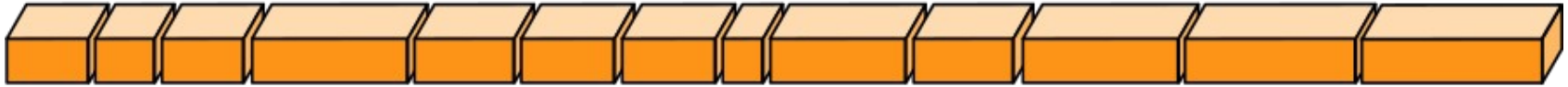
Video segments

Video structure

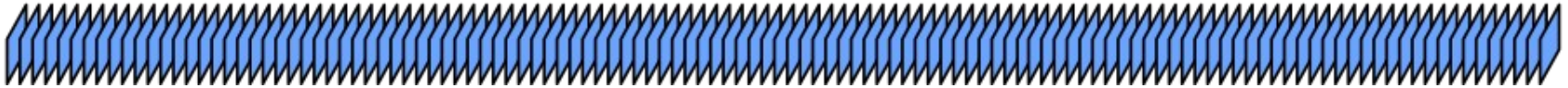
Video/movie - sequence of scenes

Scene - sequence of shots that form a semantic unit

Shot - sequence of frames from begining to the end of camera recording



Frame - single image

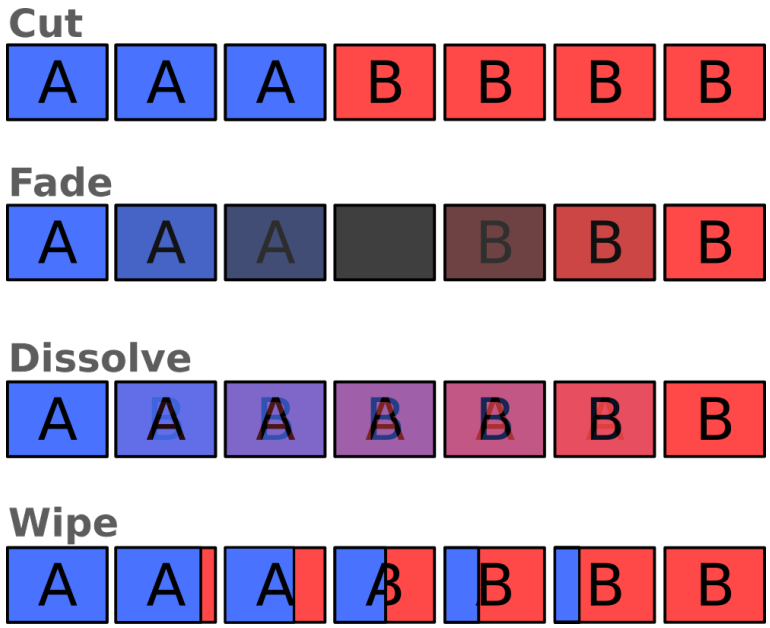


Video as sequence of shots

- Shots are useful start to detect scenes
  - Grouping shots into semantic units
  - Enable semantic retrieval in video
- Manual segmentation of video into shots is slow
  - About 10 hours per 1 hour of video (for a movie)
  - Easier if edit decision list is available (unreliable)
- Automatic detection of shots
  - Detecting boundaries - transitions

Transition types

- Cut
  - Sharp transitions between shots
  - Sudden change of all pixels in the frame
- Fade
  - Fade-out – gradual transition to color
  - Fade-in – gradual transition from color
  - Dissolve – gradual transition between shots
  - Wipe – gradual erase

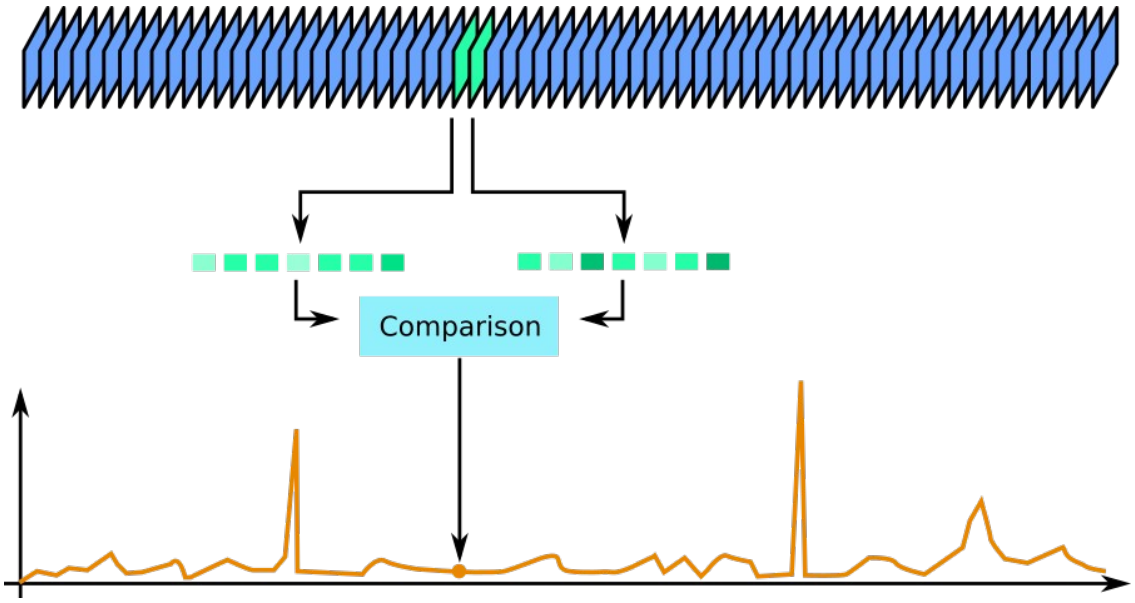


Detecting transitions

- Describe frame content
  - Features: color, texture, edges, etc.
- Measure difference
  - Two frames
  - Multiple frames
- Difference large enough
  - Threshold
  - Adaptive measures

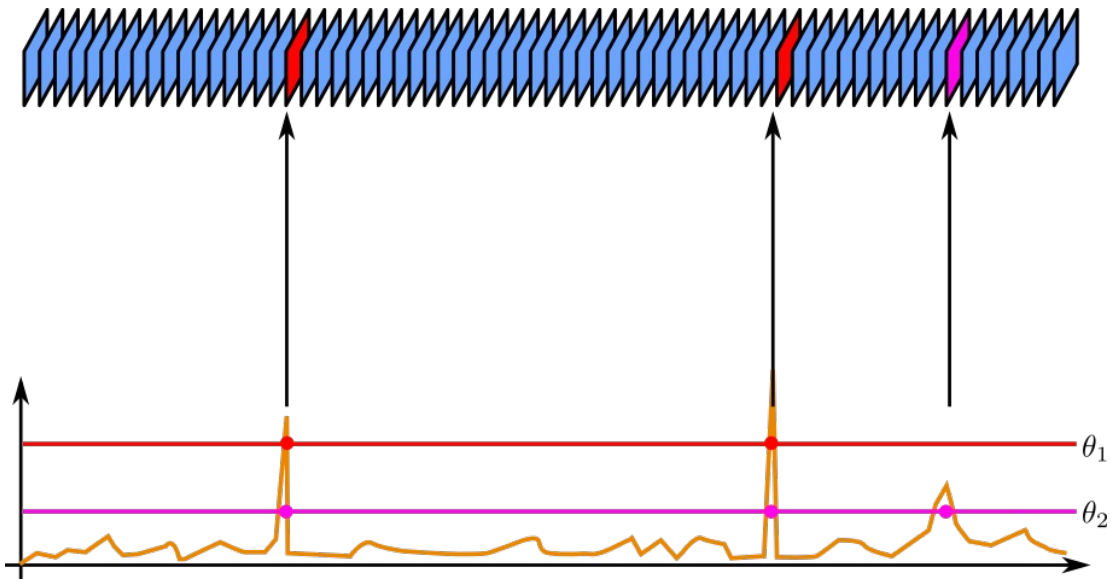
Detecting cuts

- Assumptions
  - Almost stationary
  - Almost constant scene
  - Constant illumination
- Cut if significant change
  - Color
  - Intensity
- Descriptors
  - Gaussian model
  - Histograms



Setting a threshold

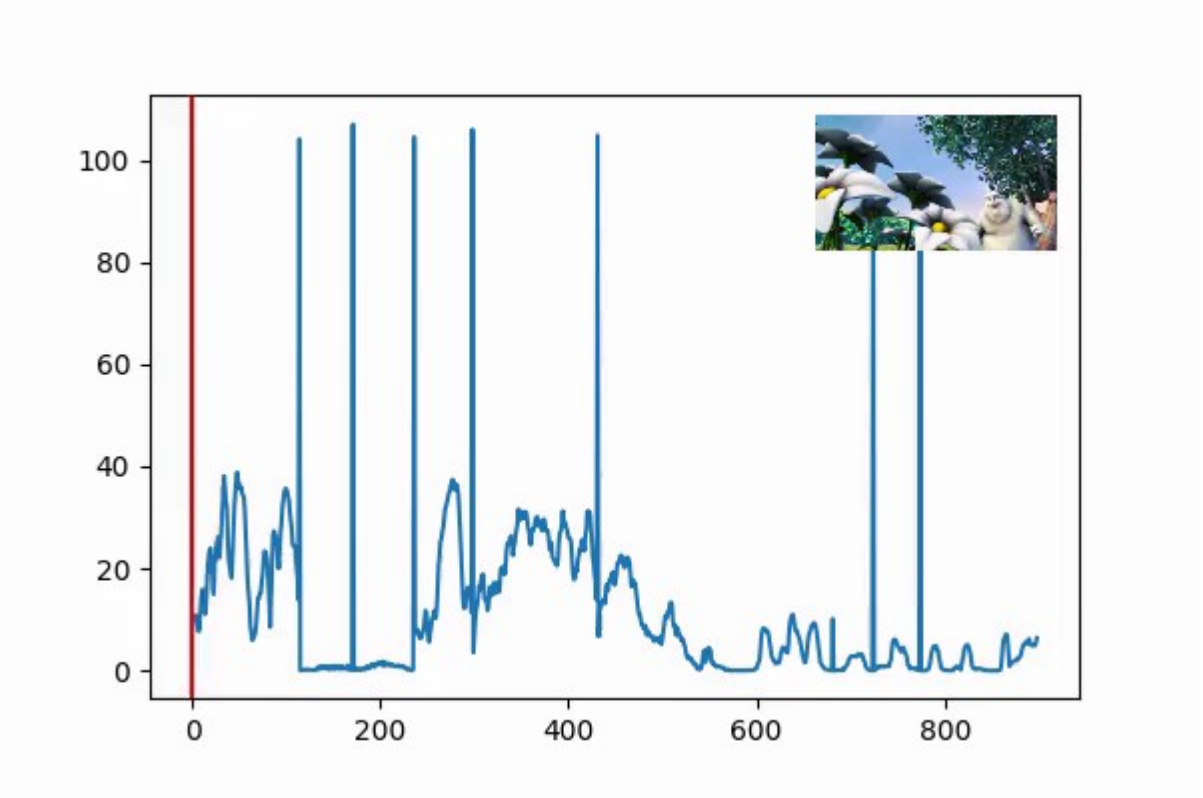
- Distance between consecutive frames
- How to set cut detection threshold?
  - Global methods
  - Adaptive methods



Y. Yusoff, W. Christmas, J. Kittler, „Video Shot Cut Detection Using Adaptive Thresholding“, BMVC 2000

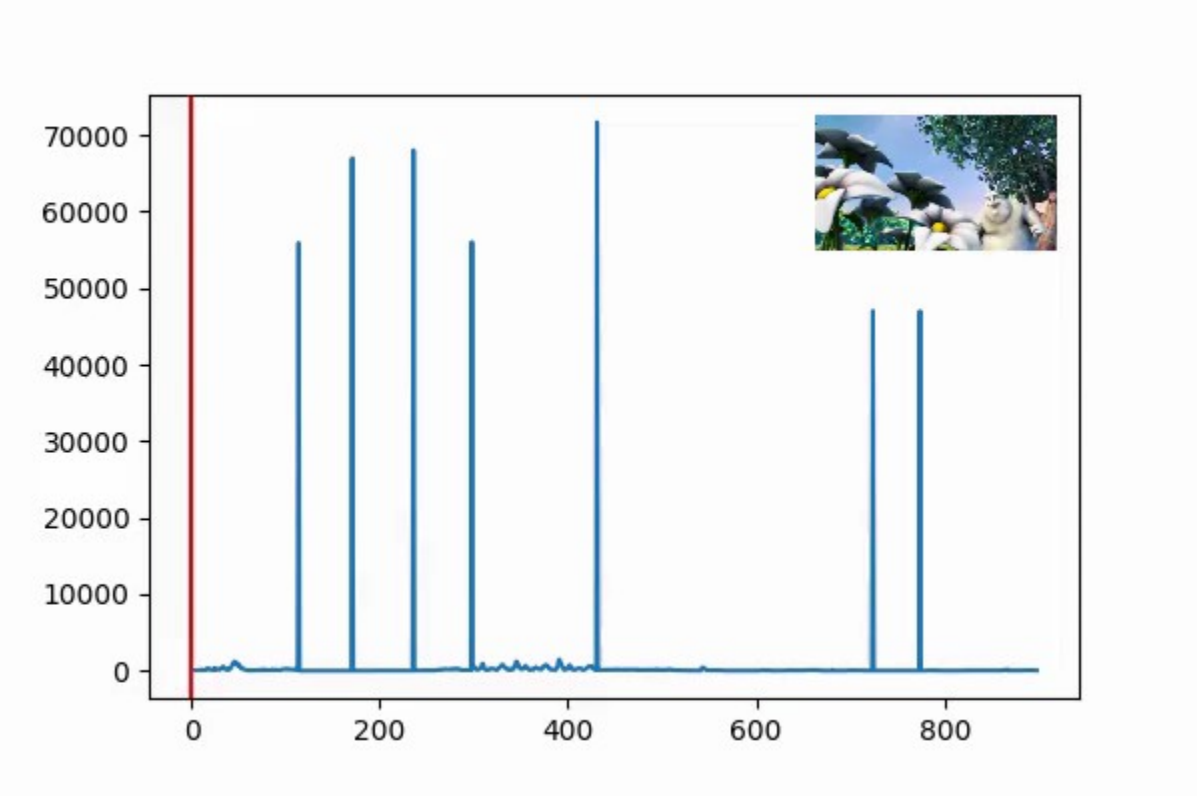
Detecting cuts with MSE

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Xi - Yi)^2$$



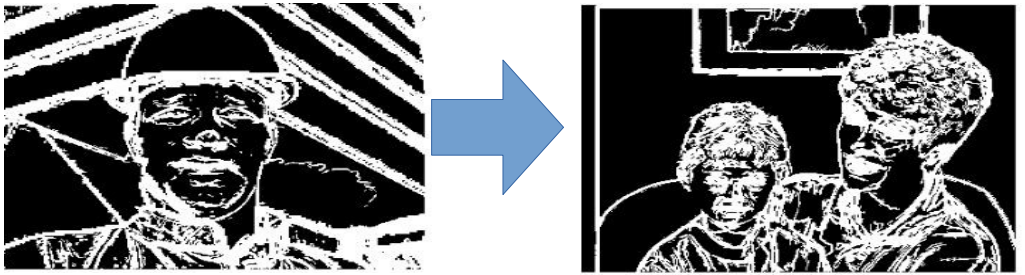
Detecting cuts with histograms

$$X^2 = \frac{1}{2}\sum_{i=1}^{B}\frac{(x_i - y_i)^2}{(x_i + y_i)}$$



Detecting cuts with edges

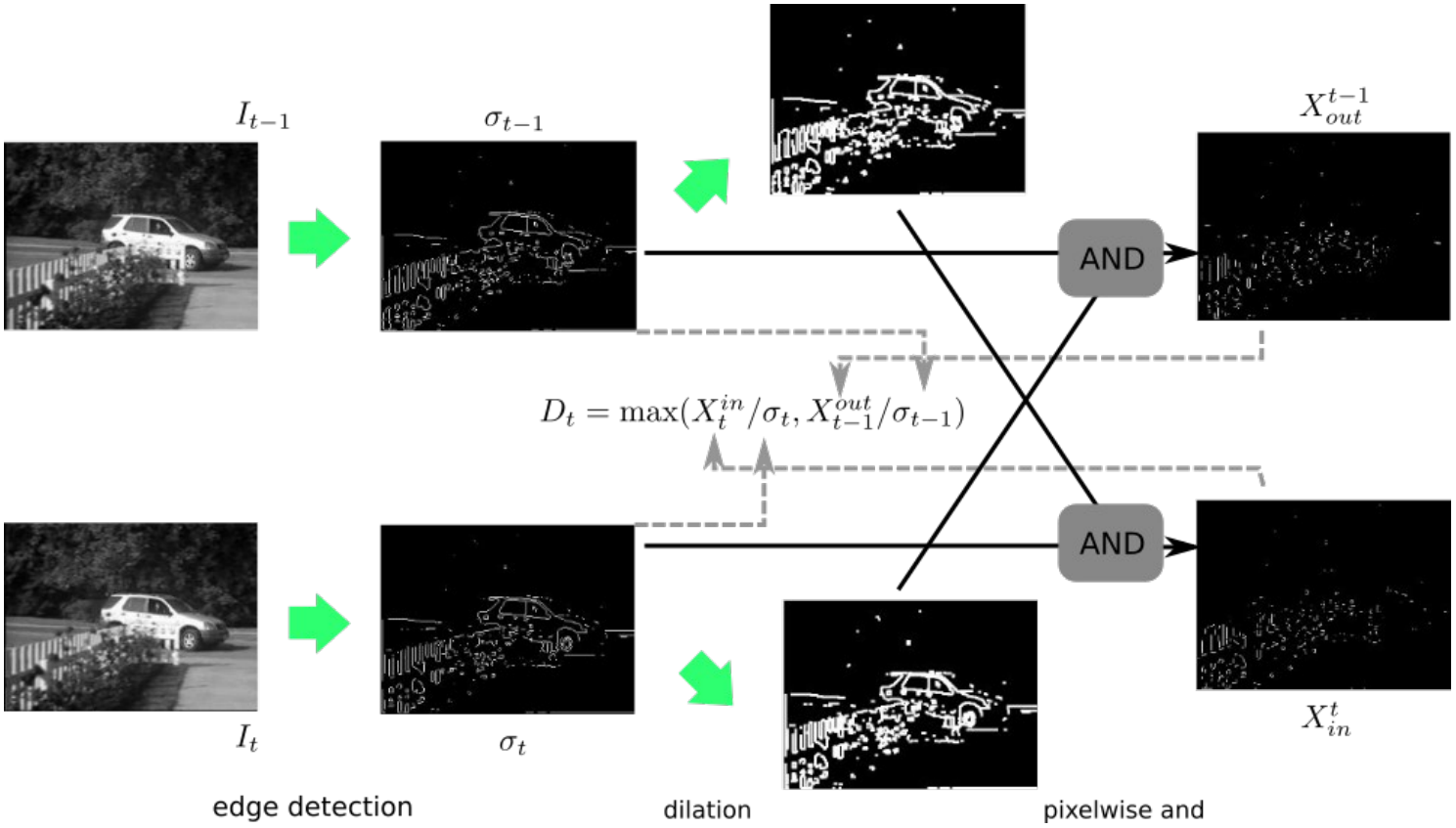
- Color methods are not robust to illumination changes
- Compare edge pixels
  - How many appeared
  - How many vanished



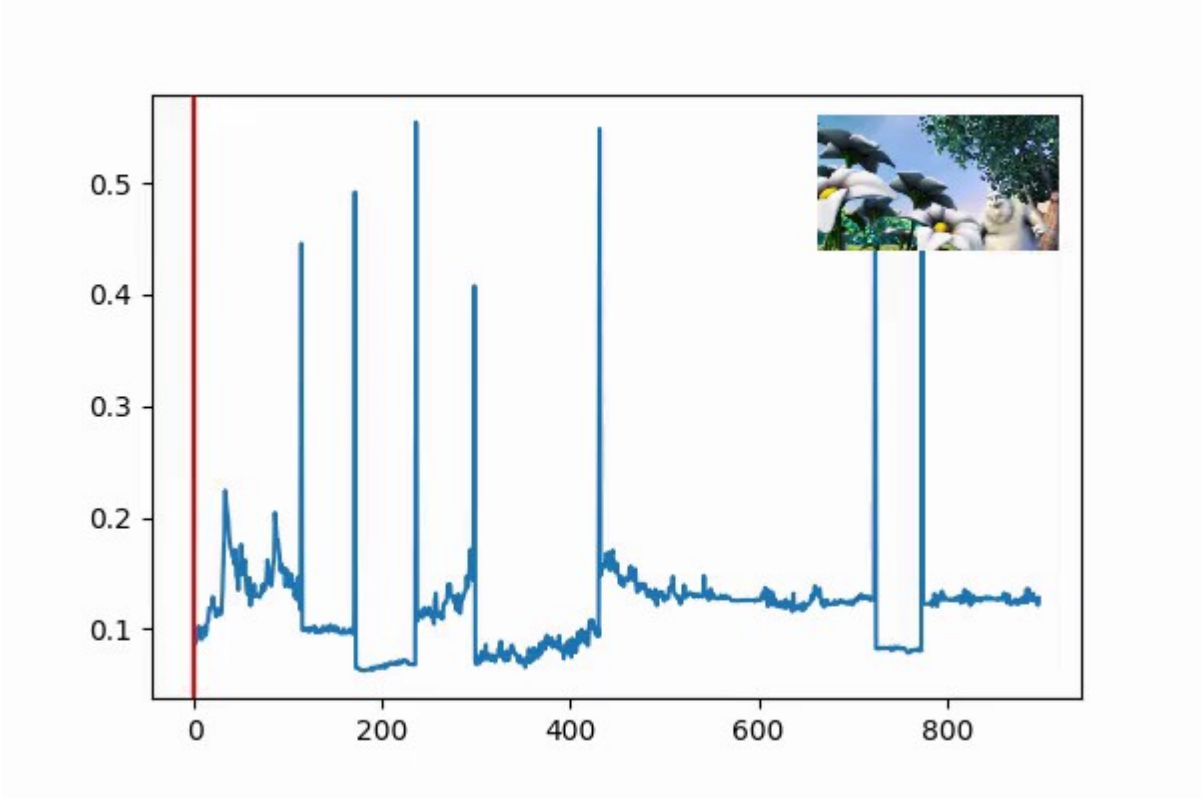
$$D_t = \max(X_t^{in}/\sigma_t, X_{t-1}^{out}/\sigma_{t-1})$$

$X_t^{in} \cdots$ number of new edges at time t
$X_{t-1}^{out} \cdots$ number of vanished edges at time $t-1$
$\sigma_t \cdots$ number of all edges at time t
$\sigma_{t-1} \cdots$ number of all edges at time $t-1$

Algorithm


I_{t-1}
σ_{t-1}
X_{out}^{t-1}
AND
$D_t = \max(X_t^{in}/\sigma_t, X_{t-1}^{out}/\sigma_{t-1})$
AND
I_t
σ_t
X_{in}^t
edge detection
dilation
pixelwise and

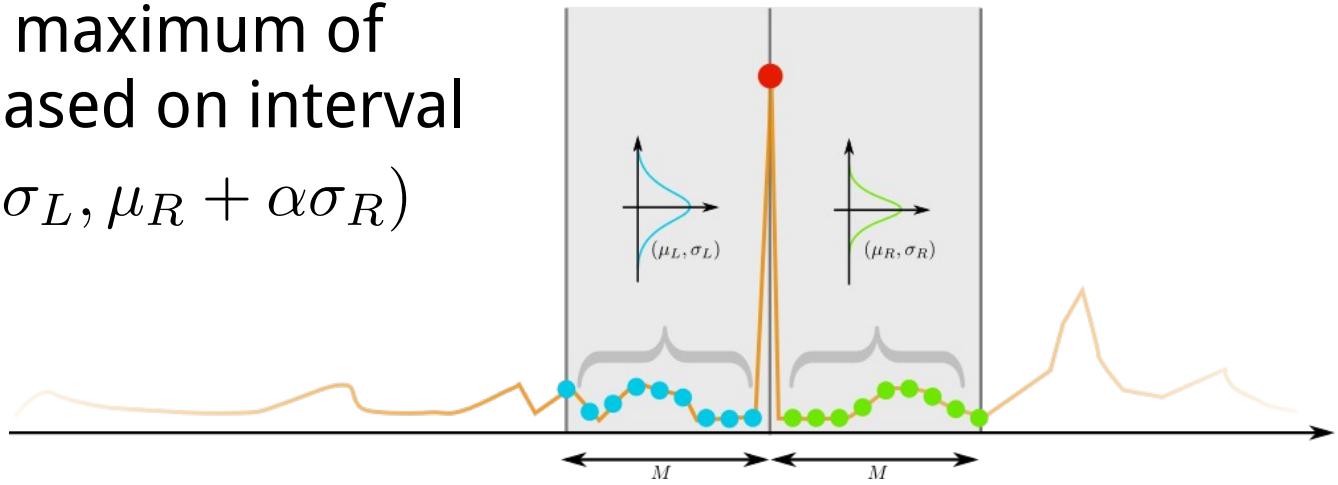
Detecting cuts with edges



Adaptive threshold

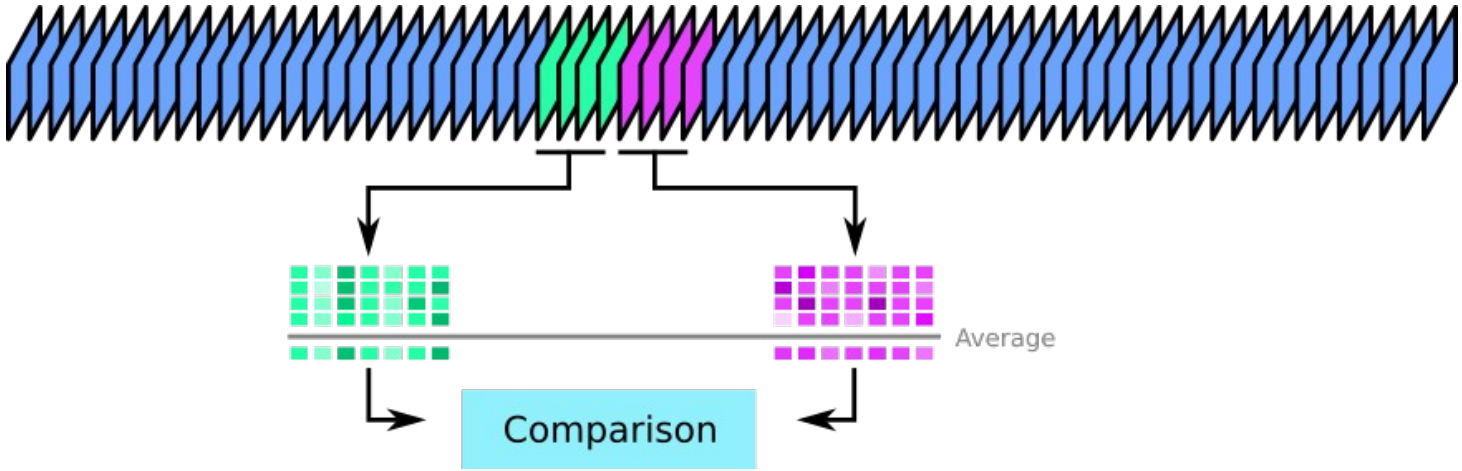
- Cut changes result in sharp peaks
- Frame t is a cut frame if D_t
  - is the largest in interval $[t-M, t+m]$
  - is larger than the maximum of scaled variance based on interval

$$D_t > \max(\mu_L + \alpha\sigma_L, \mu_R + \alpha\sigma_R)$$



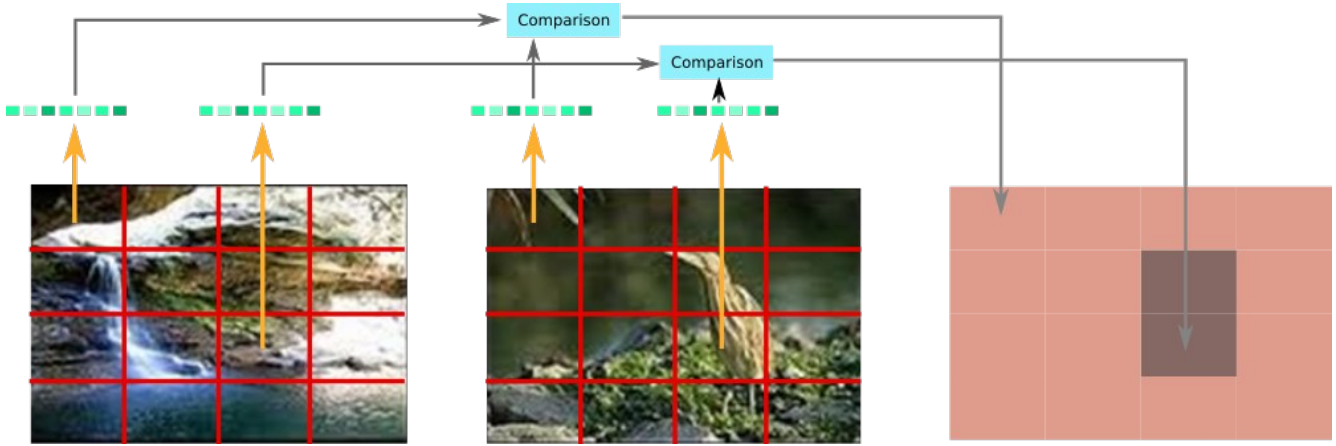
Temporal averaging

- Not enough or too much change between two frames
- Average several consecutive descriptors



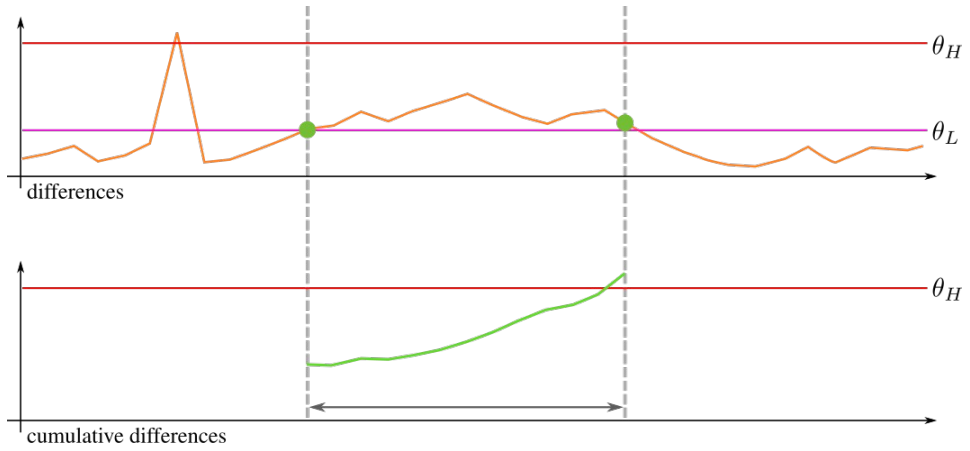
Partial changes

- Global descriptors do not consider locality of changes
- Compute distances between frames for blocks
  - Ignore change if less than N blocks change
  - Compute overall distance



Detecting fades

- Not a lot of change between two frames
- Two stage threshold
  - Low threshold – potential fade start
  - Comparing to the start frame
  - Measure difference until it is increasing
  - Compare to the high threshold



B. Shahraray, “Scene change detection and content-based sampling of video sequences” Digital Video Compression: Algorithms and Technologies, 1995