


# Content-based Retrieval from Multimedia Databases



Web Age Information Management  
(WAIM-2001)  
Xi'an, CHINA  
July 9th, 2001




John R. Smith  
IBM T. J. Watson Research  
Center  
30 Saw Mill River Road  
Hawthorne, NY 10532 USA  
jrsmith@watson.ibm.com

## Multimedia Databases

- Environment:
  - Growing amounts of non-structured digital image, video, audio, and multimedia data
  - Need for multimedia content management systems (storage, access, querying, and retrieval of media objects with metadata)
  - Applications: multimedia search engines, set-top box filtering, universal access (pervasive, mobile)
- Problems:
  - Difficulty in effectively searching and filtering multimedia
  - Lack of consistent, standards-based annotation and multimedia content description framework
- Emerging Solutions:
  - Content-based retrieval (CBR) and similarity search
  - MPEG-7 multimedia content description standard
  - Integrated multimedia searching (attributes, descriptors, XML) in database systems
  - Open benchmark for multimedia retrieval systems

## Evolution of Multimedia Technologies

Signals → Features → Semantics → Knowledge

Recent past



Near future

MPEG-1,-2,-4 Video storage Broadband Streaming video delivery	MPEG-4,-7 Multimedia databases Content-based retrieval Multimedia filtering Content adaptation	MPEG-7 Semantic-based retrieval and filtering Intelligent media services Interactive TV (ITV)	MPEG-21 Multimedia framework e-Commerce
Compression Coding Communications	Similarity searching Object- and feature-based coding	Modeling and classification Personalization and summarization	Media mining Decision support IFMP

## Outline

- Applications:
  - Multimedia Database Requirements
  - Multimedia Content Models
- Standards:
  - MPEG-7 Multimedia Content Description
  - MPEG-21 Multimedia Framework
- Problems and Technologies:
  - Multimedia Search Problems
  - Multimedia Knowledge Representation
  - Multimedia Content Adaptation
- Benchmarking:
  - TREC Video Retrieval
- References and Links

# Multimedia Applications and Requirements

## Multimedia Database Applications & Requirements

- **Multimedia Content Management:**
  - Examples: Digital Libraries, Media Asset Management
  - Requirements: query and retrieval; browsing and navigation; storage – archival, preservation; versioning, editing, workflow
- **Web Content Management:**
  - Requirements: Authoring tools – creation and production; interactive content, presentation; searching
- **Interactive Personalised Content Delivery:**
  - Examples: Interactive TV (digital)
  - Requirements: Personalization, filtering, summarization
- **Standards-based Representation:**
  - Essence – MPEG-1, -2, 4; JPEG, JPEG-2000
  - Metadata – MPEG-7, TV Anytime, SMPTE
- **Multimedia Content Modeling:**
  - High level semantics – objects, events, people, places
  - Low-level features – perceptual attributes

### Challenges in Multimedia Database Systems

- Multimedia applications require storage, access, retrieval, searching, filtering, editing, repurposing, ...
- Traditional multimedia objects:
  - Text, Image, Audio, Digital Video, Graphics, Web
- Object-based multimedia (recent):
  - Objects, Regions, Segments, Scenes, 3D Models, MPEG-4
- Many file formats:
  - GIF, JPEG, MPEG, WAV, MIDI, MP3, PDF, PS, BMP, ...
- Presentation, synchronization, quality of service (QoS), distributed multimedia
- Content-based retrieval (CBR), similarity searching, multi-dimensional indexing

### Multimedia Content Modeling Approaches

- **Content-based models:**
  - Audio-visual perceptual features (color, texture, shape, motion, ...)
- **Image data models:**
  - Image objects (regions), spatial relationships, visual features (color, texture, shape)
- **Video data models:**
  - Video objects (scenes, shots, segments, frames), temporal relationships, temporal features (motion)
- **Semantic models:**
  - Real world scene description (objects, events, people, Places), knowledge bases

### Example Multimedia Content Models

- **Four-layer model** (Gupta, et al, '91)
  - Picture description model: image layer, image object layer, semantic object layer and semantic event layer (**Structure / semantics**)
- **PDL** (Leung, et al, '92)
  - Picture Description Language (PDL) based on an entity-attribute-relationship model (**Relationships**)
- **OVID** (Oomoto, et al, '93)
  - Video-object system allows arbitrary attribute structures and attribute-value inheritance based on temporal interval inclusion relationships (**Attributes**)
- **EMIR** (Lahlou, '95)
  - Extended Model for Information Retrieval (EMIR) which models objects, relationships and concept categories comprised of descriptions, compositions and topologies (**Spatio-temporal**)
- **MPEG-7 Conceptual Model** (MPEG-7)
  - Identification and modeling of 192 principal content-description concepts from multimedia domain (**MPEG-7 Multimedia Content Description Standard**)

### Image and Video Specific Content Models

- **Image content models:**
  - Describe image regions / objects and spatial relationships
  - Examples: symbolic images (maps), color photographs (color regions)
  - Related work: 2-D strings (S. K. Chang, et al),  $\theta$ -R (Gudivada, et al), SaFe – spatial and feature (Smith and Chang, '99)
  - Spatial relationship based indexing and matching
- **Video content models:**
  - Scenes, Shots, frames, key-frames
  - Shot detection: segmentation into temporal units (shots)
  - Key-frame selection: selection of salient frame(s) from each shot
  - Scene analysis: feature extraction from shots
  - Multi-modal analysis: joint analysis of audio/video

### Advanced Video Content Models

- **Segmentation Trees:**
  - Video Table of Contents – hierarchical segmentation (segments and sub-segments)
- **Summaries:**
  - Key-frame summaries – audio and slideshow
  - Fast playback – adaptive sampling based on salient features
  - Hierarchical summaries
- **Visualizations:**
  - Mosaics – stitching of multiple frames, extraction of moving object regions
- **Temporal Models:**
  - Scene-transition graphs – clustering of recurring shots (scenes), transition frequencies

### Content-based Retrieval (CBR)

- Similarity Search based on visual- and audio-features
- **Content analysis and feature extraction:**
  - Automatic extraction of descriptors of low-level features – colour, texture, shape, ...
- **Content-based Queries:**
  - Result in ranked lists based on similarity score
  - Similarity computation requires distance metric (domain dependent, subjective)
- **Traditional databases:** correctness of matching, optimization of query efficiency
- **Content-based databases:** varying precision vs. recall, high-dimensional feature spaces, multi-dimensional indexing, query pre-filtering

**Content-Based Retrieval (CBR) Systems**

- QBIC (IBM, '92 - '97) - Query by Image Content
  - Images - color, texture, shape, spatial
  - Query by Example, relevance feedback)
- VIMSYS - Virage Search Engine
  - DB Objects at 4 levels
  - Image Representations, Image Objects, Domain Objects, Domain Events
- Photobook (MIT, '95)
- VisualSEEK and WebSEEK (Columbia, '96)
- SPIRE (IBM, '99)

**Example: IBM QBIC: Query by Color**

The screenshot shows the IBM QBIC web interface. At the top, there's a search bar with the word "color" entered. Below the search bar, there's a grid of 12 image thumbnails representing search results. The interface includes navigation buttons like "Home", "Search", and "Help".

**Example: IBM QBIC: Query by Texture**

The screenshot shows the IBM QBIC web interface with the word "texture" entered in the search bar. The grid of image thumbnails displays various textures like wood, stone, and fabric. The interface layout is consistent with the previous example.

**Example: Content-based Query by Video Events, Objects, Scenes (I.e., goal scores)**

Two screenshots of the IBM QBIC web interface. The left one shows a search for "Hammock" with a grid of video thumbnails. The right one shows a search for "Basketball" with a grid of video thumbnails. Both show various scenes and objects related to the search terms.

**MPEG-7 Multimedia Content Description Standard**

**MPEG-7: XML for Multimedia Content Description**

- **MPEG-7 Normative elements:**
  - Descriptors and Description Schemes
  - DDL for defining Description Schemes
  - Extensible for application domains

The diagram illustrates the MPEG-7 standard structure. It shows a central "MPEG-7" box connected to various components: "DDL" (Descriptor Definition Language), "Application domain", and "Descriptors". Below this, there's a diagram showing "Video Segments" and "Keyframes" with "Motion Regions" and "Spatial-Temporal Locators".

- **Example MPEG-7 Descriptions:**
  - Video segments (shots, keyframes, text and semantics)
  - Moving regions (spatio-temporal locators)
  - Audiovisual features, object tracking, scene description

### ISO MPEG-7 Standard (2001)

- **What is MPEG-7 about?**
  - Specification of a "Multimedia Content Description Interface"
  - Developed by *International Standards Organization (ISO)* and *International Electrotechnical Commission (IEC)*
  - Standardized representation of multimedia metadata in XML (*XML Schema Language*)
  - Describes audio-visual content at a number of levels (*features, structure, semantics, models, collections, immutable metadata*)
  - Designed for fast and efficient searching and filtering of multimedia content
  - MPEG-7 is to be released by ISO as Intl. Standard in Sept. 2001

### MPEG-7 Key Points

1. MPEG-7 is not a video coding standard
2. MPEG-7 is a metadata standard for describing multimedia content
3. MPEG-7 defines an industry standard schema (*Descriptors and Description Schemes*) using XML Schema Language (~450 simple and complex types)
4. MPEG-7 produces XML descriptions
5. MPEG-7 also provides a binary compression system for MPEG-7 descriptions (called *BIM*)
6. MPEG-7 descriptions can be embedded in the video streams or stored separately in files or databases

### ISO MPEG-7 Multimedia Content Description Standard (XML)

### Overview of MPEG Standards

### MPEG-7 Roadmap

- "Multimedia Content Description Interface"
- Start of competitive tests 02/1999
- Final Committee Draft (FCD) 03/2001
- Final Draft Int'l Standard (FDIS) 07/2001
- International Standard (IS) 09/2001
- MPEG-7 standard has the following parts:
  - MPEG-7 Systems
  - MPEG-7 Description Definition Language (DDL)
  - MPEG-7 Visual (Visual Descriptors)
  - MPEG-7 Audio (Audio Descriptors)
  - MPEG-7 Multimedia Description Schemes (MDS)
  - MPEG-7 Reference Software

### MPEG-7 Application Types

- Pull Applications (Search and Browsing)
  - Internet search engines and multimedia databases
  - **Advantages:** queries based on standardized descriptions, interoperability
- Push Applications (Filtering)
  - Broadcast video and interactive television
  - **Advantages:** intelligent software agents filter content / channels based on standardized descriptions
- Universal Multimedia Access (Perceptual QoS)
- Specialized Professional and Control Applications, i.e., digital video recording

### MPEG-7 Application Domains

- Education (e.g., distance learning)
- Journalism (e.g. searching for speeches by voice or face)
- Cultural services (history museums, art galleries, etc.)
- Entertainment (e.g. searching a game, karaoke)
- Investigation services (human characteristics recognition, forensics)
- Geographical information systems (GIS)
- Remote sensing (cartography, ecology, natural resources management, etc.)
- Surveillance (traffic control, surface transportation)
- Bio-medical applications
- E-commerce and shopping (e.g. searching for clothes/patterns)
- Architecture, real estate, and interior design
- Social (e.g. dating services)
- Film, video and radio archives

### Example: MPEG-7 Description (simple)

■ MPEG-7 Video description (MPEG-7 XML) – Structured Annotation:

```

<Mpeg7 type="complete">
  <ContentDescription xsi:type="ContentEntityType">
    <MultimediaContent xsi:type="VideoType">
      <TextAnnotation>
        <StructuredAnnotation>
          <Who><Name> Sammy Sosa </Name></Who>
          <WhatObject><Name> Baseball </Name></WhatObject>
          <WhatAction><Name> Homerun </Name></WhatAction>
          <Where><Name> Chicago </Name></Where>
        </StructuredAnnotation>
      </TextAnnotation>
    </MultimediaContent>
  </ContentDescription>
</Mpeg7>
    
```

### Example: MPEG-7 Description (complex)

■ MPEG-7 ClusterModel description (MPEG-7 XML):

```

<ClusterModel confidence="0.75" reliability="0.5">
  <Label><Name> Nature scenes </Name></Label>
  <Collection xsi:type="DescriptorCollectionType">
    <Descriptor xsi:type="ScalableColorType">
      <Coefficients dim="16"> 1 2 . . . 16 </Coefficients>
    </Descriptor>
  </Collection>
  <DescriptorModel>
    <Descriptor xsi:type="ScalableColorType">
      <Coefficients dim="16"> 1 2 . . . 16 </Coefficients>
    </Descriptor>
    <Field> ./Descriptor/Coefficients </Field>
  </DescriptorModel>
  <ProbabilityModel xsi:type="ProbabilityDistributionType" confidence="1.0" dim="16">
    <Mean dim="16"> 0.5 0.5 . . . 0.5 </Mean>
    <Variance dim="16"> 0.25 0.75 . . . 0.25 </Variance>
  </ProbabilityModel>
</ClusterModel>
    
```

### MPEG-7 Meta-data for Content Description

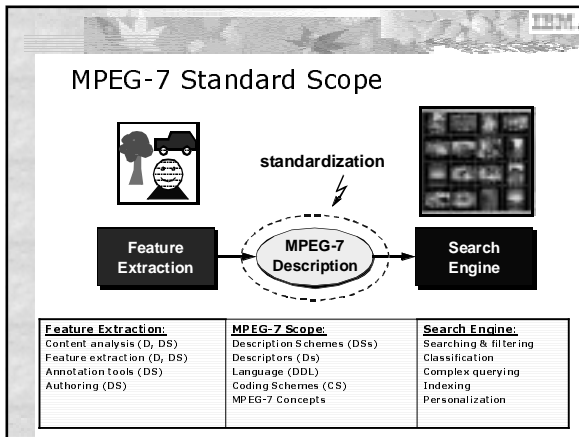
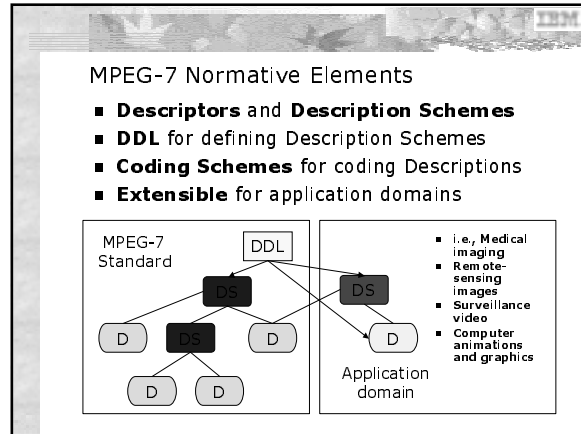
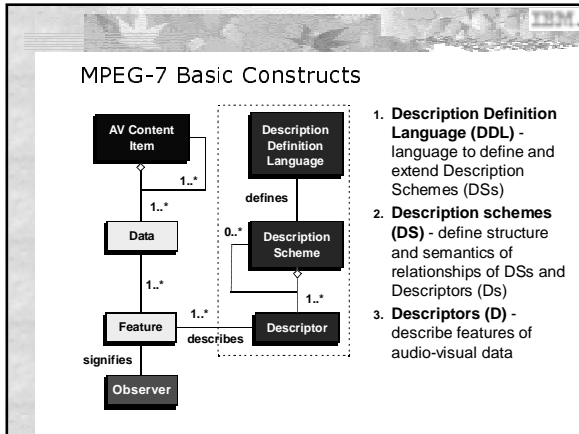
Data	Structure	Features	Models	Semantics
Images Video Audio Multimedia Formats Layout	Regions Segments Grids Motion Relationships (Space-temporal)	Color Texture Shape Motion Speech Timbre Melody	Clusters Classes Analytic models Probability models Classifiers	Objects Events Actions People Labels Relationships

### MPEG-7 Meta-data for Content Management

### MPEG-7 Lifecycle of Multimedia Content and Metadata

■ Content and Metadata Life Cycle:

- Commission** – consumer content requirements
- Elaboration** – pre-production of content concept
- Capture** – shooting and acquisition of raw content
- Analysis** – feature extraction and content analysis
- Synthesis** – merging of content and description data
- Composition** – cut-down and editing
- Packaging** – final preparation for distribution
- Delivery** – transfer of production form to delivery vehicle
- Consumption** – viewing and receiving content
- Interaction** – hyperlinking, transactions, visualization



## MPEG-7 Description Definition Language

### MPEG-7 Description Definition Language (DDL): Key Concepts

- **Description Definition Language (DDL):**
  - Standardized, yet flexible language to define Description Schemes (DS) and Descriptors (D)
  - Extends W3C XML Schema Language (W3C Recommendation in May, 2001)
  - Expresses relations, object orientation, composition, and partial instantiation
- **MPEG-7 Systems:**
  - Specifies means for binarizing MPEG-7 descriptions
  - Specifies methodology for carrying descriptions as streams
  - Specifies means for accessing and synchronously consuming data
  - Specifies management and protection of data

### MPEG-7 DDL Requirements

- MPEG-7 DDL is the language for defining Description Schemes (DS) and Descriptors
- Expresses spatial, temporal, structural, and conceptual relationships
- Provides a rich model for links and references between one or more descriptions and the data that it describes
- The DDL is platform and application independent and generates definitions that are human- and machine-readable
- The DDL Parser validates Description Schemes (content and structure) and Descriptor data types, both primitive (integer, text, date, time) and composite (histograms, enumerated types)

### Example: MPEG-7 DDL Definition

- MPEG-7 Structured Annotation Type (DDL):

```

<complexType name="StructuredAnnotationType">
  <sequence>
    <element name="Who" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
    <element name="WhatObject" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
    <element name="WhatAction" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
    <element name="Where" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
    <element name="When" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
    <element name="Why" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
    <element name="How" type="mpeg7:TermUseType"
      minOccurs="0" maxOccurs="unbounded" />
  </sequence>
  <attribute ref="xml:lang" use="optional" />
</complexType>
    
```

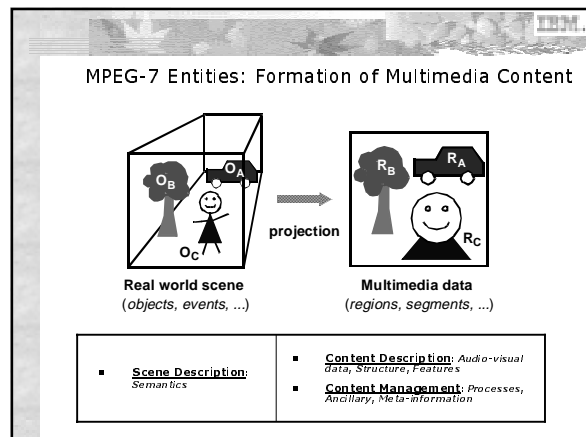
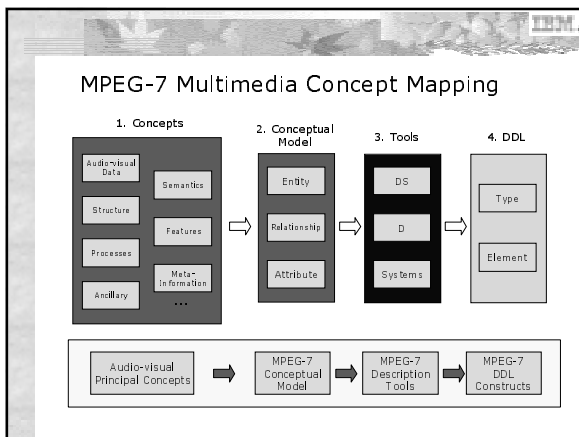
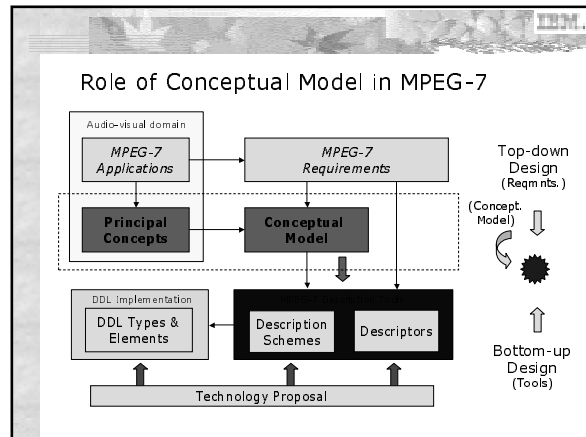
### Example: MPEG-7 DDL Definition

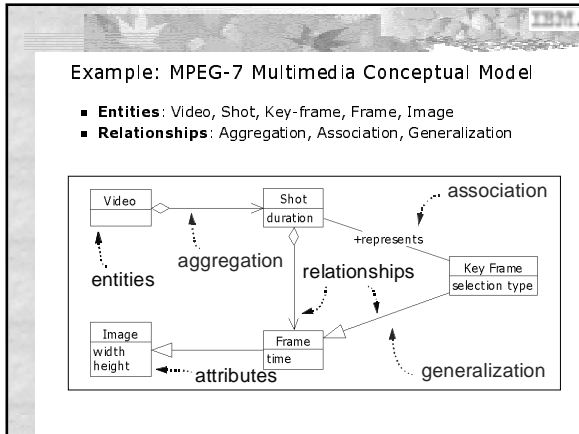
- MPEG-7 Moving Region Type (DDL):

```

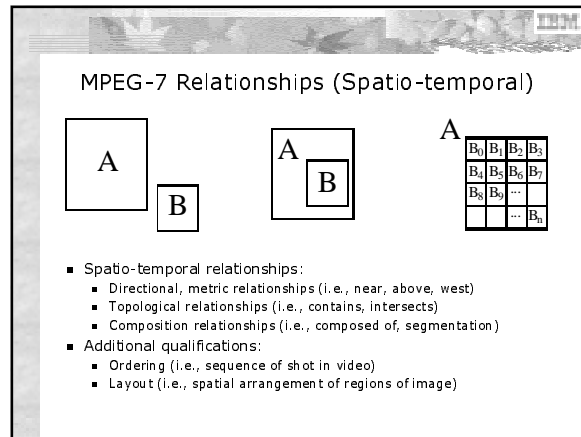
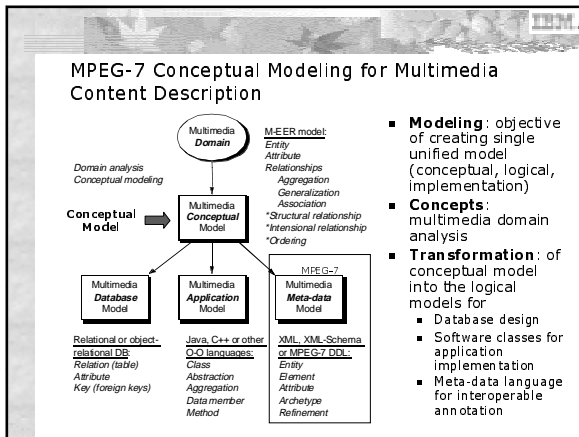
<complexType name="MovingRegionType">
  <complexContent>
    <extension base="mpeg7:SegmentType">
      <sequence>
        <element name="SpatioTemporalLocator" type="mpeg7:SpatioTemporalLocatorType"
          minOccurs="0" />
        <element name="SpatioTemporalMask" type="mpeg7:SpatioTemporalMaskType"
          minOccurs="0" />
        <choice minOccurs="0" maxOccurs="unbounded">
          <element name="VisualDescriptor" type="mpeg7:VisualDescriptorType" />
          <element name="VisualDescriptionScheme" type="mpeg7:VisualDescriptionSchemeType" />
        </choice>
        <choice minOccurs="0" maxOccurs="unbounded">
          <element name="SpatialDecomposition"
            type="mpeg7:SpatialDecompositionType" />
          <element name="TemporalDecomposition"
            type="mpeg7:TemporalDecompositionType" />
          <element name="SpatioTemporalDecomposition"
            type="mpeg7:SpatioTemporalDecompositionType" />
          <element name="MediaSourceDecomposition"
            type="mpeg7:MediaSourceDecompositionType" />
        </choice>
      </sequence>
    </extension>
  </complexContent>
</complexType>
    
```

## MPEG-7 Conceptual Model

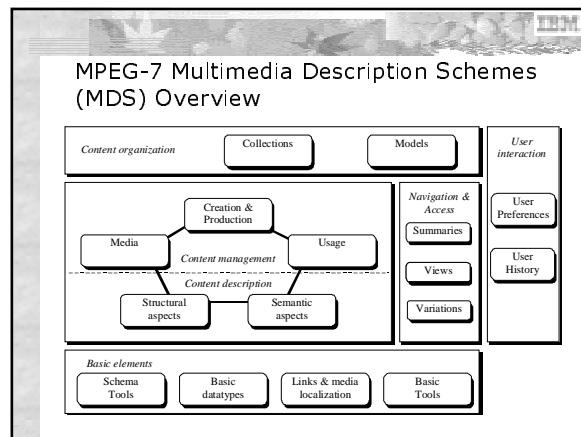




- ### MPEG-7 Principal Concepts (193)
- **Scene Description:**
    - **Semantics:** Action, Character, Conversation, Dialog, Episode, Event, Face, Object, Point of view, Pose, Scene, Script, Story, ...
  - **Content Description:**
    - **Audio-visual data:** Audio, Edition, Film, Graphic, Image, Mix, Mosaic, Music, Program, Promotion, Rush, Synchronization, Sound effect, Speech, Stream, Summary, Symbolic audio, Trademark-Image, Variation, Version, Video, View, ...
    - **Structure:** Animation, Audio spectrum, Background, Composition Effect, Connectivity, Cut, Duration, Edit, Embedded text, Frame, Image type, Internal Transition Effect, Key frame, Linguistic structure, Locator, Region, Segment, Sequence, Shot, Spatial geometry, Spatial relationship, Synchronization, Transition Effect, ...
    - **Features:** Audio features, Color, Deformation, Melody, Motion, Noisiness, Shape, Silence, Sketch, Texture, Timbre, Visual features, Volume, ...
  - **Content Management:**
    - **Processes:** Acquisition, Coding, Creation, Delivery, Editing, Manipulation, Presentation, Production, Publication, Segmentation, Storage, Streaming, Transcoding, ...
    - **Ancillary:** Bandwidth, Camera, Client, Instrument, Medium, Modality, Person, Recorder, Terminal, Usage, Usage history, User, User preference, ...
    - **Meta Information:** Annotation, Archive, Author, Classification, Context, Contract, Copyright, Date, Financial, Format, IPMP, Language, Locality, Market, Organization, Owner, Place, Quality, Rating, Rights, Technical staff, Text, Title, Translation, Unique identifier, ...



# MPEG-7 Multimedia Description Schemes

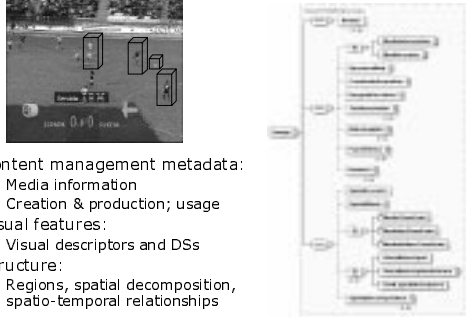




### Examples: MPEG-7 MDS Description Tools

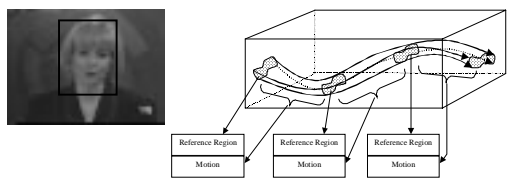
- **Audio-visual content:**
  - Images, video, audio, collections
- **Media and meta-information:**
  - Format, title, author, producer, date, address
- **Content and signal structure:**
  - Segments, segment decompositions, spatio-temporal relationships
  - Summaries, space and frequency views, variations
- **Audio-visual features:**
  - Visual: color, texture, shape, motion, camera motion
  - Audio: audio energy, melody, spoken content
- **Models:**
  - Descriptor model, analytic model, classification model
- **Semantics:**
  - Objects, events, concepts, places, states, abstractions

### Example: MPEG-7 Image Description



- Content management metadata:
  - Media information
  - Creation & production; usage
- Visual features:
  - Visual descriptors and DSs
- Structure:
  - Regions, spatial decomposition, spatio-temporal relationships

### MPEG-7 MDS: Moving Regions

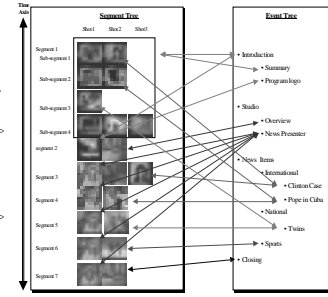


- **Moving Regions:**
  - Region Feature description (shape, color, texture)
  - Spatio-Temporal Locator
  - Regions Motion Trajectory and Tracking


### MPEG-7 MDS: Video Segment

```

<VideoSegmentId = "VS1">
  <MediaTimeMask NumOfIntervals = "2">
  <MediaTimeMask>
  <SegmentDecomposition Gap = "true"
    Overlap = "true"
    DecompositionType = "temporal">
  <VideoSegmentId = "VS2">
  <MediaTime
    <MediaTimePoint>
      <ch>0</ch> <m>0</m> <st>0</st>
    </MediaTimePoint>
  </MediaTime>
  </VideoSegment>
  </MediaTime>
  <MediaTimePoint>
    <ch>0</ch> <m>0</m> <st>0</st>
  </MediaTimePoint>
  </MediaTime>
  </VideoSegment>
  </SegmentDecomposition
  </VideoSegment>
    
```

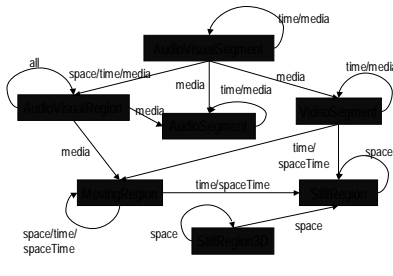


### MPEG-7 MDS: Analytic Video Segment

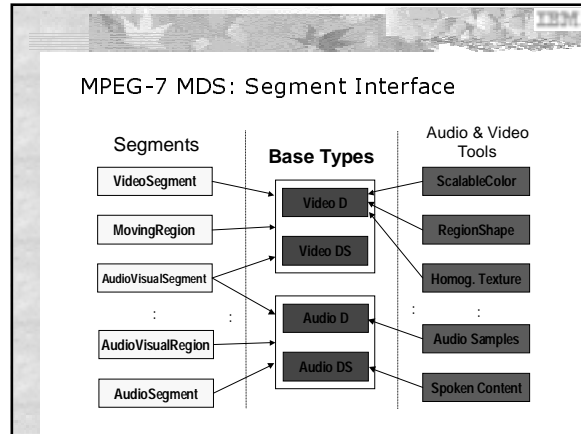
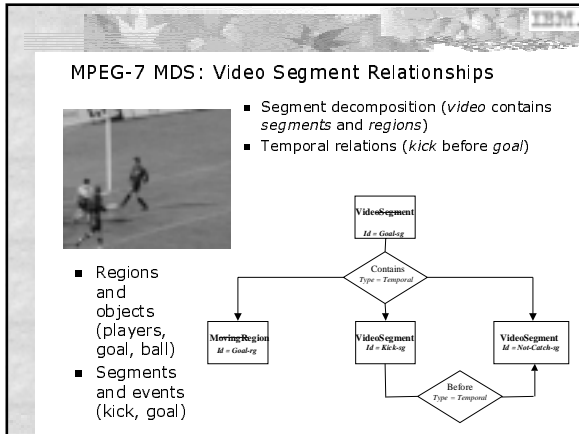


Level	Video Editing	Transition
1	Shot	Global Transition
2	Composition Shot	Composition Transition
3	Intra Composition Shot	Internal Transition

### MPEG-7 MDS: Segment Decomposition

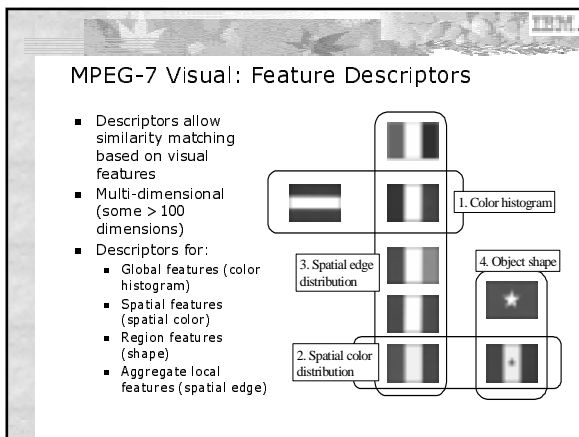


- **Spatial** (image decomposition into regions)
- **Temporal** (video decomposition into segments)
- **Spatio-temporal** (video decomposition into moving regions)
- **Media** (video decomposition into audio and video tracks)



# MPEG-7 Audio and Visual Descriptors

- ### Summary of MPEG-7 Visual Descriptors (multi-dimensional features) – 25 tools
- Color (7):**
    - ColorLayout
    - ColorSpace
    - ColorStructure
    - ColorQuantization
    - DominantColor
    - GoGoPCColor
    - ScalableColor
  - Shape (4):**
    - EdgeHistogram
    - ContourShape
    - RegionShape
    - Shape3D
  - Texture (2):**
    - HomogeneousTexture
    - TextureBrowsing
  - Composition (5):**
    - GridLayout
    - RegularTimeSeries
    - Spatial2DCoordinateSystem
    - TemporalInterpolation
    - IrregularTimeSeries
  - Locator (3):**
    - RegionLocator
    - SpatioTemporalLocator
    - MultipleView
  - Motion (3):**
    - MotionActivity
    - MotionTrajectory
    - ParametricMotion
  - Semantic (1):**
    - FaceRecognition



### Example: MPEG-7 Visual Descriptor Definition (Scalable Color)

- MPEG-7 Scalable Color Type (DDL):

```

<complexType name="ScalableColorType" final="#all">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <sequence>
        <element name="Coefficients" type="mpeg7:integerVector"/>
      </sequence>
      <attribute name="numberOfCoefficients" type="mpeg7:unsigned3"/>
      <attribute name="numberOfBitplanesDiscarded" type="mpeg7:unsigned3"/>
    </extension>
  </complexContent>
</complexType>
    
```

### Example: MPEG-7 Visual Descriptor Definition (Dominant Color)

- MPEG-7 Dominant Color Type (DDL):

```

<complexType name="DominantColorType" final="#all">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <sequence>
        <element name="ColorSpace" type="mpeg7:ColorSpaceType" minOccurs="0"/>
        <element name="ColorQuantization" type="mpeg7:ColorQuantizationType" minOccurs="0"/>
        <element name="SpatialCoherency" type="mpeg7:unsigned5"/>
        <element name="Values" maxOccurs="8">
          <complexType>
            <sequence>
              <element name="Percentage" type="mpeg7:unsigned5"/>
              <element name="ColorValueIndex" . . . </element>
              <element name="ColorVariance" minOccurs="0" . . . </element>
            </sequence>
          </complexType>
        </element>
      </sequence>
      <attribute name="size" _ _ </attribute>
    </extension>
  </complexContent>
</complexType>
    
```

### MPEG-7 Visual: Color Spaces

HSV color space

HMMD color space

### MPEG-7 Visual: Texture Descriptors

- Homogeneous Texture:
  - Rotation and scale invariance

- Texture Browsing
- Edge Histogram

### Example: MPEG-7 Visual Descriptor Definition (Homogenous Texture)

- MPEG-7 Homogeneous Type (DDL):

```

<complexType name="HomogeneousTextureType" final="#all">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <sequence>
        <element name="Average" type="mpeg7:unsigned8"/>
        <element name="StandardDeviation" type="mpeg7:unsigned8"/>
        <element name="Energy" type="mpeg7:textureListType"/>
        <element name="EnergyDeviation" type="mpeg7:textureListType" minOccurs="0"/>
      </sequence>
    </extension>
  </complexContent>
</complexType>

<simpleType name="textureListType">
  <restriction base="mpeg7:unsigned8"/>
  <list itemType="mpeg7:unsigned8"/>
  </simpleType>
  <length value="30"/>
</restriction>
</simpleType>
    
```

### MPEG-7 Visual: Shape Descriptors

- 2D contour based descriptors:
  - Rotation and scale invariance and invariance to small non-rigid deformations
- 2D area based descriptors:
  - Complex shapes such as trademarks

- 3D shape descriptors:

### Example: MPEG-7 Visual Descriptor Definition (Region Shape)

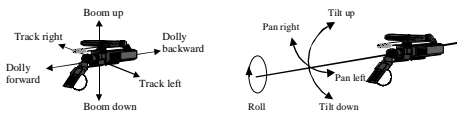
- MPEG-7 Region Shape (DDL):

```

<complexType name="RegionShapeType" final="#all">
  <complexContent>
    <extension base="mpeg7:VisualDType">
      <element name="ArtDE">
        <simpleType>
          <restriction base="mpeg7:listOfUnsigned4Type">
            <length value="35"/>
          </restriction>
        </simpleType>
      </element>
    </extension>
  </complexContent>
</complexType>
    
```


Angular Radial Transform basis functions (35 dimensions)

### MPEG-7 Visual: Motion Descriptors

- Camera motion description:**
  - Camera parameters
 
- Motion trajectory
- Parametric motion
- Motion activity
  - Intensity, Direction, Spatial and Temporal

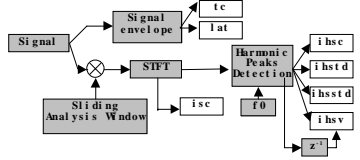
### MPEG-7 Audio: Melody Description

- Targeted application: Query-by-humming
- Note Lattice DS:
  - Notes labeled with time, duration and pitch, and optionally timbre, instrument, estimation reliability
  - Accommodates monophonic and polyphonic pitch information, derived from
- Melody DS:
  - Describes coarsely quantized pitch-interval information (5 levels of change) and rhythm (meter and beat)

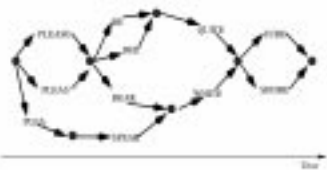


### MPEG-7 Audio: Timbre Description

- Describes perceptual features of instrument sounds (features that distinguish sounds having the same pitch and loudness)
- Relates to audio notions of "attack", "brightness" and "richness"

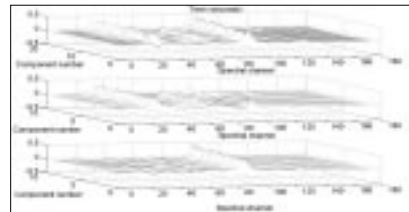


### MPEG-7 Audio: Spoken Content

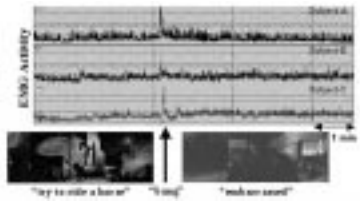
- Describes word and phone (sub-word units) lattices for speakers (audio)
- i.e., hypothetical decoding of the phrase "Please be quite sure":
 

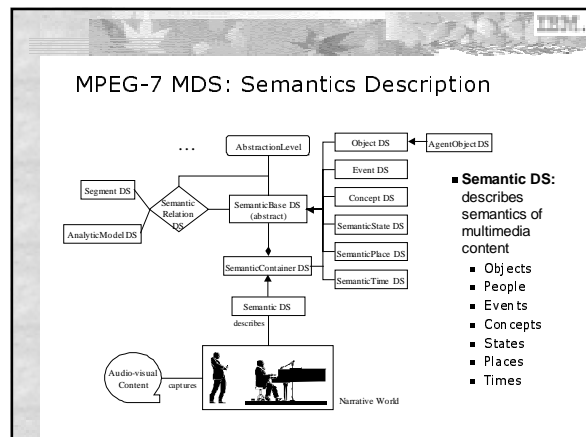
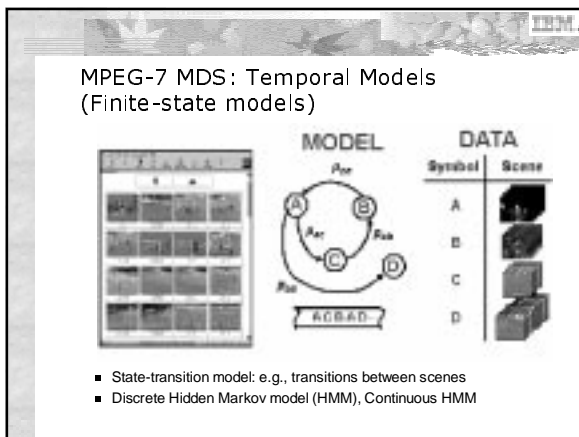
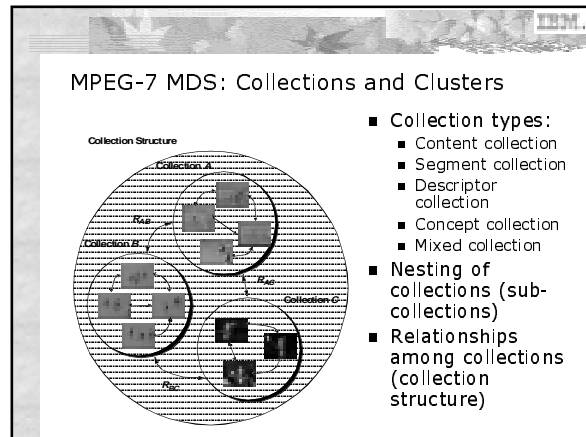
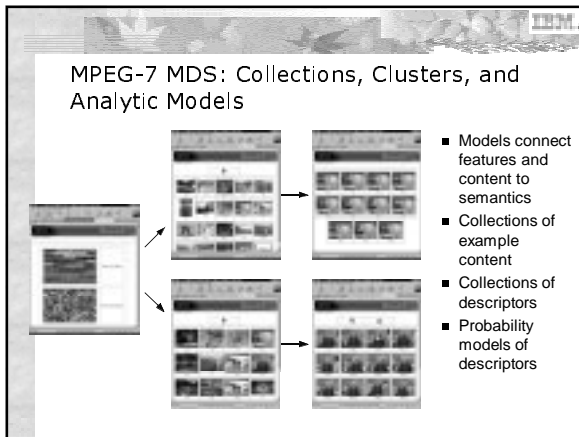
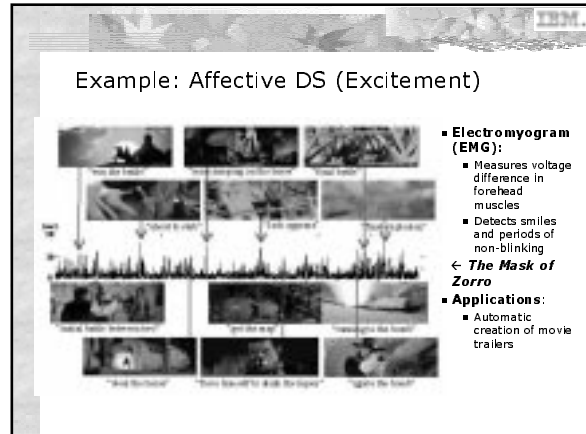
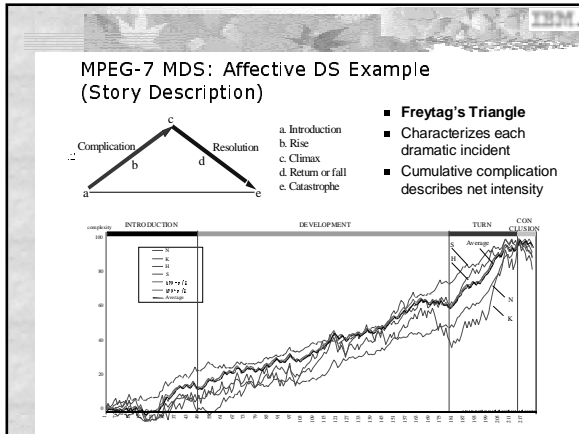
### MPEG-7 Audio: Audio Independent Components

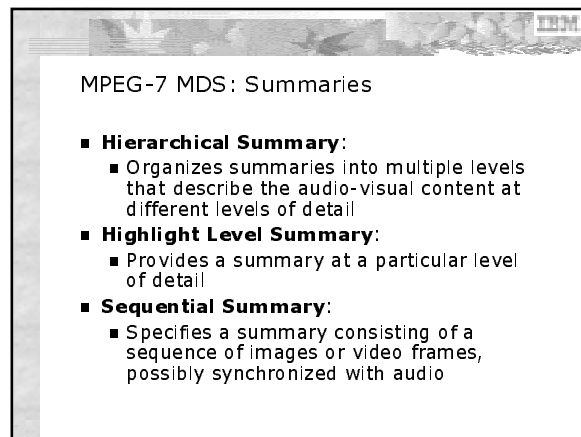
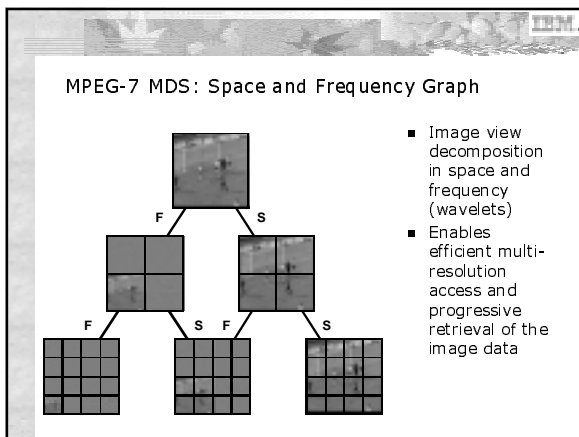
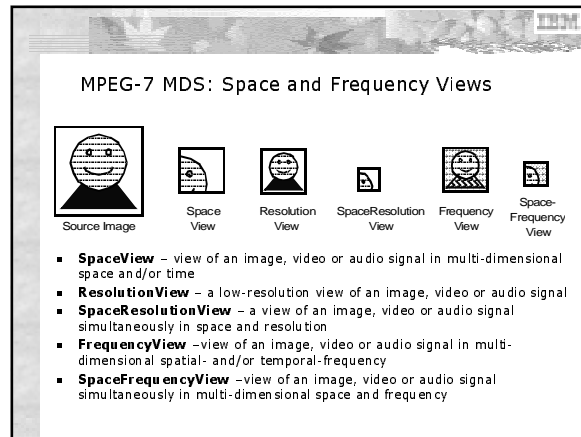
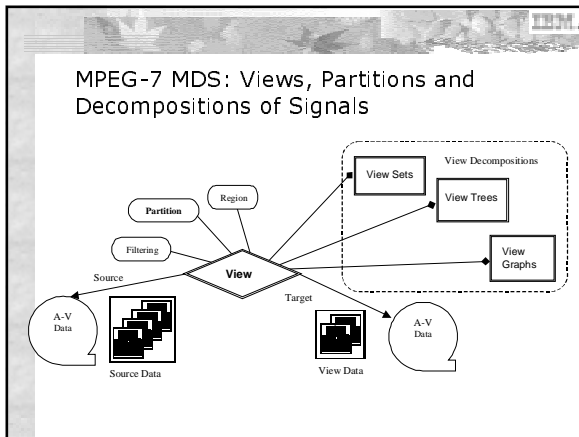
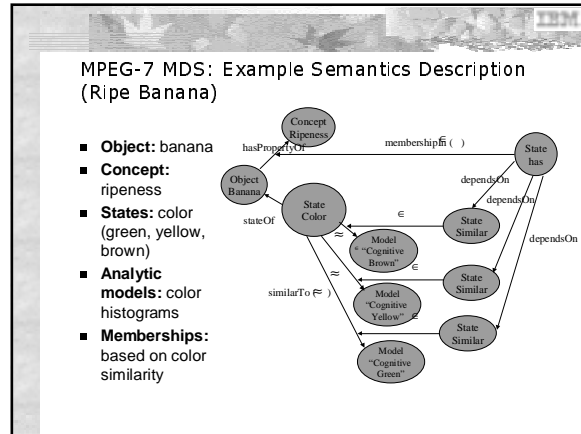
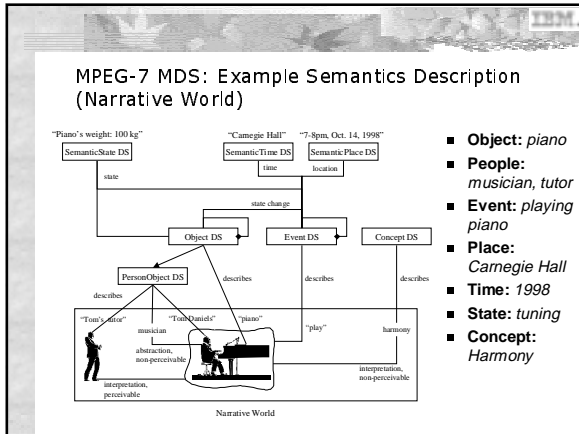
- Describes the decomposition of an audio spectrogram into a collection of statistically independent spectral and temporal features

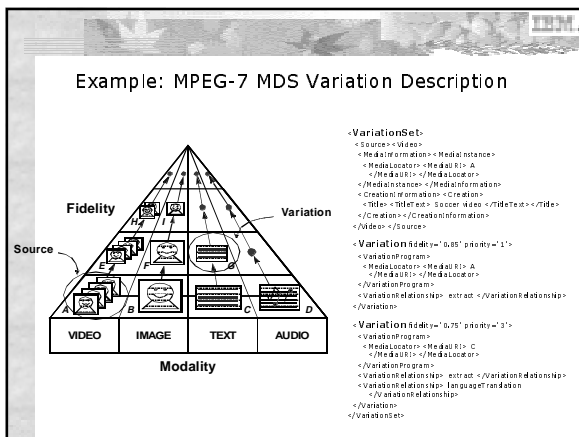
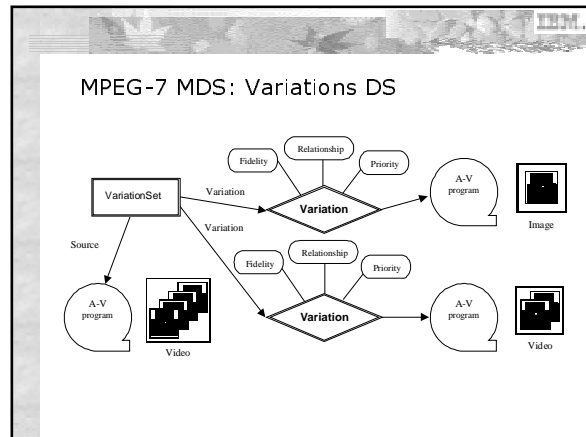
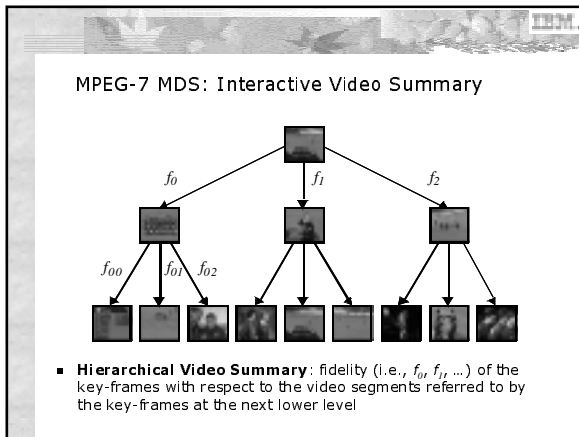
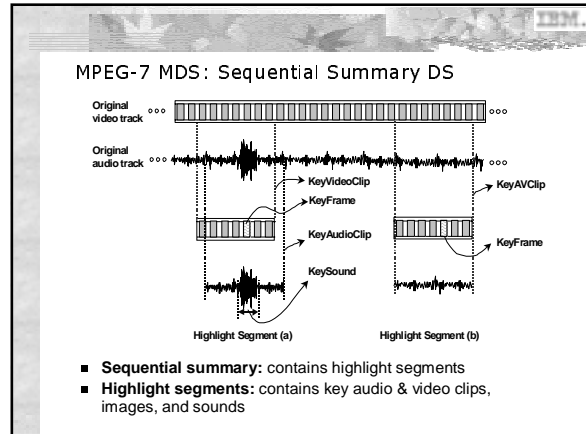
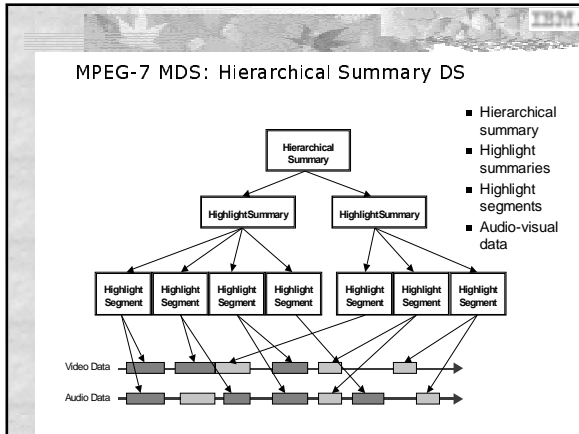


### MPEG-7 MDS: Affective DS

- Describes:** audience's affective (analytical, emotional or mood) response to multimedia content
- Example:** pattern of emotional tension
- Automatic Extraction:** via physiological measurements and analysis (real-time)
- Manual Extraction:** Audience survey (high temporal-resolution)
 







## MPEG-7 Applications

### Example MPEG-7 Applications

- **MPEG-7 Interoperable Content-Based Retrieval (CBR):**
  - Standardized Descriptors (Ds) and Description Schemes (DSs)
  - Content Description – features, structure, models, semantics
  - Content Management – meta-information: author, owner, rights; media information
- **MPEG-7 Universal Multimedia Access:**
  1. **Content Variations** – storing, selecting and delivering different variations of multimedia content
  2. **Transcoding Hints** – on-the-fly adaptation to device and network conditions
  3. **Video Summarization** – personalized interactive summaries of multimedia content
  4. **Views** – progressive retrieval of scalable representations of multimedia content

### MPEG-7 Interoperable Multimedia Search and Retrieval

**Search and Retrieval:**

- **Search:** color, texture, shape, motion, spatial
- **Browse:** video parsing, shot detection, key-frames
- **Filter:** object detection, classification

### MPEG-7 Content Analysis (Ingestion Engine)

### MPEG-7 Querying (Search Engine)

### Summary of MPEG-7

1. MPEG-7 is not a video coding standard
2. MPEG-7 is a metadata standard for describing multimedia content
3. MPEG-7 defines a standard schema (Descriptors and Description Schemes) using XML Schema Language
4. MPEG-7 produces XML descriptions
5. MPEG-7 introduces new challenges for searching (*similarity searching, multiple objects, multiple features, fuzzy relationships*)

## MPEG-21 Multimedia Framework



### MPEG-21 Multimedia Framework

- "Interoperable Multimedia Framework"
- "E-Commerce for E-Content"
- "Digital Audio-Visual Framework"
- **Vision**: "To enable transparent and augmented use of multimedia resources across a wide range of networks and devices."
- **Goal**: Integration of technologies for content *identification* and *consumption*
- **Output**: ISO technical report and technical specification (International Standard in late 2001)

### MPEG-21 Timeline

- **Oct '99** - MPEG looks forward:
  - **Study**: "Technologies for E-content"
  - **Report**: "A Multimedia Framework"
  - No new coding technology on horizon
- **Dec '99** - First steps towards MPEG-21:
  - **Statement**: "Rationale for the Digital Audio-Visual Framework"
  - **Proposal**: "Proposal for New Work Item for MPEG-21" (N3200)
- **July '01** - Current Draft Documents:
  - *Digital Item Declaration (DID)* (Working Draft)
  - *Digital Item Identification and Description (DII&D)* (Working Draft)
  - *Rights Data Dictionary (RDD) and Rights Language* (Requirements)

### Parts of MPEG-21 Standard

- **Multimedia Framework Architecture:**
  1. Digital Item Declaration (what is the unit of transaction and distribution?)
  2. Digital Item Identification and Description (what content?)
  3. Content Handling and Usage (how is content used and delivered?)
  4. Intellectual Property Management and Protection (how are rights controlled in respect to each User?)
  5. Terminals and Networks (how is content delivered?)
  6. Content Representation (synthetic/natural content, scaling?)
  7. Event Reporting (activity in any of above)

### MPEG-21 Interactions

1. Digital Item Declaration
2. Digital Item Identification and Description
3. Content Handling and Usage
4. Intellectual Property Management and Protection
5. Terminals and Networks
6. Representation
7. Event Reporting

### Summary of MPEG-21

- **MPEG-21:**
  - Challenging scope in volving standardizing framework for multimedia interoperability
  - Some gluing of existing standards (MPEG-4, MPEG-7)
  - Targeting whole content lifecycle (*creation to delivery to use to discard*)
  - Key elements involve content *identification* and *consumption*
- **MPEG-21 Market place for Multimedia content**
  - Many to many distribution
  - Everybody has role of content producer, distributor, user, matchmaker, broker, service provider - annotation, searching, filtering
  - Shift from old models - few content providers, pay per view, pub/sub, direct channels
  - New models - new content usage models - plays per purchase, agents, searching
  - Variable costing via scheduling and QoS
  - Mobile users and pervasive devices

### Multimedia Database Search Problems

### Multimedia Search Problems

- **Querying of MPEG-7 = Querying of XML**
- **Traditional Databases:**
  - Correctness of matching, optimization, efficiency
- **Multimedia Content-based Queries:**
  - Top-*k* searches with ranks based on similarity scores
  - Computation of distance metric (domain dependent, subjective)
  - Varying precision vs. recall
  - High-dimensional feature spaces and multi-dimensional indexing
  - Fuzzy relationships, arbitrary join predicates
  - Combinatorial search (i.e., among *regions, segments, objects, events*)

### Taxonomy of Multimedia Similarity Searching

- **Global features:** (*images, video, audio*)
- **Local features:** (*regions, segments, objects, events*)
  - Single region (local features)
  - Multiple regions (no relationships)
  - Multiple regions (relationships)

### Problem 1: Similarity Matching of Global Features

- Feature extraction from query and target images
- Similarity measured by distance between descriptors
- Multi- and high-dimensional indexing
- Weighting of multiple features
- Relevance feedback searching

### Example: Global Image Features

- a. **Source Image**
- b. **Texture:** spatial-frequency energy (9 dimensions)
- c. **Color:** color histogram (166 dimensions)
- d. **Color composition:** grid color histogram (890 dimensions)
- **Non-normative:**
  - Histogram normalization
  - Distance metrics

### Example: Histogram Normalization

- Removes dependency on image size (# samples)
- L2 norm (Euclidean): vectors on surface of hypersphere
- L1 norm: vectors on surface of simplex

### Example: Color Histogram Matching

- Minkowski Metrics (i.e., L1, L2)
- Direct comparison of histogram bins (like colors only)
- Quadratic Metrics
- Cross comparison of colors (color similarity or correlation)

### Multimedia Top-k Similarity Search (Global Features)

```

<?xml?>
<ContentDescription xsi:type="ContentEntityType">
<MultimediaContent xsi:type="ImageType">
<VisualDescriptor xsi:type="ColorHistogramType">
<Bins dim="256"> 0.02 0.01 ... 0.07 </Bins>
</VisualDescriptor>
<VisualDescriptor xsi:type="TextureType">
<Coeffs dim="9"> 0.04 0.03 ... 0.02 </Coeffs>
</VisualDescriptor>
<VisualDescriptor xsi:type="ColorCompositionType">
<Coeffs dim="80"> 0.01 0.03 ... 0.01 </Coeffs>
</VisualDescriptor>
<MatchingHint reliability="1">
<Hint value="0.5">
  xpath="..."//VisualDescriptor[1]*/
<Hint value="0.3">
  xpath="..."//VisualDescriptor[2]*/
<Hint value="0.2">
  xpath="..."//VisualDescriptor[3]*/
</MatchingHint>
</MultimediaContent>
</ContentDescription>
</?xml?>
    
```

- Top-k similarity search
- $n$  images
- $m$  features
- Sequential search:
  - $O(nm)$
- Multi-search (Fagin):
  - $O(n^{(m-1)/m} k^{1/m})$
- Quick combine (Guntzer et al):
  - Reduction by  $\frac{m}{\sqrt[m]{m!}}$

### Global Feature Similarity Search: Retrieval Effectiveness

- Precision:** proportion of retrieved documents that are relevant
- Recall:** proportion of relevant documents that are retrieved
- A more effective system shows higher precision for all values of recall

### Multi-search problem

- Problem Definition:** Similarity search over  $N$  database objects:
  - Each object has  $m$  attributes (i.e., color, shape, texture)
  - Have sorted list for each attribute (i.e., ranked match results)
  - Have monotone aggregation function or combining rule (i.e., *min*, *average*)
- Query:** determine top  $k$  objects:
  - Naive algorithm – examines all  $N$  objects
  - Fagin’s algorithm (FA) – optimal with high probability for some aggregation functions in the worst case
  - Threshold algorithm (TA) – optimal for **ALL** monotone aggregation functions in **ALL** cases (instance optimal – worst, best, average)
  - No random access algorithm (NRA) – optimal combining of ranked streams

### Related Work on Multi-search

- Fagin (*ACM PODS '96, ACM PODS '98*) – Fagin’s Algorithm (FA)
- Nepal and Ramakrishna (*IEEE ICDE '99*)
- Guntzer, Balke, Kiessling (Quick Combine, *VLDB '2000*) – (TA)
- Guntzer, Balke, Kiessling (Stream Combine, *IEEE ITCC 2001*) – (NRA)
- Fagin (*ACM PODS 2001*) – Instance optimality of TA

### Fagin’s Multi-Search Algorithm (FA) – ACM PODS '96 and ACM PODS '98

- Example:** Find top-2 objects with (Color = 'red') AND (Shape = 'square')

### Fagin’s Multi-Search Algorithm (FA)

- FA Procedure:**
  - Sorted access:** Do sorted access in parallel on  $m$  sorted lists. Wait until at least  $k$  matches ( $k$  objects seen in all  $m$  lists)
  - Random access:** For each object  $R$  seen in any list, do random access in all lists to retrieve all  $m$  attributes
  - Compute score for each object  $R$  and return top  $k$
- Correct for monotone aggregation functions
- If sorted lists are probabilistically independent, then middleware cost is  $O(N^{(m-1)/m} k^{1/m})$
- Worst case optimal if aggregation function is “strict” (i.e., optimal for *conjunction*, not optimal for *max*)

### Threshold Algorithm (TA) – Guntzer, et al., VLDB 2000 (“Quick Combine”)

- Intuition: Halt as soon as top- $k$  answers have appeared
- Procedure:
  - Sorted access:** Do sorted access in parallel on  $m$  sorted lists.
  - Random access:** As an object  $R$  is seen in any list, do random access in all lists to retrieve all  $m$  attributes and compute score. If score is one of top- $k$  so far, remember object  $R$  and its score.
  - Define threshold score  $T$  that is computed from score of combined last objects in each list. As soon as  $k$  objects retrieved with score at least equal to  $T$ , STOP.
  - Return top  $k$
- Fagin shows correctness for monotone aggregation functions, and that always stopping rule for TA < FA
- Fagin shows that TA is instance optimal over all algorithms that correctly find top- $k$  answers, over all classes of databases (excluding algorithms that make wild guesses)

### Problem 2: Similarity Matching of Local Features (Regions and Segments)

- Region extraction from query and target images
- Similarity measured by combined distance of regions
- Multiple matches of regions
- Possible scoring of combinations of regions

### Example: Local Image Features (single region)

```

<Mpeg7>
<ContentDescription xsi:type="ContentEntityType">
<MultimediaContent xsi:type="ImageType">
<SpatialDecomposition>
<StillRegion id="reg0">
<SpatialLocation>
<Box<<Coords dim="2 2"> x y w h
</Coords>/Box>
</SpatialLocation>
<VisualDescriptor
xsi:type="ColorHistogramType">
<Bins dim="256"> 0.01 0.03 ... 0.01 </Bins>
</VisualDescriptor>
<VisualDescriptor xsi:type="TextureType">
<Coeffs dim="9"> 0.02 0.01 ... 0.03
</Coeffs>
</VisualDescriptor>
</StillRegion>
</SpatialDecomposition>
</MultimediaContent>
</ContentDescription>
</Mpeg7>
    
```

- Multiple features:
  - color, texture, composition
- Spatial features:
  - Location, size, shape, boundary features, contour

### Example: Multiple Regions with features

- Multiple regions
  - reg0 ("sky")
  - reg1 ("building")
  - reg2 ("water")
- Multiple features (color, texture, composition)
- Spatial features (i.e., size, location, shape)

### Top-k Similarity Search (Multiple Regions with no Relationships)

```

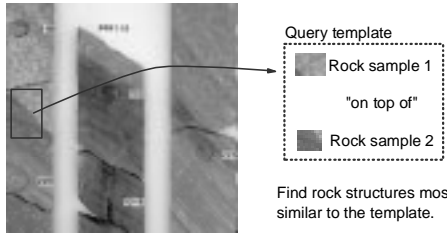
<Mpeg7>
<ContentDescription
xsi:type="ContentEntityType">
<MultimediaContent
xsi:type="ImageType">
<SpatialDecomposition>
<StillRegionRef idref="reg0">
<StillRegionRef idref="reg1">
<StillRegionRef idref="reg2">
</SpatialDecomposition>
</MultimediaContent>
</ContentDescription>
</Mpeg7>
    
```

- Top- $k$  similarity search
- $n$  images;  $m$  features
- $n_q$  query regions;  $n_t$  target regions per image
- Nested multi-search (lossy):
  - $O(n_q n_t^{(m-1)/m} k^{1/m})$
- Sequential search:
  - $O(n n_q n_t)$
- Example: ( $n=10^4$ ,  $n_q=10$ ,  $n_t=3$ ):
  - $O(3 \times 10^7)$  evaluations

### Example: Multiple Regions with Relationships

- Multiple regions
  - reg0 ("sky")
  - reg1 ("building")
  - reg2 ("water")
- Multiple features (color, texture, composition)
- Spatial features (i.e., size, location, shape)
- Fuzzy and crisp relationships
  - Relationships (spatial):
    - reg0 ("sky") **right** of reg1 ("building")
    - reg1 ("building") **above** reg2 ("water")
    - reg2 ("water") **below** reg0 ("sky")

### Fuzzy Cartesian Spatial Search: Example



Query template

- Rock sample 1
- "on top of"
- Rock sample 2

Find rock structures most similar to the template.

### Example Constraints (Spatial, temporal; fuzzy, crisp)

- By type
  - Topological (i.e., inclusion, union, intersection, negation)
  - Directional (i.e., before/after, left/right, above/below)
  - Metric (i.e., close/far, 0.1m away, 5 minutes before)
- By precision
  - Fuzzy (i.e., soon) vs. crisp (i.e., within 1 hour)
- By arity
  - Unary (i.e., shape is square), binary, n-ary (i.e., mapped by an affine transform)
- By dimension
  - 1-D (i.e., left of, before, smaller, close to), 2-D (i.e., 30 degrees NW of, top-left, near), k-D (i.e., overlapping, within certain distance)
- By domain
  - Boolean, spatial, scale, temporal, multimedia

### Top-k Similarity Search (Multiple Regions with Relationships)

```

<?xml?>
<ContentDescription xsi:type="ContentEntity">
  <MultimediaContent xsi:type="Image">
    <SpatialDecomposition>
      <StillRegionRef idref="reg0"/>
      <StillRegionRef idref="reg1"/>
      <StillRegionRef idref="reg2"/>
    </SpatialDecomposition>
    <MultimediaContent>
      <Relationships>
        <Relation xsi:type="DirectionalRelationType"
          name="right" source="reg0" target="reg1"/>
        <Relation xsi:type="DirectionalRelationType"
          name="above" source="reg1" target="reg2"/>
        <Relation xsi:type="DirectionalRelationType"
          name="below" source="reg2" target="reg0"/>
      </Relationships>
    </MultimediaContent>
  </ContentDescription>
</?xml?>
    
```

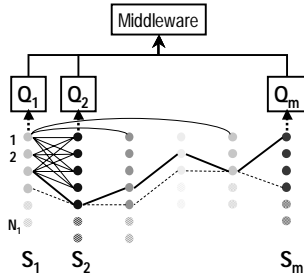
- Top-k similarity search
- $n_q$  query regions;  $n_t$  target regions per image
- $O(n^{n_q})$  combinations of regions for each of  $n$  images
- Naïve search over regions (exhaustive):
  - $O((nq)^{nq})$  evaluations
- Lossy search (pruning region lists to  $l_q$  regions):
  - $O(l_q^{n_q})$  evaluations
- Example: ( $n=10^2$ ,  $n_t=10$ ,  $n_q=3$ ,  $l_q=10^3$ ):
  - $O(10^9)$  evaluations

### Related Work on Fuzzy Cartesian Search

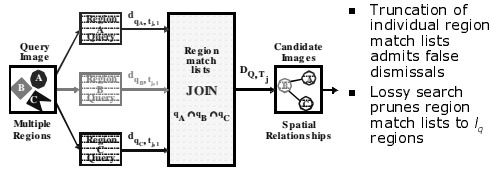
- Smith and Chang (*ACM Multimedia '97*) - SaFe (spatial and feature query algorithm)
- Li, et al (*SPIE Multimedia Databases '98*) - SPROC
- Natsev, et al (*VLDB 2001*) - J\* Algorithm

### Fuzzy Cartesian Query Model

- $m$  ranked streams
- $p$  join constraints
- Score aggregation
- $k$  desired outputs
- Middleware cost:
  - Sorted access
  - Random access
  - Predicate access

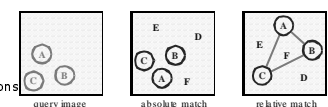


### Region-based Queries (SaFe '97)



- Truncation of individual region match lists admits false dismissals
- Lossy search prunes region match lists to  $l_q$  regions

- Absolute match:
  - Similar location of regions
- Relative match:
  - Similar layout of regions



### SPROC Algorithm in SPIRE (Li, et al, '98)

- Represents matches by weighted nodes and fuzzy relationships by weighted edges
- Linearizes relationship graph by unwinding cycles and branches through node replication
- Prunes set of edges and set of nodes that need to be maintained for correctness

SA SAB SB SBC SC SCD SD SDE SE SEF SF

### Algorithm J\* (Natsev, et al 2001)

- A\*-like algorithm**
  - Streams – variables; tuples – possible variable values
  - States – sets of variable assignments (or instantiations)
  - Solutions – complete states
  - Potential(state) = max score if solution agrees with given state
- Algorithm J\*:**
  - Maintain priority queue of states based on their potential
  - If state at the head of the queue is complete, return
  - Else, expand state and reinsert

### J\* Matching Algorithm: Example

matches			
Score	A	B	...
0.90	1	1	...
0.83	2	1	...
0.67	3	2	...

Score A B		
1.0	1	1
0.93	1	1
0.93	2	1
0.90	1	1
0.90	1	2
0.86	2	1
0.86	3	1

matches			
Score	ID	...	...
0.90	1	...	...
0.80	2	...	...
0.70	3	...	...

matches			
Score	ID	...	...
0.90	1	...	...
0.60	2	...	...

On top of

A1 A2 A3  
B1 B2 B1 B2 B1 B2  
0.90 0.83 0.67

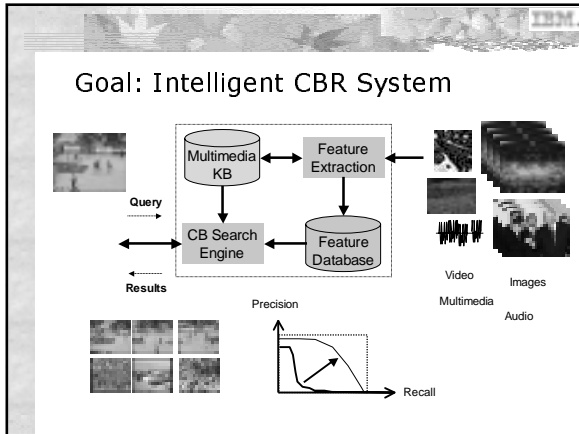
### Algorithm J\*: Heuristics

- Heuristics for choosing next free variable (stream):
  - Variable that has smallest number of values considered so far?
  - Variable that can contribute the most to the overall score?
  - Variable that is expected to contribute the most?
  - Variable that is most constrained?
  - Variable that is least constraining?

### Summary of Multimedia Search Problems

- Multimedia Content-based Queries:**
  - Single features (I.e., color, texture, shape), metrics, multi-dimensional indexing
  - Multiple features (multi-search)
  - Multiple objects and multiple features
  - Multiple objects with multiple features and fuzzy and crisp relationships (I.e., spatial, temporal)

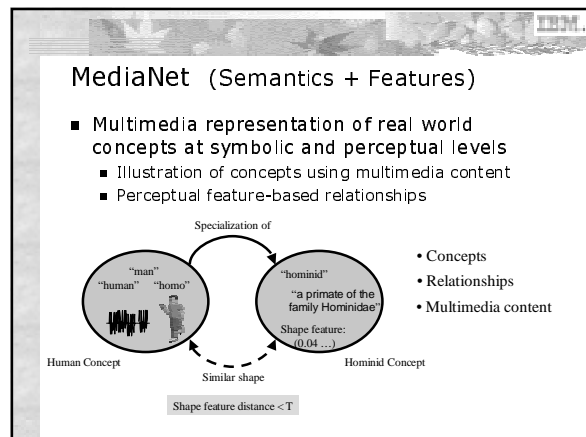
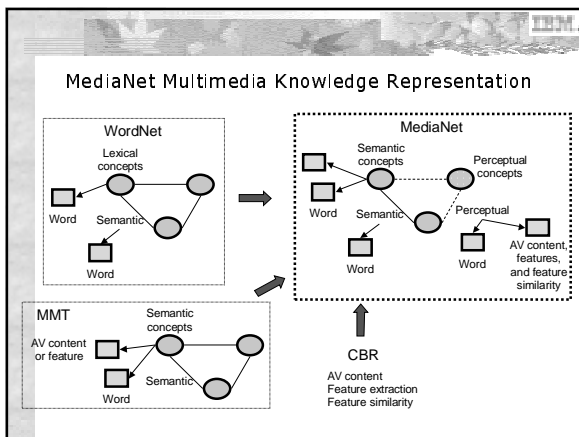
### Multimedia Knowledge Representation



- ### Pathways Towards Intelligence
- Artificial Intelligence:
    - Logics and knowledge representation schemes
    - Dimensions of AI: symbolic vs. non-symbolic, behavior vs. thought, human performance vs. rationality
  - Logics (e.g., First-Order-Logic):
    - Language, facts (knowledge), rules (deduction)
    - Semantic objects and relationships
  - Knowledge Representation:
    - Concepts and relationships, Semantic Nets, Frames, Scripts
    - Cyc: everyday life knowledge (encyclopedia) plus context
    - WordNet: electronic lexical system for English
    - Visual pattern libraries, multimedia thesauri

- ### WordNet (Semantics)
- Electronic lexical system for English:
    - Organizes words in sets of synonyms or "lexicalized" concepts
    - One "sense" per synonym set
    - Semantic relationships between synonym sets
      - Antonymy: To be opposite
      - Hypernymy/hyponymy: To be an generalization
      - Meronymy/holonymy: To have part, member, or substance
      - Troponymy: To be a manner of
      - Entailment: To entail

- ### Multimedia Thesaurus (Features)
- Multimedia Thesaurus (MMT) [Tansley'98]:
    - Concepts as abstract real-world entities
    - Semantic relationships among concepts (generalization)
    - Media representations of concepts (content and features)
  - Texture Image Thesaurus [Ma'98]:
    - Texture pattern library for CBR
  - SAFE: Visual Feature Library [Smith'97]:
    - Library of color and texture patterns
    - Region extraction via back-projection



### MediaNet Constructs

- Nodes:
  - Concepts and multimedia content
  - Relationships represented by multimedia content
- Edges:
  - Relationships among concepts
  - Associations among concepts and multimedia content

### MediaNet Entities

- Real world entities (objects and events):
  - Inanimate objects (Rock)
  - Living entities (Fish)
  - Events (Basketball game)
  - Properties (Blue)
- Types of objects:
  - Classes (Shack)
  - Identified entities (Ronald Reagan)
  - Abstract (Beauty)
  - Perceptual (Texture pattern)

### MediaNet Relationships

- Semantic relationships: (Extended Entity Relationship Model, WordNet)
  - Generalization and specialization (hominid, human)
  - Aggregation and composition (martini, gin)
  - Opposites (white, black)
  - Entailment (divorce, marry)
  - Manner of (whisper, speak)
- Perceptual relationships: (CBR feature similarity)
  - Audio-visual relationships illustrated by audio-visual content
  - Examples:
    - To sound similar (images of a bird and a frog)
    - To have similar shape (images of two different birds)

### MediaNet Content

- MediaNet Multimedia Content: Media representations
  - Audio-visual content (images, video, audio, regions, segments, objects, text, words, ...)
  - Audio-visual features (color histogram, contour shape, tamura texture, parametric motion, ...)
  - Feature similarity (Euclidean distance, L1 distance, ...)
- Not all media representations relevant for all concepts or concept relationships
  - Color feature of concept Jazz ?
  - Audio representation of concept Sky?

### MediaNet Implementation

- MediaNet Construction: Semi-automatic construction
  - Text annotations, WordNet, image network of examples, automatic extraction tools
- Intelligent CBIR using MediaNet:
  - Refine, expand, and translate queries across modalities

### Multimodal Query Translation and Expansion



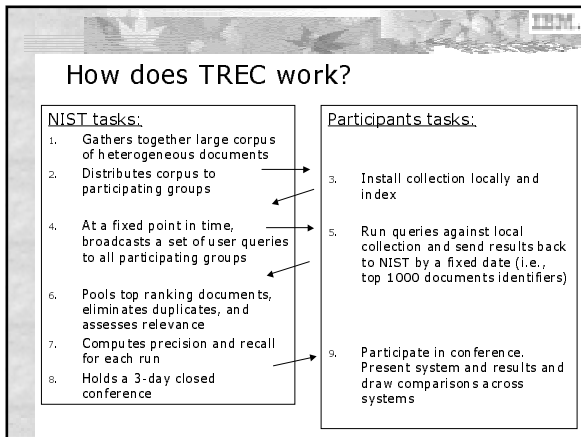
# Multimedia Retrieval Benchmarking

## What is TREC?

- The NIST coordinated Text Retrieval Conference (TREC)
- Series of annual information retrieval benchmarking exercises on large text collections
- Spawned in 1992 from the DARPA TIPSTER information retrieval project
- TREC has become important forum for evaluating and advancing the state-of-the-art in information retrieval
- Tracks for tracks for spoken document retrieval, cross language retrieval, and Web document retrieval

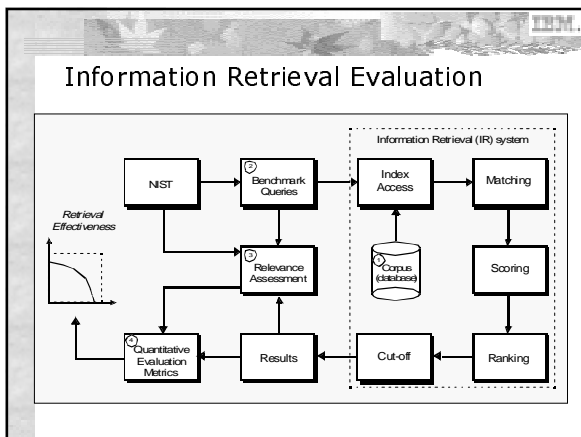
## TREC: Goals and Overview

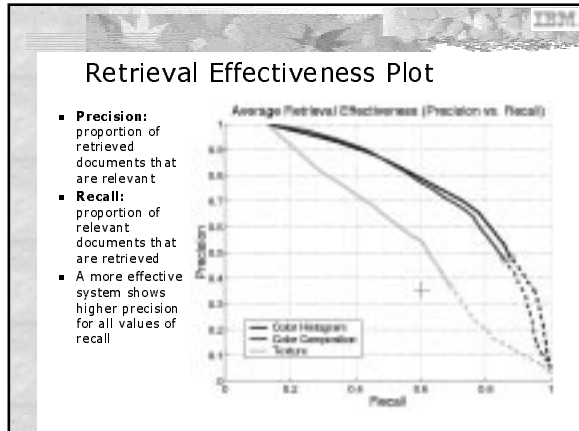
- Goals of TREC:
  - Increase research in IR on large-scale test collections
  - Increase communication among academia, industry and gov't through open forum
  - Increase technology transfer between research and products
  - Provide a state-of-the-art showcase of retrieval methods for TREC sponsors
  - Improve evaluation techniques
- What makes TREC notable?
  - Document collections are huge compared to previous ones
  - Groups participating represent who's who of IR research
    - 10-12 commercial companies (I.e., Excalibur, Lexis-Nexis, Xerox)
    - 20-30 university / research groups
    - 70% participation from US



## Anatomy of TREC test bed

1. Large corpus of documents (i.e., 2GB text)
2. Benchmark queries (topics):
  - Contributed from participants with some aggregation and weeding
3. Relevance assessments:
  - NIST does not exhaustively assess relevance of all documents to all topics
  - Pooling of results from across runs to assess relevance of top retrieved documents to each topic
  - Distorts evaluation of precision for recall > 0.4
4. Quantitative evaluation metrics:
  - Retrieval effectiveness (precision vs. recall, 3pt score, 11pt score)





- ### TREC's Missing Evaluation Criteria
- Efficiency
  - Usability
  - Interoperability
  - Goal satisfaction
  - Support for distributed collections
  - Integration with existing software / systems
  - Platform support

- ### Range of TREC Approaches
- Statistical and probabilistic term weightings
  - Passage or paragraph retrieval
  - Combining results of independent searches
  - Combining sub-document retrieval scores
  - NLP based and statistically based phrase indexing
  - Bayesian inference networks
  - Hyperlink based retrieval
  - Probabilistic indexing
  - Boolean and soft Boolean retrieval
  - Use of syntactic structures
  - Query reduction and query expansion
  - Concept recognition
  - Indexing by word pairs
  - Dictionary-based stemming
  - N-gram based indexing and retrieval

## TREC Video Retrieval Benchmark

- ### TREC Video Retrieval
- Planning of TREC video retrieval track started in 1999
  - **Coordinators:**
    - Paul Over, NIST
    - Alan Smeaton, Dublin City University
  - **Goal:**
    - Promote progress in Content-based retrieval (CBR) from digital video via open, metrics-based evaluation
  - **Tasks:**
    - Identify shot boundaries (automatic)
    - Given statement of information need, return ranked list of shots which best satisfy the need
    - Interactive and fully automatic approaches are allowed

- ### Motivation for Initial Tasks
- **Shot boundary detection:**
    - Needed for higher-level tasks
    - Easier entry due to existing base of work and software
  - **Known item search:**
    - Reflects significant type of user need ("I know it is there somewhere")
    - Lower evaluation costs, Human assessors not needed since "answers" identified by topic author
  - **General statements of information need:**
    - Most diverse, toughest for systems, costliest to evaluate
    - Ultimately, most important for real users

### Example Types of Needs

- I'm interested in video material / information about:
  - A specific person (e.g., Ronald Reagan)
  - One or more instances of a category of people (e.g., men wearing hardhats)
  - A specific thing (e.g., Hoover Dam or OGO satellite)
  - One or more instances of a category of things (e.g., helicopters)
  - A specific event / activity (e.g., Reagan's speech about space shuttle)
  - One or more instances of a category of events / activities (e.g., rockets taking off)

### TREC Video Retrieval Topics

- Participants contribute topics (5 topics per group minimum)
- **Topics include:**
  - Description of information need in text (input to systems and guide for human assessment of relevance)
  - Optional examples of image, audio, video content expressing user's need
- **Topics should be:**
  - In the realm of *doable* for current systems
  - *Realistic* in intent and expression
  - Should NOT be satisfiable using SDR/IR alone

### TREC Video Evaluations

- **Shot boundary detection:**
  - Pooled boundaries detected by systems
  - Human adjudication of major conflicts
  - Human enhancement for good boundary type mix
- **Known item search**
  - Automatic comparison to reference
- **General statements of information need:**
  - Human assessment per shot of whether the shot meets the need or not

### TREC Video Retrieval Evaluation Timetable

Date	Milestone
Jan 1, 2001	TREC video retrieval plan posted for comment
Feb 1, 2001	Groups intending to participate send application to NIST
Mar 1, 2001	Participating groups post list estimating number and types of topics they will contribute
April 1, 2001	Participating groups submit planned test topics to NIST (NIST will pool them)
May 1, 2001	Remaining details of guidelines completed, including schedule for distribution of test topics and for evaluation

### TREC Video Corpus

- 15 – 20 hours of MPEG-1 video
- NIST Digital Video Collection, Vol. 1
  - NIST educational and promotional video
  - Subjects include: overview of NIST programs, report on technologies for detecting aircraft hangar fires,
- Open Video Project Videos (UNC)
  - Subjects include: California wetlands, canyons, space travel, science lectures
- BBC stockshot video material
  - Raw, un-produced video

### Example Video Content

- *The Colorado*
  - Documentary about Colorado River narrated by Charleton Heston
- *Challenge at Glen Canyon*
  - Speech, music, scene noise, sound effects
- *To Build A Dream*

**Example Video Content**

- *Wetlands Regained*
  - *Documentary about recovery of California wetlands*
- *New Horizon*
  - *Documentary about Dams in Western United States*
- *Giant on the Bighorn*

**Example Video Content**

- *NASA 25th-Anniversary-Show*
- *Segment 5*
- *Segment 6*
- *Segment 9*
- *Segment 10*

**Example Video Content**

- *Spaceworks*
- *Episode 5*
- *Episode 6*
- *Episode 8*

**Example Video Content**

- *Sense and Sensitivity*
  - *Lecture 4*
  - *Classroom lecture about senses and perception*
  - *Instructor footage, graphics and illustrations, speech*

**Example Video Content**

- *Sense and Sensitivity*
  - *Lecture 3*
  - *Instructor footage, graphics and illustrations, speech*


**Example Video Content**

- *NIST: Aircraft Hangar Fires Fire Protection Improvements*
- *NIST: You Don't Have To Be There Telepresence Microscopy*
- *NIST: A Uniquely Rewarding Experience*

### TREC Video Retrieval Topic Breakdown

	Known Item	General Search
Interactive	1	3
Automatic	12	12
Interactive or automatic	24	6

### Example Query Topics



- Find video of the Statue of Liberty which can provide an unambiguous answer to the question "how many spikes are on the head of the Statue of Liberty?"
- Above example image is provide

### Example TREC Video Query Topics

Topic	Examples	# Known Items	Processing
Find shots of an astronaut diving a lunar rover across the surface of the moon with full view of the lunar rover	2 video segments	5	Interactive & automatic
Find shots of Harry Hertz, Director of the National Quality Program, NIST	2 images	3	Interactive & automatic
Find other examples of rocket and shuttle launches	7 video + 7 audio segments	N/A	Interactive
Find all shots should be extracted which contain monologues. Monologues are all shots containing a single person in the image facing the camera who is speaking to the viewer or an audience. Voice over while a person is in view does not count as a monologue.	2 video + 2 audio	N/A	Automatic
Other clips during the lecture showing and explaining the example graphic	9 video segments	N/A	Automatic

### TREC Video Topic Categories

Category	#	Examples
Space (outer space, space travel, rocket launch)	13	Mars, Jupiter, Rocket launch, Moon
Transportation vehicle (airplane, boat, helicopter)	15	Cars, Ski-boats, Moon-rover, Plane taking off
People in some activity	21	people being interviewed, people talking, people skiing
Wildlife (animals)	2	Deer, Birds
Manmade objects (dams, statues, cityscapes, forts)	7	White fort, Hoover dam
Activity/Events (not involving humans directly)	8	blasting a hill, explosion, fire
Lectures, interviews, meetings, monologues, testimonials	4	interview, speech
Particular personalities or objects, i.e. proper nouns	22	Statue of Liberty, Hoover dam, Reagan speaking, John Deere tractor
Low-level visual effects	4	zoom, pan, slow fading shots
Miscellaneous	3	Environmental degradation, water planning

### Topic Video Media Support

Media Support	Number
Image only	20
Audio only	1
Video only	34
Video + audio	7
Image + audio	2
Video + image	6
Not known	2

- ### Multiple TREC Video Retrieval Approaches
- Content-to-content matching:**
    - Generic feature matching, no training
    - Audio and visual feature extraction
    - MPEG-7 Descriptors
  - Semantic modeling and classification:**
    - Joint audio-visual analysis
    - MediaNet semantic net knowledge base
    - Graph-structured classification
    - MPEG-7 Description Schemes
  - Speech + feature + semantics:**
    - Automatic topic detection
    - Learning through linguistic association

### Summary

- **Content-based Retrieval from Multimedia Databases:**
  - Growing amounts of non-structured digital image, video, audio, and multimedia data
  - Need for multimedia content management (storage, access, querying, and retrieval of media objects and metadata)
- **Emerging Solutions:**
  - MPEG-7 multimedia content description standard
  - Content-based retrieval (CBR) and similarity search
  - Complex search - multiple objects, multiple features, fuzzy relationships
  - Multimedia knowledge representations
  - Multimedia retrieval benchmarking

# Universal Multimedia Access

### MPEG-7 Universal Multimedia Access

**Content Adaptation for Universal Access**

audio, images, video, data

hand-helds, set top boxes, smart phones, watch pads, wearables

Growing mismatch resource requirements vs. device capabilities

Increasing amounts of rich data (structured, non-structured, multimedia)

**MPEG-7 - ISO standard for multimedia content description meta-data**

Growing diversity of pervasive computing devices

### MPEG-7 UMA Approaches

① Variations: Fidelity, Variation

② Transcoding Hints

③ Summaries:  $f_0, f_1, f_2$

④ Views

### MPEG-7 Universal Multimedia Access

Server: MPEG-7 Selection, Content Servers, MPEG-7 Filter/Big

Proxy: Transcoding, Compositing, Summarization, Personalization

Client: Browsing, Visualization


- On-line processing at server:
  - Variation selection (optimized, in pyramid)
- On-line processing at server (MPEG-7):
  - Automatic processing (Variations)
  - Human annotation (Transcoding Hints)
  - Semi-automatic feature extraction (Semantics and Summarization)
- On-line processing at proxy (MPEG-7):
  - Video transcoding (Transcoding Hints)
  - Video summarization and personalization

# Links, References, and Demos

References (MPEG and IBM Research)

- MPEG Web site
  - <http://www.cselt.it/mpeg>
- MPEG-7.com Web site
  - <http://www.mpeg-7.com>
- MPEG-7 special issues:
  - *IEEE Trans. CSVT*, 2001 (To appear)
  - *Signal Processing: Image Communication*, Aug. 2000
- IBM Research MPEG-7 alphaWorks demos and downloads:
  - <http://pmedia.l2.ibm.com:8000/mpeg7>
  - <http://www.alphaworks.ibm.com>
  - MPEG-7 Visual Annotation Tool
  - Image Transcoding Proxy
  - VideoZoom Progressive Video System
  - SFGraph High-Resolution Image Zooming

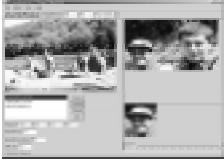
MPEG-7 Demo: Image and Video Content-based Retrieval (CBR)




- Similarity searching based on color, texture, color composition, and other audio-visual features
- Multimedia semantics modeling, classification, and search

MPEG-7 Universal Multimedia Access (UMA) Demos


① MPEG-7 Image Annotation Engine



② Video Transcoding



③ Semantic Video Transcoding



- MPEG-7 demos:
  - MPEG-7 Image Annotation Engine
  - Video Transcoding (formats and devices)
  - MPEG-7 Semantic Video Transcoding (personalized summaries, user preferences)