Conceptual Modeling for Multimedia Applications and the Role of MPEG-7

Uma Srinivasan, CSIRO Mathematical and Information Sciences
John R. Smith, IBM T. J. Watson Research Center

ACM Conceptual Modeling 2000
(ER-2000)

Outline

- Introduction:
  - Challenges in multimedia systems
  - Application requirements and multimedia content modeling
  - Related work
- Multimedia Conceptual Modeling:
  - Multimedia description concepts
  - Multimedia Extended-Entity Relationship (MM-EER) model
  - MPEG-7 Multimedia Content Description Standard
- Case 1: Design of Multimedia Content-Description Databases
- Case 2: Design of Multimedia Content-Description Software
Challenges in Multimedia Systems

- Multimedia applications require storage, access, delivery, 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

Multimedia Application Requirements

- 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 Digital TV
  - Requirements: Personalization, filtering, summarization
- Content Modeling:
  - High level semantics - objects, events, people, places
  - Low-level features - perceptual attributes
Content Modeling Approaches

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

Recent Modeling Approaches

- **Four-layer model** (Gutpa, 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 CM AHG** (Smith, Srinivasan, Rising, et. al, '99 - )
  - Identification and modeling of 192 principal content-description concepts from multimedia domain (MPEG-7 Multimedia Content Description Standard)
Content-based Multimedia Model

- Enables content-based similarity search
- Example queries: query-by image, query-by-sketch, query-by humming
- Content analysis and feature extraction:
  - Automatic extraction of descriptors of low-level features – colour, texture, shape, ...
- Similarity search driven by visual descriptors

Similarity Matching and Searching

- 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, multi-dimension indexing, high-dimensional feature spaces, query filtering
Image Data 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), η-R (Gudivada, et al), SaFe - spatial and feature (Smith and Chang, '99)
- Object and relationship based indexing and matching

Video Data Models

- Analysis and Modeling of Video Structural
  - Video parsing:
    - Shot detection: segmentation into temporal units (shots)
    - Key-frame selection: selection of salient frame(s) from each shot
    - Scene analysis: feature extraction from shots (Srinivasan et al. '97)
    - Multi-modal analysis: analysis of audio/video (Informedia, CMU, Srinivasan and Lindley, '99)
  - Model based on semantics
    - Identification of meaningful segments (Srinivasan et al.'99)
    - cohesive unit from a narrative point of view
Advanced Video Functions

- Visualization:
  - Mosaics – stitching of multiple frames, extraction of moving object regions
  - Scene-transition graphs – clustering of recurring shots (scenes), transition frequencies

- Navigation and browsing:
  - Video skimming of sequential summaries
  - Browsing of hierarchical summaries (TOC)
  - Key-frame navigation

Example Multimedia Query 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

- VIDEOSTAR
- FRANK
  - Text, video synchronisation
Multimedia Semantic Models

- Textual analysis:
  - Source: closed captioning, transcripts, embedded text, audio analysis speech detection
  - Content description – categorical, narrative structures
  - Synchronisation between text and video (Yap, Simpson-Young, Srinivasan, ‘97)

- Semantic relationships:
  - Generalization, association, aggregation

- Knowledge representation:
  - Frames, Semantic nets, Cyc knowledge-base

Multimedia Description Concepts
Conceptual Modeling of MM Objects

Integration: Semantic-level (objects, events), feature-level (color, texture), structure (regions, segments), relationships (spatio-temporal, semantic)

Concepts: model audio-visual content including attributes / features of actual physical objects

Formation of Visual Data via Projection of Real World Scenes

Real world scene (objects, events, ...) 

Multimedia data (regions, segments, ...)

projection
Multimedia Description Concepts

- **Semantics** - refers to the real-world entities that are depicted in or are partially captured by the multimedia data
  - Examples: objects, events, actions, real world scenes

- **Structure** - refers to how the multimedia data is composed from its constituent elements
  - Examples: regions, segments, spatio-temporal organization

- **Features** - refer to the low-level perceptual aspects of the multimedia data used for discrimination and classification
  - Examples: color, texture, shape, motion

- **Depictions** - refer to the projections of the real-world entities that form the multimedia data
  - Examples: camera acquisition and other types of recordings

- **Deductions** - refer to derived relationships which may be inferred at query time
  - Examples: near, left-of, similar-to, above

- **Ordering** - ordering (1-D) in the composition and aggregation of data entities
  - Example: temporal ordering - a video is composed of a sequence of shots in a particular order

- **Layout** - layout (2-D, 3-D) in the composition and aggregation of data entities
  - Example: spatial layout - an image is composed of regions in a particular layout
Multimedia-EER (MM-EER) Model

MM-EER Model constructs:
- Entities, Attributes, Relationships
- Generalization, Aggregation, Association
- Structural, Intensional Relationships, Qualifications
**Multimedia-EER (MM-EER) Model**

- **Entity** - an entity in the multimedia domain that can be distinctively identified based on a study of the principal objects in the domain
  - Examples: include structural entities such as **region**, **shot**, **frame**, **video**, **segment**, **image**, and semantic entities such as **object**, **event**, **action**
  - A class can be used to describe the set of entities that have the same characteristics

**Multimedia-EER (MM-EER) Model**

- **Attribute** - information about an entity or relationship that has descriptive properties
  - There are two types of attributes:
    - **Identifiers** - which uniquely identify entities
    - **Descriptors** - which describe entities.
  - Examples: from the multimedia domain include perceptual feature attributes such as **color**, **texture**, **shape**, **motion**
Multimedia-EER (MM-EER) Model

<table>
<thead>
<tr>
<th>Relationship</th>
<th>an association among one or more entities, as follows:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalization</td>
<td>is-a</td>
</tr>
<tr>
<td>Aggregation</td>
<td>has-a</td>
</tr>
<tr>
<td>Association</td>
<td>(semantic, structural)</td>
</tr>
<tr>
<td>* Structural relationship</td>
<td>(spatio-temporal)</td>
</tr>
<tr>
<td>* Intensional relationship</td>
<td>(derived)</td>
</tr>
<tr>
<td>* Qualified relationship</td>
<td>(ordering, layout)</td>
</tr>
</tbody>
</table>

Generalization - a relationship that partitions the entity class into mutually exclusive subclasses, also called “kind-of” or “is-a” or “specialization” relationship

Example: a frame is an image

Aggregation - is an assembly-component relationship, also called a “part-of” or “has-a” relationship

Example: a body semantic object is part-of a person semantic object
Example EER Model of Audio-Visual Content

Enhanced Semantics of MM-EER Model
Spatio-Temporal Relationships

- Spatio-temporal association (STA)
- Spatio-temporal encapsulation / aggregation (STE)
- Spatio-temporal composition (STC)

Example: in the multimedia domain a region is-near another region is modeled by:
- A spatio-temporal association (STA) in which the region is related to the other region via spatial association
Spatio-Temporal Relationships

- **Spatio-temporal aggregation (STE)** - a relationship along the spatio-temporal dimension in which one entity encapsulates or aggregates other entities.
  - Example: in the multimedia domain a key-frame belongs to a shot is modeled by:
    - A temporal aggregation relationship in which the shot temporally encapsulates the key-frame.

- **Spatio-temporal composition (STC)** - a relationship along the spatio-temporal dimension in which one entity is strictly composed of other entities.
  - Example: in the multimedia domain a video is composed of shots is modeled by:
    - A temporal composition relationship in which the video is temporally composed of a sequence of shots.
Intensional Relationships

- **Intensional relationship** - corresponds to a relationship among entities that is not stored but is inferred from attribute values, such as:
  - Spatio-temporal intensional relationships
  - Propagated intensional relationships
  - Perceptual intensional relationships

\[ \text{sem}(O_i, O_j) \]

\[ \text{struct}(R_i, R_j) \]

- **Spatio-temporal intensional relationship** - a relationship based on spatio-temporal topology or geometry
  - Example: a region near another region is not necessarily stored explicitly, but can be derived or inferred from the spatial attributes of the two regions
Perceptual Intensional Relationship

- **Perceptual intensional relationship** - a relationship based on similarity among attributes or entities.
  - Example: consider a certain “grainy” texture that is somewhat similar to a “coarse” texture
  - The degree of similarity is not explicitly stored, but is derived from the attributes of the two textures and a similarity function

Intensional Relationships

- **Propagated intensional relationship** - a relationship propagated to entities based on intensional relationships among attributes or aggregates of the entities
  - Example: in the multimedia domain consider two images that have similar color attributes
  - That similarity can be propagated, perhaps with consideration of the weighted combination of other features such as texture and shape, to deem the images to be similar
Qualifications of Relationships

- **Qualifications** - adds information about the many end of the relationship, as follows:
  - Cardinality
  - Existence
  - Sequence
  - Layout

- **Cardinality** - indicates the number of entities of one class that are related to entities of the other class
  - Characterized by degree of connectivity (one to one, one to many, many to many).
  - Example: in the multimedia domain a video contains one or more shots

- **Existence** - indicates the degree of participation in relationships, such as by allowing optional participation
  - Example: in the multimedia domain a shot may optionally contain a key-frame
Qualifications of Relationships - Layout

- **Depictions:**
  - Image depiction of scene
  - Regions depiction of objects

- **Layout:**
  - Scene layout of objects
  - Image layout of regions

---

Qualifications of Relationships

- **Sequence** - indicates the ordering of entities or attributes participating in a relationship
  - Example: in the multimedia domain a video contains a sequence of shots

- **Layout** - indicates the spatio-temporal arrangement of entities or attributes participating in a relationship
  - Example: in the multimedia domain an scene consists of a spatio-temporal layout of objects
MPEG-7 Multimedia Content Description Standard

MPEG-7 Multimedia Content Searching / Filtering

metadata
author
producer
copyright
semantics
objects
events
people
structure
regions
segments
features
colors
textures
shapes

Audio-Visual Data Searching and Filtering
MPEG-7 Standard

What is MPEG-7 about?
- Specification of a “Multimedia Content Description Interface”
- Enabling fast and efficient searching and filtering of audio-visual material
- Describing audio-visual content at a number of levels (features, structure, semantics, models, collections, immutable metadata)
- Satisfying content description requirement of a large number of audio-visual applications

What does MPEG-7 Promise?
- A standardized interface for **Description** and **Annotation** of multimedia content:
  - Association of MPEG-7 metadata with audio-visual data (carried in audio-visual content headers, embedded in audio-visual streams, stored in multimedia databases)
  - MPEG-7 can be extended to define new or non-standardized Descriptors and Description Schemes
  - Does not standardize feature extraction methods
  - Does not standardize search engines, matching methods or query languages
Overview of MPEG Standards

- **MPEG-1**: Storage/retrieval of moving pictures/audio on storage media (Nov. 92)
- **MPEG-2**: Digital television (Nov. 94)
- **MPEG-4**: Coding of natural and synthetic media objects for multimedia applications (Oct. 98), Ver. 2 (Dec. 99), Ver. 3 & 4 (Dec. 2000)
- **MPEG-7**: Multimedia content description for searching and filtering (Oct. 2001)
- **MPEG-21**: Digital audiovisual framework - integration of technologies for content identification and consumption (late 2001)
MPEG-7 Scope

Feature Extraction: Content analysis, Feature extraction, Annotation tools, Authoring

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

Search Engine: Searching and filtering, Classification, Manipulation, Summarization Indexing

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

Example: MPEG-7 Applications

- Image and video search engines
- Digital libraries and scientific image databases
- Universal multimedia access

Audio-visual archive

Query engine
Access
Index
DATA
META-DATA

Search
Filter
WWW

Playback

Basic Constructs of MPEG-7

- **Description Definition Language (DDL)**
  - Standardized, yet flexible language to specify Description Schemes (DS) and Descriptors (D)
  - It is derived from eXtensible Markup Language (XML) Schema
  - It expresses relations, object orientation, composition, partial instantiation,

- **Description schemes (DS)**
  - Define structure and semantics of relationships of DSs and Descriptors (Ds)

- **Descriptors (D)**
  - Describe features of audio-visual data

---

MPEG-7 Description Definition Language (DDL)

- **MPEG-7 Description Definition Language (DDL):**
  - Standardized, yet flexible language to specify Description Schemes (DS) and Descriptors (D)
  - It is derived from eXtensible Markup Language (XML) Schema
  - It expresses relations, object orientation, composition, partial instantiation,

- **MPEG-7 Systems:**
  - Specifies means for binarizing DDL data
  - Specifies methodology for carrying descriptions as streams
  - Specifies means for accessing and synchronously consuming data
  - Specifies management and protection of data
MPEG-7 Multimedia Description Schemes (MDS) Overview

Datatype & structures  Basic elements  Link & media localization  Basic DSs

Content organization  Collection & Classification  Model

User interaction

Creation & production  Media  Content management  Usage

Navigation & Access

Summary  Variation

Structural aspects  Conceptual aspects

Basic DSs

Overview

MPEG-7 Meta-data for Content Description

Data  Signal structure  Features  Model  Semantics

Data
- Images
- Video
- Audio
- Multimedia
- Formats
- Layout

Structure
- Regions
- Segments
- Grids
- Mosaics
- Relationships (Spatio-temporal)

Features
- Color
- Texture
- Shape
- Motion
- Speech
- Timbre
- Melody

Models
- Clusters
- Classes
- Collections
- Probabilities
- Confidences

Semantics
- Objects
- Events
- Actions
- People
- Labels
- Relationships

MPEG-7 Meta-data for Content Management

1. Acquisition
2. Editing
3. Coding
4. Analysis
5. Archival
6. Publication

Scene attributes
Parameters
Correlation

MPEG-7 Visual Descriptors

- Grid layout (spatial structure)
- Color:
  - Color space, dominant color, color layout
- Texture:
  - Homogeneous texture, texture browsing
- Shape:
  - Contour-based shape
- Motion:
  - Motion activity
Example: MPEG-7 Moving Region Description

- Moving Regions:
  - Reference regions - bounding boxes
  - Features - color, texture, shape (boundary)
  - Spatio-temporal locators
    - Figure Trajectory
    - Parameter Trajectory

MPEG-7 Audio

  - Audio Description Framework
  - Spoken Content DS
  - Timbre DS
  - Audio Independent Components

- Core Experiments (ongoing):
  - Melody
  - Sound Effects
Example: MPEG-7 Spoken Content

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

![Diagram of word and phone lattices]

Design of MPEG-7 DSs and Ds

- **MPEG-7 Design Issues:**
  - How to characterize MPEG-7 design problem:
    - Database design, software design, domain modeling?
  - What is appropriate design methodology:
    - E-R or O-O modeling? Need extensions?
  - What is the requisite discipline?
  - What is the scope of MPEG-7 design work:
    - Conceptual, functional, logical, physical modeling tasks?
Conceptual Modeling in MPEG-7

- **Conceptual model**: defines logical model independent of implementation
- **MPEG-7 Description schemes (DS)**: implement specific description structures
- **MPEG-7 DDL**: DS definition language
- **MPEG-7 Descriptions**: instantiate the DSs

---

MPEG-7 Concept Properties

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal Concept</td>
<td>Concept name</td>
</tr>
<tr>
<td>Definition</td>
<td>Definition in words</td>
</tr>
<tr>
<td>Conceptual Model Construct</td>
<td>(Attribute, Entity, Function, Relationship, Type)</td>
</tr>
<tr>
<td>Domain</td>
<td>(Audio, Generic, Video)</td>
</tr>
<tr>
<td>Type</td>
<td>(Ancillary, Audio-visual data, Description, Feature, Language, Meta, Model, Process, Semantics, Structure, Syntax/ Semantics)</td>
</tr>
<tr>
<td>MPEG-7 DDL Construct</td>
<td>(D, DDL)</td>
</tr>
<tr>
<td>MPEG-7 DDL Construct</td>
<td>(Element, Attribute, Type)</td>
</tr>
<tr>
<td>Description</td>
<td>Description of development as D or DDL or Systems</td>
</tr>
<tr>
<td>Related principal concepts</td>
<td>Related principal concepts</td>
</tr>
<tr>
<td>Related secondary concepts and terms</td>
<td>Related secondary concepts and terms</td>
</tr>
</tbody>
</table>
## MPEG-7 Principal Concepts List

(MPEG-7 Requirements)

<table>
<thead>
<tr>
<th>Principal Concept</th>
<th>Definition</th>
<th>Mode of Conceptualization</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstraction Level</td>
<td>The particular level of detail at which data is represented</td>
<td>MPEG-7</td>
<td>Multiresolution Pyramid DS (XM) specifies a hierarchy of views of data. Summary DS (WD) is used to specify a set of summaries to enable rapid browsing, navigation, visualization, and sonification of audio-visual content. A summary is an audio-visual abstract of the content.</td>
</tr>
<tr>
<td>Acquisition</td>
<td>The process of acquiring audio or visual data from a source</td>
<td>MPEG-7</td>
<td>CreationMaterial (WD) describe the devices and elements used for the creation of the content (e.g., types of device, lens, instruments, settings, etc.).</td>
</tr>
<tr>
<td>Action</td>
<td>A semantically identifiable behavior of an object or group of objects, e.g., soccer player kicking ball</td>
<td>MPEG-7</td>
<td>Annotation DS (WD) contains the descriptive data (DCs) needed for a simple and structured description of persons, objects, events, etc. Semantic Relationships (XMs) describe the relations among persons, objects, events, or states among two or more elements. Examples of such relations are “is-located” and “is-touched”.</td>
</tr>
<tr>
<td>Aggregation</td>
<td>Grouping of items such as objects, regions, or audio-visual data</td>
<td>MPEG-7</td>
<td>Cluster DS ( WD) describe the arbitrary grouping of audio-visual data items or syntactic elements.</td>
</tr>
</tbody>
</table>

### MPEG-7 Application Scenario

- **Application Domain**
- **Conceptual Model**
- **Annotation Tool**
- **Description Generator**
- **MPEG-7 Description**
- **MPEG-7 Coded Description**
- **Decoder**
- **Search/Query**
- **Filtering Agent**
- **User/System**

Case 1: Multimedia Database Design

Conceptual Modeling of Audio-Visual Content

- **Basic Problems:**
  - Modeling of multimedia content description data
  - Design of corresponding database schema, software classes, and metadata schema

- **Proposed Solution:**
  - Multimedia-Extended Entity Relationship (MM-EER) model
  - Conceptualizes content description data requirements
  - Provides new model constructs for structural, depictional, intensional and perceptual relationships
  - Additional qualifications that denote ordering and layout

- **Applications** (automatic MM-EER transformation):
  - Logical database schema (i.e., relational databases)
  - Software classes (i.e., Java classes)
  - Content description meta-data (i.e., XML / MPEG-7 meta-data)
Objective: Modeling Methodology for Multimedia Content Description

- Single unified modeling framework
- **Modeling**: rich conceptual model built for multimedia content description data
- **Transformation**: conceptual model is transformed into the logical models for
  - Database design
  - Software classes for application implementation
  - Meta-data language for portable annotation and communication

MM-EER Relational Database Design Methodology (Transform Rules)

- Transform each MM-EER entity into a table
- Transform each isA relationship into an entity table with a foreign key of the parent entity
- Transform each depictional relationship into a table with foreign keys of the related entities
- Transform each ST-aggregation (STE) relationship into a table with foreign keys of the related entities
- Transform each ST-composition (STC) relationship into a table with foreign keys of the related entities and consider order and layout as follows:
  - For temporal composition add a table attribute that denotes order
  - For spatial composition add two table attributes to denote spatial position
  - For ST-composition add table attributes to denote spatio-temporal position
- Transform each ST-association (STA) relationship into a table with foreign keys of the related entities. Add a table attribute that denotes distance
- Transform each perceptual attribute (attribute) into an attribute of the entity relation. Add a table with foreign keys for pairs of the entities. Add an attribute that denotes similarity
Example Conceptual Model - Diagram

- STC(video, shot) - a video is temporally composed of a sequence of shots
- STC(shot, frame) - a shot is temporally composed of a sequence of frames
- isa(frame, image) - a frame is an image,
- isa(key-frame, frame) - a key-frame is a frame,
- STE(video, segment) - a video temporally aggregates segments
- STA(segment, segment) - a segment temporally aggregates segments
- STA(event, event) - an event temporally aggregates events
- STC(image, region) - an image spatially aggregates regions
- STA(region, region) - a region is spatially associated with regions
- STE(region, region) - a region spatially aggregates regions
- Depicts(region, object) - a region depicts an object
- STA(object, object) - an object is spatially associated with objects
- STC(object, object) - an object is spatially composed of and aggregates objects
MM-EER Transformation (Entities)

- Entity transformation (tables):
  - IMAGE (IMAGEID, COLOR, TEXTURE)
  - REGION (REGIONID, SHAPE, LOCATION)
  - FRAME (FRAMEID, IMAGEID)
  - KEYFRAME (KEYFRAMEID, FRAMEID)
  - VIDEO (VIDEOID)
  - SHOT (SHOTID, SEGMENTID)
  - SEGMENT (SEGMENTID)
  - OBJECT (OBJECTID)
  - EVENT (EVENTID)

MM-EER Transformation (Relationships)

- Depiction transformation (tables):
  - DEPICTION.SEGMENT.EVENT (DSEID, SEGMENTID, EVENTID)
  - DEPICTION.REGION.OBJECT (DROID, REGIONID, OBJECTID)

- ST-aggregation (STE) transformation (tables):
  - STE.VIDEO.SEGMENT (STEVSID, VIDEOID, SEGMENTID)
  - STE.SHOT.KEYFRAME (STESKID, SHOTID, KEYFRAMEID)
  - STE.IMAGE.REGION (STERRID, IMAGEID, REGIONID)
  - STE.REGION.REGION (STERRID, REGIONID, REGIONID)
  - STE.SEGMENT.SEGMENT (STESSID, SEGMENTID, SEGMENTID)
  - STE.OBJECT.OBJECT (STEEOID, OBJECTID, OBJECTID)
  - STE.EVENT.EVENT (STEEVID, EVENTID, EVENTID)
Case 2: Multimedia Object Software Design

UML-Based Modeling

- System and notation for modeling
- Design of software
- Model constructs:
  - Class
  - Abstraction
  - Aggregation
  - Data member
  - Methods
Multimedia Object Classes

- Multimedia Objects → software classes
  - Characterised using 3 distinct categories of attributes:
    - Content-bearing (colour, moving object, etc), created using some decoding procedures
    - Descriptive (metadata about the content/object)
    - Structural information - temporal

Multimedia Attributes

- Multimedia attributes → data members
  - Automatically extracted features:
    - Image - visual features
    - Video - motion, activity, object tracking
    - Audio - loudness, energy, music features
  - Spatio-temporal attributes:
    - Location, size, duration
  - Semantic attributes:
    - Manual annotation
  - Complex value domains
## Multimedia Relationship Types

- Multimedia Relationships → software classes (association, generalization), data members (aggregation)
- Complex semantic and structural relationships

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Relationship Type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brightness</td>
<td>Visual, based on light intensity</td>
<td>Brighter-than, Dimmer, same</td>
</tr>
<tr>
<td>Loudness</td>
<td>Audio, based on sound level</td>
<td>Louder, softer, same</td>
</tr>
<tr>
<td>Spatial-Position</td>
<td>Spatial, based on position</td>
<td>(Intersect, Contained-in, Disjoint, Above...)</td>
</tr>
<tr>
<td>Temporal</td>
<td>Temporal, based on intervals</td>
<td>(Before, After, Overlap, Meets...)</td>
</tr>
</tbody>
</table>

## Relationship Modeling

- Relationship examples:
  - Temporal relationships:
    - After, before, meets
  - Spatial relationships:
    - Intersection, above, disjoint
  - Audio relationships:
    - Loudness comparison
  - Visual relationships:
    - Color similarity
Summary

- Introduced challenges and applications of multimedia modeling
- Presented a Multimedia Extended-Entity Relationship (MM-EER) model
- MPEG-7 Multimedia Content Description Interface Standard
- Case 1: Design of Multimedia Databases
- Case 2: Design of Multimedia Software

Demo: MPEG-7 Visual Annotation Tool
IBM MPEG-7 Visual Annotation Tool

References

References

- A. Gupta, T. E. Weymouth and R. Jain, Semantic Queries with Pictures: the VIMSYS Model, Intern. Conf on Very Large Databases (VLDB), Sept. 3 - 6, 1991

MPEG-7 References

- MPEG-7 Principal Concept List (V0.9), ISO/IEC JTC1/SC29/WG11 MPEG2000/N3413, Geneva, CH, June 2000

© 2000
ACM Conceptual Modeling (ER-2000)
Software References

- IBM MPEG-7 Visual Annotation Tool