
Compact Lecture

Multimedia Coding: Methods & Applications

Part 4: Video Coding Basics

4.3 Model Based Video Coding

Dr. Klaus Illgner

Dr. Uwe Rauschenbach

Model Based Coding

Aim:

Modeling the original 3D scene and determining the model parameters from the 2D image data (compare Vocoder for speech coding)

→ **scene description**

image reconstruction by animation

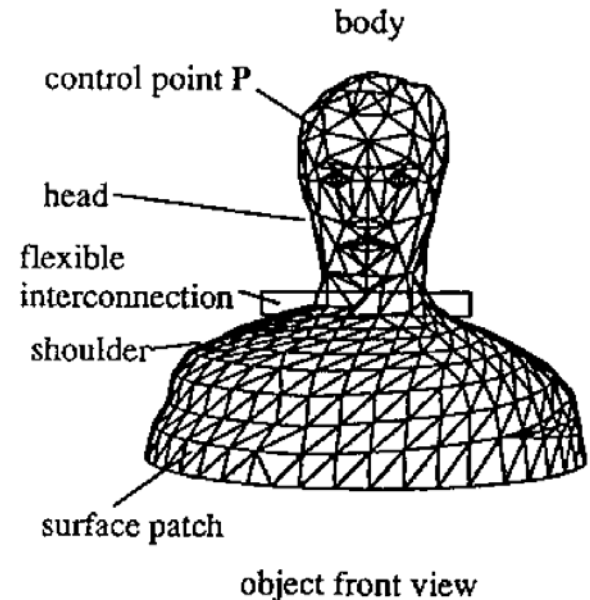
→ **advantage of very low data rate**

only applicable for specific scenes

Parameter estimation

e.g. via analysis by synthesis (compare CELP)

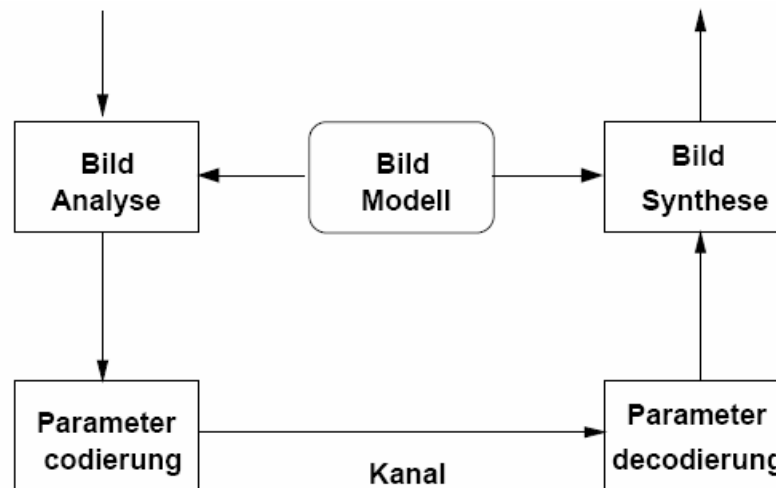
Model	Surface	Volume
parametric	splines	Generalized cylinder
Non-parametric	Wire model	splines



Model based Coding (2)

Coding

- Explicit object models → semantic coding approach
e.g. animated, talking heads
- Implicit models → object based coding approach
modeling the scene only to an extent, which allows to describe the 2D scene
→ e.g. analysis to synthesis coding



Example of a MPEG-4 Scene



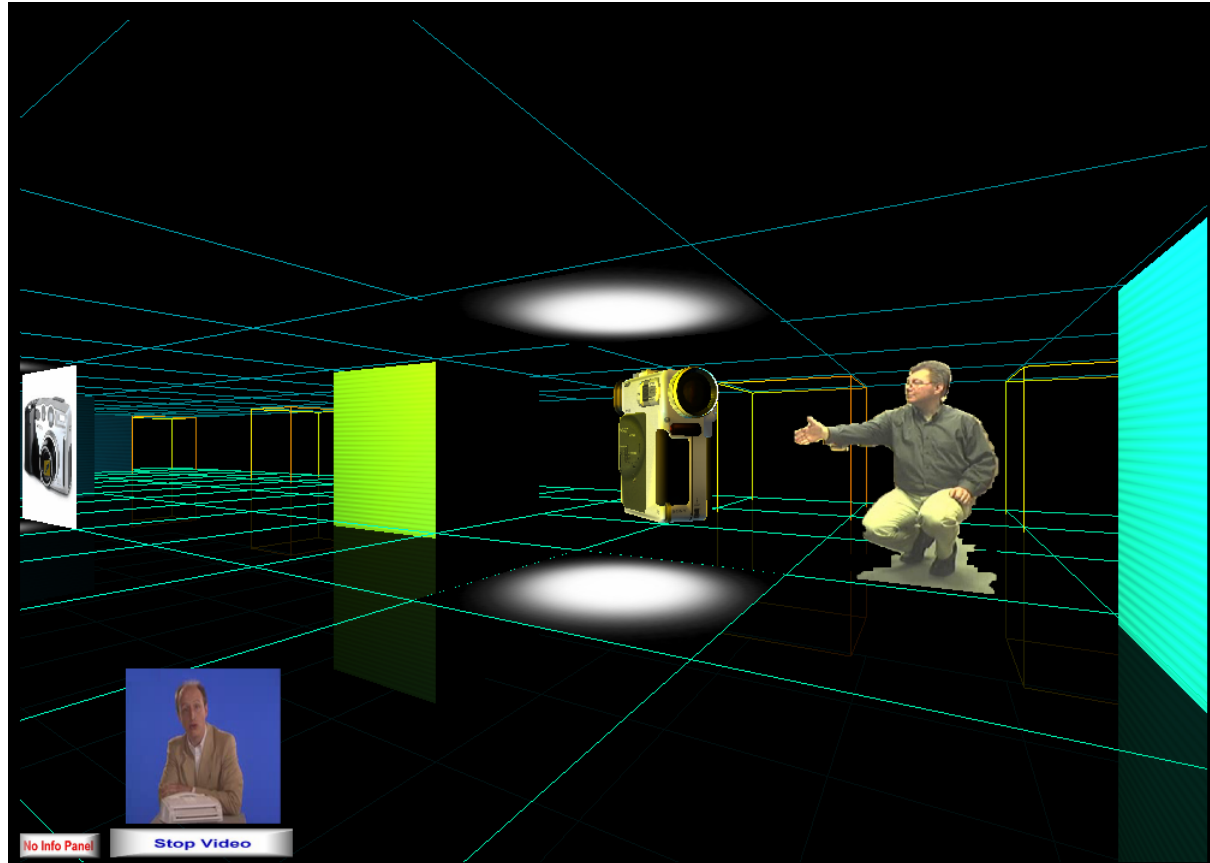
Elements of a Scene

- virtual 3D room
- virtual objects
- manipulable 3D objects
- A/V object
- arbitrarily shaped A/V objects
- hyperlinks

Functionality within a scene

- Navigating in a room
- Interacting with 3D objects
- retrieve and control videos
- live-streaming of arbitrarily shaped 2D video objects
- retrieving hyperlinks

Another Example: IST Project SoNG (2002/2003)



Communicating a MPEG-4 Scene



Transmitting the scene



Video streaming



Communications



3D object



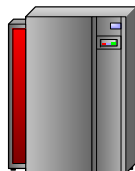
Authoring

„Playout“

Transmission

Composition

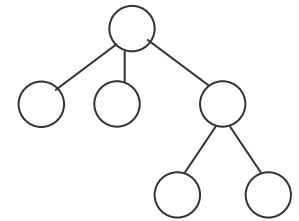
Rendering



Aspects for a Technical Implementation

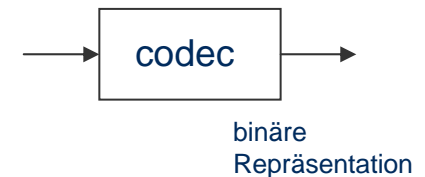
Description:

- 3D scene
- Objects of the scene (2D, 3D, A/V,)
- events (interactions, animations)
- arrangement of objects in space and time



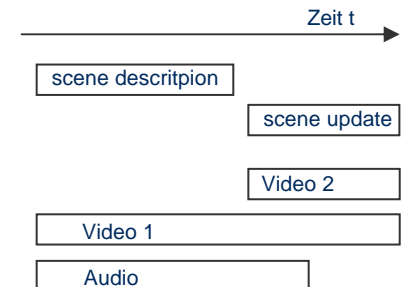
Coding:

- Description (e.g. 3D scene)
- natural media elements (audio, images, video)
- synthetic objects



Transmission:

- Multiplex of elementary stream
- synchronization
- Transport mechanisms (streaming, download)
- Methods for dynamic modification



The Standard ISO/IEC MPEG-4

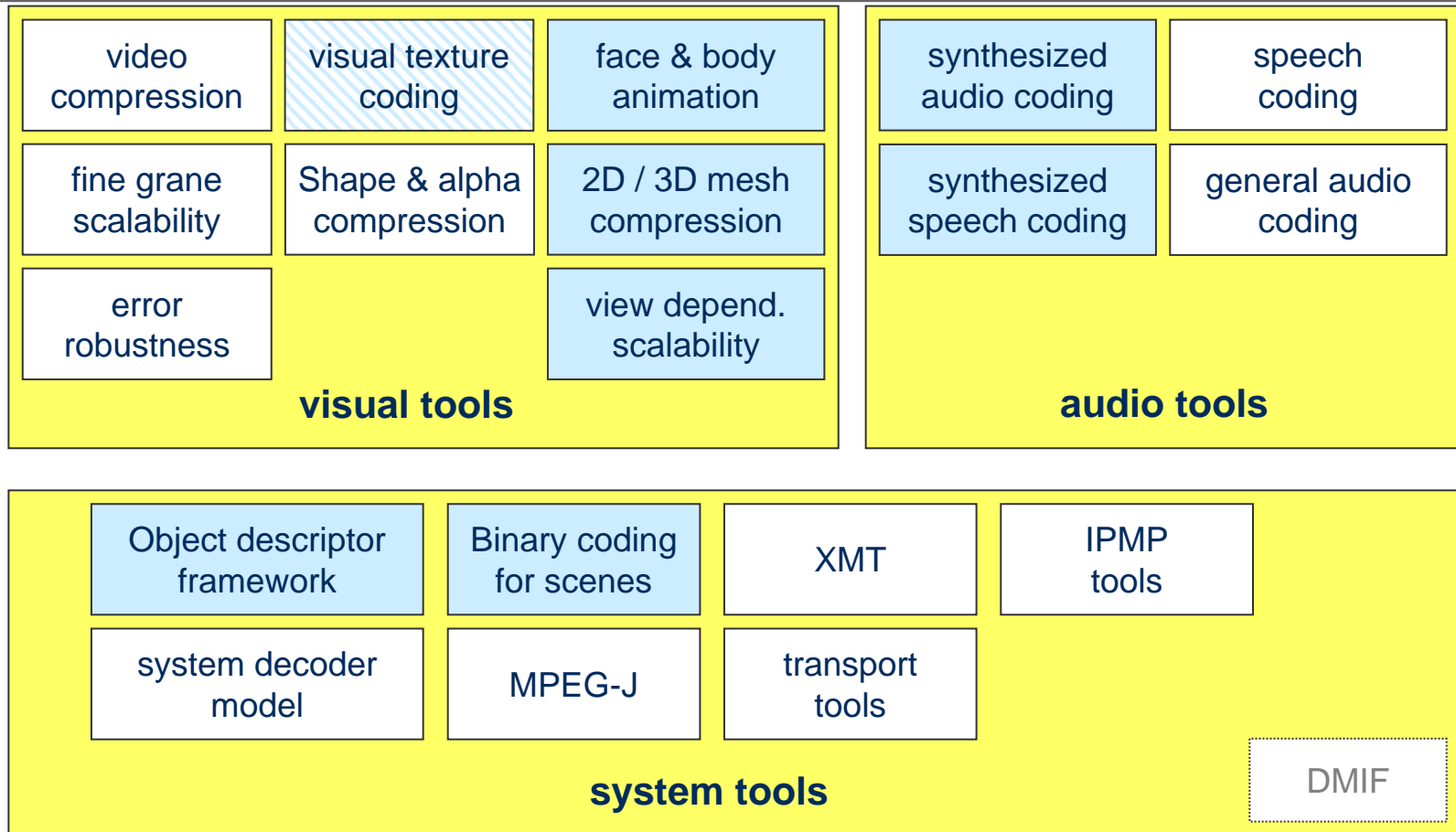
Idea:

- scene is a composition of various A/V objects
e.g. text, images, video, 2D/3D artificial or natural objects
- individual coding approaches for different object types
- opportunity to interact with the objects of the scene
- suitable mechanisms for the various transport and delivery requirements, according to the target application

SNHC – Synthetic and Natural Hybrid Coding

- describing synthetic and natural A/V objects
- temporal and spatial composition of the A/V objects
- coding the description
- coding the specific media components beyond audio and video
(text-to-speech, animation parameter, graphical elements,)
- defining specific visual and auditory profiles

Elements of MPEG-4 Standards



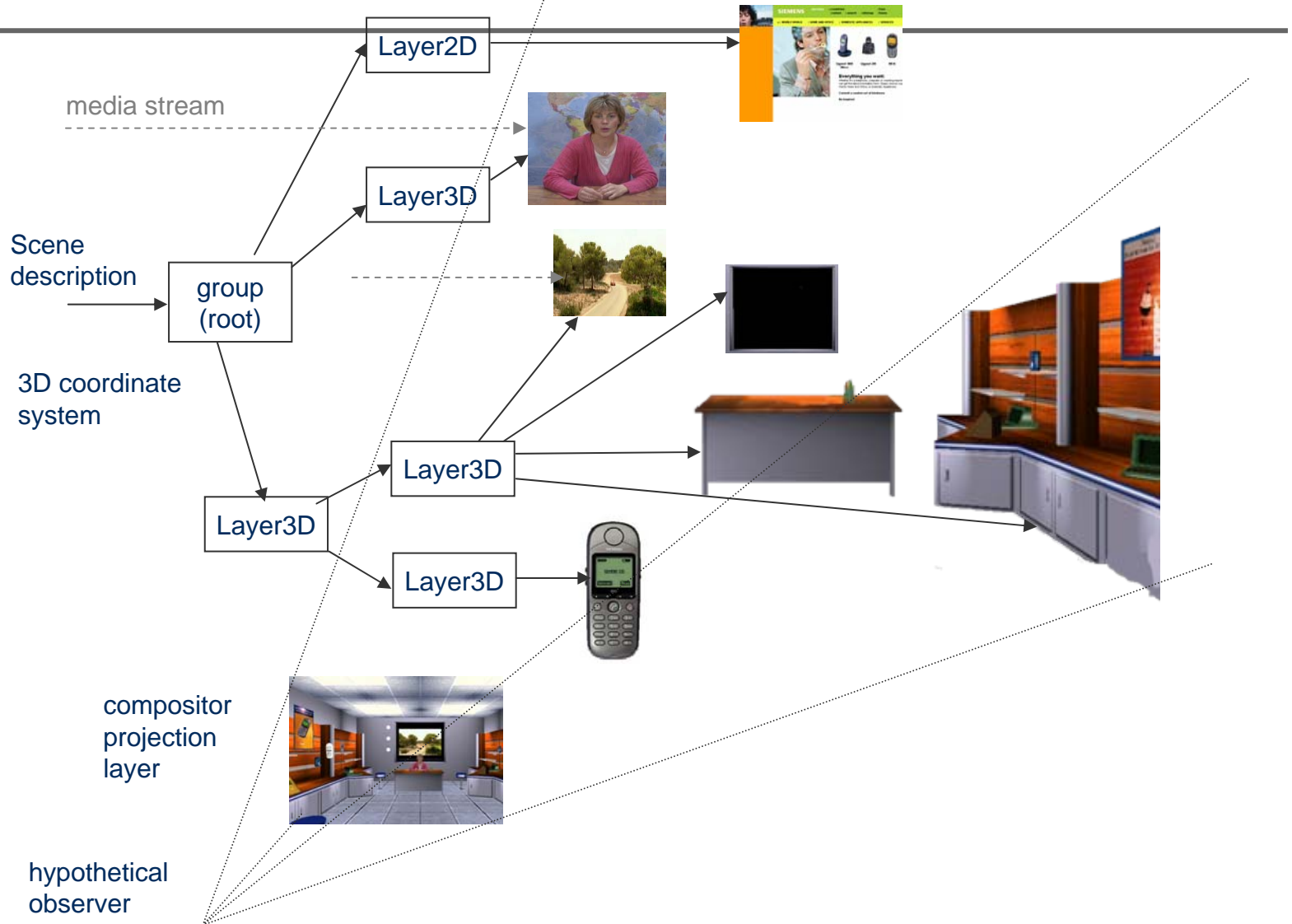
Legend: SNHC components and respectively contain relevant elements

Scene Description:

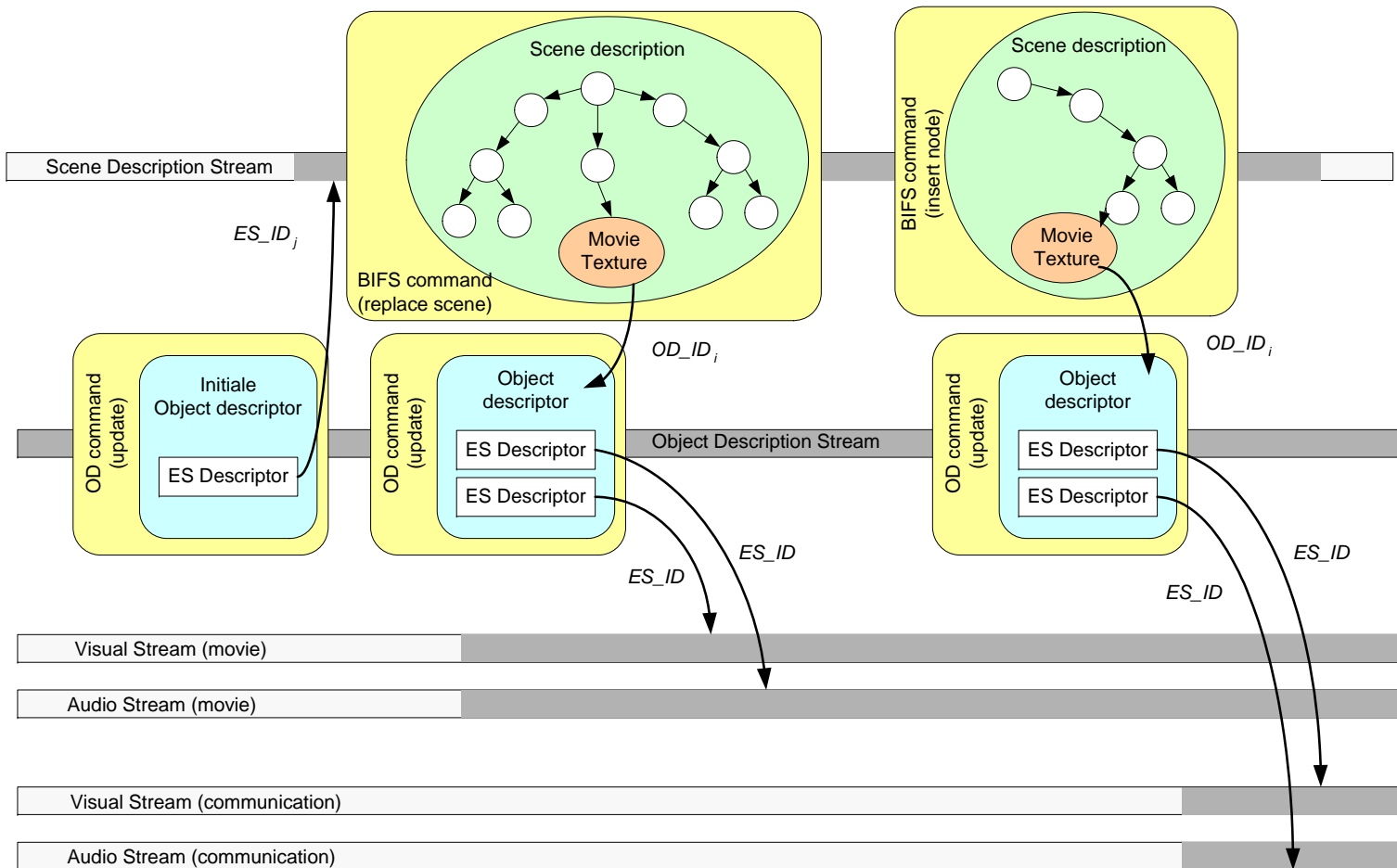
BIFS – Binary Format for Scenes

- **Extending VRML**
- **Embedding audio-visual object with extended parameters**
 - A/V Objects are leaves of a scene graph, e.g. 2D Graphics, Sound2D, AudioSource, . . .
 - Complementing visual tools / nodes, e.g. FBA, 2D/3D Mesh
 - Extended functionality, e.g. mixing audio sources, synthesis of audio
- **Combination of 2D and 3D Scenes**
 - e.g. Layer2D, Layer3D, CompositeTexture2D nodes
- **Concepts for interactivity**
 - Extend via `Script` and `ConditionalNode`
 - Animation „continuous interpolation“
- **Streaming and download of scenes**
 - Commands for synchronized scene updating → modifying of node parameters at a specific point of time (replace nodes, routes, ...)
 - `AnimationStream` nodes, for continuously stream scene animations
- **Binary representation (coding) of the description**

Example for a MPEG-4 Scene Description



Example: Object Descriptor



DMB reanimates BIFS

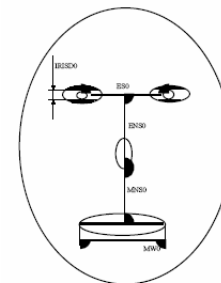
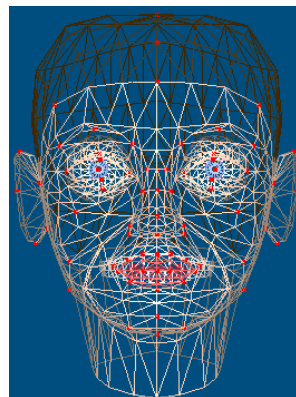
- DMB (Digital Multimedia Broadcast) used for mobileTV services
- Utilizes MPEG-4 BIFS for interactive applications
- Based on Core2D Profile and XML description



SNHC Tools „Face and Body Animation“: Face Animation

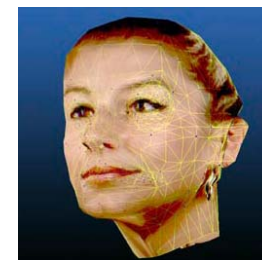
Description:

- 84 feature points describe the face anatomy (face in neutral position)
- 68 face animation parameters – FAP; moving feature points for reaching a specific expression; expressions divided in 10 categories
- FAP units allow “by normalization” to model the intensity of an animation independent of the face model (6 relations of „feature points“ to each other)



Face and head model

- Describing the face and head, e.g. as a wireframe model
- Mapping the feature points onto the face model
- Mapping the face texture
- Animation tables (animation rules) – several nodes (points) of the face are animated jointly for each FAP



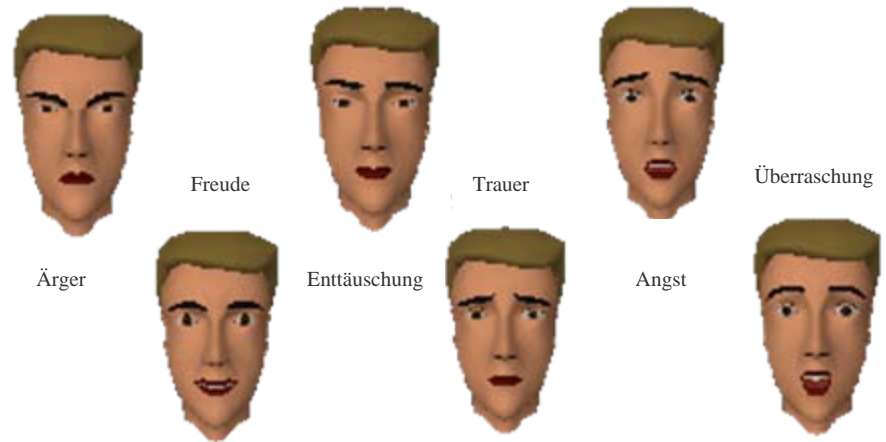
Extended Animation Options of the Face

Approach:

Animation is a composition of elementary FAPs

Expressing emotions

6 fundamental expressions



Animation for speech playback

14 viseme: visual analogon to phonemes

Text	Graphem	Phonem/Visem
A	A	a: / 10 (A:)
v	v	v / 2 (f,v)
a	a	a: / 10 (A:)
t	t	t: / 4 (d,t)
a	a	a: / 10 (A:)
r	r	r / 9 (r)

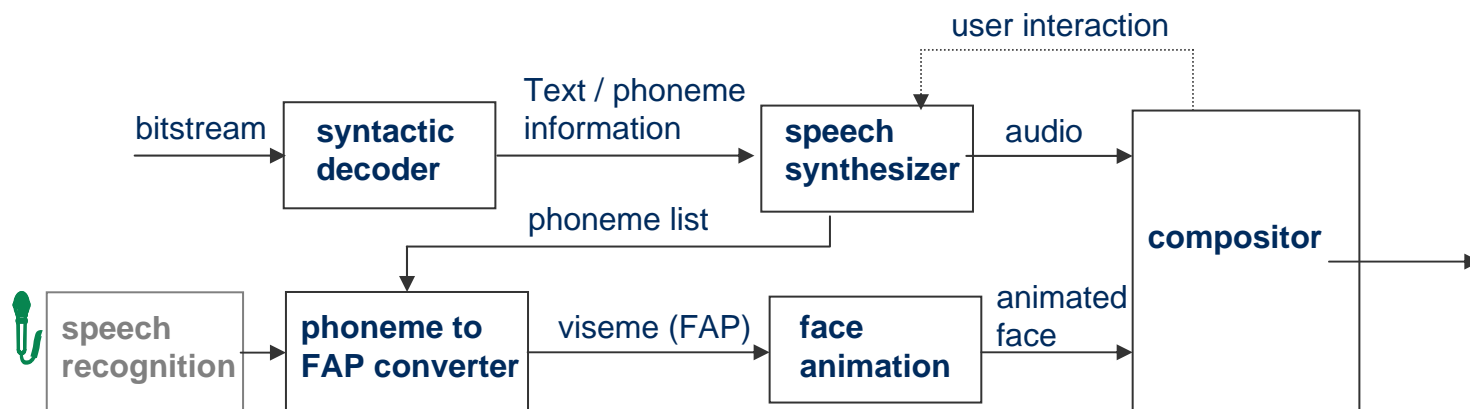
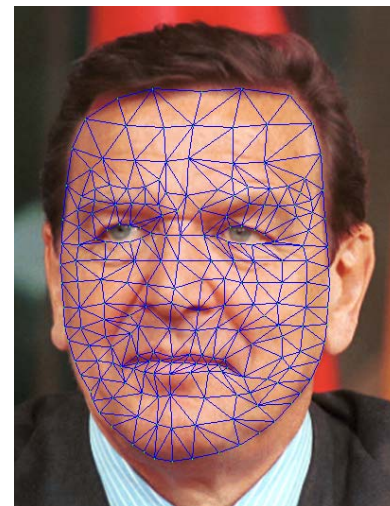
Synthesized Speech -- TTS

Principle:

- **Transmitting a sentence**

- **complemented with**

- Speaker related information
- Prosody
- „Lip shape“ information
- Speech code (ID)
- Emoticons (Smilie)
- The control parameters for animating the face



Summary

MPEG-4 SNHC

- defines an approach for a consistent representation and the temporal and spatial composition of synthetic objects and natural A/V objects in 2D and 3D scenes
- defines an efficient coding for the description
- defines coding approaches for media formats beyond audio and video (text-to-speech, animation parameter, graphical elements)
- provides functionality to stream and dynamically update scene description
- defines specific visual and auditive profiles

Applications

Virtual video conferences -- Avatars, e.g. for user interfaces (GUI) --
distant learning – e-Learning – e-commerce portals – simulations –
interactive multi-user environments (games) -- --