MULTIMEDIA CODING

ESEIAAT- UPC

Course presentation

Unit 1

Introduction

Verónica Vilaplana (Slides)
Albert Mosella-Montoro (Lecturer)

Introduction outline

- Motivating multimedia compression
- How does compression work?
  - Redundancy and irrelevance
- Lossless and lossy compression
  - Compression efficiency
  - Quality measures
- Coding systems
  - Importance of standards

Motivating multimedia compression

Motivating multimedia compression

- **Uncompressed multimedia** (graphics, audio and video) data requires considerable storage capacity and transmission bandwidth.

- Despite rapid progress in mass-storage density, processor speeds, and digital communication system performance, **demand for data storage capacity and data-transmission bandwidth** continues to outstrip the capabilities of available technologies.

- The continuous growth of **data intensive multimedia-based applications** have not only sustained the need for more efficient ways to encode signals and images, but have made compression of such signals central to storage and communication technology.

Motivating multimedia compression

- Size of text
  - For an average word
    - 6 characters/word, 7 bits/character: 42 bits ~= 5 bytes
- Size of a typical color image
  - For display
    - 640 x 480 x 24 bits = 7.372.800 bits = 92160 bytes ~ 0,9 MB
  - For current mainstream digital camera (12 Mpixels)
    - 4032 x 3024 x 24 bits = 292.626.432 bits = 36.578.304 bytes ~35 MB
- Bit rate: bits per second for storage
  - CD Music
    - 2 channels x 44100 samples/second x 16 bits/sample ~ 1.4 Mbps
      4min music: 338 Mb
  - HDTV digital video
    - 1920 x 1080 x 24 bits/pixel x 24 frames/second: ~ 1.2 Gbps
      1 hour TV episode: 4300 Gb = 4,3 Tb ~500 GB
  - 4K UHDTV
    - 3840 x 2160 x 24 bits/pixel x 24 frames/second: ~ 4.8 Gbps
      1 hour TV episode: 17200 Gb = 17 Tb ~ 2,1TB
  - 8K
    - 7680 x 4320 x 30 bits/pixel x 50 frames/second: ~ 47 Gbps
      1 hour TV episode: 21 TB

Your average Internet connection: 100~500 Mbps (0.1~0.5 Gbps)!

Introduction outline

- Motivating multimedia compression
- How does compression work?
  - Redundancy and irrelevance
- Lossless and lossy compression
  - Compression efficiency
  - Quality measures
- Coding systems
  - Importance of standards

How does compression work?

Two fundamental components of compression are redundancy and irrelevance reduction.

- **Redundancy reduction** aims at removing duplication from the signal source (text/ image/ video/ audio).

How does compression work?

Two fundamental components of compression are redundancy and irrelevance reduction.

- **Redundancy reduction** aims at removing duplication from the signal source (text/ image/ video/ audio). In general, three types of redundancy can be identified:
  - **Spatial Redundancy** or correlation between neighboring values.
  - **Inter-channel Redundancy** or correlation between different color channels, spectral bands, or audio channels.
  - **Temporal Redundancy** or correlation between adjacent frames in a sequence of images (in video applications), adjacent samples in audio.

How does compression work?

Two fundamental components of compression are redundancy and irrelevance reduction.

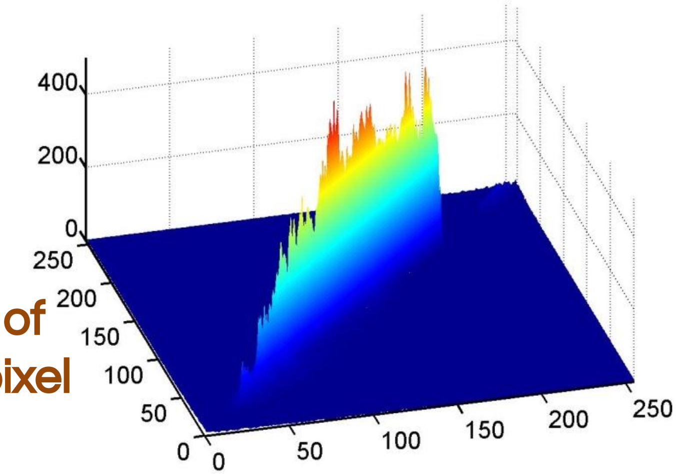
- **Irrelevance reduction** omits parts of the signal that will not be noticed by the signal receiver
  - The human auditory system (**HAS**) or
  - The human visual system (**HVS**)

Spatial Redundancy: image

- In the spatial domain, there is usually a high correlation **between pixels** (samples) that are close to each other, i.e. the values of neighboring samples are often very similar.

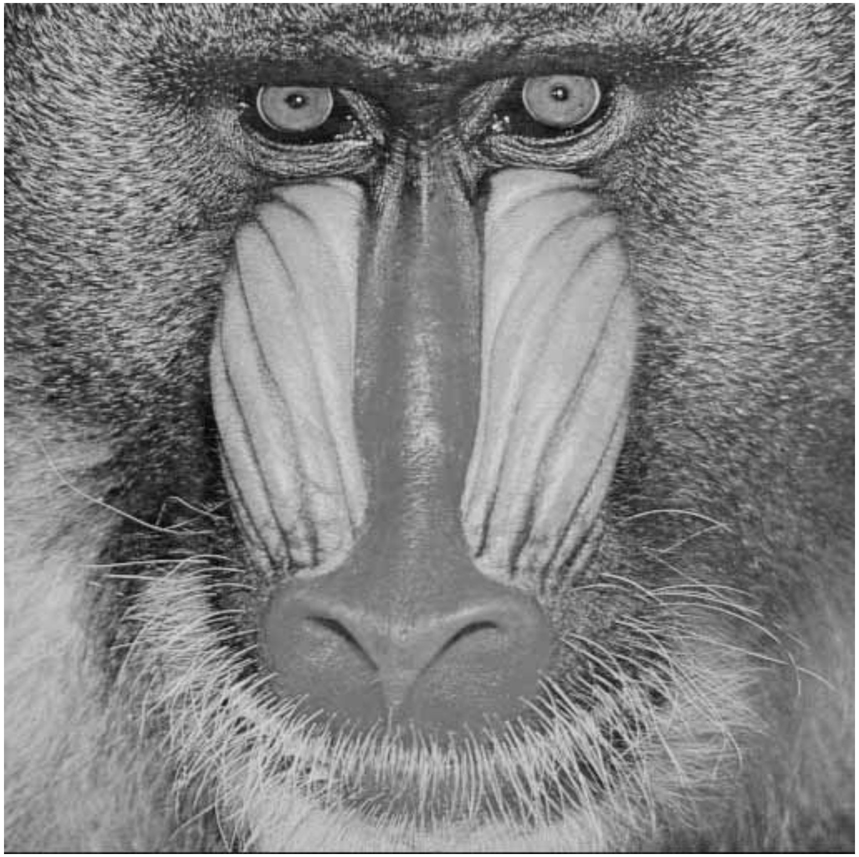
Joint histogram of two horizontal adjacent pixels

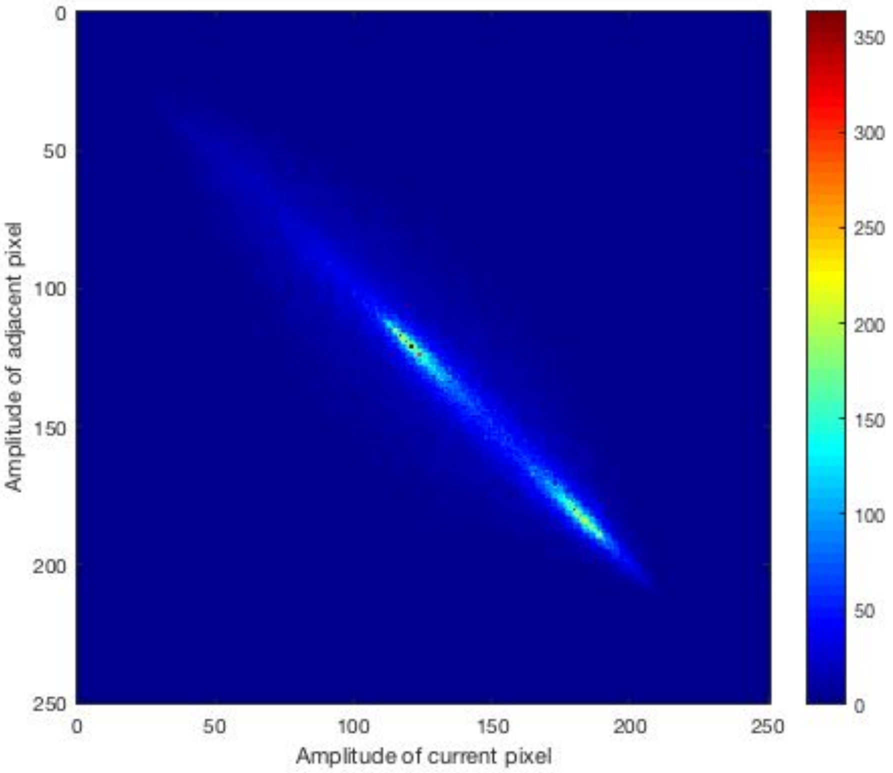
Gray level image, 8bbp



Spatial Redundancy: image

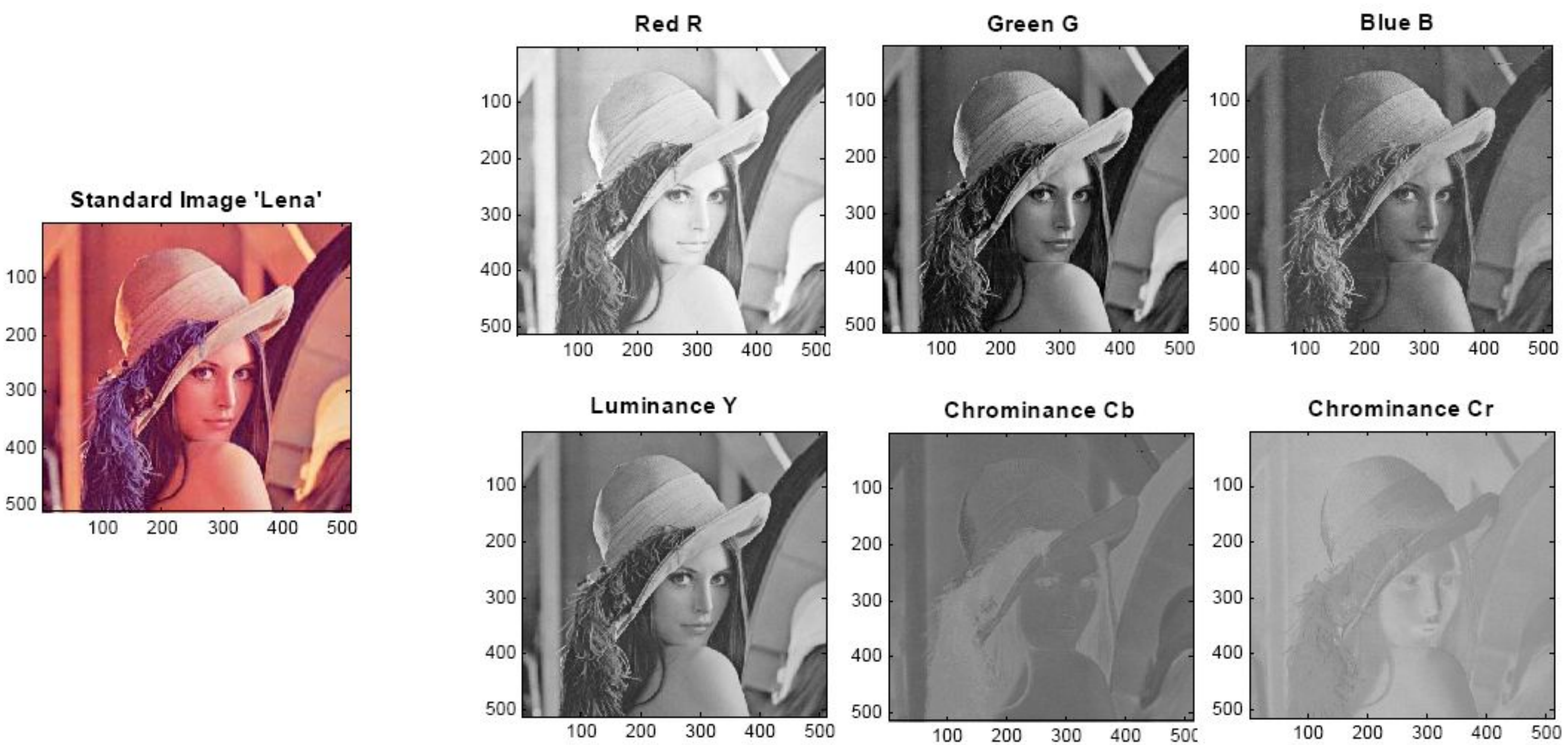
Ex. Baboon (lab1)





Inter-channel redundancy: image

- There is usually a high correlation between **color image components**

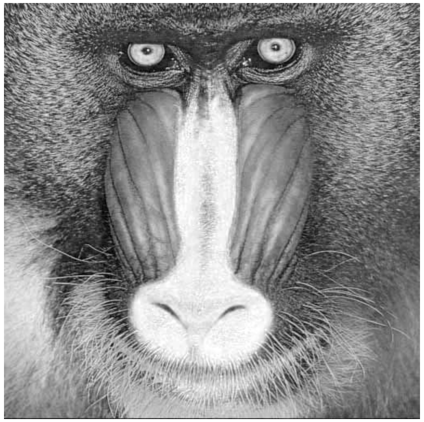


Statistical dependence between R,G,B is stronger

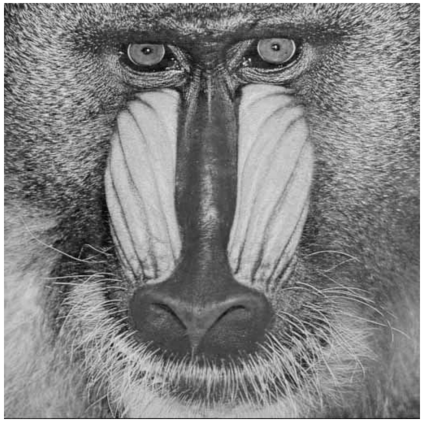
Spatial Redundancy: image

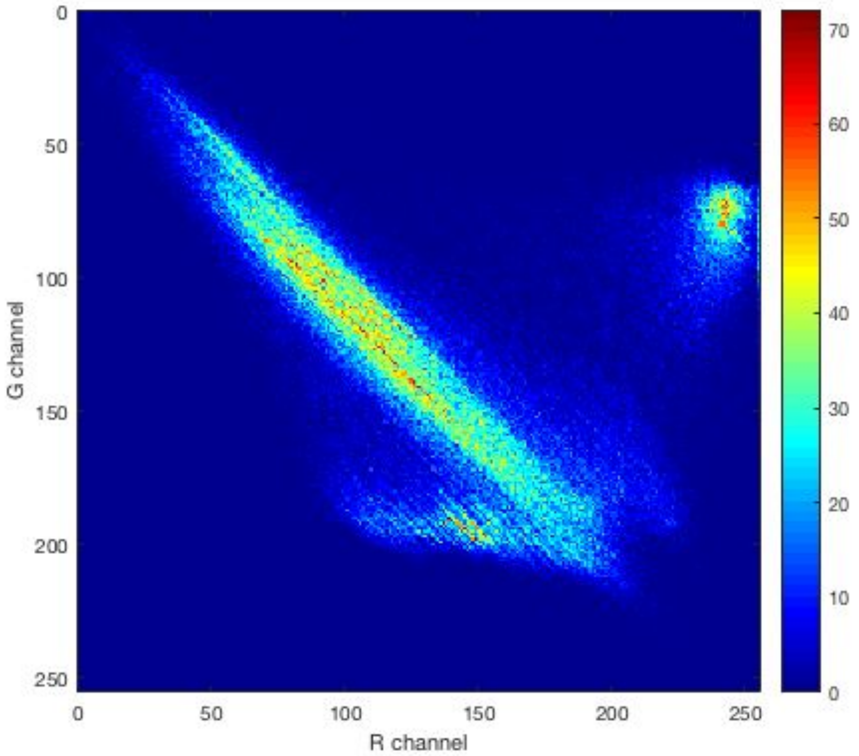
Ex. Baboon (lab1)

R channel



G channel





Temporal Redundancy: video

- In the temporal domain, there is usually a high correlation (similarity) **between frames** of video that are captured at around the same time, especially if the temporal sampling rate (the frame rate) is high.

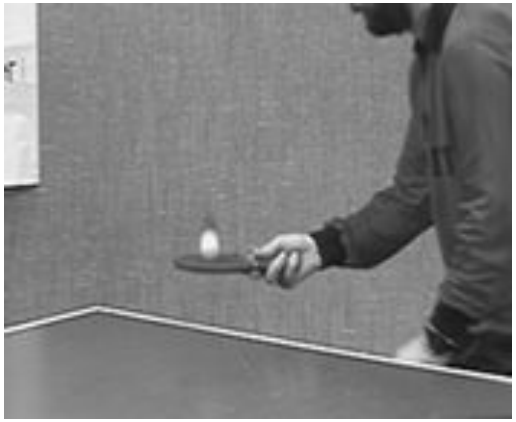


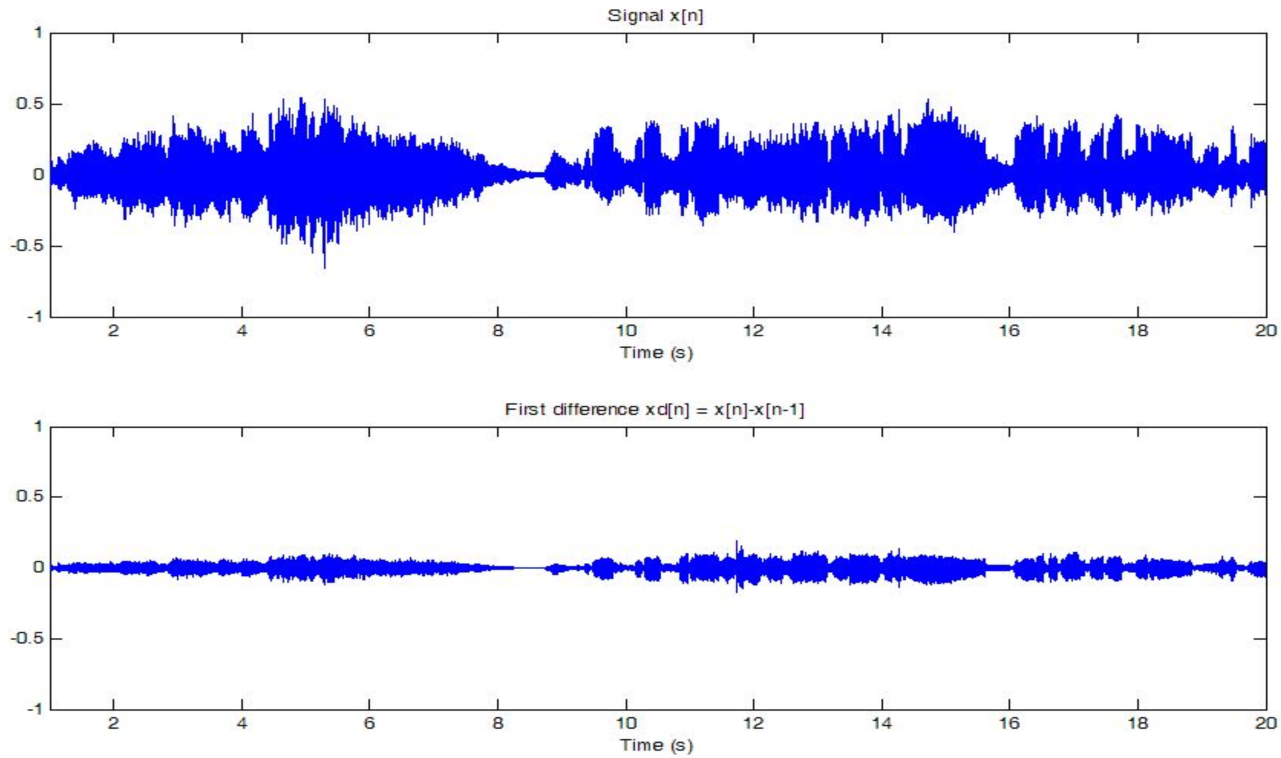
Table tennis sequence frame #40

Table tennis sequence frame #41

Table tennis sequence frame #42

Temporal Redundancy: audio

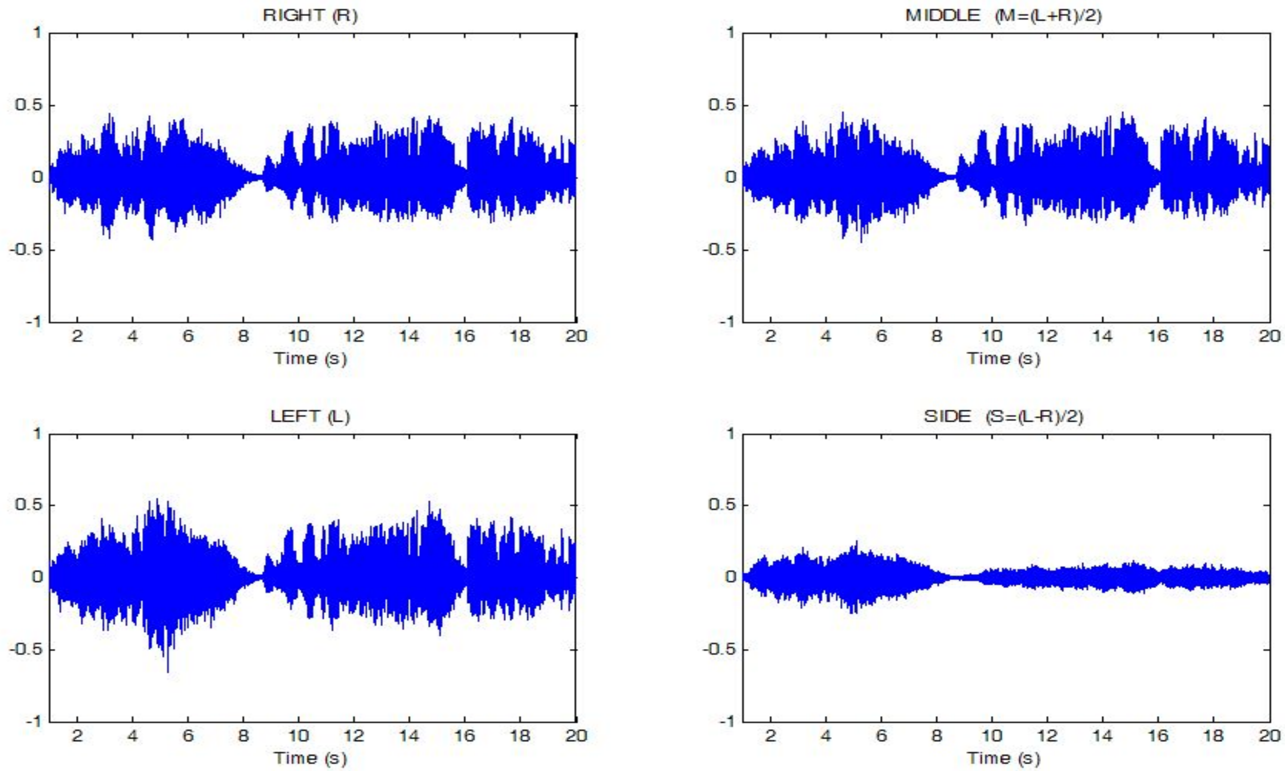
- Audio signals with high harmonic contents have a strong correlation between consecutive samples.



Found in voiced sounds of speech and many sounds from musical instruments
Noisy sounds (like unvoiced sounds of speech) show much less correlation

Inter-channel Redundancy: audio

- For some stereo audio signals, there is a strong correlation between channels.



Ex. recorded sounds coming from the front (as in live concerts)

Irrelevance

- For audio, images, video signals to be listened or viewed by humans, no need to represent more than the **information that can be perceived** in
  - Space
  - Time
  - Brightness
  - Color
  - Loudness
- The required resolution might **depend on audio/image content** ("masking")
- Irrelevance can be **application dependent**:
  - Ex: only a specific region of the image might be relevant for some task, e.g. in medical imaging or military imaging. Regions of the image which do not contribute to the task may be taken to be irrelevant.

Exploiting limitations of HVS

- Human visual system has **much lower acuity for color hue and saturation** than for brightness
- Use color transform to facilitate exploiting that property

Luminance component

Chrominance components {

$$\begin{pmatrix} x_Y \\ x_{Cb} \\ x_{Cr} \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.813 \end{pmatrix} \cdot \begin{pmatrix} x_R \\ x_G \\ x_B \end{pmatrix}$$

RGB-components

- Cb and Cr often sub-sampled 2:1 relative to Y

Y

Cb

Cr

Example: JPEG2000 compression

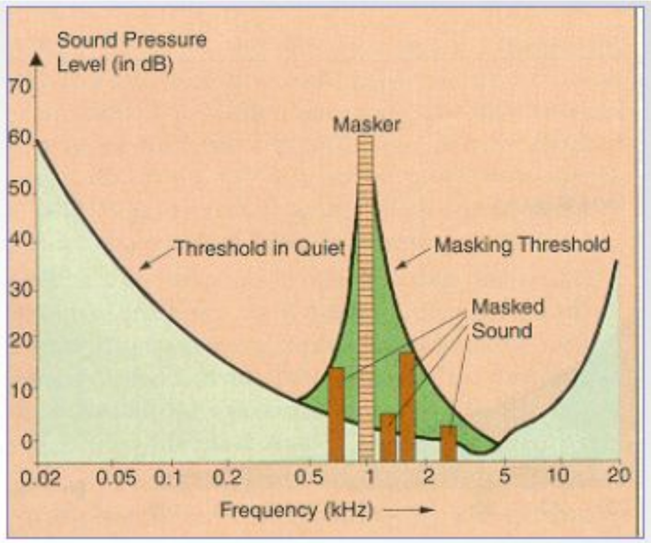
512x512 pixels, 24 bits/pixel RGB

Uncompressed
786 Kbytes

75:1 compression
10.6 Kbytes

Exploiting limitations of HAS

- Human ear can only hear sound louder than a frequency dependent threshold (20Hz-20kHz)
- **Masking**: some signal components are irrelevant to human perception
  - Depends on spectral characteristics, temporal characteristics and intensity

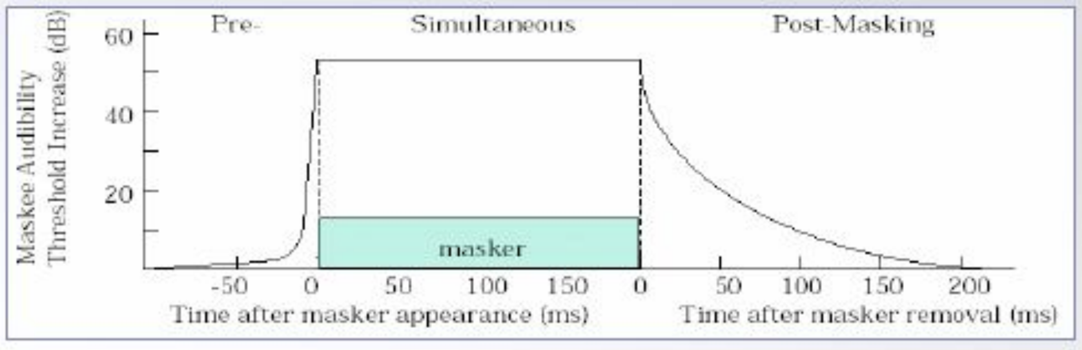


1. Can occur **simultaneously** with the masking signal **(frequency masking)**
   - Loud sounds mask soft ones, louder sound happens at the same time

Exploiting limitations of HAS

- **Masking:** some signal components are irrelevant to human perception
  - Depends on spectral characteristics, temporal characteristics and intensity

2. Can occur **before or after** the masking signal **(temporal masking)**
   - The auditory system needs an integration time which is higher for lower sounds



Signals which are not audible do not need to be transmitted

Perceptual lossless: assign number of bits so that the quantization noise is not audible

Introduction outline

- Motivating multimedia compression
- How does compression work?
  - Redundancy and irrelevance
- Lossless and lossy compression
  - Compression efficiency
  - Quality measures
- Coding systems
  - Importance of standards

Lossless and lossy compression

- **Lossless compression:**
  - The reconstructed data at the output of the decoder is a perfect copy of the original data.
  - Only **statistical redundancy** is exploited.
  - Lossless compression of image, video, audio gives only a **moderate amount of compression.**
  - Useful for file compression (rar, zip, arj), medical or satellite images

- **Lossy compression:**
  - The decompressed data is not identical to the source data.
  - Both **statistical and visual/psychoacoustic redundancy** are exploited.
  - Much **higher compression ratios** can be achieved at the expense of a loss of visual/audio quality.

Performance

- How to evaluate the performance of a compression algorithm?

- Performance is based on several factors:
  - Algorithm complexity
  - Memory requirements
  - Run time
  - **Compression efficiency**
  - **Quality measures**
    - Objective measures
    - Subjective measures

Efficiency (Measures of compression)

- If b and b' are the number of bits (or information-carrying units) used in two representations of the same information,
- **Compression ratio:** $$C = \frac{b}{b'}$$
  - **Example:** grayscale image of size 256x256 at 1 byte per pixel. If the compressed image occupies 16384 bytes, the compression ratio is 4

$$C = \frac{256\text{x}256\text{x}8}{16384\text{x}8} = \frac{65536}{16384} = \frac{4}{1} = 4$$

- **Compressed bit rate:** number of bits used to represent each sample (for images, in **bpp**) or number of bits used per second (for audio/video, in **bps** o **b/s**)
  - **Example:** for the previous image, the compressed bit rate is 2 bpp (bits per pixel)

$$\frac{16384\text{x}8}{256\text{x}256} = \frac{131072}{65536} = 2 \;\; bpp$$

Typical compressed bit rates

- **Color Image**
  - Lossless: 3 bpp (uncompressed 24 bpp)
  - Lossy quality:
    - Usable: 0.25 bpp
    - Moderate: 0.5 bpp
    - High: 1 bpp
- **Video (MPEG2/H.264)**
  - 16 Kb/s videophone quality (minimum necessary for a consumer-acceptable "talking head" picture)
  - 128 – 384 Kb/s business-oriented videoconferencing system quality
  - 1.25 Mb/s VCD quality
  - 5 Mb/s DVD quality (uncompressed 190 Mbps)
  - **15 Mb/s HDTV quality (uncompressed 1.2 Gbps)**
  - 36 Mb/s HD DVD quality
  - 54 Mb/s Blu-ray Disc quality
  - 36 Mb/s 4K UHDTV (uncompressed 4,8 Gbps)
  - 91 Mb/s 8k UHDTV-2 (uncompressed 18~46 Gbps)

Typical compressed bit rates

- **Audio (MP3)**
  - 32 Kb/s MW (AM) quality
  - 96 Kb/s FM quality
  - 128–160 Kb/s Standard Bitrate quality
  - 192 Kb/s DAB (Digital Audio Broadcasting) quality
  - 224–320 Kb/s Near CD quality

- **Other audio formats**
  - 800 b/s minimum necessary for recognizable speech
  - 8 Kb/s telephone quality
  - 0.5 – 1 Mbit/s lossless audio
  - 1411 Kb/s PCM sound format of Compact Disc Digital Audio

Quality measures: image and video

- Objective measures

  - Mean Square Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=0}^{N-1}\left|X_i - \hat{X}_i\right|^2$$

  - Mean Absolute Difference (MAD)

$$MAD = \frac{1}{N}\sum_{i=0}^{N-1}\left|X_i - \hat{X}_i\right|$$

  - Max Error

$$MaxError = \max_i\left\{\left|X_i - \hat{X}_i\right|\right\}$$

  - Peak Signal-to-Noise Ratio (PSNR)

$$PSNR = 10\log_{10}\frac{M^2}{MSE};$$

PSNR is useful to compare images with different dynamic ranges:
M = Maximum peak-to-peak value of the representation

Quality measures: image and video (II)

- Subjective measures

  - **Group of observers- ITU-RStandard BT 500**
    - Time consuming, expensive, off-line

  - **Model the fundamental properties of Human Visual System**
    - Several attempts, **no universal solution yet**
    - Lower levels of processing (eye imaging system, retina sampling, spatial masking, etc.) can be modeled, but higher level in the brain visual cortex is still to complex for detailed description
    - Commercial systems, like PQA500 (Tektronix)

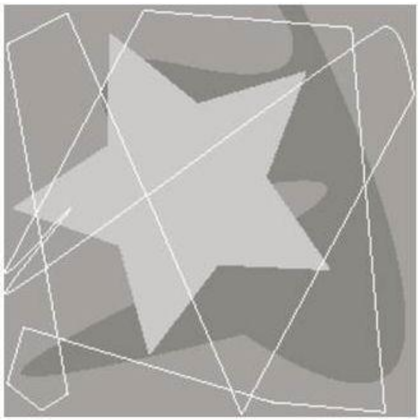
Quality measures: image and video (III)

- **Example of subjective quality measure:** absolute rating scale of the Television Allocation Study Organization

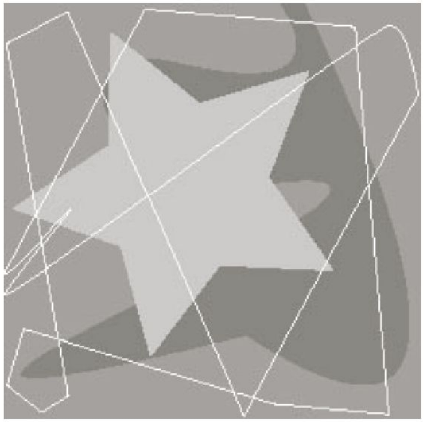
Value	Rating	Description
1	Excellent	An image of extremely high quality, as good as you could desire.
2	Fine	An image of high quality, providing enjoyable viewing. Interference is not objectionable.
3	Passable	An image of acceptable quality. Interference is not objectionable.
4	Marginal	An image of poor quality; you wish you could improve it. Interference is somewhat objectionable.
5	Inferior	A very poor image, but you could watch it. Objectionable interference is definitely present.
6	Unusable	An image so bad that you could not watch it.

R. González and R. Woods, *Digital Image Processing*, Prentice-Hall, 2008.

Quality meas. : subjective vs. objective

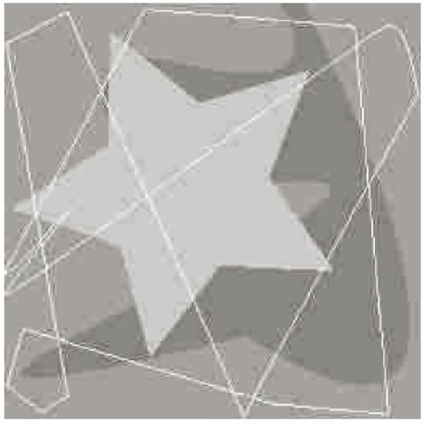


Original



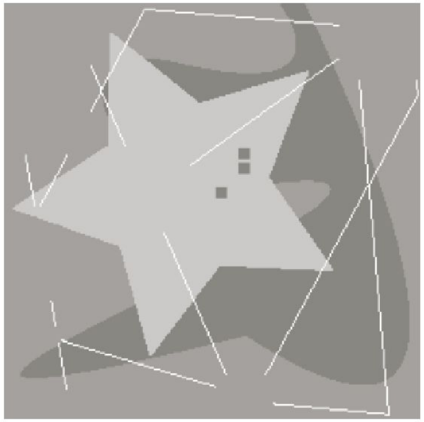
Compressed and reconstructed, very good quality

MSE= 5.17



Compressed and reconstructed, noticeable degradation

MSE = 15.67



Missing visual information and with artifacts

MSE = 14.17

Quality measures: audio

- Audio coders are evaluated in terms of bit rate, complexity, delay, robustness and output quality.
  - **Objective measures** exist for computing bit rate, complexity and delay performance.
  - For quality or robustness only **subjective measures** can be applied.

- In current perceptual audio coding, uncompressed waveforms can be very different from the originals, but still they can sound quite similar.

Quality measures: audio (II)

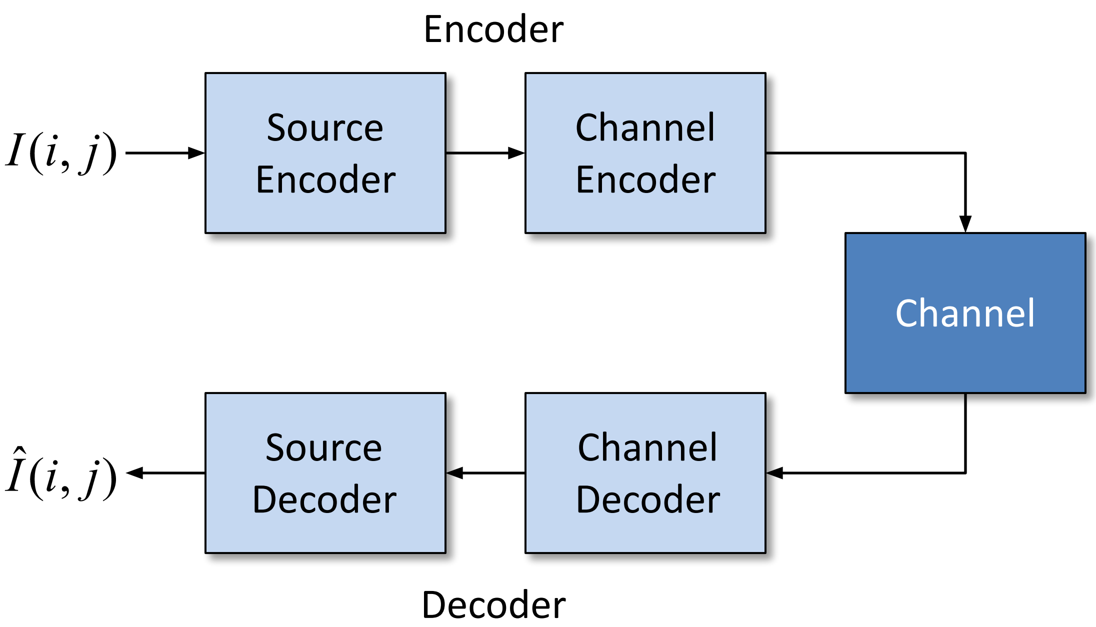
- **Objective measures**
  - Classical objective measures (SNR or THD) are inadequate, especially in perceptual coders.
  - Segmental SNR may be used for some classes of coders.

- **Subjective measures**
  - Subjective listening tests are the most reliable tools for assessing quality.
  - Standard procedures exist (e.g. ITU-T P.800/P.830, ITU-R BS.1116, ITU-R BS-1387, …)
  - Listening tests are influenced by
    - Playback level and background noise
    - Loudspeakers and headphones distortions
    - Listening room acoustics
    - Listeners' expertise

Introduction outline

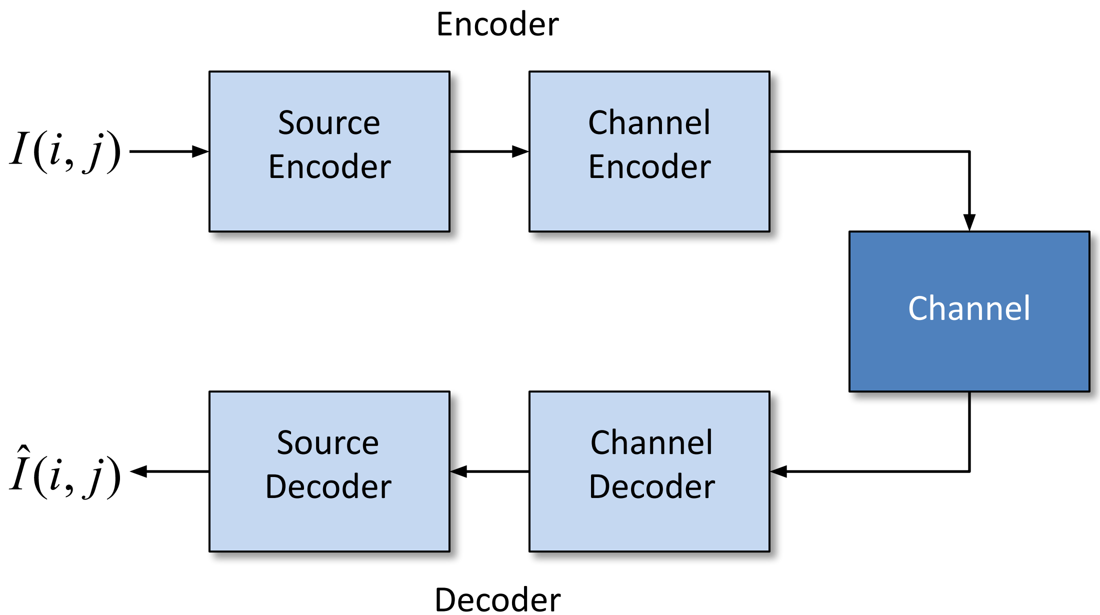
- Motivating multimedia compression
- How does compression work?
  - Redundancy and irrelevance
- Lossless and lossy compression
  - Compression efficiency
  - Quality measures
- Coding systems
  - Importance of standards

A typical coding system



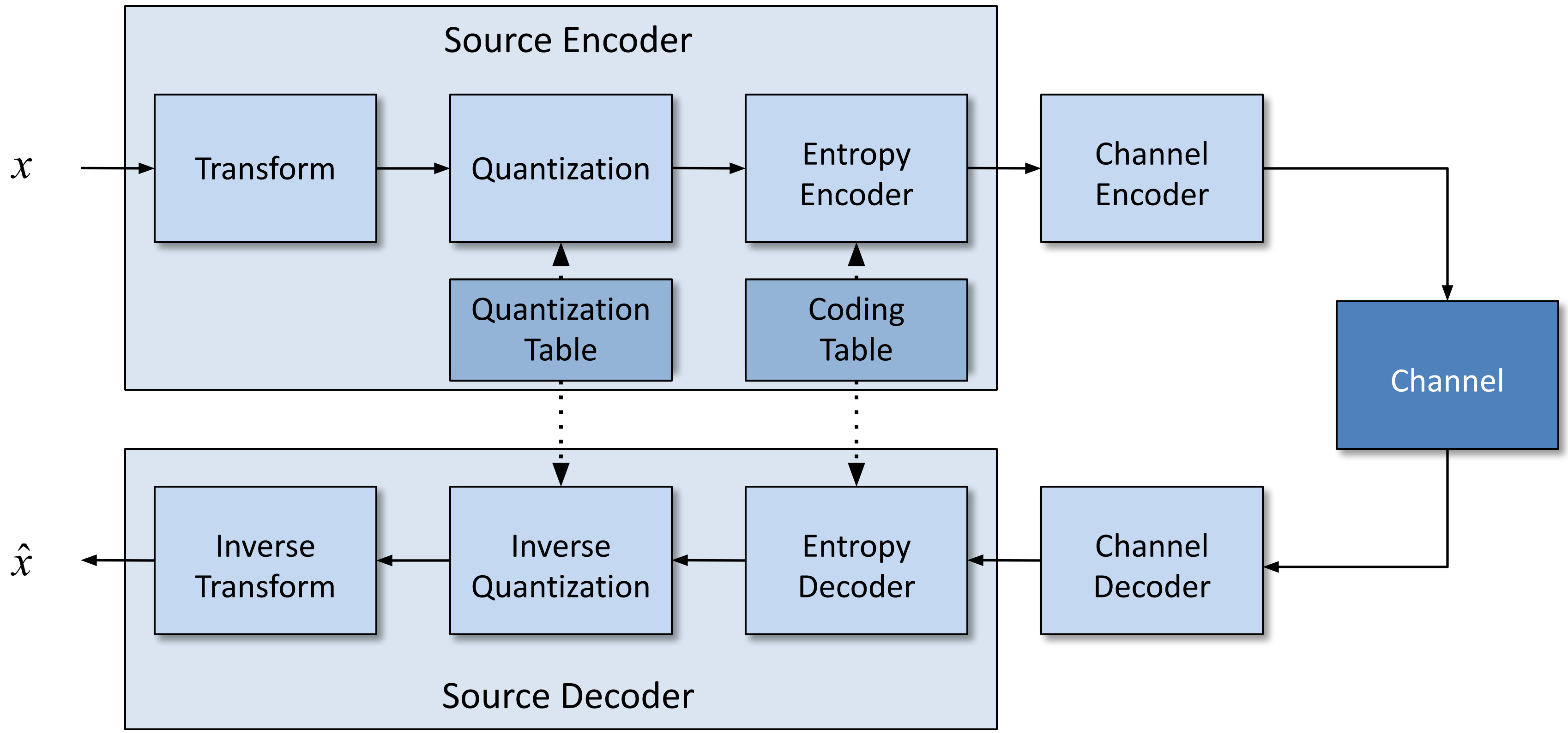
- Two different pairs of blocks: **source encoder/decoder and channel encoder/decoder**.
- The source encoder converts the source output to a **binary sequence,** and the channel encoder processes the binary sequence for transmission over the channel. The channel decoder recreates the incoming binary sequence (hopefully reliable) and the source decoder recreates the source input.

A typical coding system (II)



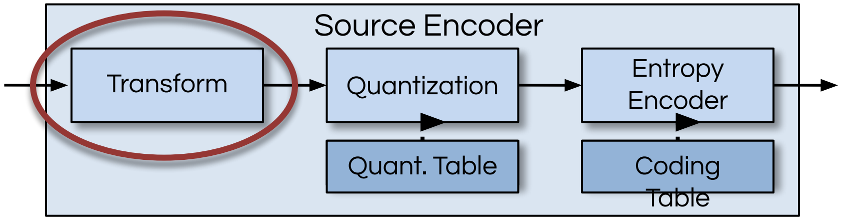
- The objective of the **source encoder** is to reduce the amount of data (average number of bits) necessary to represent the information in the signal. This data compression is mainly carried out by exploiting the redundancy in the signal information.
- The objective of the **channel encoder** is to add redundancy to the output of the source encoder to enhance the reliability of the transmission.

Typical source codec



Source encoder

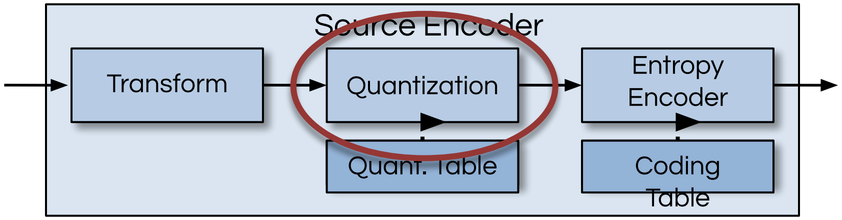
- **Transform or prediction**



  - This block **changes the description of the signal** into a less correlated set of transformed samples describing the signal; transforms the signal into a format designed to reduce its redundancy.
  - The operation is **generally reversible**. That is, the transform step does not introduce losses in the description.
  - The operation **may or may not reduce** directly the amount of data required to represent the signal.
  - Coding algorithms can be grouped into two main classes: techniques that code the original samples or use a **prediction** model (e.g. RLE) and techniques that represent the signal in a different domain and code the **transformed** coefficients (e.g. DCT)

Source encoder (II)

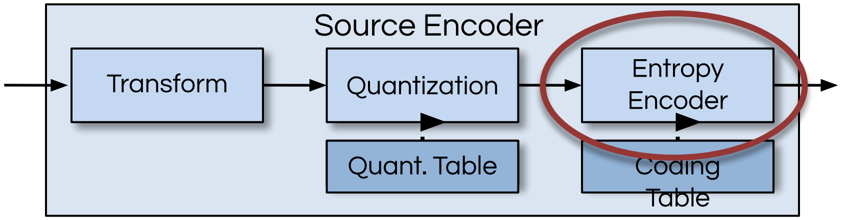
- **Quantization**



- This block **reduces the accuracy** of the transform's output in accordance with a pre-established fidelity criterion.
- The goal is **to keep irrelevant information** out of the compressed representation.
- The quantization represents a range of values of a given transformed sample by a single value in the range. The set of output values are the quantization indices. **This step is responsible for the losses in the system** and, therefore, determines the achieved compression (here, we are assuming scalar quantization; grouping a set of transformed samples into a vector and performing vector quantization is possible as well).
- This **operation is irreversible**, so it must be omitted when (lossless compression) error-free compression is desired.

Source encoder (III)

- **Entropy encoder**

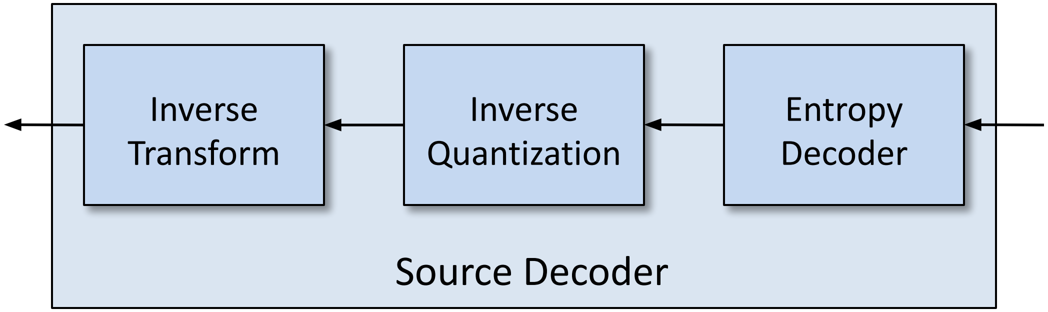


  - This block produces the **final bitstream**.
  - The entropy encoder generates a **fixed or variable length code** to represent the quantization output and maps the output in accordance with the code.
  - In many cases, a variable-length code is used: the shortest code words are assigned to the most frequently occurring quantization output values, thus minimizing coding redundancy.
  - This operation is **reversible**

- Data compression is generally performed in the quantization and entropy encoder blocks.

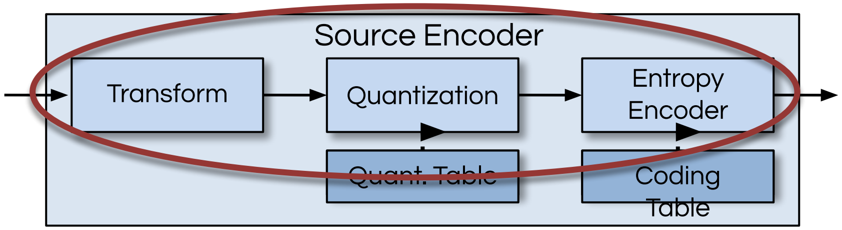
Source decoder

- **Source Decoder**
  - The decoder contains three blocks that perform, in reverse order, the inverse operations of the encoder: **entropy decoder, inverse quantization** and **inverse transform** block.



Rate/Distortion Analysis (I)

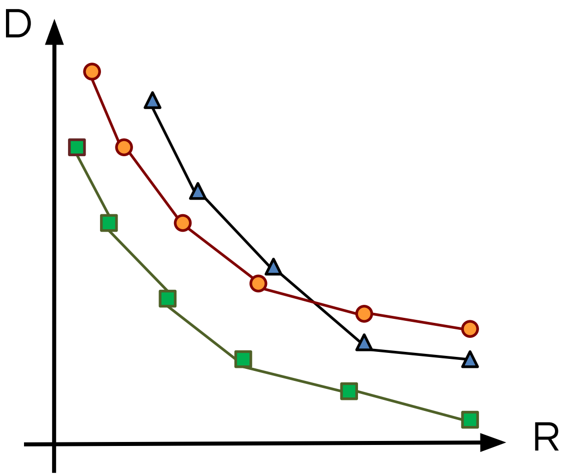
Complete system assessment



- The way to analyze the performance of a coding system is by evaluating the **quality** of the decoded signal when **varying the encoding rate**.

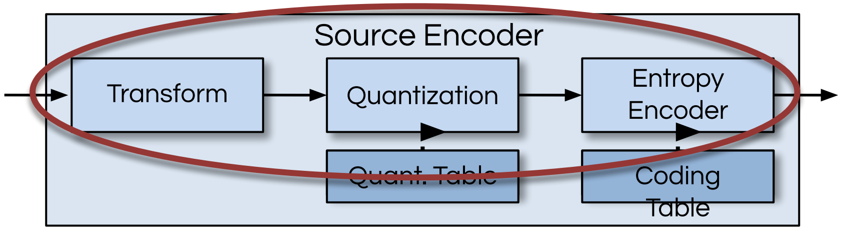
- In the case of audio signals, since good models for the HAS exist, the concept of quality refers to **perceived audio quality**.

- In the case of image and video signals, the concept of quality commonly refers to a measure of **distortion** in the decoded signal (for instance, MSE or PSNR)



Rate/Distortion Analysis (II)

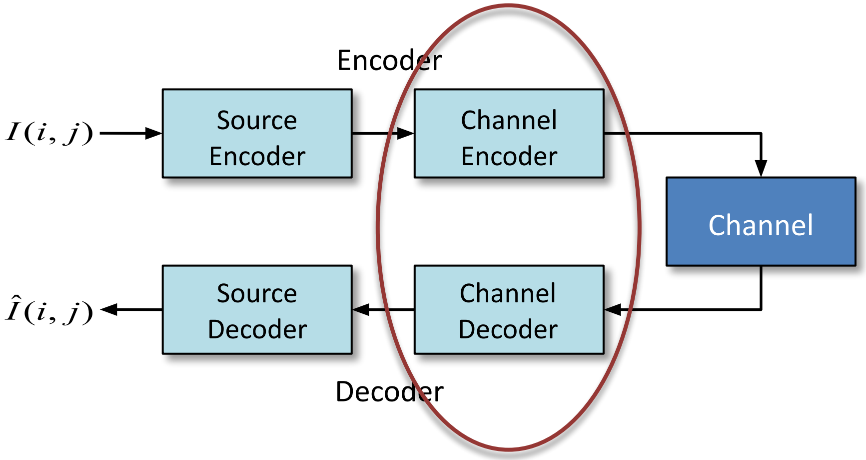
Complete system assessment



- A Rate/Distortion analysis requires the implementation of **the complete coding system**, which sometimes is not available
- The various parts of the system are commonly assessed using other (local) measures:
  - **Transform / Prediction**: Decrease in **correlation** of the output samples.
  - **Quantization**: Decrease in **entropy** of the output symbols
- Those are suboptimal measures and usually the global analysis is necessary since the different blocks can be strongly related:
  - For instance, entropy coders which are very adapted to a given source

Channel encoder and decoder

- The channel encoder and decoder play an **important role when the channel is noisy or prone to error**.
- They are design to reduce the impact of channel noise by inserting a **controlled form of redundancy** into the source encoded data.
- As the output of the source encoder contains little redundancy, it would be highly sensitive to transmission noise without the addition of this controlled redundancy.



Normalization and Coding Standards

What do standards standardize?

- In order to correctly decode the bitstream produced by the coding system, the encoder has to **share some information** with the decoder. The decoder should know
  - how the different codewords have been stored in the bitstream (**syntax definition**)
  - which actions are associated with their different decoded values (**decoder definition**).

- This shared information defines a standard
- For example, a standard may fix the specific transform to be used at a given stage of the decoding process or the range of possible values of a given quantization index.

What is left for commercial competition?

- Standards do not fix the way in which the various parameters have to be computed, opening the door to improved implementations of **standard compliant** encoders.
- A standard can introduce some additional flexibility. This is the reason for explicitly representing the Quantization and Coding Tables. Typically, these elements can be chosen at the encoder side to improve the coding performance and transmitted to the decoder in the bitstream.
- Manufacturers still have a margin for innovation and competition
  - Codec performance: parameters, coding tables, actual implementation, …
  - Added functionalities: interface, adaptation, complexity, quality, price, …

Importance of Normalization

- Normalization allows **interconnection** of equipment from different manufacturers
- Standards should arrive early in the market, to avoid *de facto* standards

Standards, file formats and containers

- Standards, Formats and Containers
  - Still Image
    - Binary
      - CCITT Group3
      - CCITT Group4
      - JBIG
      - JBIG2
      - TIFF
    - Continuous Tone
      - JPEG
      - JPEG-LS
      - JPEG-2000
      - MPEG-4
      - BMP
      - GIF
      - PDF
      - PNG
      - TIFF
      - ...
  - Video
    - DV
    - H.261
    - H.262
    - H.263
    - H.264
    - MPEG-1
    - MPEG-2
    - MPEG-4
    - MPEG-4 AVC
    - AVS
    - HDV
    - M-JPEG
    - QuickTime
    - VC1 (WMV9)
    - ...
  - Audio
    - G.711
    - G.722
    - MPEG-1
    - MPEG-2
    - MPEG-4
    - WAV
    - MP3
    - M4A
    - OGG
    - WMA
    - AAC
    - FLAC
    - ...

R. González and R. Woods, *Digital Image Processing*, Prentice-Hall, 2008.

Standards ISO/MPEG

ISO/IEC MPEG has published audio-visual standards to encourage the development of the "digital environment".

- **MPEG-1** 1992, ISO/IEC 11172

  Coding of video & associated audio for Digital Storage Media (DSM) up to 1,5 Mbit/s in error-free environments (QEF)
- **MPEG-2** 1994, ISO/IEC 13818 1996Emmy for technical excellence

  Generic video and audio coding for TV Broadcasting of Standard Definition signals (SDTV) & High Definition (HDTV)
- **MPEG-4** 1998, ISO/IEC 14496 2003JVT (ISO/ITU-T) "AVC"

  Coding of Audio-Visual Objects (AVO) oriented to manipulation (e.g. use of a "compositor" in the decoder)
- **MPEG-7** 2001, ISO/IEC 15938

  Interface for the description of multimedia content, in order to facilitate search, retrieval and access of audio-visual contents
- **MPEG-21** 2003, ISO/IEC 21000

  Structure for the development of the multimedia, management of adaptation of content to different media and devices (repurposing) & digital rights management for content creation (applications: collections, multimedia libraries...)

Course contents