CSE 126/228F Lecture 4: April 10

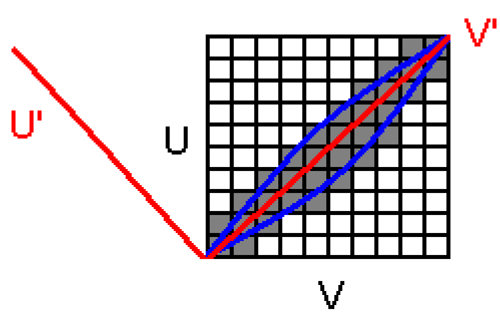
Digital Video

The bandwidth required for digital video is staggering. Uncompressed NTSC video requires a bandwidth of 20MByte/sec, HDTV requires 200MByte/sec! Various encoding techniques have been developed in order to make digital video feasible. Two classes of encoding techniques are Source Encoding and Entropy Encoding.

Source Encoding:

Source encoding is lossy and applies techniques based upon properties of the media. There are four types of source encoding:

- *Sub-band* coding gives different resolutions to different bands. E.g. since the human eye is more sensitive to intensity changes than color changes, give the Y component of YUV video more resolution than the U and V components.
- *Subsampling* groups pixels together into a meta-region and encodes a single value for the entire region
- *Predictive* coding uses one sample to guess the next. It assumes a model and sends only differences from the model (error values).
- *Transform* encoding transforms one set of reference planes to another. In the example of vector quantization from last class, we could rotate the axes 45 degrees so that fewer bits could be used to represent values on the U-axis. In this example, instead of using 4 bits to represent the ten possible values of U, we can use 2 bits for the four different values of U'.

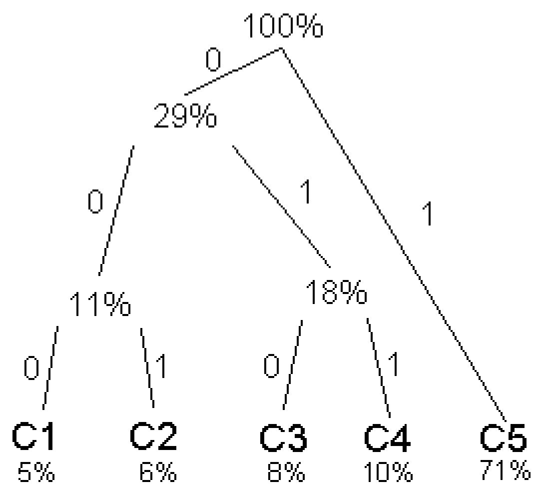


Entropy Encoding:

Entropy Encoding techniques are lossless techniques which tend to be simpler than source encoding techniques. The three entropy encoding techniques are:

- *Run-Length Encoding (RLE)* encodes multiple appearances of the same value as {value, # of appearances}. E.g. 1,1,1,1,2,2,2,3 would encode as {1,4},{2,3},{3,1}

- *Huffman Coding* looks at statistical distributions of data to provide compression. It does this by giving the smallest length code to the most frequent character and the giving the longest length code to the character which occurs least. Given a set of data with 5 values {c1,c2,c3,c4,c5}, having distributions {c1=5%, c2=6%,c3=8%,c4=10%,c5=71%}, a Huffman Tree will be built. This tree defines the codes for the 5 values {c1=000,c2=001,c3=010,c4=011,c5=1}. This tree also satisfies another important property of Huffman coding; any code cannot be a proper prefix of another code. If this property did not hold, we would be unable to decode the variable bit-length code, because one value could appear as a combination of two other values or vice versa.

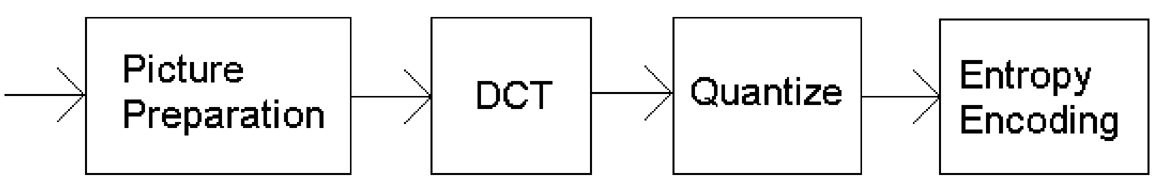


- A*rithmetic Coding* is similar to Huffman Coding, but it is more complex and provides better compression, especially for text. For images, it is not necessary.

JPEG

JPEG is an acronym for Joint Photographic Experts Group, the group that created the standard. It provides fast, efficient compression for images. It is also the basis for the MPEG video compression standard.

JPEG Stages



Picture Preparation

The first step of the picture preparation phase converts the image into components up to 256 separate components. In JPEG, compression is never done across planes. The standard set of planes for JPEG is YUV.

Now, we can use some information about the source. We know that Y is more important to humans than U or V, so we can subsample the U and V planes by a factor of four in both the x and y direction, encoding a value for sixteen adjacent pixels as one meta-region.

Now, 8x8 blocks of values will be sent through the remaining stages. Note that for the subsampled U and V planes, 64 meta-regions comprise a block. Blocks pass through the stages one at a time and an entire region is completed before the next region begins. This is called *Interleaving*, and it means that block 1 will have its Y,U,V values encoded before block 2 begins. Due to the subsampling of U and V, one U and V encoding will be sent with 16 Y encodings.

Transform

There are many transforms, most of which are very slow. This is important to consider since video demands real-time encoding and decoding. The JPEG committee took suggestions and empirically studied the use of several different transforms. Although it is not optimal; of the transforms studied, DCT (Discrete Cosine Transform) proved superior.

In JPEG, DCT operates on one block at a time. Because there are 64 elements in an 8x8 block, this is called the 64-element or 64-coefficient DCT. The DCT transform operates on this block in a left-to-right, top-to-bottom manner.

We will examine DCT next time.