Experiments from paper on Hierarchical Video Segmentation

February 17, 2016

Original paper:

Streaming Hierarchical Video Segmentation

Chenliang Xu , Caiming Xiong and Jason J. Corso

Further Experiments and Presentation: Kim Houck

Using code made available by original paper authors

Overview

- Basics of Hierarchical Video Segmentation
- Exploration of segment size on performance
- Effects of video resolution on runtime

Hierarchical Video Segmentation

- Video segmentation – image segmentation through time
  - Much more data to process
  - Consistent structure over time
- Hierarchical Segmentation merges similar regions through space and time at each layer

$$S = \underset{s}{argmin}\, E(s|video)$$

Streaming Hierarchal Segmentation

- A balance between processing whole video and frame by frame processing
- Breaks video into segments
- Uses Markov assumption

$$S = \underset{s_i}{argmin}\, E\left(s | V, S_{i-1}, V_{i-1}\right)$$

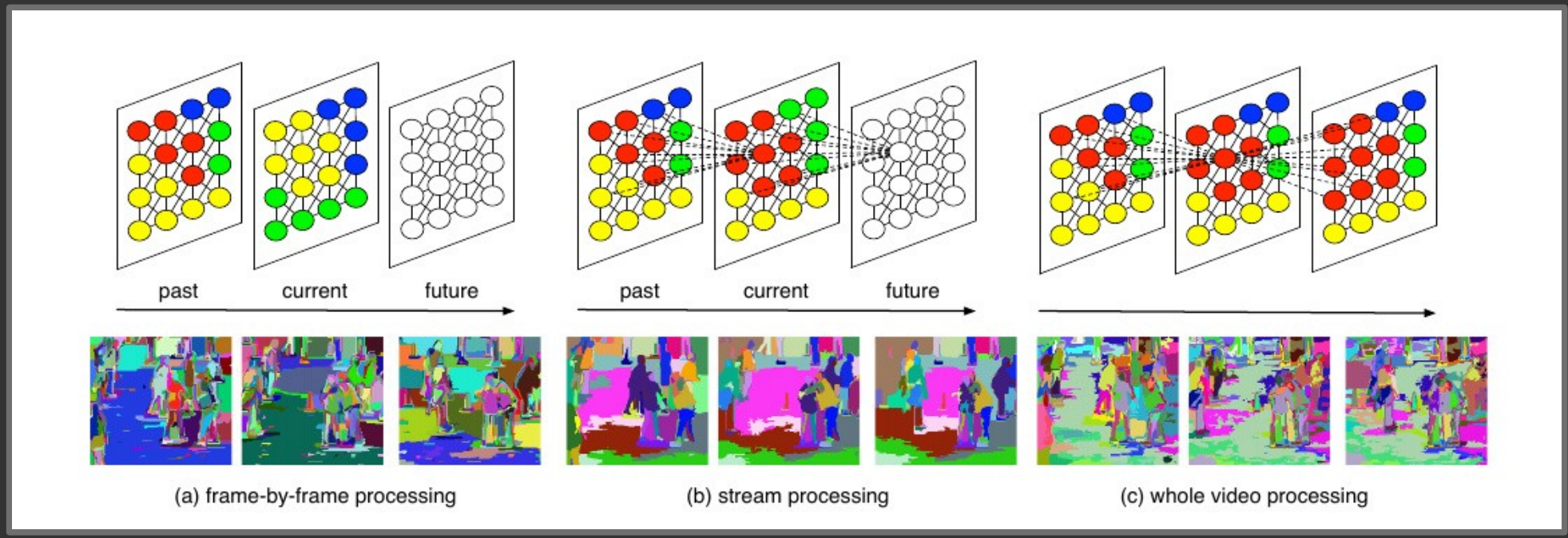


Figure: xu et al, 2012

Authors' Dataset

- 8 videos at 240x160 resolution

bus

container

garden

ice

paris

salesman

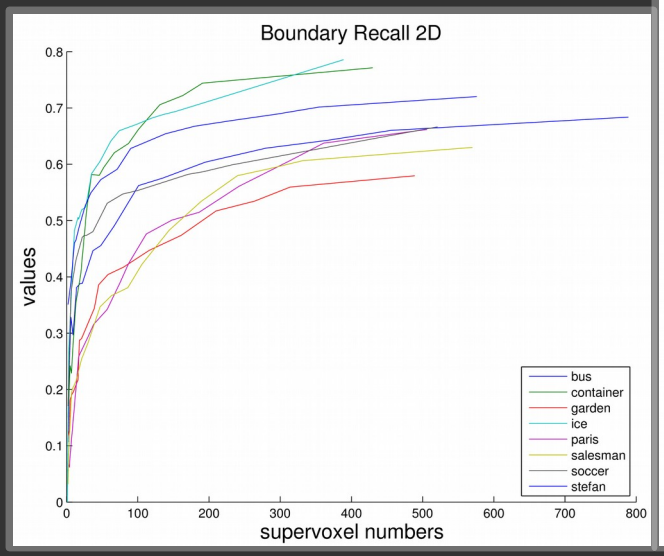
soccer

stefan

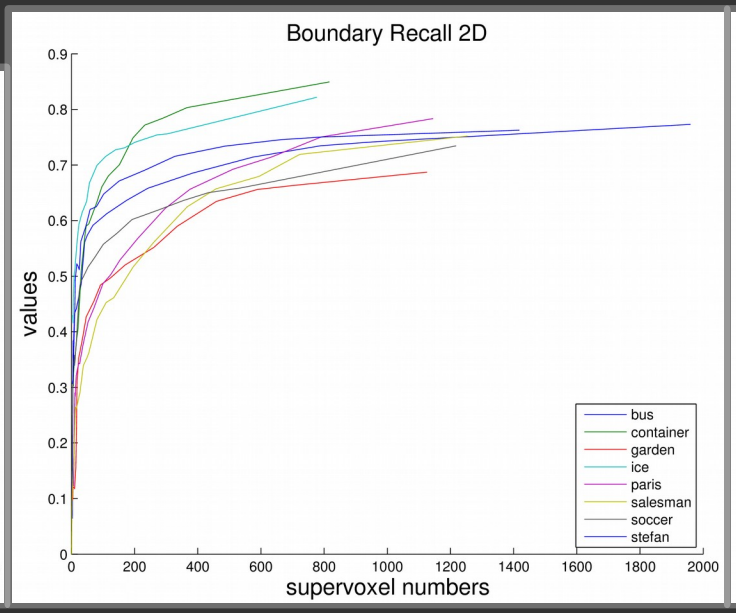
Effect of segment size

- Look at how segment size effects performance of GBH_Stream algorithm
- Documentation for libsvx recommends a sequence length of 10 frames
- Compare performance to that sequence lengths of 5 and 15 frames
- Use 8 videos from authors' dataset

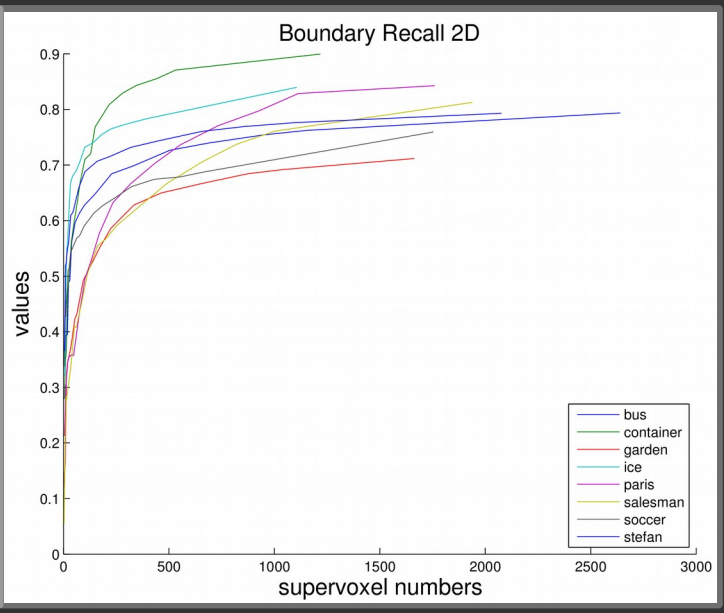
Boundary Recall - 2D



5 Frames

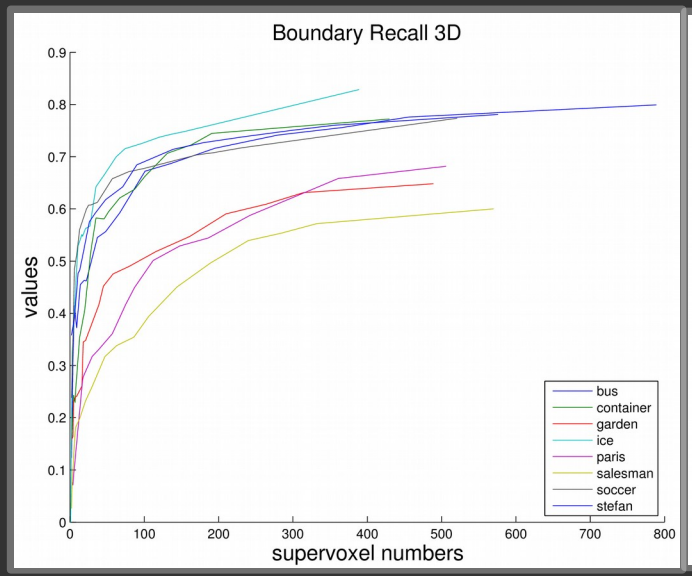


10 Frames

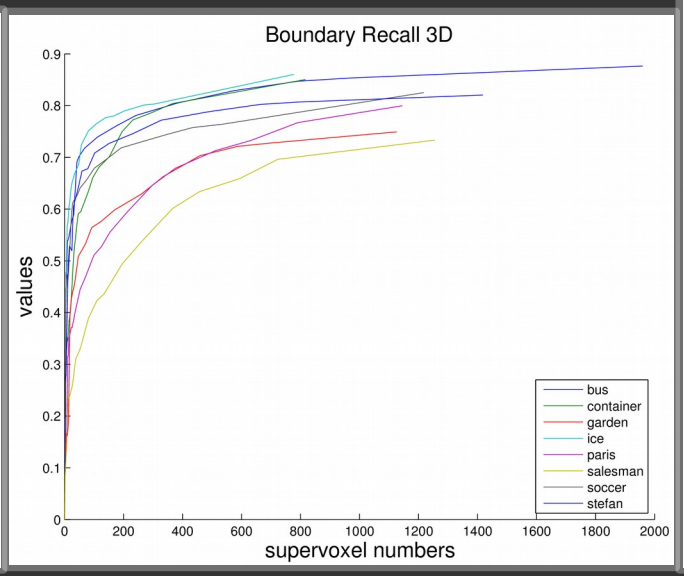


15 Frames

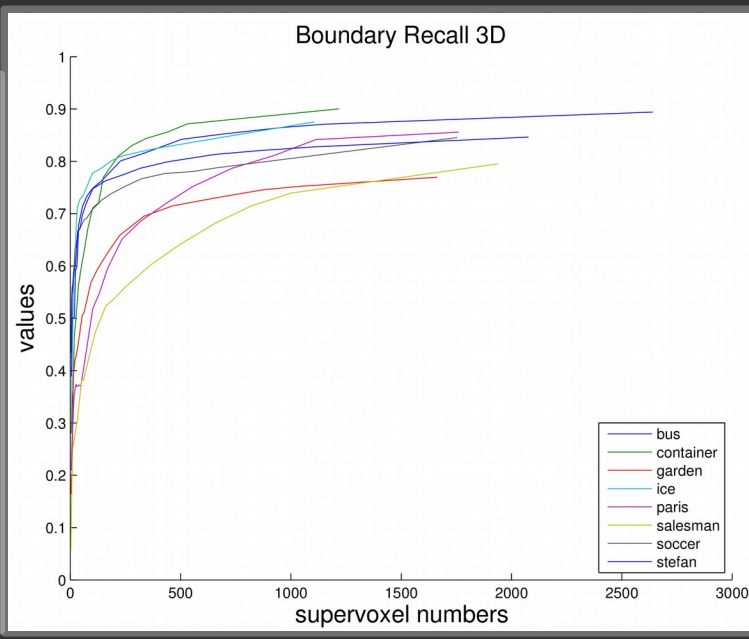
Boundary Recall - 3D



5 Frames

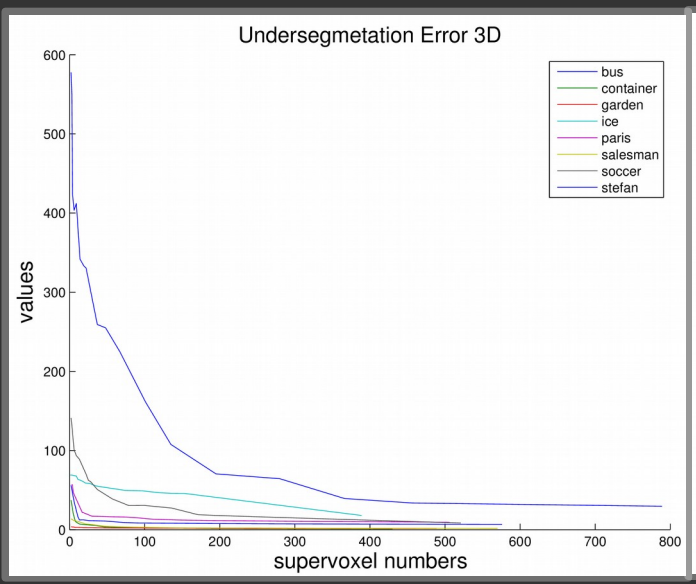


10 Frames

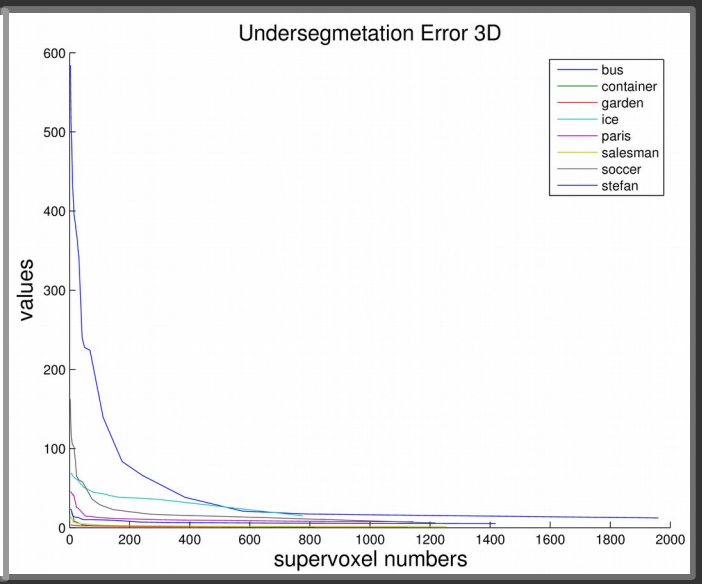


15 Frames

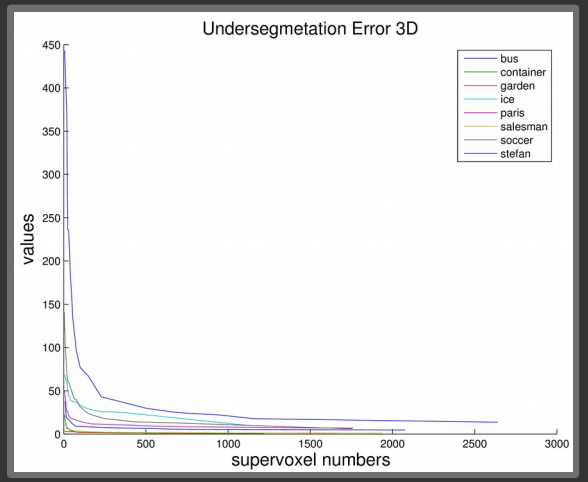
Undersegmentation error - 3D



5 Frames



10 Frames



15 Frames

Runtime on a longer/larger video

- Processing whole video at once better
  - Have the whole picture
  - Less info available when only (some) info previous to a frame is available
- Processing whole video at once often impractical
  - Too big to fit in memory
  - Not available yet (realtime processing)

Longer example video

- ~10 secs
- 246 frames
- 1920x1088 original resolution
- Test 240x136 and 480x272 resolutions

Runtime results

- 240x136: 8m 22s, 8m 28s
- 480x272: 35m 38s, 35m 37s
- Run on 3.5 GHz i7 (Haswell)
- Could not run larger sizes due to memory use

Qualitative Analysis

- This is a hard video
  - Very little contrast for main focus (dustdevil)
- Supervoxels merge after level 9 at 240x136
  - Still barely visible at level 20 at 480x272

Level 9 - 240x136

Level 18 - 480x272

Level 18

- 240x136 vs 489x272

Questions?