A Simple and Easily Parallelized Video Copy Detection Method

G. Roth, R. Laganière, M. Bouchard, T. Janati, I. Lakhmirie

School of Information Technology and Engineering (SITE)
University of Ottawa, Ottawa ON Canada

uOttawa

L'Université canadienne
Canada's university

Video Copy Detection

- Useful alternative to watermarking
- A problem with many possible solutions
  - TrecVid helps in evaluation, but is not enough
  - Need some more evaluation criteria
- Our goals: Small amount of index info per frame, search efficiently, effectively and have search process easy to parallelize

uOttawa

L'Université canadienne
Canada's university

Alternative Approaches

- Global methods
  - Descriptor from global image characteristics
  - Compact, but difficult to make effective
- Local methods
  - Find local feature points (like SIFT)
  - Effective, but difficult to make compact

Combine local and global

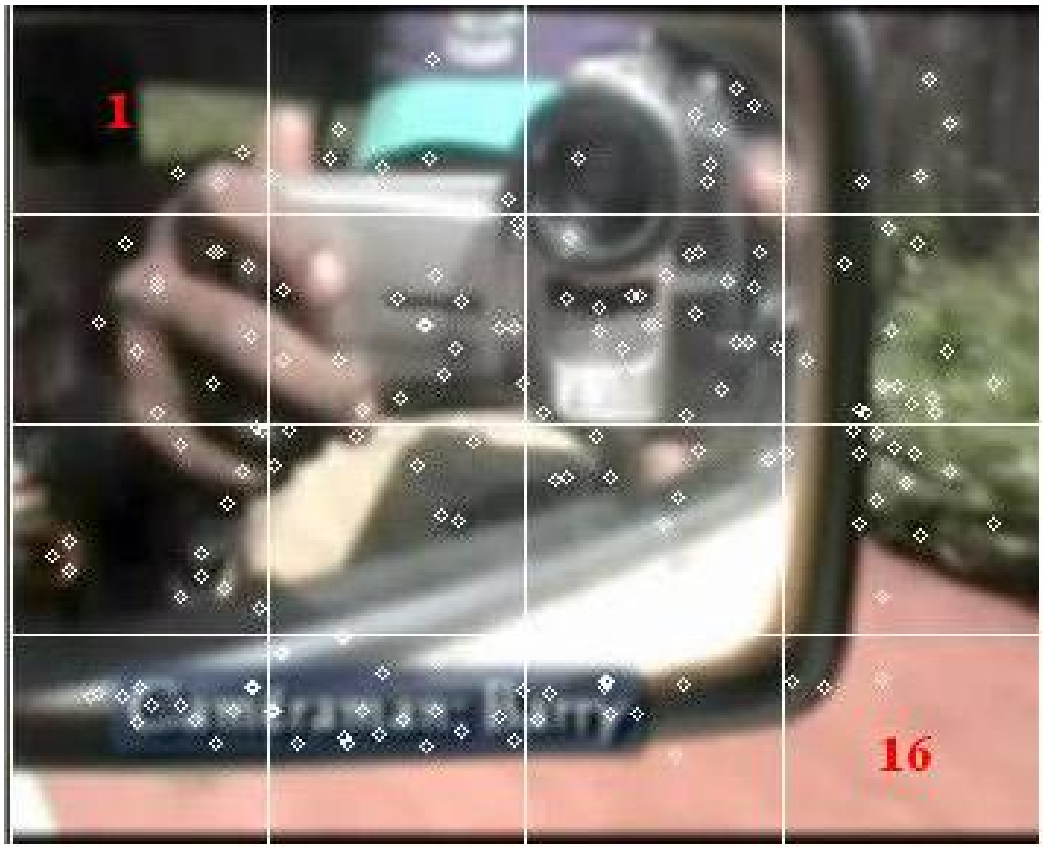
- Find all the SURF feature points in a frame
- Divide image into 4 by 4 regions
- Count feature points in each of these regions
- Descriptor for each frame is the count of the number of feature points (less than 256)
- Have a 16 byte descriptor for a video frame

uOttawa

L'Université canadienne
Canada's university

Descriptor is (1,6, …, 3)

- Tested 2x2, 4x4, and 8x8 descriptors



G. Roth 2009

What about other descriptors?

- Historically, ordinal measures are good global descriptors (invariant)
- First tried PACT, a recent ordinal descriptor

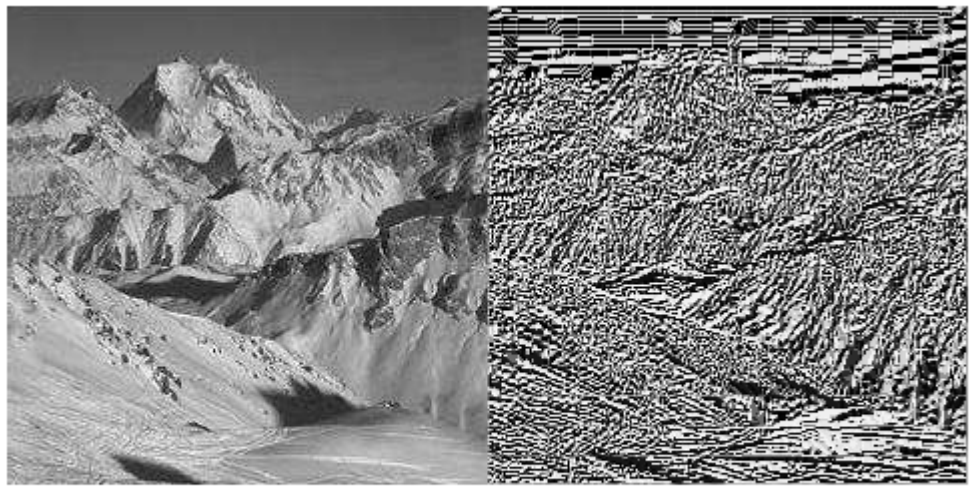
32	64	96		1	1	0
32	**64**	96	$\Rightarrow$	1	0	$\Rightarrow (11010110)_2 \Rightarrow \mathrm{CT} = 214$
32	32	96		1	1	0

uOttawa

L'Université canadienne
Canada's university

PACT Ordinal Descriptor

- Transform byte => byte for entire image
- Descriptor not compact nor effective?

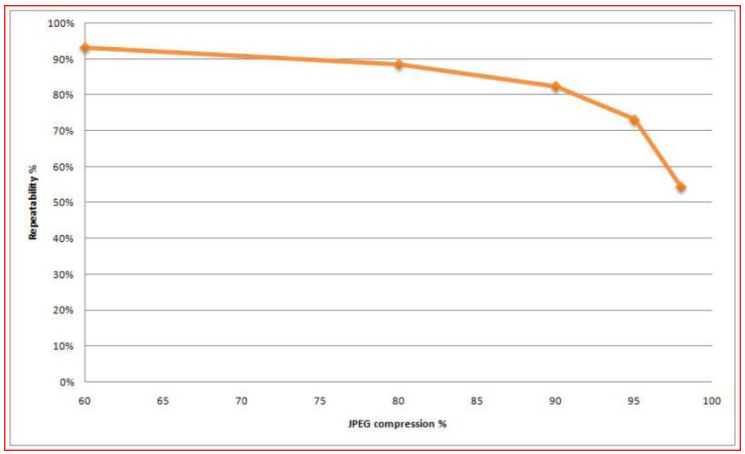


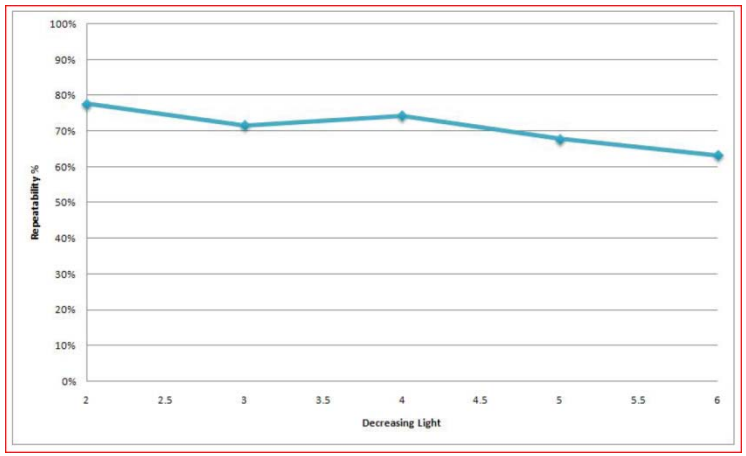
(a) mountain image (b) Transformed image

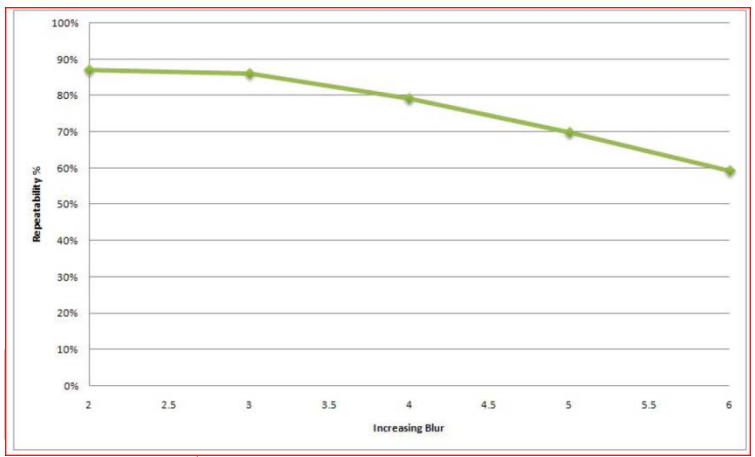
SURF Feature Points

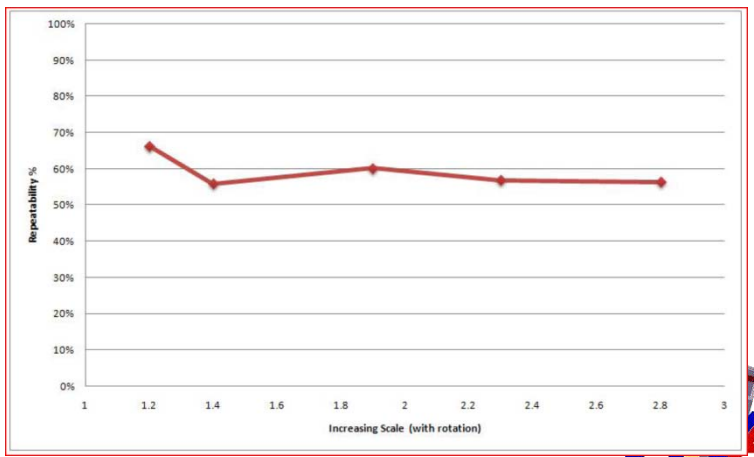
- Finds features (interest points) in an image

SURF Characteristics









Advantages of our descriptor

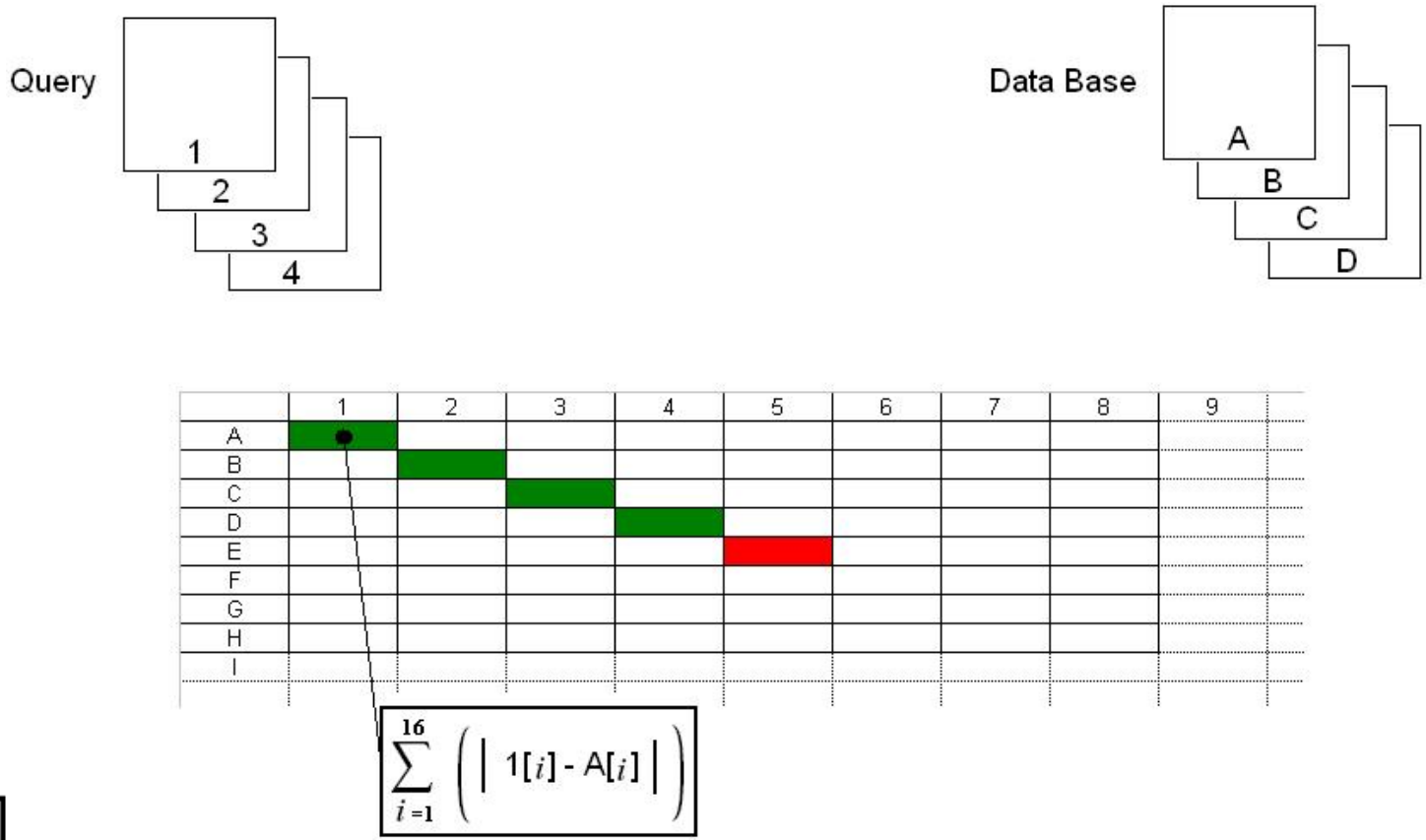
- Feature counts are very compact
  - 90,000 frames in an hour of video requires only 1,440,000 bytes (1.44 mbytes)
- Is effective
  - Use natural invariance of the SURF features
  - In video we compare a sequence of descriptors so we do not need a more powerful descriptor

uOttawa

L'Université canadienne
Canada's university

VIVA Lab

Comparing descriptors



Skipping bad matches

	A	B	C	D
1	S<thresh			
2		S<thresh		
3			S<thresh	
4				S>thresh

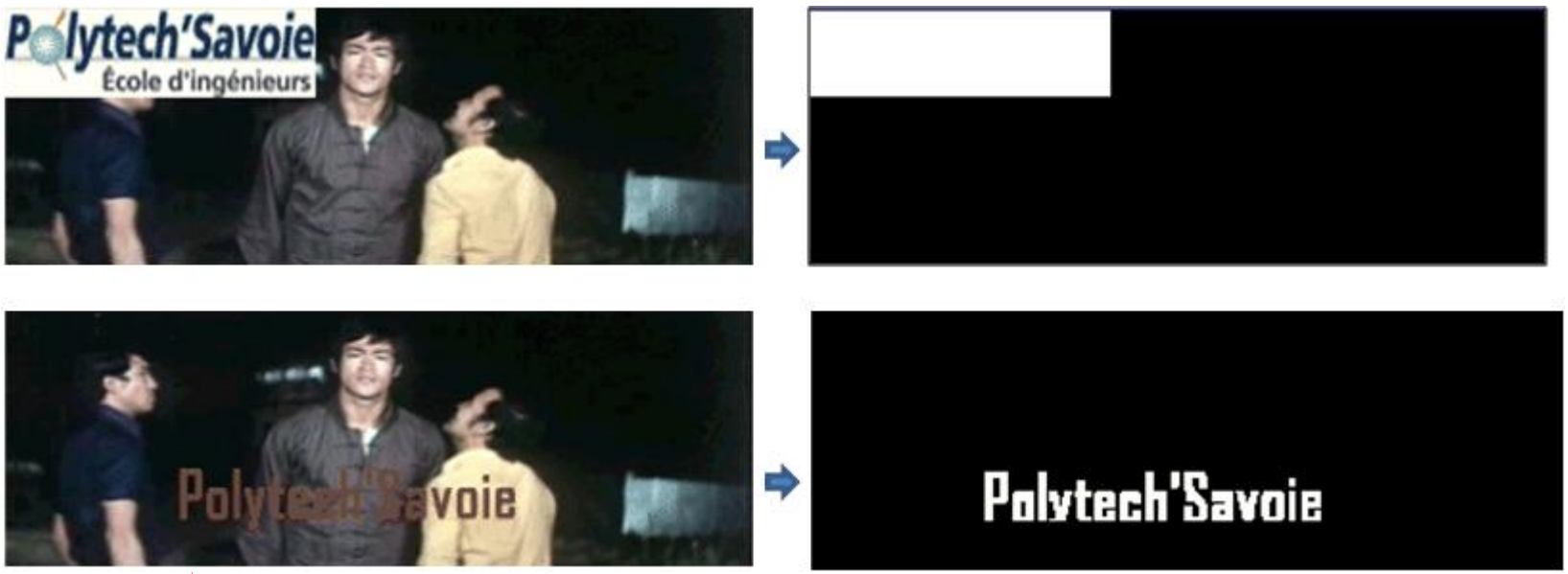
Creating masks

$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n}$$

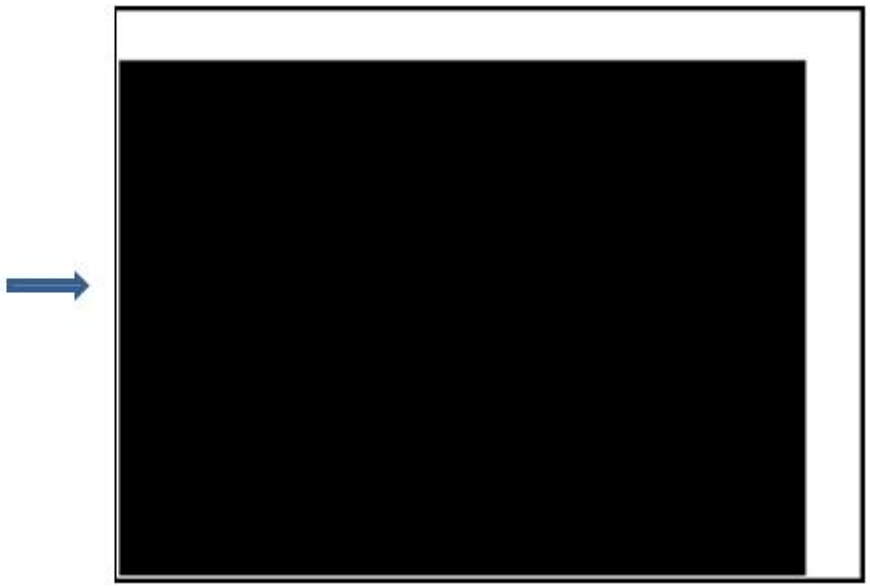
When σ = 0 draw the pixels as white

Otherwise draw the pixels as black

Text Insertion Mask



Shift Mask



Mirror Transform

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

→

4	3	2	1
8	7	6	5
12	11	10	9
16	15	14	13

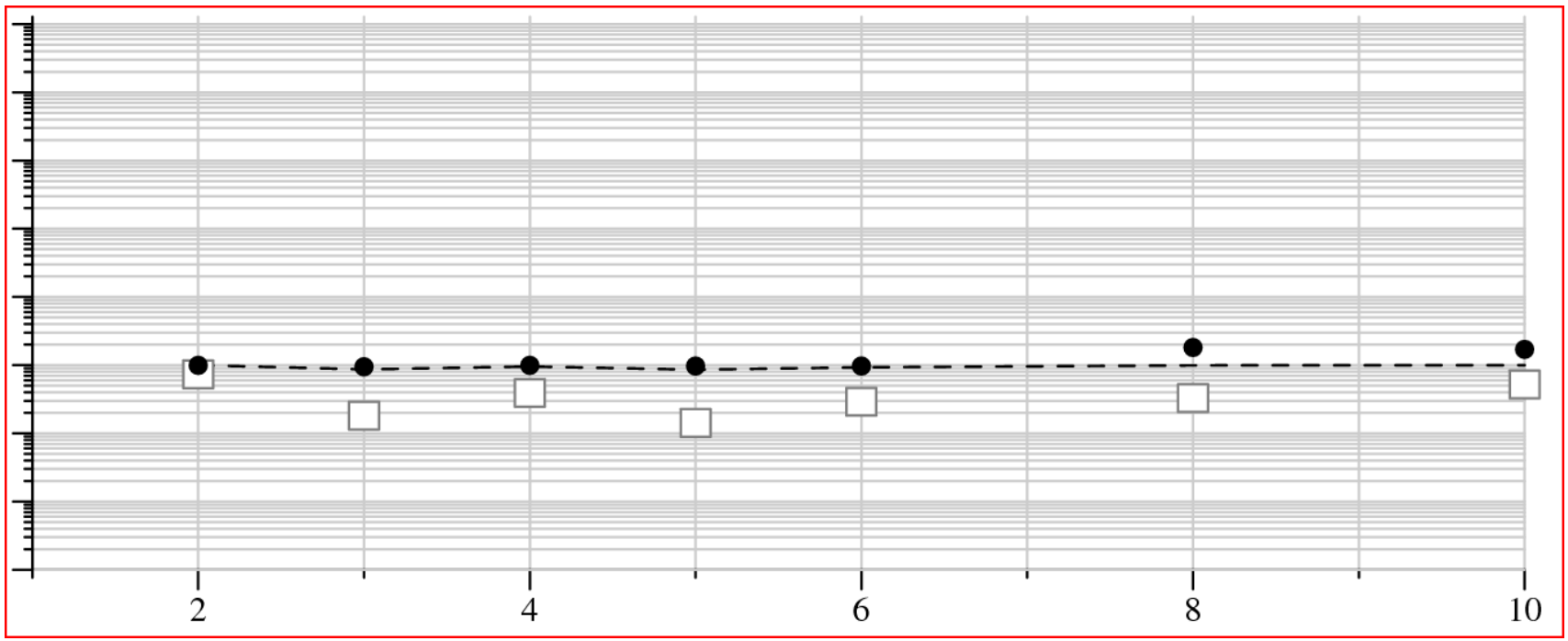
Audio Matching

- Based on coherence function using intermediate features in ITU-R BS.1387 Perceptual Evaluation of Audio Quality
- Idea of using PEAQ features was to include psychoanalytic effects such as critical bands, frequency masking and loudness

Performance

- Video only – NDCR around the median, while the F1 (localization) is near the top
- Audio only – slightly worse than median NDCR, low false pos., but high false neg.
- Combined – audio only boots the video, not very good results (not clear why?)

Video NDCR – Balanced Insert



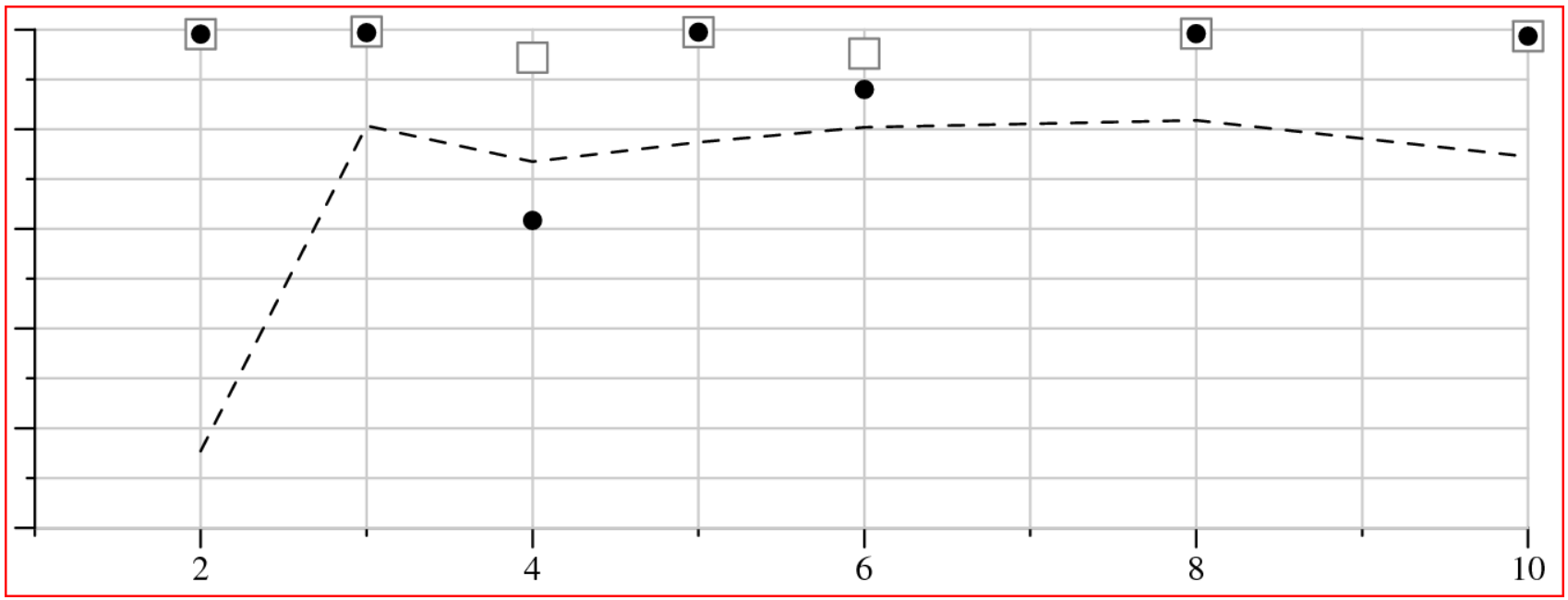
uOttawa

L'Université canadienne
Canada's university

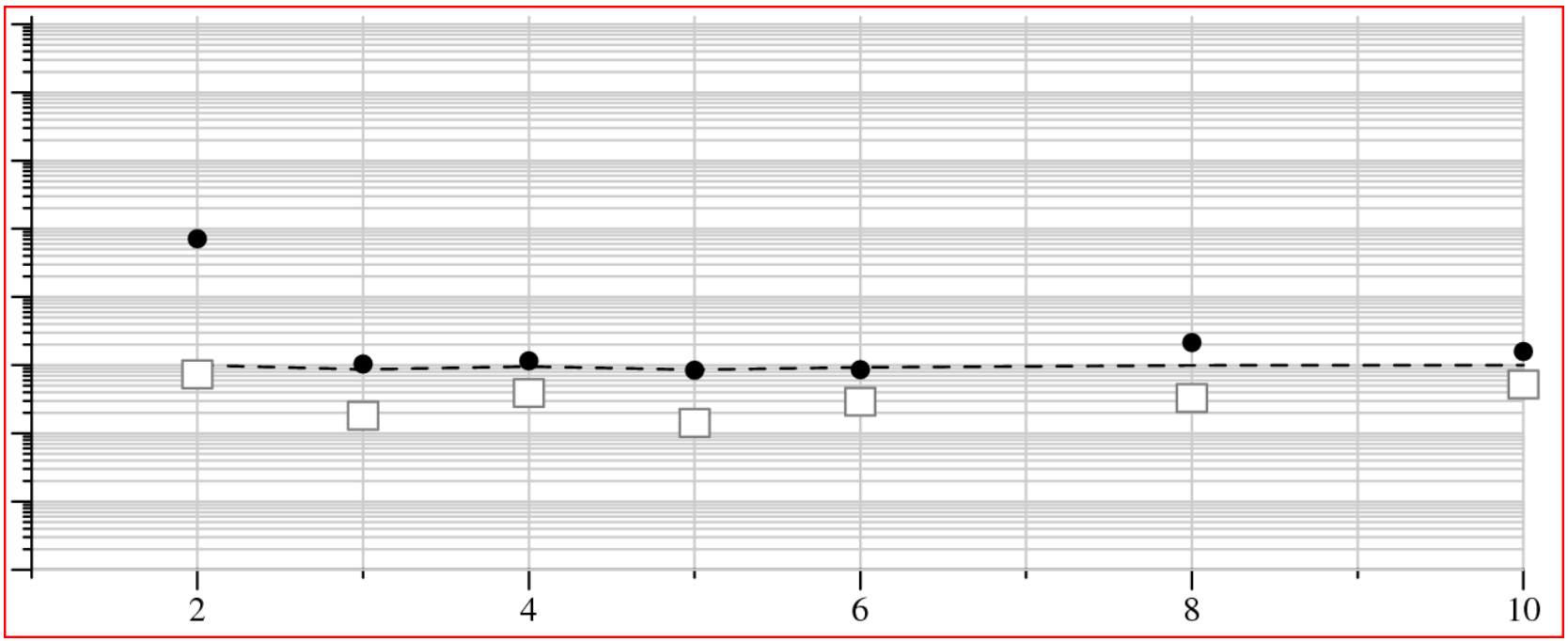
G. Roth 2009

VIVA Lab

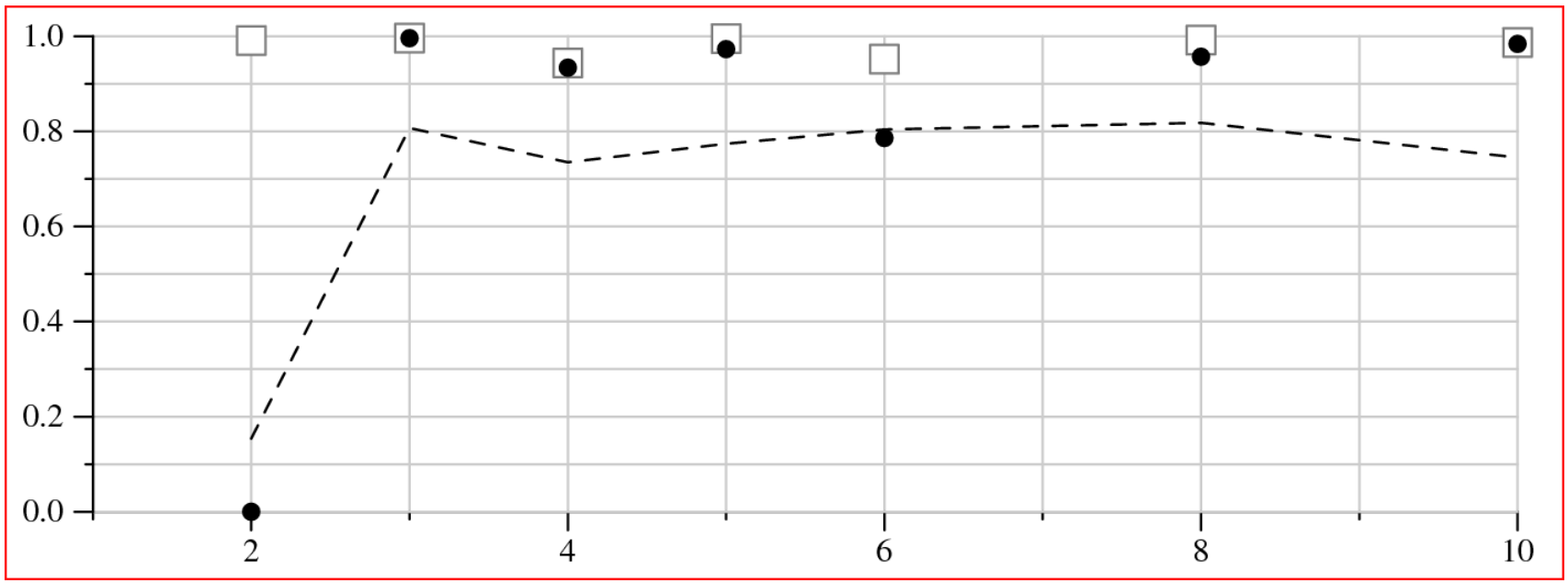
Video F1 – Balanced Insert



Video NDCR – Balanced No Insert



Video F1 – Balanced No Insert

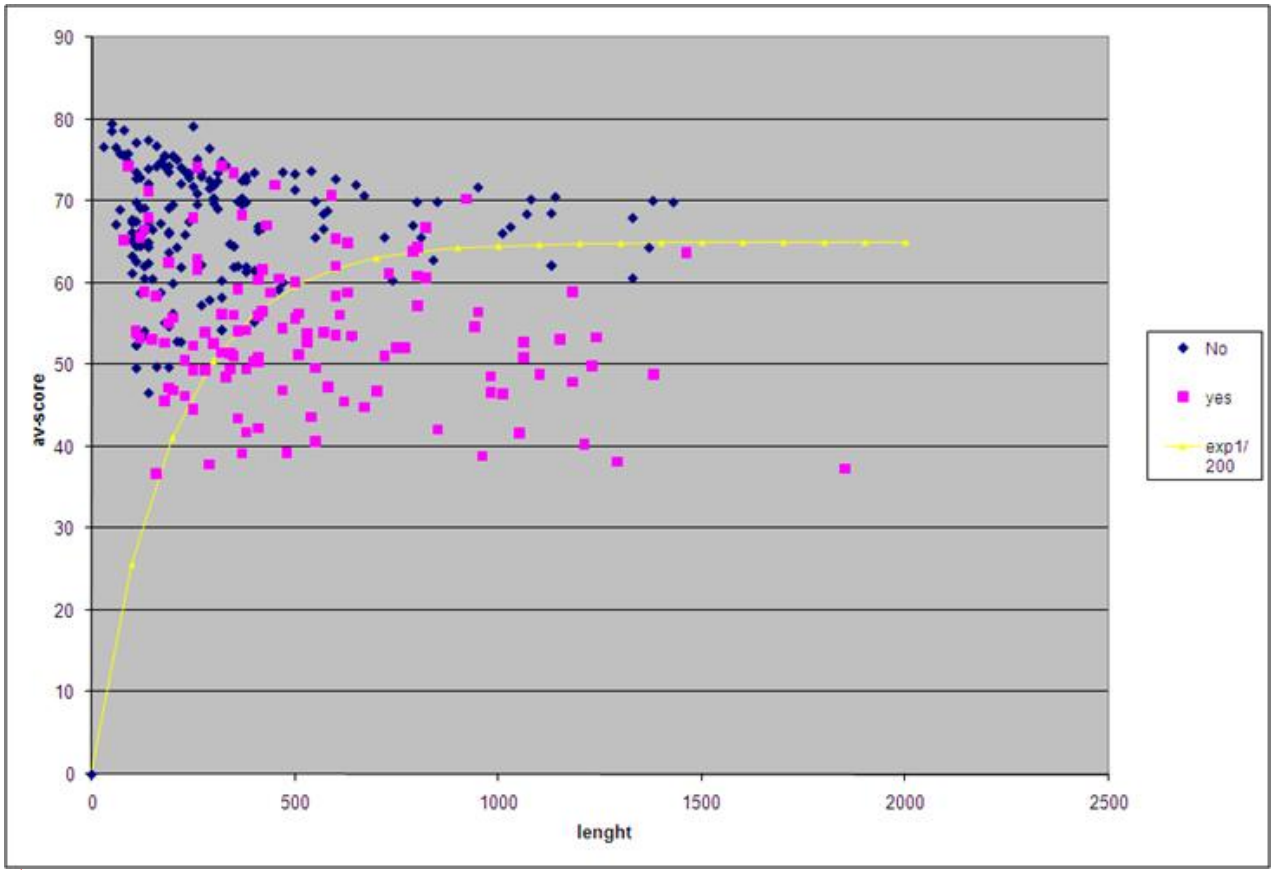


uOttawa

L'Université canadienne
Canada's university

VIVA Lab

Improved Thresholding



Parallel Processing

- Algorithms must run on parallel hardware
- What is ease of parallelization?
  - Best if no reprocessing is necessary for a different assignment of dbase files to processor
  - If you have intermediate data structures (like tree or hash table, then this not the case)
  - Our method allows trivial parallelization

uOttawa

L'Université canadienne
Canada's university

Future Work

- Implement parallelization on GPUs
- Better combination of audio and video
- Better decision thresholding (as described)
- Different feature points with this approach
  - Use real-time feature extraction (like Harris) for on-line commercial removal (simple transform)
  - Detect many commercials in real-time

uOttawa

L'Université canadienne
Canada's university