CHAPTER 5: DATA COMPRESSION

Multimedia Systems

WHY COMPRESS?

- To reduce the volume of data to be transmitted(Text, Audio,Image,Video)
- To reduce the bandwidth required for transmission.
- Data compression is the process of encoding, restructuring or otherwise modifying data in order to reduce its size.
- Fundamentally, it involves re-encoding information using fewer bits than the original representation.
- Text compression can usually succeed by removing all unnecessary characters, instead inserting a single character as reference for a string of repeated characters, then replacing a smaller bit string for a more common bit string.
- Image compression with the sequence of colors, like 'blue, red, red, red' is found throughout the image, the formula can turn this data string into smaller data stream, while still maintaining the underlying information.

AUDIO AND VIDEO COMPRESSION

- **Audio compression:**

❖ Implemented as audio codecs, compression of audio files is necessary to guarantee bandwidth and storage limits aren't exceeded.

❖ Audio compression can be either lossy or lossless, MP3 being the most ubiquitous lossy codec. FLAC is a major lossless encoding format.

- **Video compression:**

❖ Videos combine image compression with audio compression. There are usually separate codecs for each aspect of a video, which are then wrapped together as a single compression codec.

❖ Because of the high data rate required for uncompressed video, most video files are compressed using lossy compression. The most prevalent form of (lossy) video compression is MPEG.

MODES FOR COMPRESSION

- Compression in multimedia system is subject to certain constraints:
- The complexity of the technique used should be minimal to avoid high cost.
- The processing of compression algorithms must not exceed certain time spans.
- The Two Modes are:

1. **Dialogue Mode**

2. **Retrival Mode**

MODES FOR COMPRESSION

1. Dialogue Mode

- *"Interaction among human users via multimedia information."*
- Requirements for compression and decompression:
  - End-to-End delay lower than 150ms
  - End-to-End delay of 50ms for face to face dialogue application

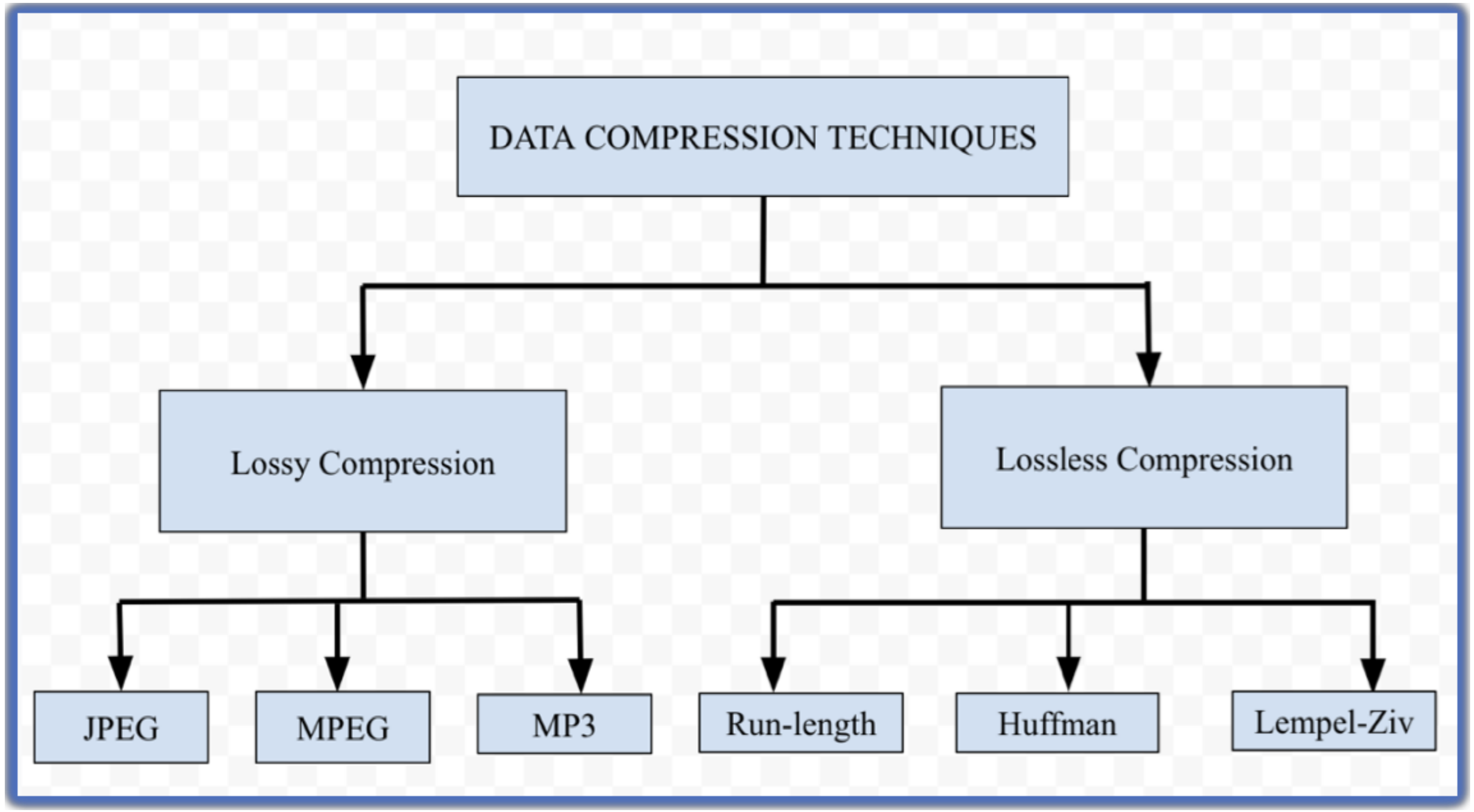
Retrival Mode

- "Retrival of information from multimedia database"
- Requirements:
  - Forward and backward data retrival with simultaneous display
  - Fast search of information in multimedia database
  - Random access to images and audio files with access time less than 0.5 seconds

REQUIREMENTS FOR DIALOGUE AND RETRIVAL MODE

- Support scalable video in different systems independent of frame size.
- Must be able to synchronize audio and video datas
  - *Lip synchronization*
- To make an economical solution to compression.
  - *Use of software (cheap,may be low quality) or Hardware (Expensive, High Quality)*
- It should be possible to generate multimedia data in one system and reproduce it in another system
  - *Compatibility between systems.*

VARIOUS COMPRESSION TECHNIQUES



LOSSY VS LOSSLESS

- **Lossy**
- Lossy data compression, which only allows an approximation of the original data to be reconstructed.
- Typically, there is some distortion between the original and reproduced data.
- Lossy compression is most commonly used to compress multimedia data (audio, video).
- Lossy compression will often have a Higher compression ratio
- DCT(Discrete Cosine Transform), FFT(Fast Fourier Transform) are used.
- Example: Joint Photographic Experts Group (JPEG) , Motion Picture Experts Group (MPEG), MP3

- **Lossless**
- Lossless data compression is a class of data compression algorithms that allows the exact original data to be reconstructed.
- The compressed-then-decompressed data is an exact replication of the original data.
- Lossless compression is , used in cases where it is important that the original and the decompressed data be identical.
- Lossless compression will often have a smaller compression ratio
- Huffman Coding, RLE(Run length Encoding), LZW(Lempel–Ziv–Welch) are used
- Example such as FLAC, PNG

LOSSLESS COMPRESSION

- **1. Huffman coding:**
  - Huffman Coding is a technique of compressing data to reduce its size without losing any of the details. It was first developed by David Huffman.
  - Huffman Coding is generally useful to compress the data in which there are frequently occurring characters.
  - Huffman coding works by assigning shorter codes to more frequently occurring characters and longer codes to less frequent characters.
- **2. Run Length Encoding**
  - Run-length encoding (RLE) is a lossless compression method where sequences that display redundant data are stored as a single data value.
  - RLE can make the file size bigger rather than smaller. Therefore it is important to understand the content and whether this algorithm will help or hinder compression.

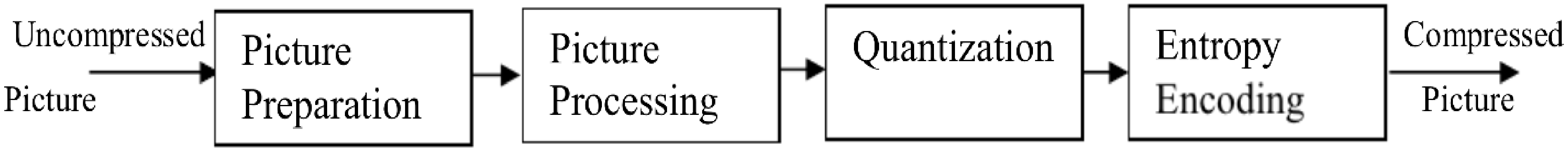
COMPRESSION TECHNIQUES IN MULTIMEDIA

Coding Type	Basis	Technique
Entropy Encoding	Run-length Coding	
	Huffman Coding	
	Arithmetic Coding	
Source Coding	Prediction	DPCM
		DM
	Transformation	FFT
		DCT
	Layered Coding	Bit Position
		Subsampling
		Sub-band Coding
	Vector Quantization	
Hybrid Coding	JPEG	
	MPEG	
	H.263	
	Many Proprietary Systems	

ENTROPY ENCODING, SOURCE CODING AND HYBRID CODING

- 1. Entropy Encoding
  - It is a lossless process, which is used regardless of the media's specific characteristics.
  - The data stream to be compressed is considered to be a simple digital sequence and the semantics of the data are ignored..
  - Decompression process regenerates the the data completely.
- 2. Source Coding
  - It takes into account the semantics of data and the degree of compression depends on the contents of the data.
  - A content prediction technique helps to reduce spatial redundancies. Using DCT, sorting can be done among data having higher and lower frequencies.
- 3. Hybrid Coding
  - This technique uses both the methods of entropy encoding and source coding for the compression of data.
  - This can be seen in JPEG, MPEG, H.261 (px64) compression.

GENERIC STEPS IN DATA COMPRESSION



STEPS

1. **Picture Preparation:**
   - This is the preprocessing step.
   - The image is prepared for compression, often by converting it into a suitable format or representation (e.g., converting RGB to YCbCr color space).
   - It may also involve filtering or dividing the image into smaller blocks for easier processing. An image is divided into blocks of 8x8 pixels, and represented by a fixed number of bits per pixel

2. **Picture Processing:**
   - The image is transformed into a format that is more efficient for compression.
   - Commonly, mathematical transformations like the **Discrete Cosine Transform (DCT)** or **Wavelet Transform** are applied to break the image into frequency components (e.g., separating low and high-frequency data).
   - This step helps distinguish important information (like smooth color changes) from less important details (like small variations or noise).

STEPS..CONTD

3. Quantization:

- This is the **lossy compression step**, where less important details are simplified or removed.
- The transformed data (from the previous step) is rounded to fewer levels, reducing precision and data size.
- For example, small color or brightness differences that are less noticeable to the human eye might be merged.

4. Entropy Encoding:

- This is a **lossless compression step** that encodes the quantized data into a compact format.
- Techniques like **Huffman Coding** or **Arithmetic Coding** are used to assign shorter codes to frequent data and longer codes to less frequent data.
- This reduces redundancy in the data representation.

5. Compressed Picture:

- The output is the final compressed version of the image, ready for storage or transmission.

JPEG COMPRESSION

- Why JPEG?
- Its implementation is independent of image size and appropirate to any image and pixel density.
- JPEG can be applied to color as well as greyscale image.
- It can handle large image content of any complexity.
- The JPEG standard specification is state-of-the-art regarding the compression factor and achieved image quality.
- The software can run on many avaliable standard processors.
- The user is able to select the quality of the reproduced image, the compression processing time and size of compressed image by choosing approporate parameters.

JPEG COMPRESSION

- In JPEG, the lossy compression is based on two psychovisual principles:

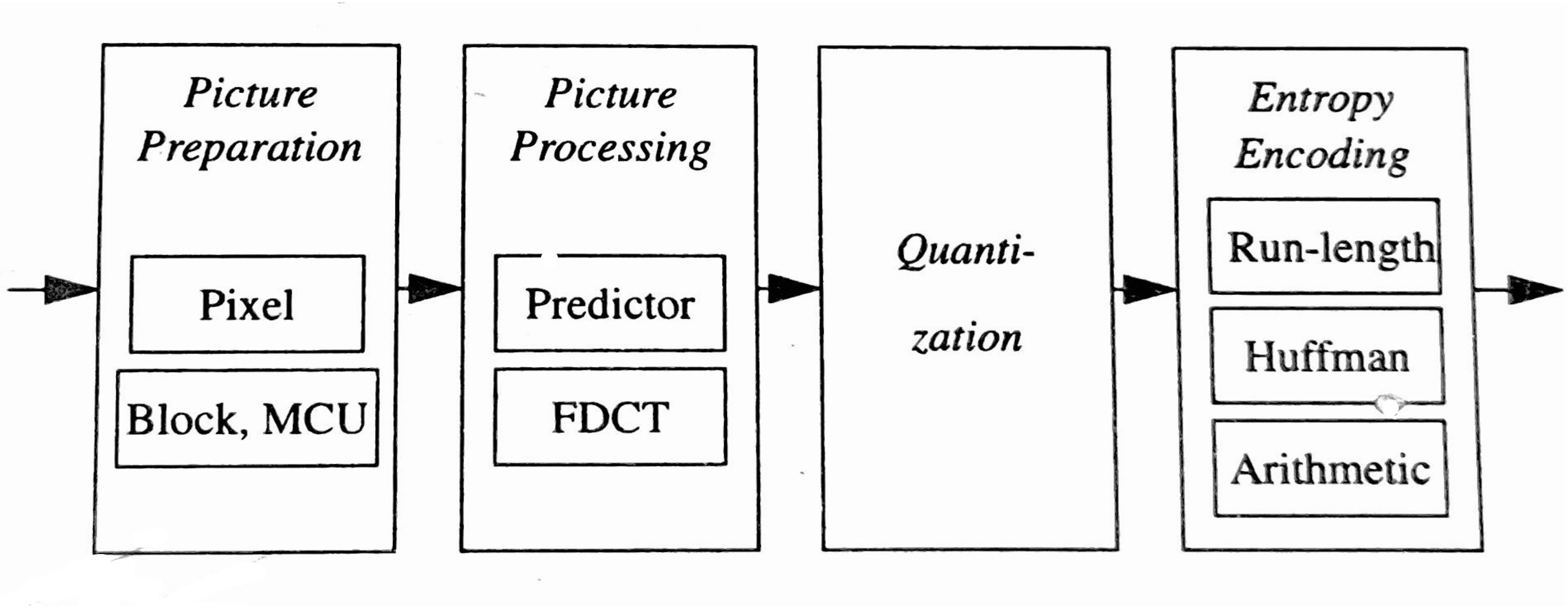
✓ Changes in brightness are more important than changes in color: the human retina contains about 120 million brightness-sensitive rod cells, but only about 6 million color-sensitive cone cells.

✓ Low-frequency changes are more important than high-frequency changes. The human eye is good at judging low-frequency light changes, like the edges of objects.

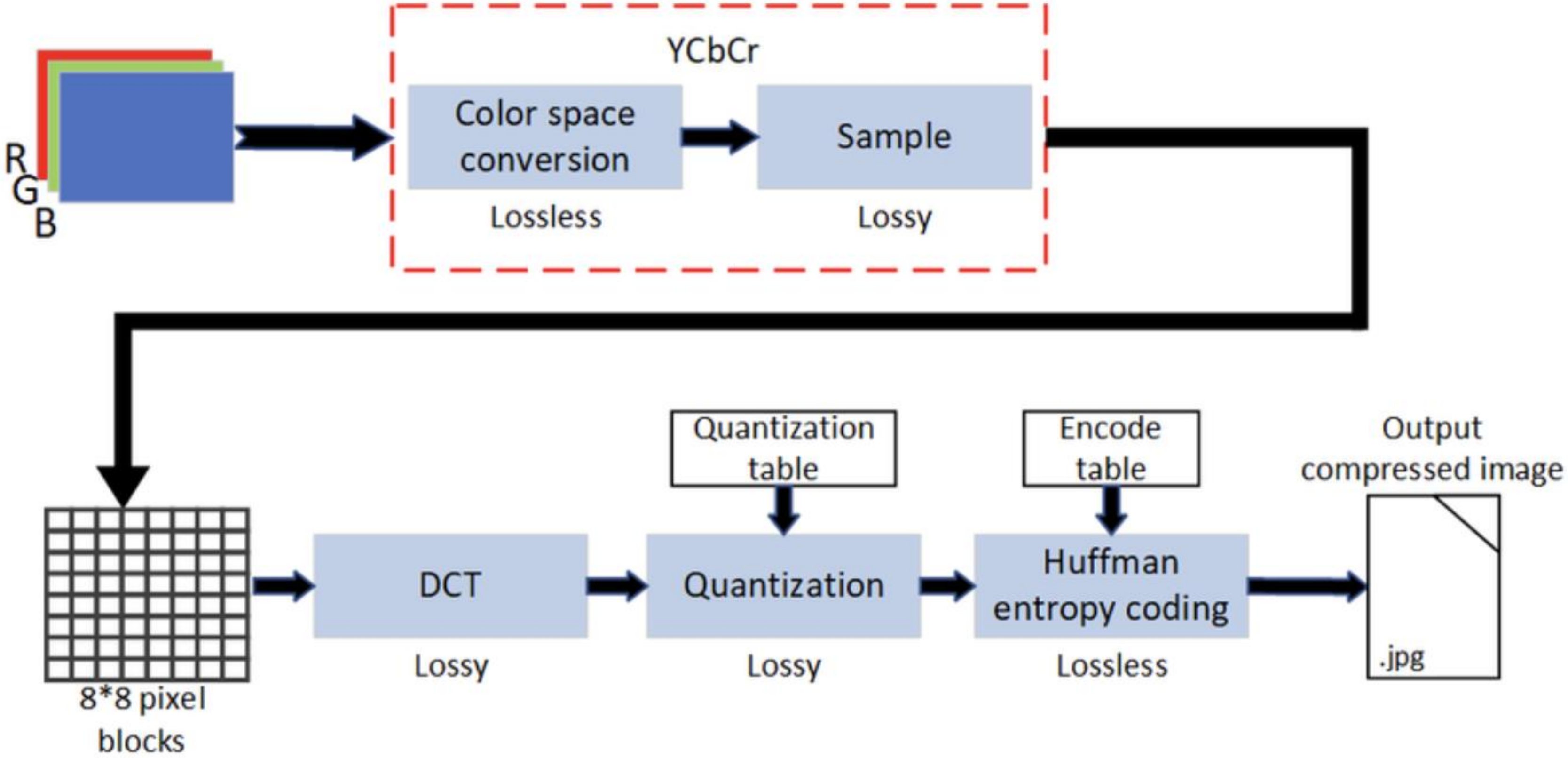
✓ It is less accurate at judging high-frequency light changes, like the fine detail in a busy pattern or texture.

JPEG AND H.261 COMPRESSION VIDEO

JPEG COMPRESSION

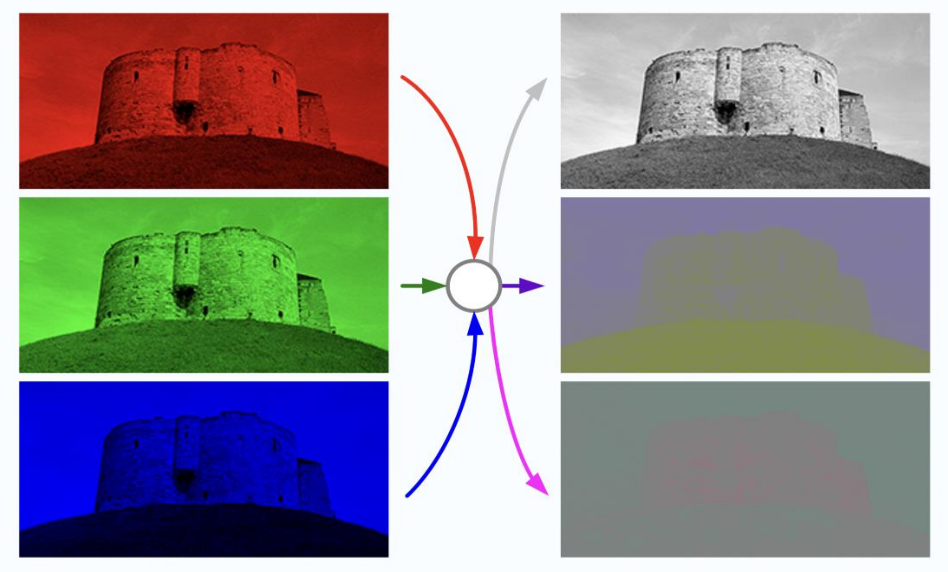


STEPS IN JPEG COMPRESSION



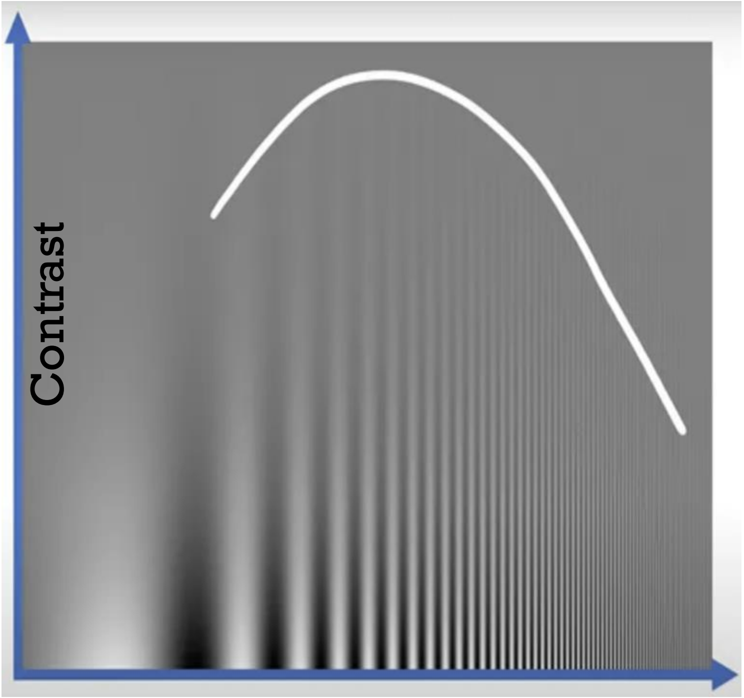
1. PICTURE PREPARATION

- **Color Space Conversion**
- In the first step of the JPEG image compression algorithm, the original image's color space is converted to YCrCb format.
- A YCbCr image also has three channels, but it stores all of the brightness information in one channel (Y) while splitting the color information between the other two (Cb and Cr).
- Throw away some color information. JPEG throws away some of the color information by scaling down just the Cb and Cr (color) channels while keeping the important Y (brightness) channel full size. **(Down sampling)**
- This is also called chroma subsampling.



2. PROCESSING

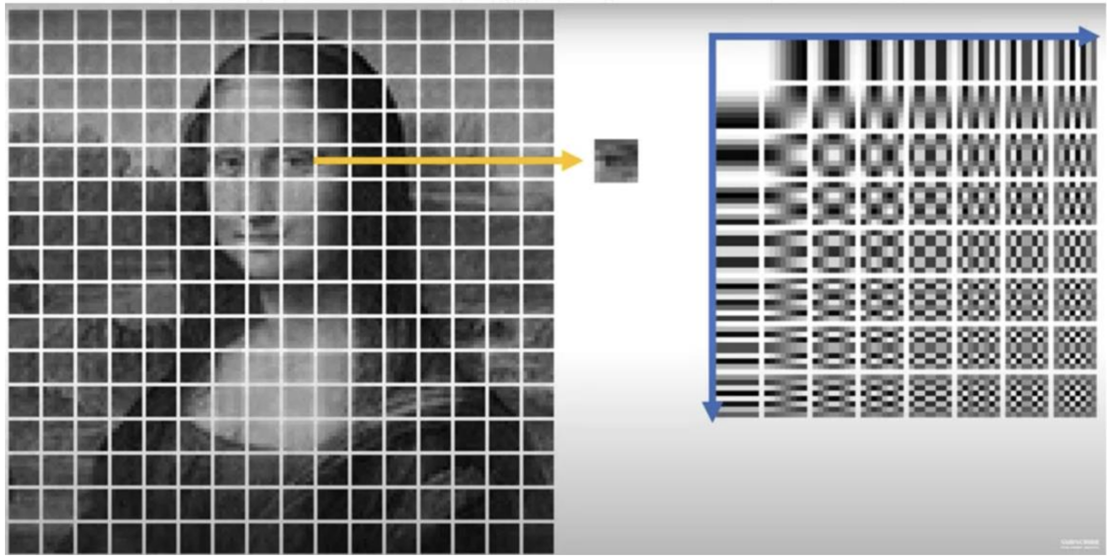
- The processing is carried out by dividing each of the Y, Cb, and Cr channels up into 8×8 blocks of pixels. Then transform each of these blocks from the spatial domain to the frequency domain.
- The bars in figure have increasing spatial frequency from left to right and decreasing contrast from bottom to top.
- Due to the ‘frequency-dependent contracts sensitivity’ phenomena, the bars under the curve (shown in Fig) are more visible than the rest.
- The bars which are located at the low-contrast, high-spatial frequency are barely visible.
- Therefore, those less visible frequencies give the JPEG algorithm some room for compression.



Frequency

8X8 BLOCKS AND APPLYING DCT

- The discrete cosine transformation is then applied to each of the 8x8 blocks.
- In this transformation, each 8x8 pixel is compared with 64 frequency patterns shown in first image where the spatial frequency increases from left to right and top to bottom.
- Ultimately each pixel in an 8x8 block which previously represented the brightness of each pixel gets transformed into another 8x8 block in which each pixel represents the particular frequency component (as shown in Fig below).

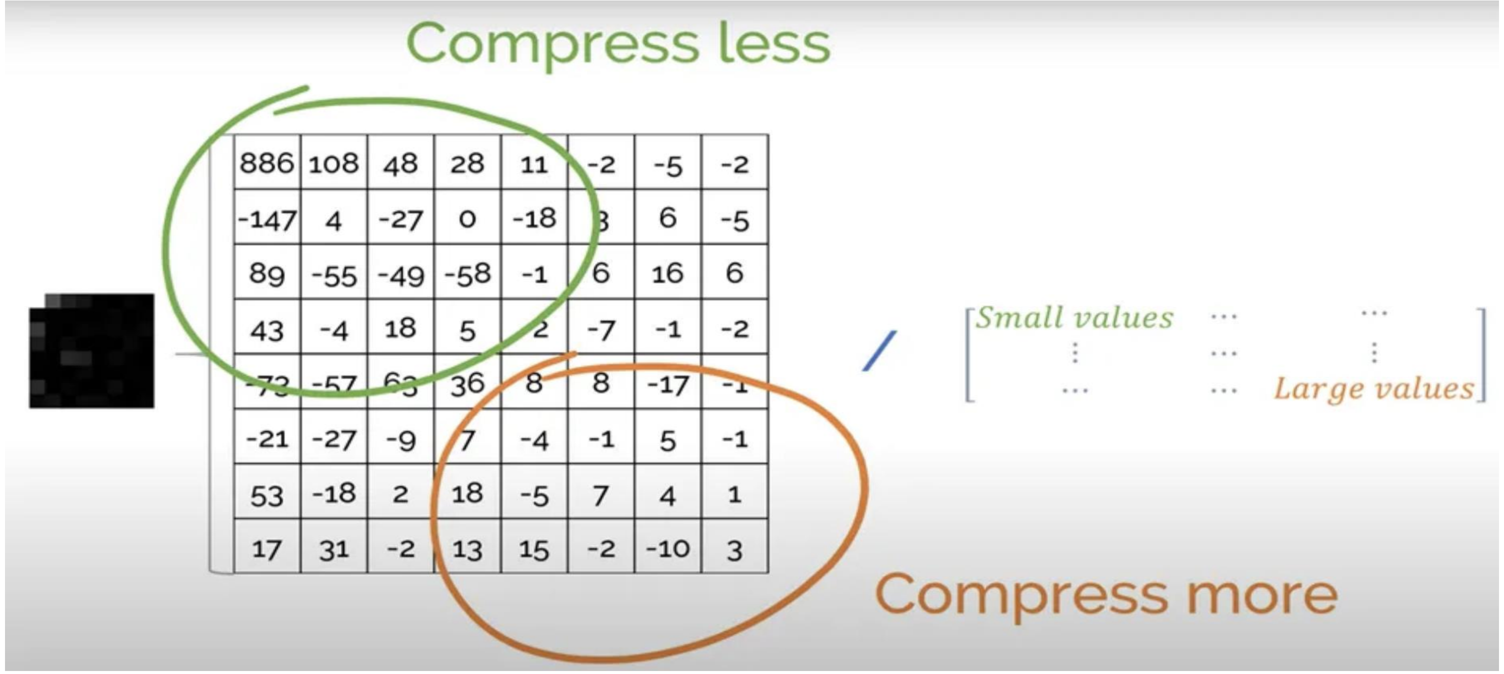


98	92	94	93	93	107	109	107
102	121	138	118	89	75	59	67
142	127	116	111	105	95	71	52
117	89	39	39	51	72	76	62
189	145	61	68	85	74	89	80
180	167	147	130	130	136	125	126
156	157	158	150	142	116	106	124
153	134	133	136	132	125	149	160

886	108	48	28	11	-2	-5	-2
-147	4	-27	0	-18	3	6	-5
89	-55	-49	-58	-1	6	16	6
43	-4	18	5	-2	-7	-1	-2
-73	-57	63	36	8	8	-17	-1
-21	-27	-9	7	-4	-1	5	-1
53	-18	2	18	-5	7	4	1
17	31	-2	13	15	-2	-10	3

3.QUANTIZATION

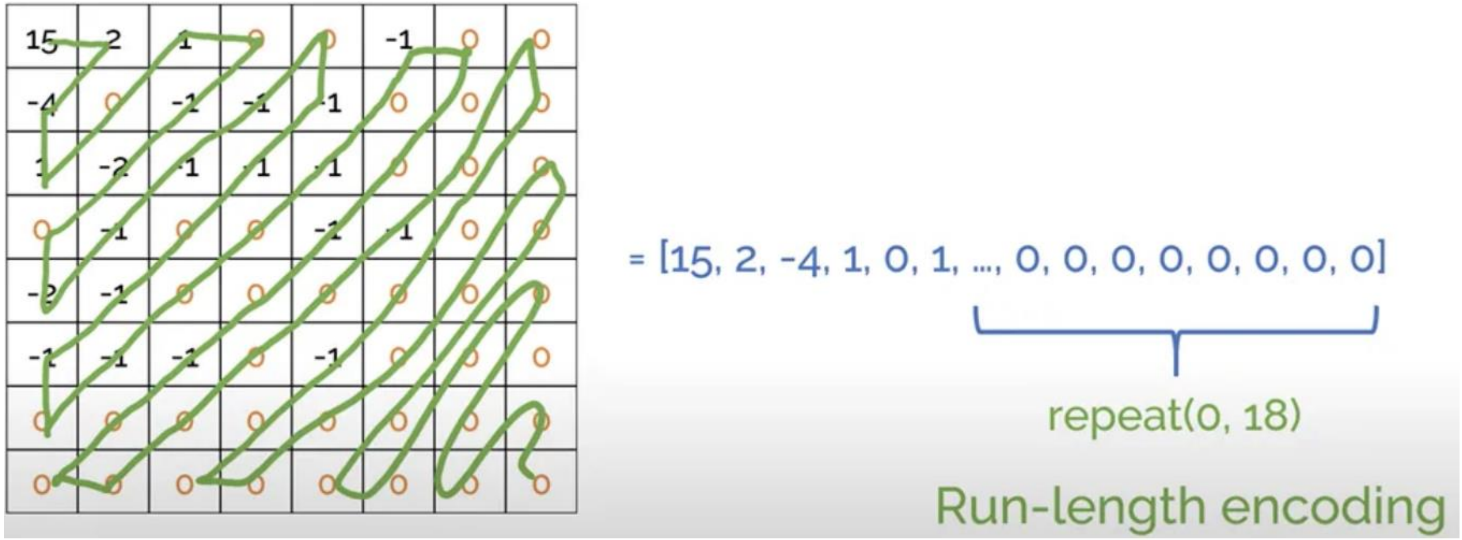
- The next step is quantization where the frequencies that are less visible to the human eye .(which is spread more towards the bottom right corner of the image as shown in Figure)
- It will be compressed more by dividing the frequency values with some constants and then quantizing them. **The human eye is not very good at perceiving high-frequency elements in an image.**
- To achieve this, the frequency components that are less sensitive to the human visual system get divided by larger constants as compared to the ones that humans are more sensitive to.



APPLY RUN-LENGTH

- After diving the result gets rounded to the nearest integers and this makes most of the high-frequency component (spread in the bottom right-hand corner) values become zero.
- The algorithm uses ‘run-length encoding’ to achieve efficient storage of the same amount of data in a block (without losing data).
- In this step, each data item in an 8x8 block is rearranged in a **zig-zag order** from top left to bottom right. This allows the grouping of the zeros together and storing only one zero element followed by the number of times that zero element occur in the zigzag order.

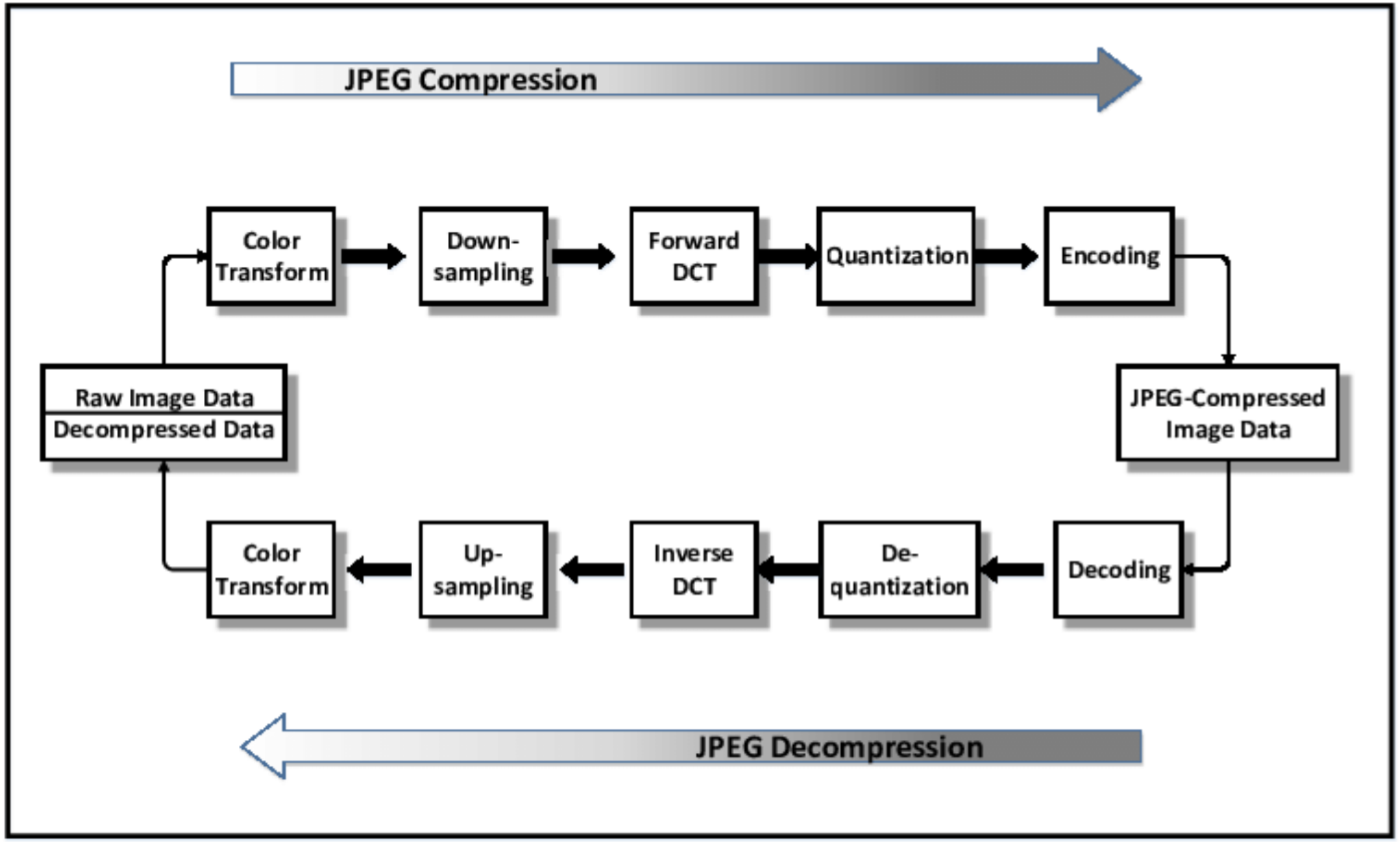
15	2	1	0	0	-1	0	0
-4	0	-1	-1	-1	0	0	0
1	-2	-1	-1	-1	0	0	0
0	-1	0	0	-1	-1	0	0
-2	-1	0	0	0	0	0	0
-1	-1	-1	0	-1	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0



HUFFMAN (ENTROPY) ENCODING

- As the final step of the JPEG compression algorithm, the Huffman encoding, which is an entropy encoding algorithm, is used to further compress the data that need to be stored.
- In this step, frequently occurring values get encoded with fewer bits while less frequent values with a greater number of bits. It helps to reduce the average number of bits needed per symbol.

COMPRESSION & DECOMPRESSION



H.261(PX64) COMPRESSION

- H.261 is an International Telecommunication Union (ITU-T) video coding standard introduced in 1988 that uses ISDN(Integrated Service Digital Network)
- H.261 employs lossy compression techniques to reduce the size of video data while maintaining an acceptable level of visual quality.
- The standard focuses on low-bit-rate video communication in applications such as video conferencing, remote surveillance, and multimedia communication.
- H.261 supports various resolutions, including **Common Intermediate Format** (CIF) and **Quarter Common Intermediate Format** (QCIF).
- Has a bit rate of 64Kbit/second hence the name px64.

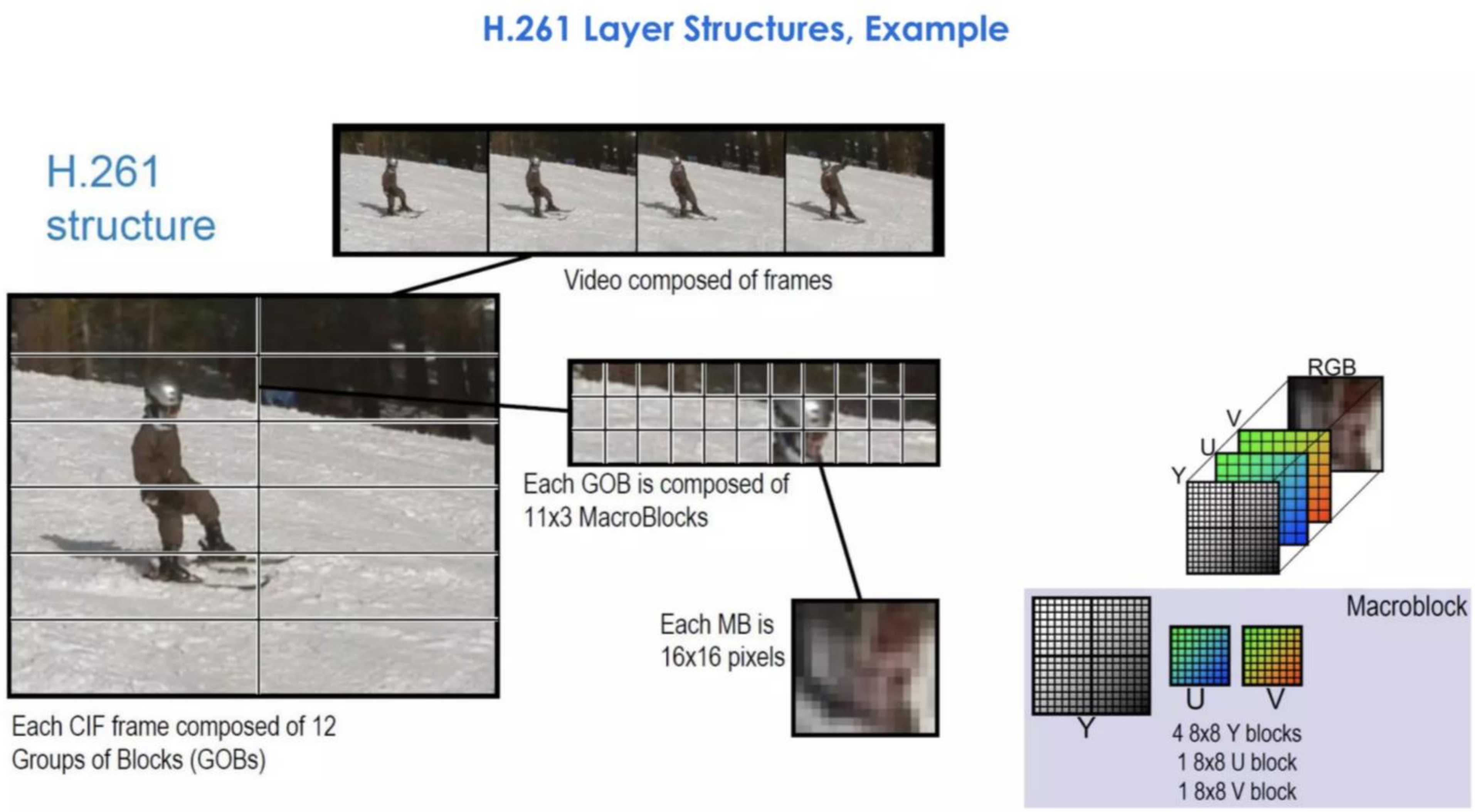
WHERE IS H.261 USED?

- H.261 has found extensive use in various domains where low-bit-rate video communication is critical. Some notable use cases include:
- **Video Conferencing** – H.261 was primarily developed for video conferencing systems. It efficiently transmits video data over narrowband ISDN connections, enabling real-time communication with acceptable video quality.
- **Surveillance Systems** – H.261's low-bit-rate capabilities and real-time performance made it suitable for remote surveillance systems. It allowed the transmission of video streams from multiple cameras with minimal bandwidth requirements.
- **Multimedia Communication over Low-Bandwidth Channels** – H.261 facilitated video communication in scenarios where network bandwidth is limited, such as early internet video streaming and teleconferencing applications.

IMAGE PREPARATION

- It uses a very precise format to work with i.e. 29 frames per second and the aspect ratio of 4:3
- As the changes between two consecutive frames is very less in most of the cases, the amount of information in the difference between those frames is a lot less.
- So the main idea behind any video compression is to send only the differences of the frames instead of sending whole frame.

H.261 STRUCTURE



H.261 STRUCTURE

- The H.261 codec is notable for being the first video codec to use macroblocks.
- Each macroblock in H.261 consisted of a 16x16 array of luma samples and two corresponding 8x8 arrays of chroma samples using 4:2:0 sampling and YCbCr color space.
- The CIF(Common Intermediate Frame) defines a luminance component of 288 lines with 352 pixels each.
- The chrominance components have a resolution with the rate of 144 lines and 176 pixels per line.
- QCIF(Quarter CIF) has exactly half of the CIF resolution. i.e. 176x144 pixels for luminance and 88x72 pixels for other components.

INTRA-FRAME AND INTER-FRAME

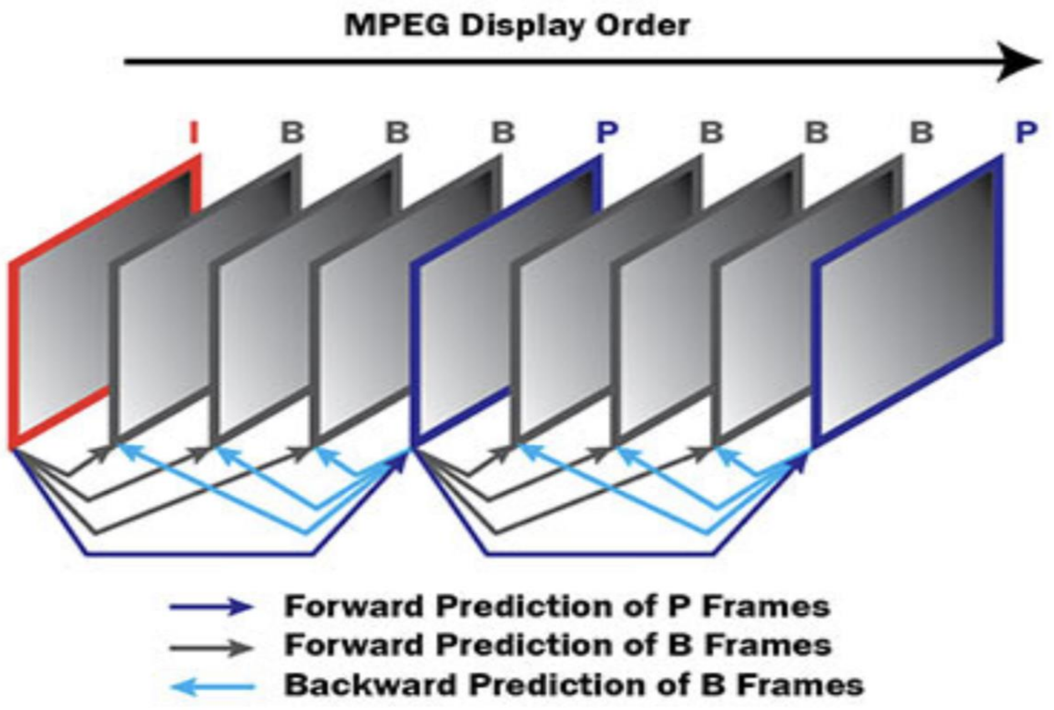
- **Intraframe Coding** and **Interframe Coding** are two fundamental techniques used in video compression to reduce the size of video files while maintaining visual quality.
- 1. Intraframe Coding:
  - Intraframe coding compresses each video frame independently, treating it as a still image. Compression is achieved using spatial redundancy within a single frame. No advantage is taken from the redundancy between frames
  - Similar to JPEG where each block of 8x8 pixels was transformed into 64-coffecients of DCT and entropy encoding is applied further.
- 2. Interframe Coding:
  - Interframe coding compresses video by exploiting temporal redundancy between consecutive frames, predicting the content of one frame based on others.
  - The prediction is based out of each macroblock of an image by comparing it with previous macroblock.

MPEG COMPRESSION

- It stands for Motion Picture Expert Group.
- It is a standard compression technique for encoding videos, movies.
- It is a lossy video compression technique where higher levels of compression can be achieved.
- Unlike H.261 compression this uses a more advanced interframe coding strategy with I-frames, P-frames, and B-frames (bidirectionally predicted frames).
- B-frames enhance compression by predicting from both previous and future frames, offering higher efficiency at the cost of increased computational complexity.
- Supports a wide range of resolutions:
- MPEG-1: Up to 352x240 (VCD-quality).
- MPEG-2: Higher resolutions, up to 720x480 (SD) and 1920x1080 (HD).

MPEG PREDICTION AND FRAMES

- 1. Forward Prediction: Prediction Based on Previous frames
- 2. Backward prediction: Prediction Based on Future frames
- 3. Bi-directional Prediction: Based on both



MPEG FRAMES

- MPEG is broken down into GOP(Group of Pictutres)
- GOPs consists of I-frame, B-frame and P-frames.
- Initially when a scene is taken, it is constructed using I-frame(Intra-coded frame)
- An I-frame basically consists of a entire first picture.
- Then the codec waits and counts for example 4 frames into the future where it will take another snapshot. This snapshot is regarded as P-frame(Predictive coded)
- The P-frame only contains the information which is different from the point in time where I-frame was taken.
- So the P-frame would contain 40-50% of the information from I-frame

CONTD..

- The B-frames(Bi-directionally coded) are basically the predictions of how objects move across the scene. This prediction is carried out by using motion vectors where each motion vector points to the best match for the block in either the past or future frame.
- Only by using B-frame compression can a high compression of 200:1 be achieved. Generally, I- frame has the lowest compression efficiency, P frame is higher, and B frame is the highest.
- Finally the images are converted into the frequency domain and quantized.
- The RLE algorithm is applied to those quantized values and Huffman encoding is used for further compression just like in JPEG.

DIFFERENCE

Feature	H.261 🟢	MPEG (MPEG-1/2) 🔵
Developer	ITU-T	MPEG (ISO/IEC)
Purpose	Video conferencing over ISDN	Digital storage, broadcasting, entertainment
Compression Efficiency	🔴 Lower	🟢 Higher
Frame Types	I-frames, P-frames	I-frames, P-frames, B-frames (bidirectional)
Resolution	CIF (352x288), QCIF (176x144)	VCD to HD (e.g., 352x240, 720x480, 1920x1080)
Motion Compensation	Simple: Integer-pixel accuracy	Advanced: Sub-pixel accuracy (e.g., half-pixel)
Applications	Real-time communication	Entertainment, streaming, and archival storage
Audio Support	🚫 Not integrated	🎵 Integrated (e.g., MP3 for MPEG-1/2 audio)

MODES IN JPEG

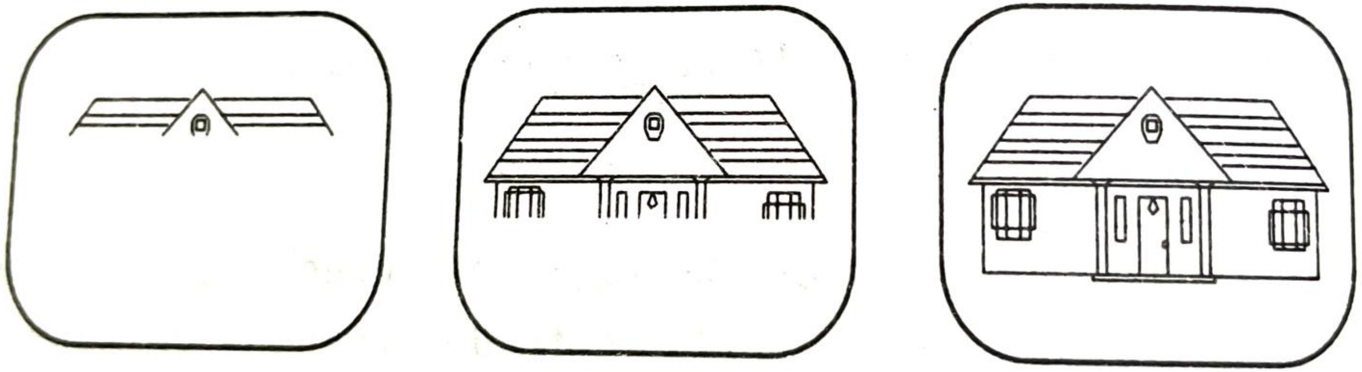
Lossy Sequential DCT model

- This is the basic and most commonly used mode in JPEG compression.
- The image is divided into small blocks (typically 8x8 pixels).
- Each block is processed one at a time, in order (sequentially).
- The compression happens sequentially, one block at a time, without overlapping blocks or additional enhancements.
- The DCT is applied to transform the pixel values into frequency components.
- The decoder reads the compressed file in one pass and reconstructs the entire image block by block.

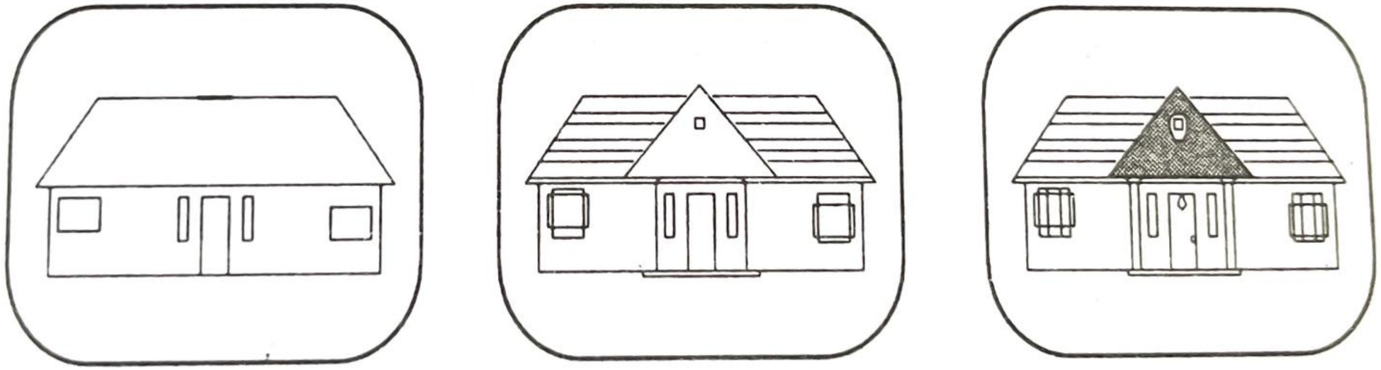
Expanded Lossy DCT model

- This mode expands on the basic lossy sequential mode to offer more advanced features for specific needs.
- It still uses DCT, but with added flexibility or customization options.
- It allows for non-sequential processing of blocks, support for progressive encoding.
- First scan: Encodes the most significant frequency components (low frequencies) for the entire image. This creates a rough, blurry version of the image.
- Subsequent scans: Add more detail by encoding finer frequency components (high frequencies) in stages.
- With each pass, the image becomes clearer as more detail is added.

MODES IN JPEG

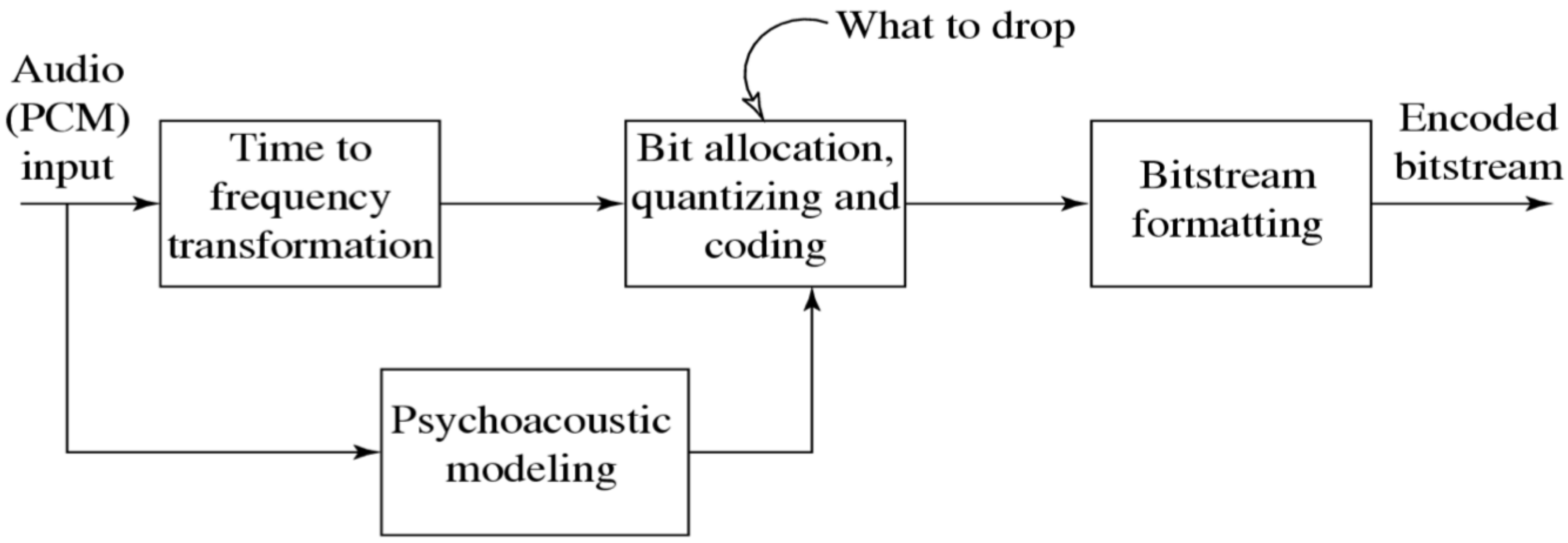


Lossy Sequential
DCT model



Expanded Lossy
DCT model

AUDIO ENCODING USING MPEG



(a) MPEG Audio Encoder

AUDIO ENCODING USING MPEG

- MPEG audio compression takes advantage of psychoacoustic models, constructing a large multi- dimensional lookup table to transmit masked frequency components using fewer bits.
- The input audio stream passes through a filter bank to break it into its frequency components.
- In parallel, the input audio stream simultaneously passes through a psychoacoustic model.
- The bit or noise allocation block uses the signal-to-mask ratios to decide how to assign the total number of code bits available for the quantization of the sub-band signals to minimize the audibility of the quantization noise.
- Finally, the last block takes the representation of the quantized audio samples and formats the data into a decodable bit stream.

END OF CHAPTER 5

- *Multimedia: Computing, Communications and Applications", Ralf Steinmetz and Klara Nahrstedt, Pearson Education Asia*
- *"Multimedia Communications, Applications, Networks, protocols ad Standards", Fred Halsall, Pearson Education Asia*
- *"Multimedia Systems", John F. Koegel Buford, Pearson Education Asia*