ELEC S347F Multimedia Technologies

Advanced Video Compression

MPEG Video Standards

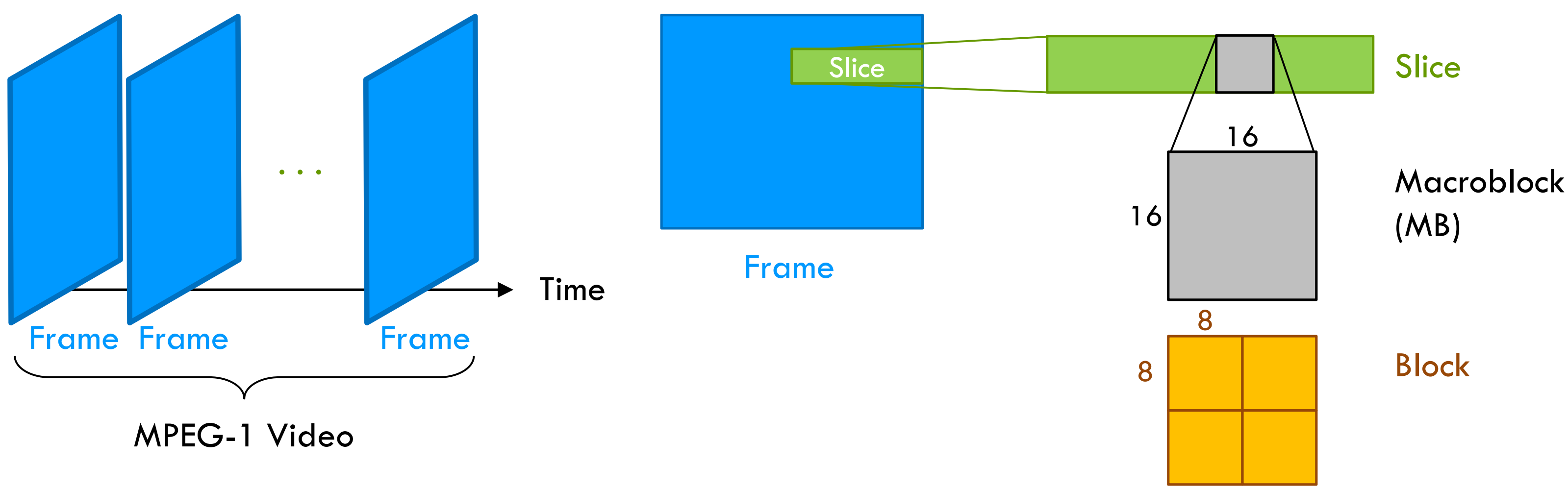
- Motion Picture Expert Group (MPEG) established around 1990 to create standard for delivery of audio and video
  - Different MPEG standards have been developed for various target domains that need different compression approaches
- MPEG-1 (1991): Target at VHS quality on a CD-ROM
  - 320x240 resolution + CD audio (1.5 Mbps)
- MPEG-2 (1994): Target at DVD video and digital TV broadcasting
  - MPEG-3: Target at HDTV (merged with MPEG-2)
- MPEG-4 (1998): Originally targeted at very low bit-rate communication (4.8 to 64 kbps), but later for wide range of bitrates
  - MVC (2008): Target at 3D Blu Ray Disc

MPEG-1 and H.261: Similarities

- MPEG-1 video is similar to the H.261 standard
  - Both use YCbCr color model and 4:2:0 Chroma subsampling
  - Macroblocks(MB)-based motion estimation and compensation
  - Intra mode and Inter mode MB compression
  - Subdivide a frame in coding to improve the robustness of a frame and allow resynchronization
    - GOB (for H.261) and Slice (for MPEG-1)

MPEG-1 and H.261: Differences

- Robustness and Resynchronization
  - H.261: Group of Blocks (GOB)
    - A GOB contains 33 MBs
  - MPEG-1: Slice
    - Slices within a frame may vary in size

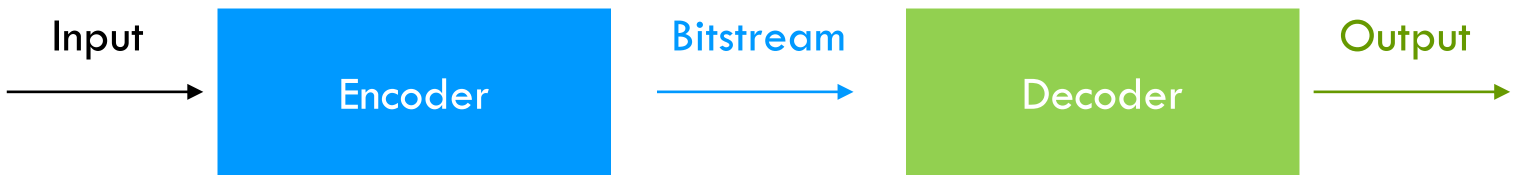


MPEG-1 and H.261: Differences

- Target applications
  - H.261: target at video telephony and conferencing on ISDN
    - Real-time two-way applications
      - Require low encoding and decoding complexity
      - Low tolerance to delay
      - Supporting to random access, video thumbnails are not required
    - Low transmission rate
      - Typical data rate of ISDN: 64 kbps (with audio)
  - MPEG-1: target at high quality video storing on CD-ROMs
    - Stored one-way video
      - Allow higher encoding complexity
      - Allow longer playback delay
      - Should support random access, fast forward/backward, video thumbnails
    - Higher data rate
      - Typical reading speed of a 1X CD-ROM: 1.2 Mbps
      - MPEG-1: 1.2 Mbps for video, 0.3 Mbps for audio

MPEG-1 and H.261: Differences

- Asymmetric Complexity
  - MPEG encoders are generally more complex than decoders



Encoder is algorithmic
i.e. it does different things according to nature of input

Complex to make
Expensive in terms of complexity

Decoder is deterministic
i.e. it always does what the bitstream tells it to do

Simple to make
Inexpensive in terms of complexity

MPEG-1 and H.261: Differences

- Intra Mode and Inter Mode Encoding
  - H.261: applied in MB level
    - i.e. some MBs are intra mode coded, some are inter mode coded
  - MPEG-1: applied in MB level and frame level
    - Intraframe-encoded Frame (I-Frame): all MBs in the frame are encoded in intra mode
      - I-Frames allow random access, fast forward/backward, video thumbnails for MPEG-1 video
    - Predicted Frame: each MB can be coded in intra mode or inter mode
      - Forward Predicted Frame (P-Frame): similar to H.261's
      - Bidirectional Predicted Frame (B-Frame): introduced in MPEG-1

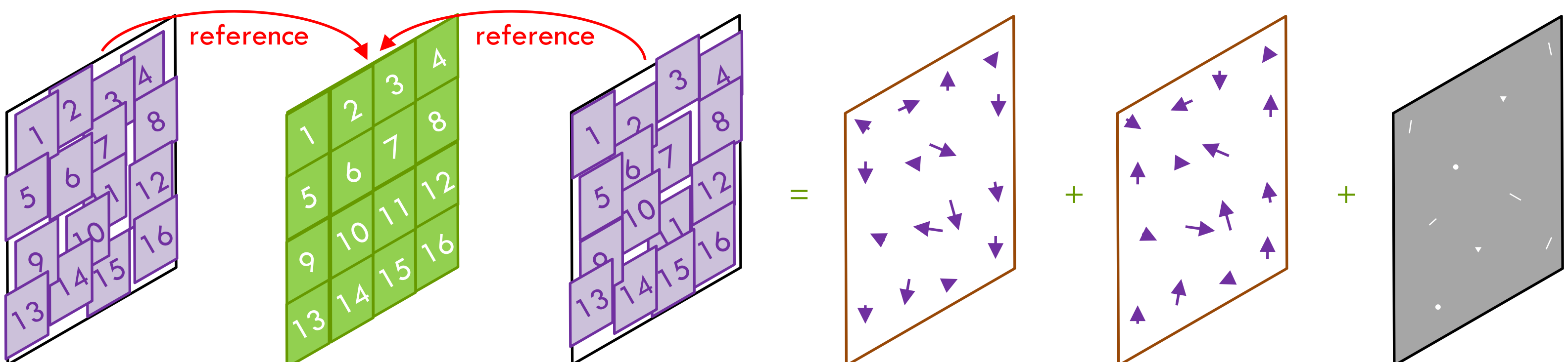
MPEG-1 and H.261: Differences

- Reference Frame
  - H.261: always reference the previous frame
    - Motion vectors are computed from the previous frame
  - MPEG-1: can reference previous or future frame or both
    - Forward Prediction (P-Frame)
      - As before in H.261, compute the motion vector for each MB from the nearest preceding I/P-frame
    - Backward Prediction (B-Frame):
      - Compute the motion vector for each MB from the nearest succeeding I/P-Frame
    - Bidirectional Prediction (B-Frame):
      - Compute two motion vectors for each MB from the nearest preceding I/P-Frame and the nearest succeeding I/P-Frame

B-Frame

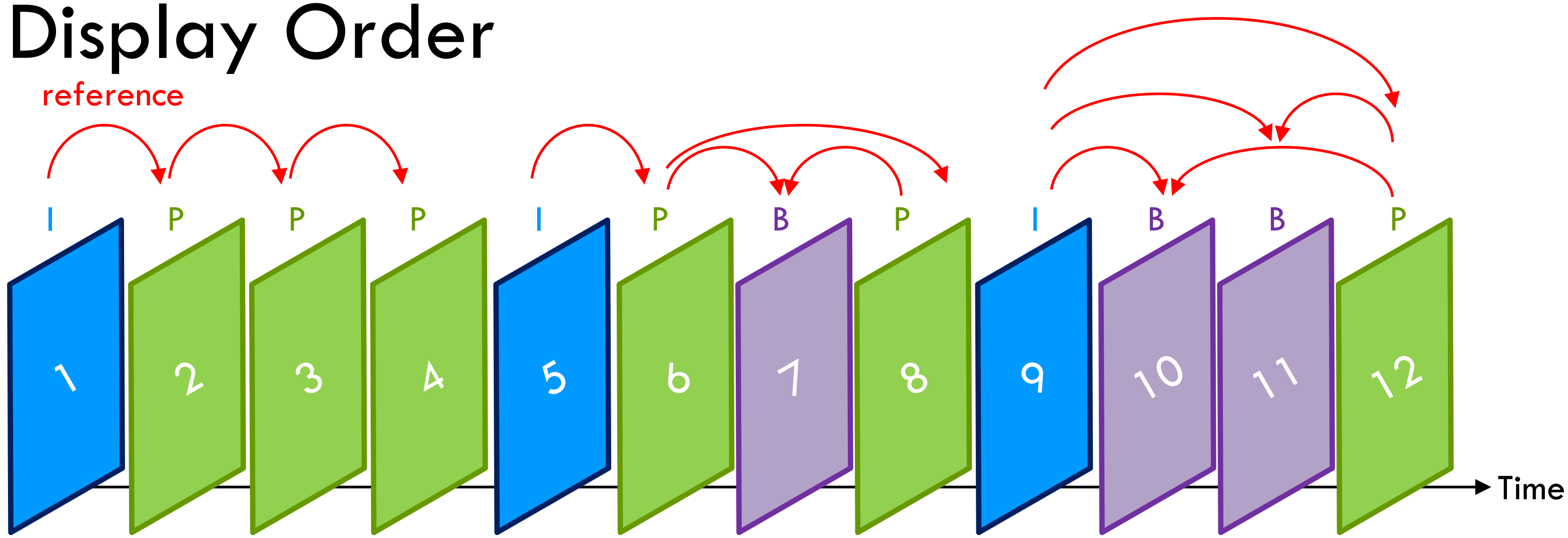
- Backward Prediction

- Bidirectional Prediction

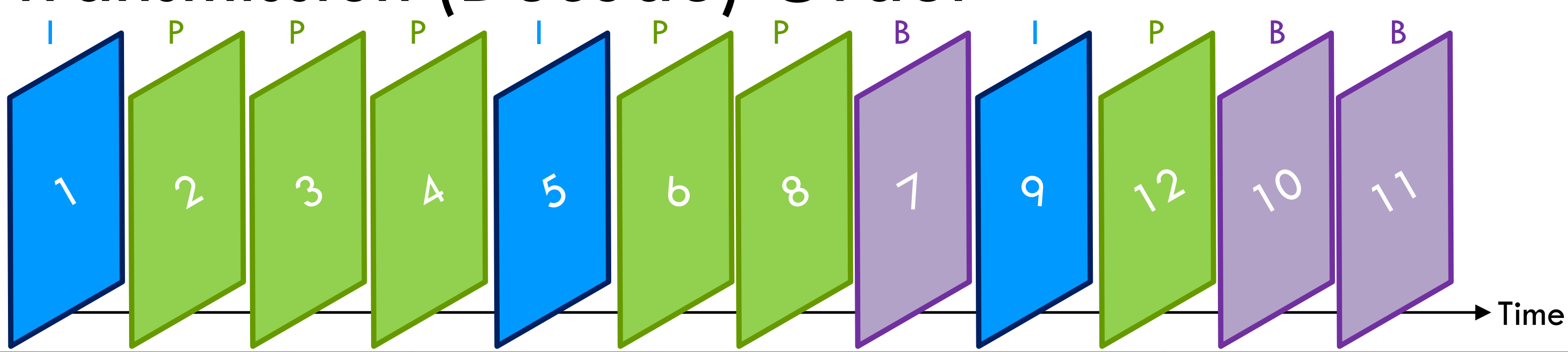


B-Frame

- Display Order



- Transmission (Decode) Order



B-Frame

- Advantages
  - Improve accuracy for motion estimation
    - More choices of MVs for each MB
    - Especially good for MBs which contain moving objects that are hided by other objects in some frames
  - Better coding efficiency
    - Typical compression ratio:
    - I-Frames: 7:1, P-Frames: 20:1, B-Frames: 50:1
  - Improve robustness
    - Errors in a B-Frame will not be propagated to future frames

B-Frame

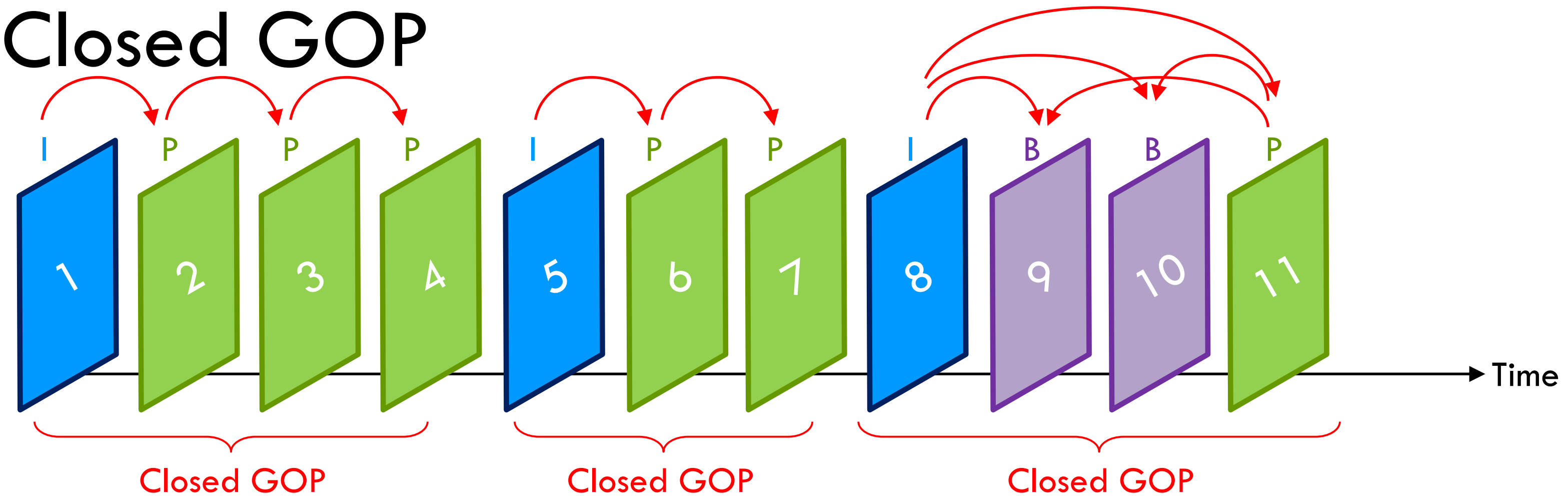
- Disadvantages
  - Introduce extra delay in playback
    - B-Frames reference future frames
    - So, require to encode and transmit reference frames first
    - Decoders will receive the frames out of order
    - E.g. consider the frame sequence in previous example
  - Higher memory requirement
    - Require more memory in encoder and decoder to accommodate the extra reference frame and frames pending for encoding
    - E.g. in previous example,
      - Decoder: requires remembering frame 9 and 12 to decode frame 10, 11
      - Encoder: require remembering frame 9, 10, 11 before encoding frame 12
  - Higher complexity
    - More MV searches required than P-Frames

Group of Pictures (GOP)

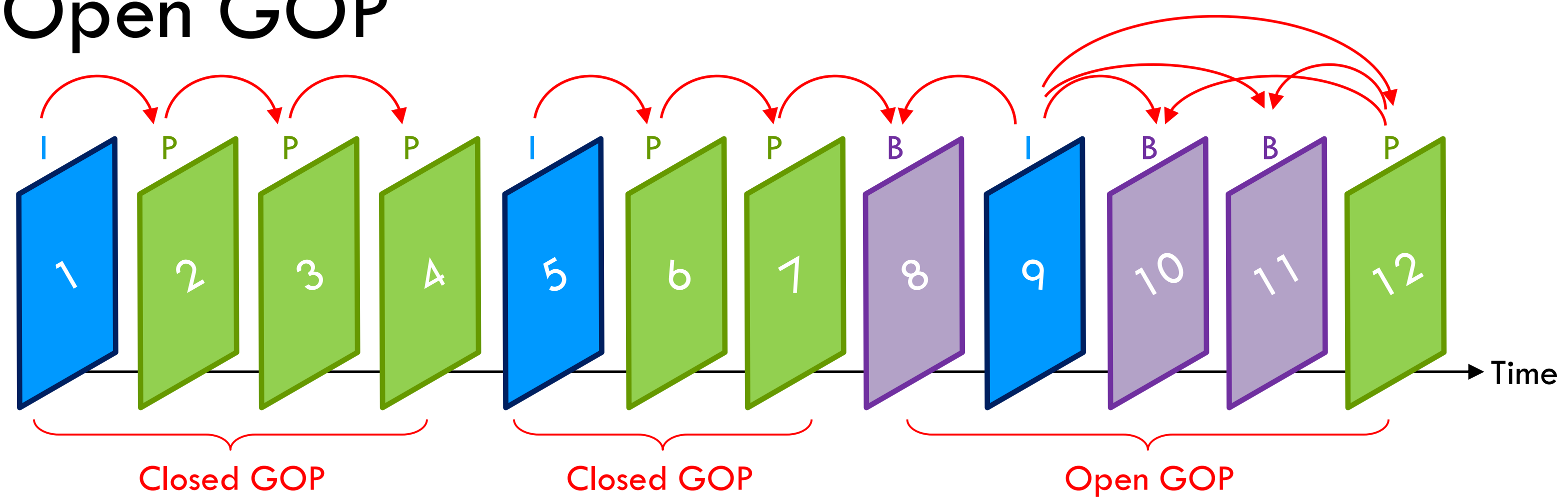
- A group of pictures (GOP) is a sequence of frames that can be removed without destroying the integrity of the video
  - A GOP begins with an I-Frame or one or more B-Frames followed by an I-Frame
  - And ends with an I-Frame or P-frame
- A closed GOP
  - Begins with an I-Frame
  - Can be decoded independently out of the video
- An open GOP
  - Begins with one or more bidirectional predicted B-Frames followed by an I-Frame
  - Can be decoded only if the preceding GOP is available

Group of Pictures (GOP)

- Closed GOP



- Open GOP



Group of Pictures (GOP)

- The frequency of I-Frame can be fixed or dynamic
  - Fixed: start a new I-Frame regularly
  - Dynamic: start a new I-Frame based on the residual between prediction and actual frame data
    - When there is scene change, the residual is getting large
    - Should start a new I-Frame
- The higher the frequency of I-Frame, the higher the random access capability, and the higher the resistance to error propagation
  - However, I-Frames demand the highest bandwidth (B-Frames the least)
  - This causes large delay variation between frames and not good for network streaming

MPEG-2 Video/H.262

MPEG-2 Video

- ISO/IEC 13818-2
  - MPEG-2 Part 2
  - Jointly developed by ITU-T and MPEG
  - Commonly known as MPEG-2 Video and H.262
  - Target at storing high quality video on DVDs
  - Later expanded to SD and HD TV broadcasting over the air or cable
  - Backward compatible to MPEG-1 Video
  - But support higher bit rate, higher resolutions and interlaced video

MPEG-2 Video vs. MPEG-1 Video

- MPEG-2 Video originally target at storing high quality video on DVDs
  - As opposed to MPEG-1's on VCDs
  - Allow higher data rate and larger frame resolution
- More options for Chroma subsampling
  - In addition to 4:2:0, support 4:4:4 and 4:2:2
  - Higher video quality but higher data rate
- MPEG-2 Video supports interlaced video
  - Search motion vectors on fields, not just frames

MPEG-2 Video vs. MPEG-1 Video

- More scalable than MPEG-1
  - Support playing the same video in different resolutions and frame rates
  - Allows a video to be coded into two layers
  - The base layer provides a coarse but low bandwidth version of the video
  - The enhancement layer when added to the base layer provides a detailed and more accurate representation
  - Especially useful for broadcasting to enhance robustness and support two versions of broadcasting using one signal
- Extended to support HDTV
  - MPEG-3 originally for HDTV, but later merged into MPEG-2

MPEG-2 Video: Layered Video

- MPEG-2 has 4 ways to generate layered video
  - 1) Data Partition
    - A video is first coded using conventional method
    - The base layer contains the motion vectors and the first few low-frequency DCT coefficients of the residual
    - The enhancement layer includes the remaining DCT coeff.
  - 2) SNR Scalability
    - A video is coded using conventional method but with a large quantization step size and form the base layer
    - Then the quantization error for the DCT coeff. are quantized again using a smaller step size and form the enhancement layer

MPEG-2 Video: Layered Video

- 3) Spatial Scalability
  - The base layer codes a subsampled version of the original video
  - The enhancement layer codes the original size
    - Using either the past coded frame in the original size, interpolated version of the current frame produced by the base layer, or a weighted sum of both for frame prediction
- 4) Temporal Scalability
  - The base layer codes the original video at a lower frame rate
  - The enhancement layer codes the skipped frame
    - Using either the coded frames in the base layer, or past coded frames in the enhancement layer for frame prediction

MPEG-4 AVC/H.264

MPEG-4 Part 2 and Part 10

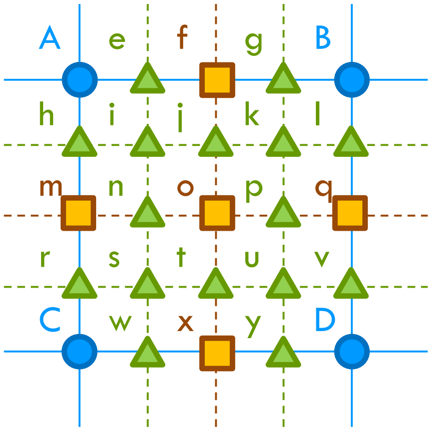
- ISO/IEC 14496-2
  - MPEG-4 Part 2
  - Commonly known as MPEG-4 Visual
  - Compatible to H.263
- ISO/IEC 14496-10
  - MPEG-4 Part 10
  - Jointly developed by ITU-T and MPEG
  - Commonly known as MPEG-4 Advanced Video Coding (AVC) and H.264
  - Target for storage of HD video, interactive multimedia, games
  - Achieve up to 30%-50% better compression than MPEG-2 Video and up to 30% over H.263+
    - Not only aim to improve compression, but also functionality and interactivity

Improvement of AVC/H.264

- Improvement on prediction accuracy
  - Multiple Forward/Backward Reference
    - AVC/H.264 allows finding the best reference in 2 possible directions (forward and backward) each contains up to 16 frames
    - As opposed to MPEG-1/2 which can reference one frame in each direction for prediction
  - Flexible Weighting of Reference Frames
    - AVC/H.264 allows prediction from reference frames with any weighting
    - While MPEG-1/2's B-Frames use averaged weighting for bidirectional prediction

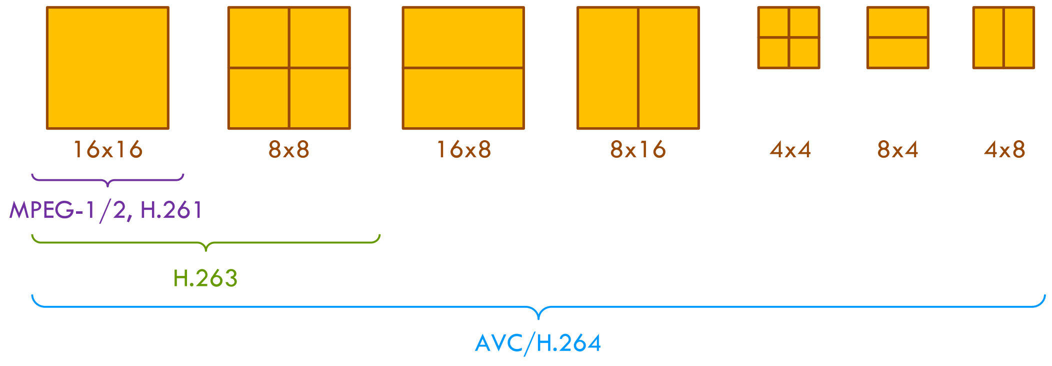
Improvement of AVC/H.264

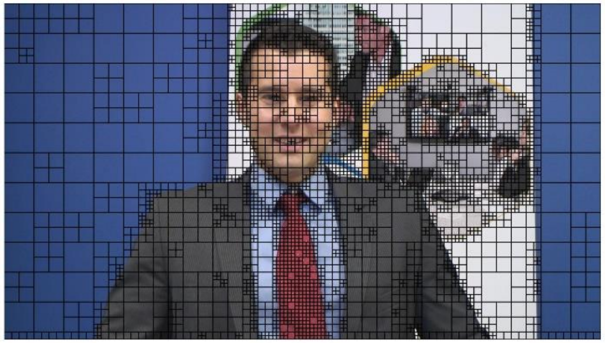
- Improvement on prediction accuracy
  - Finer Search Step Size of Motion Estimation
    - AVC/H.264 allows quarter-pixel (1/4 pixel) step size
      - 16X choices
    - As opposed to H.263's half-pixel (1/2 pixel) step size,
      - 4X choices
    - MPEG-1/2's one-pixel step size
    - Original pixel:
      - A, B, C, D
    - Interpolated half-pixel and quarter-pixel:
      - f, m, o, q, x, e, g, h, i, j, k, l, n, p, r, s, t, u, v, w, y



Improvement of AVC/H.264

- Improvement on prediction accuracy
  - More Options for Block Partition
    - AVC/H.264 supports block size of 16x16, 8x8, 16x8, 8x16, 4x4, 8x4, 4x8
    - As opposed to MPEG-1/2 and H.261's 16x16, H.263's 16x16 and 8x8





Improvement of AVC/H.264

- Intra-prediction
  - Previous: MBs within a frame are coded independently
  - AVC/H.264: exploit spatial correlation between pixels in current block and adjacent blocks of a frame
- Integer Transformation
  - AVC/H.264 uses an integer version of the DCT
  - Eliminate any numerical errors between the forward transform at the encoder and the inverse transform at the decoder
  - Also, the transform block size can be varied from block to block depending on which one gives the best representation

AVC/H.264

- MPEG-1/2 Video use frame-based/block-based coding
- MPEG-4 AVC uses object-based coding
  - A video sequence is decomposed into multiple objects and each object is coded separately
  - The information transmitted for each object includes its shape, motion, and texture
- Advantages of object-based coding
  - Higher compression ratio
    - For block-based coding, it is possible that multiple potential matches yield small prediction errors since objects usually collide with each other
    - For object-based coding, each object is of arbitrary shape (not necessary be a rectangle) and ideally will obtain a unique motion vector consistent with the actual object motion

AVC/H.264

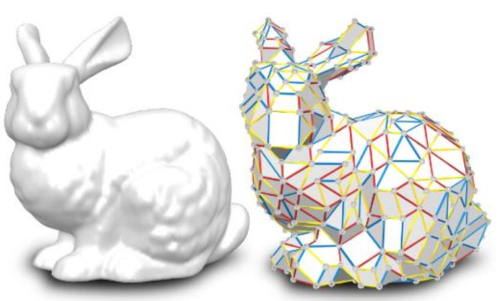
- Advantages of object-based coding
  - Allow encodes to code different objects with different accuracy
    - E.g. the foreground moving object can be coded more accurately than background
  - Easy manipulation of audiovisual scenes
    - Digital effects (e.g. fade-ins)
    - Animation and synthetic sound can be composed with natural audio and video
    - Insertions of subtitles
  - Content-based interactivity
    - A viewer can translate or remove a graphic overlay to view the video beneath it
    - Decoders can choose to compose the objects as desired (that means it can choose not to decode certain objects, or replace an object with some other pre-stored objects)

AVC/H.264

- Shape Coding
  - Bitmap:
    - A matrix in which a '1' means opaque and a '0' represents transparent of the corresponding pixel
  - Alpha map:
    - A matrix whose element values refer to the transparency of the shape ranging from 255 (opaque) and 0 (completely transparent)
- The SAD computation is modified as
  - $\mathrm{SAD}(i,j) = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+k+$

AVC/H.264

- Synthetic Object Coding
  - 2D Mesh
    - Useful for object animation
  - 3D Mesh
    - Useful for face objects and body objects which the face or body can be shaded or texture-mapped



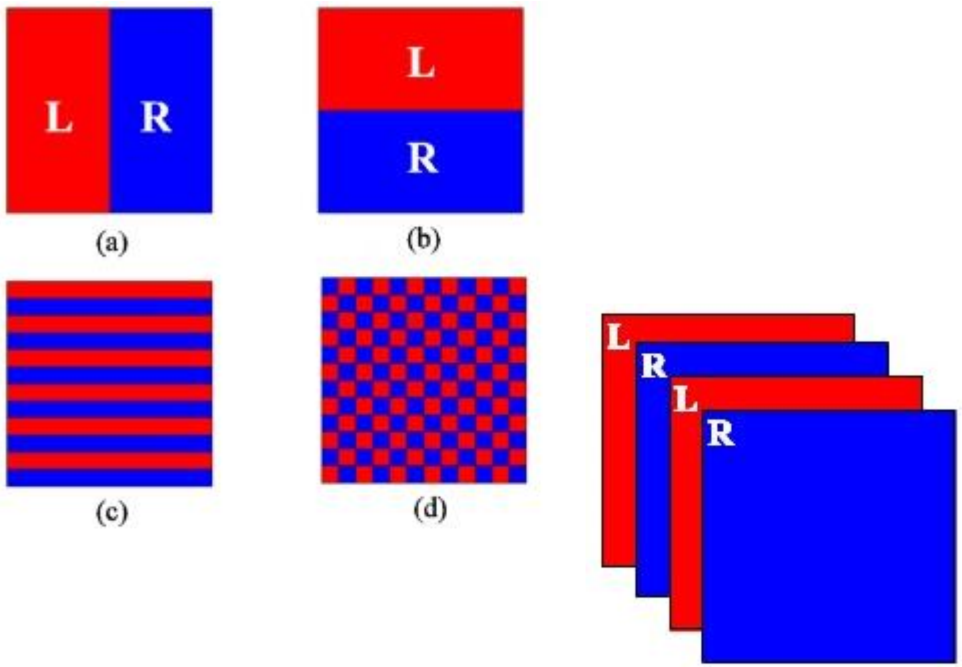
Multi View Coding (MVC)

3D Video

- 3D content is made by showing a separate image (video) to each eye
- It can be broadly classified into glasses-based and glass-free technologies
- Each of the two captured views is presented to one of the eyes
- Can be multiplexed either spatially or temporally

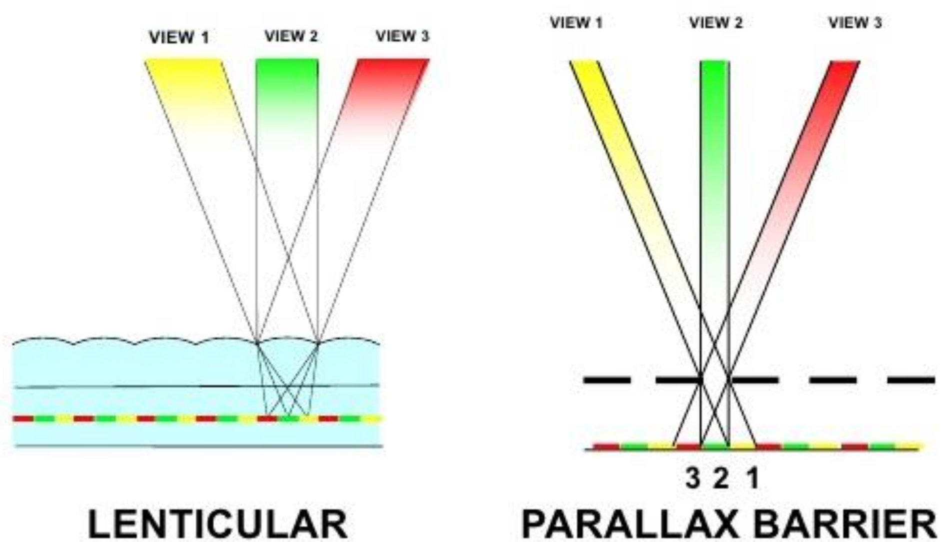
Multiplexing Stereo Video

- Spatial Multiplexing
  - Half the resolution
  - (a) Side-by-Side layout
  - (b) Top-Bottom layout
  - (c) Horizontal layout
  - (d) Checkers layout
- Temporal Multiplexing
  - Double the frame rate
  - Advantage: maintaining the full resolution of each view
  - Disadvantage: it is hardware representation dependent (acquisition process is tailored to a specific type of displays)



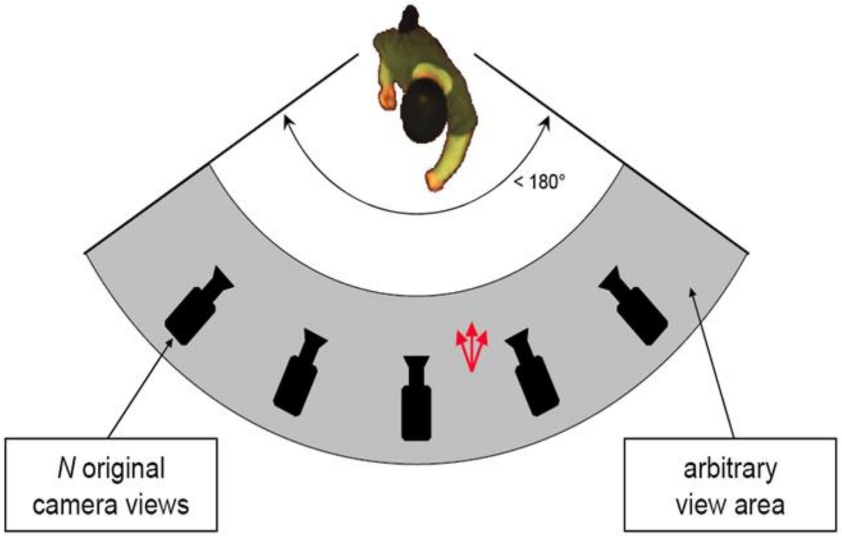
Multi-View Displays

- The most common form of multi-view displays utilize either lenticular screens or parallax barriers
  - Lenticular screens consist of a series of vertically aligned or slanted cylindrical lenses and parallax barriers vertically aligned apertures
  - Light is guided in the appropriate directions by either focusing it or by blocking unwanted rays



Multi-View Coding (MVC)

- MVC is a standard for compression of multiple video streams into one encoded streams
  - All streams are encoded as differences from the first stream
  - MVC with 2 video streams is 3D stereo



MVC Coding

- Simulcast Encoding
  - Encode each view independently
  - Advantages: enable streaming each view over separate channels, clients can request as many views as their 3D displays require
- Dependent Encoding
  - Encode views by exploiting the inter-view redundancies
  - Advantages: decrease the overall bit rate
  - Disadvantages: require to employ a special inter-view prediction structure in order to enable view-scalable and view-selective adaptive streaming

H.264/AVC 3D Encoding

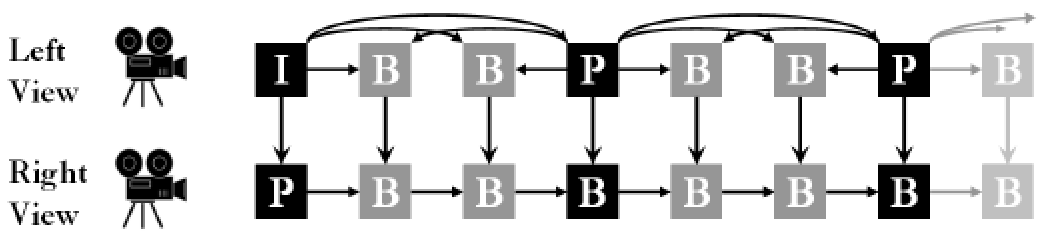
- There are two options for 3D H.264 AVC encoding
  - Regular encoding of a single video stream with two L/R frames packed into one frame
    - No signaling information is required
  - Informing the encoder the video is 3D
    - Encoder will insert Frame-Packing information by Frame Packing Arrangement (FPA) structure (e.g. side-by-side, top-bottom, interlaced, etc)
    - Decoder will switch to correct 3D mode based on its signaled frame packing information

H.264/AVC Multi-View Extension

- Multi-view Video Coding was standardized in 2008 and adopted for 3D Blu Ray in 2009
- Exploit temporal and inter-view dependencies using MPEG4-AVC/H.264
- Usage of hierarchical B pictures in temporal direction
- Usage of P or hierarchical B pictures in inter-view direction
- Frame reordering to optimize memory usage

H.264/AVC Multi-View Extension

- Coding structure allows AVC coder to select the best inter-view-temporal neighbors within the MVC sequence
  - Coding gain obtained by usage of hierarchical B-frames and by exploitation of inter-view dependencies
- MVC with only two cameras



H.264/AVC Multi-View Extension

- Enable inter-view prediction
  - Prediction structure is simplified by restricting inter-view prediction to anchor pictures only
  - Large disparity or different camera calibration affects coding efficiency