National Technical University of Athens
Department of Electrical and Computer Engineering

Efficient Content Representation in MPEG Video Databases

Yannis Avrithis, Nikolaos Doulamis, Anastasios Doulamis and Stefanos Kollias

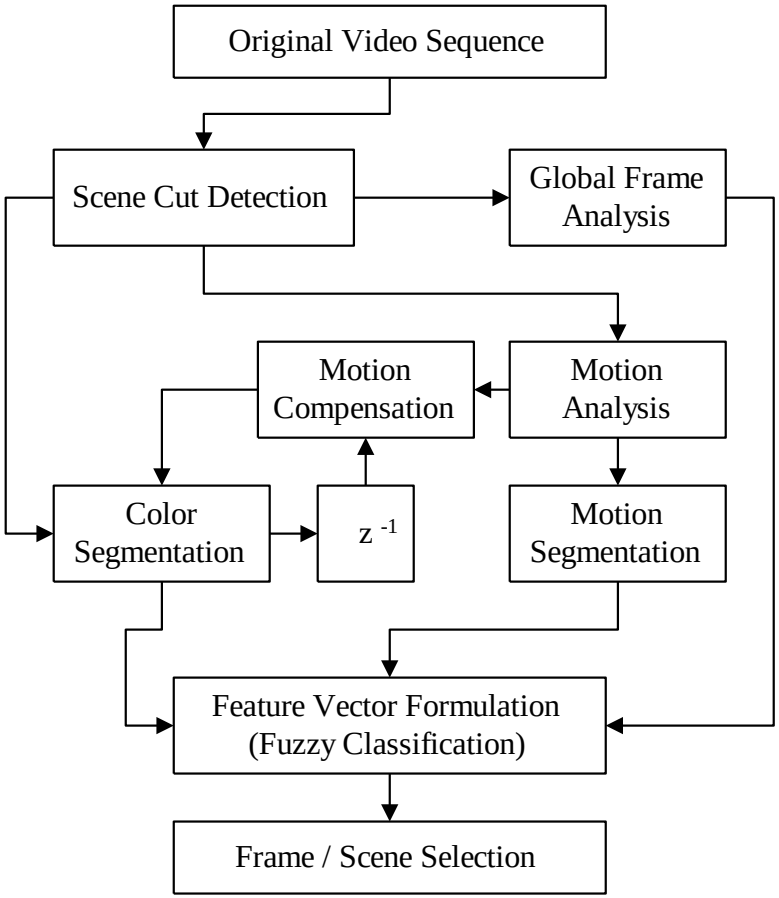
Objective

- Automatic selection of a limited number of *key frames and scenes* from MPEG video streams
- Key frames and scenes provide sufficient information about the content of video sequences
- Representation of video sequences by multi-dimensional *feature vectors* of key frames and scenes, containing color & motion information
- *Video queries* applied directly on feature vectors of key frames and scenes

Applications

- *Multimedia database management:* reduction of storage requirements for search capabilities, direct content-based retrieval, faster and more efficient video queries, improvement of user interface
- *Multimedia interactive services:* production of low resolution video clip previews (trailers) or still image mosaics, browsing of video databases on web pages

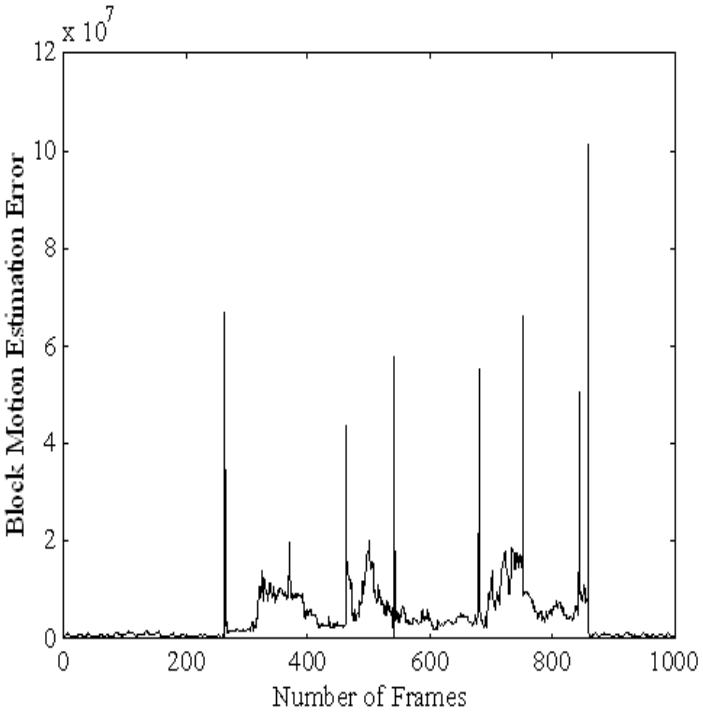
Proposed System Architecture



- Scene cut detection
- Feature extraction for each frame
- Formulation of scene feature vectors
- Selection of the most representative scenes
- Extraction of key frames for each scene

Scene Cut Detection

- Computation of the sum of the block motion estimation error
- Selection of frames for which sum exceeds a certain threshold
- Computations applied directly to MPEG-coded sequences

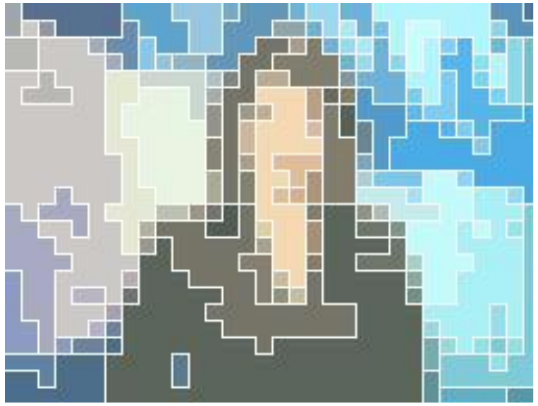


Color Segmentation

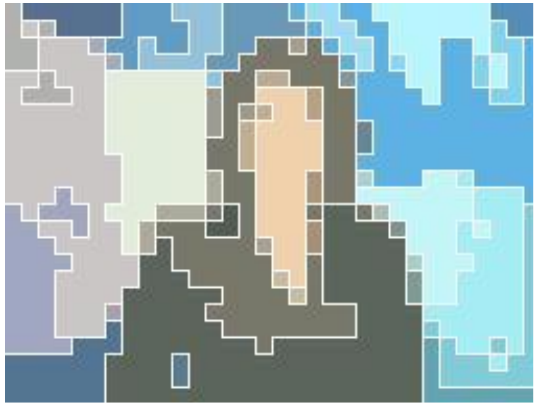
- Segmentation according to *spatial homogeneity*
- *Block resolution* (reduction of computational time, exploitation of MPEG information)
- *Hierarchical merging* of similar segments (depending on color homogeneity & segment size)
- *Color features*: number of segments, location, size & mean color of each segment
- *Object tracking:* comparison with motion compensated segmentation results of previous frames (connected regions are encouraged to remain connected in successive frames)

Color Segmentation Results

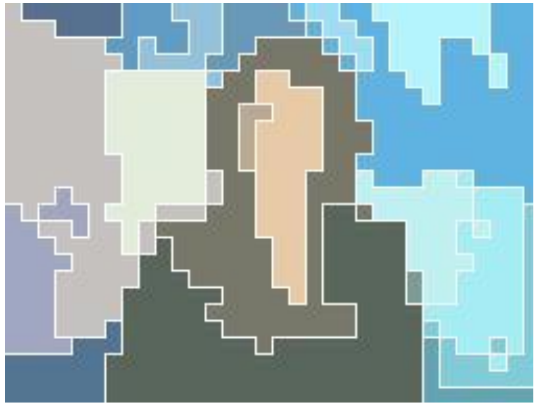
Original frame



1st iteration



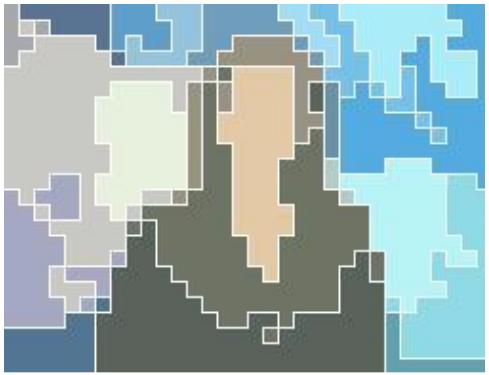
3rd iteration

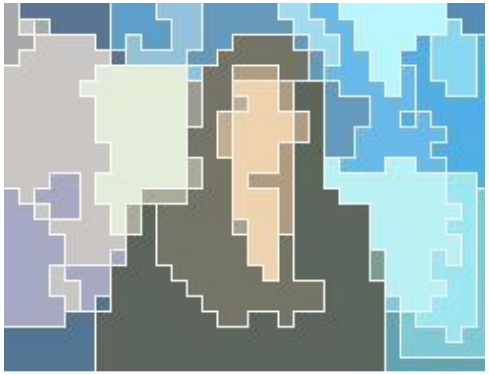


Final Result

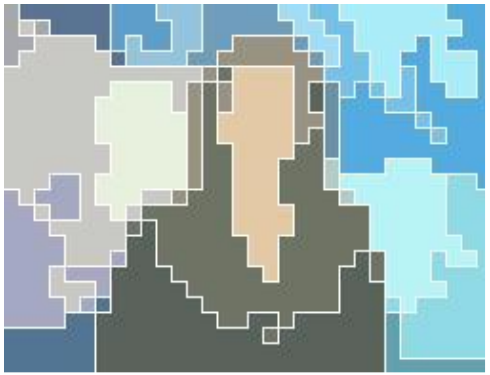
Object Tracking Capabilities

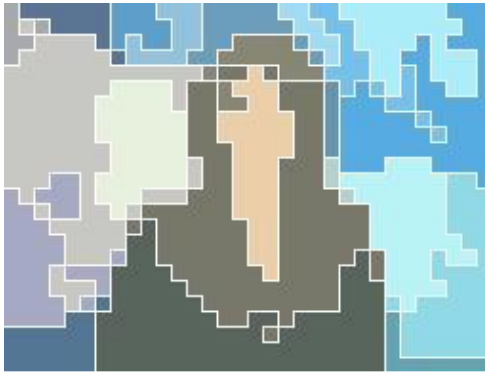
Two original successive frames





Segmentation without tracking



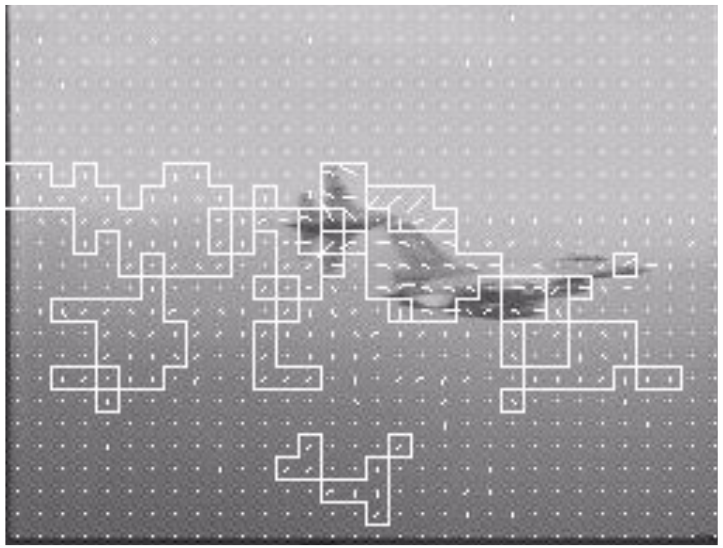


Segmentation with tracking

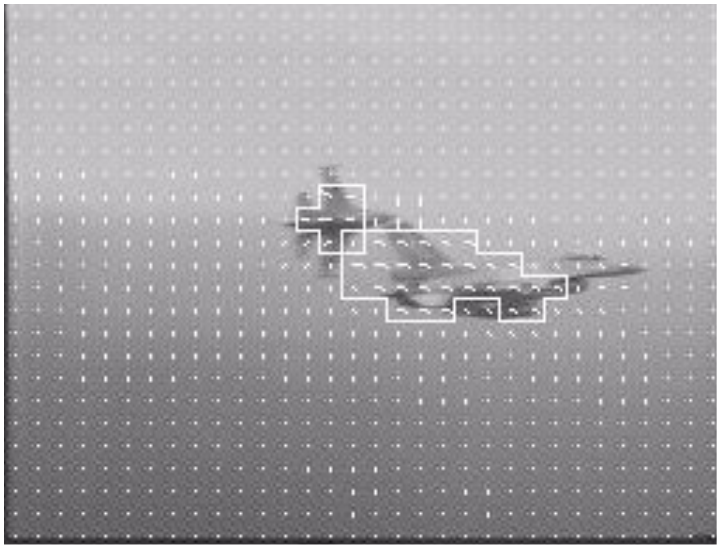
Motion Segmentation

- Segmentation according to *spatial homogeneity*
- *Block resolution* (reduction of computational time)
- Motion vectors derived from motion analysis, or directly from MPEG stream
- *Median filtering* of derived motion vectors: elimination of “noise”, preservation of “edges”
- *Motion features*: number of segments, location, size & mean motion vector of each segment

Motion Segmentation Results



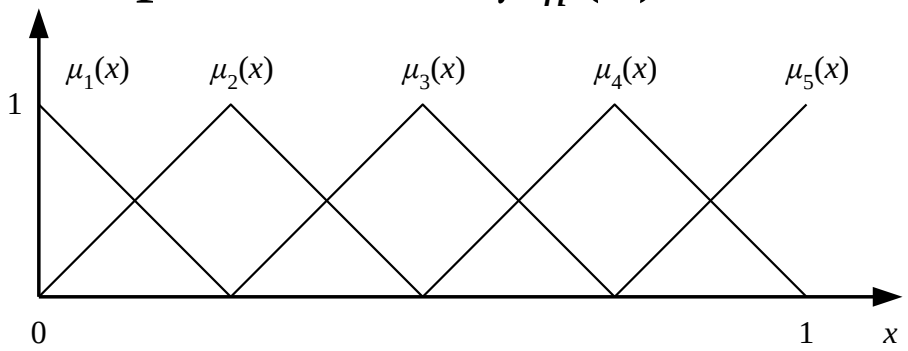
Motion segmentation
without filtering



Motion segmentation
with filtering

Feature Vector Formulation

- *Multidimensional "histogram"*: classification of color and motion segments into pre-determined classes
- *Fuzzy classification:* normalization of each feature x to [0,1], and partitioning into Q classes defined by membership functions $\mu_n(x) \in [0,1],\ n = 1,\ldots,Q$



- Histogram construction possible even with small number of samples x.

Multidimensional Fuzzy Classification

- *Degree of membership* allocated to each class

$$F(n_1,\ldots,n_L)=\sum_{i=1}^{K}\left\{\prod_{j=1}^{L}\mu_{n_j}(f_j^{(i)})\right\}$$

where $n_j \in \{1,2,\ldots,Q\}$: classification index for *j*th feature, *Q*: no. of partitions, *L*: no. of features, *K*: no. of segments, $f_j^{(i)}$: *j*th feature of *i*th segment, $\mu_n(f)$: degree of membership of feature *f* in partition *n*

- Feature vector formed by degrees of membership for all $M=Q^L$ combinations of $n_1,\ldots,n_L \in \{1,2,\ldots,Q\}$
- *Global frame characteristics* included in feature vector (color histogram, etc.)

Representative Scene Selection

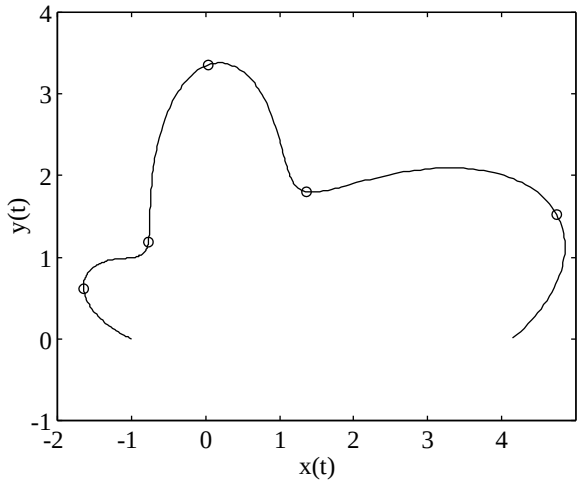
- *Scene feature vector* constructed based on frame feature vectors over duration of scene
- *Clustering* of similar scene feature vectors $\mathbf{s}_i \in \Re^M$, $i=1,\ldots,N_S$ and selection of cluster *representatives* $\mathbf{c}_i$, $i=1,\ldots,K_S$
- *Average distortion* $D(\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{K_S}) = \sum_{i=1}^{K_s} \sum_{\mathbf{s} \in Z_i} d(\mathbf{s}, \mathbf{c}_i)$ should be minimized, where

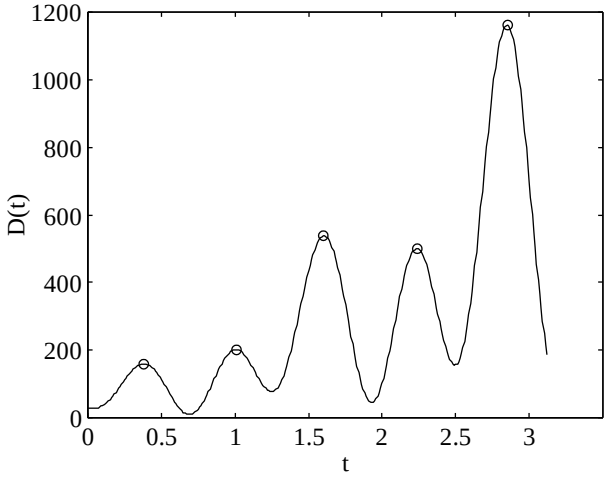
  $Z_i = \{\mathbf{s} \in S : d(\mathbf{s}, \mathbf{c}_i) < d(\mathbf{s}, \mathbf{c}_j) \; \forall \; j \neq i\}$

  is the *influence zone* of $\mathbf{c}_i$
- Minimization performed with *generalized Lloyd* or *K*-means algorithm

Key Frame Selection

- *Selection of frames* whose feature vector resides in extreme locations of feature vector trajectory $\mathbf{r}(t)$
- Magnitude of 2nd derivative $D(t)=|d^2\mathbf{r}(t)/dt^2|$ used as *curvature measure*





- Extremely fast, easy to implement in hardware

Optimization Methods for Frame Selection

- Minimization of a *correlation criterion* (key frames should not be similar to each other)
- *Correlation measure* of feature vectors $\mathbf{f}_i$, $i = x_1, \ldots, x_{K_F}$

$$R(\mathbf{x}) = R(x_1, \ldots, x_{K_F}) = \left(\sum_{i=1}^{K_F - 1} \sum_{j=i+1}^{K_F} (\rho_{x_i, x_j})^2 \right)^{1/2}$$

  where ρ_{ij}: correlation coefficient of vectors f_i, f_j and $\mathbf{x} = (x_1, \ldots, x_{K_F})$: index vector corresponding to a set of selected frame numbers
- Minimization of $R(\mathbf{x})$ w.r.t. $\mathbf{x}$ implemented by *logarithmic search* or *genetic algorithm* (exhaustive search is unfeasible)

Video Queries

- Searching and retrieval of frames based on comparisons in the feature space
- Feature space contains all essential information, while preserving a very low dimension
- Comparisons performed on key frames only
- Dramatic reduction achieved in number of frames required for indexing, browsing or retrieval
- *Adaptive video queries* possible with parametric (weighted) distance function between feature vectors and parameter adaptation according to user requirements

Scene Selection Results








8 scenes of a test video sequence



3 scenes selected as most representative

Key frame selection results

6 key frames from the 2nd representative scene

Conclusions

- *Automatic extraction* of key frames and scenes of video sequences taken from large video databases
- *Optimal selection* of representative scenes
- *Object tracking* provides smoother feature vector trajectories and more robust frame selection
- *Direct implementation* on MPEG video streams
- *Feature vector space* enables robust and efficient frame comparisons, suitable for *video queries*

Further Work

- Integration of color and motion segmentation results
- More robust object tracking
- Optimal frame selection mechanisms
- More intelligent object extraction (e.g., human faces)
- Interweaving of audio and video information
- Adaptive video queries