The Lecture Contains:

- Performance Measures
3) Subband Coding

It is known that rate-distortion theory can provide insights into the design of efficient coders. The mathematical form of the rate-distortion function suggests that an efficient coder splits the original signal into spectral components of infinitesimal bandwidth and encodes these spectral components independently. This is the basic idea behind subband coding. Subband coding was first introduced by Crochiere et al. in 1976 in the context of speech coding, and was applied to image coding by Woods and O'Neil in 1986. In subband coding the input image is passed through a set of bandpass filters to create a set of bandpass images, or subbands. Since a bandpass image has a reduced bandwidth compared to the original image, it can be downsampled (subsampled or decimated).

![Figure 8.7: A 1-D, two-band subband coding system](image)

This process of filtering and downsampling is called the analysis stage. The subbands are then quantized and coded independently. At the decoder, the decoded subbands are upsampled (interpolated), filtered, and added together to reconstruct the image. This is known as the synthesis stage. Note that subband decomposition does not lead to any compression in itself, since the total number of samples in the subbands is equal to the number of samples in the original image (this is known as critical decimation). The power of this method resides in the fact that each subband can be coded efficiently according to its statistics and visual importance. A block diagram of a basic 1-D, two-band subband coding system is presented in Figure 8.7.
Ideally, the frequency responses of the low-pass and high-pass filters should be nonoverlapping but continuous and have unity gain over their bandwidths. In practice, however, filters are not ideal and their responses must be overlapped to avoid frequency gaps. The problem with overlapping is that aliasing is introduced when the subbands are down sampled. A family of filters that circumvent this problem is the quadrature mirror filter (QMF). In the QMF, the filters are designed in such a way that the aliasing introduced by the analysis stage is exactly cancelled by the synthesis stage.

The 1-D decomposition can easily be extended to 2-D using separable filters. In this case, 1-D filters can be applied first in one dimension and direction results in four subbands: horizontal low/vertical low (LL), horizontal low/vertical high (LH), horizontal high/vertical low (HL), and horizontal high/vertical high (HH), as illustrated in Figure 8.8(a). This four-band decomposition can be continued by repetitively splitting all subbands (uniform decomposition) or just the LL subband (nonuniform decomposition). A three-stage nonuniform decomposition is illustrated in Figure 8.8(b).

Figure 8.8 Two-dimensional subband decomposition
Module 8: Video Coding Basics
Lecture 42: Sub-band coding, Second generation coding, 3D coding

Note that nonuniform decomposition results in a multiresolution pyramidal representation of the image. A commonly used technique for nonuniform decomposition is the discrete wavelet transform (DWT). The DWT is a transform that has the ability to operate at various scales and resolution levels. Having used the DWT for decomposition, various methods can be used to encode the resulting subbands. One of the most efficient methods is the embedded zero-tree wavelet (EZW) algorithm proposed by Shapiro. This algorithm assumes that if a coefficient at a low-frequency band is zero, it is highly likely that all the coefficients at the same spatial location at all higher frequencies will also be zero and, thus, can be discarded. The EZW algorithm encodes the most important information first and then progressively encodes less important refinement information. This results in an embedded bitstream that can support a range of bit rates by simple truncation. In particular, the set partitioning in hierarchical trees (SPIHT) algorithm has become the choice of most practical implementations.

One advantage of subband coding systems is that, unlike transform systems, they do not suffer from blocking artefacts at very low bit rates. In addition, they fit naturally with progressive and multiresolution transmission. One disadvantage, however, is that at very low bit rates, ringing artefacts start to occur around high-contrast edges. This is due to the Gibbs phenomenon of linear filters. To avoid this artifact, subband decomposition using nonlinear filters has been proposed.

4) Vector Quantization

Vector quantization (VQ) is a block-based spatial-domain method that has become very popular since the early 1980s. In VQ, the input image data is

![A vector quantization system](image)

**Figure 8.9:** A vector quantization system
first decomposed into k-dimensional input vectors. Those input vectors can be generated in a number of different ways; they can refer to the pel values themselves or to some appropriate transformation of them. For example, a $k = M \times M$ block of pels can be ordered to form a k-dimensional input vector $s = [s_1, ..., s_k]^T$. In VQ, the k-dimensional space $\mathbb{R}^k$ is divided into $N$ regions, or cells, $R_i$. Any input vector that falls into cell $R_i$ is represented by a representative codevector $r_i = [r_{1i}, ..., r_{ki}]^T$. The set of codevectors $C = \{r_1, ..., r_N\}$ is called the codebook. Thus, the function of the encoder is to search for the codevector $r_i$ that best matches the input vector $s$ according to some distortion measure $d(s, r_i)$. The index $i$ of this codevector is then transmitted to the decoder using at most $I = \log_2 N$ bits. At the decoder, this index is used to lookup the codevector from an identical codebook. A block diagram of a vector quantization system is illustrated in Figure 8.9.

Compression in VQ is achieved by using a codebook with relatively few codevectors compared to the number of possible input vectors. The resulting bit rate of a VQ is given by $I$ bits/pel. In theory, as $k \to \infty$, the performance of VQ approaches the rate-distortion bound. However, large values of $k$ make codebook storage and searching impractical. Values of $k = 4 \times 4$ and $N = 1024$ are typical in practical systems.

A very important problem in VQ is the codebook design. A commonly used approach for solving this problem is the Linde-Buzo-Gray (LBG) algorithm, which is a generalization of the Lloyd-Max algorithm for scalar quantization. The LBG algorithm computes a codebook with a locally minimum average distortion for a given training set and given codebook size. Entropy-constrained vector quantization (ECVQ) extends the LBG algorithm for codebook design under an entropy constraint. Another important problem is the codebook search. A full search is usually impractical, and a number of fast-search algorithms have been proposed.

There are many variants of VQ. Examples include adaptive VQ, classified VQ, tree-structured VQ, product VQ (including gain/shape VQ, mean/residual VQ, and interpolative/residual VQ), pyramid VQ, and finite-state VQ.

Theoretically, VQ is more efficient than scalar quantization for both correlated and uncorrelated data. Thus, the scalar quantizer in predictive, transform, and subband coders can be replaced with a vector quantizer.

Vector quantization has a performance that rivals that of transform coding. Although the decoder complexity is negligible (a lookup table), the high complexity of the encoder and the high storage requirements of the method still limit its use in practice. Like transform coding, VQ suffers from blocking artefacts at very low bit rates.
Module 8: Video Coding Basics
Lecture 42: Sub-band coding, Second generation coding, 3D coding

5) Second-Generation Coding
The coding methods discussed so far are generally known as waveform coding methods. They operate on pels or blocks of pels based on statistical image models. This classical view of the image coding problem has three main disadvantages. First, it puts more emphasis on the codeword assignment (using information and coding theory) rather than on the extraction of representative messages. Because the encoded messages (pels or blocks) are poorly representative in the first place, saturation in compression is eventually reached no matter how good is the codeword assignment. Second, the encoded entities (pels or blocks) are consequences of the technical constraints in transforming scenes into digital data, rather than being real entities. Finally, it does not place enough emphasis on exploiting the properties of the HVS. Efforts to utilize models of the HVS and to use more representative coding entities (real objects) led to a new class of coding methods known as the second-generation coding methods.

Second-generation methods can be grouped into two classes: local-operator-based techniques and contour/texture-oriented techniques. Local-operator-based techniques include pyramidal coding and anisotropic nonstationary predictive coding, whereas the contour/texture-oriented techniques include directional decomposition coding and segmented coding. Two commonly used segmented coding methods are region-growing and split-and-merge. Second-generation methods provide higher compression than waveform coding methods at the same reconstruction quality. They also do not suffer from blocking and blurring artefacts at very low bit rates. However, the extraction of real objects is both difficult and computationally complex. In addition, such methods suffer from unnatural contouring effects, which can make the details seem artificial.

6) Other Coding Methods
There are many other intraframe coding techniques. Example are block-truncation coding, fractal coding, quad-tree and recursive coding, multiresolution coding, and neural-network-based coding. A detailed (or even a brief) discussion of such techniques is beyond the scope of this module.
Interframe Coding

As already discussed, video is a time sequence of still images or frames. Thus, a naive approach to video coding would be to employ any of the still-image (or intraframe) coding methods discussed on a frame-by-frame basis. However, the compression that can be achieved by this approach is limited because it does not exploit the high temporal correlation between the frames of a video sequence. Interframe coding refers to video coding techniques that achieve compression by reducing this temporal redundancy. For this reason, such methods are also known as temporal redundancy reduction techniques. Note that interframe coding may not be appropriate for some applications. For example, it would be necessary to decode the complete interframe coded sequence before being able to randomly access individual frames. Thus, a combined approach is normally used in which a number of frames are intraframe coded (I-frames) at specific intervals within the sequence and the other frames are interframe coded (predicted or P-frames) with reference to those anchor frames. In fact, some systems switch between interframe and intraframe coding, within the same frame.

1) Three-Dimensional Coding

The simplest way to extend intraframe image coding methods to interframe video coding is to consider 3-D waveform coding. For example, in 3-D transform coding based on the DCT, the video is first divided into blocks of $M \times N \times K$ pels ($M$, $N$, $K$ denote the horizontal, vertical, and temporal dimensions, respectively). A 3-D DCT is then applied to each block, followed by quantization and symbol encoding, as illustrated in Figure 8.10. A 3-D coding method has the advantage that it does not require the computationally intensive process of motion estimation. However, it requires $K$ frame memories both at the encoder and decoder to buffer the frames. In addition to this storage requirement, the buffering process limits the use of this method in real-time applications because encoding/decoding cannot begin until all of the next $K$ frames are available. In practical systems, $K$ is typically set to 2-4 frames.

![Figure 8.10: A 3-D transform coding system](image)