# How JPEG Works

Shuvraneel Mitra (22EC10074), Abhirup Giri (22EC30002), and Arkaprova Haldar (22EC30008)

Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

**Abstract.** The following text is a brief overview of how the JPEG compression technique works, compiled from the texts and resources shared in the Multimedia Systems and Applications (EC60104) class.

**Keywords:** Still Image Compression · JPEG standards · Discrete Cosine Transform.

## 1 Introduction

With the rapid developments of imaging technology, image compression and coding tools and techniques, it is necessary to evolve coding standards so that there is compatibility and interoperability between the image communication and storage products manufactured by different vendors. In this text, we discuss the highlights of these standards, specifically **JPEG**, created by the Joint Photographic Expert Group. JPEG is a very simple and easy-to-use standard based on the **Discrete Cosine Transform (DCT)**.

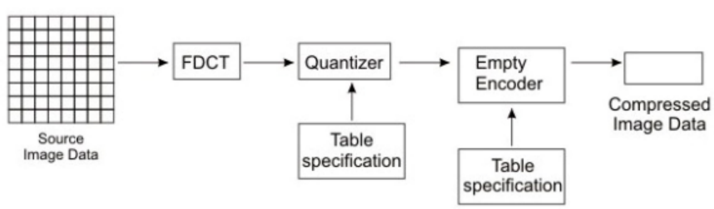


**Fig. 1.** The JPEG Encoder

**The JPEG Encoder** The JPEG encoder shown in Figure 1 has the following components:

*1. Forward Discrete Cosine Transform (FDCT)* The still images are first partitioned into non-overlapping blocks of size $8 \times 8$ and the image samples are shifted from unsigned integers with range $[0, 2^p - 1]$ to signed integers with range $[-2^{p-1}, 2^{p-1}]$, where $p$ is the number of bits (here, 8). The transformed image is then passed through a DCT algorithm. It should be mentioned that, to preserve freedom for customization within implementations, JPEG neither specifies any unique FDCT algorithm nor any unique IDCT algorithms.

*2. Quantization* Each of the 64 coefficients of the FDCT output of a block is uniformly quantized **according to a quantization table**. Since the aim is to compress the images without visible artifacts, each step size should be chosen as the perceptual threshold or for "just noticeable distortion". Psychovisual experiments have led to a set of quantization tables, in ISO-JPEG standard. The quantized coefficients are zig-zag scanned and each DC coefficient is encoded as a difference from the DC coefficient of the previous block and the 63 AC coefficients are encoded into **(run, level) pairs.**

*3. Entropy Coding* The JPEG standard specifies two entropy coding methods – Huffman and arithmetic coding. Huffman coding requires that one or more sets of coding tables are specified by the application. The same table used for compression is used needed to decompress it.

**The JPEG Decoder** The decoder uses an IDCT algorithm at the end to approximately recover the original image.

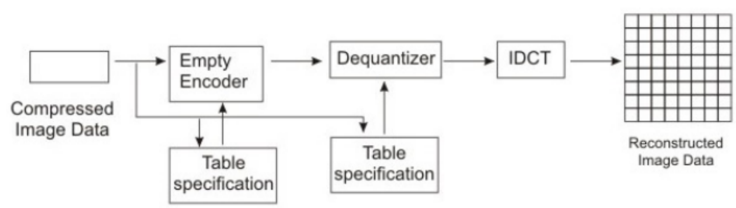


**Fig. 2.** The JPEG Decoder performs inverse operations with respect to the JPEG encoder

## Modes of Operation in JPEG

*1. Baseline encoding* Baseline/sequential coding is for images with 8-bit samples and uses Huffman coding only. In baseline encoding, each block is encoded in a single left-to-right and top-to-bottom scan. It encodes and decodes complete 8x8 blocks **with full precision** one at a time.

*2. Progressive Encoding* Unlike baseline encoding, each block in progressive encoding is encoded in multiple scans, rather than a single one. Each scan takes much less time to encode and decode, as compared to the single scan of baseline encoding, since each scan contains only a part of the complete information. With the first scan, a crude form of image can be reconstructed at the decoder and with successive scans, the quality of the image is refined. It is very convenient for browsing applications, where crude reconstruction quality at the early scans may be sufficient for quick browsing. There are two forms of progressive encoding: (a) spectral selection approach and (b) successive approximation approach.

(a) ***Spectral selection approach***: The first scan sends some specified low frequency DCT coefficients within each block. The reconstructed image obtained at the decoder from the first scan therefore appears blurred as the details in the forms of high frequency components are missing. In subsequent scans, bands of coefficients, which are higher in frequency than the previous scan, are encoded and therefore the reconstructed image gets richer with details.

(b) ***Successive approximation***: Here, each scan encodes all the coefficients within a block, but not to their full quantized accuracy. In the first scan, only the $N$ most significant bits of each coefficient are encoded (where $N$ is specifiable) and in successive scans, the next lower significant bits of the coefficients are added and so on until all the bits are sent. The resulting reconstruction quality is good even from the early scans, as the high frequency coefficients are present from the initial scans.

*3. Hierarchical Encoding* The hierarchical encoding is an encoding in which the image to be encoded is organized in a pyramidal structure of multiple resolutions, with the original (the finest resolution) image on the lowermost layer and reduced resolution images on the successive upper layers. Each layer decreases its resolution with respect to its adjacent lower layer **by a factor of two** in either the horizontal or the vertical direction or both. Hierarchical encoding is used for applications in which a high-resolution image should be accessed by a low resolution display device.

Hierarchical encoding may be regarded as a special case of progressive encoding with increasing spatial resolution between the progressive stages. We encode the reduced resolution image from the topmost layer of the pyramid. Then we decode the above reduced resolution image, then interpolate and up-sample it by a factor of two horizontally and/or vertically (using the identical interpolation filter of the decoder). We use this interpolated and up-sampled image as a predictor for encoding the next lower layer. Encode the difference between the image in the next lower layer and the predicted image. Repeat these steps till the lowermost layer of the pyramid is encoded.

*4. Lossless Encoding* The lossless mode of encoding in JPEG follows a simple predictive coding mechanism, rather than having FDCT + Entropy coder for encoding and Entropy decoder + IDCT for decoding. In lossless encoding, the $8 \times 8$ block structure is not used and each pixel is predicted based on three adjacent pixels using one of the eight possible predictor modes:

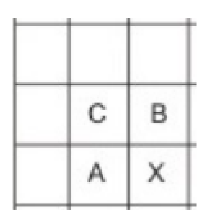


**Fig. 3.**

**Table 1.** The predictor table for lossless JPEG Mode

| Selection Value | Prediction |
|---|---|
| 0 | None |
| 1 | A |
| 2 | B |
| 3 | C |
| 4 | A+B-C |
| 5 | A+(B-C)/2 |
| 6 | B + (A-C)/2 |
| 7 | (A+B)/2 |

**YUV and Interleaving** YUV is considered a more efficient color space than RGB, because our eyes are relatively insensitive to the high frequency information from the chrominance channels and thus the chrominance components can be represented at a reduced resolution. To convert RGB to YUV, we have the following relations:

$$Y = 0.3R + 0.6G + 0.1B \tag{1}$$

$$U = \frac{B - Y}{2} + 0.5 \tag{2}$$

$$V = \frac{R - Y}{1.6} + 0.5 \tag{3}$$

The U and the V components are sub-sampled by a factor of two in both horizontal and vertical directions.

The three components may be transmitted in either an interleaved manner or a non-interleaved manner. The non-interleaved ordering can be shown as
**Scan-1**: $Y_1, Y_2, Y_3, ..., Y_{15}, Y_{16}$.
**Scan-2**: $U_1, U_2, U_3, U_4$.
**Scan-3**: $V_1, V_2, V_3, V_4$.

The interleaved ordering encodes in a single scan and proceeds like

$$Y_1, Y_2, Y_3, Y_4, U_1, V_1, Y_5, Y_6, Y_7, Y_8, U_2, V_2, ...$$

Interleaving requires much lower buffering to decode the image at the decoder.

### 1.1 Concluding Remarks

Considering color images having 8-bits/sample luminance components and 8-bits/sample for each of the two chrominance components U and V, each pixel requires 16-bits for representation, if both U and V are sub-sampled by a factor of two in either of the directions. Using JPEG compression on a wide variety of such color images, the following image qualities were measured subjectively:

| Bits/pixel | Quality | Compression Ratio |
|---|---|---|
| $\geq 2$ | Indistinguishable | 8:1 |
| 1.5 | Excellent | 10.7:1 |
| 0.75 | Very good | 21.4:1 |
| 0.5 | Good | 32:1 |
| 0.25 | Fair | 64:1 |

**Table 2.** Image Quality versus Compression Rate

We have specifically considered only the JPEG image encoding standards in the above presentation. A more advanced still image compression standard JPEG-2000 has evolved in recent times, which we intend to cover in the future.