National Technical University of Athens
Department of Electrical and Computer Engineering

Indexing and Retrieval of the Most Characteristic Frames / Scenes in Video Databases

Anastasios Doulamis, Yannis Avrithis,
Nikolaos Doulamis and Stefanos Kollias

Objective

- Automatic extraction of a small number of frames or scenes from a video sequence
- These frames or scenes provide sufficient information about the sequence

Applications

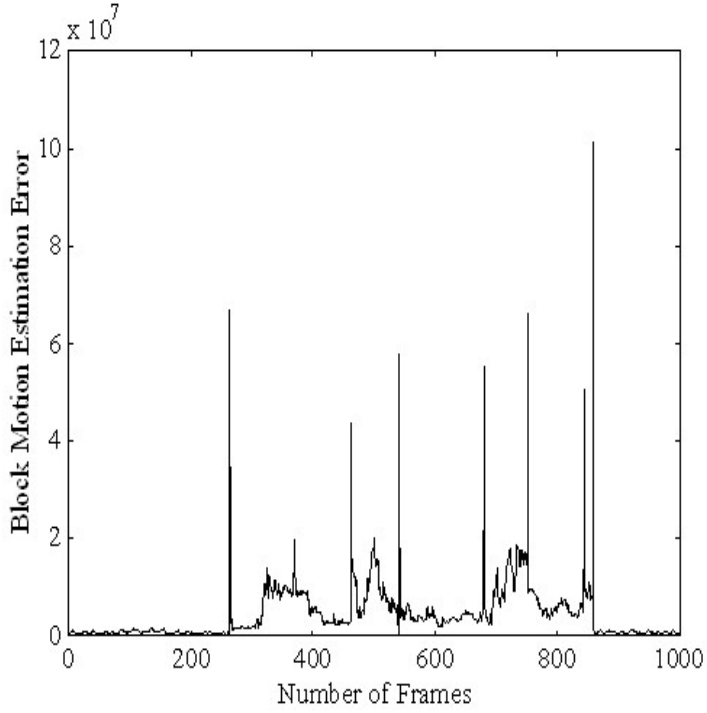
- Multimedia interactive services (browsing of video databases on web pages, production of video trailers)
- Indexing & retrieval

System Overview

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

Scene Cut Detection

- Computation of the sum o the block motion estimation error
- Selection of frames for which sum exceeds a certain threshold
- In MPEG-coded sources, direct measurement of frame size



Color Segmentation

- Segmentation according to *spatial homogeneity* criteria
- *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

Color Segmentation (cont.)

- Let $S_k^{(n)}$ the final k-th segment at the n-th iteration and $S_k^{(n)}(m)$ the k-th segment at n-th iteration after merging m neighboring segments
- Then $S_k^{(n)}(m) = S_k^{(n)}(m-1) \cup S_l^{(n-1)}$

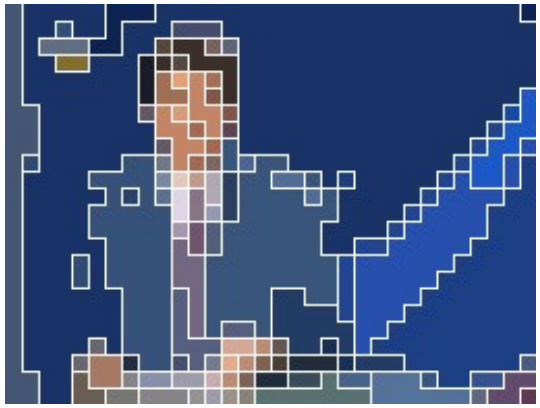
  for all $S_l^{(n-1)} \in N(S_k^{(n)}(m-1))$ such that

  $d(\mathbf{c}(S_l^{(n-1)}), \mathbf{c}(S_k^{(n)}(1)) < f_n(P(S_l^{(n-1)}), P(S_k^{(n)}(m-1))$

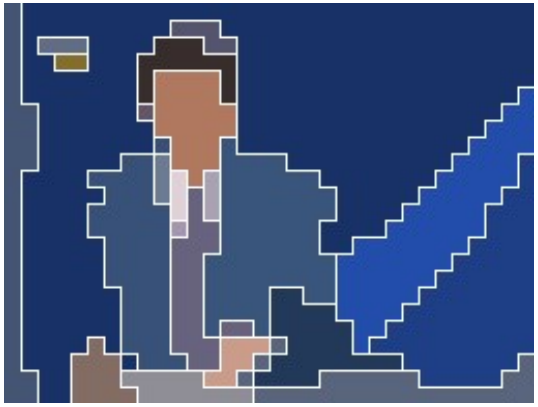
  where $N(S)$ the set of segments neighboring to S, $P(S)$ the number of blocks of S, $f_n(\ )$ a threshold function, $d(\mathbf{x},\mathbf{y})$ a metric distance, and $\mathbf{c}(S)$ the mean color of S

Color Segmentation Results

Original frame



First step of color segmentation



Last stage of color segmentation

Motion Segmentation

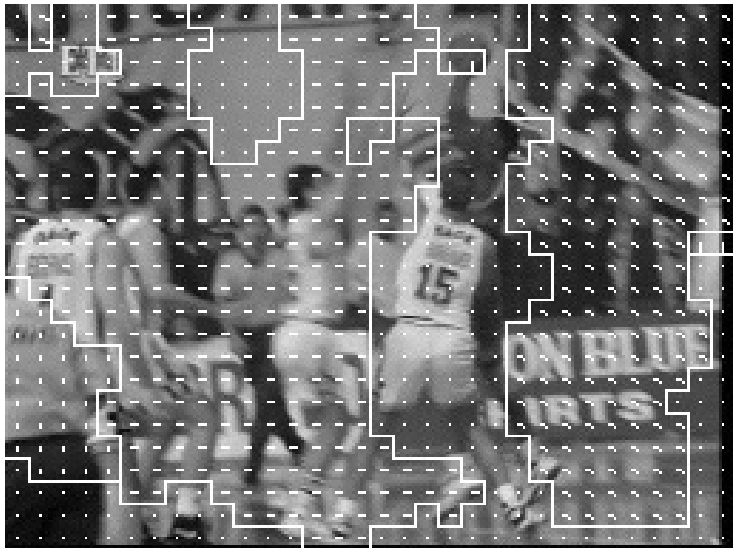
- Segmentation according to *spatial homogeneity*
- *Block resolution*
- *Noise suppression* by adding a penalty term to the equation of motion vector calculation:

$$\hat{\mathbf{v}} = (\hat{v}_x, \hat{v}_y) = \arg\min_{(v_x, v_y) \in A} \sum_{k=8*i}^{8*i+7} \sum_{l=8*j}^{8*j+7} (u^{(n)}(k,l) - u^{(n-1)}(k - v_x, l - v_y))^2 + D(v_x, v_y)$$

- *Median filtering* of derived motion vectors
- *Motion features*: number of segments, location, size & mean motion vector of each segment

Motion Segmentation Results

Motion vectors
after post-processing



Motion segments

Feature Vector Formulation

- *Multidimensional "histogram"*: classification of color and motion segments into pre-determined categories
- *Fuzzy classification*: Degree of membership allocated to each category

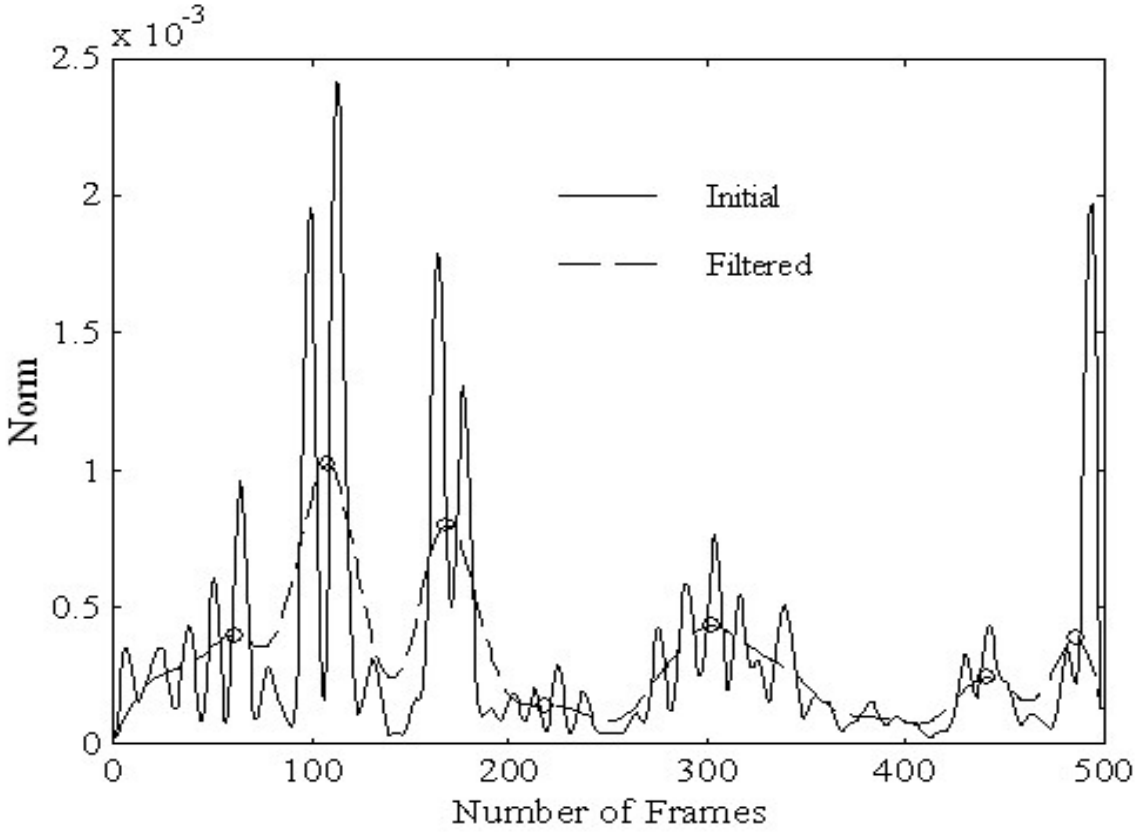
$$F(n_1,\ldots,n_L)=\sum_{i=1}^{K}\left\{\prod_{j=1}^{L}\mu_{n_j}(f_j^{(i)})\right\}$$

where $n_j \in \{0,1,\ldots,Q\}$: 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*

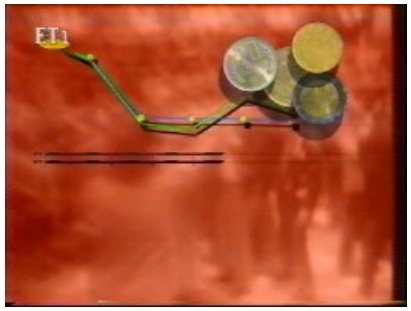
Scene and Frame Selection

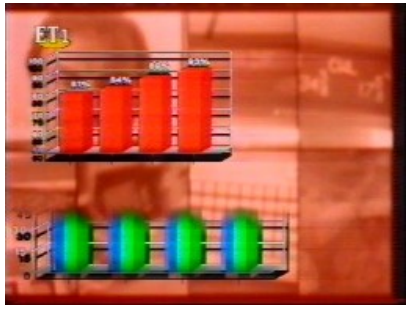
- *Scene vector* constructed based on frame feature vectors over duration of scene
- *Scene selection* accomplished by neural classifier mechanism with scene vectors as input
- *Selection of frames* whose feature vector resides in extreme locations of vector space: magnitude of 2^{nd} derivative used as curvature measure
- *Video queries*: searching and retrieval of frames based on comparisons in the feature space

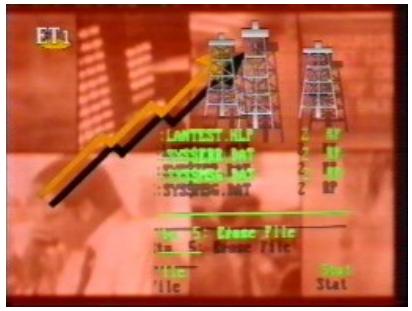
Frame Selection Results

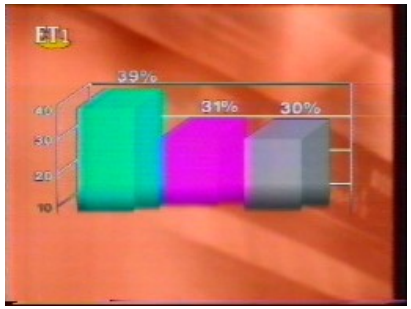


Frame Selection Results (cont.)









Conclusions

- *Automatic extraction* of most characteristic frames or scenes of video sequences taken from large video databases
- *Feature vector space* enables robust and efficient frame comparisons, suitable for *video queries*
- *Possible improvements*: more robust color or motion segmentation algorithms, enhancement of frame selection mechanism, inclusion of additional components (shape or texture information) in feature vector