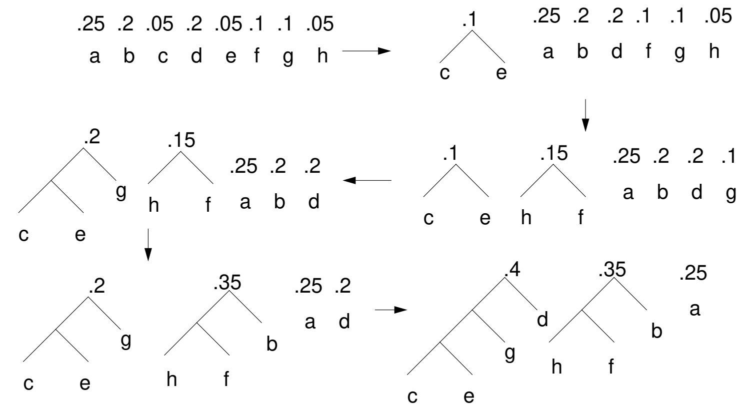


- abaabdfgae = 30 bits with 3 bits per character
- 111011111101011001001110001 = 27 bits

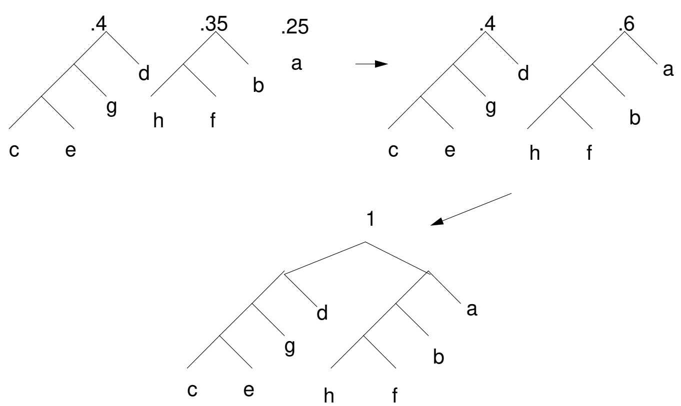
Huffman Coding

- Prefix code - no codeword is a prefix of another
- Variable length code
- Optimal prefix code if all that is known is the frequencies of the base symbols
- Algorithm to construct code is simple
  - Maintain a forest of symbol groups each with a given probability
  - Merge the two symbol groups with the smallest probabilities into a new subtree.

Huffman Example



Huffman Example (Continued)

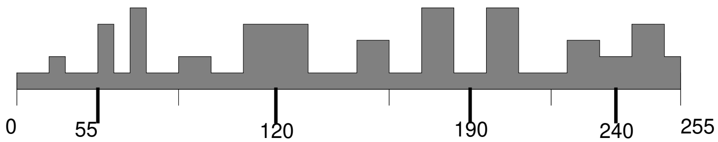


Lossy Compression

- Subsampling
- Scalar Quantization
- Vector Quantization
- Filter methods - transform the data, keep only significant coefficients and quantize those.
  - Cosine transform
  - Fourier transform
  - Wavelet transform

Scalar Quantization

- Given data in the 0 to 255 range (gray scale) find 4 levels of grey that has the least distortion. Compression ratio is 2/8 = .25.



Other Compression Issues

- Differential predictive coding - To code y, first predict x, then compress y-x. If x is actually close to y then y-x will compress well.
- Progressive transmission - some bits give low fidelity, adding more bits yields higher fidelity.
- Sound, images, video - how do the methods differ?
- Real time vs. not real time
- Negotiating the compression ratio - distortion trade-off between sender and receiver.
- Standards - MPEG, JPEG and others
- Patents