18-796

Multimedia Communications:
Coding, Systems, and Networking

Prof. Tsuhan Chen
tsuhan@ece.cmu.edu

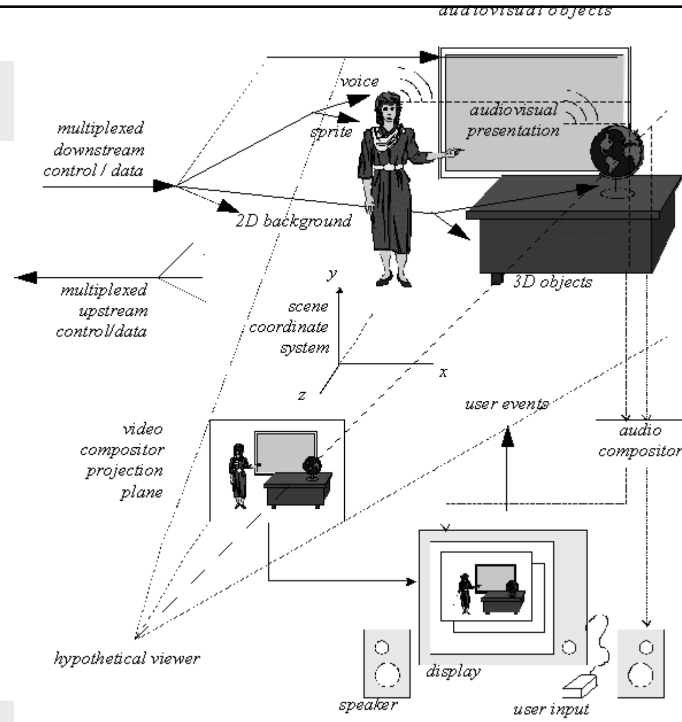
MPEG-4

MPEG-4

- Originally
  - A standard for very low bit rate coding of limited complexity audio-visual material
- In July 94, the scope was extended to
  - Functionalities not supported by other standards
    - Content-based interactivity
    - Universal access
    - High compression
  - Coding of general material for a wide bit rate range
  - Flexibility and extensibility

Content-Based Interactivity

- A scene is composed of audio-visual objects
  - Not just pixels or moving blocks
- Objects can be of different nature
  - Text or images
  - Rectangular or arbitrary shape
  - 2D or 3D objects
  - Natural of synthetic
- Different coding schemes applied to different objects
- Compositor puts objects back in a scene

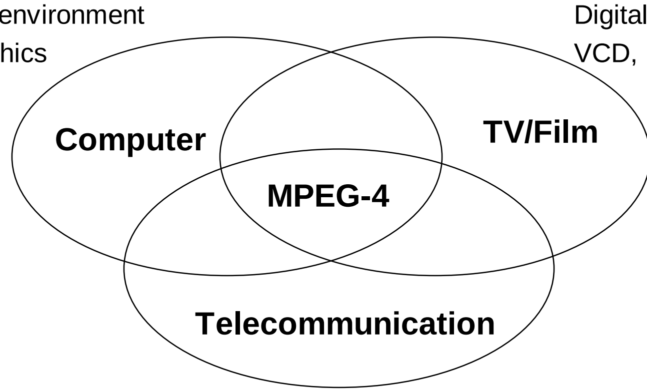


[MPEG-4 N1909, "Overview of the MPEG-4 Version 1 Standard"]

Applications

Human-machine interface
GUI, Virtual environment
Vision, Graphics

Content creation
Digital TV, HDTV
VCD, DVD



Wireless, Internet, ISDN, POTS, Cable

Parts of MPEG-4

- Part 1: Systems
- Part 2: Video
- Part 3: Audio
- Part 4: Conformance testing
- Part 5: Reference software
- Part 6: Delivery multimedia integration framework
- Others
  - Synthetic and Natural Hybrid Coding (SNHC)
  - Requirements and applications
  - Implementation Study
  - Intellectual property rights (IPR)

MPEG-4 Activities

- Competitive phase
  - Proposals and evaluations
- Collaborative phase
  - Verification model and core experiments
    1. Define Verification Model (VM-n)
    2. Define core experiments for improving VM-n
    3. Perform core experiments. Compare with VM-n
    4. n++, go to Step 1

MPEG-4 Time Table

- July 93 Started work
- Nov 95 Subjective tests and tool evaluation
- Jan 96 Define verification model (VM) and core experiments (CE)
- Mar/July/Sept/Nov 96, Feb/Apr/Jul 97
  Update VM and define a new set of CEs
- Oct 97 Committee Draft (CD)
- July 98 Final CD (FCD)
- Nov 98 Draft international standard (DIS)
- Jan 99 International standard (IS)

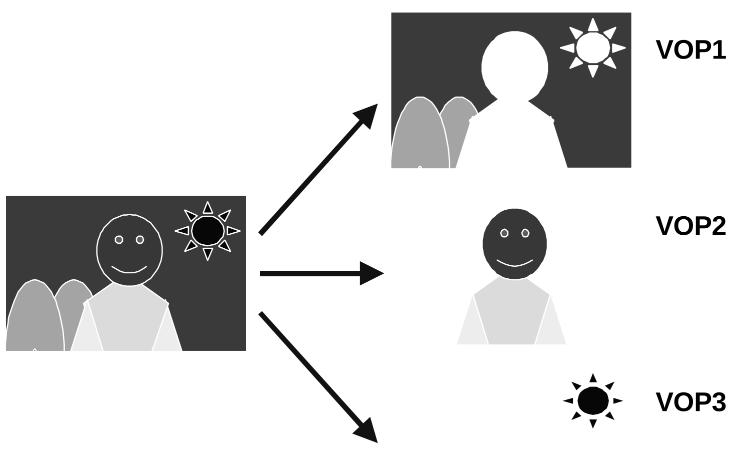
MPEG-4 Video

- General functionalities
  - Coding efficiency
    - For 5 kbit/s ~ 5 Mbit/s
    - From small images to TV resolution
    - Progressive/interlaced
  - Error resilience and robustness
  - Spatial and temporal scalabilities
- Content-based functionalities
  - Shape coding and sprites
  - Content-based scalabilities
  - Error resilience and robustness

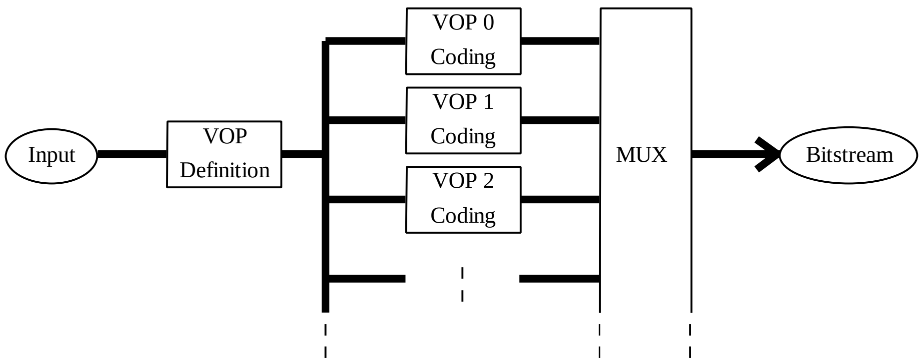
MPEG-4 Video (cont.)

- Tools
  - Motion/texture coding derived from H.263
  - Coding of video object plane (VOP): I, B, P
  - Binary and gray-scale shape coding
  - Scalabilities: temporal/spatial
  - Static sprites
  - Interlaced prediction
  - 12 bit video
  - Computational graceful degradation (CGD)

Video Object Plane (VOP)

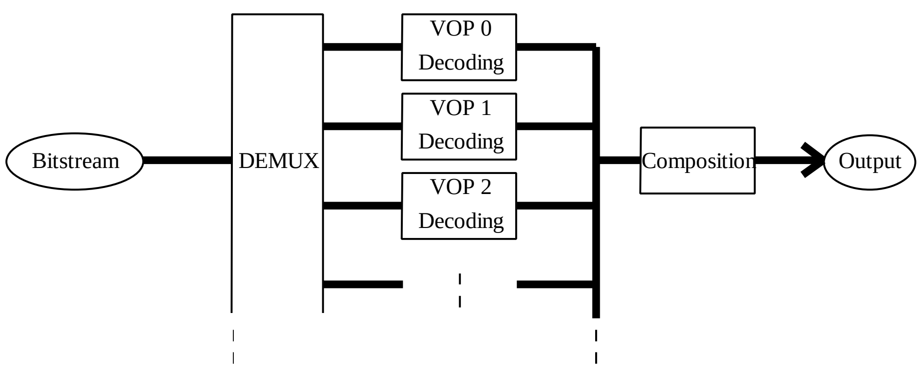


Structure of VOP Encoder



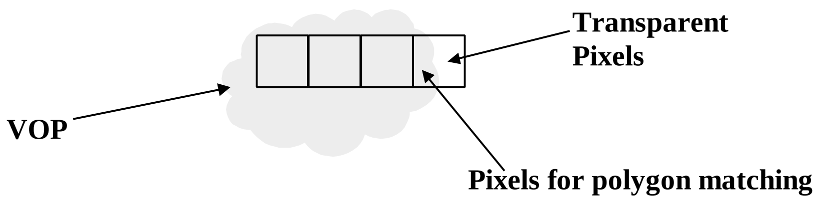
– Note: Segmentation is outside the scope of MPEG-4

Structure of VOP Decoder



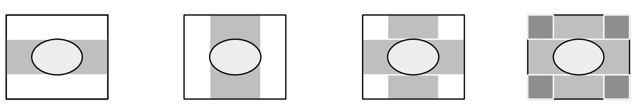
Coding of VOP

- Motion compensation and DCT
  - Similar to H.263
- Polygon matching for motion estimation



VOP Padding

- Applied to the reference VOP prior to motion estimation/compensation

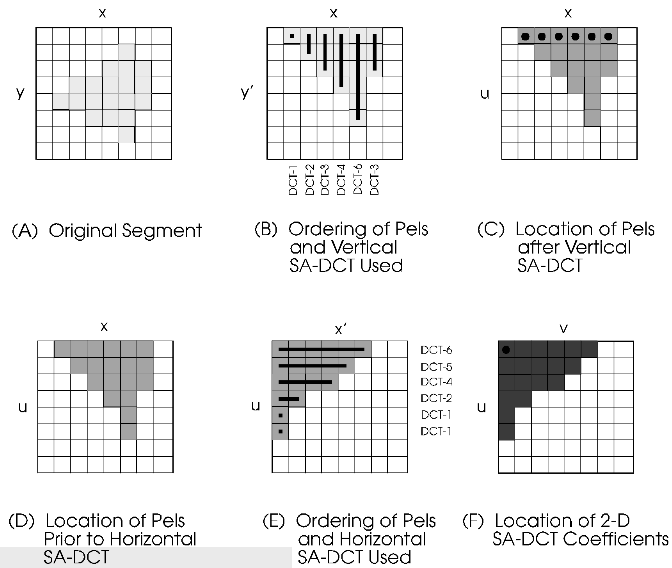


Padded Previous Frame

Current Frame

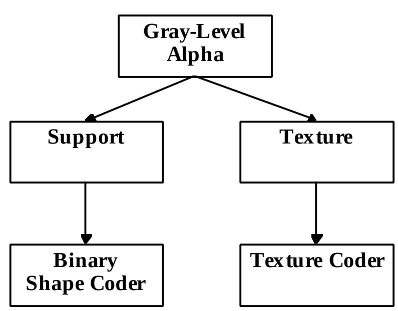
18-796/Spring 1999/Chen

Shape-Adaptive DCT for Texture Coding



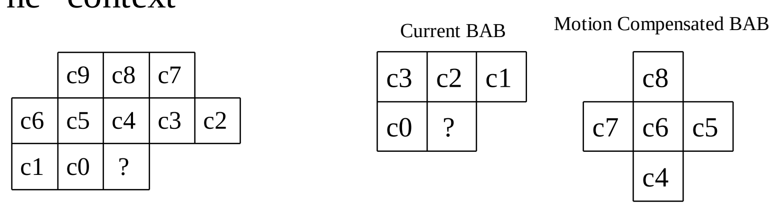
Shape Coding

- Binary shape
  - Context-based arithmetic encoding (CAE)
- Gray scale alpha plane
  - Motion compensated DCT
    - Similar to texture coding



Binary Shape Coding

- Context-based arithmetic encoding (CAE)
  - A binary shape is treated as a binary image
  - Apply CAE to each binary alpha block (BAB)
- The “context”

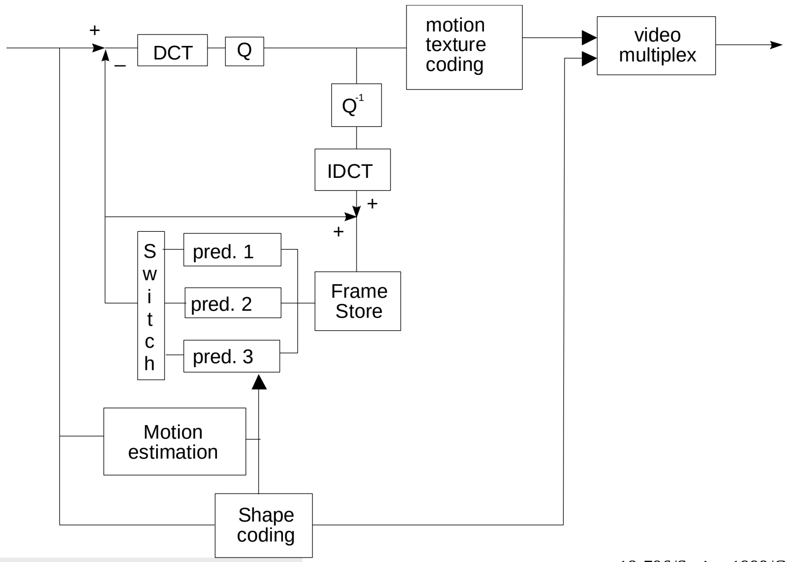


Intra   Inter

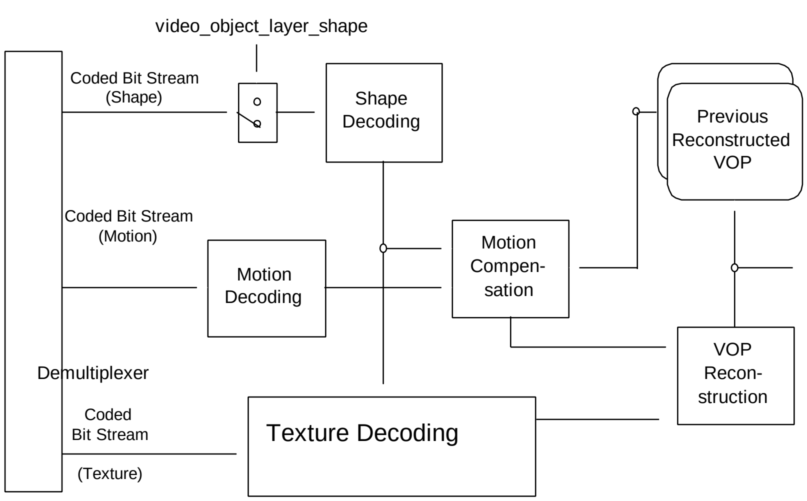
Feathering

- Feathering and translucency coding
  - No effects
  - Linear feathering
  - Constant alpha
  - Linear feathering and constant alpha
  - Feathering filter
  - Feathering filter and constant alpha

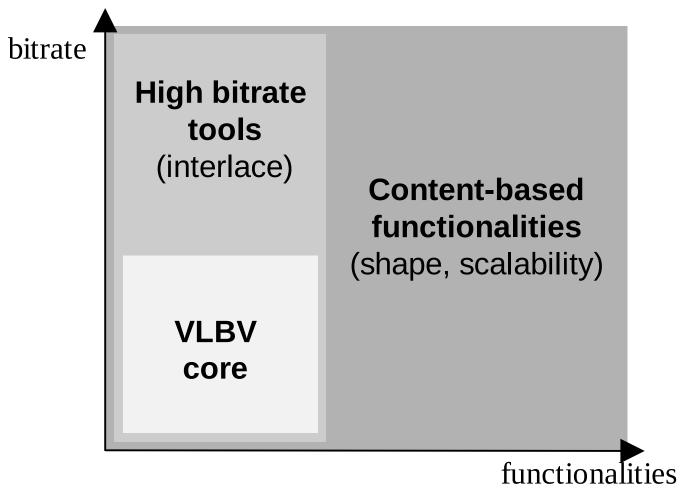
VOP Encoder



VOP Decoder

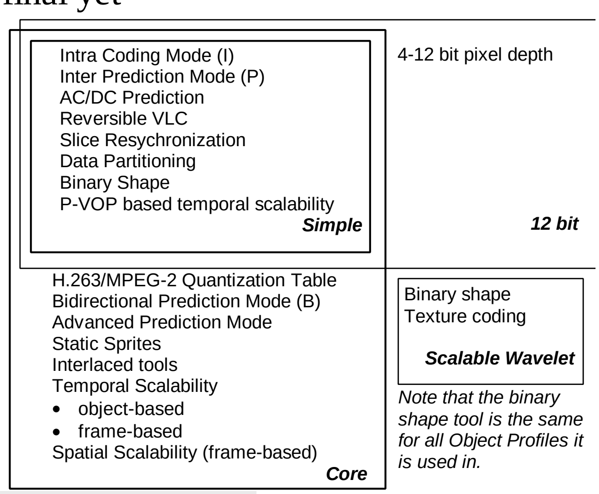


Profiles of MPEG-4

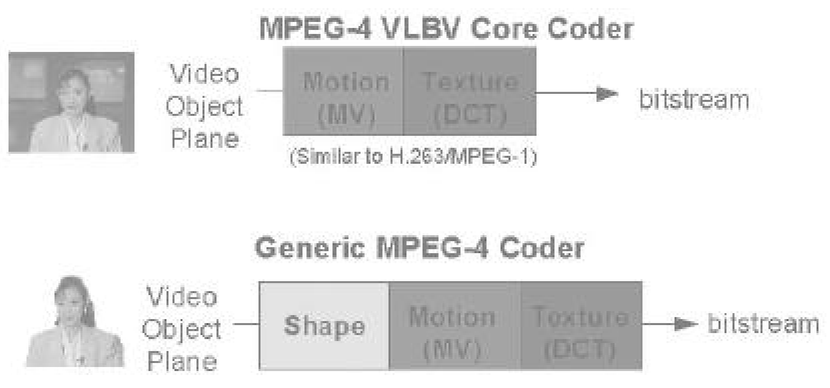


Visual Object Profiles

- Not final yet



Core vs. Generic MPEG-4



Synthetic & Natural Hybrid Coding (SNHC)

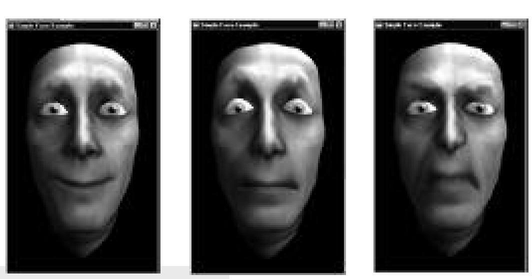
- Efficient representation and composition of synthetically and naturally generated audiovisual data
- To be integrated into MPEG-4 Video and Audio
  - Not a separate part of MPEG-4
- Applications
  - Virtual environment, conferencing, education, entertainment, media production, and real-time, interactive and broadcast media experiences

SNHC Target technologies

- Video
  - Face animation
  - 2D/3D mesh compression
  - Wavelet-based still texture coding
  - View dependent scalability
- Audio
  - Text-to-speech synthesis, structured audio, environmental auralization, 3D audio, etc.

Face Animation

- Face animation
  - 2D/3D polygon mesh for face rendering
  - Facial Definition Parameter (FDP) Set
    - Controls shape, texture, gender, age, etc.
  - Facial Animation Parameter (FAP) Set
    - 68 parameters to produce animation and to create expressions



Demo

Examples of FAP and FDP

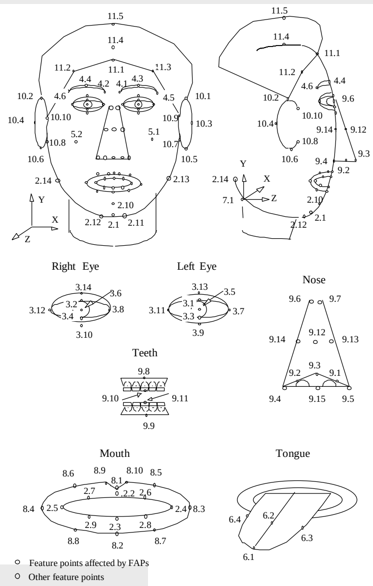
```
class FaceDefinitionParams {
  public:
     3Dmesh shape;
     3DPoint featurePoint[46];
     Image texture;
     int age;
     int gender;
     ...;
}

class FaceAnimationParams {
  public:
    int move_h_l_eyeball;
    int move_h_r_eyeball;
    int move_v_l_eyeball;
    int move_v_r_eyeball;
    int enlarge_l_pupil;
    int enlarge_r_pupil;
    ...;
}
```


Example FAPs

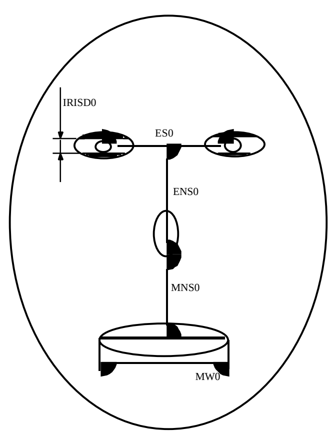
#	FAP name	FAP description	units	Uni- or Bidir	Pos motion	Grp	FDP subgrp num	Quant step size
1	viseme	Set of values determining the mixture of two visemes for this frame (e.g. pbm, fv, th)	na	na	na	1	na	1
2	expression	A set of values determining the mixture of two facial expression	na	na	na	1	na	1
3	open_jaw	Vertical jaw displacement (does not affect mouth opening)	MNS	U	down	2	1	8
4	lower_t_midlip	Vertical top middle inner lip displacement	MNS	B	down	2	2	5
5	raise_b_midlip	Vertical bottom middle inner lip displacement	MNS	B	up	2	3	5
6	stretch_l_cornerlip	Horizontal displacement of left inner lip corner	MW	B	left	2	4	5

Facial Definition Parameters



Facial Animation Parameter Units (FAPUs)

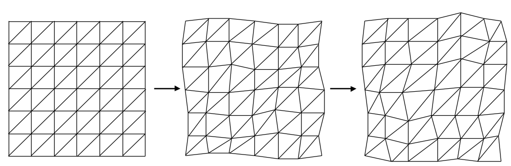
- Parameters normalized to FAPUs



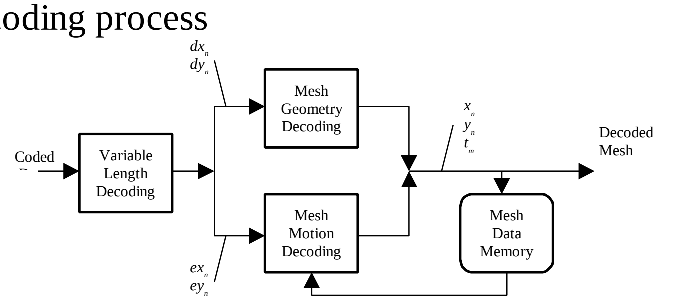
Visemes for Lip Synch

viseme_select	phonemes	example
0	none	na
1	p, b, m	put, bed, mill
2	f, v	far, voice
3	T,D	think, that
4	t, d	tip, doll
5	k, g	call, gas
6	tS, dZ, S	chair, join, she
7	s, z	sir, zeal
8	n, l	lot, not
9	r	red
10	A:	car
11	e	bed
12	I	tip
13	Q	top
14	U	book

2D Mesh Compression

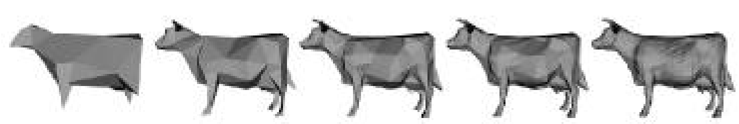


- Decoding process



3D Mesh Compression

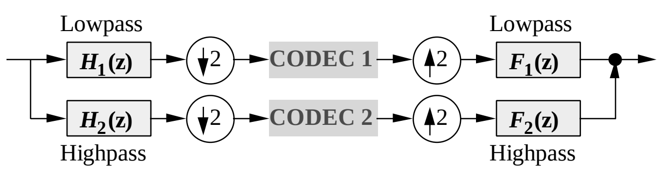
- Progressive representation
  - Streaming of 3D objects
  - Both spatially and temporally



- Indexing and retrieval of 3D meshes
  - Multiresolution databases
  - Related to MPEG-7

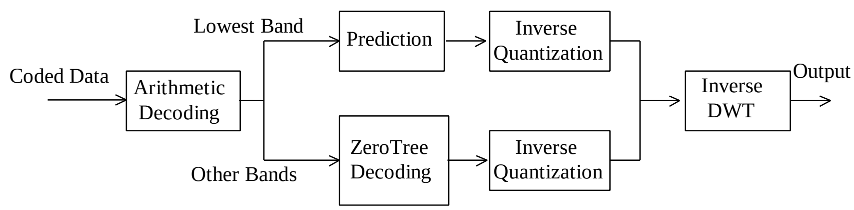
Demo

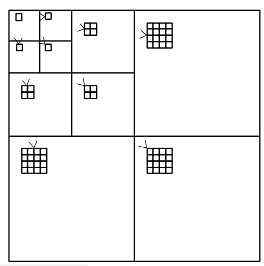
Wavelets for Scalable Texture Coding



- Decompose the signal in the frequency domain
- Critical downsampling maintains the number of samples in the subbands
- For 2D case, separable filters are often used. Decompose into four bands: LL, LH, HL, HH
- Decompose the LL band iteratively

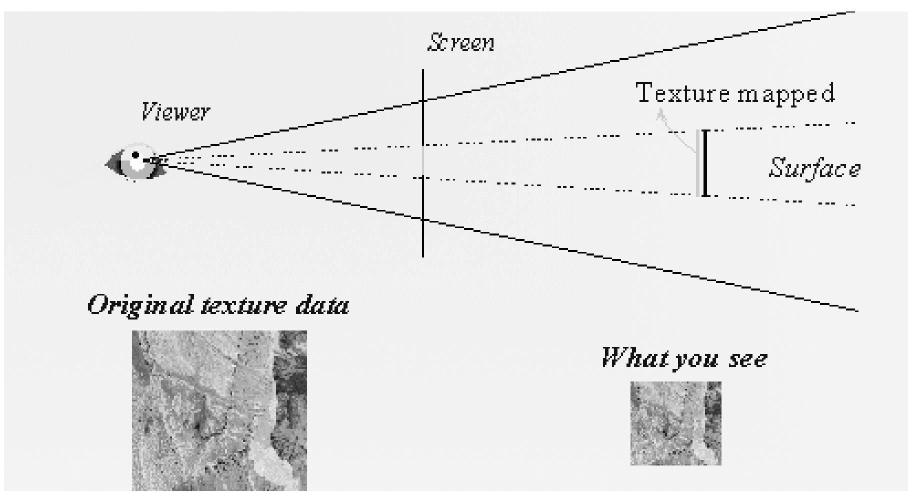
Wavelets for Scalable Texture Coding (cont.)





"Zero Tree"

View Dependent Scalability



MPEG-4 Video Test Sequences

- Class A: Low spatial detail and low amount of movement
- Class B: Medium spatial detail and low amount of movement or vice versa
- Class C: High spatial detail and medium amount of movement or vice versa
- Class D: Stereoscopic
- Class E: Hybrid natural and synthetic
- Class F: 12-bit video sequences

Test Sequences

Sequence Name	Class	Input Format	YUV files	Alpha files	Segment Mask Available
Mother & daughter	A	ITU-R 601 (60Hz)	1	0	0
Akiyo	A	ITU-R 601 (60Hz)	2+1	1	2
Hall Monitor	A	ITU-R 601 (60Hz)	1	0	3
Container Ship	A	ITU-R 601 (60Hz)	1	0	6
Sean	A	ITU-R 601 (60Hz)	1	0	3
Foreman	B	ITU-R 601 (50Hz)	1	0	0
News	B	ITU-R 601 (60Hz)	4+1	3	4
Silent Voice	B	ITU-R 601 (50Hz)	1	0	0
Coastguard	B	ITU-R 601 (60Hz)	1	0	4
Bus	C	ITU-R 601 (60Hz)	1	0	0
Table Tennis	C	ITU-R 601 (50Hz)	1	0	0
Stefan	C	ITU-R 601 (60Hz)	1	0	2
Mobile & Calendar	C	ITU-R 601 (60Hz)	1	0	0
Basketball	C	ITU-R 601 (50Hz)	1	0	0
Football	C	ITU-R 602 (60Hz)	1	0	0
Cheerleaders	C	ITU-R 601 (60Hz)	1	0	0
Tunnel	D	ITU-R 601 (50Hz)	2x1	0	0
Fun Fair	D	ITU-R 601 (50Hz)	2x1	0	0
Children	E	ITU-R 601 (60Hz)	3+1	2	3
Bream	E	ITU-R 601 (60Hz)	3+1	2	3
Weather	E	ITU-R 601 (60Hz)	2+1	1	2
Destruction	E	ITU-R 601 (60Hz)	11+1	10	0
Ti1	F	176x144 (15Hz)	1	0	0
Man1sw	F	272x136 (15Hz)	1	0	0
Hum2sw	F	272x136 (15Hz)	1	0	0
Veh2sw	F	272x136 (15Hz)	1	0	0
labview	F	176x144 (60Hz)	1	0	0
hallway	F	176x144 (60Hz)	1	0	0

MPEG-4 Version 2

- One year following Version 1
- Adds new profiles with new functionalities
- Video
  - Scalable transmission of arbitrary-shaped objects
  - Tools for additional efficiency improvements
  - Tools for improved error robustness
  - Coding of multiple views
  - Body animation
  - Coding of 3D meshes and scalabilities

References

- MPEG Home Page http://drogo.cselt.it/mpeg/
- MPEG Video Home Page http://wwwam.hhi.de/mpeg-video/
- MPEG4-SNHC http://www.es.com/mpeg4-snhc/
- T. Sikora, "MPEG digital video coding standards," *IEEE Signal Processing Magazine*, Sept. 1997, pp. 82-100
- T. Sikora, "The MPEG-4 Video Standard Verification Model," IEEE Trans. CSVT, Vol.7, No.1, Feb.1997