Vector Quantization and Subband Coding

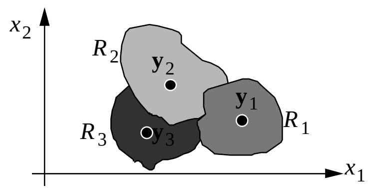
18-796 Multimedia Communications:
Coding, Systems, and Networking

Prof. Tsuhan Chen
tsuhan@ece.cmu.edu

Vector Quantization

Vector Quantization (VQ)

- Each image block has *N* pels
- Consider each image block as an *N*-D vector **x**



- Quantization: $\mathbf{x} \rightarrow \mathbf{y}_k$ if $\mathbf{x} \in R_k$
- $\boldsymbol{y}_\mathrm{k}$: codewords or code vectors
- The set of $\boldsymbol{y}_\mathrm{k}$ is called a codebook

Rate and Distortion

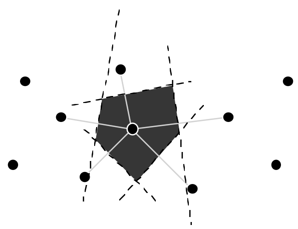
- If the number of codewords is *K*, then the number of bits required to send one vector is $\log_2 K$
- Rate *R*
  - *R* bits per pixel
  - *NR* bits for one vector, so $\log_2 K = NR$, i.e., log $K = 2^{NR}$
- Distortion *D*
  - Given the probability density function $p(\mathbf{x})$ and distortion measure $d(\mathbf{x},\mathbf{y})$, the average distortion is

$$D = \sum_{k=1}^{K} \int_{R_k} d(\mathbf{x}, \mathbf{y}_k) p(\mathbf{x}) d\mathbf{x}$$

- Given $\mathbf{y}_k$, $\boldsymbol{R}_k$ should be chosen such that

$$R_k = \{\mathbf{x}: d(\mathbf{x}, \mathbf{y}_k) \le d(\mathbf{x}, \mathbf{y}_j)\ \forall j \ne k\}$$
$$= \text{the set of } \mathbf{x} \text{ for which } \mathbf{y}_k \text{ is the nearest point}$$

- For L_2 norm, i.e., $d(\mathbf{x}, \mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2$ we get



Convex Polytope
(Voroni Cell)

$\text{Q}:$ How about L_1 or L_∞ ?

- Given $\boldsymbol{R}_k$, $\mathbf{y}_k$ should be chosen such that

$$\int_{R_k} d(\mathbf{x}, \mathbf{y}_k) p(\mathbf{x}) d\mathbf{x} \text{ is minimum}$$

- With L_2 norm, we get

$$\mathbf{y}_k = \text{centroid of } R_k = \int_{R_k} \mathbf{x} p(\mathbf{x}) d\mathbf{x}$$

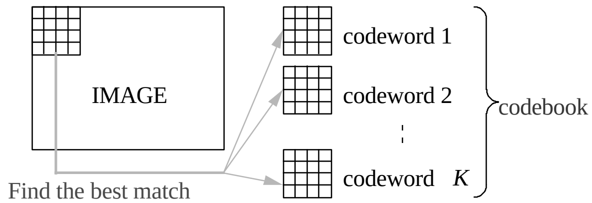
- In the discrete case, optimal $\mathbf{y}_k$ is the average of the vectors in $\boldsymbol{R}_k$

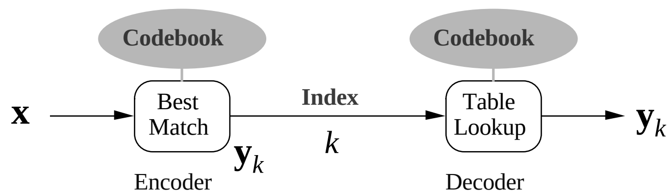
Generalized Lloyd Algorithm (LBG Algorithm, K-means Algorithm)

- Linde, Buzo, and Gray, 1980
- Given $p(\mathbf{x})$, or given a set of training vectors
  (1) Start with an initial set of $\mathbf{y}_k$, i.e., initial codebook
  (2) With the current $\mathbf{y}_k$, calculate the region R_k
  (3) Replace each $\mathbf{y}_k$ with the centroid of R_k
  (4) If the overall distortion D is lower than a threshold, stop. Otherwise, go to (2)
- Only gives local optimum. Proper choice of initial codebook is important

- Choice of initial codebook
  - A representative subset of the training vectors
  - Scalar quantization in each dimension
  - Splitting…
- Nearest Neighbor (NN) algorithm [Equitz, 1984]
  - Start with the entire training set
  - Merge the two vectors that are closest into one vector equal to their mean
  - Repeat until the desired number of vectors is reached, or the distortion exceeds a certain threshold

Image Coding





Properties of VQ

- Codebook design is very complex
  - 4x4 blocks at 1 bpp: 2^{16} codewords
  - 16 images of size 256×256: 2^{16} training vectors (4×4 each)
  - Codebook size: $2^{16} \times 4{\times}4 \times 8$ bits = 8.3 Mbits
- More useful for low bitrate
  - 4x4 blocks at 0.5 bpp: $2^8 = 256$ codewords
  - One 256x256 image: 4096 training vectors
  - Codebook size: $256 \times 4{\times}4 \times 8$ bits = 32.8 Kbits
- Simple decoder, complex encoder
  - Very good for image retrieval
- Poor performance on images not in the training set vs. overhead of sending the codebook

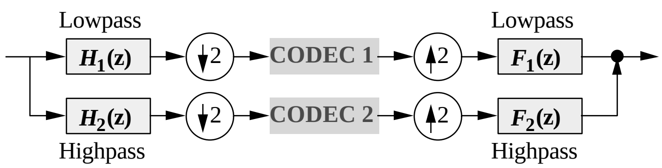
18-796/Spring 1999/Chen

VQ Variants and Improvements

- Multistage VQ
- Product Codes
  - Send mean and variance separately
- Classified VQ
  - Edges, texture areas, flat areas
- Predictive VQ
- VQ for color images
  - Exploit correlation among color components, e.g. R,G,B
  - YUV components are practically uncorrelated

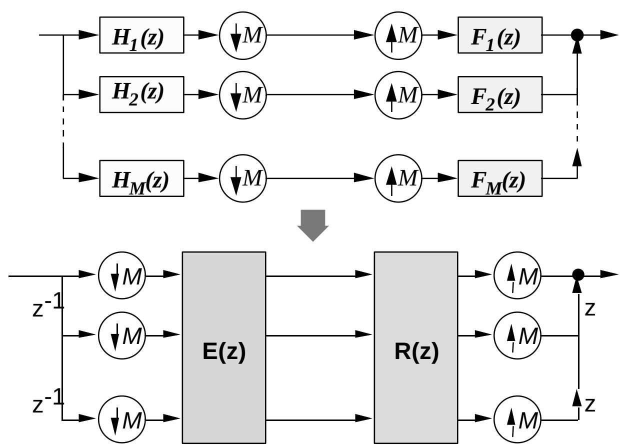
Subband Coding

Subband Coding



- Decompose the signal in the frequency domain
- Critical downsampling (maximal decimation) maintains the number of samples in the subbands
- Wavelet coding: Recursively apply subband decomposition to the low freq band
- 2-D: Separable filtering to get 4 bands: LL, LH, HL, HH

Subband Coding vs. Transform Coding



Polyphase Representation

- Perfect reconstruction (PR) is obtained when **R**(z)=**E**$^{-1}$(z)
- When **E**(z) and **R**(z) are constant matrices, subband coding degenerated to blocked-based operation, i.e., transform coding
- In particular, if **E**(z) is a DCT matrix and **R**(z) is IDCT, this becomes DCT coding
- Subband coding can be viewed as transform coding with overlapped blocks. So, it can exploit correlation of pixels at longer range
- Coding Artifacts:
  - Transform Coding: blocking
  - Subband Coding: ringing, contouring

Optimal Bit Allocation

- We can allocate different bit rates to the subbands based on their properties
- Assume that we apply scalar quantization with bitrate b_k to the subbands x_k , then the quantization error is

$$s_{q_k}^2 = c \times 2^{-2b_k} s_{x_k}^2$$

- The overall quantization error is $s_q^2 = \frac{1}{M}\sum_{k=1}^{M} s_{q_k}^2$
- The overall bitrate is

$$b = \frac{1}{M}\sum_{k=1}^{M} b_k$$

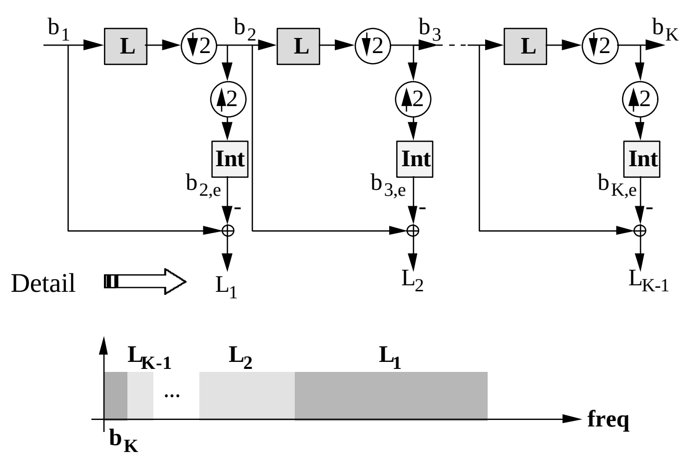
$$s_q^2 \ge \left(\prod_{k=1}^{M} s_{q_k}^2 \right)^{1/M} \quad \text{(AM-GM inequality)}$$

$$= c\left(\prod_{k=1}^{M} 2^{-2b_k} s_{x_k}^2 \right)^{1/M} = c\left(2^{-2\sum b_k / M}\right)\left(\prod_{k=1}^{M} s_{x_k}^2 \right)^{1/M}$$

$$= c \times 2^{-2b} \left(\prod_{k=1}^{M} s_{x_k}^2 \right)^{1/M} \quad \text{(a constant for given signal and filter bank)}$$

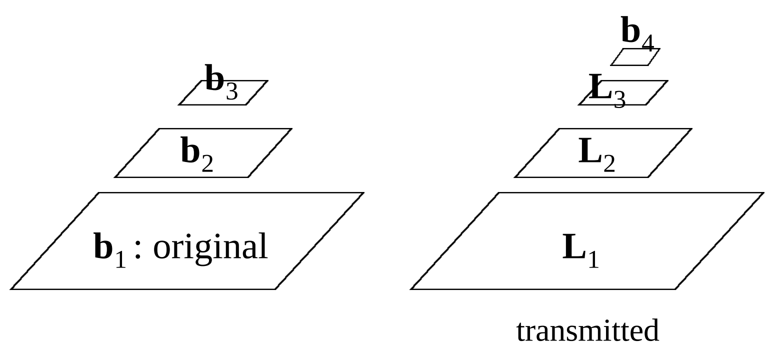
- Equality holds if and only if $s_{q_k}^2 = s_q^2 \quad \forall k$
- Optimal bit allocation $b_k = \frac{1}{2} \log \frac{c \times s_{x_k}^2}{s_q^2}$
- Gain $= \dfrac{\frac{1}{M}\sum_{k=1}^{M} s_{x_k}^2}{\left(\prod_{k=1}^{M} s_{x_k}^2\right)^{1/M}} \ge 1$ No gain if $s_{x_k}^2$ are identical

Pyramid Coding



The "Pyramid"

- Consider the 2-D case



- Number of samples is 33% more

$$N + \frac{N}{4} + \frac{N}{16} + \cdots \approx \frac{4}{3} N$$

- Non-critical sampling
- PR is always possible
  - No matter how L and Int are designed
- Progressive transmission is possible

References

- VQ
  - Allen Gersho, and Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992
- Subband
  - John Woods, ed., *Subband Image Coding*, Kluwer Academic Publishers, 1991
  - P.P. Vaidyanathan, Multirate Systems and Filter Banks, Prentice Hall, 1993
  - N.S. Jayant and Peter Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall, 1984