Bluray (http://www.blu-ray.com/faq)

- MPEG-2 enhanced for HD, also used for playback of DVDs and HDTV recordings
- MPEG-4 AVC part of the MPEG-4 standard also known as H.264 (High Profile and Main Profile)
- SMPTE VC-1 standard based on Microsoft's Windows Media Video (WMV) technology
- Video bitrate 40.0Mbps (vs ~10 Mbps for DVD)



MPEG-4

MPEG-4 adopts a object-based coding:

- Offering higher compression ratio, also beneficial for digital video composition, manipulation, indexing, and retrieval
- The bit-rate for MPEG-4 video now covers a large range between 5 kbps to 10 Mbps
- More interactive than MPEG-1 and MPEG-2



Comparison between Block-based Coding and Object-based Coding



2/11/09

CSE 40373/60373: Multimedia Systems

Composition and manipulation of object



Æ

Overview of MPEG-4

- Video-object Sequence (VS)—delivers the complete MPEG-4 visual scene, which may contain 2-D or 3-D natural or synthetic objects
- Video Object (VO) a object in the scene, which can be of arbitrary shape corresponding to an object or background of the scene
- Video Object Layer (VOL) facilitates a way to support (multi-layered) scalable coding. A VO can have multiple VOLs under scalable coding, or have a single VOL under non-scalable coding
- 4. Group of Video Object Planes (GOV) groups Video Object Planes together (optional level)
- Video Object Plane (VOP) a snapshot of a VO at a particular moment

Object oriented

- ▶ VOP I-VOP, B-VOP, P-VOP
- Objects can be arbitrary shape need to encode the shape and the texture (object)
 - Need to treat MB inside object different than boundary blocks (padding, different DCT etc)



Sprite Coding

- A sprite is a graphic image that can freely move around within a larger graphic image or a set of images
- To separate the foreground object from the background, we introduce the notion of a sprite panorama: a still image that describes the static background over a sequence of video frames
 - The large sprite panoramic image can be encoded and sent to the decoder only once at the beginning of the video sequence
 - When the decoder receives separately coded foreground objects and parameters describing the camera movements thus far, it can reconstruct the scene in an efficient manner



Global Motion Compensation (GMC)

- "Global" overall change due to camera motions (pan, tilt, rotation and zoom)
 - Without GMC this will cause a large number of significant motion vectors
- There are four major components within the GMC algorithm:
 - Global motion estimation
 - Warping and blending
 - Motion trajectory coding
 - Choice of LMC (Local Motion Compensation) or GMC.

		Typical	Bit-rate	Max number
Profile	Level	picture size	(bits/sec)	of objects
Simple	1	176 imes 144 (QCIF)	64 k	4
	2	352 imes 288 (CIF)	128 k	4
	3	352×288 (CIF)	384 k	4
Core	1	176 imes 144 (QCIF)	384 k	4
	2	352 imes 288 (CIF)	2 M	16
Main	1	352×288 (CIF)	2 M	16
	2	720×576 (CCIR601)	15 M	32
	3	1920 imes 1080 (HDTV)	38.4 M	32

2/11/09

0

#

CSE 40373/60373: Multimedia Systems

page 10

MPEG-7

The main objective of MPEG-7 is to serve the need of audio-visual content-based retrieval (or audiovisual object retrieval) in applications such as digital libraries

MPEG-7 video segment





Chapter 13: VOCODER

- Voice only coder use aspects of human hearing
 - E.g. Formant Vocoder voice is not equal represented in all frequencies because of vocal cord
 - They can produce good quality sound in 1,000 bps
- concerned with modeling speech so that the salient features are captured in as few bits as possible
 - use either a model of the speech waveform in time (LPC (Linear Predictive Coding) vocoding
 - break down the signal into frequency components and model these (channel vocoders and formant vocoders)



 Simulations are improving but still recognizable (automated phone calls)

Phase insensitivity

A complete reconstituting of speech waveform is really unnecessary, perceptually: all that is needed is for the amount of energy at any time to be about right, and the signal will sound about right.



Solid line: Superposition of two cosines, with a phase shift. Dashed line: No phase shift. wave is different, yet the sound is the same, perceptually

Linear Predictive Coding (LPC)

LPC vocoders extract salient features of speech directly from the waveform, rather than transforming the signal to the frequency domain

LPC Features:

- uses a time-varying model of vocal tract sound generated from a given excitation
- transmits only a set of parameters modeling the shape and excitation of the vocal tract, not actual signals or differences → small bit-rate



Chapter 14: MPEG Audio

Psychoacoustics

- Frequency: Remove audio that are masked anyway
 - A lower tone can effectively mask (make us unable to hear) a higher tone
 - The reverse is not true a higher tone does not mask a lower tone well
 - The greater the power in the masking tone, the wider is its influence – the broader the range of frequencies it can mask
 - As a consequence, if two tones are widely separated in frequency then little masking occurs
- Temporal Phenomenon: any loud tone will cause the hearing receptors in the inner ear to become saturated and require time to recover

14.2 MPEG Audio

- MPEG audio compression takes advantage of psychoacoustic models, constructing a large multidimensional lookup table to transmit masked frequency components using fewer bits
 - Applies a filter bank to the input to break it into its frequency components
 - In parallel, a psychoacoustic model is applied to the data for bit allocation block
 - The number of bits allocated are used to quantize the info from the filter bank – providing the compression

MPEG Layers

- Each succeeding layer offering more complexity in the psychoacoustic model and better compression for a given level of audio quality
 - Layer 1 quality can be quite good provided a comparatively high bit-rate is available
 - Digital Audio Tape typically uses Layer 1 at around 192 kbps
 - Layer 2 has more complexity; was proposed for use in Digital Audio Broadcasting
 - Layer 3 (MP3) is most complex, and was originally aimed at audio transmission over ISDN lines
- Most of the complexity increase is at the encoder, not the decoder – accounting for the popularity of MP3 players

Summary

- Apply different set of heuristics (than video), yet achieve the goal of attaining good compression by removing components that the human ear is not good at distinguishing
- More complexity at the encoding phase leads to better compression
- Humans are lot more sensitive to dropped "frames" in audio than in video. Audio should also be well synchronized - otherwise distracting
- Humans also like audio better than video, TVs/ DVDs send higher fidelity audio