ECE625
Data Compression - Overview

Chang-Su Kim

Course Information

- Course homepage
  - http://portal.korea.ac.kr
- Lecturer
  - Chang-Su Kim
  - Office: Engineering Bldg, Rm 508
  - E-mail: changsukim@korea.ac.kr
- Tutor
  - 신녕호 (nhshin@mcl.korea.ac.kr)

Course Information

- Objective
  - Study theory and practice of data compression
- Textbooks
  - Elements of Information Theory by Cover and Thomas, 2nd edition, Wiley, 2006
  - Introduction to Data Compression, by Saywood, Morgan Kaufmann, 2000
  - Vector Quantization and Signal Compression by Gersho and Gray, Kluwer Academic Publishers, 1991
  - Fundamentals of Digital Image Processing by Jain, Prentice-Hall, 1989

Course Schedule

Week	Contents	Remarks
1	Overview	
2	Information Theory	
3	Information Theory	
4	Information Theory	
5	Information Theory	
6	Huffman Coding, Arithmetic Coding	
7	Dictionary Coding, Run-Length Coding	
8		Mid Exam
9	Vector Quantization	
10	Vector Quantization	
11	Transform Coding	
12	Transform Coding	
13	Paper Presentation	
14	Paper Presentation	
15	Paper Presentation	
16		Final Exam

Paper Presentation

- 1-Page Proposal (due 28 Sept) (20%)
  - Individual project
  - Topic
    - Select a recent paper (after 2015)
      - ICLR, Neurips, CVPR, ICCV, ECCV
      - IEEE TIP, TMM, TCST
      - IEEE TPAMI, IJCV
      - ETC
    - Should be related with data compression or signal processing
  - I have the right to reject the proposal and recommend a new topic
- Presentation (30%)
  - 20 min presentation + 5 min Q&A
- Report in English (50%)
  - 1-column 2-page report
  - Evaluation of compactness, clarity
  - Direct copying of sentences in paper will reduce the mark

Course Information

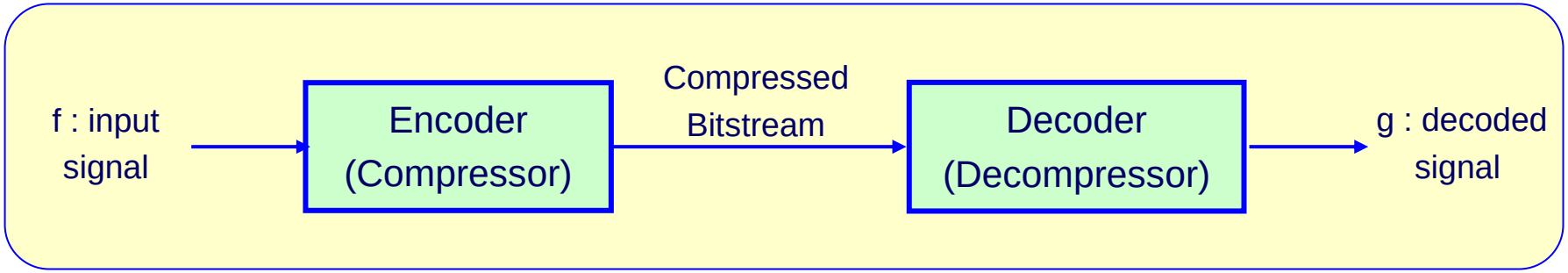
- Prerequisites
  - Probability
  - Signals and Systems
  - Discrete Signal Processing

- Assessment

Exercises	20 %
Midterm Exam	20 %
Final Exam	30 %
Paper Presentation	30 %

Data Compression

- The objective is to reduce the amount of data required to represent an input signal
- Data compression system



- Lossless compression vs. lossy compression
  - Lossless compression (f = g)
    - Low compression ratio
    - Medical imaging, satellite imaging: information loss is not tolerable
  - Lossy compression (f ≠ g)
    - High compression ratio
    - DVD: Human visual system can tolerate a modest loss of information
    - Tradeoff between compression ratio and reconstructed image quality

Compression Ratio

- Suppose that an image of size 300x200 is compressed into a 6 Kbyte file
  - File size before compression (n_1)
    - 300x200 = 60000 byte = 60 Kbyte
  - File size after compression (n_2)
    - 6 Kbyte
  - Compression ratio (C_R)
    - $C_R = n_1/n_2 = 60/6 = 10$

Entropy

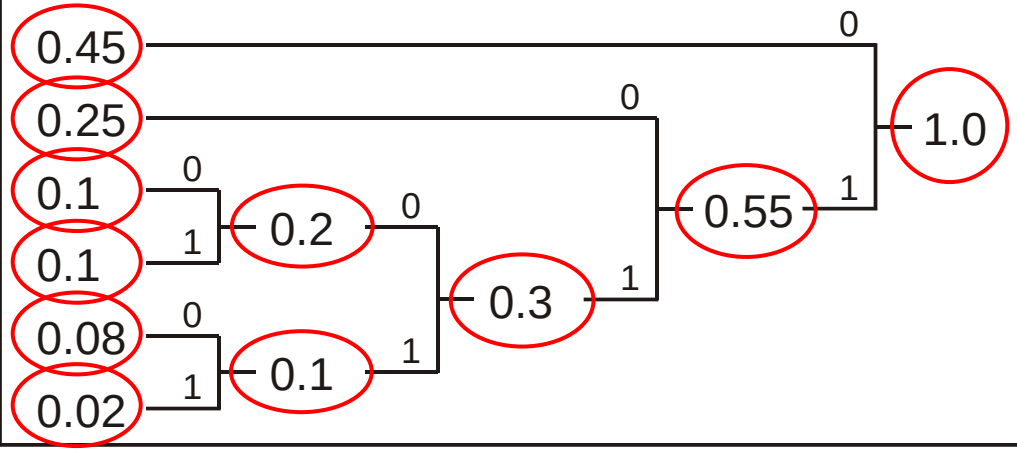
- A random data source *X*, consisting of *n* symbols
- The entropy *H*(*X*) is the average uncertainty of the symbols

$$H(X) = \sum_{i=1}^{n} p_i \log_2 \frac{1}{p_i}$$

- *H*(*X*) is the fundamental lower bound for the average length that any code can achieve
  - At least *H*(*X*) bits/sample are required to encode a data source *X* losslessly, no matter what compression scheme is employed

Huffman Coding

- Construction method for an optimal prefix code

Probability	Huffman Tree	Codeword
0.45		0
0.25		10
0.1		1100
0.1		1101
0.08		1110
0.02		1111

Arithmetic Coding

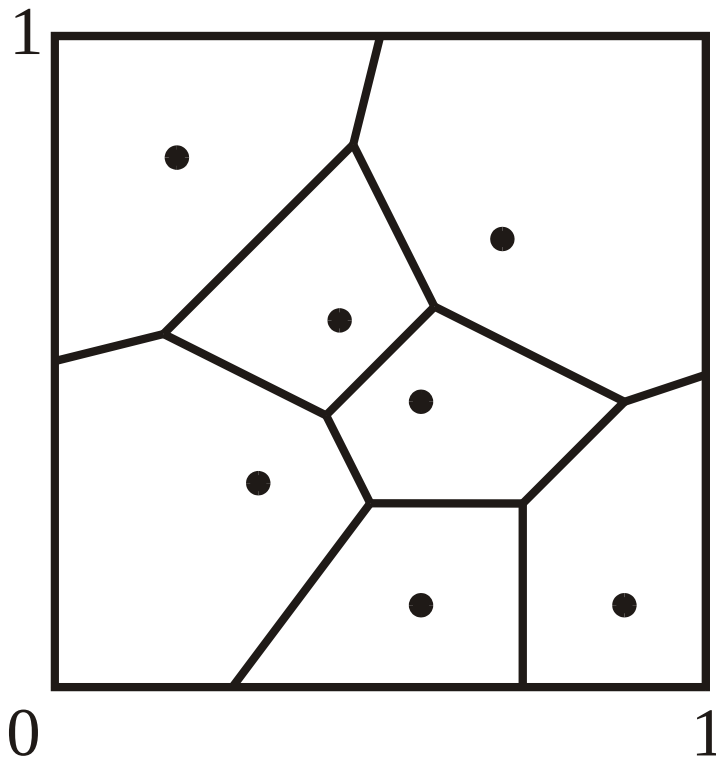
- Hard to explain in one slide but easy to use

Dictionary Coding

- Dictionary coding replaces blocks of consecutive samples with indices into some dictionary
  - The dictionary contain frequently used blocks
  - Effective only if source contains frequently occurring patterns
  - Removal of intersample redundancy
    - Ko*ea U*iv**sity
    - We have our dictionaries
- Adaptive dictionary coding
  - The dictionary is built adaptively from the source itself
  - Universal coding
  - LZ77, LZ78 – ZIP and ARJ
  - LZW
    - Unix command ‘compress’
    - V.42 bis standard for data compression over modems
    - Graphics interchange format (GIF) to compress graphical images

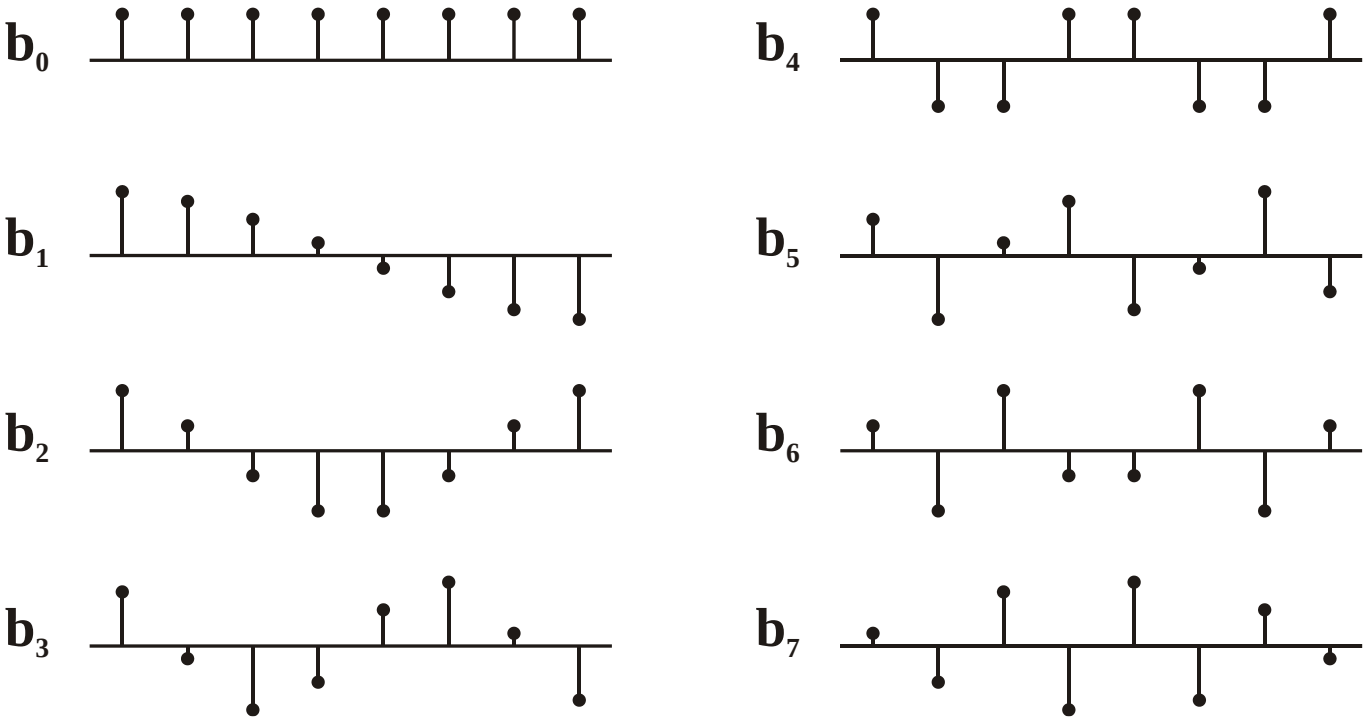
Vector Quantization

An example of a 2D vector quantizer.



Transform Coding

Basis vectors for the 8-point DCT.



Subband and Wavelet Coding