5. Video coding (Part 1)

Yi-Shin Tung
National Taiwan University

PCM

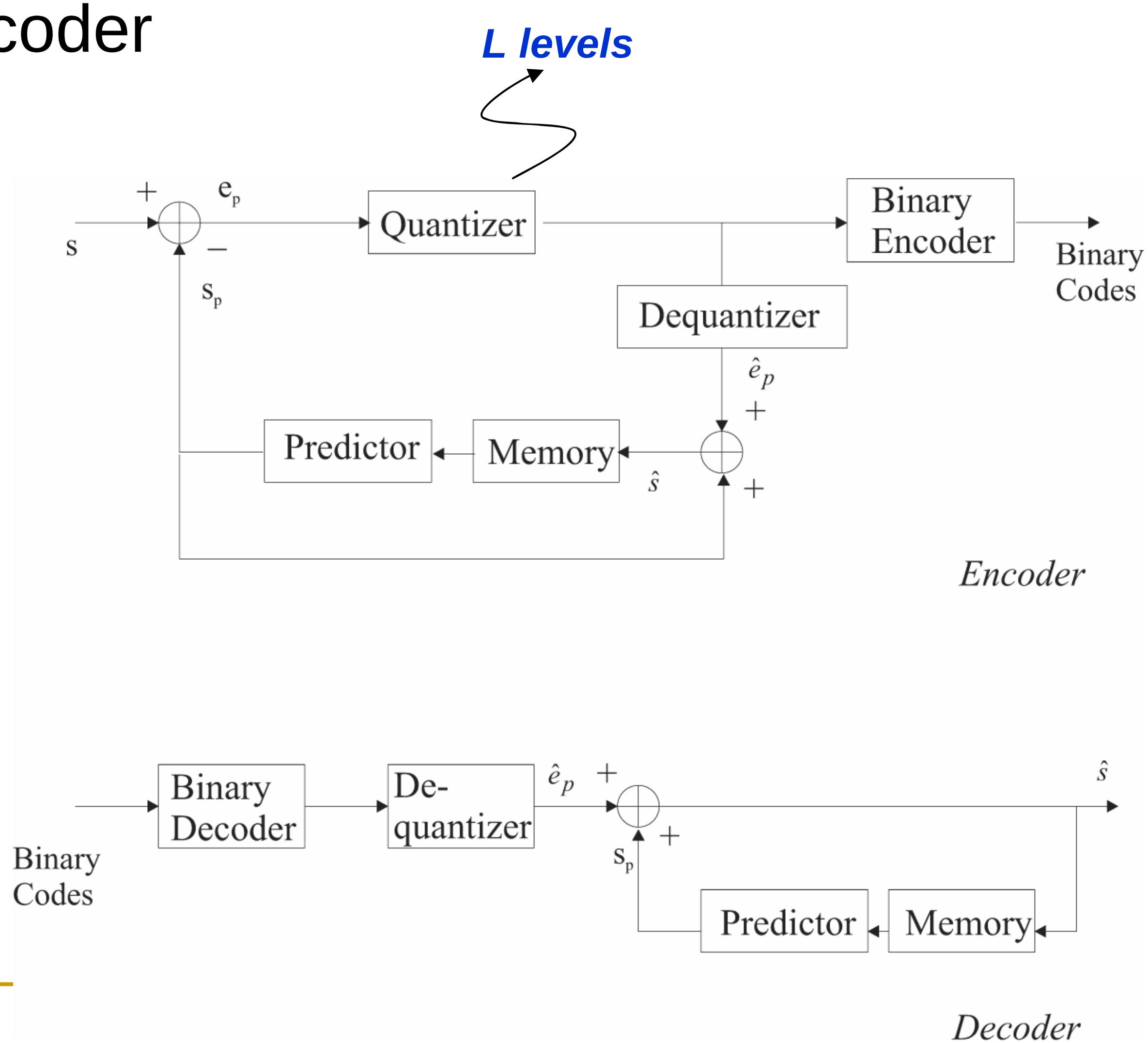
- Sampling: an array of samples on a rectangular spatial grid sampled at a frame rate, $B(x,y,t)$.
- Each sample is quantized to 2^K levels, called *pulse coded modulation*.
  - Time discreteness is provided by sampling the signal at lease at the Nyquist rate. (both spatial and temporal)
  - Amplitude discreteness is provided by using a sufficient number of quantization levels
- PCM does not remove any statistical or perceptual redundancy.

Predictive coding: DPCM

- Prediction next *pel* using values of the “previously transmitted” *pels*, and transmitting quantized prediction error, called *Differential PCM*.
  - Intra-frame prediction
  - Inter-frame prediction
    - The same spatial location in case of no motion
    - A displaced location in case of it is part of motion object.

DPCM

- Simple DPCM coder
  - Predictor
  - Quantizer
  - Entropy coder

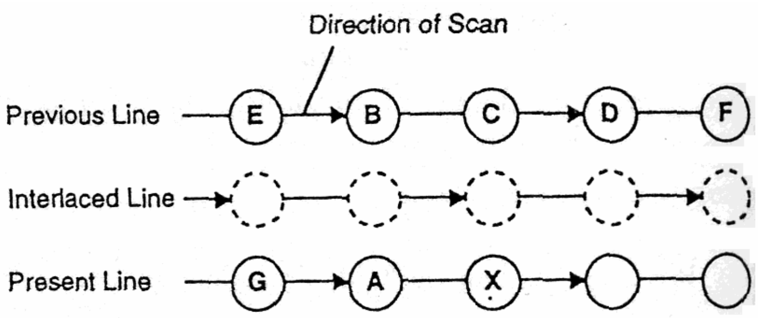


DPCM

- 1D predictor (1 pixel memory)
- 2D predictor (1 line memory)
- Temporal predictor (1 frame/2 fields memory)
- Motion compensated predictor

The cost of memory drops, the distinction becomes blurring.

The cost of computation drops, ME/MC become practical.



1D/2D predictor

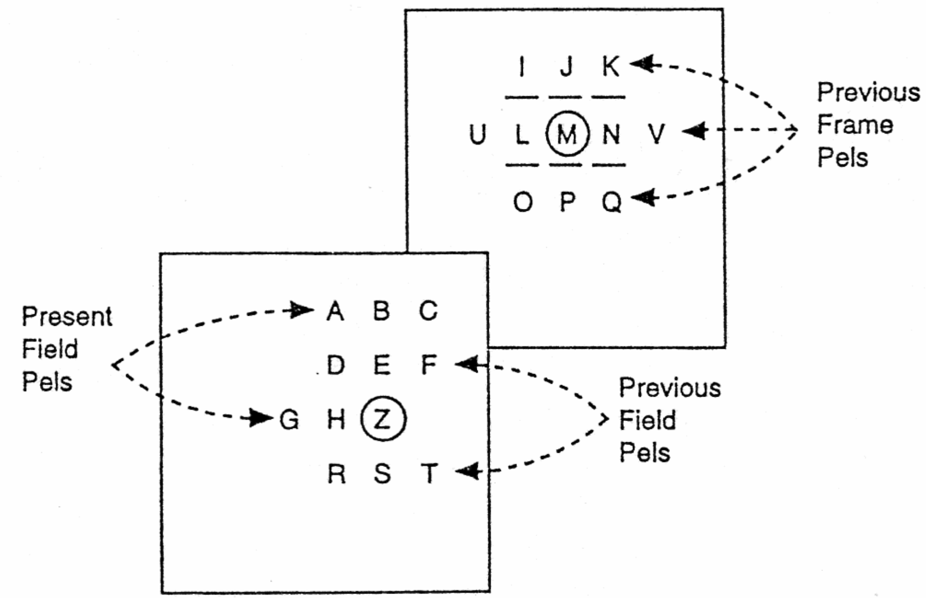
	Prediction Coefficients			Variance
Video Signal	α_A	α_B	α_C	σ_E^2
Y	1	—	—	53.1
	1	−1/2	1/2	29.8
	3/4	−1/2	3/4	27.9
	7/8	−5/8	3.4	26.3
$R–Y$	1	—	—	22.6
	—	—	1	5.6
	1/2	−1/2	1	4.9
	5/8	−1/2	7/8	4.7
$B–Y$	1	—	—	13.3
	—	—	1	3.2
	1/2	−1/2	1	2.5
	3/8	−1/4	7/8	2.5

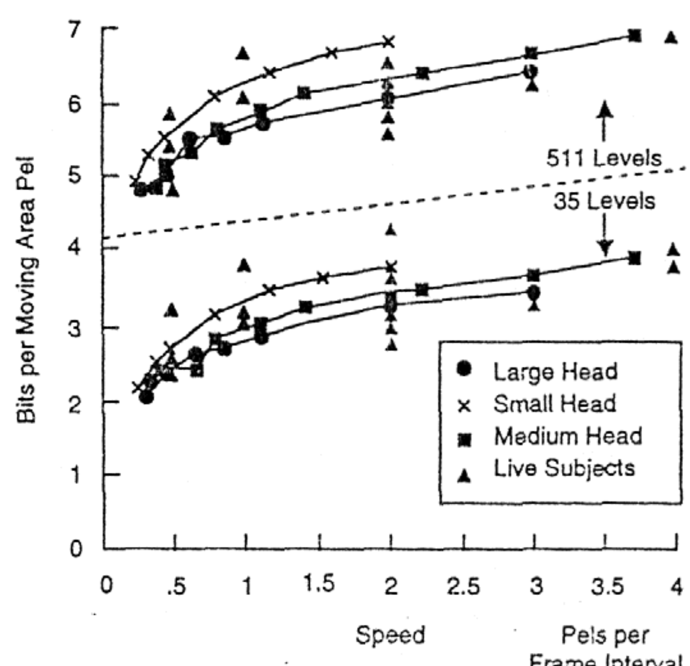
Power of 2 is easy to implement.

Temporal predictor

- Frame difference predictor becomes inefficient when motion is larger than 1~2 pels.
- Motion becomes larger, all temporal predictors fail.

Transmitted Signal $Z - \hat{Z}$	Prediction $\hat{Z}$	Entropies in Bits per Moving-Area Pel (35 Level Quantization)
Frame-difference	M	≈2.1 to 3.9
Element-difference	H	≈2.0 to 3.7
Element-difference of frame-difference	$M + H - L$	≈1.8 to 3.1
Line-difference of frame-difference	$M + B - J$	≈1.5 to 3.5
Field-difference	$(E + S)/2$	≈1.8 to 3.2
Element-difference of field-difference	$H + (E + S)/2 - (D + R)/2$	≈1.5 to 2.5



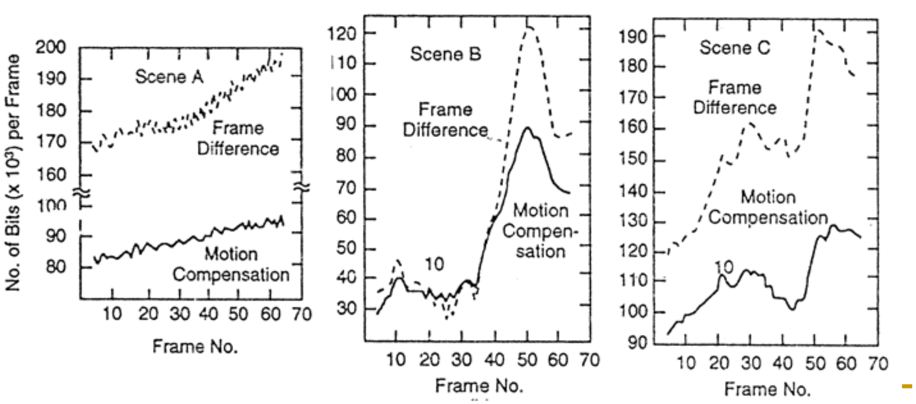


Motion compensated predictor

- If a scene contains moving objects and an estimate of their translation is available, temporal prediction can be adapted, called *MCP*.
- Real motion is complex combination of 3D translation and rotation, but 2D translational motion works well in most cases.
  - Object motion is in a plane perpendicular to the camera axis.
  - Illumination is spatially and temporally uniform.
  - Occluded object and uncovered background are relatively small.

$$b(Z,t) = b(Z - D, t - \tau)$$

b(.): brightness function
Z: spatial position
τ: the time between two frames
D: 2D translational motion



Motion estimation

- Block matching methods: assume motion is constant within $M \times N$ block of pels.

$$PE(D) = \sum_{M \times N} G(b(Z,t) - b(Z - D, t - \tau))$$

  - G(.) is distant metric such as magnitude or square function.
- Exhaustive search

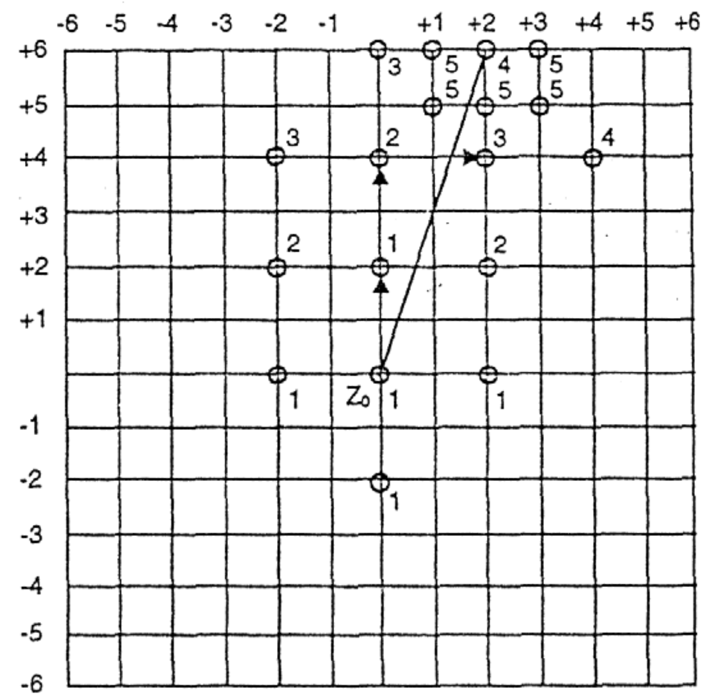
$$PE(Z_0, i, j) = \frac{1}{M \times N} \sum_{|m| \le \frac{M}{2}} \sum_{|n| \le \frac{N}{2}} \left| b(Z_{mn}, t) - b(Z_{m+i,n+j}, t - \tau \right|$$

$$where -d_{\max} \le i, j \le d_{\max}, Z_{mn} = Z_0 + [m, n]'$$

- If assuming *PE(*Z_0*,i,j)* increases monotonically as the shift *(i,j)* moves away from the minimum, fast algorithms apply.

Fast motion estimation

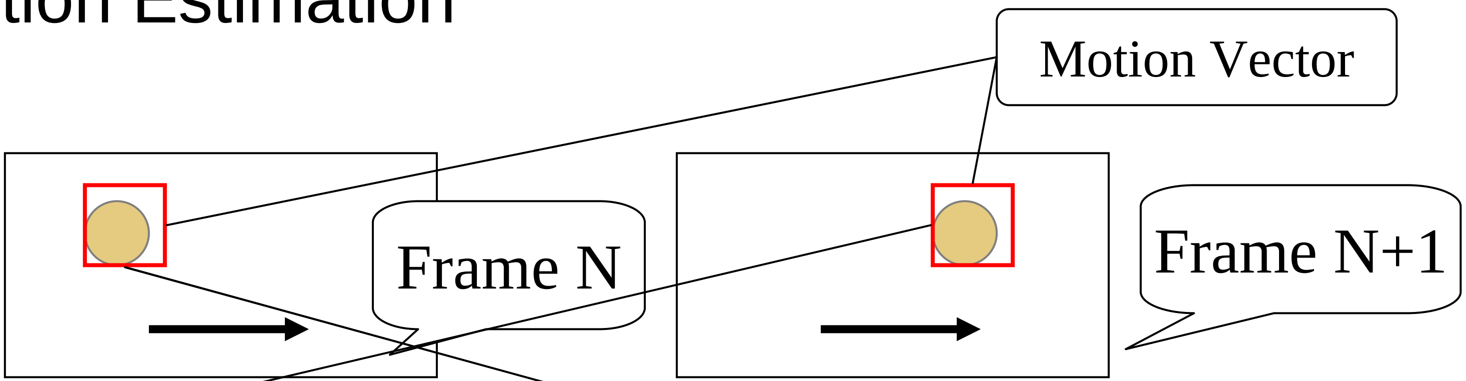
- TSS, logarithmic search
- Fractional pel accuracy
- Bits for motion vector have to be transmitted.
- Reduce the time of cost function evaluation from $(2\times d_{max}+1)^2$ to $n\times log(d_{max})$.
- For 5x5 block size, d_{max}=8, half-pel resolution,
  - Motion rate is 0.4bits/pel.



On Denotes A Search Point of Step n

Motion Estimation

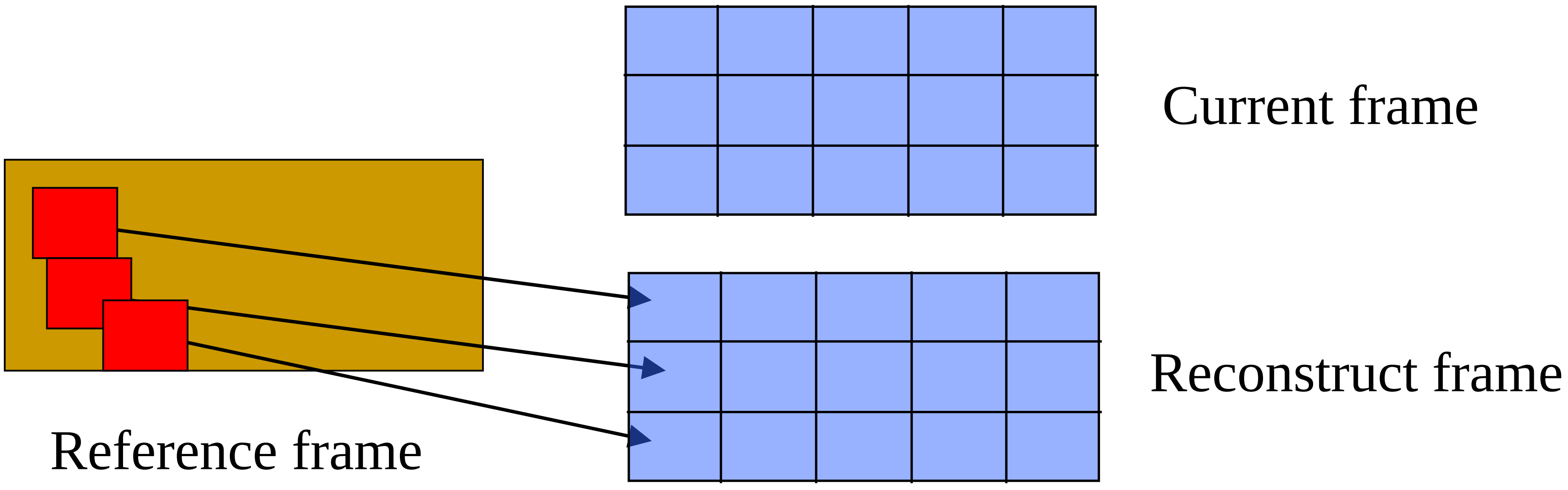
- Motion Estimation



$$
\begin{bmatrix}
88 & 84 & 83 & 84 & 85 & 86 & 83 & 82 \\
86 & 82 & 82 & 83 & 82 & 83 & 83 & 81 \\
82 & 82 & 84 & 87 & 87 & 87 & 81 & 84 \\
81 & 86 & 87 & 89 & 82 & 82 & 84 & 87 \\
81 & 84 & 83 & 87 & 85 & 89 & 80 & 81 \\
81 & 85 & 85 & 86 & 81 & 89 & 81 & 85 \\
82 & 81 & 86 & 83 & 86 & 89 & 81 & 84 \\
88 & 88 & 90 & 84 & 85 & 88 & 88 & 81
\end{bmatrix}
-
\begin{bmatrix}
84 & 82 & 83 & 81 & 85 & 86 & 83 & 81 \\
82 & 82 & 81 & 83 & 82 & 83 & 83 & 81 \\
83 & 82 & 84 & 87 & 87 & 87 & 81 & 88 \\
81 & 85 & 86 & 88 & 82 & 82 & 84 & 87 \\
81 & 84 & 85 & 87 & 85 & 89 & 84 & 81 \\
82 & 85 & 81 & 84 & 81 & 89 & 81 & 83 \\
81 & 87 & 86 & 83 & 86 & 89 & 81 & 84 \\
88 & 82 & 87 & 84 & 87 & 89 & 84 & 81
\end{bmatrix}
=
\begin{bmatrix}
-4 & -2 & 0 & -3 & 0 & 0 & 0 & -1 \\
-4 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\
0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 2 & 0 & 0 & 0 & 4 & 0 \\
1 & 0 & -4 & -2 & 0 & 0 & 0 & -2 \\
-1 & 6 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -6 & -3 & 0 & 2 & 1 & -4 & 0
\end{bmatrix}
$$

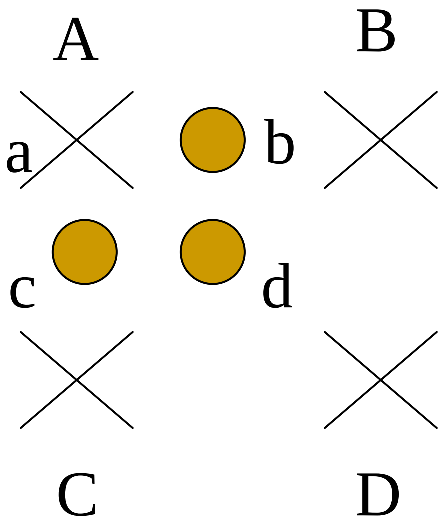
Motion Compensation

- Motion Compensation
  - using motion vector between current frame and reference frame to reconstruct the prediction of current frame



Motion Compensation (conti.)

- Motion vector is coded by DPCM.
- Half-pel motion vector is applicable by using interpolation.



$a = A$

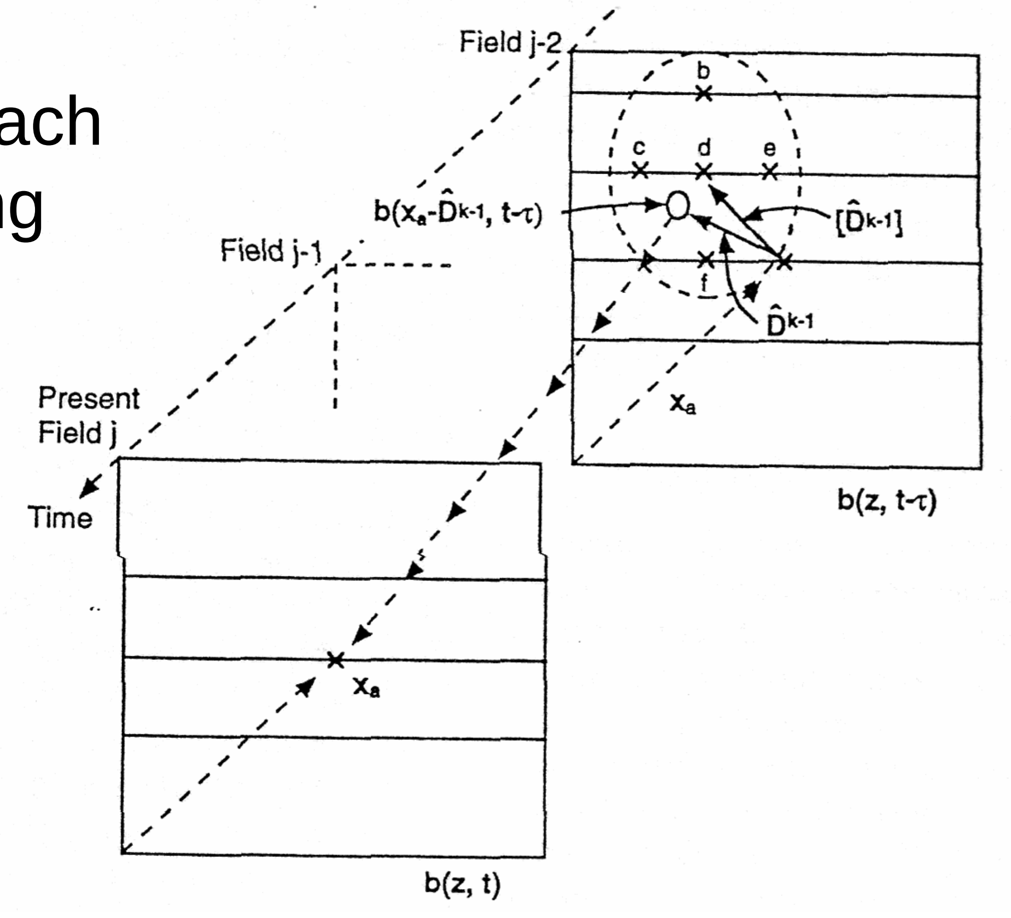
$b = (A+B+1)/2$

$c = (A+C+1)/2$

$d = (A+B+C+D+2)/4$

Motion estimation

- Recursive methods
- Minimize recursively displaced frame distortion (DFD) at each moving area pel using steepest descent algorithm.



Quantization

- DPCM achieve compression by quantizing the prediction error more coarsely than original signal.
- Properly placing quantization levels results in excellent picture quality.
- Quantization artifacts
  - Granular noise
  - Slope overload
  - Edge busyness

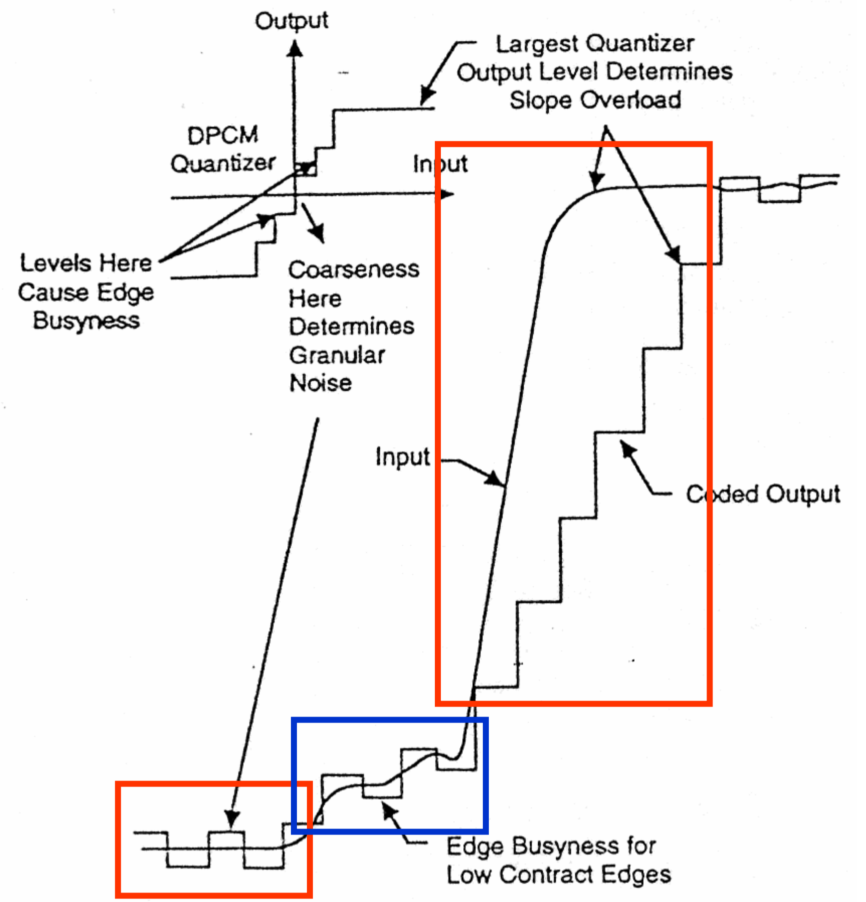


Fig. 6.12 An intuitive classification of distortion due to coarse DPCM quantization. Three classes of quantization noise are identified: granular noise, edge business, and slope overload.

Quantization design

- Statistical design

$$D = \sum_{k=1}^{L} \int_{t_{k-1}}^{t_k} f(x - l_k) \cdot p(x) dx$$

$$where\; t_0 < t_1 < \ldots < t_L \text{ and } l_1 < l_2 < \ldots < l_L$$

- Differentiating with respect to t_k and l_k

$$f(t_{k-1} - l_{k-1}) = f(t_{k-1} - l_k), k = 2 \ldots L$$

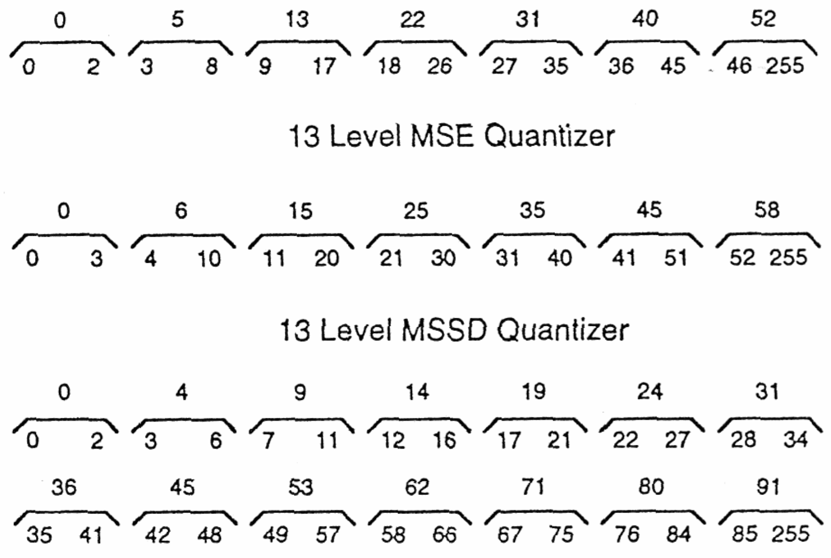
$$\int_{t_{k-1}}^{t_k} \frac{df(x - l_k)}{dx} \cdot p(x) dx = 0, k = 1 \ldots L$$

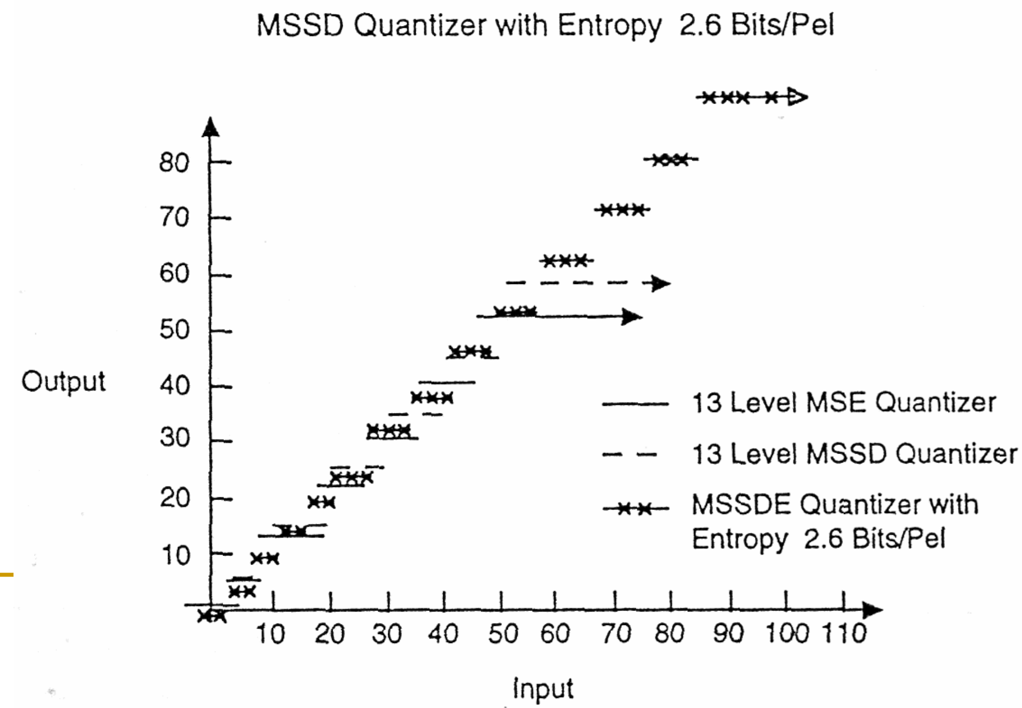
- If *f* is MSE, then…

$$t_k = \frac{1}{2}(l_k + l_{k+1}), l_k = \int_{t_{k-1}}^{t_k} x \cdot p(x) dx \Bigg/ \int_{t_{k-1}}^{t_k} p(x) dx$$

- *Lloyd-Max* iteration process solves the problem.
- *p(x)* for pel-differential coder can be approximated by a Laplacian density.

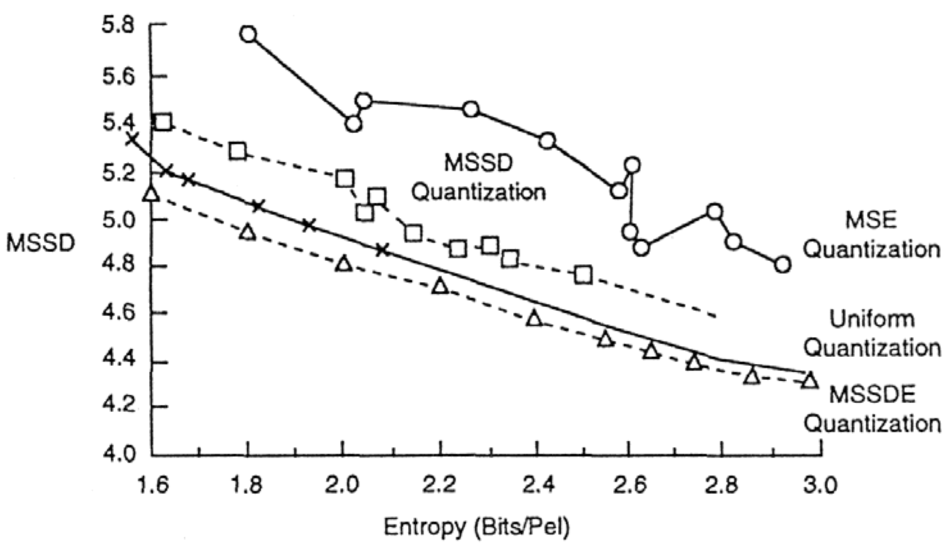
- MSE⇒companded
  - Less granular noise.
  - Frequent slope overload and edge busyness.
- Minimize quant distortion subjective to fixed entropy
  - Uniform quant

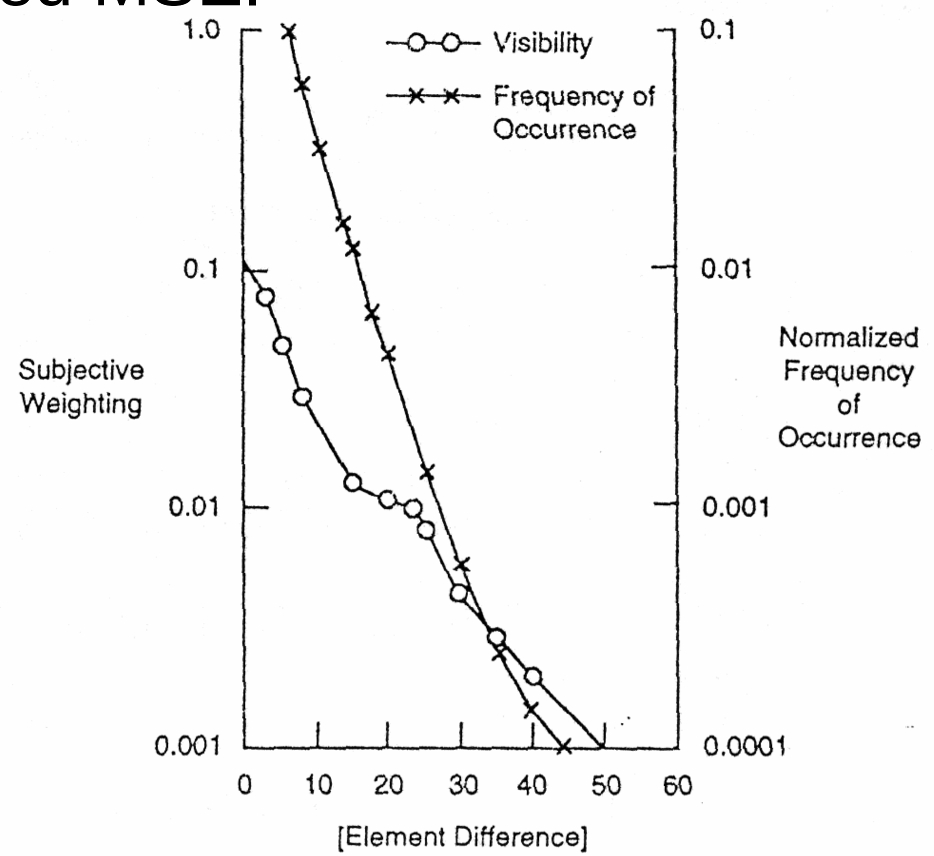




Quantization design

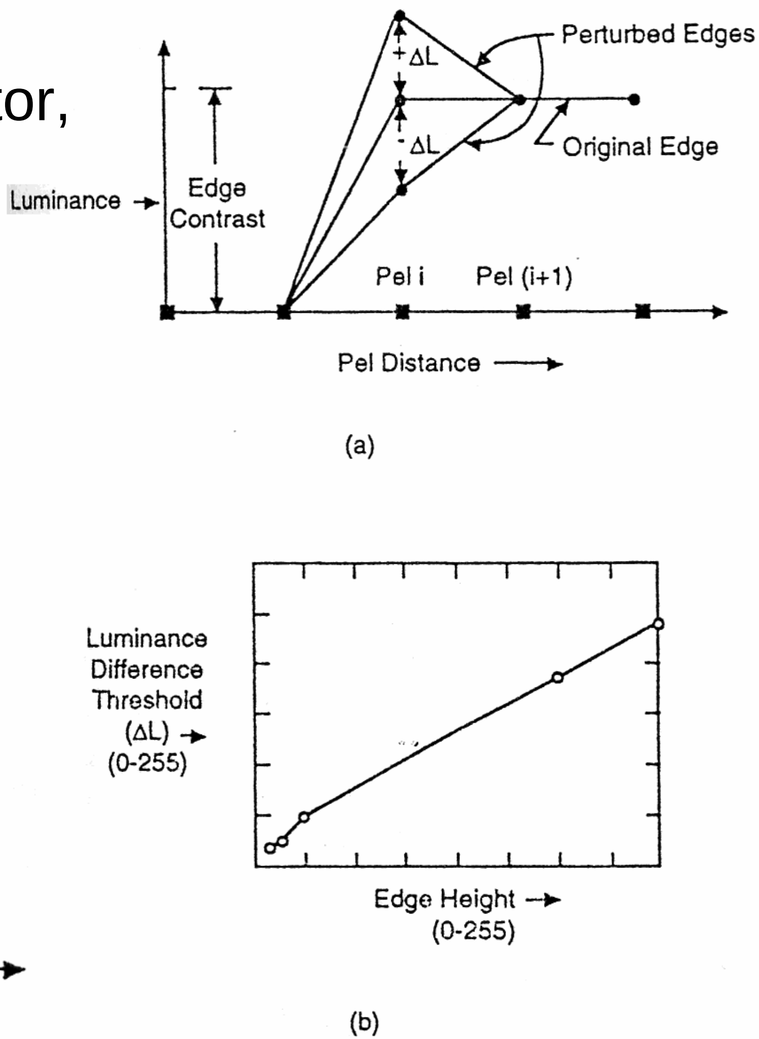
- Subjective design
  - Consider psychovisual
- Replace MSE with weighted MSE.
  - Less companded.
  - Faithful edge.

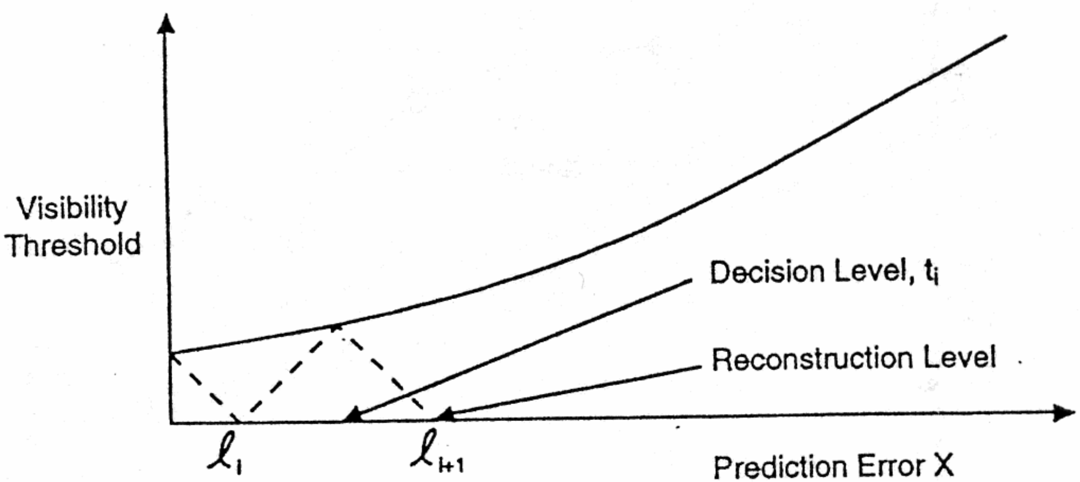




Subjective quantization design

- Make quant error below the threshold of visibility.
- Content dependent.
  - If previous pel is used as predictor, 27 levels are required for typical video conf. seq.
  - 13 levels for 2D predictor.
  - 9 and 5 for Cb and Cr.

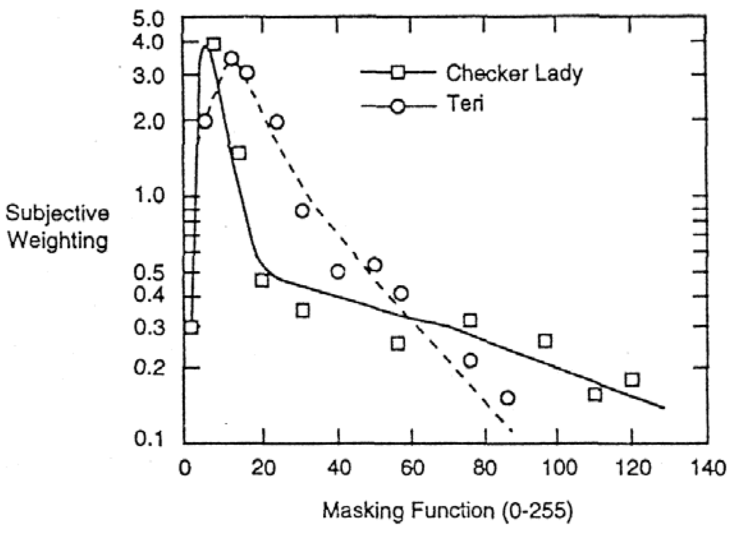


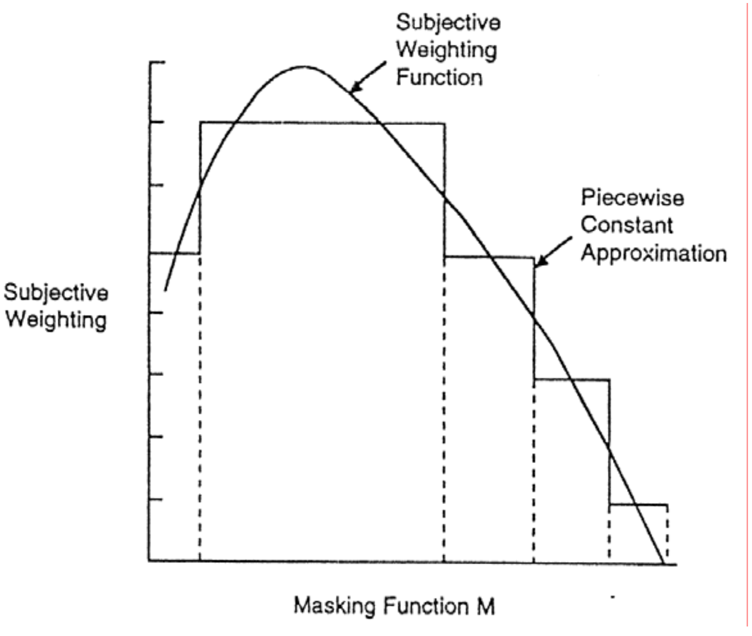


Adaptive quantization

- Adapting the DPCM quantizer usually lead to higher compression.
  - Variation of the image statistics
  - Different fidelity of reproduction
- Divide a picture into regions and then use different quantizers on each region.
- Spatial detail is measured by 3x3 matrix.

$$M(i,j)=\sum_{l,m=-1}^{+1}\alpha^{\sqrt{l^2+m^2}}\left[\left|\Delta^{H}_{i-l,j-m}\right|+\left|\Delta^{V}_{i-l,j-m}\right|\right]$$

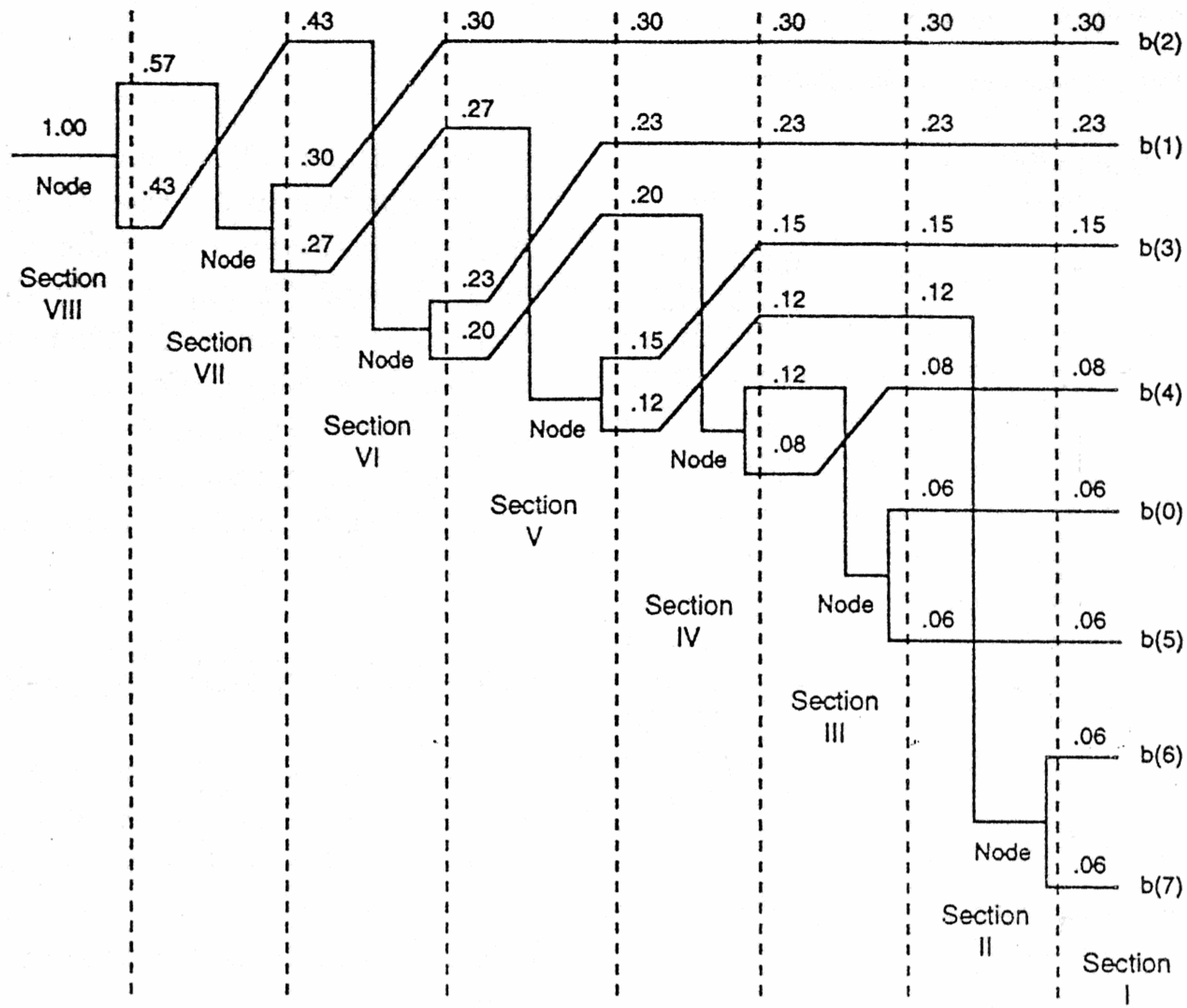




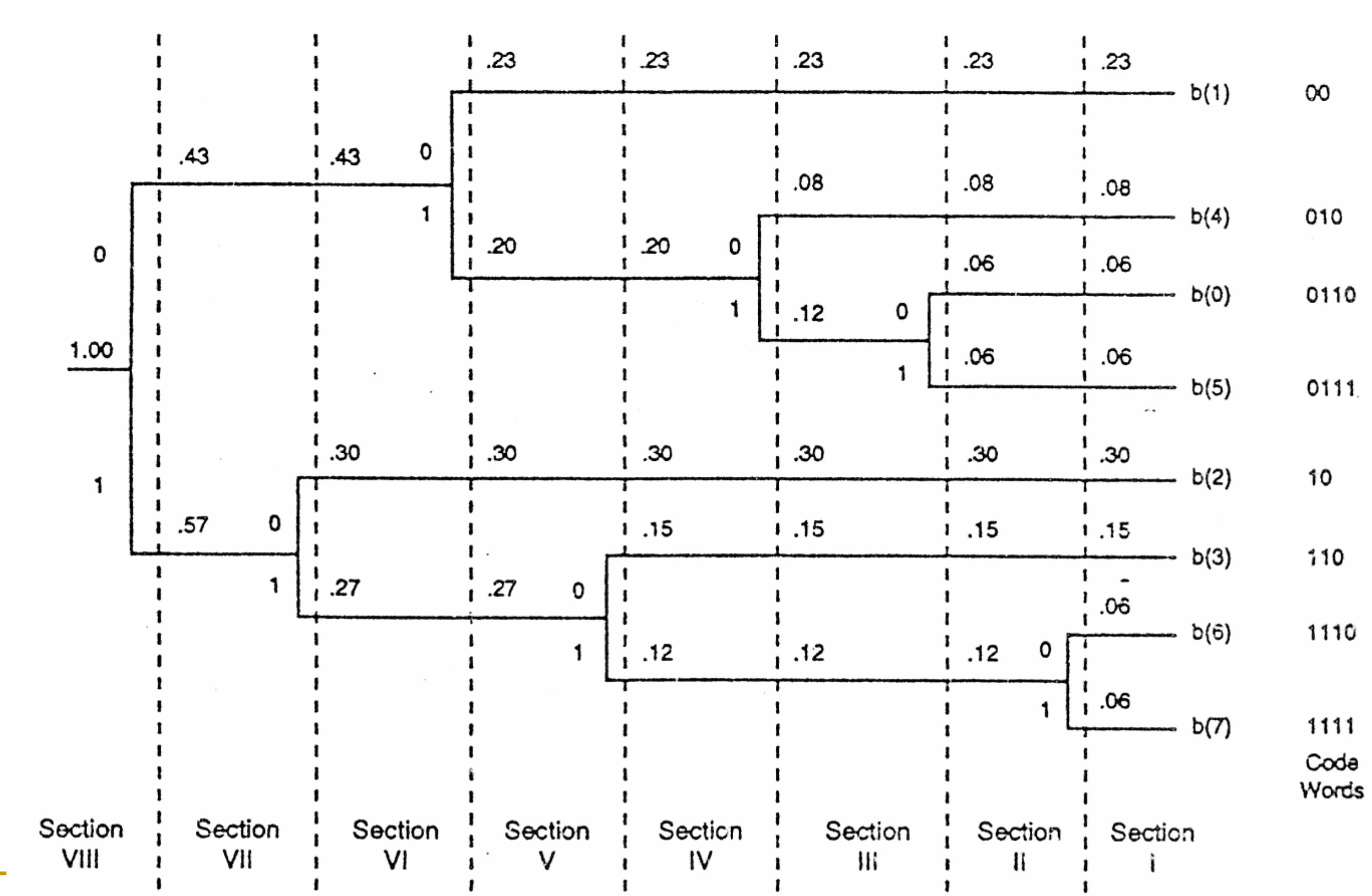
Entropy coding

- The probability distribution of quantizer output levels is highly non-uniform both in intra-coding and inter-coding.
- Huffman code is widely used.
  - Prefix rule
  - Minimize the average word length
  - Require a fairly reliable knowledge of the probability distribution
  - Result in variable bitrate
- The probability distributions for DPCM are approximately Laplacian, and Huffman code is not too sensitive to different pictures.

Construction of Huffman code



Construction of Huffman code



Transform coding

- By transforming block of pels, correlation between different transform coefficients is reduced substantially.
- Effective especially for low bitrates
  - Not all of coefficients need to be transmitted.
  - Need not be represented with full accuracy.
  - Coefficients have a very non-uniform probability.
- DCT becomes the most widely used unitary transform.
  - Ability to compress, fast and inexpensive computation.
  - Fast algorithm exists.
  - 2D block is better than 1D. 3D is too complex, requires multiple frame store and inability to incorporate motion compensation.
  - Size of 8x8 is adopted. Above 8x8 does not help too much.

Transform coding

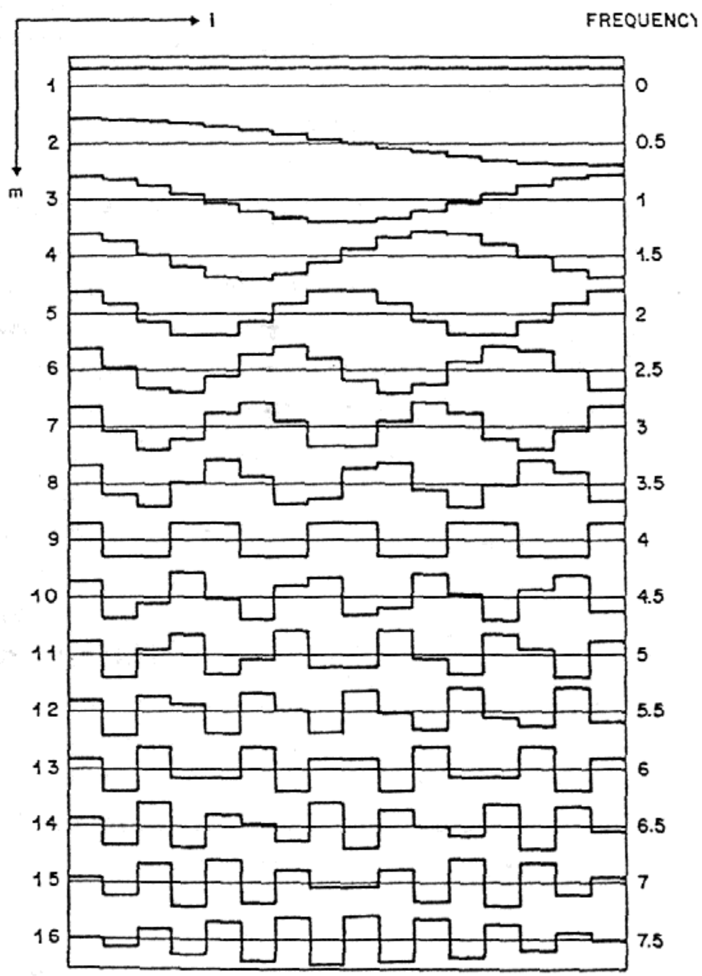


Fig. 6.25 Basis vectors for the DCT with $N = 16$.

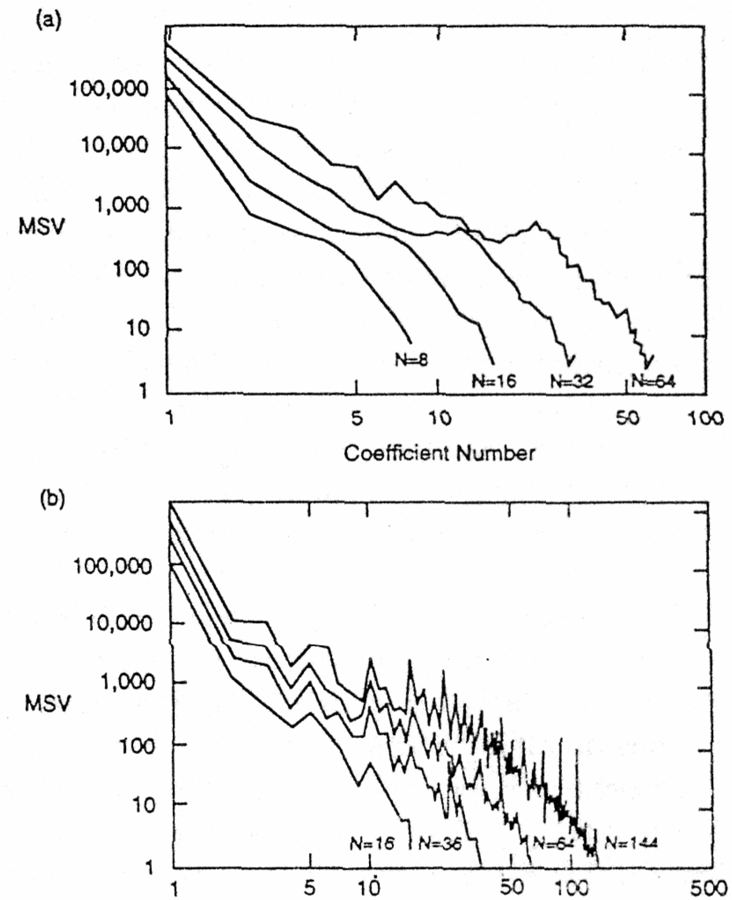
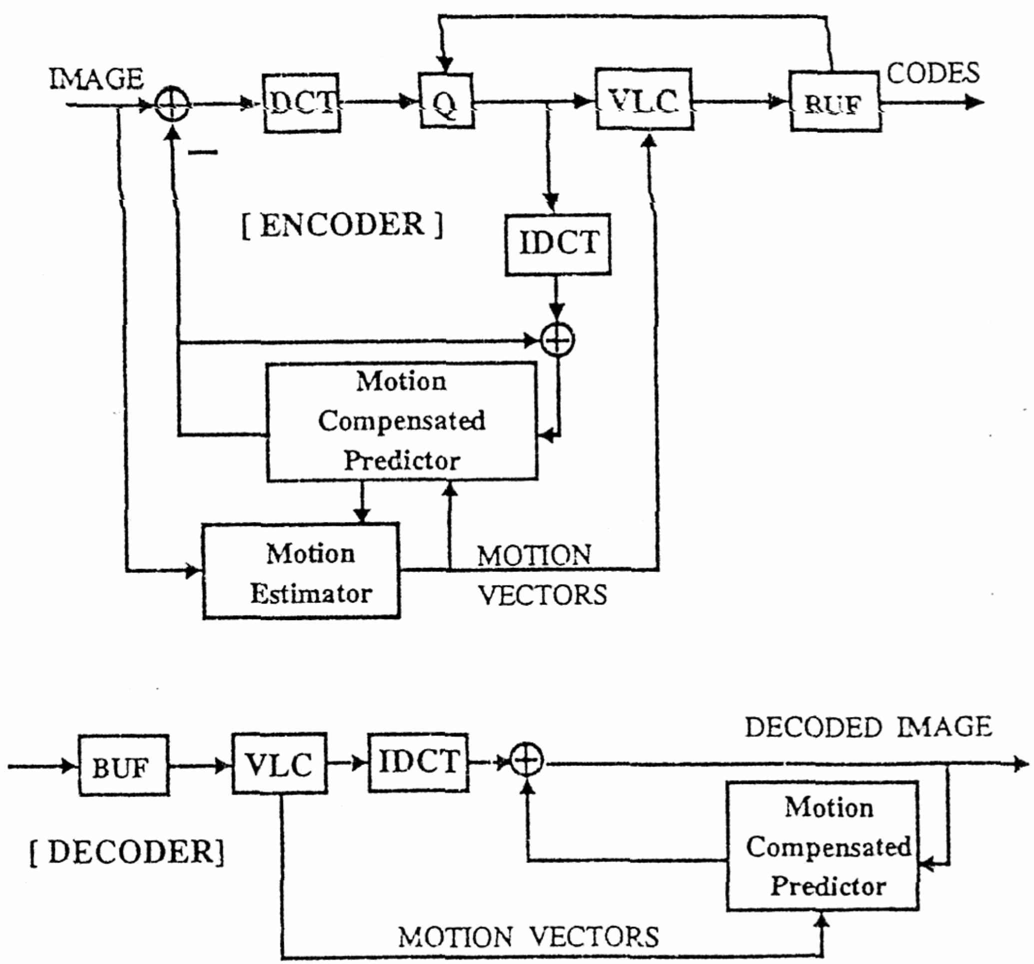


Fig. 6.27 DCT coefficient mean square values for the image "Karen" with various block sizes. (a) One-dimensional, arranged according to increasing frequency. (b) Separable two-dimensional, arranged according to increasing spatial frequency.

Quantization in transform coding

- Quantization error will spread to all pels within the block.
  - The low-energy higher order transform coefficients are set to zero, blurring or loss of sharpness occurs in regions of rapid luminance variation.
  - Coarse quantization of DC coefficient causes visible discontinuity between adjacent blocks.
  - The quantization error in higher order coefficients appears as granular noise.
- Threshold sampling
- Low detail blocks with a large number of insignificant coefficients are coded with fewer bits.
- High detail blocks are coded more coarsely.

Hybrid transform coding



Reading assignment

- Mandatory
  - Barry G. Haskell, Atul Puri, Arun N. Netravali, “Chapter 6: Digital compression: Fundamentals,” Digital video: An introduction to MPEG-2, 1997.