Slide 1

Overview of MPEG-7

Dr Junqing Yu

Digital Media Lab

Huazhong University of Sci & Tech
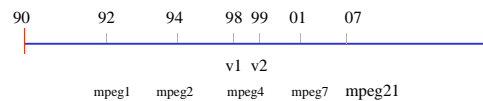
12/7/2008

Slide 2

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Digital Media Research Lab, HUST

2

Slide 3

Terms

MPEG-7 – Multimedia Content Description Interface

MPEG-7 Standard No. ISO/IEC 15938

Digital Media Research Lab, HUST

3

Slide 4

From MPEG-1 to MPEG-7

```
90        92        94      98 99   01      07
                                 v1 v2
        mpeg1    mpeg2      mpeg4    mpeg7  mpeg21
```

- MPEG-3, ever defined, but abandoned
- MPEG-5 and -6, not defined

Digital Media Research Lab, HUST

4

Slide 5

MPEG Family

MPEG-1 – Coding of moving pictures and audio for digital storage media (CD-ROM, MP3), 11/92

MPEG-2 – Generic Coding of moving pictures and audio information (DVD, Digital TV), 11/94

MPEG-4 – Coding of Audiovisual Objects for MM appls Ver1 09/98, Ver2 11/99

MPEG-7 – Multimedia content description for AV material 08/01

MPEG-21 – Digital AV framework: Integration of multimedia technologies, 2/07

Digital Media Research Lab, HUST

5

Slide 6

Why is MPEG-7 needed

- Digital audiovisual information increasing
  - more and more available contents
  - all kinds of sources of information
- Use of the digital audiovisual information
  - description of the contents
  - fast search of the contents

Digital Media Research Lab, HUST

6

Why do we need MPEG-7 ?

Support for Advanced Query

• *Fast & Accurate Access*

• *Personalized Content Production and Consumption*

N e e d +

• *Content Management*

• *Automation*

•**Visual**
•**Audio**
•**Sketch**

Objective of MPEG-7

• **Standardize content-based description for various types of audiovisual information**
  – Enable fast and efficient content searching, filtering and identification
  – Describe several aspects of the content (low-level features, structure, semantic, models, collections, creation, etc.)
  – Address a large range of applications

• **Types of audiovisual information**
  – Audio, speech
  – Moving video, still pictures, graphics, 3D models
  – Information on how objects are combined in scenes

Scope of MPEG-7

Description generation	Description	Description consumption
Research and future competition	Scope of MPEG-7	Research and future competition

• The description generation (feature extraction, indexing process, annotation & authoring tools,...) and consumption (search engine, filtering tool, retrieval process, browsing device, ...) are non normative parts of MPEG-7.

• The goal is to define the minimum that enables interoperability.

Scope of MPEG-7

standardization

Feature Extraction	MPEG-7 Description	Search Engine

Feature Extraction:
Content analysis (D, DS)
Feature extraction (D, DS)
Annotation tools (DS)
Authoring (DS)

MPEG-7 Scope:
Description Schemes (DSs)
Descriptors (Ds)
Language (DDL)
Ref: MPEG-7 Concepts

Search Engine:
Searching & filtering
Classification
Manipulation
Summarization Indexing

MPEG-7 Normative Interfaces

Access Unit
Textual Format

MPEG-7 Textual Encoder — MPEG-7 Textual Decoder

Content — Content Description

MPEG-7 Binary Encoder — MPEG-7 Binary Decoder

Access Unit
Binary Format

Abstract representation of possible applications using MPEG-7

Example: Content description

High level process — MPEG-7 Database — Low level process
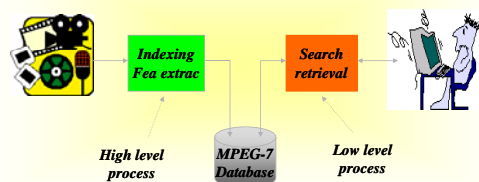
Parts of the MPEG-7 Standard

- **ISO / IEC 15938 - 1:** *Systems*
- **ISO / IEC 15938 - 2:** *Description Definition Language*
- **ISO / IEC 15938 - 3:** *Visual*
- **ISO / IEC 15938 - 4:** *Audio*
- **ISO / IEC 15938 - 5:** *Multimedia Description Schemes*
- **ISO / IEC 15938 - 6:** *Reference Software*
- **ISO / IEC 15938 - 7:** Conformance Testing
- **ISO / IEC 15938 - 8:** Extraction and use of descriptions
- **ISO / IEC 15938 - 9:** Profiles and levels
- **ISO / IEC 15938 - 10:** Schema Definition

MPEG-7 Systems

- Defines
  - the terminal architecture and the normative interfaces.
  - how descriptors and description schemes are stored, accessed and transmitted
  - tools that are needed to allow synchronization between content and descriptions

Reference Software: the XM

- XM implements
  - MPEG-7 Descriptors (Ds)
  - MPEG-7 Description Schemes (DSs)
  - Coding Schemes
  - DDL

MPEG-7 Conformance

- Includes the guidelines and procedures for testing conformance of MPEG-7 implementations

Outline of contents

- Introduction
- **Basic Components**
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Main elements of MPEG-7

- **Descriptors (D):** representations of features, that define the syntax and the semantics of each feature representation (low-level).

- **Description Schemes (DS):** that specify the structure and semantics of the relationships between their components, which may be both Ds and DSs (high-level).

- **A Description Definition Language (DDL):** based on XML Schema, to allow the creation of new DSs and Ds, and to allow the extension and modification of existing DSs

- **System tools:** to support multiplexing of descriptions, synchronization issues, transmission mechanisms, coded representations, management and protection of intellectual property

Digital Media Research Lab, HUST   19

Description Definition Language

- *Description Definition Language (DDL) is a language that define what description is valid, and allows the creation of new Description Schemes and Descriptors. It also allows the extension and modification of existing Description Schemes*
- *DDL is used to define a set of formal rules*
  - *ordering of the elements*
  - *occurrences of elements*
  - *.........*
- *XML + MPEG-7 extensions*

Digital Media Research Lab, HUST   20

XML Basics

- **XML- eXtensible Markup Language**
  - 本身不是标记语言
  - 用于创建标记语言的一套规则，是一种元语言

```
1: <? xml version="1.0"?>
2: <!--一个简单的XML文档-->
3: <message>
4: <to>Student</to>
5: <from>Teacher</from>
6: <subject>Introduction to XML</subject>
7: <body>Welcome to XML!</body>
8: </message>
```

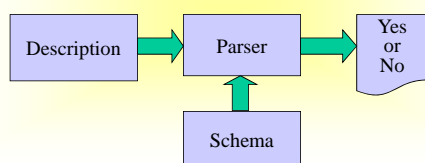*Digital Media Research Lab, HUST*   21

XML: Base for DDL

- **Why choose XML as the base for the DDL?**
  - The popularity of XML
  - The interoperability with other standards in the future

- **Why XML should be extended for MPEG-7?**
  - SGML > XML
  - Structural extensions
  - Datatype extensions

Digital Media Research Lab, HUST   22

DDL parser

DDL parser is a software to check if a description is valid

Description → Parser → Yes or No

Schema → Parser

Digital Media Research Lab, HUST   23

Integration of MPEG-7 into XML

```
<seq begin=20s dur=10s>
  <img id="Image1" dur=5s>
    <MP7: annotation>
     <Who>Fernado Morientes</Who>
     < WhatAction >Spain vs. Sweden soccer match
     </ WhatAction>
    </MP7: annotation>
  </img>
  <img id="Image2" dur=2s />
</seq>
```

Digital Media Research Lab, HUST   24

Descriptor

- **Definition**

  A Descriptor (D) is a representation of a Feature. A Descriptor defines the syntax and the semantics of the Feature representation.
- **Notes**

  A descriptor allows an evaluation of the corresponding feature via the descriptor value. It is possible to have several descriptors representing a single feature.
- **Examples**

  For example for the color feature, possible descriptors are: the color histogram, the average of the frequency components, the motion field, the text of the title, etc.

Descriptor Value

- **Definition**

  A Descriptor Value is an instantiation of a Descriptor for a given data set (or subset thereof).

- **Notes**

  Descriptor Values are combined via the mechanism of a Description Scheme to form a Description.

Descriptor Example

```
<VisualDescriptor xsi:type="DominantColorType">
    <SpatialCoherency>31</SpatialCoherency>
    <Value>
        <Percentage>31</Percentage>
        <Index>255 0 0</Index>
        <ColorVariance>0 0 0</ColorVariance>
    </Value>
</VisualDescriptor>
```

Description Scheme

- **Definition** A Description Scheme (DS) specifies the structure and semantics of the relationships between its components, which may be both Descriptors and Description Schemes.
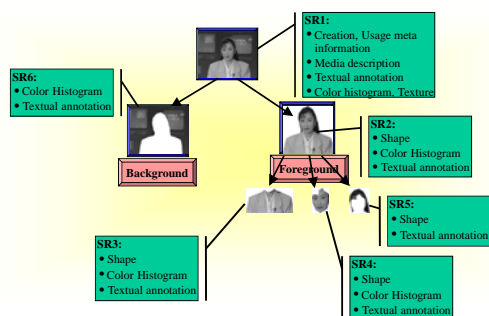- **Examples** A movie, structured as scenes and shots, including some textual descriptors at the scene level, and color, motion and some audio descriptors at the shot level.
- **Note**

  Ds contain only basic data types, and does not refer to others D or DSs.

Example

DS: XML Scheme & Extensions

- **XML Scheme**
  - Data types
  - Simple and Complex types
  - Elements
  - Inheritance, Abstract types

- **MPEG-7 extensions**
  - Array and Matrix datatype
  - Enumerated datatypes for MimeType, CountryCode, RegionCode, CurrencyCode and CharacterSetCode
  - Typed references

Description Scheme Example
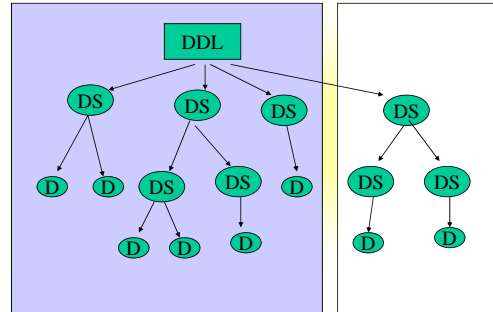
```
<DSType name="MovingRegion">
    <subDSOf type="Segment"/>
    <attributename="TemporalConnectivity"
    datatype="boolean">
    <attributename="SpatialConnectivity"
    datatype="boolean">
    <DSTypeRef type="Time"/>
    <DTypeRef type="ColorSpace" minOccurs="0"/>
    <DTypeRef type="ColorQuantization" minOccurs="0"/>
    <DTypeRef type="DominantColor" minOccurs="0"/>
    <DTypeRef type="GofGopColorHistogram"
    minOccurs="0"/>
    <DTypeRef type="MotionTrajectory" minOccurs="0"/>
    <DTypeRef type="ParametricMotion" minOccurs="0"/>
    <DTypeRef type="MotionActivity" minOccurs="0"/>
</DSType>
```

Digital Media Research Lab, HUST   31

Relations of main elements

Digital Media Research Lab, HUST   32

Illustration of descriptions

Digital Media Research Lab, HUST   33

Description Scheme Example

```
<VideoSegment id = "VS1" >
        <MediaTime>
                <MediaTimePoint> T0:0:0 </MediaTimePoint>
                <MediaDuration> PT10M </MediaDuration>
        </MediaTime>

        <StructuredAnnotation>
                This is an example of video segment
        </StructuredAnnotation>

        <GofGopColorHistogram HistogramTypeInfo = "Average">
        .....
        </GofGopColorHistogram>

        <SegmentDecomposition Gap = "true" Overlap = "true" DecompositionType = "temporal">
                <VideoSegment id = "VS2" > .... </VideoSegment>
                <VideoSegment id = "VS3" > .... </VideoSegment>
                <VideoSegment id = "VS4" > .... </VideoSegment>
        </SegmentDecomposition>
</VideoSegment>
```
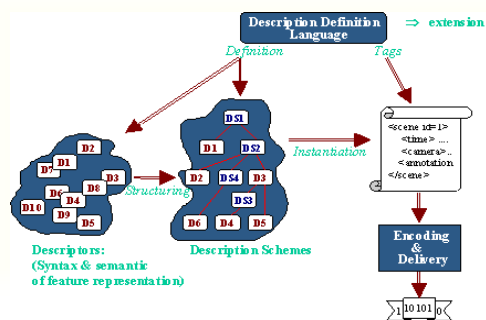
Digital Media Research Lab, HUST   34

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Digital Media Research Lab, HUST   35

Type of descriptions

- **Low level description** (features, etc)
  - *Generic and flexible*
  - *Intelligent / efficient search engine*

- **High level description** (structures, concepts,etc)
  - *Efficient and powerful*
  - *Lack of flexibility*

Digital Media Research Lab, HUST   36

Low-level Description

- **Information in the creation and production processes**
  - *director, title, short feature movie*
- **Information related to the usage of the content**
  - *copyright pointers, usage history, broadcast schedule*
- **Information on the storage features of the content**
  - *storage format, encoding*
- **Information about low-level features in the content**
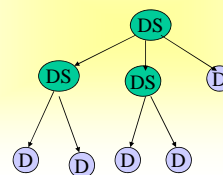 - *colors, textures, sound timbres, melody*

High-level Description

- **Structural description**
  - *video segments, frames, still and moving regions, audio segments*
  - *Segment DS (representing the spatial, temporal or spatio-temporal structure)*
- **Conceptual (semantic) description**
  - *objects, events, and notions*
  - *links of the two descriptions*
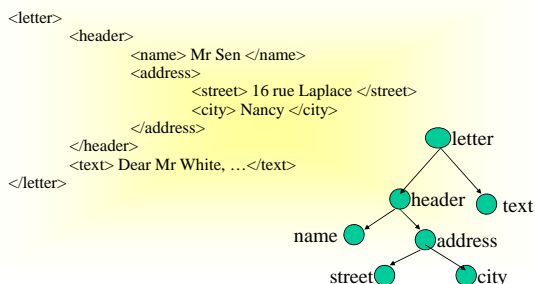

Basic description

- Elements
  - Information containers
  - containing data and other elements
  - <city> …… </city>
- Attributes
  - Attribute-value pairs used to characterize elements
  - <city population="10000"> …… </city>

Structured descriptions

- Structured descriptions are trees
- Trees are suitable for retrieval and search

Description trees

```
<letter>
     <header>
          <name> Mr Sen </name>
          <address>
               <street> 16 rue Laplace </street>
               <city> Nancy </city>
          </address>
     </header>
     <text> Dear Mr White, …</text>
</letter>
```

Example: Audio description

```
<Mpeg7Main>
     <DescriptionMetadata>
          <Version>1.0</Version>
     </DescriptionMetadata>
     <ContentDescription>
          <AudioContent xs1:type="AudioType">
               <Audio>
                    <CreationInformation>
                         <Creation>
                              <Title> The daily news </Title>
                         </Creation>
                    </CreationInformation>
               </Audio>
          </AudioContent>
     </ContentDescription>
</Mpeg7Main>
```
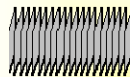
7

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Digital Media Research Lab, HUST 43

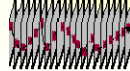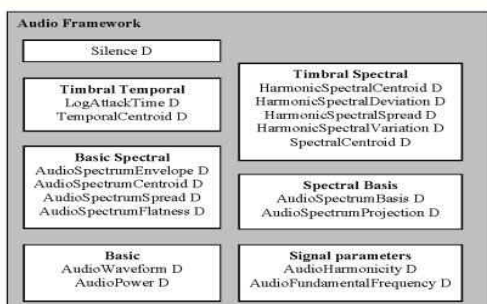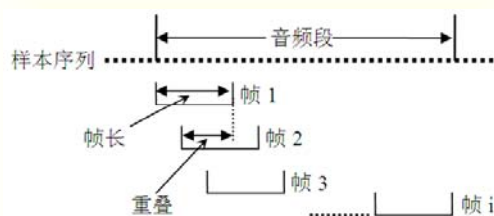## Low level AV descriptors

Video segments
- Color
- Camera motion
- Motion activity
- Mosaic

Still regions
- Color
- Shape
- Position
- Texture

Moving regions
- Color
- Motion trajectory
- Parametric motion
- Spatio-temporal shape

Audio segments
- Spoken content
- Spectral feature
- Timbre

Digital Media Research Lab, HUST 44

Audio Framework

Audio Framework

Silence D

Timbral Temporal
LogAttackTime D
TemporalCentroid D

Timbral Spectral
HarmonicSpectralCentroid D
HarmonicSpectralDeviation D
HarmonicSpectralSpread D
HarmonicSpectralVariation D
SpectralCentroid D

Basic Spectral
AudioSpectrumEnvelope D
AudioSpectrumCentroid D
AudioSpectrumSpread D
AudioSpectrumFlatness D

Spectral Basis
AudioSpectrumBasis D
AudioSpectrumProjection D

Basic
AudioWaveform D
AudioPower D

Signal parameters
AudioHarmonicity D
AudioFundamentalFrequency D
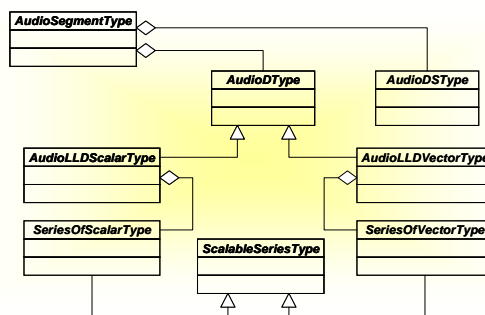
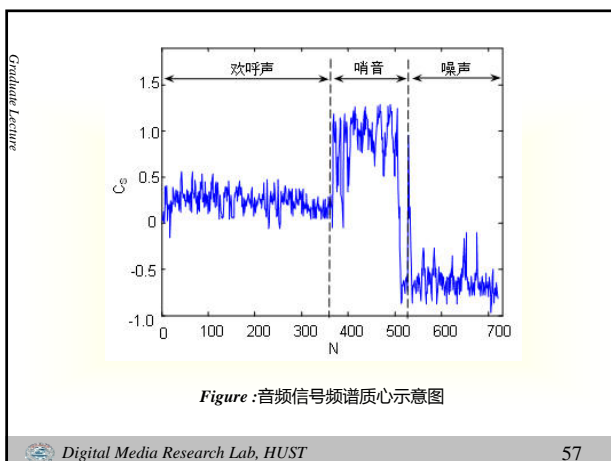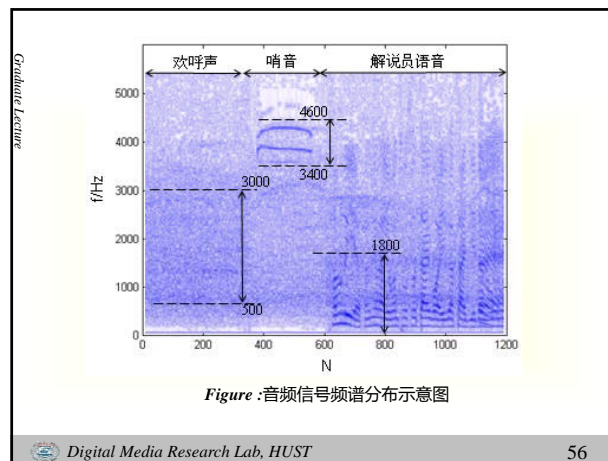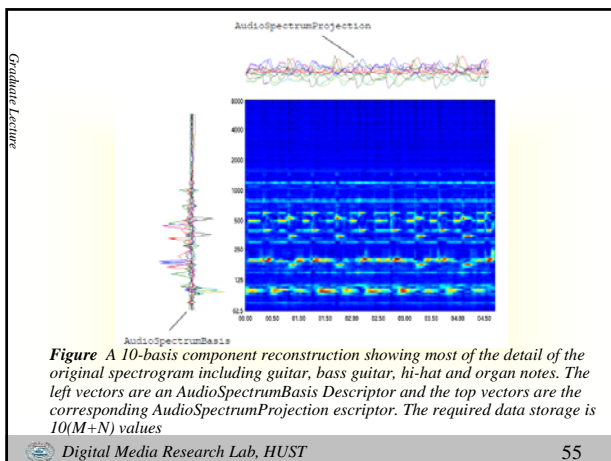*Digital Media Research Lab, HUST* 45

Audio description

- Low-level Description
  - spectrum, parametric, and temporal features
- High-level Description
  - Audio signature Description Scheme
  - Instrument timbre Description Schemes
  - The melody Description Tools
  - Sound recognition and indexing Description Tools
  - Spoken Content Description Tools

Digital Media Research Lab, HUST 46

Audio Frame & Segment

样本序列　音频段

帧长　帧 1
　　　帧 2
重叠　帧 3
　　　帧 i

Digital Media Research Lab, HUST 47

Structural types in Audio Framework

AudioSegmentType

AudioDType

AudioDSType

AudioLLDScalarType

AudioLLDVectorType

SeriesOfScalarType

ScalableSeriesType

SeriesOfVectorType

Digital Media Research Lab, HUST 48

An illustration of the scalable series

original series	●●●●●●●●●●●●●●●●●●●●●●●●●●●●●											
scaled series	O	O	O	O		O	OOO	O	O	O	O	O
k (index)	1	2	3	4	5		6 7 8	9	10	11	12	13
ratio	2			6			1	2				
numOfElements	3			2			2	6				
totalNumOfSamples	31											

Digital Media Research Lab, HUST   49

Audio descriptor: Basic

- Two basic audio Descriptors
  - AudioWaveform Descriptor
    - describes the audio waveform envelope (minimum and maximum)
  - AudioPower Descriptor
    - describes the temporally-smoothed instantaneous power

Digital Media Research Lab, HUST   50

AudioWaveform Descriptor

```
<!-- ################################################## -->
<!-- Definition of AudioWaveform D                      -->
<!-- ################################################## -->
<complexType name="AudioWaveformType">
  <complexContent>
    <extension base="mpeg7:AudioLLDScalarType">
      <attribute name="minRange" type="float" use="optional"/>
      <attribute name="maxRange" type="float" use="optional"/>
    </extension>
  </complexContent>
</complexType>
```

Digital Media Research Lab, HUST   51

AudioPower Descriptor

```
<!-- ##################################### -->
<!-- Definition of AudioPower D            -->
<!-- ##################################### -->
<complexType name="AudioPowerType">
  <complexContent>
    <extension base="mpeg7:AudioLLDScalarType"/>
  </complexContent>
</complexType>
```

Digital Media Research Lab, HUST   52

Audio descriptor: Basic Spectral

- AudioSpectrumEnvelope Descriptor
  - describes the short-term power spectrum
- AudioSpectrumCentroid Descriptor
  - describes the center of gravity of the log-frequency power spectrum
- AudioSpectrumSpread Descriptor
  - describing the second moment of the log-frequency power spectrum
- AudioSpectrumFlatness Descriptor
  - describes the flatness properties of the spectrum

Digital Media Research Lab, HUST   53

Figure : AudioSpectrumEnvelope description of a pop song. The required data storage is NMvalues where N is the number of spectrum bins and M is the number of time points

Digital Media Research Lab, HUST   54

9

Figure *A 10-basis component reconstruction showing most of the detail of the original spectrogram including guitar, bass guitar, hi-hat and organ notes. The left vectors are an AudioSpectrumBasis Descriptor and the top vectors are the corresponding AudioSpectrumProjection escriptor. The required data storage is 10(M+N) values*

*Figure :*音频信号频谱分布示意图

*Figure :*音频信号频谱质心示意图

Audio Signature Description

- AudioSignature Description Scheme provides a unique content identifier for the purpose of robust automatic identification of audio signals
- Applications include
  - audio fingerprinting
  - identification of audio
  - locating metadata for legacy audio content

Instrument Timbre Description

- Timbre is defined as the perceptual features that make two sounds having the same pitch and loudness sound different.
- Timbre Description describes the perceptual features with a reduced set of Descriptors
  - HarmonicInstrumentTimbre Descriptor
  - LogAttackTime Descriptor
  - PercussiveIinstrumentTimbre Descriptor
  - Combination with *Basic Spectral Descriptors*

Melody Description Tools

The melody Description Tools is to facilitate efficient, robust, and expressive melodic similarity matching

- MelodyContour Description Scheme
  - 5-step contour representation
  - basic rhythmic information representation
- MelodySequence Description Scheme
  - supporting an expanded descriptor set and high precision of interval encoding

Visual description

- Color Descriptors
- Texture Descriptors
- Shape Descriptors
- Motion Descriptors for Video

Basic Structures

- Grid layout
- Time series
  - RegularTimeSeries
  - IrregularTimeSeries
- Multiple view
- Spatial 2D coordinates
- Temporal interpolation.

Color Descriptors
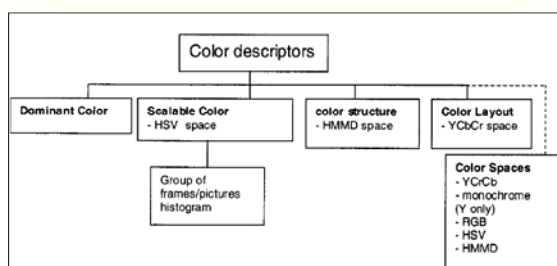
Scalable Color Descriptor

- A color histogram in HSV color space
- Encoded by Haar Transform

Dominant Color Descriptor

- Clustering colors into a small number of representative colors
- It can be defined for each object, regions, or the whole image
- $F = \{ \{c_i, p_i, v_i\}, s \}$
  - c_i : Representative colors
  - p_i : Their percentages in the region
  - v_i : Color variances
  - s : Spatial coherency

Dominant Color Descriptor
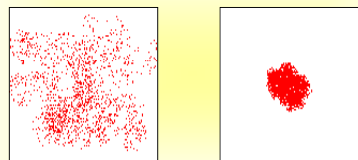
Color Layout Descriptor

- Clustering the image into 64 (8x8) blocks
- Deriving the average color of each block (or using DCD)
- Applying DCT and encoding
- Efficient for
  - Sketch-based image retrieval
  - Content Filtering using image indexing

Color Structure Descriptor

- Scanning the image by an 8x8 pixel block
- Counting the number of blocks containing each color
- Generating a color histogram (HMMD)
- Main usages:
  - Still image retrieval
  - Natural images retrieval

GoF/GoP Color Descriptor

- Extends Scalable Color Descriptor
- Generates the color histogram for a video segment or a group of pictures
- Calculation methods:
  - Average
  - Median
  - Intersection

Texture Descriptors

- Homogenous Texture Descriptor
- Non-Homogenous Texture Descriptor (Edge Histogram)

Homogenous Texture Descriptor

- Partitioning the frequency domain into 30 channels (modeled by a 2D-Gabor function)
- Computing the energy and energy deviation for each channel
- Computing mean and standard variation of frequency coefficients
- $F = \{f_{DC}, f_{SD}, e_1,\ldots, e_{30}, d_1,\ldots, d_{30}\}$
- An efficient implementation:
  - Radon transform followed by Fourier transform

2D-Gabor Function

- It is a Gaussian weighted sinusoid
- It is used to model individual channels
- Each channel filters a specific type of texture
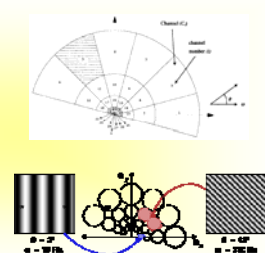
Radon Transform

- Transforms images with lines into a domain of possible line parameters
- Each line will be transformed to a peak point in the resulted image
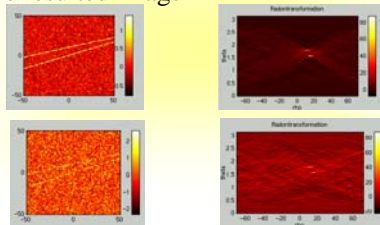
Non-Homogenous Texture Descriptor

- Represents the spatial distribution of five types of edges
  - vertical, horizontal, 45°, 135°, and non-directional
- Dividing the image into 16 (4x4) blocks
- Generating a 5-bin histogram for each block
- It is scale invariant

Non-Homogenous Texture Descriptor
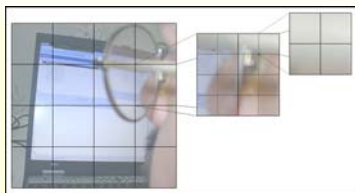
Shape Descriptors

- Region-based Descriptor
- Contour-based Shape Descriptor
- 2D/3D Shape Descriptor
- 3D Shape Descriptor

Region-based Descriptor

- Expresses pixel distribution within a 2-D object region
- Employs a complex 2D-Angular Radial Transformation (ART)
- Advantages:
  - Describes complex shapes with disconnected regions
  - Robust to segmentation noise
  - Small size
  - Fast extraction and matching

Region-based Descriptor (2)
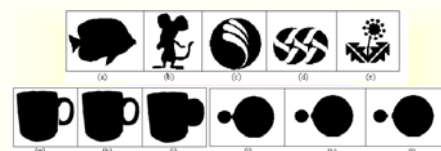
- Applicable to figures (a) – (e)
- Distinguishes (i) from (g) and (h)
- (j), (k), and (l) are similar

Contour-Based Descriptor

- It is based on Curvature Scale-Space representation

Curvature Scale-Space

- Finds curvature zero crossing points of the shape's contour (key points)
- Reduces the number of key points step by step, by applying Gaussian smoothing
- The position of key points are expressed relative to the length of the contour curve
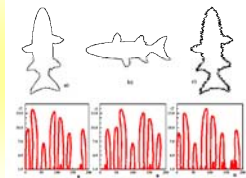
Curvature Scale Space (2)
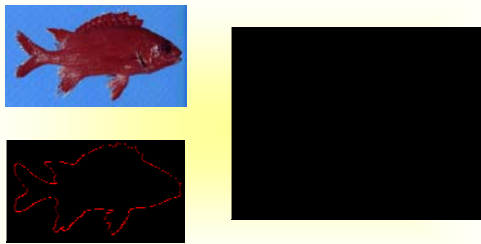
Contour-Based Descriptor

- It is based on Curvature Scale-Space representation
- Advantages:
  - Captures the shape very well
  - Robust to the noise, scale, and orientation
  - It is fast and compact

Contour-Based Descriptor (2)
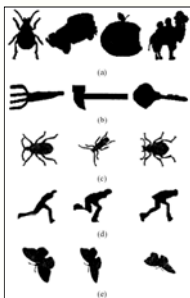
- Applicable to (a)
- Distinguishes differences in (b)
- Find similarities in (c) - (e)

Comparison
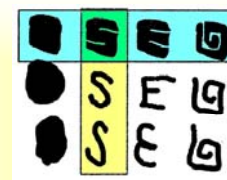
- Blue: Similar shapes by Region-Based
- Yellow: Similar shapes by Contour-Based

14

2D/3D Shape Descriptor

- A 3D object can be roughly described by snapshots from different angles
- Describes a 3D object by a number of 2D shape descriptors
- Similarity Matching: matching multiple pairs of 2D views

3D Shape Descriptor

- Based on Shape spectrum
- An extension of Shape Index (A local measure of 3D Shape to 3D meshes)
- Captures information about local convexity
- Computes the histogram of the shape index over the whole 3D surface

Motion Descriptors

- Motion Activity Descriptors
- Camera Motion Descriptors
- Motion Trajectory Descriptors
- Parametric Motion Descriptors

Motion Activity Descriptor

- Captures '*intensity of action*' or '*pace of action*'
- Based on standard deviation of motion vector magnitudes
- Quantized into a 3-bit integer [1, 5]

Camera Motion Descriptor

- Describes the movement of a camera or a virtual view point
- Supports 7 camera operations
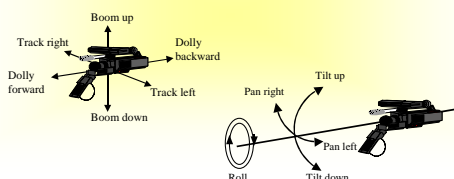
Motion Trajectory

- Describes the movement of one representative point of a specific region
- A set of key-points (x, y, z, t)
- A set of interpolation functions describing the path

15
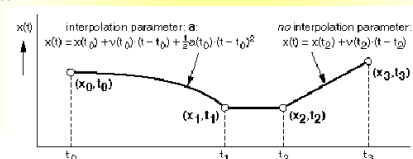
Parametric Motion

- Characterizes the evolution of regions over time
- Uses 2D geometric transforms
- Example:
  - Rotation/Scaling:
    - $D_x(\mathbf{x,y}) = a + b\mathbf{x} + c\mathbf{y}$
    - $D_y(\mathbf{x,y}) = d - c\mathbf{x} + b\mathbf{y}$

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- **Multimedia Description Schemes**
- XM and Applications
- More Information

Multimedia DSs

Multimedia Description Schemes are metadata structures for describing and annotating audio-visual (AV) content

- Basic Elements
- Content Management
- Content Description
- Content Organization
- Navigation and Access
- User Interaction

Organization of Multimedia DSs
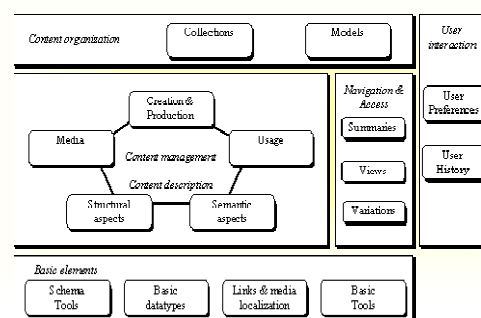
Basic Element

- Schema tools
- Basic datatypes
- Links & media localization
- Basic tools

Basic elements of DS

- Basic data types
  - a set of extended data types
  - vectors and matrices
- Constructs for linking media files
- Localizing pieces of content
- Describing
  - time, places, persons, individuals, groups, organizations, and textual annotation, etc
  - Who? What object? What action? Where? When? Why? and How?

Schema tools

- Base types
- Root Element
- Top-level types
- Multimedia Content Entities
- Packages
- Description Metadata

Base Types
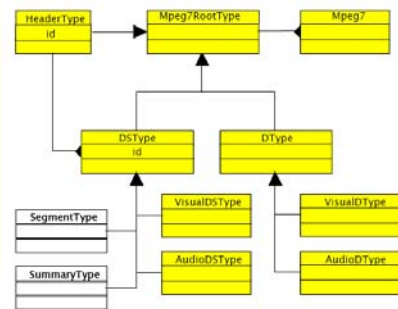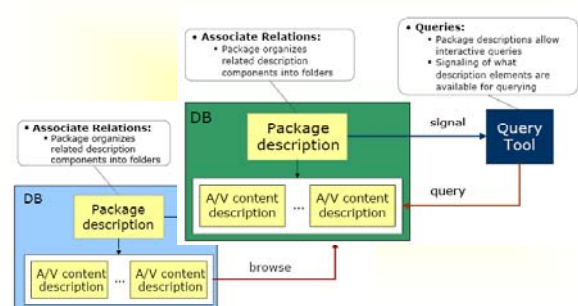

Root Element
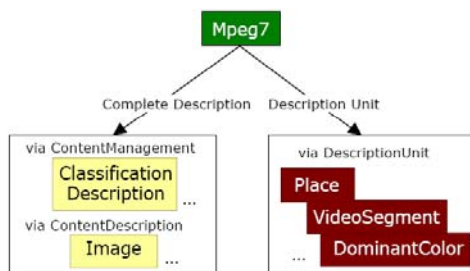
Top-level types
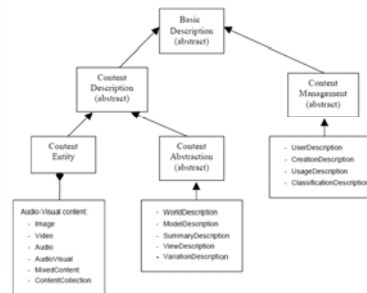
Multimedia Content Entities

- Image Type
- Video Type
- Audio Type
- AudioVisual Type
- Multimedia Type
- MultimediaCollection type
- MultimediaProgramType
- Signal Type
- ElectronicInkType
- VideoEditiing Type

Packages

Description Metadata

```
<VideoSegment>
    <Header xsi:type="DescriptionMetadataType">
        <Creator>
            <Role href="creatorCS"><Name>Creator</Name></Role>
            <Agent xsi:type="OrganizationType">
                <Name> MDS Group </Name>
            </Agent>
        </Creator>
        <CreationTime>2001-01-13T21:30</CreationTime>
    </Header>
```

Digital Media Research Lab, HUST — 103

Basic datatypes

- defines datatypes that represent different kinds of constrained types.
  - Integer
  - Real
  - Matrix
  - String
  - countryCode

Digital Media Research Lab, HUST — 104

Links & media localization

- References datatype
  - refer to a part of the description
- Unique Identifier
  - allows the identification of the multimedia or other media content under description.
- Time description tools
  - YYYY-MM-DDThh:mm:ss:nnnFNNN+hh:mm
- Media localization tools

Digital Media Research Lab, HUST — 105

Time description tools

Digital Media Research Lab, HUST — 106
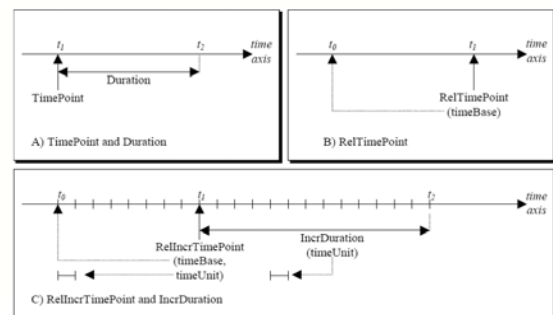
Content Management

- Creation and production information
  - Creation information
    - title, textual annotation, creators, and dates
  - Classification information
    - genre, subject, purpose, language
- Media coding, storage and file formats
  - format, compression, and coding
- Content usage
  - usage rights, usage record

Digital Media Research Lab, HUST — 107

Model for content, profile and instance

Digital Media Research Lab, HUST — 108
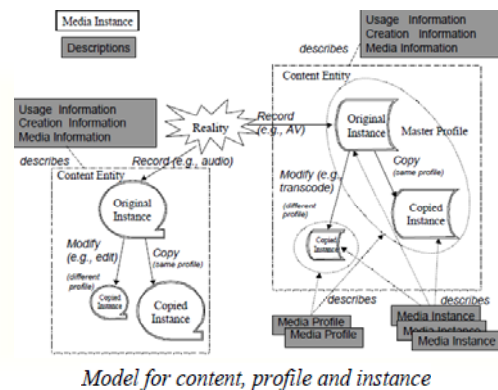
Content Description
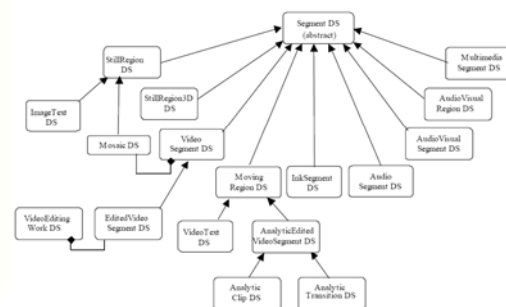
- Structural aspects
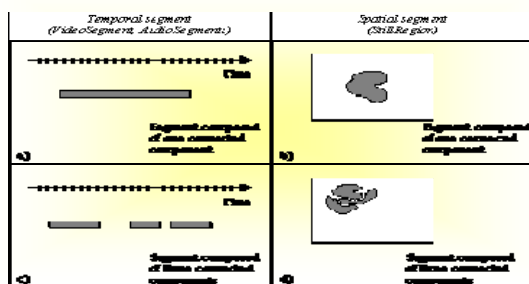- Semantics aspects

Structural aspects
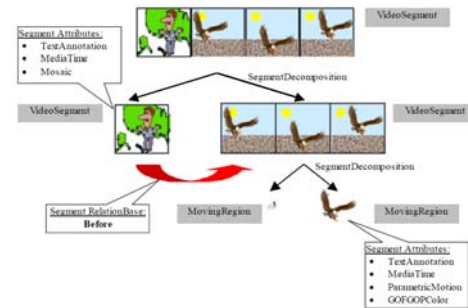
Structural aspects

- Segment entity description tools
- Segment attribute description tools
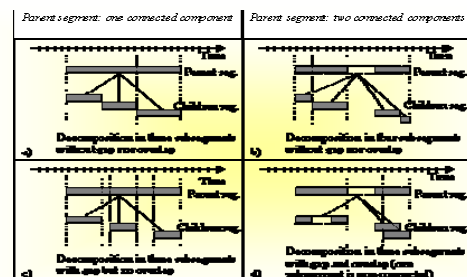- Segment decomposition tools
- Segment relation description tools

Segment entity description tools

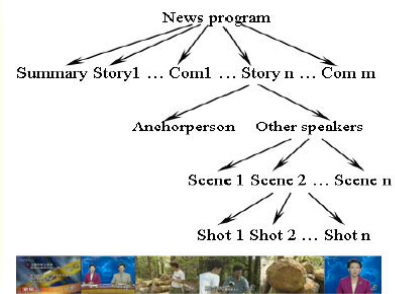# Examples: T/S segments

Segment decomposition tools

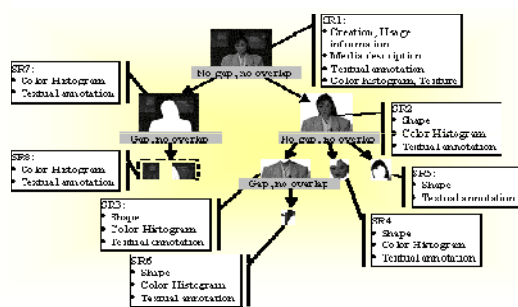Examples of Segment Decomposition: a) and b) Segment Decompositions without gap nor overlap; c) and d) Segment Decompositions with gap or overlap.

Segment relation description tools
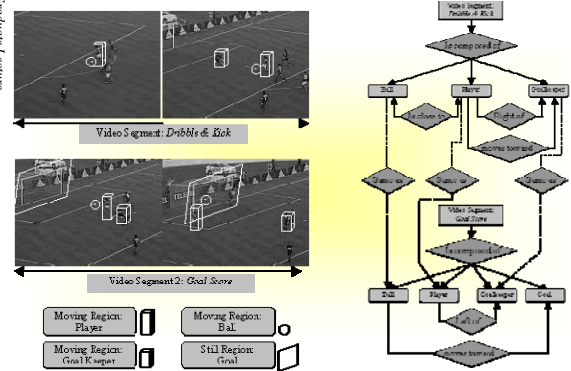
- Hierarchical Segment Tree
- Graph

Digital Media Research Lab, HUST   115

Example: Segment trees

Digital Media Research Lab, HUST   116

Example: Segment trees

Digital Media Research Lab, HUST   117

Example: Graph

Digital Media Research Lab, HUST   118

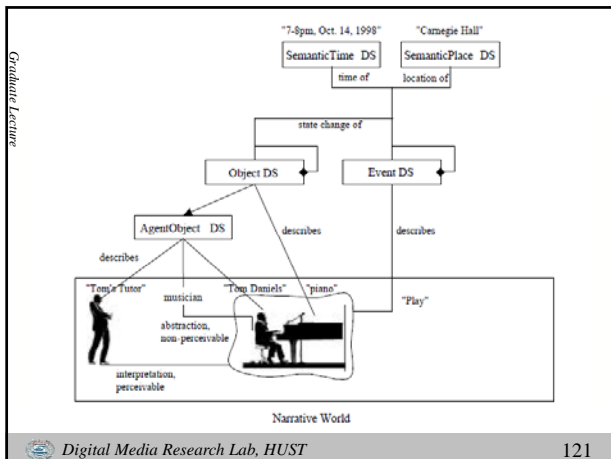## Semantic aspects

- Semantic Entity
- Semantic Attribute
- Semantic Relation

Digital Media Research Lab, HUST   119

Semantic Entity

Digital Media Research Lab, HUST   120

20

"7-8pm, Oct. 14, 1998"   "Carnegie Hall"

| SemanticTime DS | SemanticPlace DS |

time of   location of

state change of

| Object DS | | Event DS |

AgentObject DS   describes   describes

describes

"Tom's Tutor"   musician   "Tom Daniels"   "piano"   "Play"

abstraction, non-perceivable

interpretation, perceivable
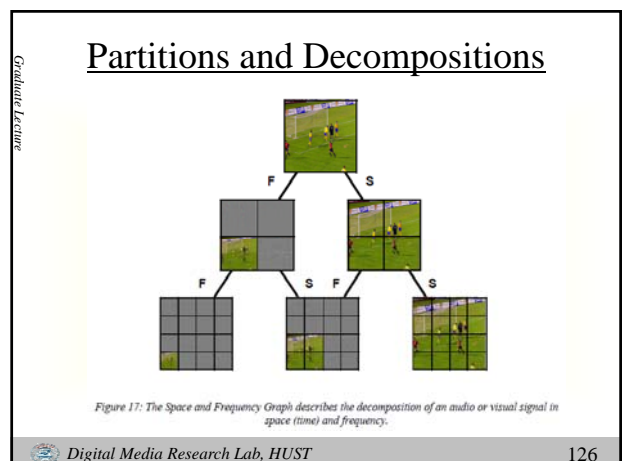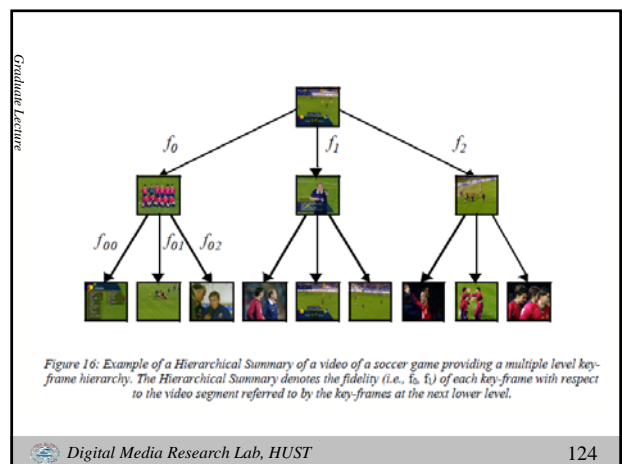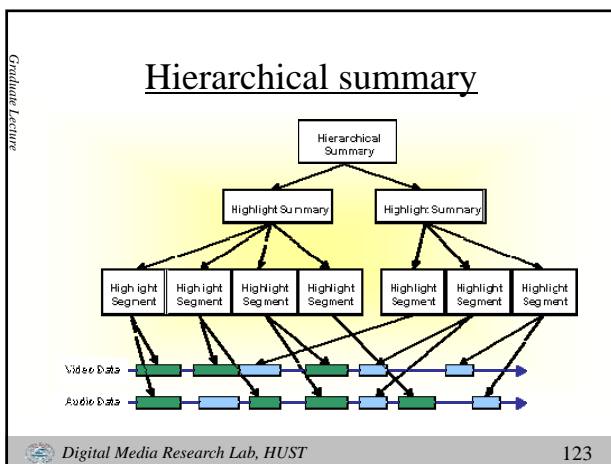
Narrative World

Digital Media Research Lab, HUST   121
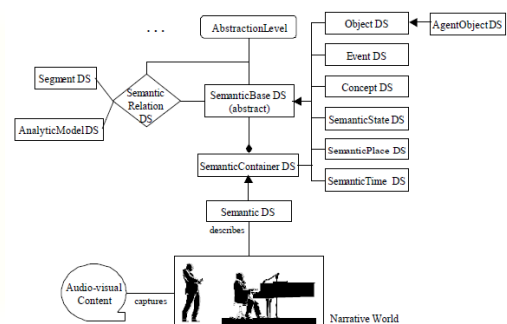
Navigation and Access

- Summaries
  - *hierarchical summaries*
  - *sequential summaries*
- View, Partitions and Decompositions
  - *decompositions in space, time and frequency*
  - *used in multi-resolution access and progressive retrieval*
- Variations
  - *selection of the most suitable of an AV program*
  - *adapt to the different capabilities of terminal devices, network conditions or user preferences*

Digital Media Research Lab, HUST   122

Hierarchical summary

| Hierarchical Summary |

| Highlight Summary | | Highlight Summary |

| High ight Segment | Highlight Segment | Highlight Segment | Highlight Segment | Highlight Segment | Highlight Segment | Highlight Segment |

Video Data

Audio Data

Digital Media Research Lab, HUST   123

---

f_0 f_1 f_2

f_{00} f_{01} f_{02}

Figure 16: Example of a Hierarchical Summary of a video of a soccer game providing a multiple level key-frame hierarchy. The Hierarchical Summary denotes the fidelity (i.e., f_0, f_1) of each key-frame with respect to the video segment referred to by the key-frames at the next lower level.

Digital Media Research Lab, HUST   124

Sequential summary

| Frame Property | | Frame Property | | Frame Property |

| Sequential Summary | | Text Property | | Text Property |

| Sound Property | | Sound Property | | Sound Property |

Visual Data

Textual Data

Audio Data

Digital Media Research Lab, HUST   125

Partitions and Decompositions

F   S

F   S   F   S

Figure 17: The Space and Frequency Graph describes the decomposition of an audio or visual signal in space (time) and frequency.

Digital Media Research Lab, HUST   126
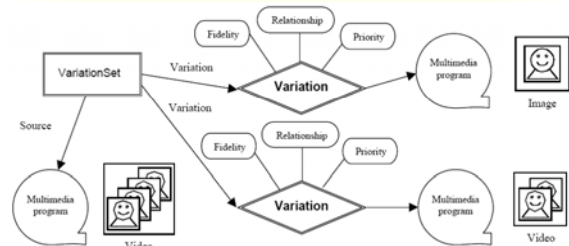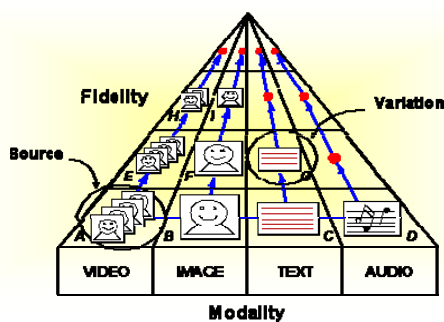
Views

Illustration of variations

## Illustration of variations

Content Organization

- Collections
  - *group the contents into clusters*
  - *describes statistics and models of the attribute values*
  - *describe relationships among collection clusters*
- Models
  - *model the attributes and features of AV content*
  - *Probability Model*
    - *specify statistical functions and structures*
  - *Analytic Model*
    - *specify semantic labels*
    - *specify the confidence*
    - *build classifiers*

Collection Structure
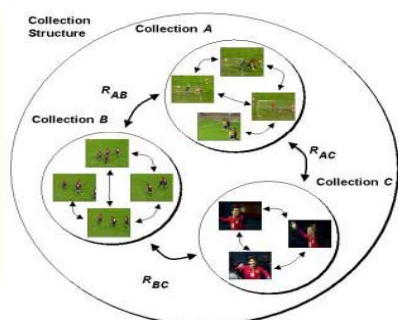
User Interaction

- User Preference
  - context dependency in terms of time and place
  - relative importance of different preferences
  - privacy characteristics of the preferences
  - preferences update by agent or user
- Usage History
  - history of actions
  - used to determine the user's preferences
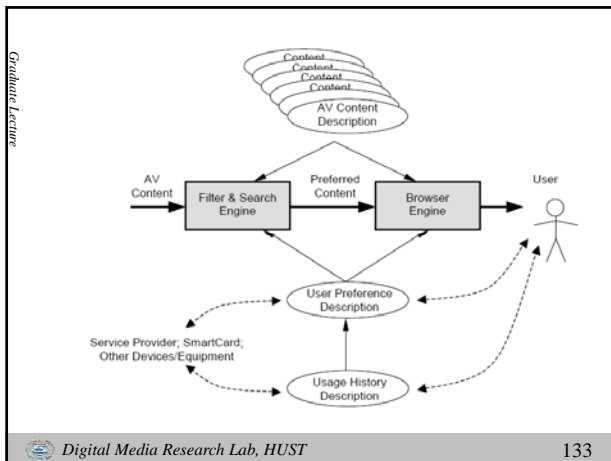
Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- **XM and Applications**
- More Information

eXperimentation Model(XM)

- **Simulation platform for:**
  - Ds, DSs, CSs, DDL
- **XM applications:**
  - the server (extraction) applications
  - the client (search, filtering and/or transcoding) applications
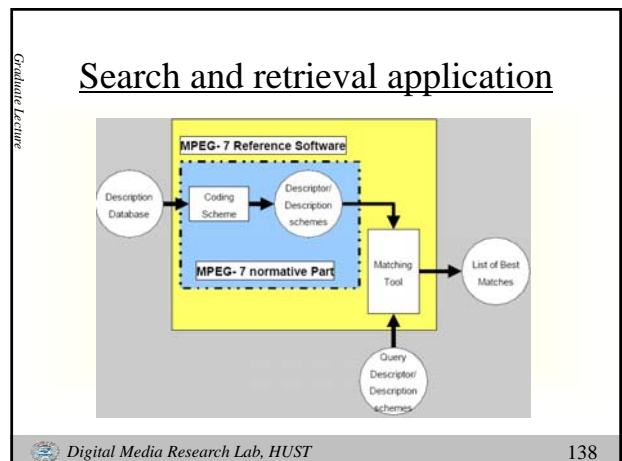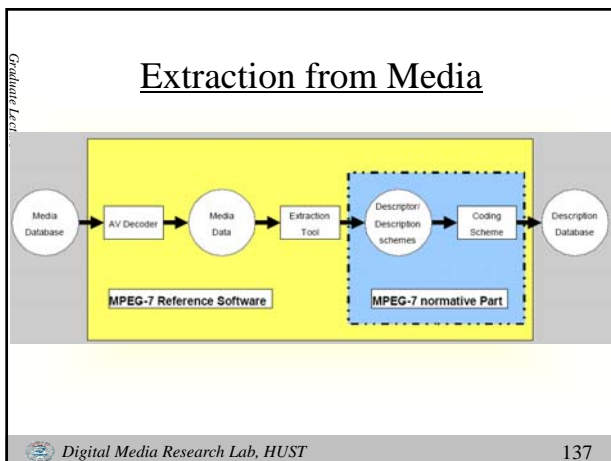
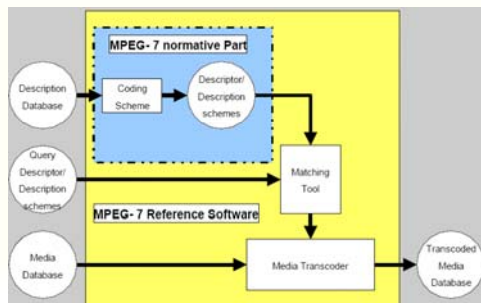*CS: Coding Schemes*

The XM applications

- Extraction from Media
  - all low-level Ds or DSs should have an application class of this type
- Search & Retrieval Application
- Media Transcoding Application
- Description Filtering Application

Extraction from Media

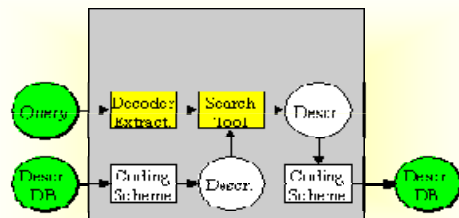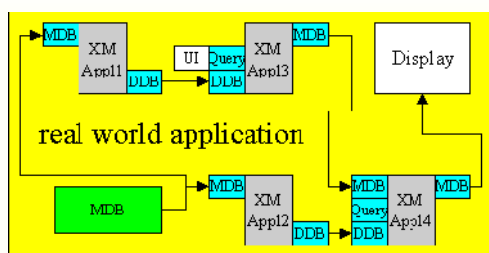## Search and retrieval application

Media transcoding application

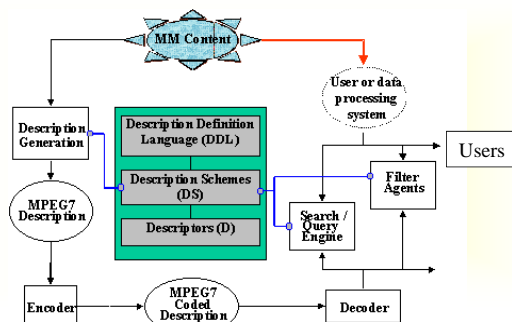## Description Filtering Application

Real world application

MDB = media database, DDB = description database.
First, from a media database two features are extracted. Then, basing on the first feature,
relevant media files are selected from the media database.
The relevant media files are transcoded basing on the second extracted feature.

MPEG-7 application areas

- Storage and retrieval of audiovisual databases (image, film, radio archives)
- Broadcast media selection (radio, TV programs)
- Surveillance (traffic control, surface transportation, production chains)
- E-commerce and Tele-shopping (searching for clothes / patterns)
- Remote sensing (cartography, ecology, natural resources management)
- Entertainment (searching for a game, for a karaoke)
- Cultural services (museums, art galleries)
- Journalism (searching for events, persons)
- Personalized news service on Internet (push media filtering)
- Intelligent multimedia presentations
- Educational applications nBio-medical applications

Illustration of applications

Information Flow
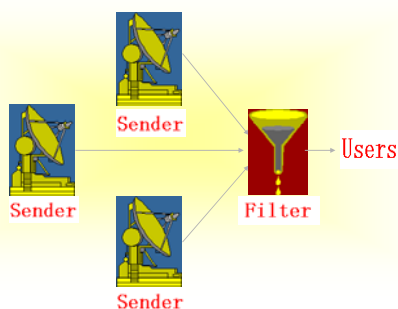
24
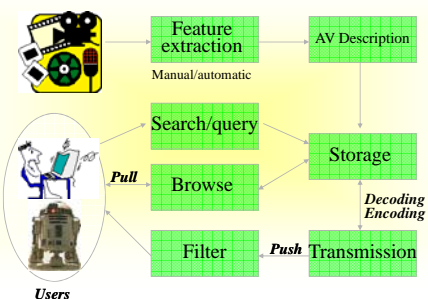
Push and Pull applications

- Push applications
  - Example: Search engines for internet and DBs
  - Advantage: Many search engines work on standardized descriptions
- Pull applications
  - Example: Broadcast of video, Interactive TV
  - Advantage: Intelligent agents filter standardized descriptions

Digital Media Research Lab, HUST   145

Example: Pull application

Digital Media Research Lab, HUST   146
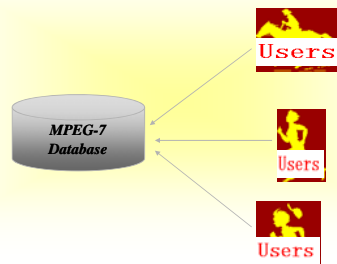
Example: Push application

Digital Media Research Lab, HUST   147

Example: queries

- Text (keywords):
  - Find AV material with subject corresponding to some keywords
- Semantic description:
  - Find AV material corresponding to a specified semantic
- Image as an example:
  - Find an image with similar characteristics (global or local)
- A few notes of music:
  - Find corresponding musical pieces or movies
- Low level features (example: motion):
  - Find video with specific object motion trajectories

Digital Media Research Lab, HUST   148

Outline of contents

- Introduction
- Basic Components
- Content Description
- Audiovisual (AV) Descriptions
- Multimedia Description Schemes
- XM and Applications
- More Information

Digital Media Research Lab, HUST   149

MPEG-7 and other Standards

- MPEG-1, -2, and -4 are designed to represent the information itself, while MPEG-7 is meant to represent information about the information.
- MPEG-1, -2, and -4 make content available, while MPEG-7 allows you to find the content you need.

Digital Media Research Lab, HUST   150

Ultimate ambition of MPEG-7

Graduate Lecture

- To make the web as searchable for multimedia content as it is searchable for text today

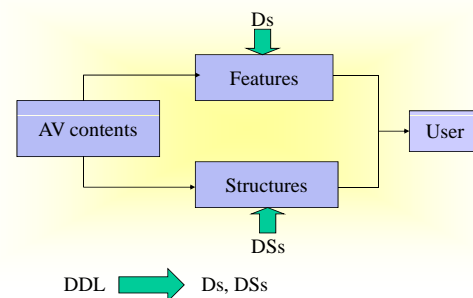- To improve the use of computer systems as easy as possible

Digital Media Research Lab, HUST   151

MPEG-7 beyond

Graduate Lecture

- To mould computers around human requirements and not humans around computer requirements
- To enable content disclosure based on facts, rather than on human annotations
- To find information by rich spoken queries, hand-drawn images and address what most people expect computers to be able to do

Digital Media Research Lab, HUST   152

More Information on WWW

Graduate Lecture

- **http://www.chiariglione.org/mpeg/**
- **http://www.mpegif.org/**
-

Digital Media Research Lab, HUST   153

Conclusion

Graduate Lecture

Ds

Features

AV contents

User

Structures

DSs

DDL → Ds, DSs

Digital Media Research Lab, HUST   154

Conclusions on MPEG-7

Graduate Lecture

- **MPEG-7:**
  - AV content description for interoperable application
- **Description Definition Language:**
  - XML Schema (flexibility) + Binary version (efficiency)
- **Description Schemes:**
  - Library of description tools
  - Covers a wide range of generic needs

Digital Media Research Lab, HUST   155

Graduate Lecture

Thanks

Digital Media Research Lab, HUST   156