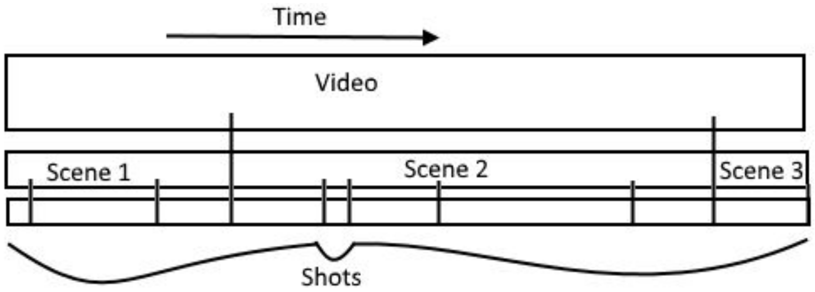
Video Segmentation

Ercan Alp Serteli

Video Segmentation

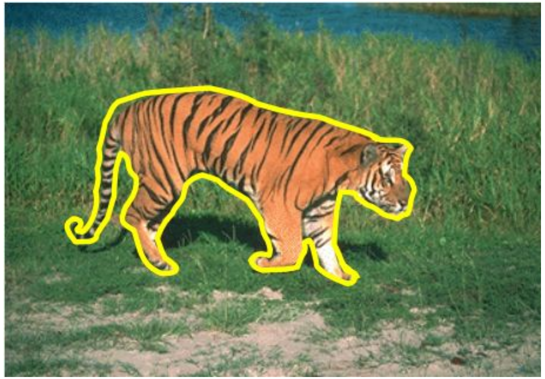
Temporal Segmentation

Partitioning in time



Spatial Segmentation

Partitioning in space, like image segmentation



[1]

Image vs. Video

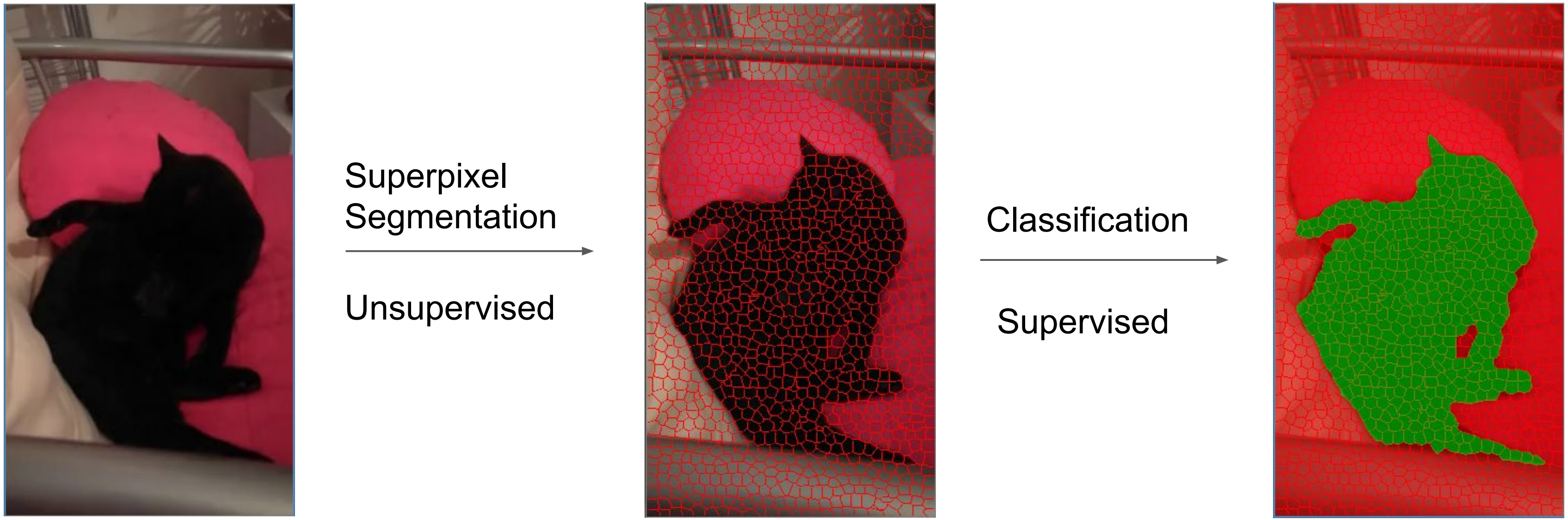
- Frames change continuously in time => New frame can use info from previous one

- Many frames per video => Needs to be performant

- Performance is even more critical in real time streaming applications

Approach Used in This Project

- One unsupervised step, one supervised step
- Superpixels (SLIC)
- Classification (SVM)



Superpixel Segmentation

Unsupervised

Classification

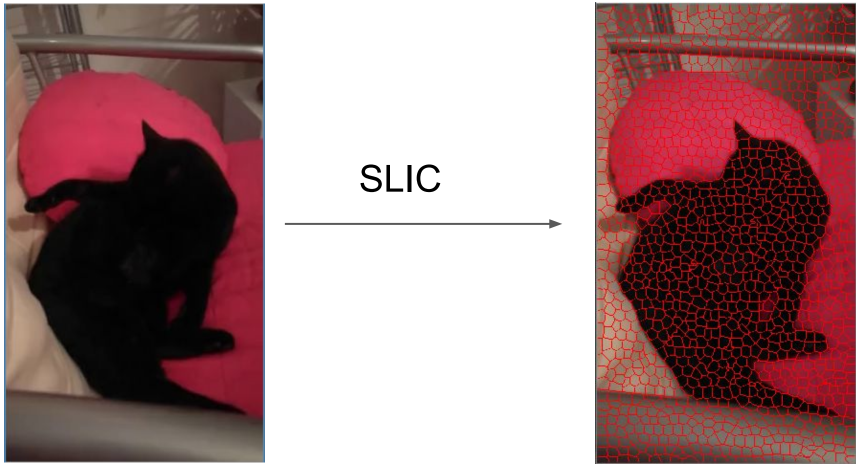
Supervised

SLIC Superpixels

- Each pixel is represented in 5D *labxy* space

  lab -> color of the pixel in CIELAB color space [2]

  xy -> position of the pixel in image



- Cluster centers initialized as a grid
- Pixels are iteratively clustered based on distance in the 5D space

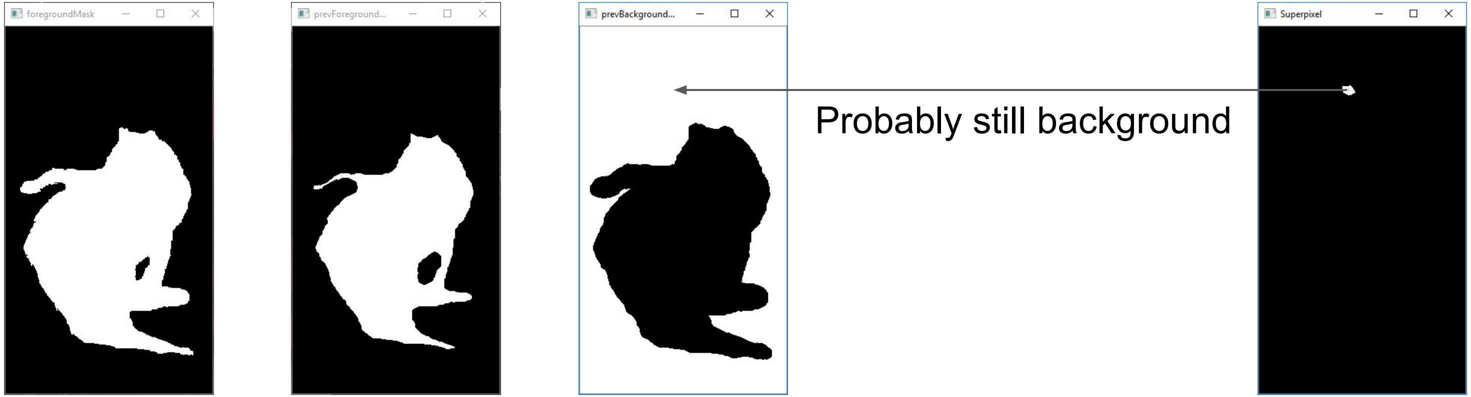
  Similar to K-means

Classification Step

- Each superpixel is converted into a feature vector
  - Mean color
  - RGB histogram → Color
  - LBP histogram → Texture
- First frame of the video is annotated by the user
- SVM is trained with the annotated superpixels
- In all frames, superpixels are classified by the same SVM

Additional Features

- Temporal Inertia
    - Eroded versions of foreground and background masks are stored for the next frame
    - If the centroid of a superpixel lies in a mask, its probability to be classified as such increases



Probably still background

Foreground mask

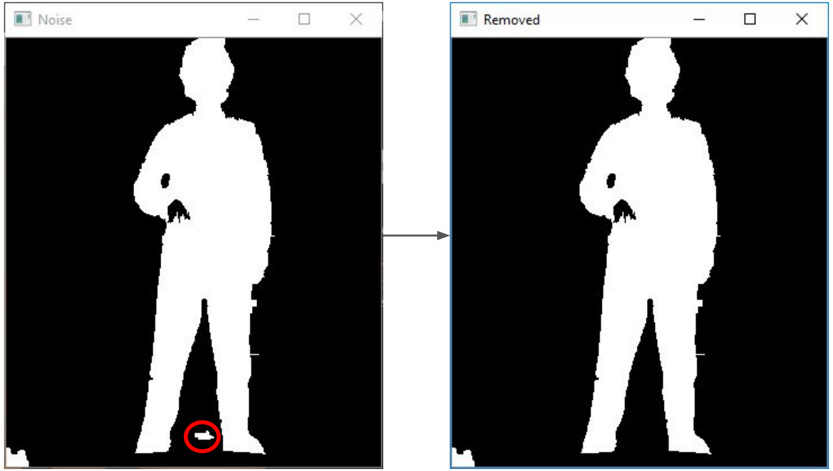
Eroded

Inverted and Eroded

Superpixel in next frame

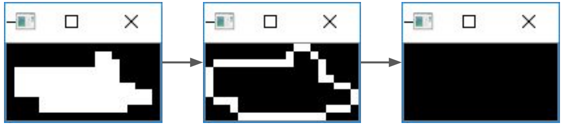
Additional Features

- Noise Reduction
  - A single superpixel completely surrounded by oppositely labeled superpixels is probably noise
  - Such superpixels are detected and converted to the opposite label



Outline of a superpixel anded with the foreground mask is completely black

Has no foreground neighbors (vice versa for background)



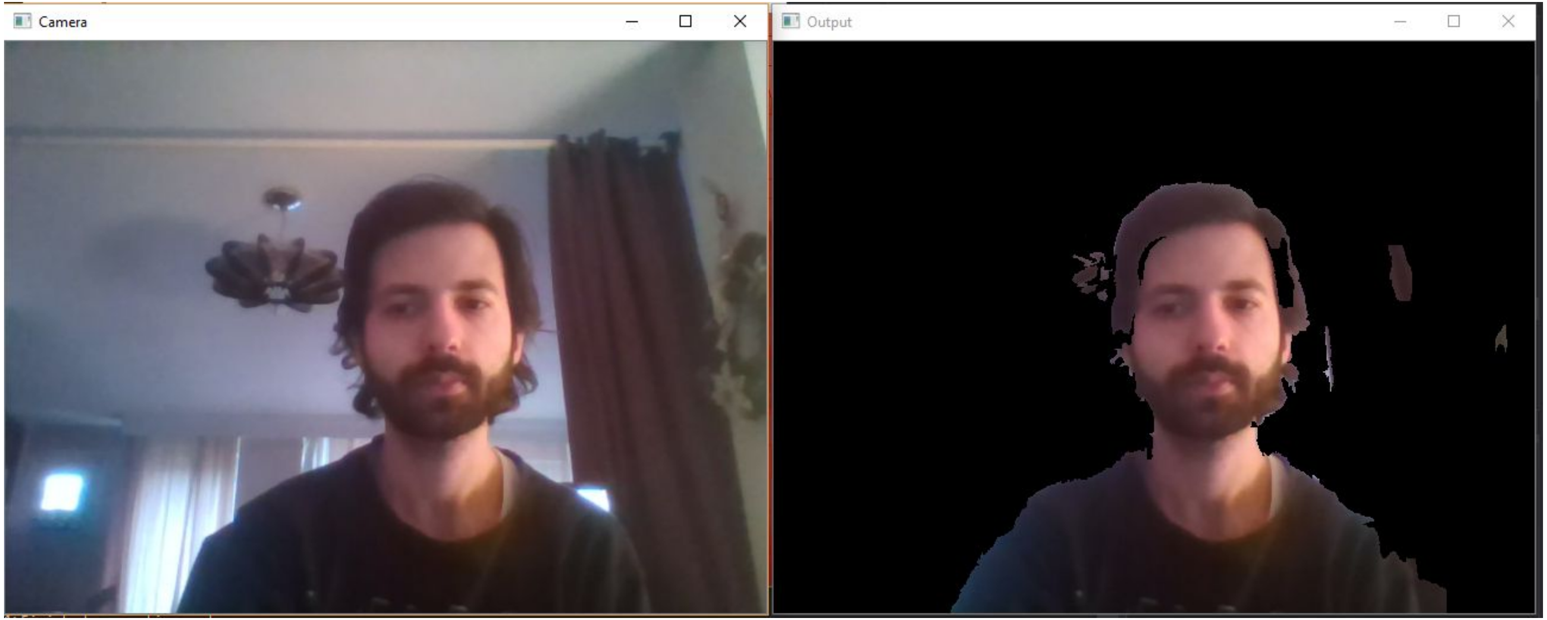
Mask

Outline

Anded

Additional Features

- Real Time Segmentation
  - Input is directly taken from a camera stream, instead of a file



Implementation

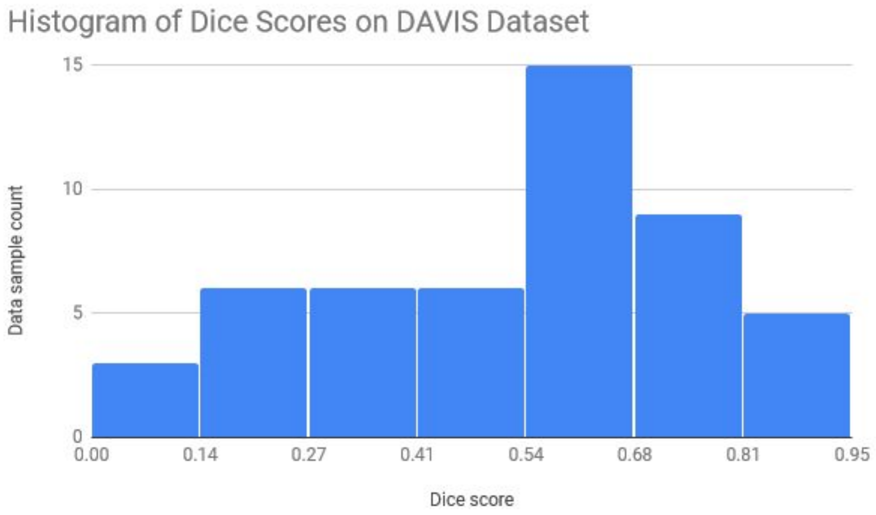
- C++
- OpenCV
- OpenCV implementations of SVM and SLIC

Limitations

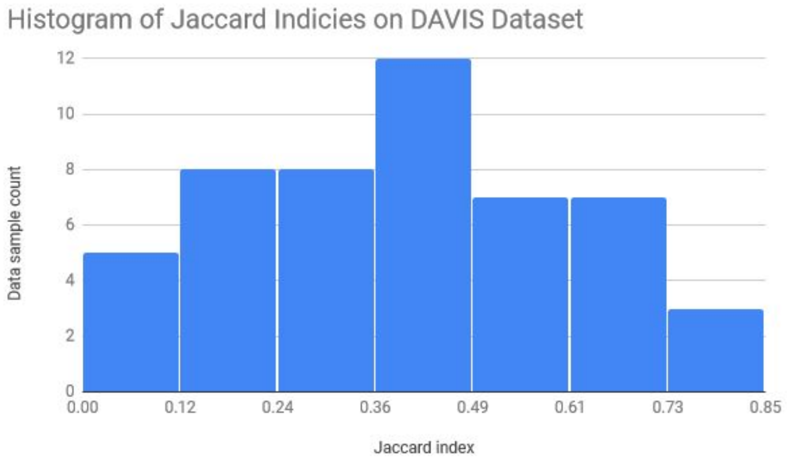
- The segmentation will perform badly if:
  - Background is too similar to the foreground in terms of color and texture
  - Color intensities of an object change too much over time
  - Different looking objects enter the scene
- Segments everything into only two classes
  - But adding multiclass support would be simple
- Parameters such as number of SLIC iterations are limited for performance
  - Parallelization and GPU utilization could be used to run it faster

Evaluation

- Testing done on DAVIS 2016 dataset
  - DAVIS: Densely Annotated VIdeo Segmentation [3]
- First annotation frame used for training
- Rest of the annotations used after testing for evaluation



Histogram of Dice Scores on DAVIS Dataset



Histogram of Jaccard Indicies on DAVIS Dataset

References

[1] Lecture slides - Segmentation

[2] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., & Süsstrunk, S. (2010). Slic superpixels (No. EPFL-REPORT-149300)

[3] DAVIS: Densely Annotated VIdeo Segmentation. (n.d.). Retrieved January 9, 2019, from https://davischallenge.org/davis2016/code.html