# Video Sequence Segmentation

J.R. Casas, J. Llach, P. Salembier
C. Dorea, M. Pardàs, F. Marqués

Grup d'Imatge, Departament TSC
ETSETB — UPC

## Image Processing Group

### Staff

**Professors**

- Josep R. Casas
- Antoni Gasull
- Ferran Marqués
- Albert Oliveras
- Montse Pardàs
- Philippe Salembier
- Elisa Sayrol

**Assistant Professors**

- Xavier Giró
- Ramon Morros
- Javier Ruiz
- Verónica Vilaplana

**Research Assistants**

- Mathieu De Craene
- Joachim Neumann

### PhD Students

- Felipe Calderero
- Cristian Cantón
- Camilo Chang
- Mariella Dimiccoli
- Christian Ferran
- José Luis Landabaso
- Miriam Leon
- Joel Solé

---

- Pedro Correa (UCL)
- Jose Carlos Pujol (UCIII)

# Research activity … Today

Basic tools ⟷ Applied

| Name | Nonlinear image processing | Segment. | Tracking | Face detection & recognition | Emotion recognition | Body analysis | Text analysis | Compres-sion | Indexing | Water-marking | Remote sensing | Biomedical |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Felipe Calderero | | | | | | x | | | | | | |
| Josep Ramon Casas | | x | x | | | x | | x | x | | | x |
| Cristian Canton | | | x | | | x | | | | | | |
| Camilo Chang Dorea | | x | x | | | | | | | | | |
| *Pedro Correa (UCL)* | | | | | | x | | | | | | |
| Mathieu De Craene | | | | | | | | | | | | x |
| Mariella Dimiccoli | x | x | | | | | | | | | | |
| Christian Ferran | | x | | | | | | | | | | |
| Toni Gasull | | x | | | | | x | | | | x | x |
| Xavier Giro | | | | | | | | | x | | | |
| José Luis Landabaso | | x | x | | | | | | | | | |
| Miriam Leon | | | | | | | x | | | | | |
| Ferran Marqués | | x | x | x | | x | | x | x | | x | x |
| Ramon Morros | | | | | | | | x | | | | |
| Albert Oliveras | | | | | | | | | | | | x |
| Montse Pardàs | | x | x | | x | x | | | | | | |
| *Jose Carlos Pujol (UCIII)* | | x | x | | | | | x | | | | |
| Julio C. Rolon | | | | | | | | x | | | | |
| Javier Ruiz | | x | x | | | | | x | x | | | |
| Philippe Salembier | x | x | x | | | | | x | x | | x | x |
| Elisa Sayrol | | | | | | | | | | x | | |
| Joel Solé | x | | | | | | | x | | | | |
| Veronica Vilaplana | | | | x | | | | | | | | |

# Region based model

I. Introduction
- Image sequence segmentation
- Segmentation steps:
  - Feature selection
  - Decision space

II. Transition based techniques
- Transition based temporal segmentation
  - Shot detection
- Transition based spatial segmentation
  - Rough object detection
- Transition based spatial/temporal segmentation
  - Snakes

III. Homogeneity-based techniques
- Homogeneity based temporal segmentation
  - Shot detection
- Homogeneity based spatial segmentation
  - Region matching
- Homogeneity based spatial/temporal segmentation
  - Robust motion estimation

IV. Summary

# Region based model

I. **Introduction**
- **Image sequence segmentation**
- **Segmentation steps:**
  - **Feature selection**
  - **Decision space**

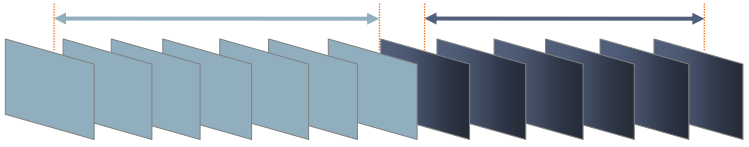
II. Transition based techniques
- Transition based temporal segmentation
  - Shot detection
- Transition based spatial segmentation
  - Rough object detection
- Transition based spatial/temporal segmentation
  - Snakes

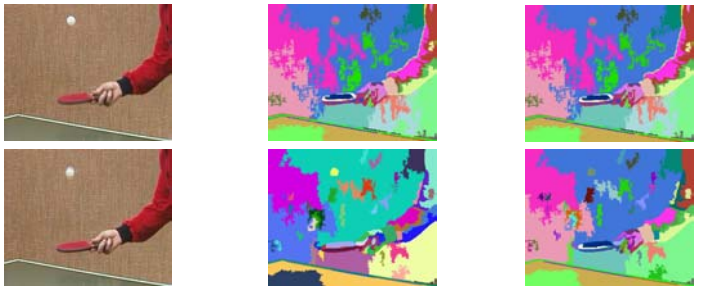
III. Homogeneity-based techniques
- Homogeneity based temporal segmentation
  - Shot detection
- Homogeneity based spatial segmentation
  - Region matching
- Homogeneity based spatial/temporal segmentation
  - Robust motion estimation

IV. Summary

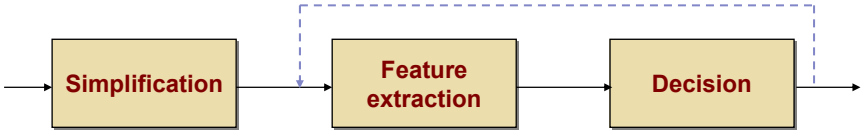
# Image sequence segmentation

**Different meanings**:
- Shot detection: temporal segmentation
- 2D Region segmentation: spatial segmentation
- 3D Region segmentation: spatial/temporal segmentation





# Segmentation steps




- Remove useless or annoying information
- Preserve shape information

- Feature space:
  Gray level/Color
  Texture
  Motion
  Depth
  Frame difference
  DFD
  Histogram
- Elements belonging to the same regions should correspond to homogeneous features

- Decision space:
  1D: Temporal
  2D: Spatial
  3D: Spatial/temporal
- Partition definition:
  → Transition
  → Homogeneity

# Region based model

I. Introduction
- Image sequence segmentation
- Segmentation steps:
  - Feature selection
  - Decision space

II. **Transition based techniques**
- **Transition based temporal segmentation**
  - **Shot detection**
- **Transition based spatial segmentation**
  - **Rough object detection**
- **Transition based spatial/temporal segmentation**
  - **Snakes**

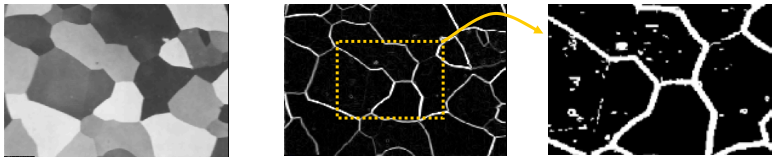
III. Homogeneity-based techniques
- Homogeneity based temporal segmentation
  - Shot detection
- Homogeneity based spatial segmentation
  - Region matching
- Homogeneity based spatial/temporal segmentation
  - Robust motion estimation

IV. Summary

# Introduction: Transition based techniques

## Spatial image segmentation:

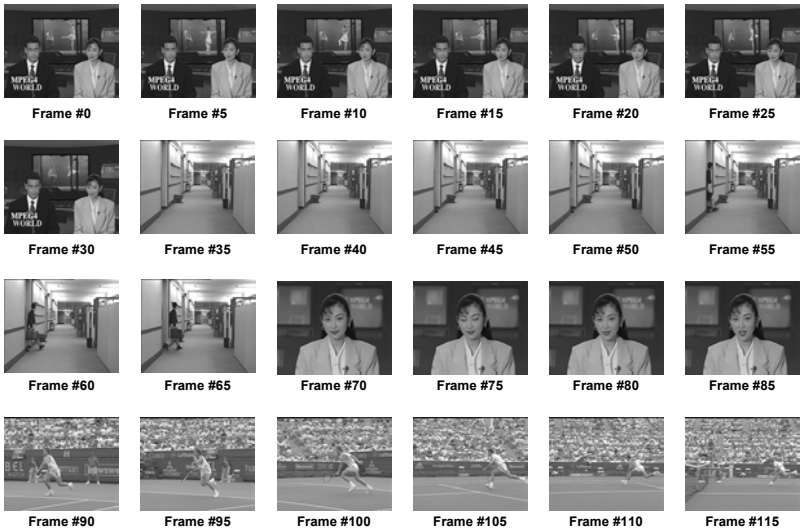
- Straightforward strategy: Gradient binarization



- Difficulty to define the threshold
- Noise may appear: Cleaning with opening
- Contours may not be closed: Closing or gradient extrapolation
- Contours may be thick: remove pixels without breaking the contours
- Not very robust!

# Transition based temporal segmentation (I)

## Temporal segmentation: an example (I)



# Transition based temporal segmentation (II)

**Temporal segmentation:**

- A video shot is a sequence of frames captured by one camera in a single continuous action in time and space.
- Straightforward strategy: Gradient binarization
  (Gradient as feature differences between consecutive frames)
- Gradient on which features?
  - Raw information:
    - Frame difference (FD)
    - Histogram comparison: grey and color information
      *Transition based color temporal segmentation*
    - Displaced frame difference (DFD)
    - Optical flow: estimation of its continuity.
      *Transition based motion temporal segmentation*
  - Coded information:
    - Global camera motion: estimated from the BM result.
    - DC components information: a spatially reduced image version
    - Comparison (inner product) between DCT coefficients
    - Transitions at different accuracy: I-frames, P-frames, B-frames

# Transition based temporal segmentation (III)

**Temporal segmentation:**

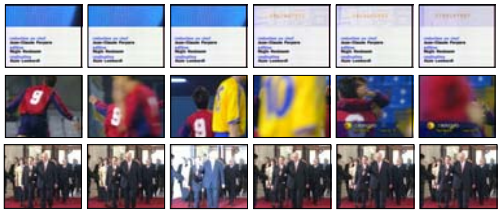
- Typical problems:
  - Difficulty to define the threshold: It still exists
  - Noise may appear: Estimation of shot duration
  - Contours may not be closed: 1D signal
  - Contours may be thick: Shrink contours
  - Edited contents: Gradual Transitions:
    - Fade: Progressive darkening or illumination (fade-out or fade-in)
    - Dissolve: Gradual transition between two shots
    - Geometric effects

# Transition based temporal segmentation (IV)

## Example of shot transitions:

**The changes correspond to shot transitions**

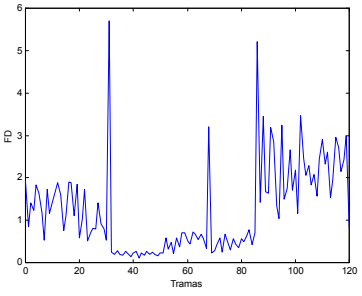
## Example of internal transitions or motion:



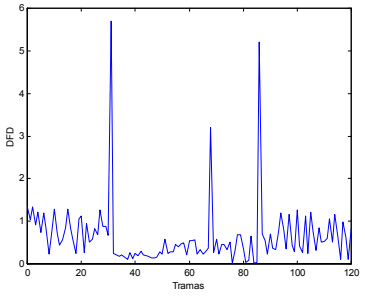
**The changes do *not* correspond to shot transitions**

# Transition based temporal segmentation (V)

## Temporal segmentation: an example (II)




FD comparison




DFD comparison

The DFD comparison allows predicting the current image with the previous information leading to a more robust estimation of the shot transition

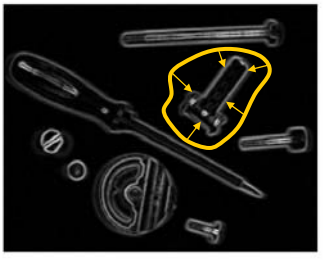
# Transition based spatial segmentation

- Not really used:
  - Difficulty to define the threshold
  - Noise may appear: Cleaning with opening
  - Contours may not be closed: Closing or gradient extrapolation
  - Contours may be thick: remove pixels without breaking the contours
- Possible application: Rough segmentation:
  - Rough definition and tracking of the objects in the scene.
  - Although the decision space is the 2D image, usually they take advantage of the motion (or position) information:

    It is not a pure *Transition based color spatial segmentation*
  - Indexing application: Rough indexing.

# Transition based spatial/temporal segmentation (I)

- Robust strategy: Active contours (Snakes)
  - Evolution of a close curve towards the points of high gradients



  - How to define the curve evolution?: Partial Derivate Equations
  - How to implement it?: Implicit implementation: Level sets
  - How to define the initial curve?: Application dependent

## Transition based spatial/temporal segmentation (II)

- Active contours (Snakes) in video sequences
  - The initial curve at every frame is obtained by a motion based prediction relying on the previous frame final curve
  - The decision is performed on the 2D space

    It is not a pure *Transition based color spatial/temporal segmentation*

- Other approaches
  - Inclusion of a motion based term in the energy

    Almost a *Transition based color-motion spatial/temporal segmentation*
  - Definition of 3D active surfaces

    *Transition based color spatial/temporal segmentation*

## Region based model

I. Introduction
   - Image sequence segmentation
   - Segmentation steps:
     - Feature selection
     - Decision space

II. Transition based techniques
   - Transition based temporal segmentation
     - Shot detection
   - Transition based spatial segmentation
     - Rough object detection
   - Transition based spatial/temporal segmentation
     - Snakes

III. **Homogeneity-based techniques**
   - **Homogeneity based temporal segmentation**
     - **Shot detection**
   - **Homogeneity based spatial segmentation**
     - **Region matching**
   - **Homogeneity based spatial/temporal segmentation**
     - **Robust motion estimation**

IV. Summary

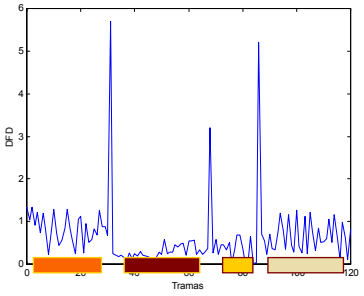
# Introduction: Homogeneity based techniques

Two types of algorithms:

- Feature space segmentation and image classification
  - Statistical analysis and threshold

    Not commonly used in image sequence segmentation
- Direct image segmentation
  - Criteria
    - Texture/Contour/Motion
    - Deterministic and Statistical: Both are commonly used
  - Optimization algorithm
    - Split & Merge: Usually over the motion information
    - Region growing: Initial regions are estimated using motion information

# Homogeneity based temporal segmentation (I)

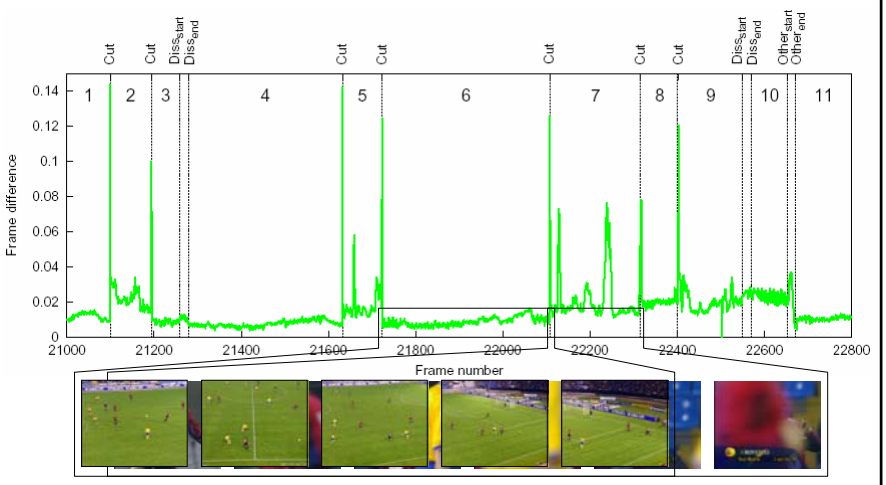



DFD comparison

- Direct signal segmentation:
  - Criteria: DFD comparison
  - Optimization: Region growing
    - Initial region selection: Zones with similar activity

## Homogeneity based temporal segmentation (II)

Color DFD for a portion of a soccer sequence:




## Homogeneity based temporal segmentation (III)

### Segmentation of the frame difference curve:

- Strategy:
  - Detect the homogeneous portions of the curve, and
  - Expand the detected regions until the whole curve is covered.
- Segmentation process based on 2D/3D segmentation techniques:
  1. Simplification: Curve simplification
  2. Feature extraction: Marker extraction
  3. Decision: Marker growing
- At this point of the process, we prioritize detecting all the transitions over avoiding false positives.

# Homogeneity based temporal segmentation (IV)

## Simplification step:

- The goal of the simplification process is to remove the oscillations that do not correspond to shot transitions.
- We rely on morphological filters:
  - Temporal filtering (morphological closing):

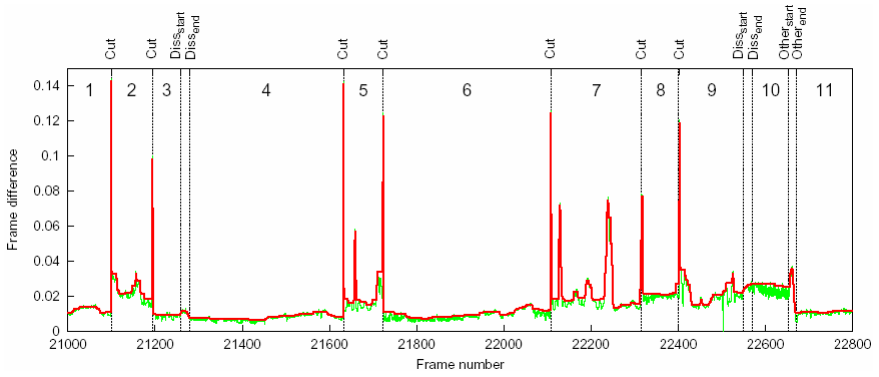
$$\mathrm{FD}_{\mathrm{temporal}}(t) = \varphi_w(\mathrm{FD}(t))$$

  - Positive contrast filtering (morphological reconstruction by geodesic dilation):

$$\gamma^{\mathrm{rec}}(f, f-c) = \delta_{\infty}^{f}(f-c)$$

# Homogeneity based temporal segmentation (V)

- Resulting curve after the **temporal and positive contrast** filtering:



# Homogeneity based temporal segmentation (VI)

## Feature extraction:

- Detect segments of the frame difference curve that represent homogeneous content.
- Detection of the local minima using a negative contrast filter:
  - Negative contrast filter: reconstruction of ***f(t)+c*** over ***f(t)*** by geodesic erosion:

$$\varphi^{rec}(f, f+c) = \varepsilon_{\infty}^{f}(f+c)$$

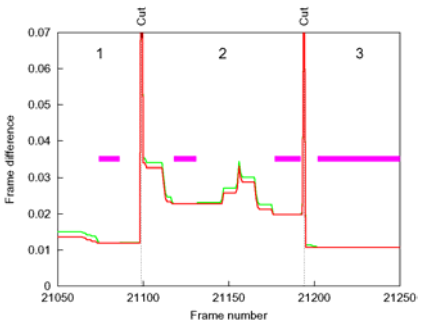
  - The markers are those connected components that verify:

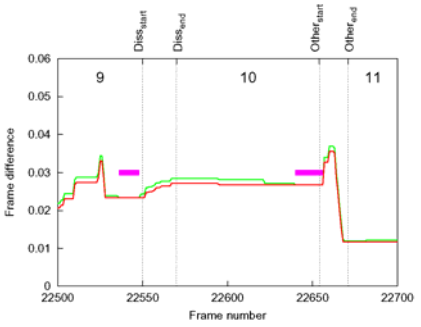
$$\varphi^{rec}(f, f+c) = f+c$$

# Homogeneity based temporal segmentation (VII)

Examples of resulting markers:

- All transitions are detected
- Several markers can be detected in a single shot.

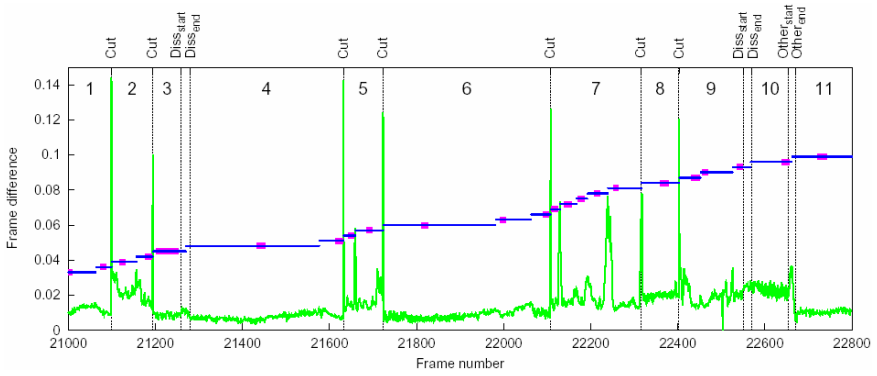


# Homogeneity based temporal segmentation (VIII)

Examples of final decision:

- All transitions are detected
- It may produce an over-segmentation

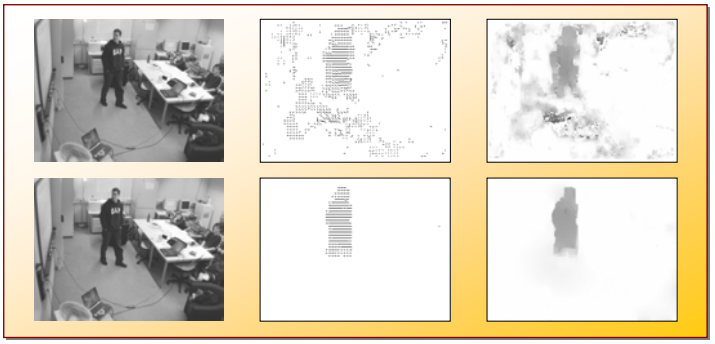



# Region based model

# Homogeneity based spatial segmentation (I)

Direct signal segmentation:

- Criteria: Texture/Contour/Motion/Depth
- Motion segmentation:
  - Motion information has to be estimated: Contours are not well defined.
  - Optical flow may be very noisy: Robust estimators
  - Motion information is a vector field: Criteria usually require scalar functions

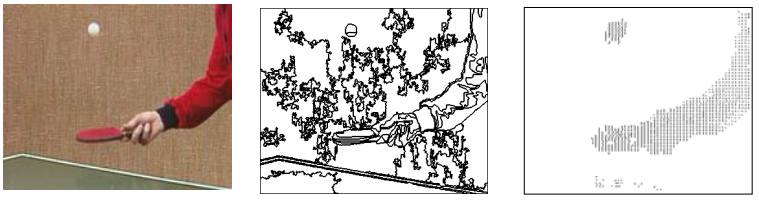


# Homogeneity based spatial segmentation (II)

Direct signal segmentation:

- Combination of criteria: Texture/Contour/Motion

  *Homogeneity based motion-texture spatial segmentation*
  - Compute the motion estimation relying on a texture partition.
  - Merge texture homogenous regions with similar motion
  - Contour information can be added to both criteria



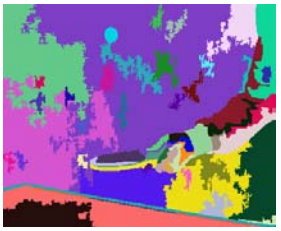
- Detection of objects in the scene:

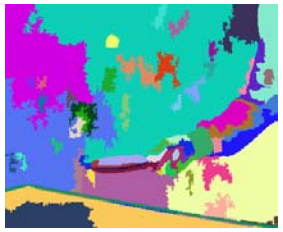
  *An object is a set of color homogenous regions that move following a given motion model.*
  - Texture information is not enough to define an object
  - The motion model has to be given.

# Homogeneity based spatial segmentation (III)

Direct signal segmentation:

- To obtain the temporal link between regions a region matching step is necessary:
  - Position and graph analysis: Sub-graph matching





# Region based model

## Homogeneity based spatial/temporal segmentation (I)

Use of 3D decision space: 2D Space + 1D Time

- Segmentation algorithms require the definition of neighborhoods.
- Neighborhoods are defined in the 3D space:
  - Spatial connectivity:
    - Four- or Eight-connected neighborhoods
  - Temporal connectivity:
    - Collocated pixels in subsequent images:
      Object velocity versus object size:
      Instances of the same object disconnected through time
      It may require an object/region matching algorithm
    - Pixels related through motion estimation:
      Motion has to be estimated:
      Inaccurate connectivity between spatial components
      It may require defining uncertainty areas

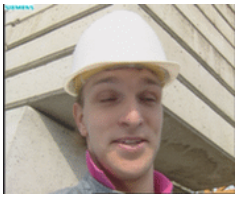
## Homogeneity based spatial/temporal segmentation (II)

Spatial/Temporal Region Growing:

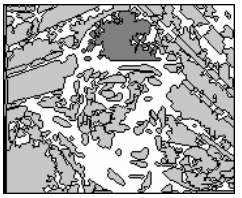
- Growing of 3D regions
  - More commonly used in real 3D (volume) data.
- Motion compensation of regions in previous images:
  - Initial regions: Correctly compensated previous regions
    - The compensated regions have to approximate the current partition.
  - Partition projection:
    - Partial fitting of projected regions in the current texture partition
    - Growing of the fit projected regions (initial regions or markers) taking into account the information in the current image (texture/contour/motion).
  - Temporal coherence of the partition sequence:
    Tracking of labels

## Homogeneity based spatial/temporal segmentation (III)

- Example of Partition Projection (I):
  - Partial fitting based on spatial and texture information.


Frame #0


Frame #1


Spatial fit markers

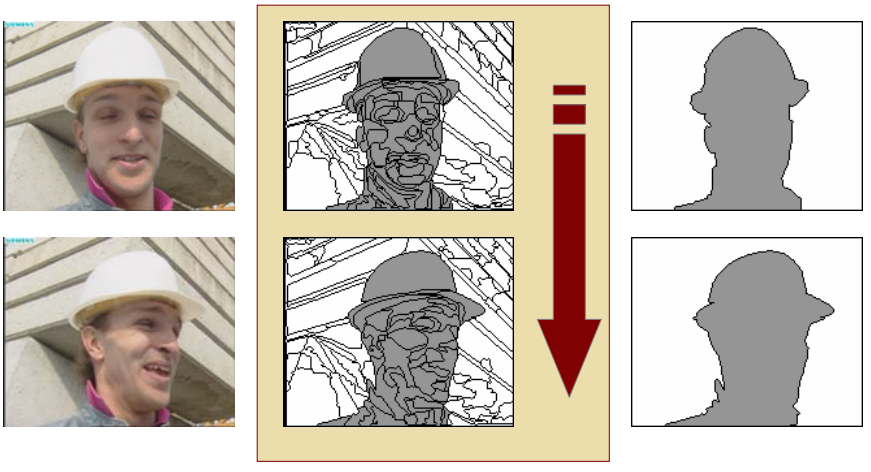

Partition at #0


Fine texture partition at #1


Spatial/Texture fit markers

## Homogeneity based spatial/temporal segmentation (IV)

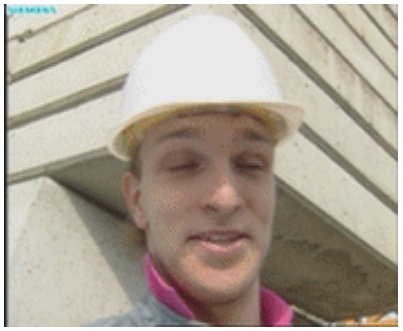
- Example of Partition Projection (II):
  - Final result obtained by a (2D or 3D) region growing of the markers.
  - It allows tracking (previously defined) objects in the scene.



## Homogeneity based spatial/temporal segmentation (V)

- Example of Partition Projection (III):

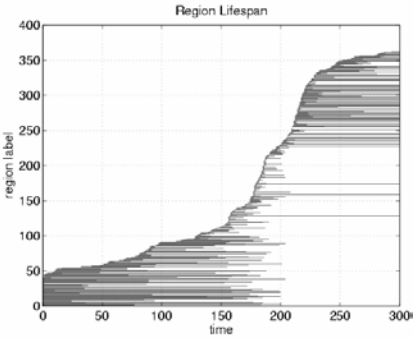
  *Homogeneity based texture space/temporal segmentation*.



## Homogeneity based spatial/temporal segmentation (VI)

- Example of Partition Projection (IV):

  *Homogeneity based texture space/temporal segmentation*.



## Homogeneity based spatial/temporal segmentation (VII)

- Example of Partition Projection (V):

  *Homogeneity based texture space/temporal segmentation*.

Frame #0

Frame #10

Frame #35

Frame #105

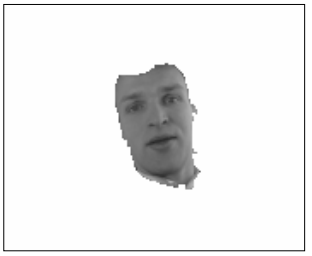
Frame #145

Frame #185

## Homogeneity based spatial/temporal segmentation (VIII)

- Example of Partition Projection (IV):

  *Homogeneity based texture space/temporal segmentation*.



## Homogeneity based spatial/temporal segmentation (IX)

Detection of objects in the scene:

*An object is a set of color homogenous regions that move following a given motion model.*

- Texture information is not enough to define an object
- The motion model has to be given.

Previous approach:

- Use of motion for projection:
  - No use as feature or decision space.
- Tracking of objects:
  - Objects have to be previously identified.

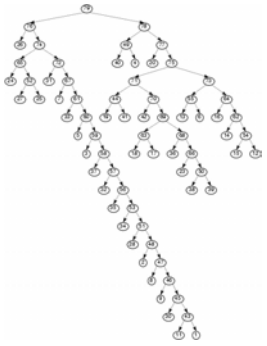
Other approaches:

- Combine motion and texture as feature and/or decision spaces:
  - In a single joint criteria: Typically in statistical approaches
  - In successive criteria: Typically in deterministic approaches
- Directly detect objects imposing a motion model:
  - Typically affine motion models

## Homogeneity based spatial/temporal segmentation (X)

Example of combination of successive criteria (I):

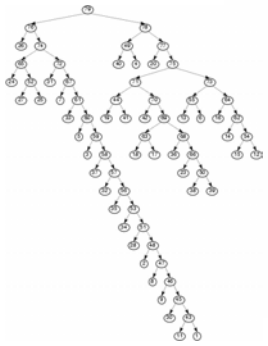
- Motion estimation and merging on a texture homogenous partition:
  *Homogeneity based texture/motion spatial/temporal segmentation*

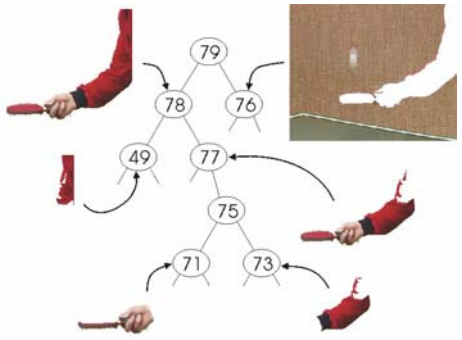


## Homogeneity based spatial/temporal segmentation (XI)

Example of combination of successive criteria (II):

- Motion estimation and merging on a texture homogenous partition:
  *Homogeneity based texture/motion spatial/temporal segmentation*



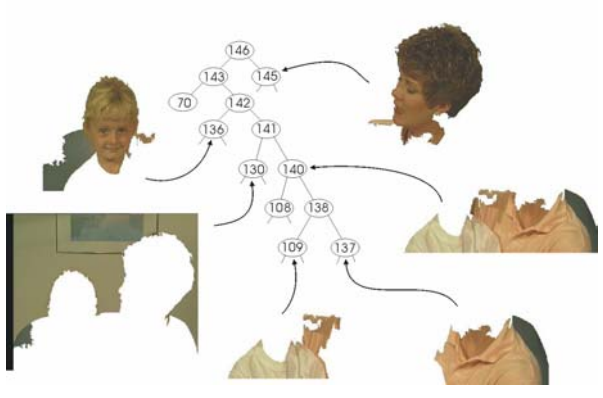



## Homogeneity based spatial/temporal segmentation (XI)

Example of combination of successive criteria (II):

Initial partition, frame 38

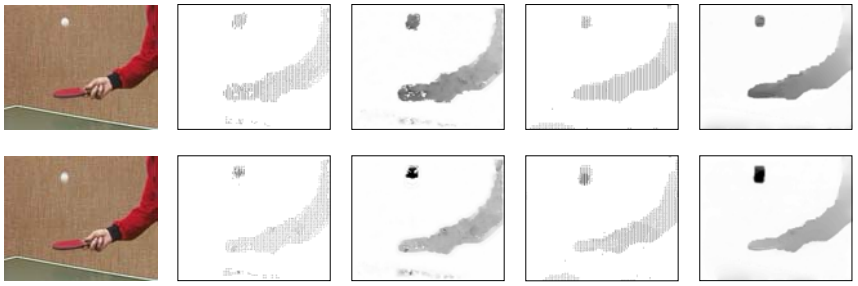
Optical flow (zero motion)



## Homogeneity based spatial/temporal segmentation (XII)

Example of combination of successive criteria (III):

- Motion estimation and merging on a texture homogenous partition:
  *Homogeneity based texture/motion spatial/temporal segmentation*

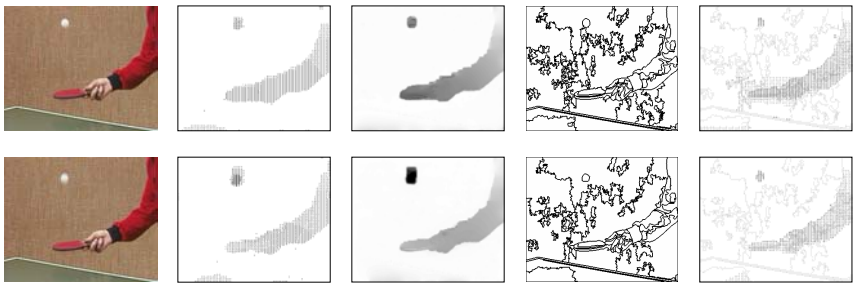


Non robust motion estimation | Robust motion estimation

## Homogeneity based spatial/temporal segmentation (XIII)

Example of combination of successive criteria (IV):

- Motion estimation and merging on a texture homogenous partition:
  *Homogeneity based texture/motion spatial/temporal segmentation*

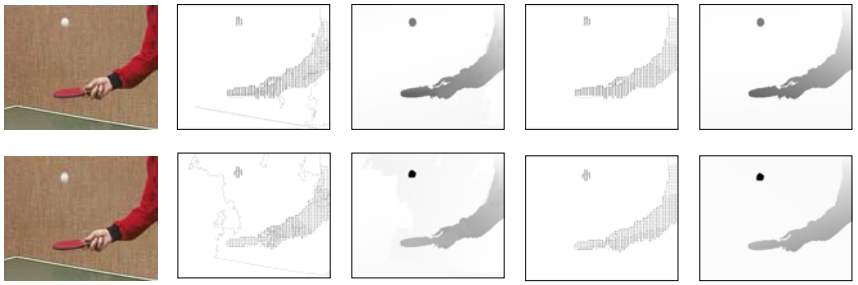


Robust motion estimation | Texture partition and affine motion estimation

## Homogeneity based spatial/temporal segmentation (XIV)

Example of combination of successive criteria (V):

- Motion estimation and merging on a texture homogenous partition:
  *Homogeneity based texture/motion spatial/temporal segmentation*



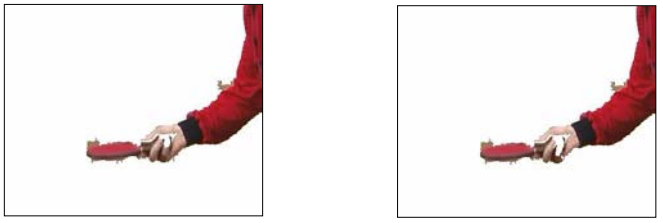
Texture/Motion partition (20 regions) Texture/Motion partition (3 regions)

## Homogeneity based spatial/temporal segmentation (XV)

Example of combination of successive criteria (VI):

- Motion estimation and merging on a texture homogenous partition:
  *Homogeneity based texture/motion spatial/temporal segmentation*

Use of long term motion information (Example with 3 regions)



Single-frame motion similarity Long-term motion similarity

# References


- Region-based analysis
  - P. Salembier and F. Marqués. Region-based representations of image and video: Segmentation tools for multimedia services . IEEE Trans. on Circuits and Systems for Video Technology, 9(8):1147-1169, December 1999.
- Binary Partition Tree
  - P. Salembier, L. Garrido. Binary Partition Tree as an Efficient Representation for Image Processing, Segmentation, and Information Retrieval, IEEE Transactions on Image Processing, 9(4):561-576, April 2000.
- Transition-based spatial/temporal segmentation
  - M. Pardàs, E. Sayrol. Motion estimation based tracking of active contours, Pattern Recognition Letters, 22:1447-1456, 2001
- Partition projection and object tracking
  - F. Marqués, V. Vilaplana. Face segmentation and tracking based on connected operators and partition projection. Pattern Recognition, 35(3):601-614, 2002.
- Homogeneity based spatial/temporal segmentation
  - C. Dorea, M. Pardàs, F. Marqués, A Motion-based Binary Partition Tree Approach to Video Object Segmentation, Proceedings of the ICIP-05, pp. 430 - 433, Genova, Italia, September 2005.
  - C. Dorea, M. Pardàs, F. Marqués, Generation of long-term color and motion coherent partitions, Proceedings of the ICIP-06, Atlanta, USA, octubre de 2006


# Region based model

I. Introduction
- Image sequence segmentation
- Segmentation steps:
  - Feature selection
  - Decision space

II. Transition based techniques
- Transition based temporal segmentation
  - Shot detection
- Transition based spatial segmentation
  - Rough object detection
- Transition based spatial/temporal segmentation
  - Snakes

III. Homogeneity-based techniques
- Homogeneity based temporal segmentation
  - Shot detection
- Homogeneity based spatial segmentation
  - Region matching
- Homogeneity based spatial/temporal segmentation
  - Robust motion estimation

IV. Summary