Video Content Description:
From Low-Level Features to Semantics

A. Murat Tekalp

University of Rochester

www.ece.rochester.edu/~tekalp

PhD Students

Ahmet Ekin, A. Mufit Ferman,

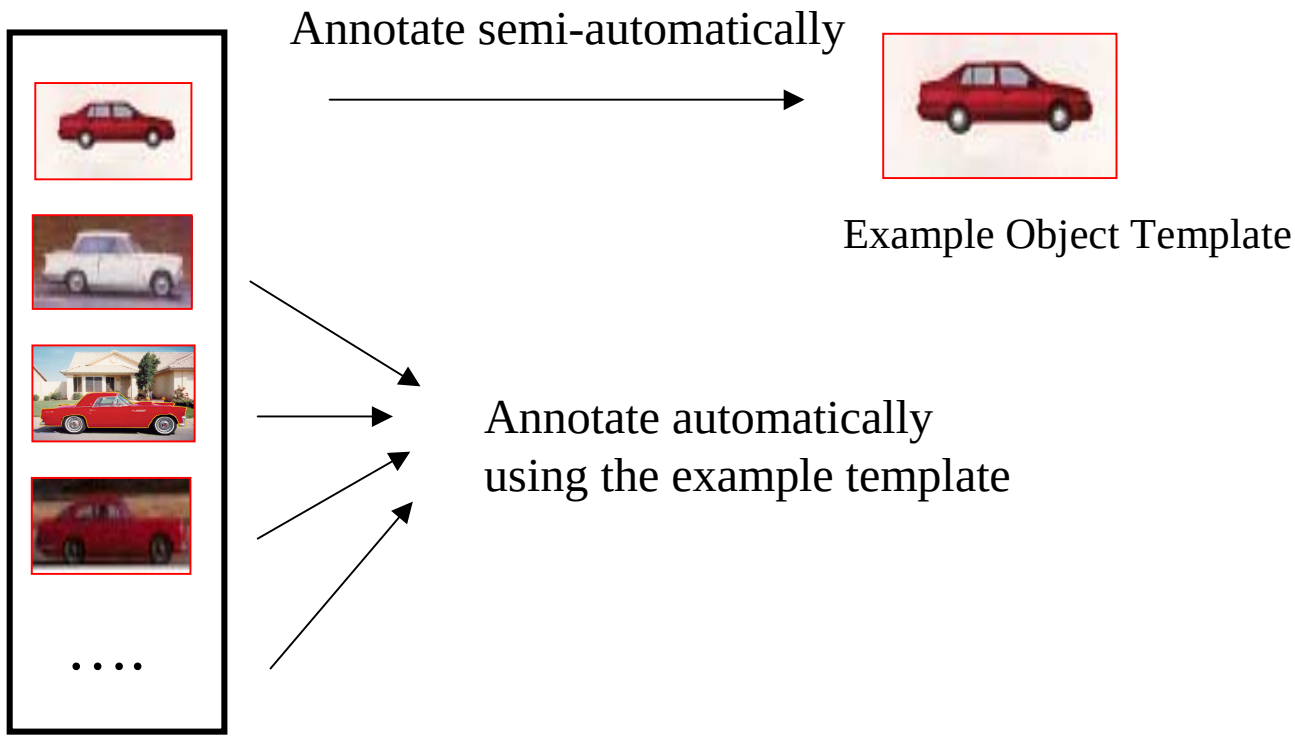
Yaowu Xu

Overview

- Learning from Examples
- Background: MPEG-7
- Low-Level and Semantic-Level Modeling of Object Motion
- An Integrated Semantic-Syntactic Video Model and Model-Based Query Processing
- Automatic Frame-Based Video Summarization: From Low-Level Features to Semantic-Level

Learning from Examples

Application: **Personal Image Library System**



1. "Learn" from examples to extract and annotate semantic objects automatically
2. Perform low-level feature, such as color and shape, between database images and example template

Formation of the Initial Content Hierarchy

Color Edge Detection

Color Segmentation

Low-Level Image Analysis

Split and Merge

Segmentation Map

Construct Content hierarchy

Scene graph

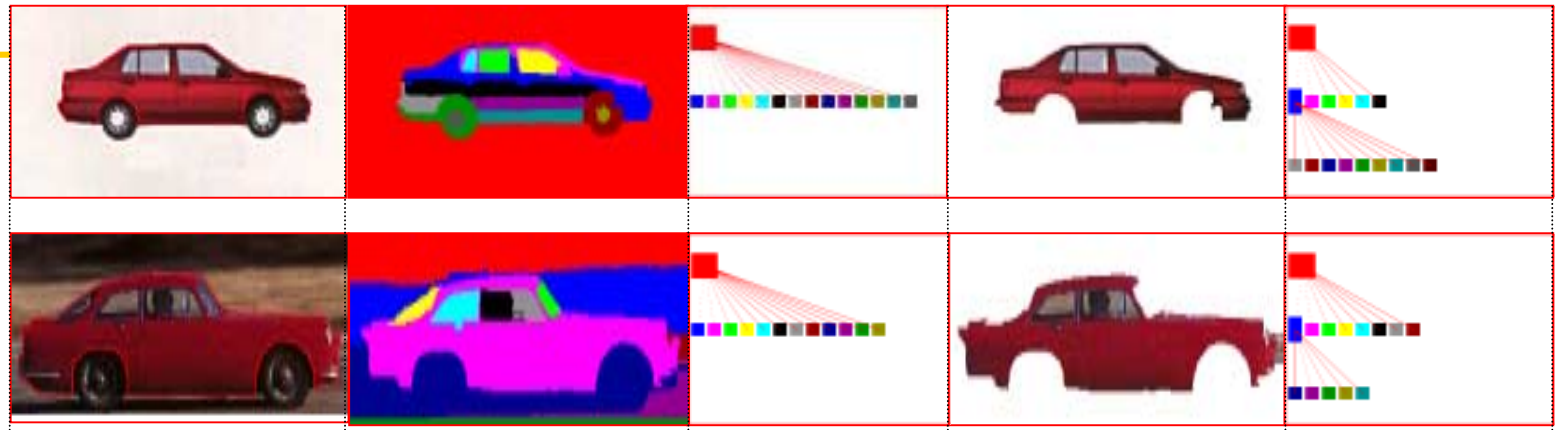
Generate Node Attributes

Node Attributes

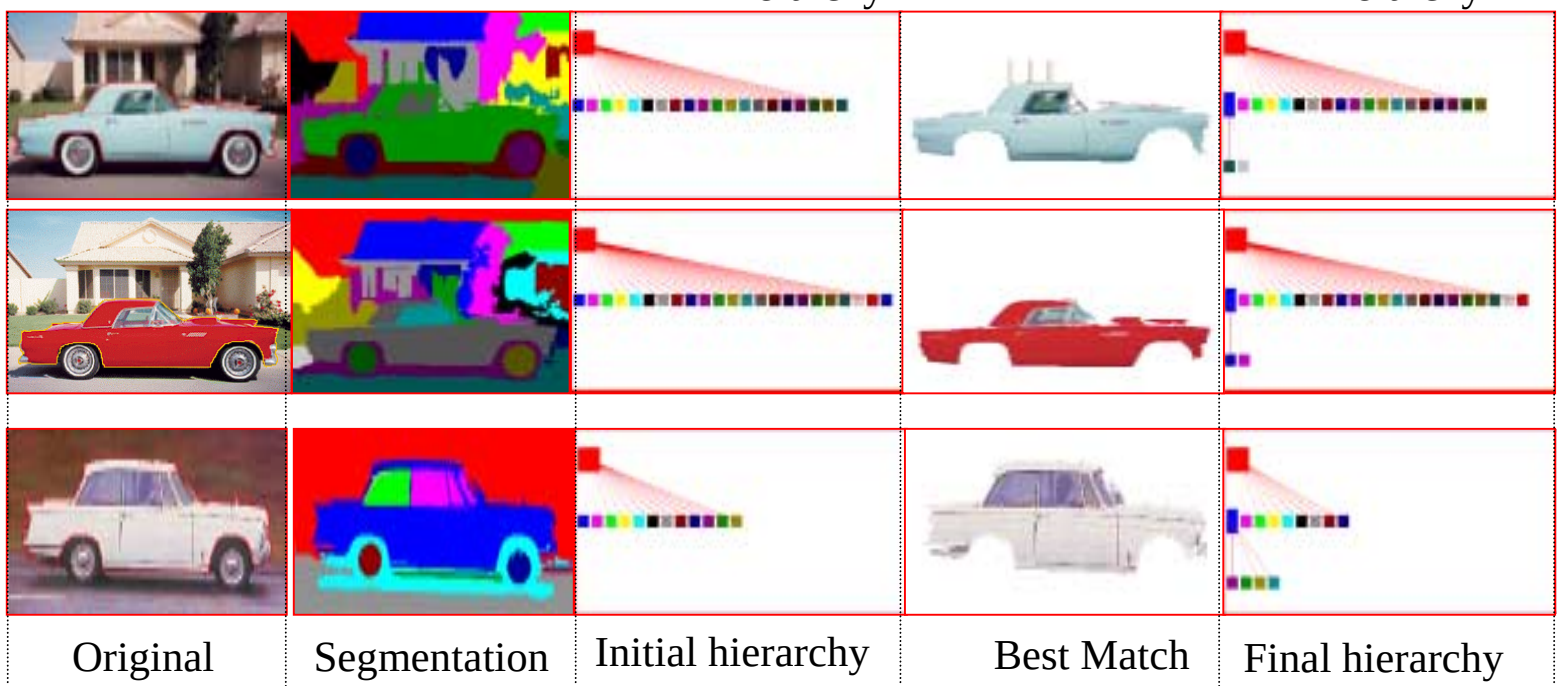
- Color Edge Detection
- Color Segmentation
- Region Formation by integration of multiple cues
  - Split regions containing edges into multiple regions
  - Merge regions using a highest confidence decision method
- Initialization of the Content Hierarchy

Examples - Shape Similarity Matching

Query template



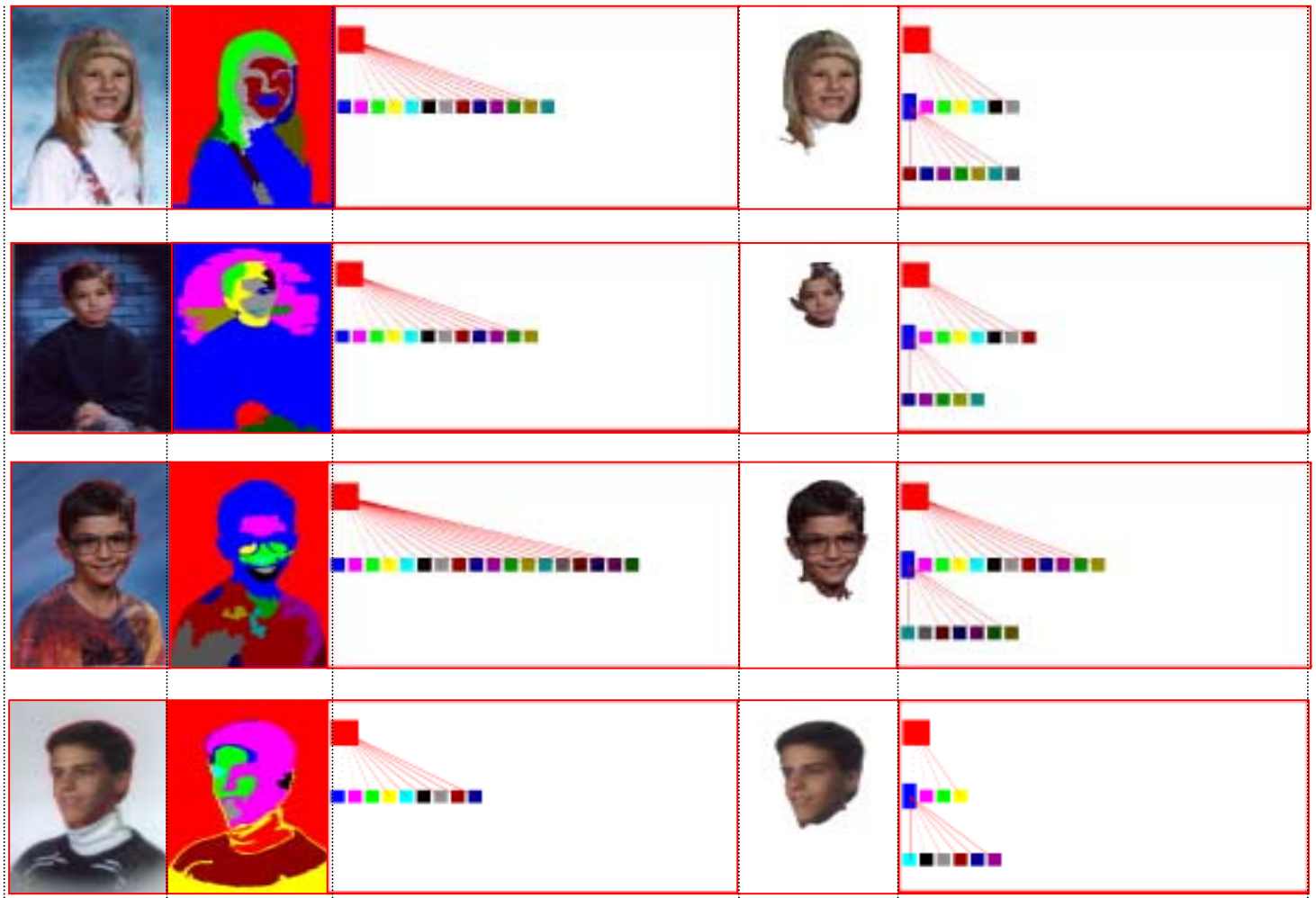
Original | Segmentation | Initial Hierarchy | Best Match | Final Hierarchy



Original | Segmentation | Initial hierarchy | Best Match | Final hierarchy

Examples - Color Similarity Matching

Query template



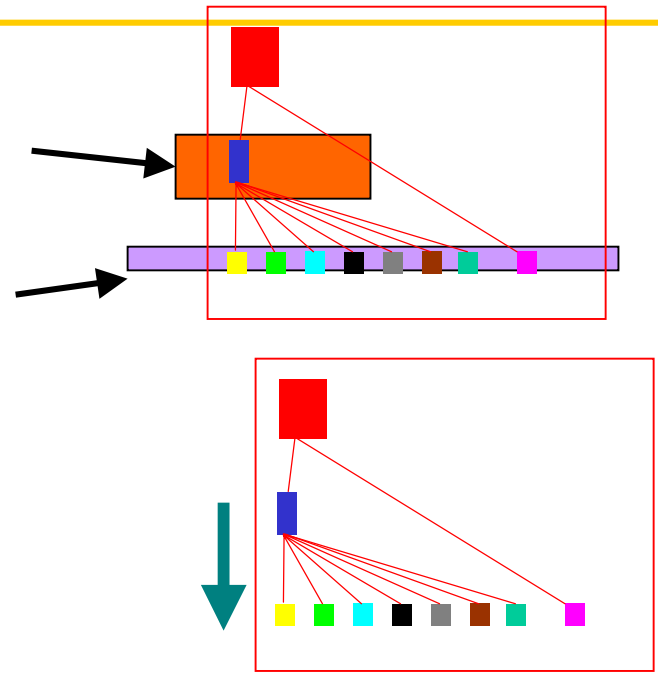
Original | Segmentation | Initial hierarchy | Best match | Final hierarchy

Hierarchical Content Matching

- Query Modes
  - High-level queries at the object level

    Example: Find “cars”
  - Low-level queries at the region level

    Example: Find a blue color region
- Matching Measure
  - Color histogram intersection
  - Hausdorff distance
- Hierarchical Content Matching
  - Top-down fashion
  - Highest-level (composite) nodes first.
  - No match in higher level, go to lower level



Semantic Segmentation

Images

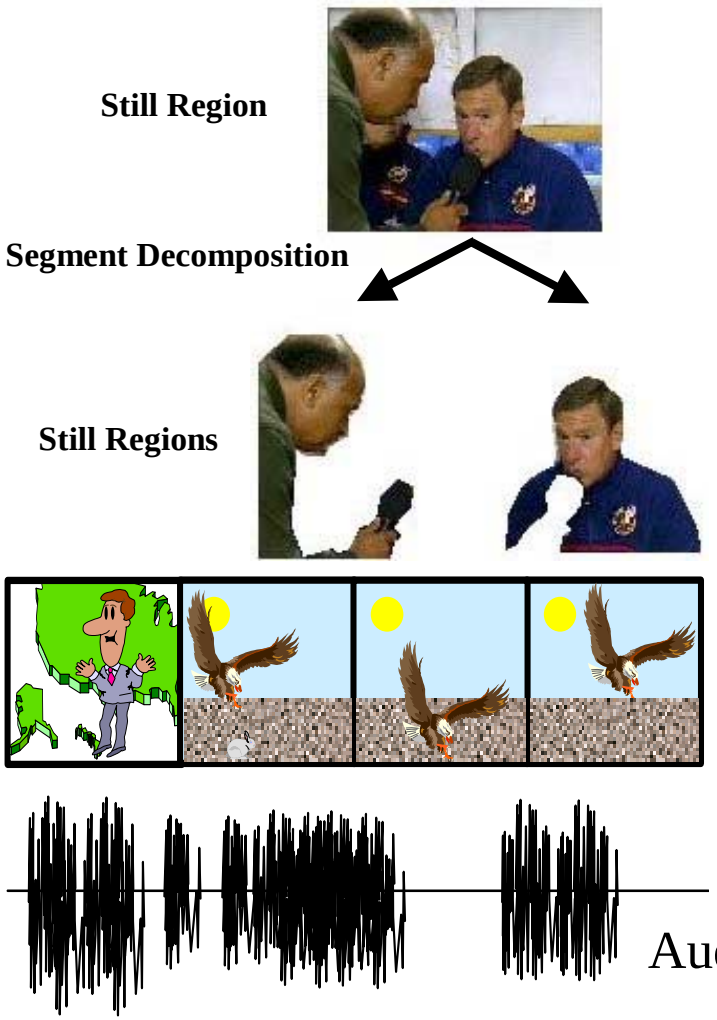
Segmentations

Low Level

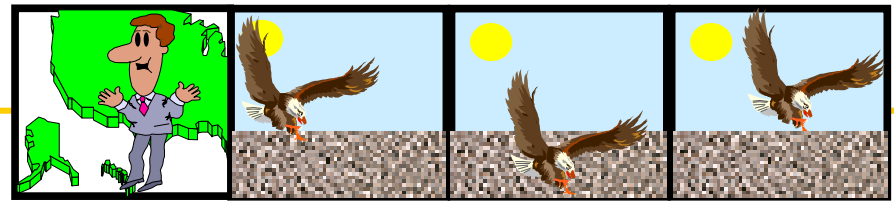
Semantic Level

Overview of Some Elements of MPEG-7

MPEG-7 Segment DS



Feature	*Video Segment*	*Still Region*	*Moving Region*	*Audio Segment*
Time	X	.	X	X
Shape	.	X	X	.
Color	X	X	X	.
Texture	.	X	.	.
Motion	X	.	X	.
Camera motion	X	.	.	.
Mosaic	X	.	.	.
Audio features	.	.	X	X

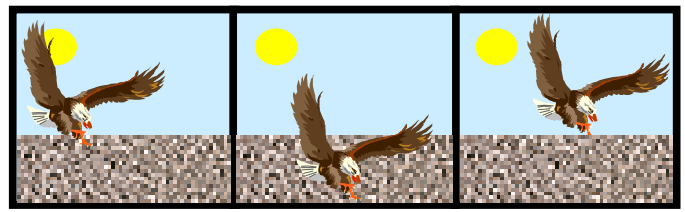


Video Segment

Segment Features

- Text Annotation
- Time
- Mosaic

SegmentDecomposition



Video Segments

Segment Relation

before

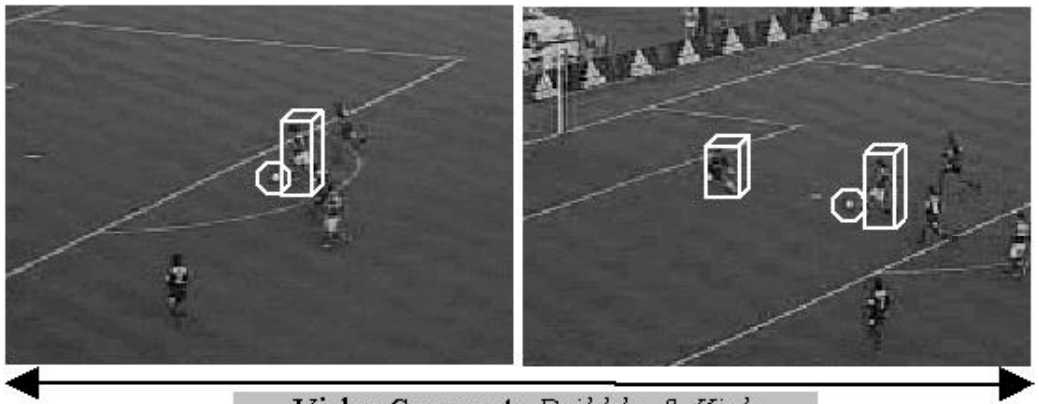
Relation	Inverse Relation
before	after
meets	metBy
overlaps	overlappedBy
during	contains
strictDuring	strictContains
starts	startedBy
finishes	finishedBy
equal	equal

SegmentDecomposition

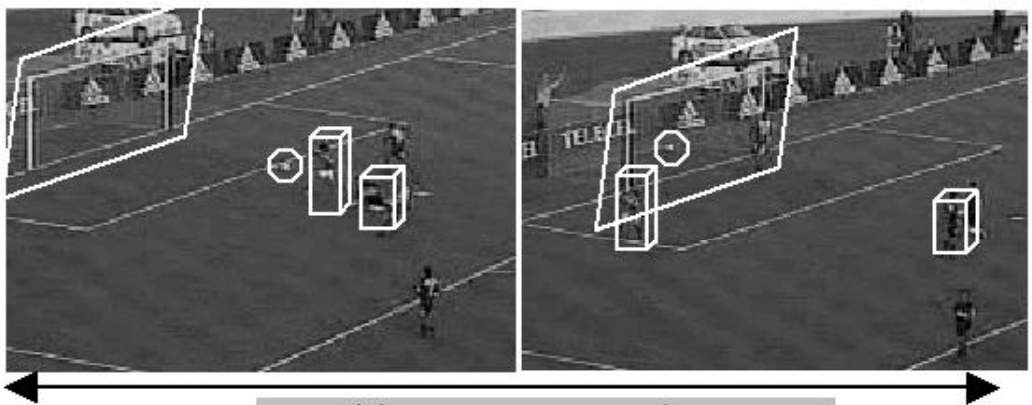
Moving Regions

SegmentFeatures

- TextAnnotation
- Time
- ParametricMotion
- GofGopColor



Video Segment: *Dribble & Kick*



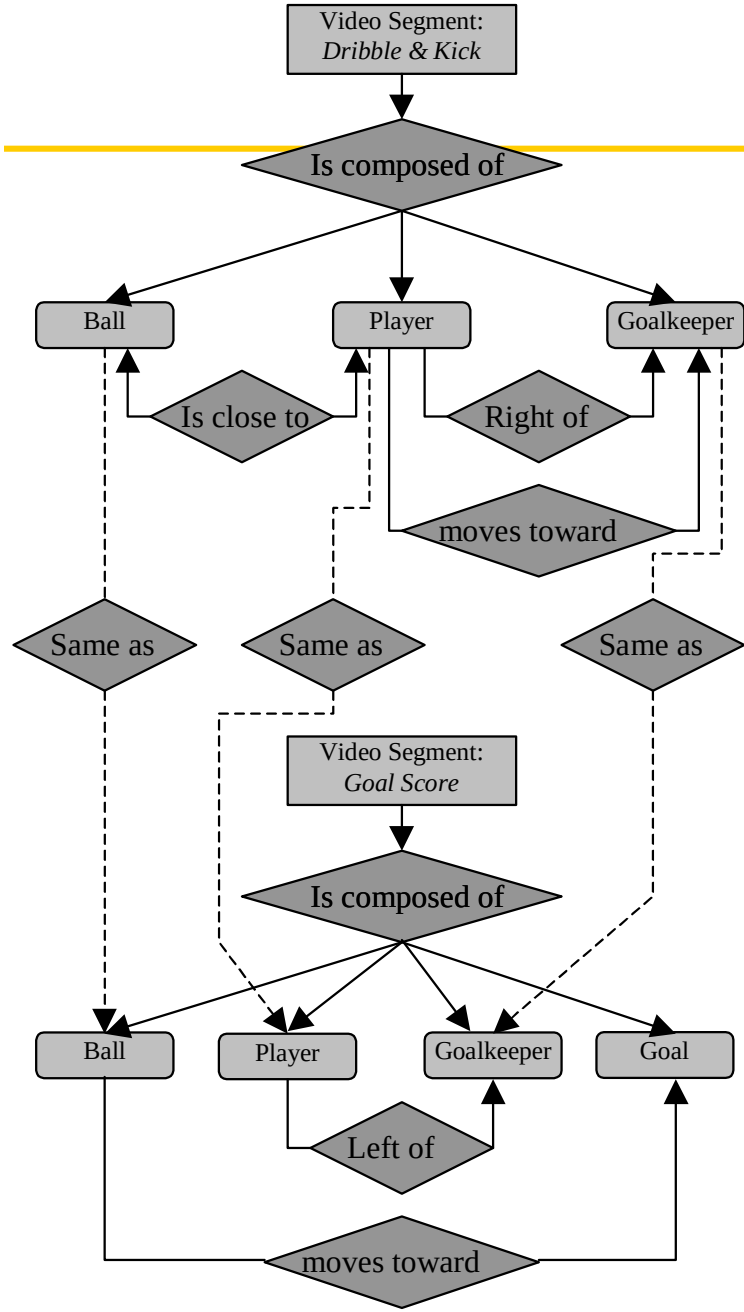
Video Segment 2: *Goal Score*

Moving Region: Player

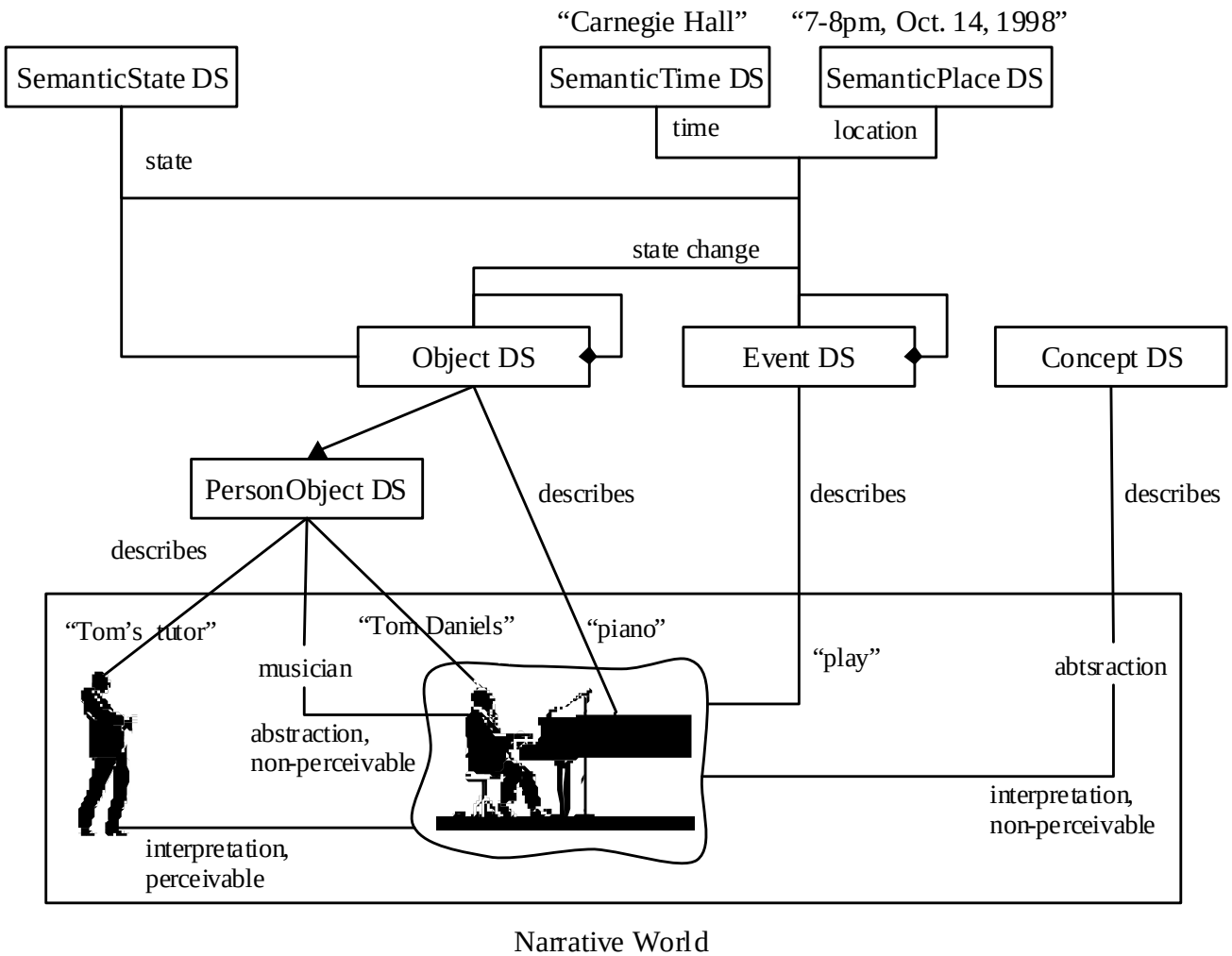
Moving Region: Ball

Moving Region: Goal Keeper

Still Region: Goal



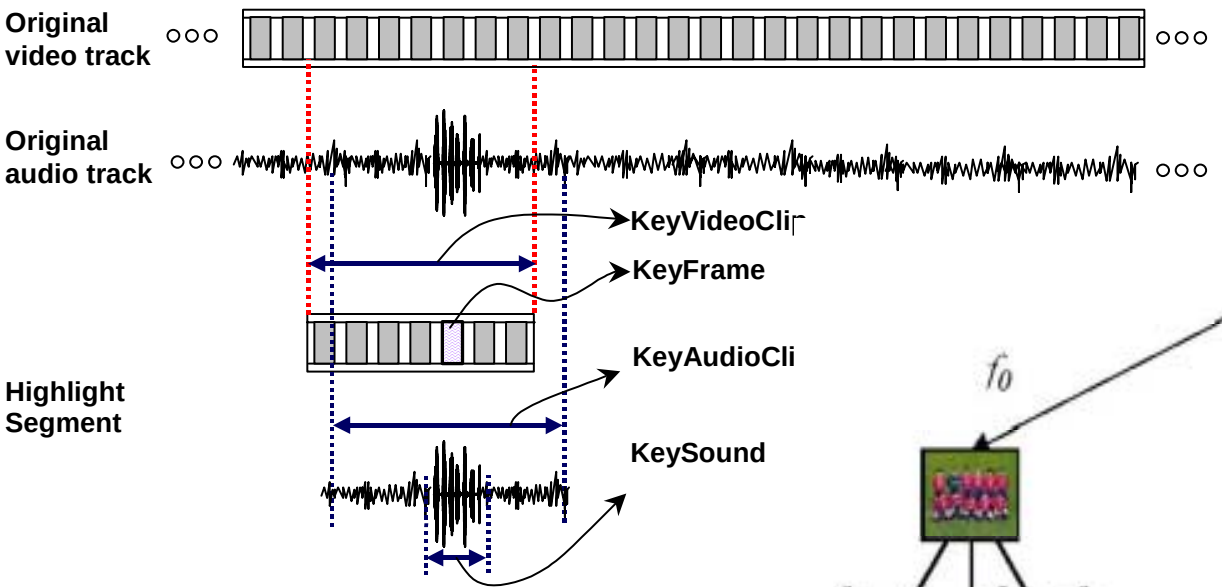
MPEG-7: Semantic DS

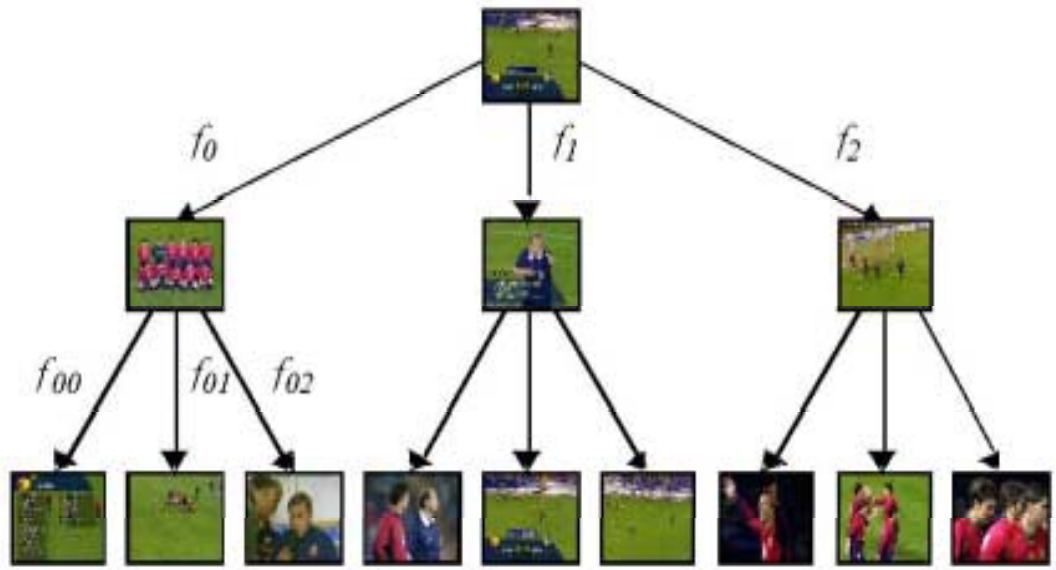


Semantic Relations:
- Object to object
- Object to event
- Event to event
- STime to event
- SPlace to event
- SBase to segment

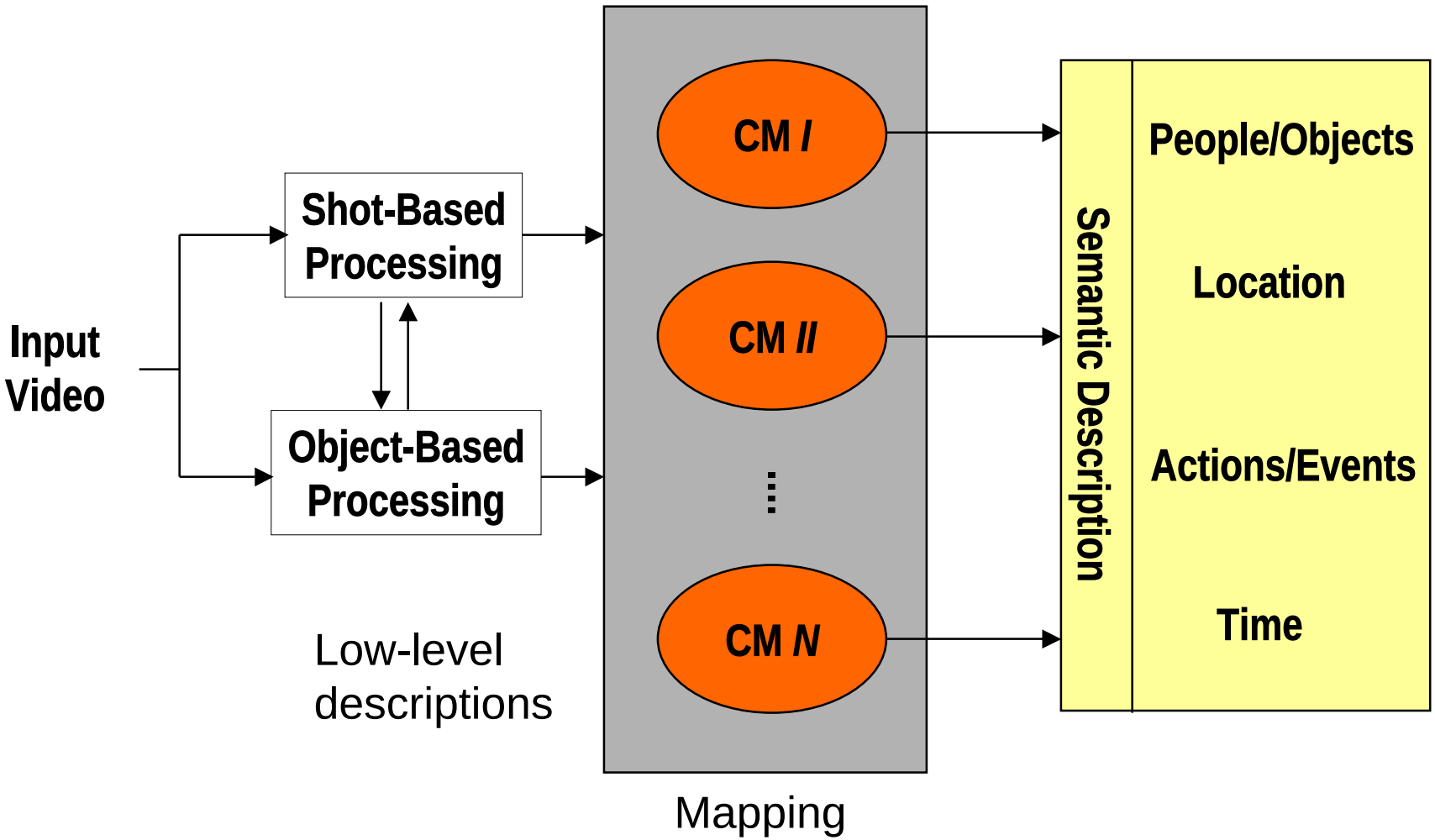
MPEG-7: Summarization DS

Key frames; Key video clips; Key audio clips; Key events; mixed.



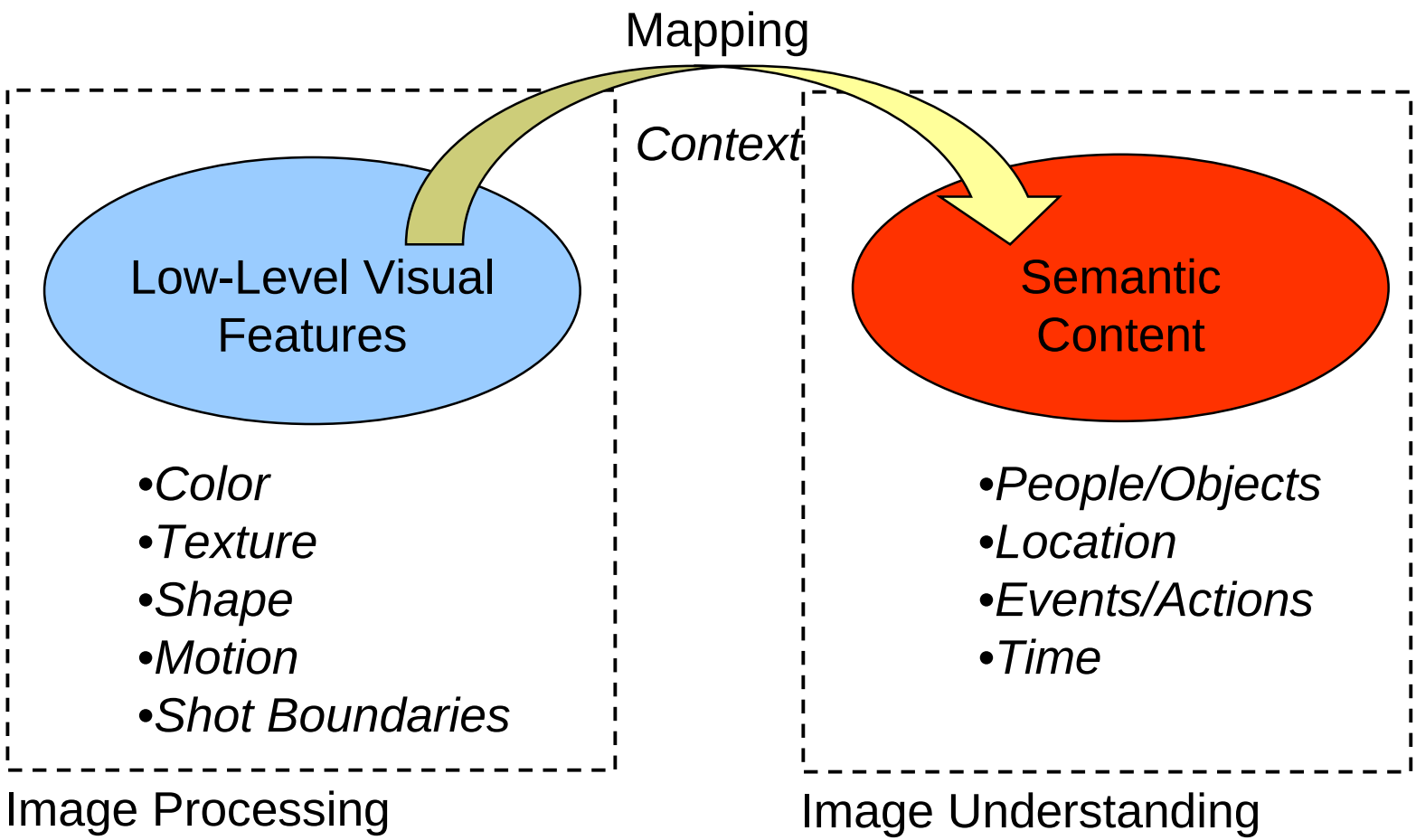


Video Analysis / Feature Extraction



Low-Level and Semantic-Level Modeling of Object Motion

From Low-Level to Semantic Level



Goal

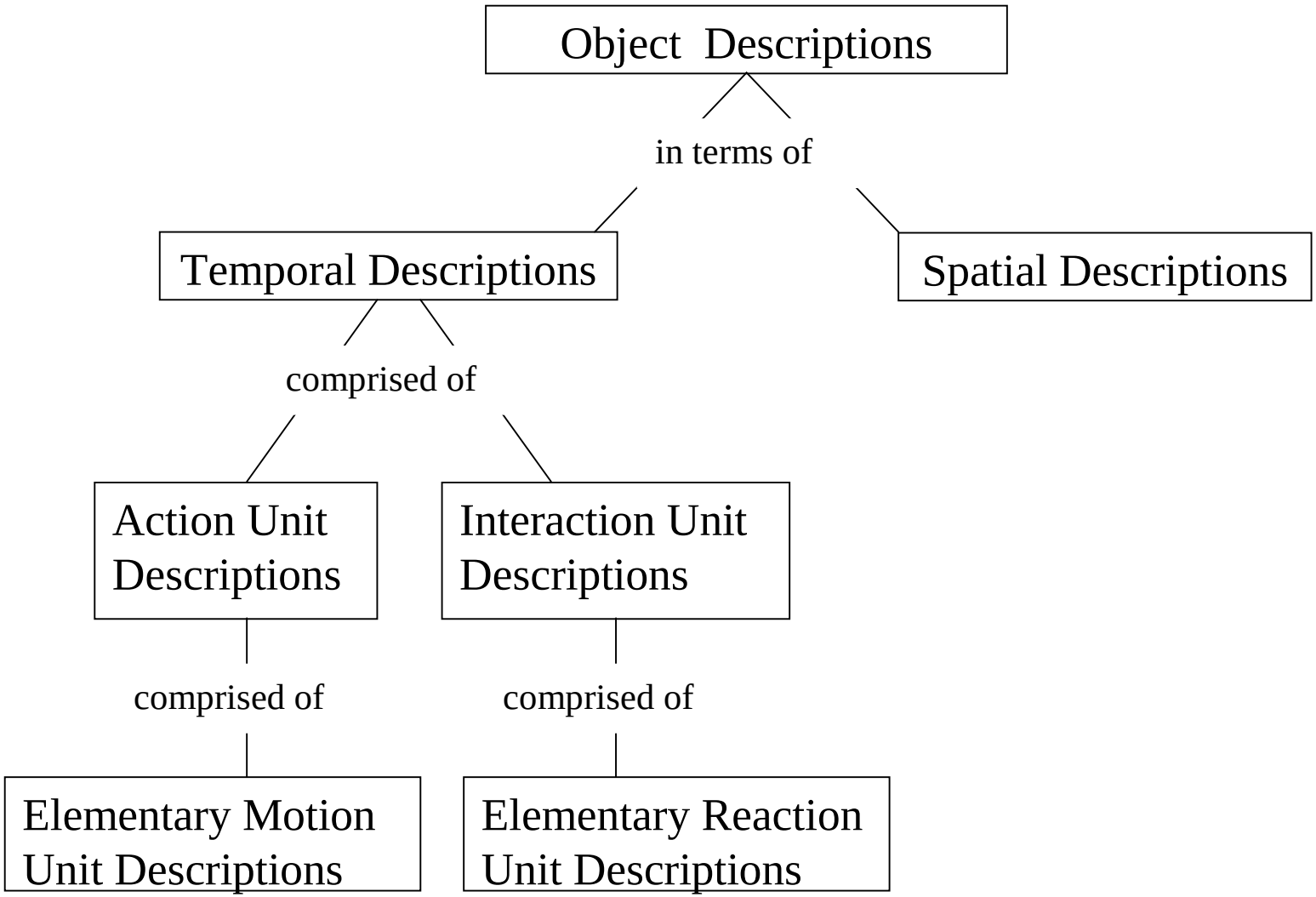
- Object-based motion description at the low level
  - Parametric motion (PM) - Dominant motion
  - Motion trajectory (MT)
  - Motion activity (MA)

- Object-based motion description at the semantic level
  - Actions
  - Interactions
  - Events

Problems

- The lifetime of a video object or even a shot is too coarse a temporal resolution to describe its motion both semantically and at the low level.
- We define segments which enable meaningful description of semantic and low-level motion of objects and interactions between them.
- We describe scene motion (events) by composing object actions and object-to-object interactions.

Object-Based Motion Description



Parametric Motion Descriptor

ModelCode	Meaning	Number of parameters
0	Translational model	2
1	Rotation/scale model	4
2	Affine model	6
3	Perspective model	8
4	Quadratic model	12

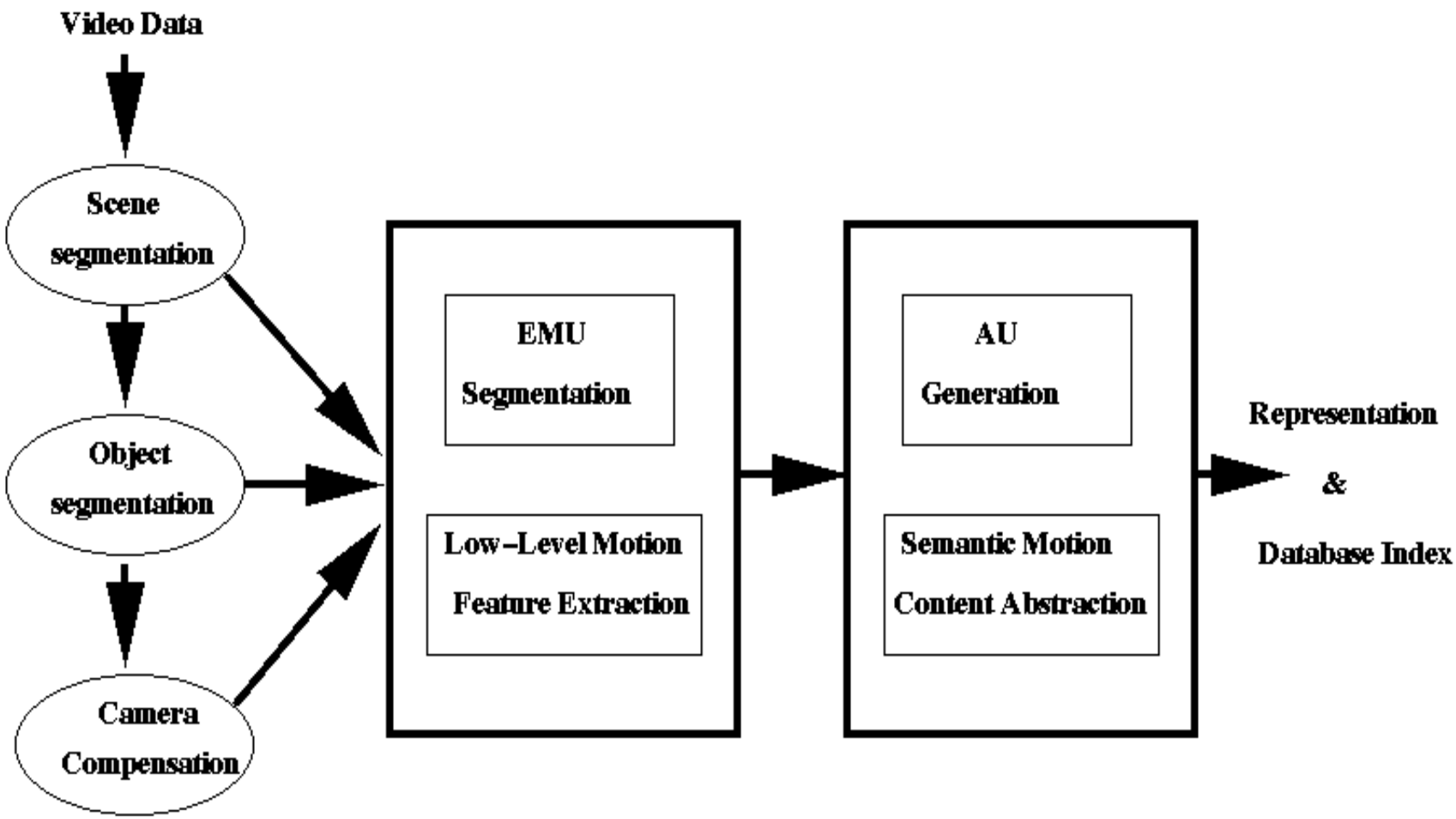
StartTime specifies the beginning of the temporal interval.
EndTime specifies the end of the temporal interval.

MotionParameters[] is a floating point array that keeps the values of the model parameters. Its size depends on the motion model specified by ModelCode.

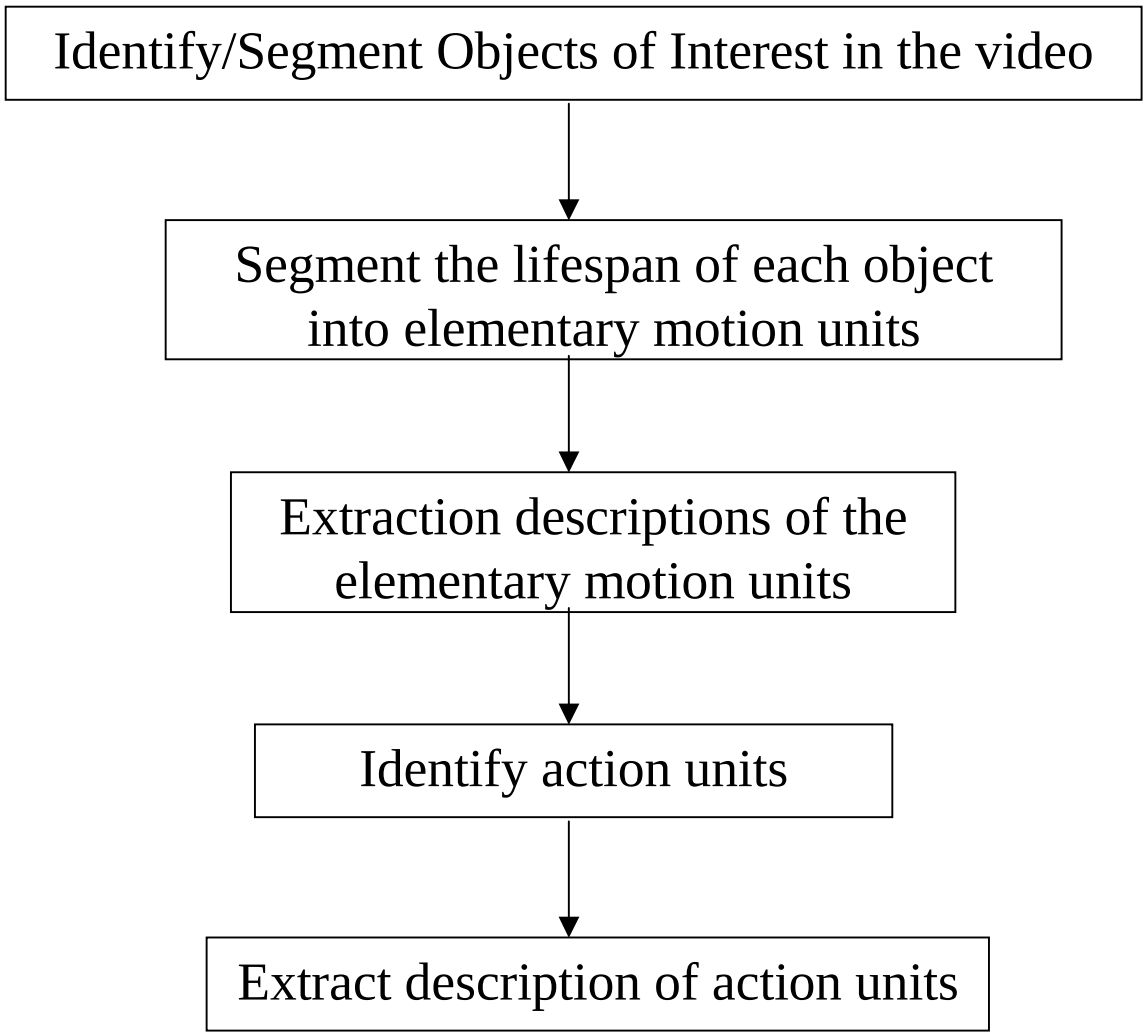
SpatialRegion a pointer to the spatial region the model is associated with.

Xorigin, Yorigin are the coordinates of the origin of the spatial reference with respect to the image coordinates.

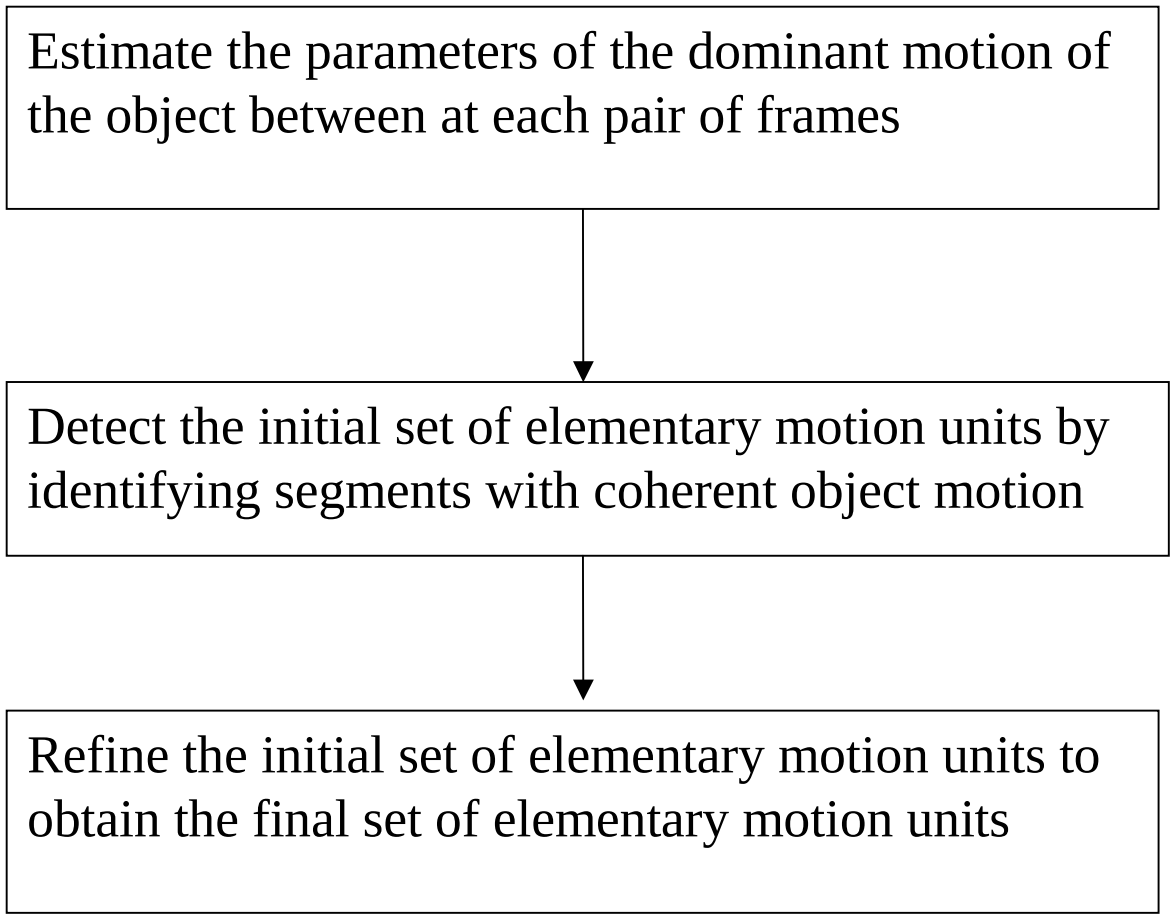
Extraction of Motion Descriptions



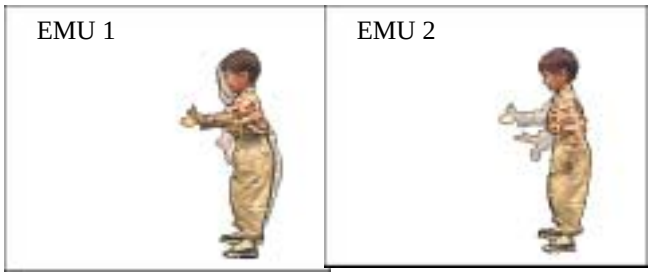
Procedure


Identify/Segment Objects of Interest in the video
Segment the lifespan of each object into elementary motion units
Extraction descriptions of the elementary motion units
Identify action units
Extract description of action units

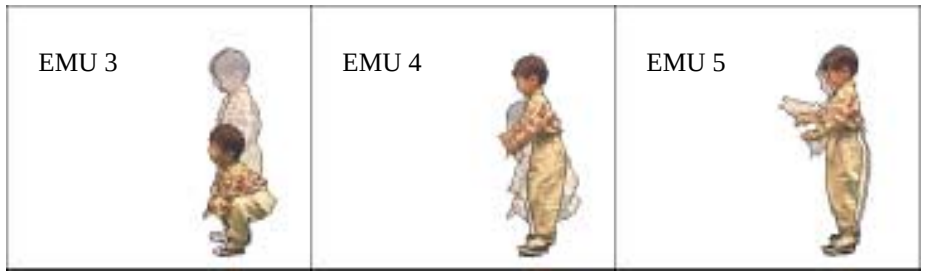
Detection of EMUs



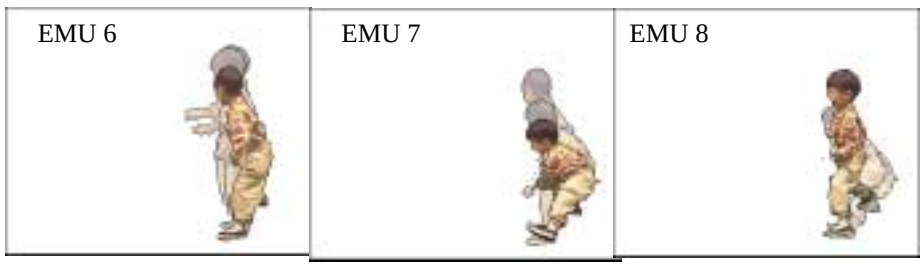
Example: Children Sequence



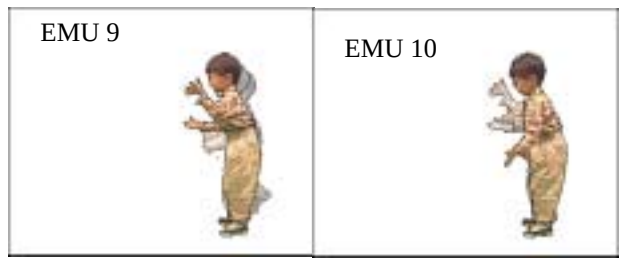
AU “throw the ball”



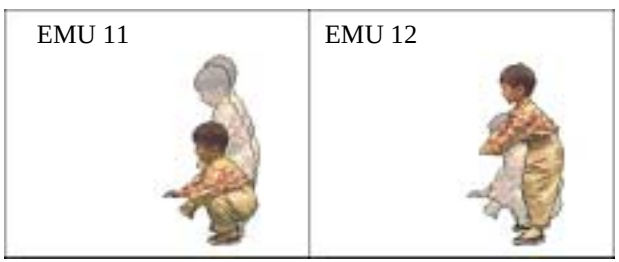
AU “pick up the ball and throw it



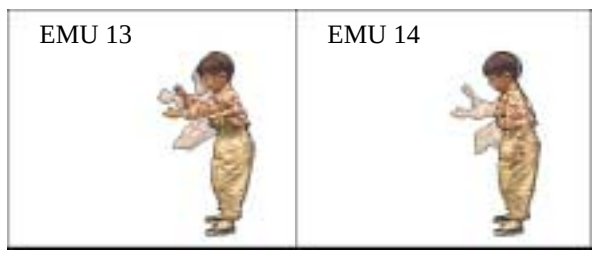
AU “pick up the ball”



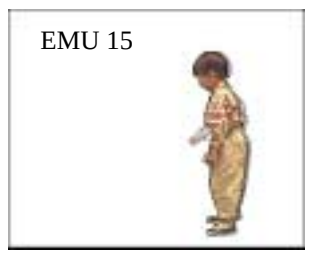
AU “throw the ball”



AU “pick up the ball”



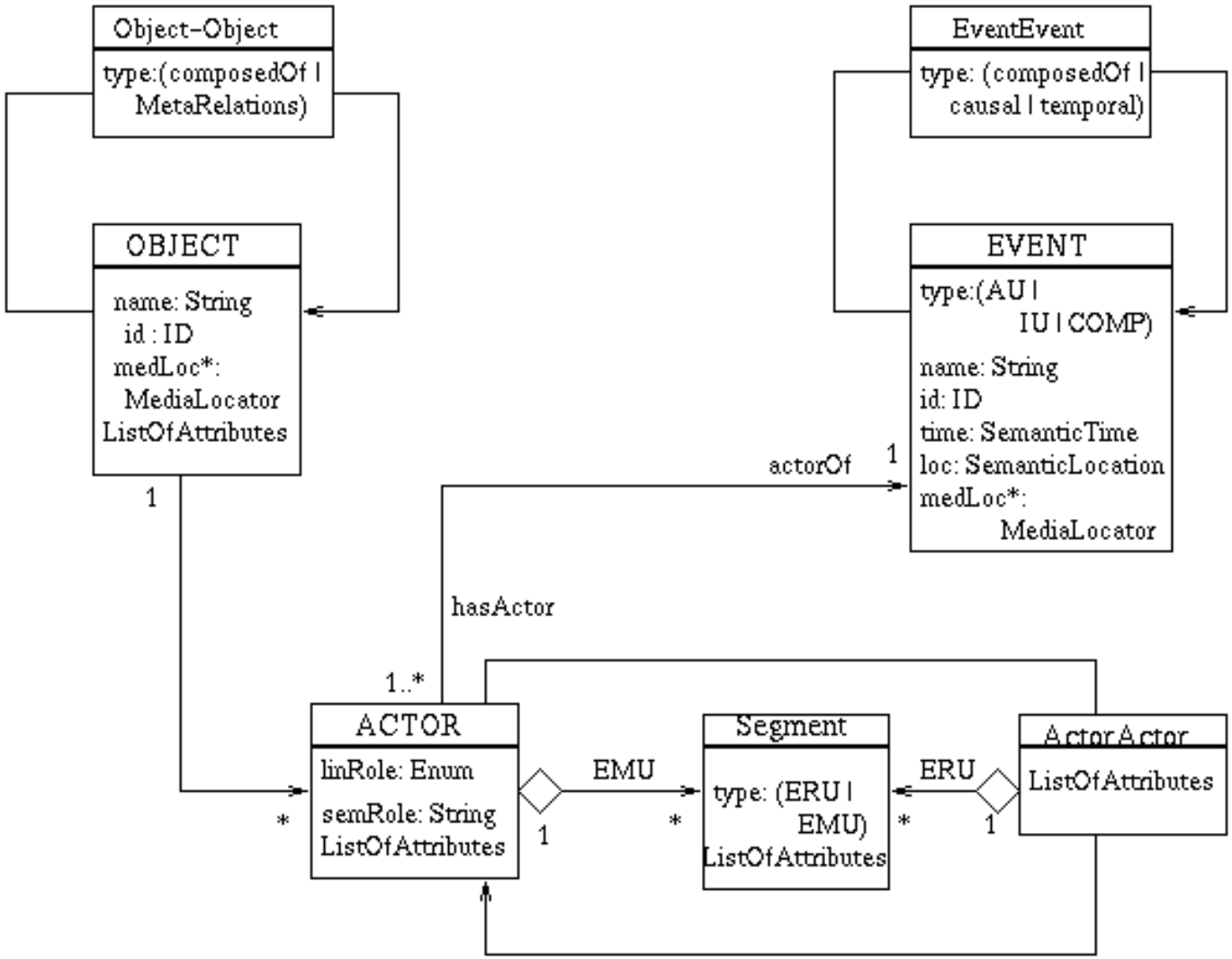
AU “throw the ball”



AU “stand”

An Integrated Semantic-Syntactic Video Description Model

Integrated Semantic-Syntactic Model



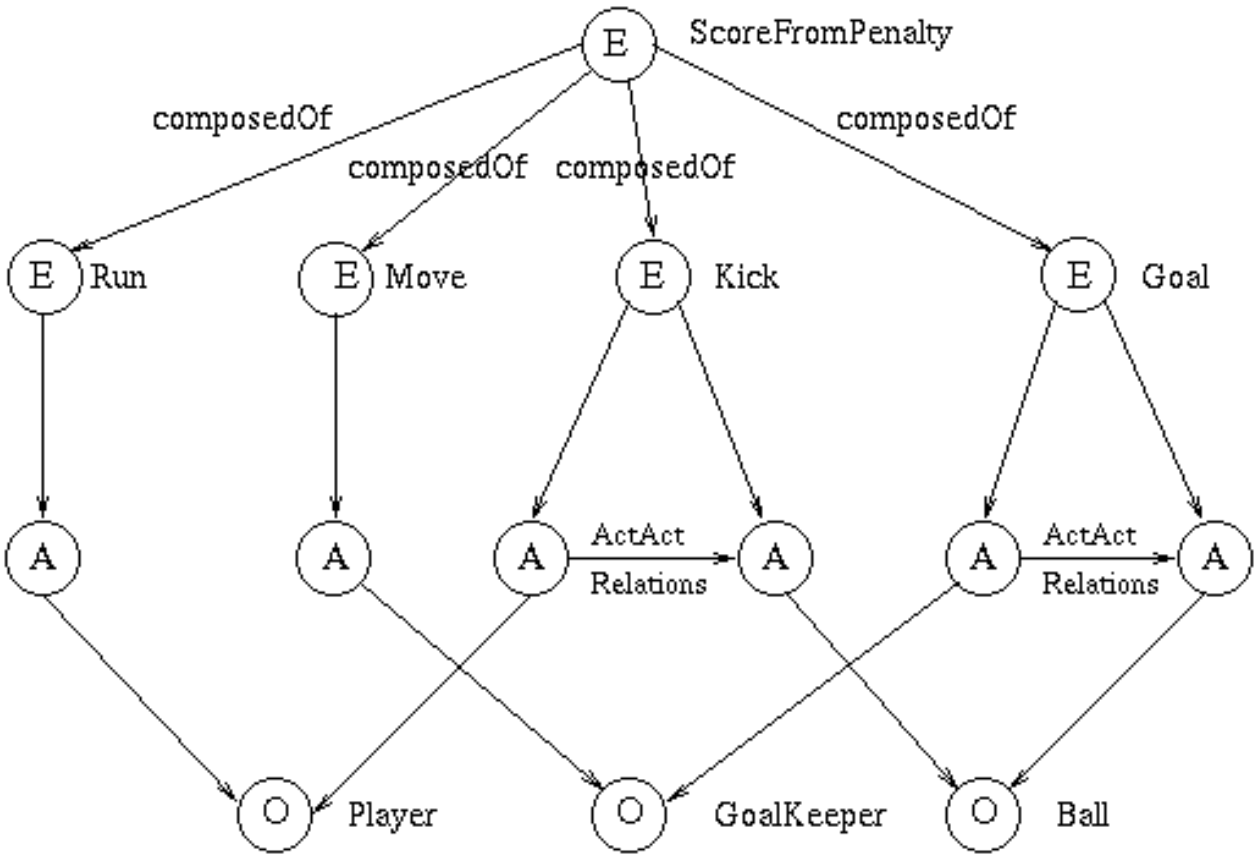
Contributions

- **Integrated semantic-syntactic model:** Enables efficient mixed-level query processing; "Find all penalty kicks shot to the left of the goalie"

- **Actor entity:** Enables incorporation of context in object-event relationships

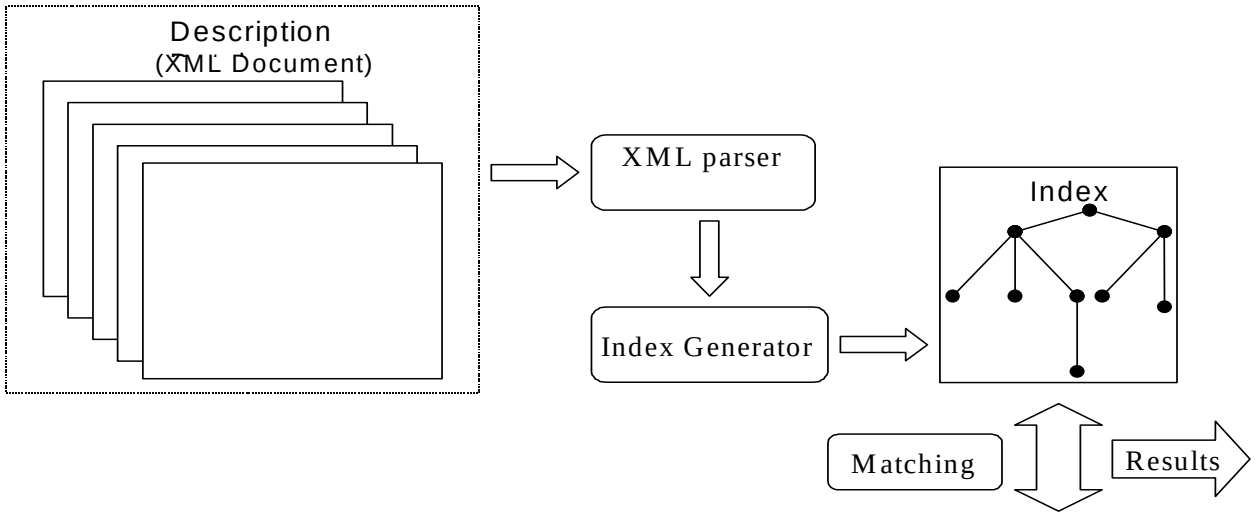
- **Model-based query processing:** Graph/tree matching; e.g.,"Find all clips similar to this one"

Abstract Event Model

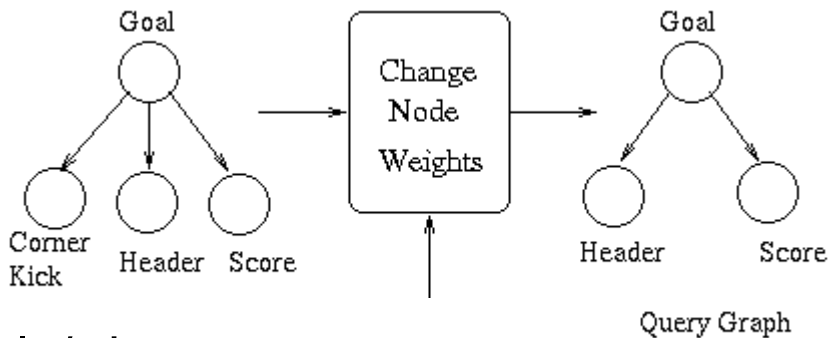


Query Processing

Database



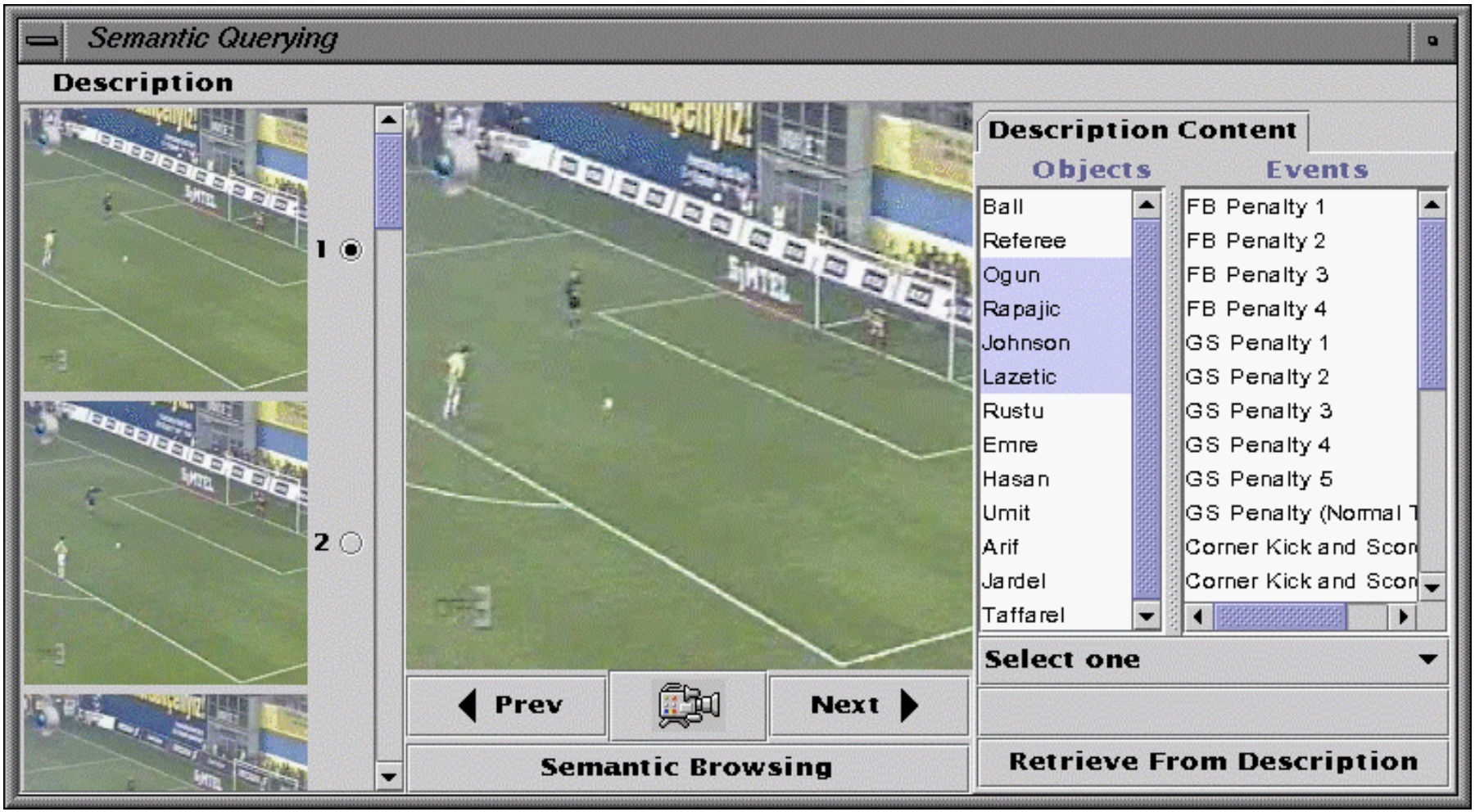
Query Formation



Example/Abstract Description

User

User Interface



Video Processing: Trajectory Estimation

Frame-Based Video Summarization and Shot Classification

Framework

- Shot Boundary Detection
- Extraction of Low-Level Shot Features
  - GoF Color
  - Spatial Layout
  - Motion Activity
  - Shot Duration, ...
  - Key Frame Extraction - for visual summarization
- Fuzzy Clustering to Generate Domain Models
- Analyze New Content in the Domain using the Generated Domain Model (similar to VQ)

GoF Color Representation

- Common approach: Describe visual and color content of a shot with key frames and key frame histograms, respectively.
  - The provided color description may vary significantly with key frame selection criterion
- Alternative: Consider color content of all frames for representative histogram computation
  - Robust color histogram descriptors that are unaffected by outlier frames due to camera movement, occlusion, text/graphics overlays, brightness variations, etc.

Alpha-Trimmed Average Histograms




Key Frame Selection

- For each GoF *I*, compute

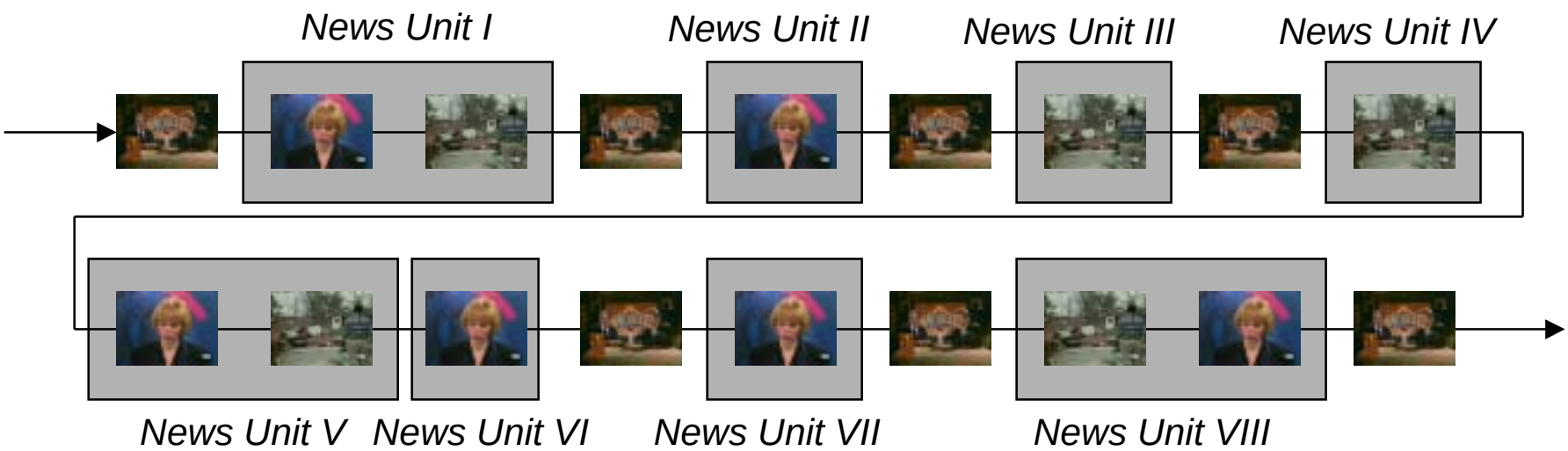
$$E_i = \left\| H_i - H_l \right\| \quad i = 1, \ldots, N.$$

- The frame *r* that minimizes *E* is the key frame

E_{min}

E_{max}

News Unit Generation



 Commercial

 Anchor

 Footage

Location-Based Browsing Using Establishing Shots