Explain MPEG compression process in detail.

MPEG compression removes two types of redundancies:

Spatial redundancy:

- Pixel values are not independent, but are correlated with their neighbors both within the same frame and across frames. So, to some extent, the value of a pixel is predictable given the values of neighboring pixels.

- It is removed with the help of DCT compression.

Temporal redundancy:

- Pixels in two video frames that have the same values in the same location (some objects repeated again and again in every frame).

- It is removed with the help of Motion compensation technique Macroblock

- Each macroblock is composed of four 8*8 Luminance (Y) blocks and two 8*8 Chrominance (Cb & Cr) blocks.

- This set of six blocks is called a macro block.

- It is the basic hierarchical component used achieving high level of compression.

- The key to achieving a high rate of compression is to remove much redundant information as possible.

- Entropy encoding and Huffman coding are two schemes used for encoding video information.

- MPEG takes the advantage of the fact that there exists a correlation between successive frames of moving pictures.

MPEG constructs three types of pictures namely:

- Intra pictures (I-pictures)

- Predicted pictures (P-pictures)

- Bidirectional predicted pictures (B-pictures)

The MPEG algorithm employs following steps:

Intra frame DCT coding (I-pictures):

The I-pictures are compressed as if they are JPEG images.

- First an image is converted from RGB color model to YUV color model.

- In general, each pixel in a picture consists of three components: R (Red), G (Green), and B (Blue).

- But (R, G, B) must be converted to (Y, Cb, Cr) in MPEG-1, then they are processed.

- Usually we use (Y, U, V) to denote (Y, Cb, Cr).

Apply DCT

DCT is performed on small blocks of 8*8 pixels to produce blocks of DCT coefficients.

The NXN two-dimensional DCT is defined as:

F(u,v)=2NC(u)C(v)∑x=0N−1∑y=0N−1f(x,y)cos(2x+1)uπ2Ncos(2y+1)vπ2N

$$F(u,v) = \frac{2}{N}C(u)\,C(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y)\,\cos\frac{(2x+1)\,u\pi}{2N}\cos\frac{(2y+1)\,v\pi}{2N}$$

C(u),C(v)={12−√foru,v=0

$$C\left(u\right),C\left(v\right) = \{ \frac{1}{\sqrt{2}} foru, v = 0$$

=1otherwise

$$= 1otherwise$$

The inverse DCT (IDCT) is defined as:

f(x,y)=2N∑u=0N−1∑v=0N−1C(u)C(v)cos(2x+1)2Ncos(2y+1)vπ2N

$$f\left(x,y\right) = \frac{2}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C\left(u\right)C\left(v\right)\cos\frac{\left(2x+1\right)}{2N}\cos\frac{\left(2y+1\right)v\pi}{2N}$$

Where x, y are spatial co-ordinates in the image block u, v are co-ordinates in the coefficient block.

Apply Quantization

- Quantization is process that attempts to determine what information can be safely discarded without significant loss in visual fidelity.

- MPEG uses a matrix called quantizer (Q[i,j]) to define quantization step. Every time when a pixels matrix (X[i,j]) with the same size to Q[i,j] comes ,use Q[i,j] to divide x(i,j) to get quantized value matrix Xq[i,j].

- Quantization Equation Xq[i,j] = Round(X[i,j)/Q[i,j])

- After Quantization, perform Zig-zag scanning to gather even more consecutive zeroes.
- Then various compression algorithms are applied including Run-length and Huffman encoding.
- In Huffman coding, we give shorter keywords to more frequently coefficients & longer keywords to least frequently occurring coefficients.
- Hence achieving final level of compression.

Motion-compensated inter-frame prediction (P-pictures):

- In most video sequences there is a little change in the contents of image from one frame to the next.

- Most video compression schemes take advantage of this redundancy by using the previous frame to generate a prediction of current frame.

- It removes temporal redundancy by attempting to predict the frame to be coded from previous frame.

- This is based on current value to predict next value and code their difference called as prediction error.

- Motion compensation assumes that current picture is some translation of previous frame.

- The frame to be compared is split in to blocks first and then best matching block is searched.
- Each block uses previous picture for estimating prediction.
- This search process is called as prediction.

Motion-compensated inter-frame prediction:

- By reducing temporal redundancy, P-pictures offer increased compression compared to I-pictures.

- Motion Estimation is to predict a block of pixel value in next picture using a block in current picture. The location difference between these blocks is called Motion Vector. And the difference between two blocks is called prediction error.

- In MPEG-1, encoder must calculate the motion vector and prediction error. When decoder obtains this information, it can use this information and current picture to reconstruct the next picture. We usually call this process as Motion Compensation.

B-frame (Bidirectional predictive frame):

-Frames can also be predicted from future frames. Such frames are usually predicted from two directions, i.e. from the I- or P-frames that immediately precede or follow the predicted frame.

-These bidirectionally predicted frames are called B-frames. A coding scheme could, for instance, be IBBPBBPBBPBB.

-B-pictures uses the previous or next I-frame or P-frame for motion compensation and offers the highest degree of compression.

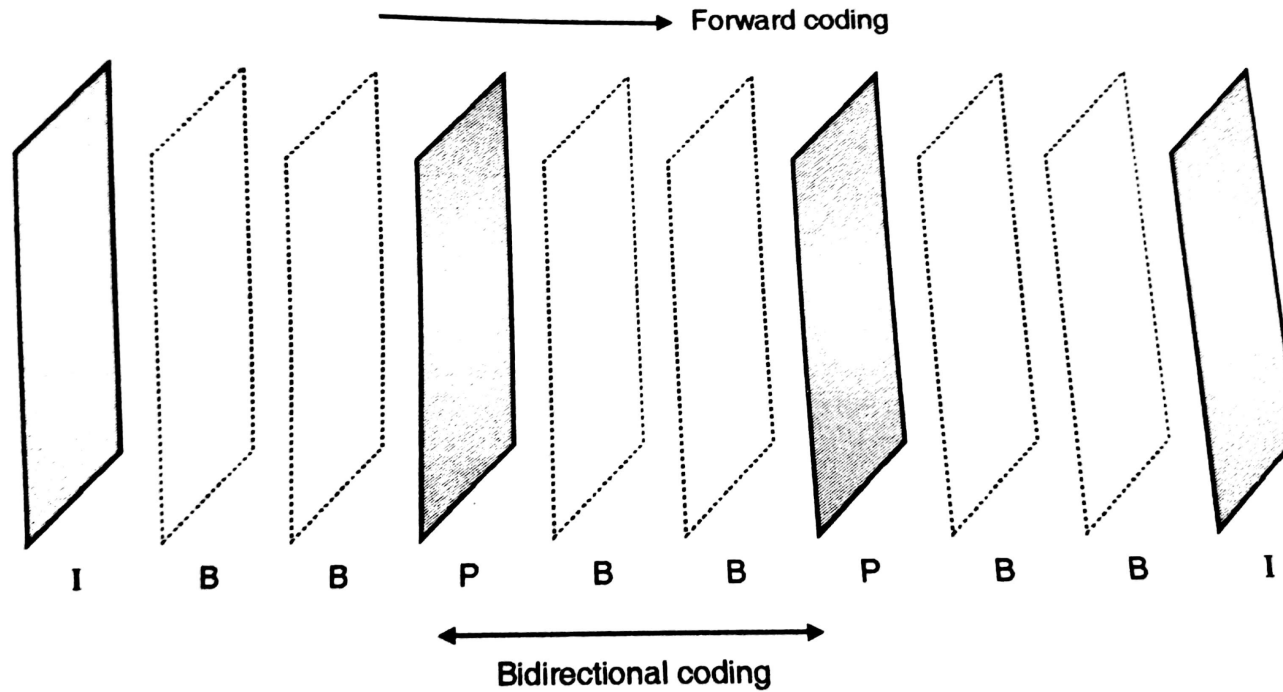Each block in a B-picture can be forward, backward or bidirectionally predicted.
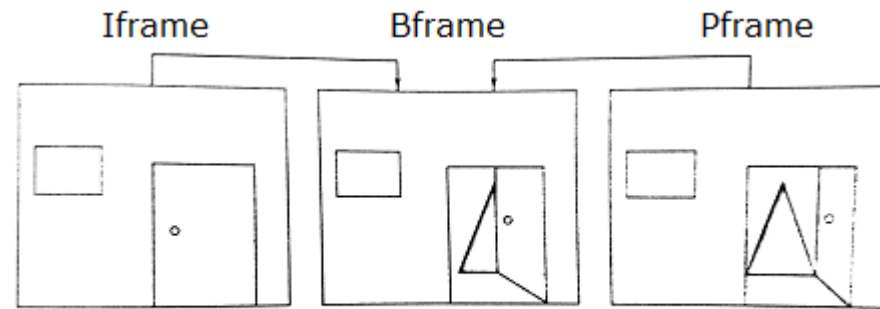

Fig. : Illustrated use of three picture types in a typical GOP

I-, P- and B-frames

Bidirectional predicted pictures (B):

- Bidirectional predicted pictures utilize three types of motion compression techniques.
- Forward motion compensation - uses past picture information.
- Backward motion compensation - uses future picture information .
- Bidirectional compensation - uses the average of the past and future picture information.

MPEG encoder: