Part 2: Video Coding Techniques

Vincent Roca and Christoph Neumann
{firstname.name}@inrialpes.fr

Planète project; INRIA Rhône-Alpes
MIPS'03, Napoli, November 2003

Outline

- Image formats
- Compressor architecture and its algorithms
- Video coding solutions to limit errors
- Scalability schemes
- Video standards
- (In)adequacy to network transmissions

Raw Image formats

- Every Pixel of an image is represented in
  - RGB: Red Green Blue

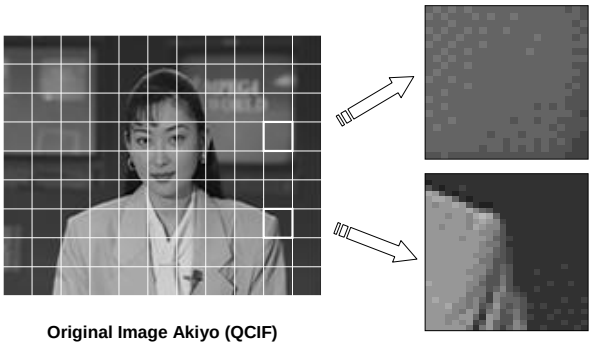
  or
  - YCbCr: Luminance Chrominance blue Chrominance red, often called YUV
- Conversion RGB to YCbCr:

$$\begin{pmatrix} Y \\ Cb \\ Cr \end{pmatrix} = \begin{pmatrix} 0.587 & 0.114 & 0.299 \\ -0.331 & 0.500 & 0.169 \\ -0.419 & -0.081 & 0.500 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

- 4:2:2
  - CIF: resolution 352x288
    - **Horizontal: 352 Y and 176 Cb and 176 Cr**
    - **Vertical: 288 Y and 144 Cb and 144 Cr**
  - QCIF: resolution 176x144
- 4:2:0: half the number of pixels for Cr and Cb

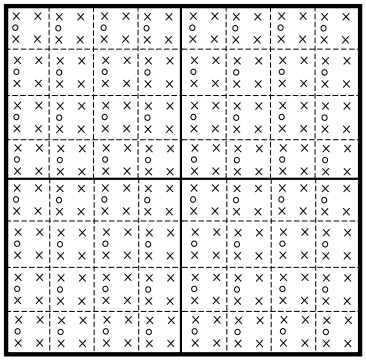
Macroblocks

- **Frames are divided into 16X16 pixel areas (4 8x8 blocks).**



Original Image Akiyo (QCIF)

Macroblocks, 4:2:0 format



- **Each 16x16 pixel area is a macroblock which contains data:**
  - 4 8x8 blocks for luminance Y
  - 1 8x8 block for chrominance Cr
  - 1 8x8 block for chrominance Cb

X: luminance
O: chrominance

Outline

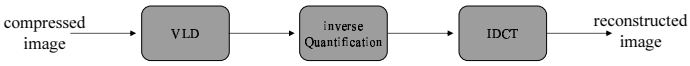
- Image formats
- Compressor architecture and its algorithms
- Video coding solutions to limit errors
- Scalability schemes
- Video standards
- (In)adequacy to network transmissions

Intra coded frame

- DCT transformation of each Macroblock
- Quantification
- VLC / entropy coding with « zigzag » scan

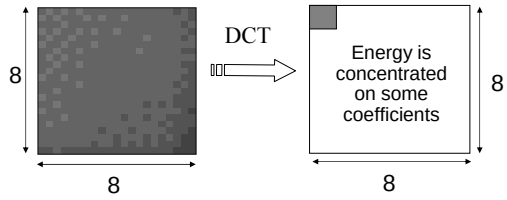
- Compression

- Decompression



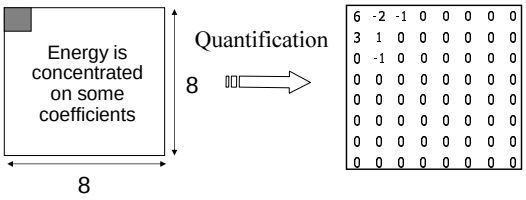
DCT

- Transforms image representation in frequencies and amplitudes
- DCT does not reduce the amount of data, but prepares it for following processes
- Low frequencies situated on left upper corner, high frequencies on right bottom corner



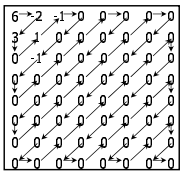
Quantification

- Filter on frequencies (adapt frequencies to human eye)
- Ex.: High frequencies are less significant for human eye
- Leads to more 0 in representation
- Reduces amount of data



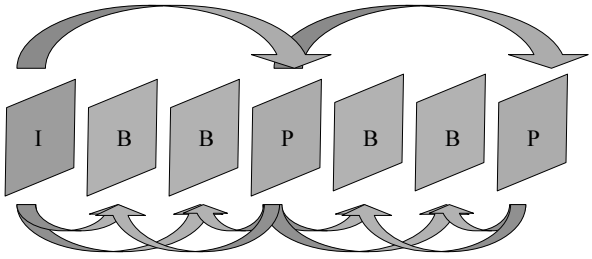
Variable length coding (VLC)/ Entropy Coding

- Many codecs use Huffman coding
- « zigzag » scan (low frequencies situated on left upper corner, high frequencies on right bottom corner)



Frame types

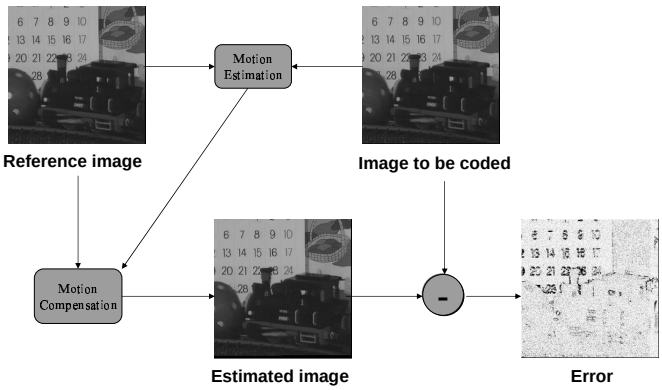
- Three frame types: I, P, B
  - I: Intracoded frames
  - P: Predicted frames
  - B: Bidirectional predicted frames



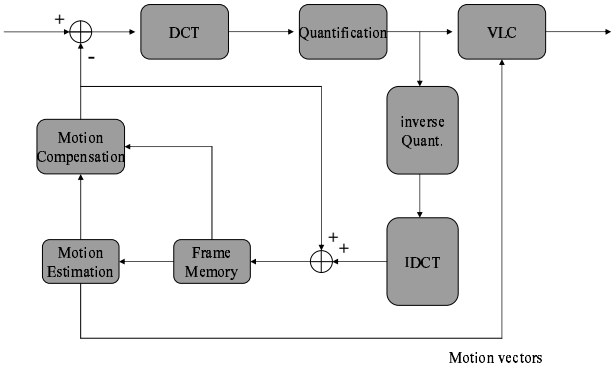
- Display order: I B B P B B P B B I B B P …
- Bit-stream order: I P B B P B B I B B P B B …

Motion compensation

- For predicted frames (P- or B-frames)



Compressor architecture

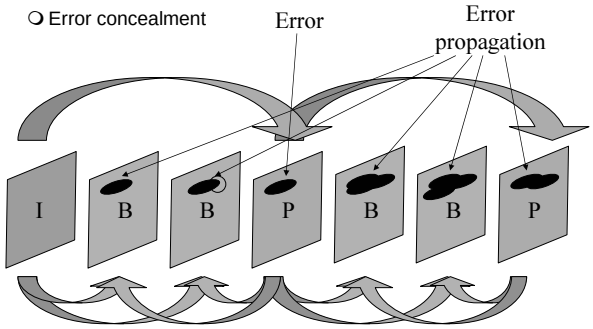


Outline

- Image formats
- Compressor architecture and its algorithms
- Video coding solutions to limit errors
- Scalability schemes
- Video standards
- (In)adequacy to network transmissions

Error propagation

- Video coding solutions to limit errors
  - Error-resilient video coding
  - Error concealment



Video coding solutions to limit errors

- Error-resilient video coding
  - Resync markers within video bitstream
  - RVLCs: reversible variable length codes. No need to jump to next resync marker if error occur. Decoder is able to continue to decode the data. But is less efficient than VLC
  - Data partitioning: data close to a resync marker are more likely to be accurately decoded than those further away
- Error concealment
  - Action only taken by the decoder
  - Estimate missing region
    - **temporal interpolation**
    - **spatial interpolation**
    - **motion-compensated temporal interpolation**

Outline

- Image formats
- Compressor architecture and its algorithms
- Video coding solutions to limit errors
- Scalability schemes
- Video standards
- (In)adequacy to network transmissions

Scalability schemes

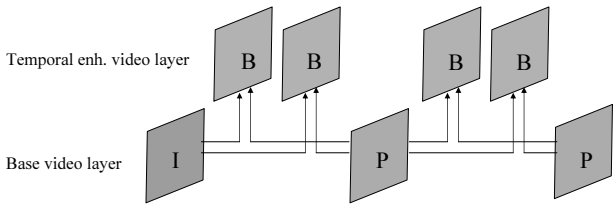
- AKA hierarchical encoding
- One base video layer + one or more cumulative enhancement layers
- Quality of video increases with number of decoded video layers
- Higher layers need lower layers in order to be decoded
- Often proposed as a solution for rate control

Scalabity schemes

- Types of scalability:
  - Temporal
    - **Frame rate**
  - Spatial
    - **Resolution**
  - SNR
    - **SNR**
- Special case FGS:
  - Partial reception of enhancement layer possible
  - Quality of video increases proportional to the number of received bits in the enhancement layer
  - FGS
    - **SNR**
    - **Temporal (FGST)**

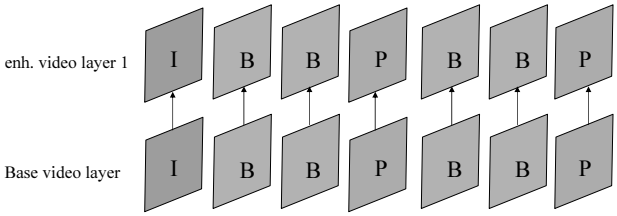
Temporal scalability

- Enhancement layer(s) increases decoded video frame rate (frames per second)
- Example (other scheduling of frames are possible):



Spatial scalability & SNR scalability

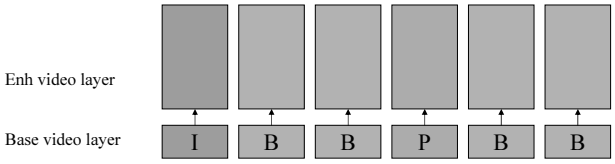
- Enhancement layer(s) increases decoded video resolution
- Example:
  - Base layer only: QCIF
  - Base layer + one enh. layer: CIF



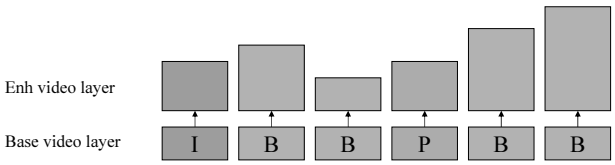
- Same dependencies for SNR scalability
  - Enhancement layer(s) increases decoded video SNR

FGS

- At the sender

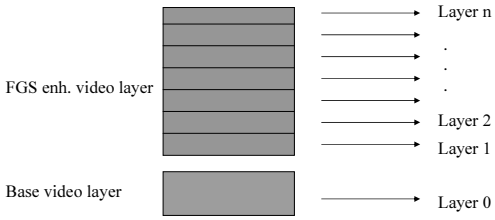


- At the receiver



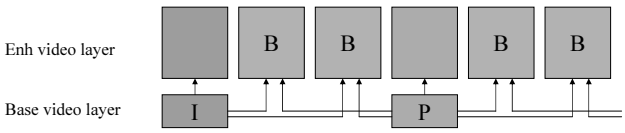
FGS

- FGS as layered video

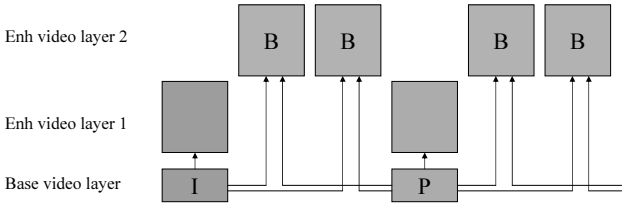


FGST

- Add temporal refinement
- Variant , single enh. layer:



- Variant 2, FGST separate layer:

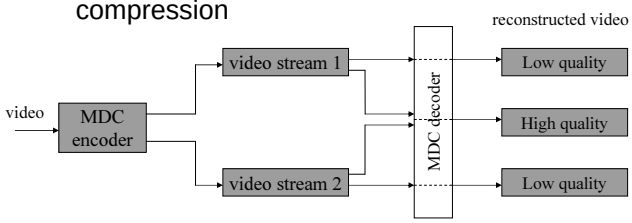


Issues of layered scalability

- Availability: Most video codecs only support one (!!!) enhancement layer
  - e.g. ISO MPEG-4 reference codec
- Compression: More overhead than single layered stream (less compression)

MDC

- Several independent video layers
- Each layer can be watched independently from others
- Increased quality if several layers available
- More overhead than cumulative layered compression



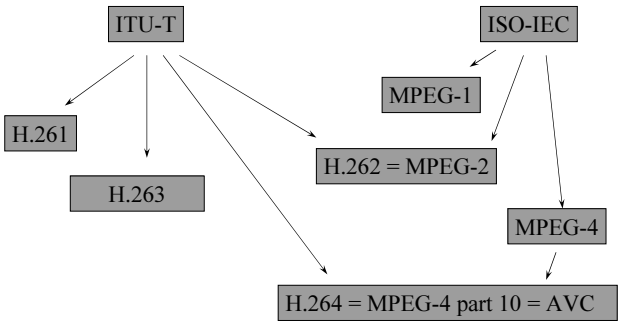
Outline

- Image formats
- Compressor architecture and its algorithms
- Video coding solutions to limit errors
- Scalability schemes
- Video standards
- (In)adequacy to network transmissions

Video Standards

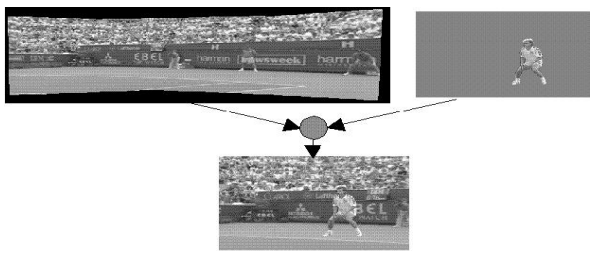
- H.261
  - Video telephony and teleconferencing over ISDN
  - p*64kb/s
- MPEG-1
  - Video CD
  - 1.5 Mb/s
- MPEG-2, H.262
  - Digital Television, DVD
  - 2-20 Mb/s
- H.263
  - Video telephony over PSTN
  - 33.6 kb/s and up
- MPEG-4
  - Interactive multimedia applications: Visiophonie, Internet Streaming, TV Broadcast, ...
  - Object-based coding, synthetic content, interactivity …
  - Bitrate variable
- MPEG-4 part 10 (AVC), H.264
  - Like mpeg-4 with improved compression
  - 10's to 100's of kb/s better
- MPEG-7, MPEG-21
  - Like mpeg-4 with content description/indexation

Standards Overview

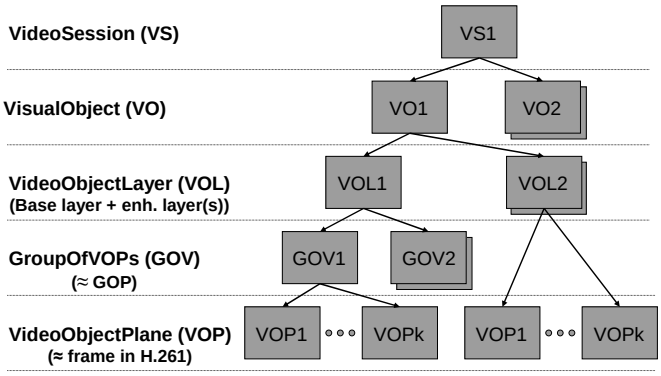


MPEG-4

- Scene decomposition
  - Decomposition in Visual Objects (VO), that are coded separately



MPEG-4



Outline

- Image formats
- Compressor architecture and its algorithms
- Video coding solutions to limit errors
- Scalability schemes
- Video standards
- (In)adequacy to network transmissions

(In)adequacy to network transmissions

- Problems due to network transmissions:
  - Bandwidth
  - Losses
  - Delay Jitter
- Video coding only solutions:
  - VBR, CBR
  - Error-resilient video coding
  - Error concealment
- Not sufficient!!!
- Solutions based on transmission mechanism needed