

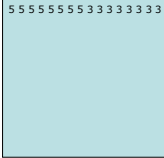
CmpE 464 Image Processing

Lecture 6 Image and Video Compression

Image Compression

- Images occupy a large space. Often, image data is highly redundant.

5 5 5 5 5 5 5 3 3 3 3 3 3 3 3 3



8 5's then 8 3's
Send 4 codewords instead of 16
"Runlength coding"

Runlength coding is lossless
In other words, image data
can be reconstructed exactly.

Image Compression

- The still image and motion images can be compressed by *lossless coding* or *lossy coding*.
- Principle of compression:
 - reduce the redundant information, e.g.,
 - coding redundancy*
 - interpixel redundancy*
 - psychovisual redundancy*

Image Compression: Coding redundancy

Variable length coding (entropy coding)

- Use code-words with different lengths to losslessly represent symbols with different probabilities
- Why?
 - Use small number of bits to represent more frequent symbols and use large number of bits to represent less frequent symbols so that the length of overall symbols can be reduced.
- How:
 - Huffman coding
 - Arithmetic coding

Image Compression: Coding redundancy

- Code length
 - fixed length
 - variable length
- The average code-length is calculated as:

$$L_{avg} = \sum_{k=0}^{L-1} l(r_k) p_k(r_k)$$

$P_k(r_k)$ is the probability of the occurrence of event r_k ,
 $l(r_k)$ is the code length of event r_k

Image Compression: Coding redundancy

- Example of variable length coding

r(k)	p(r(k))	code 1	L1	Code 2	L2
r(0)=0	0.19	000	3	11	2
r(1)=1/7	0.25	001	3	01	2
r(2)=2/7	0.21	010	3	10	2
r(3)=3/7	0.16	011	3	001	3
r(4)=4/7	0.08	100	3	0001	4
r(5)=5/7	0.06	101	3	00001	5
r(6)=6/7	0.03	110	3	000001	6
r(7)=1	0.02	111	3	000000	6

$L_{(avg)}=2(0.19)+2(0.25)+2(0.21)+3(0.16)+4(0.08)+5(0.06)+6(0.03)+6(0.02) = 2.7\text{bits}$
 Redundancy: $R(D)=1-2.7/3 = 0.1 = 10\%$

Information Theory

- Information measurement $I(E) = \log \frac{1}{P(E)} = -\log P(E)$

Note:

Event E which occurs with probability P(E)
contains I(E) units of information

e.g., P(E) = 1 \rightarrow I(E) = 0 (which means there is no uncertainty
for event E, it always happens)

e.g., If we set the base as 2 $\rightarrow -\log_2 P(E)$
if we flip a coin $\rightarrow P(E) = 1/2$,
then I(E) = $-\log_2(1/2) = 1$ (\rightarrow "bit", unit of information)
(e.g., flipping a coin)

Information Theory

- Entropy of the source (uncertainty):

$$H(z) = -\sum_{j=1}^J P(a_j) \log P(a_j)$$

H(Z) is the average amount of information;

H(Z) \uparrow \rightarrow uncertainty \uparrow \rightarrow information \uparrow

If the source symbols occur with equal probability,
the entropy is maximum

- Theoretically, H(Z) is the minimum code length for each symbol,
which is the so-called Shannon's first theorem for noiseless image coding

Image Compression: Coding redundancy

- If some bit patterns are more likely to appear than others, represent them with shorter codewords:

Example:

bit pattern	probability	codeword
00	50 %	0
01	30 %	10
10	15 %	110
11	5 %	111

Average codeword length: $0.5 \times 1 + .25 \times 2 + .20 \times 3 + .05 \times 3$
 $= 1.7$ bits/2 bits

- This is "Huffman coding" and it is used in many compression routines, including JPEG. One disadvantage is that probabilities may change over time.

Huffman code

- Variable-length coding

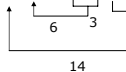
-- Huffman coding generates the smallest possible number of code
-- example:

Symbol	Probability	1	2	3	4 (source reduction)	code
A2	0.4	0.4	0.4	0.4	$\rightarrow 0.6$ (code 0)	1
A6	0.3	0.3	0.3	0.3	$\rightarrow 0.4$ (code 1)	00
A1	0.1	0.1	$\rightarrow 0.2$	$\rightarrow 0.3$		011
A4	0.1	0.1	$\rightarrow 0.1$			0100
A3	0.06	$\rightarrow 0.1$				01010
A5	0.04					01011

Image Compression: Coding redundancy

- One disadvantage of Huffman coding is that probabilities may change over time. A good idea is to build up a dictionary of common bit patterns in time and to refer to entries in the dictionary.

- LZ77: she sells sea shells by the sea shore



- LZ77 used for winzip, pkzip
- LZ78: more advanced version
- LZW latest version: Unix compress utility, GIF image files

Image Compression: Interpixel redundancy

- Inter-pixel correlation

- pixel values can be guessed (predicted) on the basis of the value of its neighbors

- The following redundancy can be reduced:
 - spatial redundancy
 - geometric redundancy
 - inter-frame redundancy

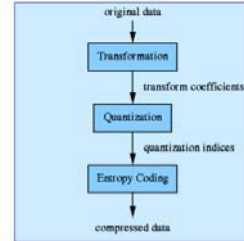
- Human perception redundancy

Image Compression: Interpixel redundancy

- Transform techniques: One idea is to transform the image into another space where important information may be condensed into a few coefficients.
- Examples: Discrete Fourier Transform (DFT), Discrete Cosine transform (DCT), Discrete Wavelet transform (DWT), Karhunen Loeve Transform (KLT)
- Transform based methods are generally lossy
- JPEG uses DCT

Image Compression

Elements of Image/Video Compression

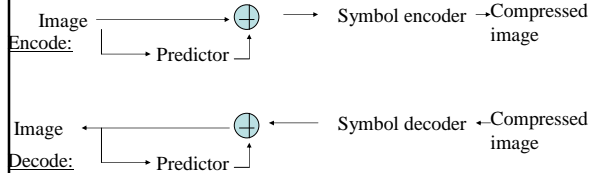


Three stage image/video coding structure.

Example coder-decoder

- Lossless predictive coding
- Lossy predictive coding
- Transform coding

-- Example of lossless predictive coding:



Example coder-decoder

- Transform coding

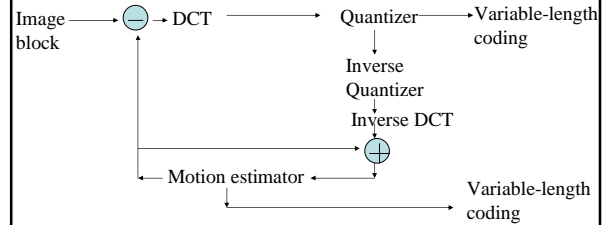


Image Compression: Psychovisual redundancy

- Quantization

-- Map a large number of input amplitude levels to a small number of amplitude levels with non-recoverable loss of quality

-- Why?

Reduce the amplitude levels can reduce the number of bits to represent each pixel

-- How:

- scalar quantization (SQ)
- vector quantization (VQ)

Image Compression

Example of SQ

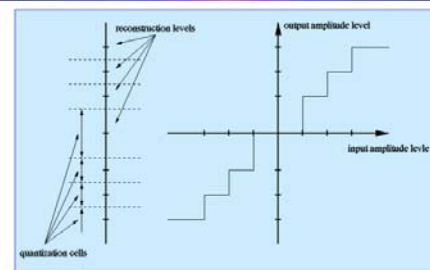


Image Compression

Examples of VQ

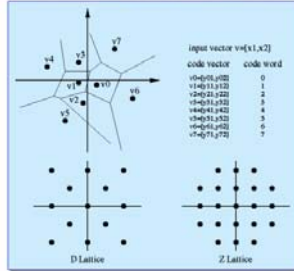


Image Compression

Example of (8x8) 2-D DCT

$$X[k_1, k_2] = \frac{1}{4} \sum_{n_1=0}^7 \sum_{n_2=0}^7 C_{k_1} C_{k_2} x[n_1, n_2] \cos\left(\frac{(2n_1+1)k_1\pi}{16}\right) \cos\left(\frac{(2n_2+1)k_2\pi}{16}\right)$$

$$x[n_1, n_2] = \frac{1}{4} \sum_{k_1=0}^7 \sum_{k_2=0}^7 C_{k_1} C_{k_2} X[k_1, k_2] \cos\left(\frac{(2n_1+1)k_1\pi}{16}\right) \cos\left(\frac{(2n_2+1)k_2\pi}{16}\right)$$

$$C_{k_1}, C_{k_2} = \begin{cases} \frac{1}{\sqrt{2}} & \text{for } k_1, k_2 = 0 \\ 1 & \text{otherwise} \end{cases}$$

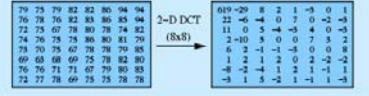
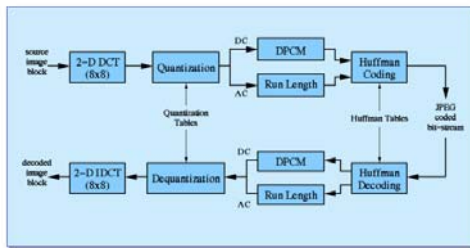
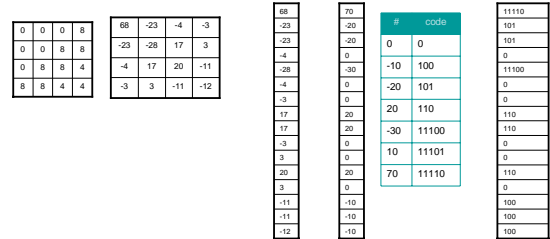
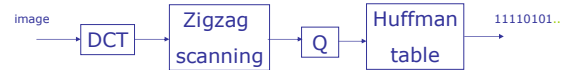


Image Compression

Joint Photographic Experts Group (JPEG)



JPEG Compression



JPEG Decompression

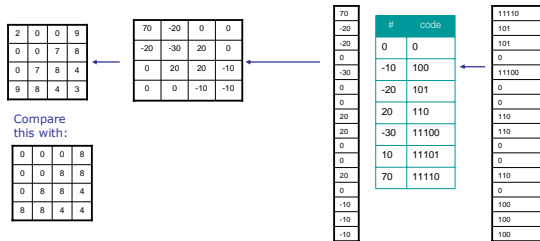


Image Compression

Examples of Compressed Images (I)

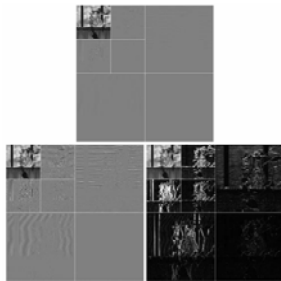
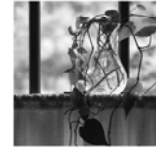
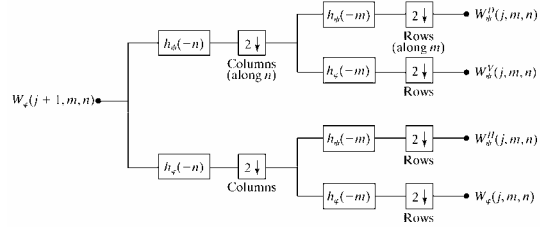


JPEG Coder

JPEG 2000 Coder

Bit Rate 0.25 bpp ⇒ compression ratio 32:1

Wavelet Transforms



Pyramid coding

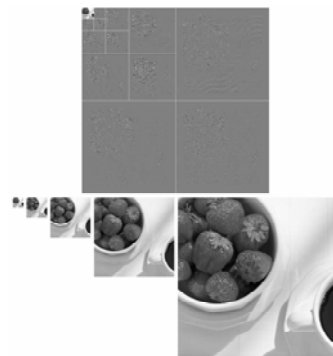
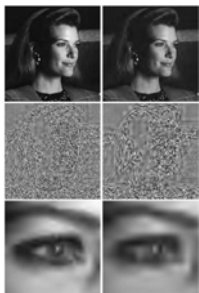


FIGURE 7.8 Progressive reconstruction: (a) A four-scale wavelet transform; (b) the fourth-level approximation image from the upper-left corner; (c) a refined approximation incorporating the fourth-level details; (d) through (f) further resolution improvements incorporating higher-level details.

JPEG 2000



a b
c d
e f

FIGURE 8.16 Left column: JPEG 2000 approximations of Fig. 8.4 using five scales and implicit quantization with $\mu_0 = 8$ and $\epsilon_0 = 8.5$. Right column: Similar results with $\epsilon_0 = 7$.

Video bit rates

- NTSC video: $640 \times 480 \times 3$ bytes \times 30
= 26 Mbytes/s
- PAL video: $768 \times 576 \times 3 \times 25$
= 31 Mbytes/s
- CIF: 360×88 for Y
 180×88 for U and V
frame rate 30, 15, 10, 7.5
37.3 Mbps
- QCIF: $180 \times (144 \text{ or } 72)$ for Y
 $90 \times (144 \text{ or } 72)$ for U and V
frame rate 30, 15, 10, 7.5
9.35 Mbps

Video Compression Standards

- H.261 (1990): videoconferencing
- MPEG-1 (1992): multimedia, storage
- MPEG-2 (1994): all-digital TV
- MPEG-4 (1998): networked multimedia applications

MPEG-1

- Storage of video on CD-ROM, DAT, disk and optical drives
- Input: CIF format video and audio; other formats also supported with some restrictions
- Rate: about 1.5 Mbps for CIF
- 1.2 Mbps compressed CIF video has quality of VHS video
- Video compression algorithm has three modes: Inter and Intra
- Intra mode: Similar to JPEG; block-based DCT
- Inter mode: temporal prediction with motion compensation followed by DCT encoding of the prediction error.
- MPEG-1 also offers random access capability and coding/decoding delay of about 1 sec.

Image Compression

Example of Motion Compensation (1)

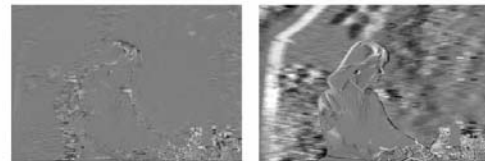


Previous frame

Current frame

Image Compression

Example of Motion Compensation (3)

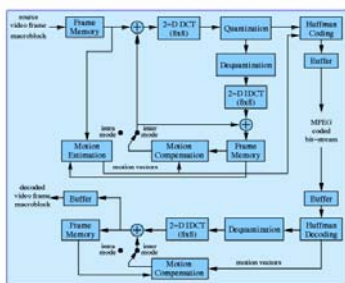


Residual frame with MC

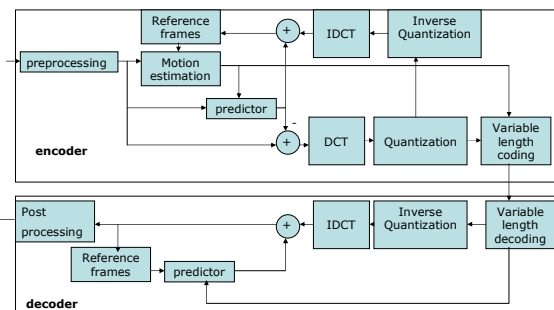
Residual frame with direct difference

Image Compression

Moving Pictures Experts Group (MPEG)

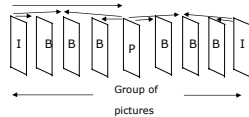


MPEG-1 Block Diagram

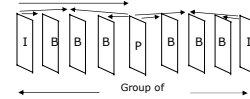


MPEG-1 Compression Modes

- I-pictures: intra-frame DCT encoded using a JPEG-like algorithm
- P-pictures: forward predicted relative to other I- or P-pictures
- B-pictures: predicted backward, forward or bidirectionally relative to other I- or P-pictures
- D-pictures: contain only the DC component of each block; serve browsing purposes.



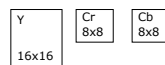
MPEG-1 Compression Modes



- N: # pictures from one I to the next
- M: # pictures from one anchor point (I or P) to the next
- In this example; N=8; M=4; in general, variable.
- Typical ratio of bit rates for I:P:B = 5:3:1
- Transmit order in this example: I0 B-2 B-1 B1 P4 B2 B3 B5 I8 B6 B7

MPEG-1 Data Structure

1. Sequences: several group of pictures. Defines picture size, rate, expected buffer sizes, nondefault quantizer matrices.
2. Group of pictures: made up of pictures.
3. Pictures: I, B or P
4. Slice: unit of synchronization: group of macroblocks followed by a resynchronization pattern
5. Macroblock: The unit of motion compensation 16 x 16 block of Y combined with the corresponding 8 x 8 chroma components. The macroblock header gives the prediction mode for each block, macroblock address, coded block pattern, and possibly, a quantizer step size variable.
6. Block: Unit of DCT. 8 x 8 collection of pixels
7. Macroblocks:



MPEG-1 Intraframe Compression Mode

- I-pictures encoded similar to JPEG
- DCT coefficients are quantized with a uniform quantizer obtained by dividing the DCT coefficient value by the quantization step size and then rounding the result.
- MPEG allows for spatially-adaptive quantization. Macroblocks containing busy, textured areas can be quantized more coarsely.
- Default quantization matrix used
- Redundancy among the quantized DC coefficients is reduced via DPCM; the resulting signal is VLC coded with 8 bits.
- Quantized AC coefficients are zig-zag scanned and converted to (run, level) pairs as in H.261
- Truncated Huffman code used for all blocks. No provision for custom tables

MPEG-1 Interframe compression mode

- B-pictures: Transform the prediction error $I_1(x) - I_1'(x)$ where:

$$I_1'(x) = 128$$

$$I_1'(x) = I_0'(x + mv_{01}) \text{ for forward predicted}$$

$$I_1'(x) = I_2'(x + mv_{21}) \text{ for backward predicted}$$

$$I_1'(x) = 0.5[I_0'(x + mv_{01}) + I_2'(x + mv_{21})] \text{ bidirectionally predicted}$$

- Quantization for prediction errors: uniform quantization is used for all DCT coefficients because low frequencies are removed by the prediction.

MPEG-1 Summary

MPEG-1 Encoder Algorithm:

- Decide on the labeling of I-, P- and B-pictures in a GOP
- Estimate a motion vector for each MB in the P- and B-type pictures
- Determine the compression mode
- Set the quantization scale, if adaptive quantization is selected.

Audio Basics

Question: What is the bandwidth of audio?

Answer: speech 10,000Hz
 telephone quality speech 4000 Hz
 quality music 20,000 Hz.

- Some common sampling rates:
 - 8 kHz
 - 11.025 kHz
 - 11.127 kHz
 - 22.05 kHz
 - 44.1 kHz CD quality
 - 48 kHz

Audio Basics

Quantization: Each sample must be represented with a codeword of finite length. The number of bits in the codeword is called the "bit depth".

- 12 bits - speech
- 16 bits - CD quality music

Channels: Audio may have one or more channels:

- 1 channel: Mono
- 2 channels: Stereo
- 4 channels: quadrophnic
- 6 channels: surround sound

Audio Compression

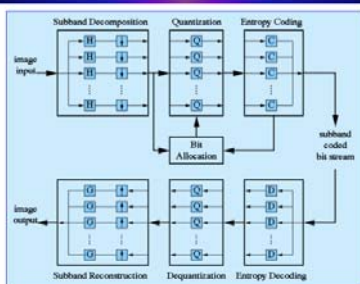
- 60 minutes of CD quality audio occupies
 - 2 channels x 44100 samples/s x 16 bits/sample = 1.411 Mbps
 - $1.411 \times 60 \times 60 = 5079.6$ Mbits = 634.95 Mbytes
- The time to transmit only 30 seconds of the above source on a 64kbps channel: 661.4 s ~ 11 minutes
- Compression is a MUST
- Lossy compression: The decompressed signal is not the same as the original; but the ear may not hear it
- Compression ratio:
 - CD quality audio has BW 1.411 Mbps
 - mp3 compresses that to 2 x 64kps
 - $1.411 \times 10^6 / 64 \times 10^3 = 22/2 = 11$
 - This is expressed as 10:1

Audio Compression

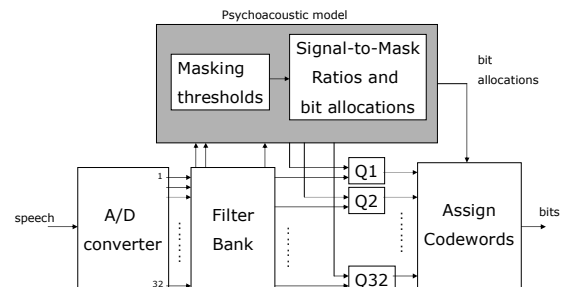
- Compression makes use of 3 things:
 - Intersample redundancy: transform techniques, predictive coding, etc. exploit this
 - Coding redundancy: Codes may not be equally probable: LZW, Huffman coding exploit this
 - Perceptual redundancy: The ear may not hear certain errors
- The ear is more sensitive to errors in silence: Companding makes use of this
- Frequency sensitivity: The ear is most sensitive to signals in the 2-5 kHz range
- Frequency masking
- Temporal masking

Audio Compression

Subband Image Coding



MPEG Audio Coder



MPEG Layers 1,2,3

Layer	Application	Compressed bit rate	Quality
1	Digital Audio Cassette	32 - 448 kbps	hi-fi quality at 192 kbps per channel
2	Digital audio broadcast	32 - 192 kbps	near CD quality at 128 kbps per channel
3	CD quality audio over low bitrate channels	64 kbps per channel	CD quality at 64 kbps

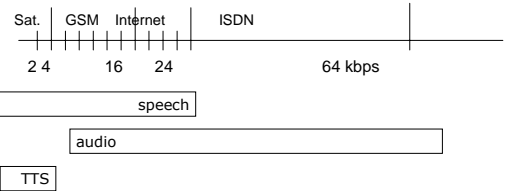
MPEG-2

- Video compression standard for broadcast TV
- Different bit rates: layers and profiles
- Extension of MPEG-1 which allows:
 1. interlaced inputs; alternative way of subsampling chroma channels
 2. Scalability
 3. Improved quantization and coding options
- Three chroma subsampling formats: 4:2:0 (same as MPEG-1)
4:2:2 (horizontal mode)
4:4:4 (no chroma subsampling)
- Two new picture formats: frame-picture and field-picture
- Field/frame DCT option per MB for frame pictures
- New MC prediction modes for interlaced video

MPEG-4

- Finalized in 1998
- For very low bit rate multimedia applications
- Aims:
 1. Compression efficiency: very low bit rate (5-64 kbits/s)
 2. Object based: scalability based on objects: different objects encoded at different temporal and spatial scales
 3. Error robustness: must support mobile channels
 4. Synthetic Natural Hybrid Coding (SNHC)
 5. Downloadability: ability to download tools

MPEG4 Overview



MPEG4: What is new?

Content Providers: Reuseability, flexibility, copyright

Networks: Streaming; Embedded information; signalling (e.g. for Qos)

End users: Interaction with content; access on low bandwidth (e.g. mobile) channels

MPEG4 achieves these by:

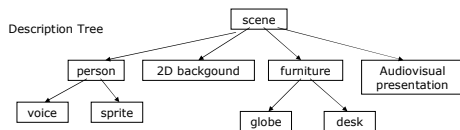
1. Content is represented by *media objects*: natural or synthetic.
2. Description tree like in VRML: you can create compound media objects
3. Multiplex and synchronize the data associated with media objects
4. End user can interact with the audiovisual scene

MPEG4 Media Objects

- Still images (background)
- Video objects (e.g. a person talking w/o background)
- Audio objects (the voice of a person)
- Associated graphics
- Text
- Talking synthetic heads
- Synthetic speech
- Synthetic sound
- Structured audio orchestra language

Composition of Media Objects

- One can define compound media objects
- Place media objects anywhere in a coordinate system
- Apply transforms to them
- Apply streamed data to change attributes (e.g. to add sound, to change texture, animation for a synthetic face)
- Change viewing and listening points interactively



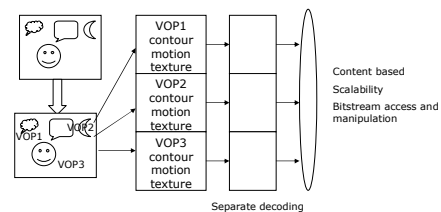
Streaming Data

- Quality of Service: maximum bit rate, bit error rate, priority
- Object content information, intellectual property rights
- Synchronization, time stamping

Interaction

- Interaction level specified by author
- Possibilities:
 - Navigation through a scene by changing viewpoint / listening point
 - Drag and drop objects
 - Trigger an event by clicking on an object
 - Select desired language when available

MPEG-4



MPEG-4

- Syntax defined by MPEG4 System Description Language (MSDL)
- Toolbox Approach:
 1. Tools: address a module of the coding system (e.g. DCT, SBC, etc)
 2. Algorithms: address one or more functionalities (such as improved compression)
 3. Profiles: Standardized set of tools for certain functionalities
- Video Objects (VO): can be manipulated separately
- Arbitrary shape object coding
- Composition of different objects (alpha channel contains transparency information)
- Binary alpha planes: indicate the shape and location of object
- Gray-scale alpha planes: Transparency

MPEG-4 Coding of shape, motion, texture

- Block based hybrid/DPCM transform coding
- Video Object Planes:
 1. I-VOP: Intraframe VOP
 2. P-VOP: Predicted VOP
 3. B-VOP: bidirectionally predicted VOP
- Motion compensation based on macroblock basis
- DCT \Rightarrow Quantization \Rightarrow Runlength coding \Rightarrow Entropy coding
- Shape coding: alpha planes

MPEG-4 SNHC

- Synthetic Natural Hybrid Coding
- Representation and coding of natural and synthetic objects.
- Example: weather forecast
 - Anchorperson: real video object: sprite
 - Satellite weather map
 - Graphics on top of weather map
 - Synthetic set
 - Real or synthesized voice

MPEG-4 SNHC

- Human face and body description and animation
- Integration of animated text and graphics
- Coding of scalable textures
- 2D and 3D mesh coding
- Video planes and shapes as separate scalable objects
- Hybrid scalable text-to-speech
- Synthetic audio coding
- 2D and 3D synthetic graphical constructs
- The ability to build scene compositions from instances of the elementary streams mentioned above

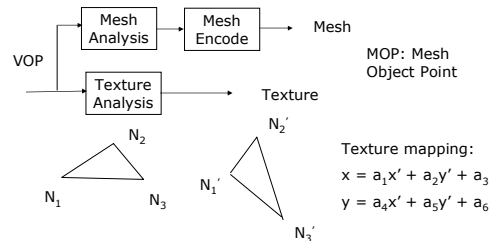
2D Mesh Animation

- Improved coding efficiency
- Editing texture
- Content based indexing
- Augmented reality

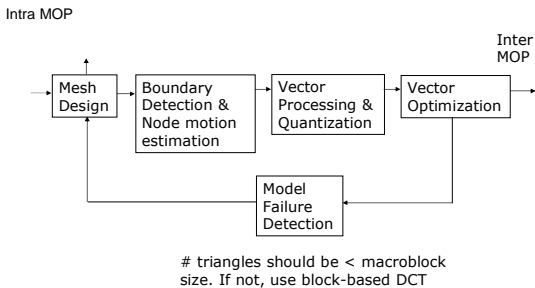
Uniform mesh vs. Content based



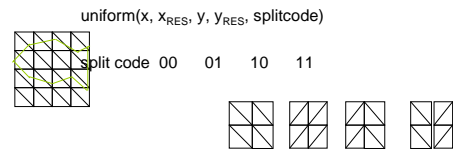
2D Mesh Animation



2D Mesh Analysis



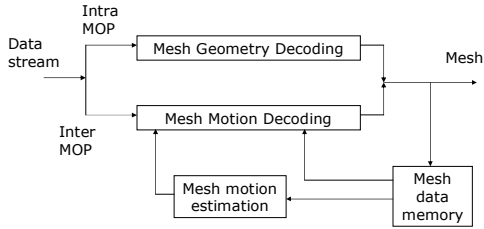
Uniform Mesh vs Content based Mesh



Content based: Delaunay triangulation

- Minimize total edge length
- Maximize (min angle in patches)
- N_b : boundary nodes
- N_i : interior points: high gradient points, corner points

2D Mesh Based Animation Decoder



MPEG-4 SENTETİK-DOĞAL KARIŞIK KODLAMA

MPEG-4 SNHC

- MPEG-4 Sahne Düzeni
- Yüz Parametreleri
- Yüz Canlandırma

AGU Sistemi

- Ses Analizi
- Yüz Kasları ve Fiziksel Yüz Katmanları

MPEG-4 SNHC Uygulaması

- Uygulamanın Yapısı



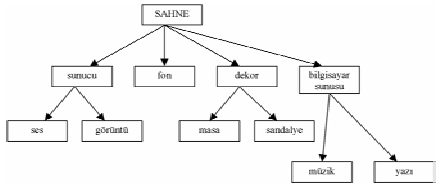
MPEG-4 Katmanlı Sahne Yapısı

Nesneler

- Video
- 2 veya 3 boyutlu değişken şekiller
- Gerçek yada sentetik ses
- Sentetik Yüz

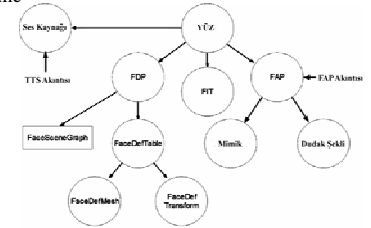
Transformasyonlar

- Nesneye özel tanımlanmış canlandırma teknikleri
- Yüz canlandırma



MPEG-4 Sentetik Doğal Kodlama (SNHC)

- Her MPEG-4 terminali içinde bir yüz modeli
- Yüz Kalibrasyonu
 - Yüz modeli uyarlama
 - Doku yükleme
- FAP akıntısı

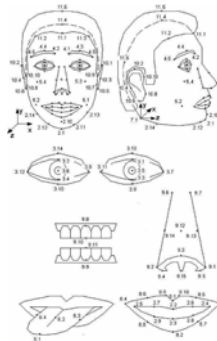


Yüz Tanımlama Parametreleri (FP)

Nötr Yüz Tanımlaması

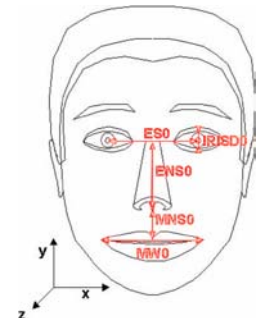
- Bakış yönü z yönündedir.
- Bütün yüz kasları rahatlamış haldedir.
- Göz kapakları irise dik durumdadır.
- Göz bebeği gözün 1/3ü büyüklüğündedir.
- Dudaklar kapalı ve dudak çizgisi doğrusaldır.
- Dişler birbirine değmektedir.
- Dil düzdür, ucu dişlerin değdiği noktaya değmektedir.

=> 84 adet referans noktası



Yüz Canlandırma Parametresi Birimi (FAPU)

IRISD0	İris çapı	IRISD=IRISD0/10 24
ES0	Göz açıklığı	ES=ES0/1024
ENS0	Göz-burun açıklığı	ENS=ENS0/1024
MNS0	Ağız-burun açıklığı	MNS=MNS0/1024
MW0	Ağız genişliği	MW=MW0/1024
AU	Açı birimi	10E-5 rad



Yüz Canlandırma Parametreleri (FAP)

Grup	FAP Sayısı	
1: dudak şekli ve mimik	2	
2: çene ve dudaklar	16	
3: göz bebekleri ve kapakları	12	Üst seviye Parametreler
4: kaş	8	
5: yanak	4	FAP1 : 15 Dudak Şekli
6: dil	5	FAP2 : 6 Mimik
7: kafa pozisyonu	3	
8: dudak kenarları	10	
9: burun	4	
10: kulak	4	
	+	
	68	



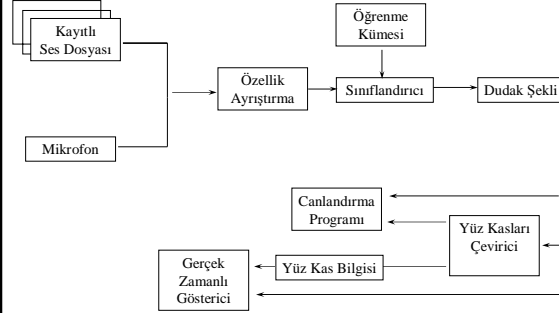
Yüz Canlandırma Tablosu (FAT)

FAP #			
	FP#	yön vektörü	FAPU cinsinden miktarı

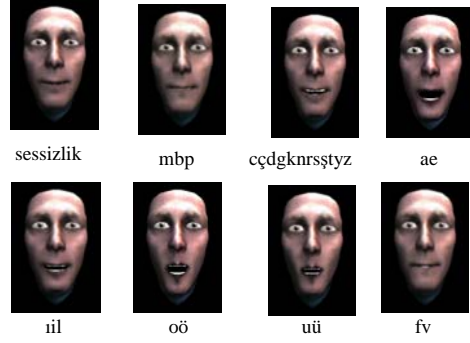
FAP interpolasyon tablosu (FIT) => bant genişliğini azaltmak için

- Simetrikler (sağ ve sol yüz yarısı)
- Sol kaş yukarı kalkınca sağ kaşta kalkıyor
- Örnek: Dudak kenarı yukarı kalkınca alt dudakta yukarı çıkıyor

AGU Sistem Yapısı



Türkçe İçin Dudak Şekilleri



Sınıflandırıcı

- 20 ms pencereler (10 ms örtüşen)
- 12 mel cepstral parametre + log enerji
- Tek konuşmacılı mükemmel öğrenme kümesi
- İleri ağaç sınıflandırıcısı
 - 3NN, FuzzyNN ve parametrik sınıflandırıcılar

=> %76 başarı (tek konuşmacı)



Hata Düzeltici

Dudak şekilleri bir süre korunmalıdır.

=> Kısa süreli dudak sınıflandırmaları potansiyel hatalı seçimdir

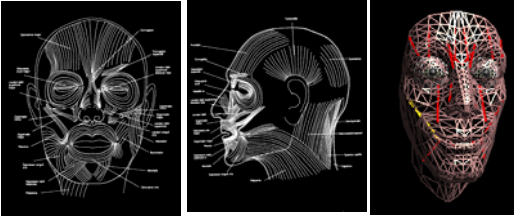


sınıflandırma hataları

Sınıflandırıcı sonuçlarından median filtre geçirilir



Yüz Kasları



10 doğrusal ve 1 eliptik kas dudak çevresinde modellendi
Toplam 21 yüz kası modellendi



Yüzün Fiziksel Yapısı



Epidermis
Yağ Tabakası
Kemik

$$F = s \Delta x \quad s : \text{yay sertliği}$$

Kaslar ilk yağ tabakasını etkiler. Uygulanan gerilim bağlantılı katmanlar arasında yayılır, epidermisi (asıl 3D model) etkiler



“O” Harfinin Söylenişi



Obicularis Oris olmadan



Obicularis Oris ile



Alt deri katmanları hesaplanarak

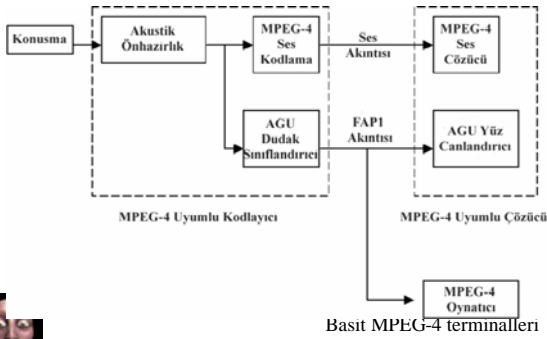


FAP1 - AGU

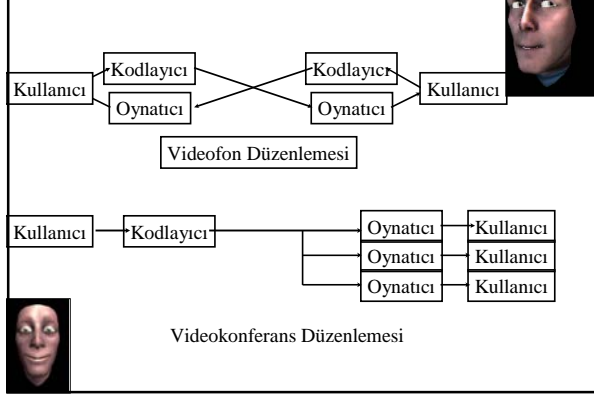
Dudak Şekli #	fonem	örnek
0	none	na
1	p, b, m	put, bed, mill
2	f, v	far, voice
3	T, D	think, that
4	t, d	tip, doll
5	k, g	call, gas
6	tS, dZ, S	chair, join, she
7	s, z	sir, zeal
8	n, l	lot, not
9	r	red
10	A:	car
11	e	bed
12	I	tip
13	Q	top
14	U	book



MPEG-4 Uygulaması 1/2



MPEG-4 Uygulaması 2/2



H.264

- H.264, MPEG4 v10, JVT, AVC: Same things
- 500-600 kbps at entertainment quality
- Content adaptive video coding
- Allocate more bits to relevant content: relevance feedback needed