Topics Covered Before First Exam

- Raster video representation
  - Interlaced vs. progressive scan
  - Spectral characteristics
- Analog TV transmission
  - Color coordinate (RGB -> YIQ)
  - Multiplexing of YIQ into composite video (QAM)
  - Multiplexing with other TV channels (VSB)
- Digitization of raster video
  - Sampling rate selection
  - Color-subsampling (4:2:2, 4:2:0, etc)
  - Deinterlacing
- Video coding
  - Motion compensated prediction
  - Motion estimation using block matching
  - Different coding modes (I-, P-, B-modes)
Topics Covered After First Exam

- Video coding standards
  - Difference among their applications and techniques
- Digital TV development
  - Why All digital approach win?
  - US DTV vs. Europe DVB standards
  - Policy and business issues involved
- Multimedia networking
  - Challenges for multimedia services
  - Streaming vs. interactive services
    - Difference in QoS requirements
    - Protocols: RTP, RTSP, SIP
- Watermarking and Stegnography
  - Application and requirements
  - Simple algorithms
Video Coding Standards

- **H.261:**
  - First video coding standard, targeted for video conferencing over ISDN
  - Uses block-based hybrid coding framework with integer-pel MC and loop filtering
- **H.263:**
  - Improved quality at lower bit rate, to enable video conferencing/telephony below 54 b CPS (modems, desktop conferencing)
  - Half-pel MC and other improvement
- **MPEG-1 video**
  - Video on CD and video on the Internet (good quality at 1.5 mbps)
  - Half-pel MC and bidirectional MC
  - Use GOP structure to enable random access
- **MPEG-2 video (H.262)**
  - SDTV/HDTV/DVD (4-15 mbps)
  - Focus on motion estimation and DCT coding for interlaced video
  - Different scalability modes
- **MPEG4**
  - Object based coding
- **H264/MPEG4-AVC JVT**
- You should know the major differences between above standards in both targeted applications and techniques used
H264/AVC

- Higher coding efficiency
- Motion estimation
  - Up to 1/8 pel motion estimation
  - Multiple reference frames
  - Variable block size (down to 4x4)
- Integer transform 4x4 to 16x16
- Intra prediction
- Deblocking filtering
- Significantly higher coding efficiency
- Significantly higher complexity
Advanced TV System Development

- Compatible vs. simulcast approaches
- Analog HDTV Development in Japan and Europe
- All digital approach win over in US!
  - Grand Alliance System becoming the DTV standard
- DTV system in US
- DVB system in Europe
DTV System in More Detail

- Video encoding: Support both SDTV and HDTV
  - MPEG2 video (mp@ml for SDTV, mp@hl for HDTV)
- Audio encoding: Dolby AC3 (5.1 channel)
  - Perceptual audio coding
- Multiplex and transport: MPEG2 transport stream syntax
- Channel coding and modulation:
  - Concatenated channel coding (RS + convolution codes)
  - ASK + VSB for modulation
  - Terrestrial: 8-level VSB, enabling 19 mbps payload in 6 MHz channel
  - Cable: 16-level VSB, enabling 38 mbps payload in 6 MHz channel
MPEG-2 Video Coding

- Benefit of having I-, P-, and B-frames in GOP structures
  - Coding efficiency
  - Support for random access and channel surfing
    - Know the relation between distance between I, P, B pictures and the delay in channel switching
  - Error resilience
- Must handle interlaced video sequences
  - Special motion estimation mechanisms
  - Special DCT scan format
The GOP structure allows random access within the same video stream (for DVD). It also enables channel-surfing. After a viewer changes to a different TV channel, decoding starts at the first encountered I-picture in the new channel. For DTV, the recommended GOP length is \( \leq 0.5 \) sec, to have acceptable delay after channel-change.
AC3 Audio Coding Block Diagram

From Fig. 6.2, Guide to ATSC DTV standard (A54)
Perceptual Audio Coding: Main steps

- Decompose a signal into separate frequency bands by using a filter bank or a transform.
- The quantization step-size for each frequency band is set so that the quantization noise is just below the masking level of the human auditory system.
- The masking level is determined based on the signal level at this frequency (threshold-in-quiet) and on the signal levels at adjacent frequencies in the same time interval (frequency masking) and the signal levels at adjacent frequencies in the neighboring time interval (temporal masking).
- For each block of data, a bit-allocation unit is used to compute the necessary number of bits for each transformed coefficient based its own magnitude as well as the magnitudes of neighboring frequencies in this block, as well as those in adjacent blocks.
- Should understand the meaning of each masking property and how to make use of these to determine quantization stepsizes for different frequency components.
Audio and Video Coding in DVB

- Also uses MPEG-2 video coding
  - But with more focus on delivering SDTV
- Uses MPEG-2 audio coding as well
  - Only require MPEG-2 BC (equivalent to MPEG-1 layer 2, stereo sound)
Channel Coding: Basics

• To enable detection and correction of bit errors in the received bit stream
• CRC: allow error detection only
• FEC: allow correction of up to a certain number of errors, depending the code used and the redundancy introduced
  – Channel code rate $r = \frac{k}{n}$, $k$ information bits, $n$ coded bits
  – Block codes vs. convolutional codes
  – Concatenated codes
    • Why concatenate
• Data interleaving
  – randomize error bursts so that decoder only see randomly distributed errors
Deinterleaver randomize error bursts due to residual errors after inner decoder, so that the outer decoder only see randomly distributed errors
Channel Coding in DTV and DVB

- DTV Uses a concatenated channel coder
  - An RS outer coder, RS (207,187), symbol length=8. Can correct up to 10 symbol errors, $r=\frac{187}{207}$
  - An inner convolutional interleaver
  - A trellis inner coder (incorporates an inner block interleaver), $r=\frac{2}{3}$
  - Total channel code rate = $\frac{187}{207} \times \frac{2}{3} = 0.60225$
- DVB also uses a concatenated code
  - Outer code: RS
  - Inner code: punctured convolutional code
Modulation of Digital Signals

- For transmission of digital bits over analog channels
  - Convert group of digital bits into analog waveforms (symbols)
  - The analog waveforms are designed according to the desired carrier frequency
  - An analog channel of bandwidth $B$ can carry at most $2B$ symbols/s. For reduced inter-symbol interference, lower than $2*B$ symbol rate is used typically
  - Main techniques:
    - ASK (amplitude modulation)
    - QAM (phase and amplitude modulation)
Digital Modulation and Multiplexing

- Modulation:
  - Mapping digital bits into analog symbols (each being a sinusoidal waveform)
  - ASK (amplitude modulation)
  - QAM (phase and amplitude modulation)
- Multiplexing
  - Multiplex multiple channel signals from transmission over the same physical media (over the air, through a cable, etc)
  - Bandlimiting each signal, and shifting each to a different central frequency
  - VSB
  - OFDM
- Modulation and multiplexing can be accomplished simultaneously by choosing the frequency of the sinusoidal waveforms appropriately
Amplitude Shift Keying (ASK)

M-ary ASK: each group of $\log_2 M$ bits generates a symbol. The number corresponding to the symbol controls the amplitude of a sinusoid waveform. The number of cycles in the sinusoid waveform depends on the carrier frequency.

4-ASK: 2 bits/symbol (00=-3, 01=-1, 11=1, 10=3)

Example: Given a sequence: 01001011…, what is the analog form resulting from 4-ASK?

Symbol representation: “-1”, ”-3”, ”3”, ”1”

Waveform:
Quadrature Amplitude Modulation (QAM)

M-ary QAM uses symbols corresponding to sinusoids with different amplitude as well as phase, arranged in the two-dimensional plane.

Ex. 4-QAM (only phase change):

\[
00 = \cos(\omega_c t - \pi/4) \\
01 = \cos(\omega_c t - 3\pi/4) \\
10 = \cos(\omega_c t - 7\pi/4) \\
11 = \cos(\omega_c t - 5\pi/4)
\]
Example of 4-QAM

Example: Given a sequence: 01001011…, what is the analog form resulting from 4-ASK?

Using the previous mapping, the analog waveform for the above sequence is
16-QAM, etc.

16 QAM (4 bits/symbol):

64-QAM (6 bits/symbol)

- If the average energy per symbol is kept the same, more bits/symbol means more bits/joule
- But the distance between adjacent symbols differing by one bit is shorter \(\rightarrow\) less noise resistant
Modulation in ATSC System: 8-VSB

- 8-level amplitude shift keying
  - Each 3 bits generates a pulse, whose amplitude depends on the number corresponding to the 3 bits
- VSB: retaining only a small portion of the lower sideband, in addition to the upper band
  - The pulse train is multiplied with a sinusoidal wave at the carrier frequency
  - A shaping filter is applied to retain only a small portion of the LSB to save bandwidth
- Channel capacity and information rate
  - A 6MHz channel can deliver at most 12 Msymbols/s.
    - Recall a signal sampled at $f_s$ sample/s has a maximum bandwidth of $f_s/2$
  - But ATSC system uses 10.762 Msymbols/s.
  - With 8-ASK, each symbol carries 3 bits. Thus the date rate is $10.762 \times 3 = 32.286$ Mbps (including channel coding redundancy)
  - Information bit rate = overall data rate * channel code rate
    $= 32.286 \times 0.60225 = 19.44$ Mbps
DSB vs SSB vs VSB
Orthogonal Frequency Division Multiplexing (OFDM)

- Basic ideas:
  - Split the original signal with \( N \) samples/s into \( M \) sub-channels, each with \( N/M \) samples/s and occupying a different frequency band (subband decomposition)
  - The sub-signals must be synchronized so that remultiplexing is possible at the demodulator
  - The modulation waveforms for the individual sub-channels should be orthogonal

Fig.13.34 in Arnold
Example with 4 sub-channels

System Block Diagram:

N sample/s

DEMUX

Freq. Shift, \( f_0 \)
Freq. Shift, \( f_1 \)
Freq. Shift, \( f_2 \)
Freq. Shift, \( f_3 \)

N/4 sample/s, per sub-channel

Original Spectrum

Combined spectrum of the 4 sub-channels

From Fig.13.34 in [Arnold]
Modulation in DTV and DVB

- **DTV**
  - 8-level ASK (3 bits/symbol)
    - 6MHz channel, transmit 10.762 Msymbols/s; Raw date rate is 10.762*3=32.286 Mbps
    - Information bit rate = overall data rate * channel code rate =32.286*0.60225=19.44 Mbps
  - VSB multiplexing: each channel occupying only 6 MHz, including a small portion of the lower sideband, in addition to the upper band
- **DVB**
  - QAM: 4-QAM, 16-QAM, 64-QAM
  - OFDM (2048 or 8096 sub-carries/TV channel)
Channel Coding and Modulation and Multiplexing in DVB

- Also uses a concatenated code
  - Outer code: RS
  - Inner code: punctured convolutional code
- Use very different modulation technique
  - For conversion to complex symbols:
    - QPSK (=4-QAM, 2 bit/symbol)
    - 16 QAM (4 bit/symbol)
    - 64 QAM (6 bit/symbol)
  - Resulting complex symbols are split into multiple sub-channels, each using slightly shifted carrier frequency (OFDM)
  - Coded OFDM: refer to combination of the concatenated channel code and OFDM
8-VSB vs. C-OFDM

- **OFDM**
  - Fading and interference in one sub-carrier frequency (sub-signal) does not affect others
  - Use of interleaving/deinterleaving at multiplexer makes the burst error on one sub-channel spread randomly over the original signal, which can then be corrected by FEC decoding more effectively.
  - Also cause reduced interference to adjacent channels
  - Much better than 8-VSB to dynamic multipath fading (mobile receivers)

- **8-VSB**
  - Less costly to implement
  - Require lower transmitter power to achieve the same coverage area
  - More immune to impulse noise
Multimedia Networking

- Basic differences between TCP and UDP
- Different types of multimedia applications and their QoS requirements
- Protocols developed for different applications
Multimedia Applications

- Major differences from data services
  - Has stringent delay requirement (application dependent)
  - But can withstand certain degree of data loss
  - Bandwidth required depend on application
- Streaming (multicast or unicast, video on demand)
  - Player need to perform decoding in realtime
  - Can accept up to 10 second playout delay
  - Need a large buffer to smooth out network jitter
  - May need to support users with different access bandwidth (rate-scalable coding desired)
  - Pre-recorded media: Need to support fast forward/rewind
  - Live media: Server must encode video in real-time
- Interactive services (audio-visual conferencing/telephony, Internet gaming)
  - Each user need to perform both encoding and decoding in real time
  - Delay: 150 ms good, 400 ms acceptable
  - Head-shoulder type of sequences typically
RTP/RTCP

- **RTP is a packetization protocol**
  - RTP header contains: payload type identification, packet sequence numbering, timestamping
  - RTP packets are encapsulated in UDP segments
- **RTCP is a companion control protocol**
  - Each participant in a RTP session periodically transmits RTCP control packets (sender and receiver reports) to all other participants.
  - Control packet report statistics useful to application, including number of packets sent, number of packets lost, interarrival jitter, etc.
  - Sender may modify its transmissions based on feedback
- **RTP does not** provide any mechanism to ensure timely delivery of data or provide other quality of service guarantees.
- But together with RTCP, it allows monitoring of QoS so that sender and receiver can adjust their operations appropriately
Video Streaming

• What is network jitter?
• How to smooth jitter?
  – Use a large smoothing buffer
  – Buffer size vs. delay
• RTSP protocol
  – Application layer, governing client-server interactions (call set-up/termination, play, fast forward, rewind)
  – RTSP control messages use different port numbers than the media stream: out-of-band.
  – Does not restrict to specific audio-video codecs
  – Does not restrict how streamed media is transported; it can be transported over RTP/UDP, UDP or TCP
  – Does not specify how the media player buffers audio/video
Streaming Multimedia: Client Buffering

- Client-side buffering, playout delay compensate for network-added delay, delay jitter
Internet Audio-Visual Phone

- Also uses a buffer to smooth network jitter
  - But the buffer size cannot be too large because of the delay constraint
  - Late packets are considered lost
- SIP
- H.323
SIP Services

- Setting up a call
  - Provides mechanisms for caller to let callee know she wants to establish a call
  - Provides mechanisms so that caller and callee can agree on media type and encoding.
  - Provides mechanisms to end call.

- Determine current IP address of callee.
  - Maps mnemonic identifier to current IP address
  - Through SIP registrar and SIP proxy

- Call management
  - Add new media streams during call
  - Change encoding during call
  - Invite others
  - Transfer and hold calls
ITU-T H.323 Standard

- A protocol suite defining multimedia conferencing over Internet (non-QoS LAN)
  - Control and data channels sent using TCP
  - Audio and video streams sent using RTP/UDP
  - Specifies the operation of gatekeeper (registra), gateway, multipoint control unit (MCU)

<table>
<thead>
<tr>
<th>system</th>
<th>Audio coding</th>
<th>Video coding</th>
<th>Signaling</th>
<th>Control</th>
<th>Packetization/synchronization</th>
</tr>
</thead>
<tbody>
<tr>
<td>H.323</td>
<td>G.7xx</td>
<td>H.261/3</td>
<td>Q.931</td>
<td>H.245</td>
<td>H.225.0</td>
</tr>
</tbody>
</table>
SIP vs. H.323

- H.323 is an ITU standard for multimedia conferencing. It is a complete, vertically integrated suite of protocols: signaling, registration, admission control, transport (uses RTP/RTCP) and audio/video codecs.
- SIP is a single component, covering signaling, registration, admission control. Works with RTP, but does not mandate it. Works with various types of codecs.
- H.323 comes from the ITU (telephony).
- SIP comes from IETF: Borrows much of its concepts from HTTP. SIP has a Web flavor, whereas H.323 has a telephony flavor.
- SIP uses the KISS principle: Keep it simple stupid.
Data Hiding

• What is data hiding, its application and requirements?
  – Embedding other data in a cover media, e.g. data casting in Television, captioning, copyright protection, ownership assertion
  – Embed as many bits as possible while not altering appearance of cover media and being resistant to modification
    • Robustness, imperceptibility and payload (capacity)
  – Main applications:
    • watermarking, fingerprinting, authentication, and steganography
    • How do these differ in their goals and technical requirements?
Watermarking

- The watermark signal is used for intellectual property protection of the cover media
  - ownership assertion, copy prevention, change detection, etc.
- Watermark should be detectable after common image operations (compression/decompression, cropping, scaling, etc) as well as malicious attacks
  - Robust watermark
- Some basic techniques
  - Spatial domain modification
    - Watermark is embedded in the lowest (or a few lowest) bit planes of selected samples
  - Transform domain modification
    - Watermark is embedded in the mid-frequency coefficients of selected blocks
- Corresponding watermark detection techniques
Authentication

- The hidden data (a type of watermark) enables to carry verify the sender or integrity of the cover media.
- The watermark should withstand a certain common signal processing operations, but not malicious attacks.
  - Fragile watermark
  - What operations should be tolerable should be specified
  - Embed watermark in a way that is invariant under the specified signal processing operations
Fingerprinting

- How is it different than watermarking?
  - Associates the cover media to its holder
  - In addition, it requires the ability to resist collusion
- This is achieved by an intelligent design of the watermark
  - Many users cooperate to generate a new version of the cover through processing (averaging, copy-pasting, etc.) each individual’s copy
  - Each watermark is a codeword as in error correcting codes
  - These codes (when colluded) enable tracing colluders
**Steganography**

- What is steganography, its application and requirements?
  - Steganography: Embed secret messages in a cover media
  - Minimize the probability of detection of secret messages
  - Active vs. passive steganography
- What is steganalysis, its application and requirements?
  - Detect the presence of a secret message (not necessarily decipher the message)
  - Maximize the probability of detection of secret messages
  - Two types of steganalysis techniques and the trade-offs
    - Techniques designed for a specific embedding algorithm
    - Universal steganalysis
Watermarking vs. Steganography

- Relation between watermarking and steganography
  - Both embed data in a cover media, special cases of data hiding
  - Embedded data in watermarking is used to protect the cover media
    - Robust watermarking
    - Authentication (fragile watermarking)
    - Fingerprinting
  - Embedded data in steganography is unrelated to the cover media. The cover media is chosen randomly, or chosen to facilitate data embedding
Logistics

- Time: 12/8 (Thursday) 11-12:50
- Closed book, 1 sheet of notes (both sides) OK
- Office hour:
  - 12/2 (Tuesday) afternoon: 2:00-3:30
  - Wed: By email
- What to prepare for:
  - Should be able to answer all questions in HWs without referring to lecture notes/handouts
  - For complex block diagrams, you don’t need to memorize the actual block diagrams, but you should understand the functionality of each block, if a block diagram is given.