# Hierarchical Mode of the JPEG Standard

The hierarchical encoding encodes an image in multiple resolutions. For e.g., one could provide 320x240, 640x480 and 1280x960 versions of an image; the decoder at the receiving end can choose the optimum resolution depending on the target's capabilities. Thus, high-resolution images can be easily viewed in lower resolution devices. This is particularly relevant to small portable terminals and for conferencing where multiple smaller images need to share the screen with full size images at different times.

In hierarchical coding scheme (refer to figure 1), the image is first sub-sampled by 2 in (both) dimension(s). This new reduced size image is encoded using one of the sequential, progressive or loss-less modes described previously. Then the encoded reduced-size image is decoded and upsampled by 2 horizontally and/or vertically. This upsampled image is used as a prediction of the original image at this resolution and the difference image is computed. The difference image (called the differential frame) is then encoded and finally, the last two steps are repeated until the original image at full resolution has been encoded. Since the higher-resolution images are coded as differences from the next smaller image, they require fewer bits than they would if sent independently.

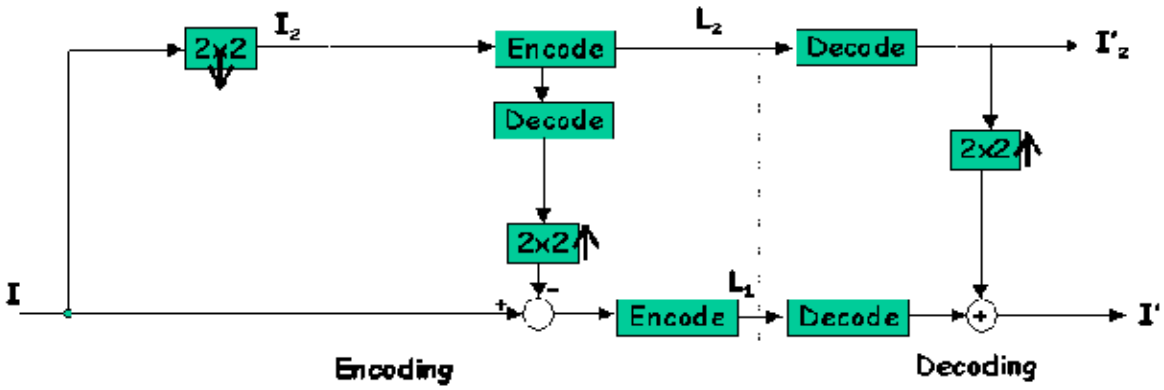


*Figure 1. Two-level Hierarchical mode Encoder/Decoder*

Figure 2 shows the block diagram of a three-level hierarchical encoder/decoder. From the original image I, we generate two subsampled versions: $I_2$, where the image is subsampled by a factor of two on both dimensions; and $I_4$, where the image is subsampled by a factor of four on both dimensions. The encoded image is coded at three different frame resolutions: $L_1$, $L_2$ and $L_4$. Image $L_4$ is just the image $I_4$ compressed. Using only $L_4$, the decoder can extract a low-resolution estimate of the original image ($I_4'$). $L_2$ (uncompressed) is the difference image between $I_2$ and an estimate of $I_2$ ($I_2'$) after upsampling $I_4$ by a factor of two. Using only $L_4$ and $L_2$, the decoder can extract a medium resolution estimate of the input I ($I_2'$). Similarly, $L_1$ (uncompressed) is the difference image between I and an estimate of I based on $I_2$ and $I_4$.

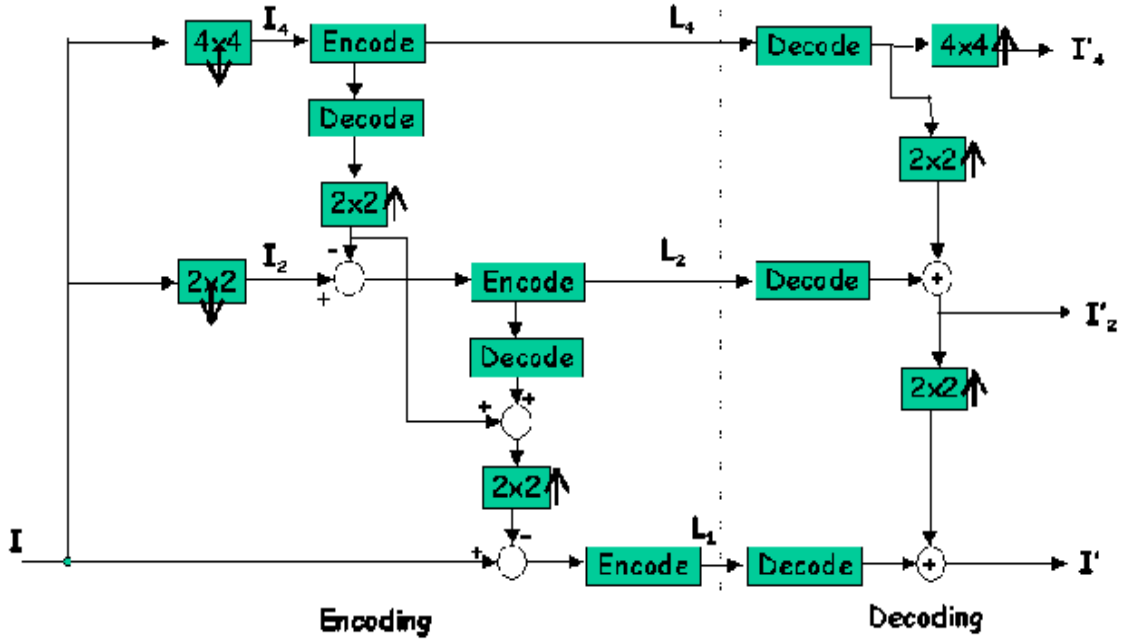


***Figure 2. Three-level Hierarchical Encoder/ Decoder***

The reason we need to add the component of $I_4$ (coded and upsampled by 2) to the component of $I_2$ during the encoding process, is that at the destination (decoder), we see that the image is constructed back mainly from $I_4'$, along with the difference components of the images in higher resolutions.

**At the Encoder:**

$I \xrightarrow{\downarrow 4} I_4$ (downsample original image I by 4)

$I \xrightarrow{\downarrow 2} I_2$ (downsample original image I by 2)

$E(I_4) \longrightarrow L_4$ (Encode $I_4$)

$D(L_4) \longrightarrow I_4'$ (Decode $L_4$; $I_4' = I_4 + l_1$ )

$I_4' \xrightarrow{\uparrow 2} I_2''$ (upsample $I_4'$ by 2)

$E(I_2 - I_2'') \longrightarrow L_2$ (Encode the difference)

$D(L_2) \longrightarrow L_2'$ (Decode $L_2'$; $L_2' = I_2 - I_2'' + l_2$)

$I_2'' + L_2' \longrightarrow I_2'$ (where $I_2' = I_2 + l_2$)

$I_2' \xrightarrow{\uparrow 2} I''$ (upsample $I_2''$ by 2)

$E(I - I'') \longrightarrow L_1$ (Encode the difference (I - I''))

**At the Decoder:**

$D(L_4) \longrightarrow I_4'$ (decode $L_4$; $I_4' = I_4 + l_1$)

$I_4' \xrightarrow{\uparrow 2} I_2''$ (upsample $I_4'$ by 2)

$D(L_2) \longrightarrow L_2'$ (decode $L_2$; $L_2' = I_2 - I_2'' + l_2$)

$I_2'' + L_2' \longrightarrow I_2'$ (here, $I_2' = I_2 + l_2$)

$I_2' \xrightarrow{\uparrow 2} I''$ (upsample $I_2'$ by 2; $I'' = I + l_2$)

$D(L_1) \longrightarrow L_1'$ (thus, $L_1' = (I - I'') + l_3$)

$I'' + L_1' \longrightarrow I'$ (here, $I' = I + l_3$)

($l_1$, $l_2$ and $l_3$ are losses introduced due to encoding and decoding)

Example of sampling an image:

sampled by 2

sampled by 2