Knowledge-Assisted Video Analysis for Content-Adaptive Coding and Transmission

V. Mezaris, N. V. Boulgouris, I. Kompatsiaris

Informatics and Telematics Institute / Centre for Research and Technology Hellas
&
Department of Electronic Engineering, Division of Engineering, King's College London

Outline

- Introduction – problem definition
- Proposed approach
- Domain knowledge
- Knowledge-assisted video analysis
  - Initial segmentation
  - Descriptor extraction
  - Genetic algorithm
- Content-adaptive coding and transmission
- Experimental results
- Conclusions

Introduction – problem definition

- Enormous amount of digital video content available on the internet, more content created at astonishing rates
- Video search, retrieval and transmission
  - Becomes everyday practice (e.g. YouTube)
  - Is not yet efficient enough
    - Search and retrieval based on keywords
    - Coding and transmission insensitive to content semantics
  - Can greatly benefit from taking into account the semantics of video content
    - More efficient search and retrieval schemes
    - Content-dependent coding and transmission

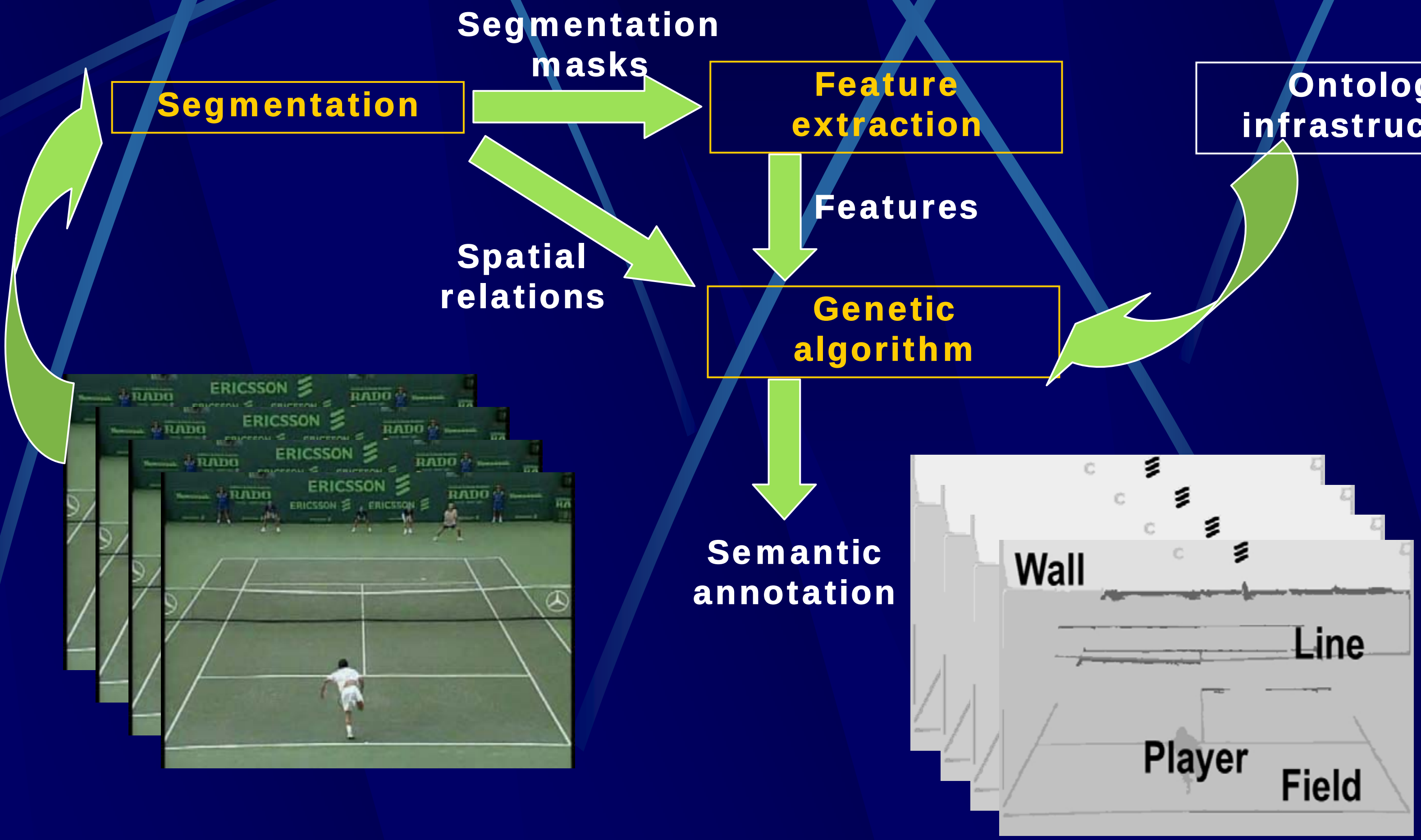
Proposed approach

- Knowledge-assisted, domain-specific video analysis framework for content-adaptive video coding and transmission
  - Domain knowledge defines domain objects and their features, relations, importance factors
  - Extraction of video semantic interpretation
    - Initial segmentation to generate atom regions
    - Feature extraction for atom regions
    - Genetic algorithm to assign regions to domain objects
  - Content-adaptive coding and transmission
    - Object source and channel coding dependent upon defined importance factors for each object

Domain knowledge

- Encoded in the form of an ontology
- Employed knowledge includes
  - Domain objects of interest (e.g. for Tennis domain, Player, Field, Lines etc.)
  - Detection-oriented information
    - Object representative low-level features (e.g. mpeg-7 descriptors) or detection methods (e.g. trained SVM classifiers) for linking the information at the signal level with the previously defined objects
    - Object spatial relations, that can be used to disambiguate detection results, particularly in the case of different objects with similar visual characteristics
  - Coding-oriented information
    - Domain-specific object importance factors
    - Application-specific object importance factors
    - User-specific object importance factors

Knowledge-assisted Analysis architecture



Segmentation

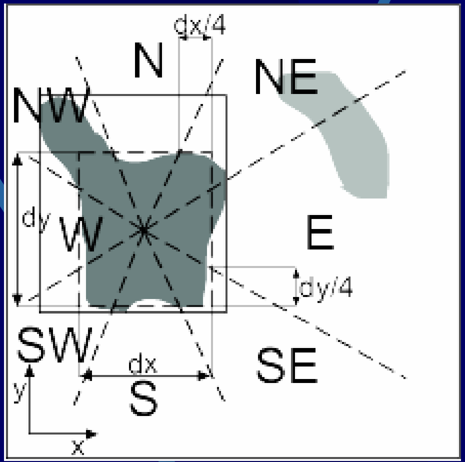
- Color segmentation
  - K-means-based
  - Generates color-homogeneous regions in each frame
- Motion segmentation
  - Using block-matching motion information
  - Iterative block rejection using bilinear motion model
  - Enforcement of temporal consistency constraints
  - Block clustering to connected regions
  - Post-processing using smoothing spline active contours to extract pixel-accuracy segmentation mask
- Combination of masks to extract color- and motion-homogeneous atom regions

Descriptor extraction

- Descriptors
  - Color
    - mpeg-7 dominant color
    - mpeg-7 scalable color
  - Motion
    - norm of average global-motion-compensated motion vectors
  - Shape
    - compactness
  - More descriptors could be extracted and introduced to the subsequent matching procedure to improve object detection performance

Spatial relations

- Supported relations
  - 8 directional relations
    - North (N)
    - East (E)
    - South (S)
    - West (W)
    - South-east (SE)
    - South-west (SW)
    - North-east (NE)
    - North-west (NW)



- Computation
  - A reduced box from the objects minimum bounding rectangle, based on compactness
  - Eight cone-shaped regions formed using this box
  - Percentage of pixels of second region included in each cone-shaped region determines the degree to which each spatial relation is satisfied
  - In this work, only the most dominant relation is taken into account

Genetic algorithm

- Optimization problem formulation
  - Each chromosome represents a possible interpretation of the scene
    - Each gene of a chromosome represents a possible interpretation of an atom region
  - A population of such chromosomes is initially generated
  - The fitness of each chromosome is evaluated using a fitness function, formulated using two key functions
    - One providing the degree of matching between an atom region and an object in the ontology, based on low-level features
    - One providing the degree of matching between the relation of two atom regions and the relation defined in the ontology for the objects associated by the chromosome with them
  - Using selection based on fitness, mutation and crossover, the population evolves until a solution is reached

Genetic algorithm

- Fitness function

$$Fitness(g_t) = (\sum_{i}^{N_R} \mathcal{I}_M^t(r_i, o_m)) \prod_{i}^{N_R} \prod_{j \in S_i} \mathcal{I}_R^t(r_i, o_m, r_j, o_l)$$

- Function $\mathcal{I}_M^t(r_i, o_j)$ estimates degree of region-object matching using visual features
- Function $\mathcal{I}_R^t(r_i, o_j, r_k, o_l)$ estimates spatial relation validity

Content-adaptive coding and transmission

- Coding scheme
  - Combination of source and channel coding
  - Each object compressed using an embedded coding method
    - Object stream can be decoded at arbitrary source rates depending on required quality
- Distortion function defined as a weighted sum of object distortion functions, object importance factors defined in the ontology serve as weights

$$D = \sum_{i=0}^{N-1} f_i D_i(\rho_i, n)$$

where $D_i(\rho_i,n)$ is the decoded object quality, ρ_i is the source+channel bitrate devoted to i-th object and n the BER of the channel

Content-adaptive coding and transmission

- Total bitrate for the video sequence

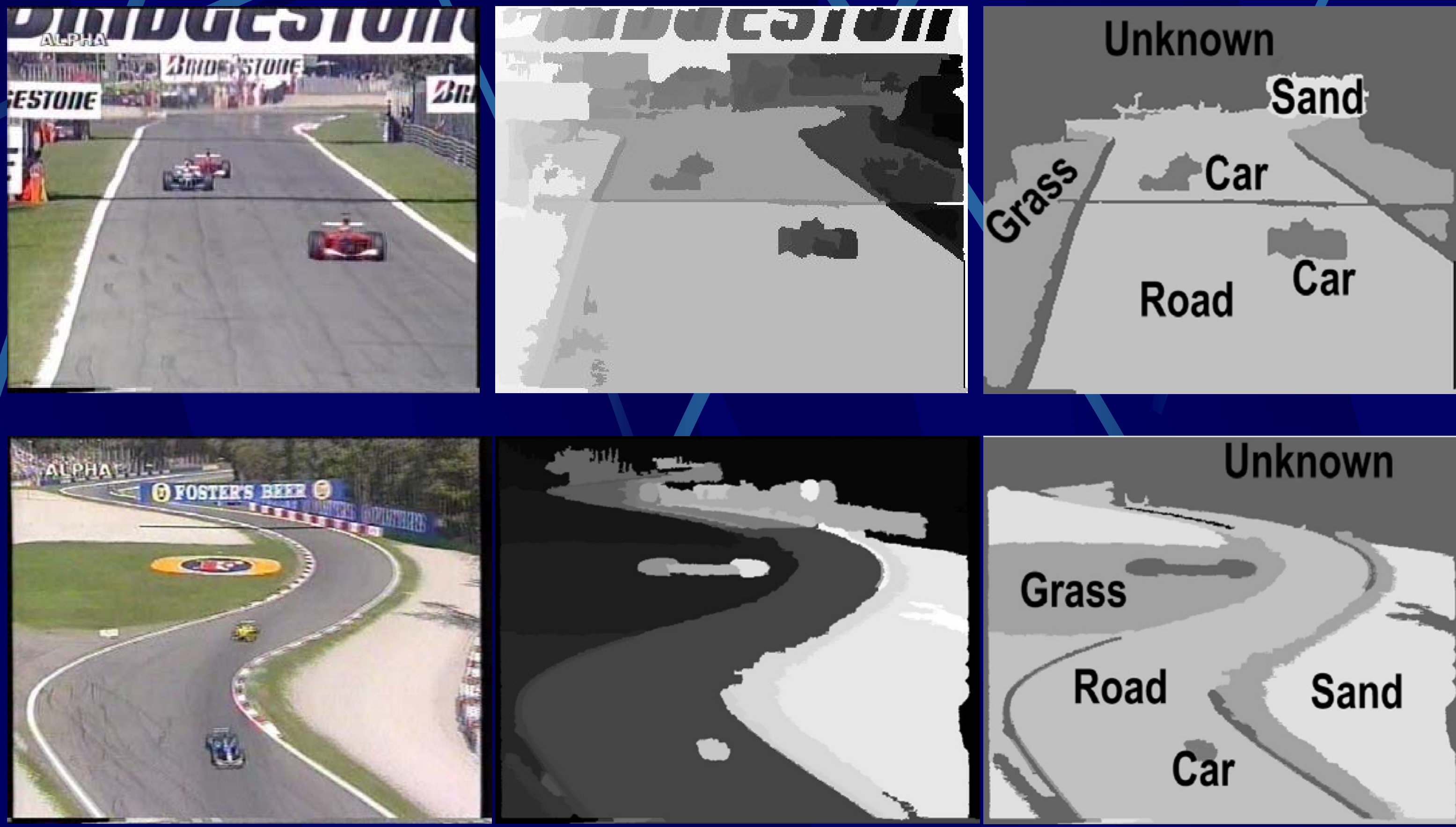
$$R = \sum_{i=0}^{N-1} \rho_i \qquad R = \sum_{i=0}^{N-1} (s_i + c_i)$$

- Source and channel bits for the coding of the i-th object determined by object-wise rate allocation optimization
  - Minimization of objective function F

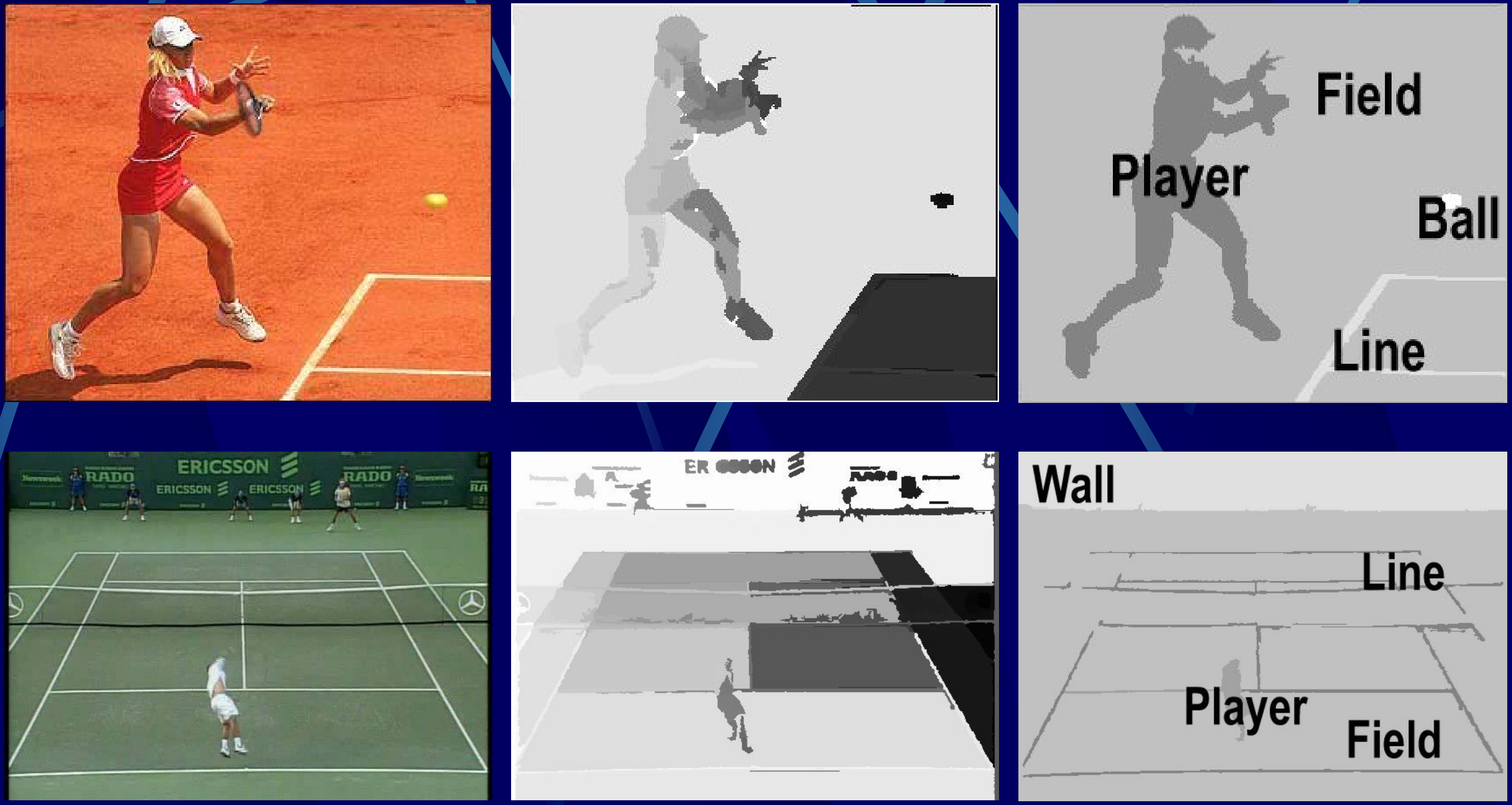
$$F = D + \lambda_L R$$

  where λ_L is a Lagrange multiplier
  - Employed optimization method similar to one developed for block-wise image coding

Experimental results



Experimental results



Conclusions

- Coding and transmission
  - Preliminary experiments based on analysis results
  - Objects designated as being of higher importance were consistently decoded at higher qualities than those of the other objects in the ontology
- Analysis
  - Extracted knowledge can also support other tasks such as efficient semantic level indexing and retrieval of video