ENGN 2560
Progress Presentation

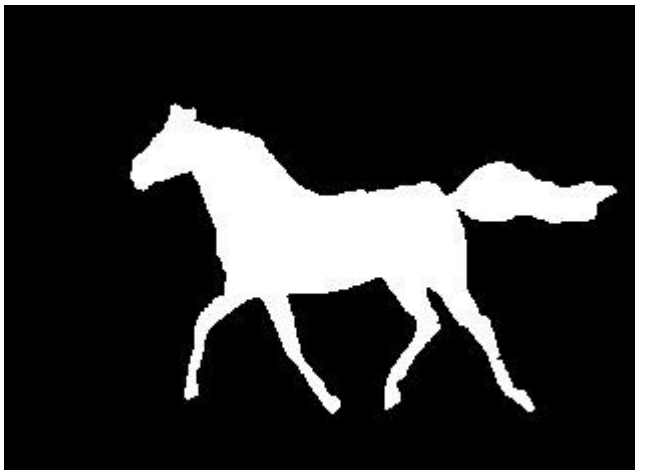
Combined Top-Down/Bottom-Up Segmentation

Firat Kalaycilar
firat@lems.brown.edu

Based on PAMI'08 Paper by Borenstein and Ullman

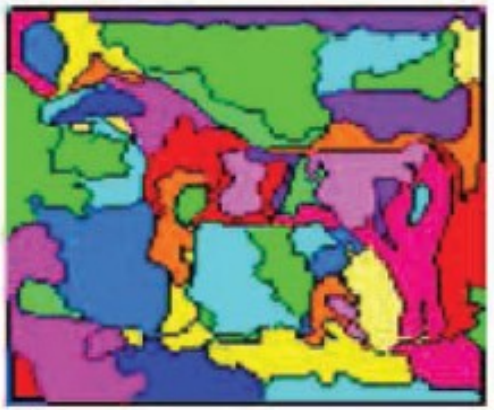
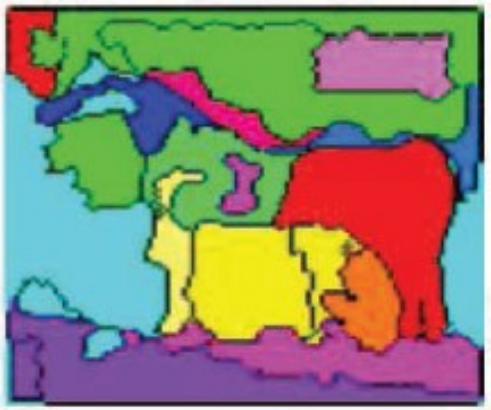
Problem

- Figure-ground image segmentation
  - The task of finding two sets of pixels:
    - A set of figure (object) pixels
    - A set of ground (background) pixels



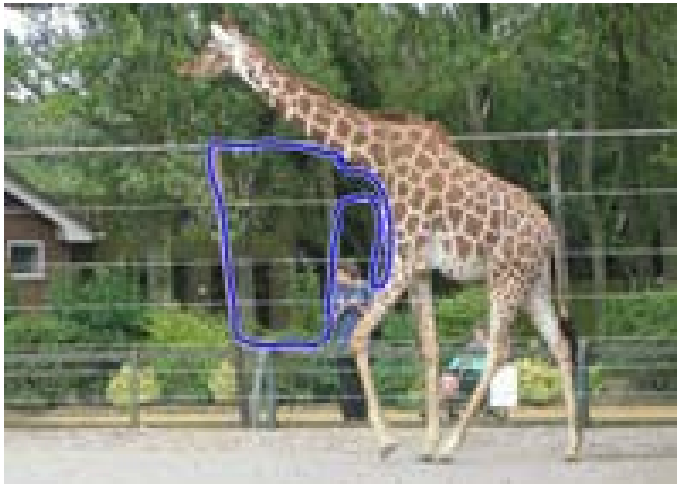
Bottom-Up Segmentation

- Segmentation as a bottom-up process
  - No stored object representation.
  - Segment image into homogenous regions using some image based criteria.
- Advantage: Informative about the exact object boundaries
- Disadvantage: Over/Undersegmentation

Top-Down Segmentation

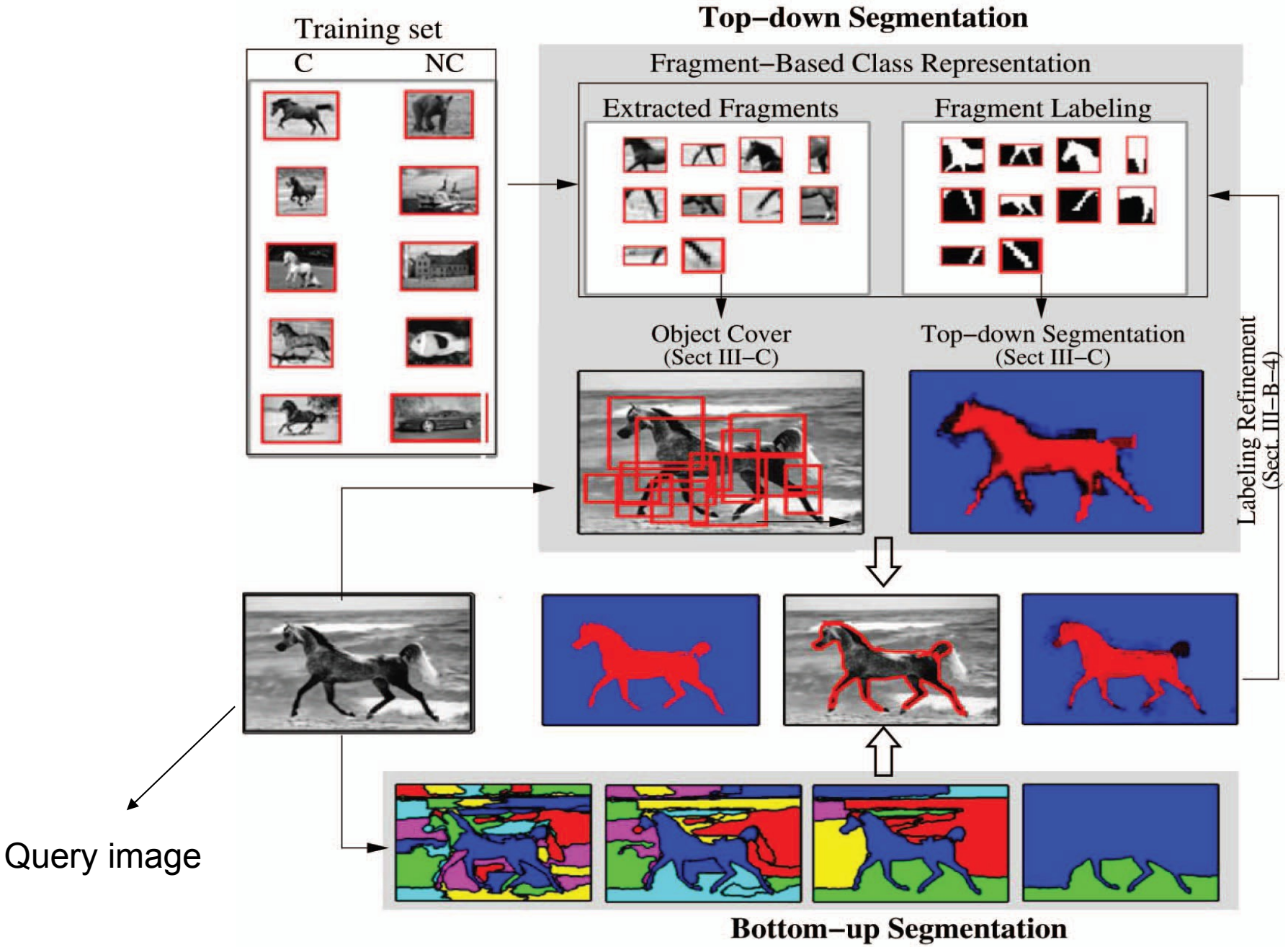
- Segmentation as a top-down process
  - Stored object representations.
  - Look for the object using prior knowledge about its possible appearance and shape.
- Advantage: We know what we are looking for.
- Disadvantage: Object representation cannot model the entire class perfectly.



Combined TD/BU Segmentation

- Idea is to make use of the advantages of both approaches.
  - TDS can produce a preliminary segmentation.
  - BUS can refine it to obtain a better delineation of the object boundary.

Combine TDS and BUS

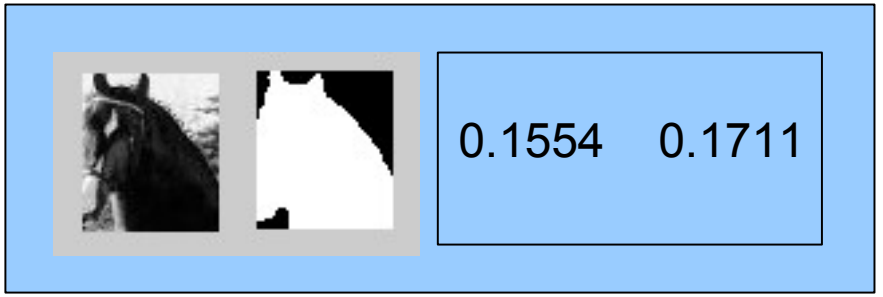


Components

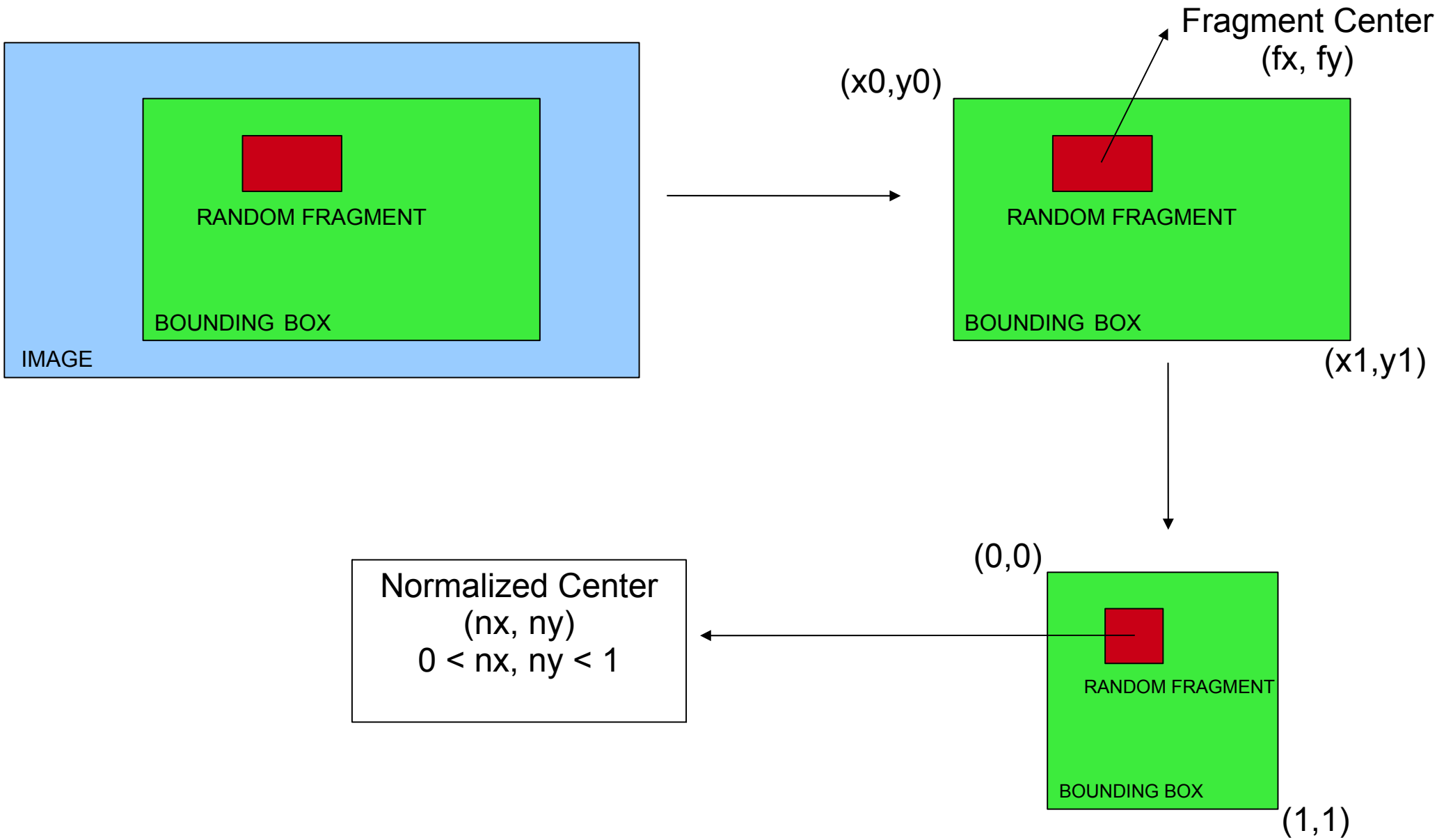
- Fragment Bank Construction (Training)
- Bottom-up segmentation (BUS)
- Top-down segmentation (TDS)
- Combining BUS and TDS (CSEG)

Fragment Bank Construction

- Divide the dataset into two parts: Training and Test sets
- In this stage, only use the training images.
- For each training image
  - Randomly extract N rectangular patches
  - Extract N corresponding groundtruth patches using the groundtruth figure-ground segmentations of the training images.
  - For every fragment, compute and record its normalized center.

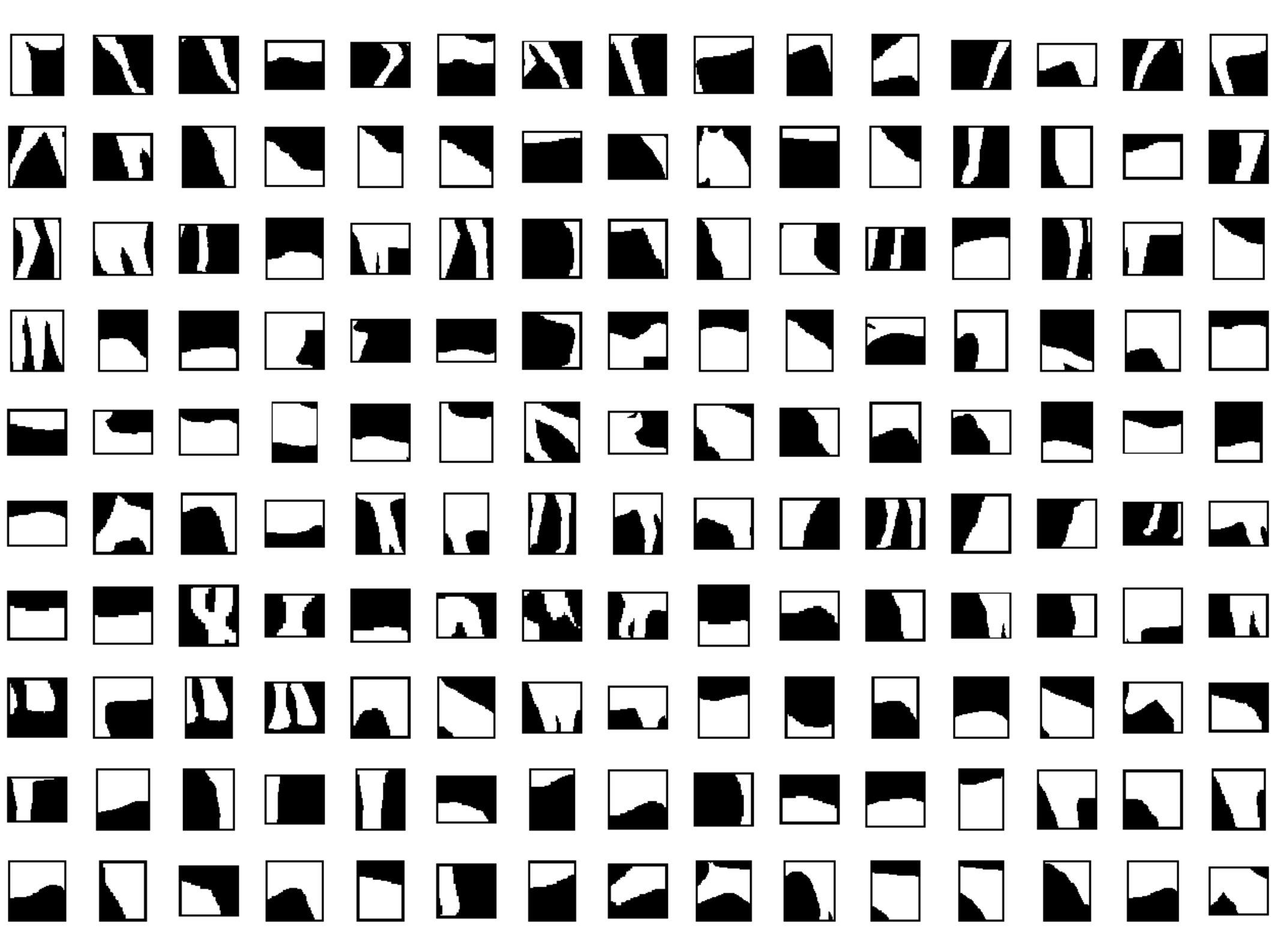


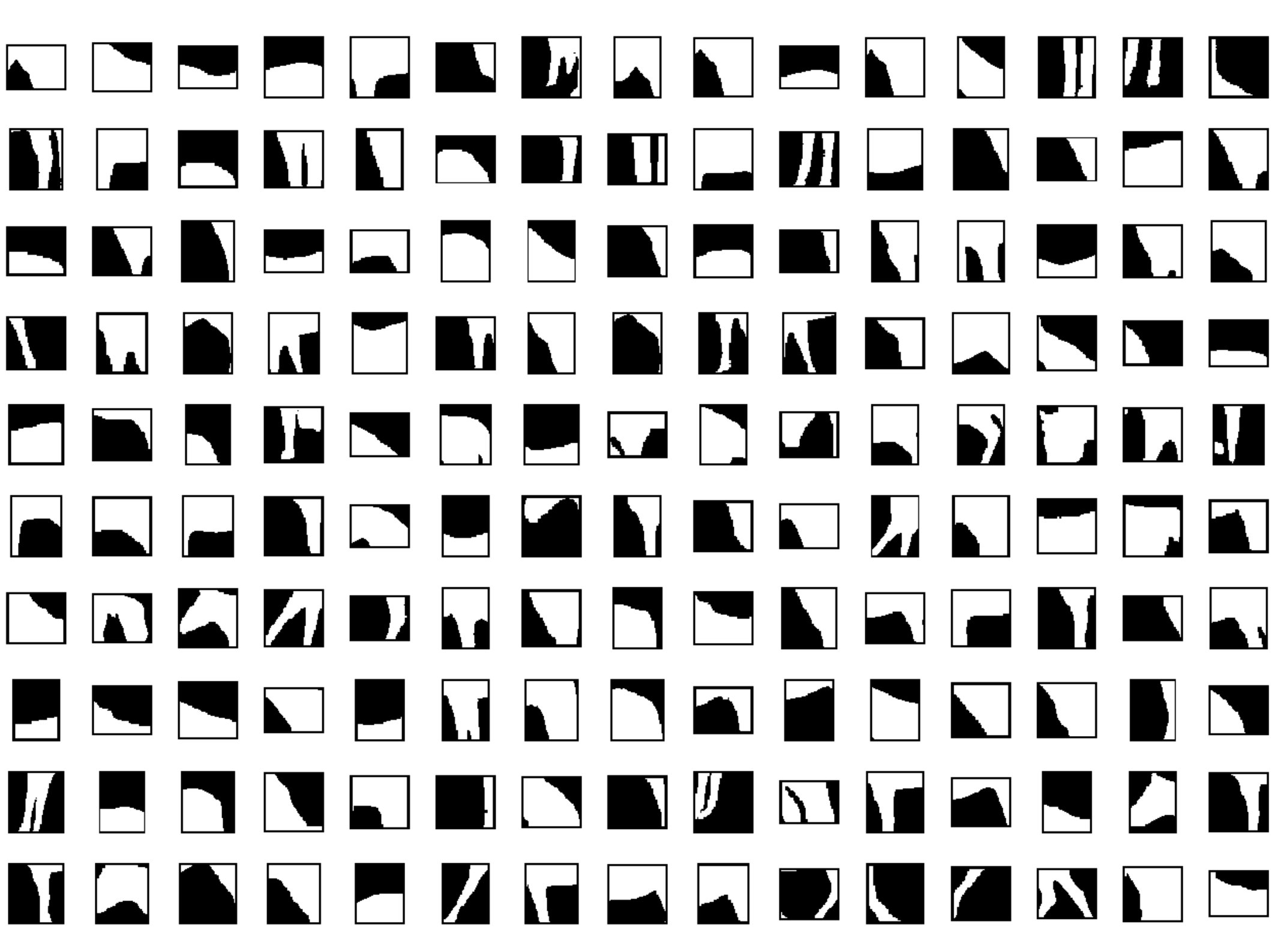
Normalized Center



Fragment Bank Construction

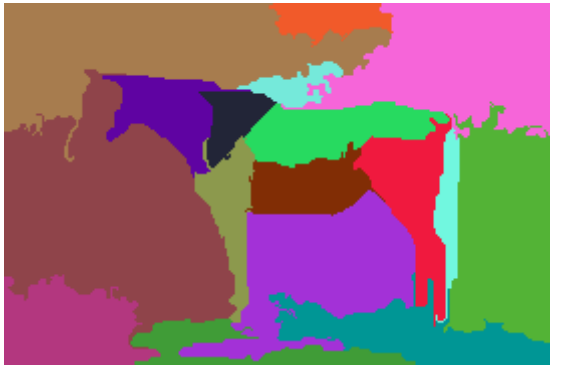
- Make sure that each gray level patch and its corresponding groundtruth figure-ground segmentation satisfies the following criteria:
  - 0.7 <= Aspect ratio <= 1.43
  - 0.15 <= Relative area of the figure (white) <= 0.85
  - 1000 <= Width x Height <= 3000
- We can reduce the number of fragments by eliminating them randomly
  - For example; first I extracted 600 (20 frags/horse x 30 horses) fragments. Then I reduced the fragment bank size to 300.





Bottom-up Segmentation

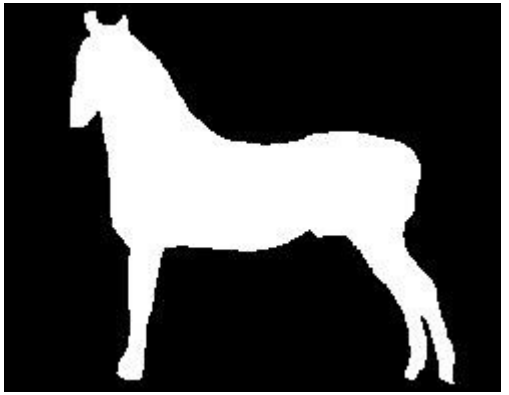
- “Efficient Graph Based Image Segmentation” by Felzenszwalb and Huttenlocher, IJCV 2004
  - Very fast
  - Has only 3 parameters
  - A hierarchical segmentation can be obtained by just varying the parameters.
- Currently, I use flat segmentation.



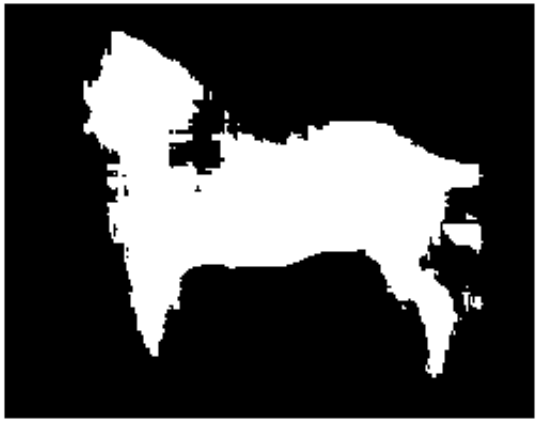
Top-down Segmentation

- In this stage, only the test images are used.
- For each test image
  - For each fragment of the fragment bank
    - Compute normalized cross correlation.
    - Use non-maximum suppression to find the locations of high responses.
    - Keep the locations having responses (R) greater than a threshold.
    - Skip if there is no location left. Else continue.
    - Normalize each location.
    - Compute the squared distance (D) between the stored normalized center and the normalized location. (1 – D/2) is the spatial consistency score (SC).
    - Compute the final score as
      - $FS = \alpha \times R + (1 - \alpha) \times SC$
      - $0 <= \alpha <= 1$.
    - Transfer the groundtruth figure-ground labeling of the fragment to the location yielding the highest FS.
  - Select the pixels covered by sufficient amount of fragments as foreground.

input

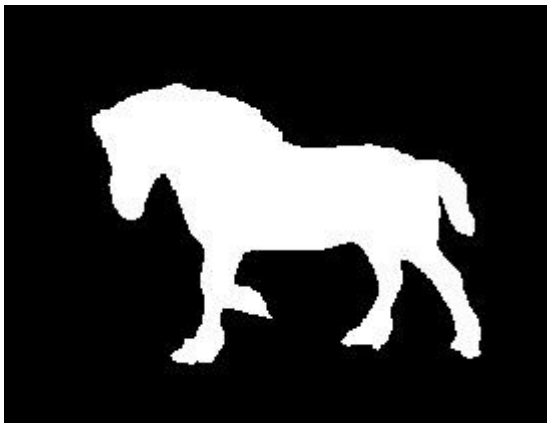


groundtruth

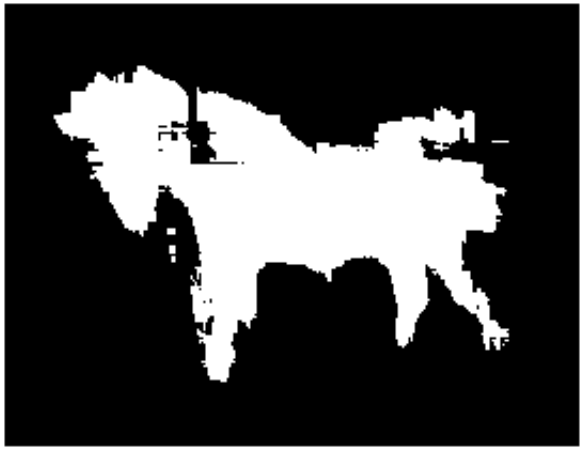


TDS output

input



groundtruth

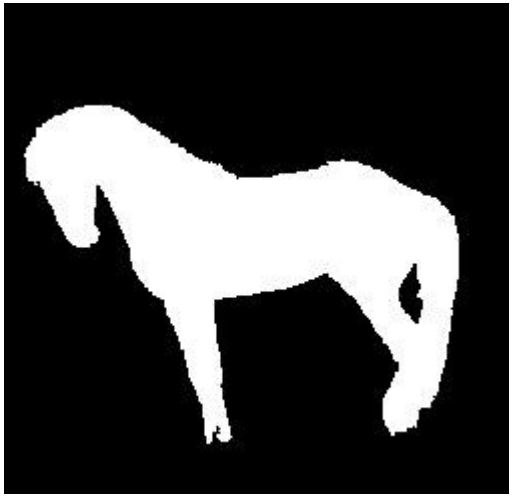


TDS output

Combining TDS and BUS

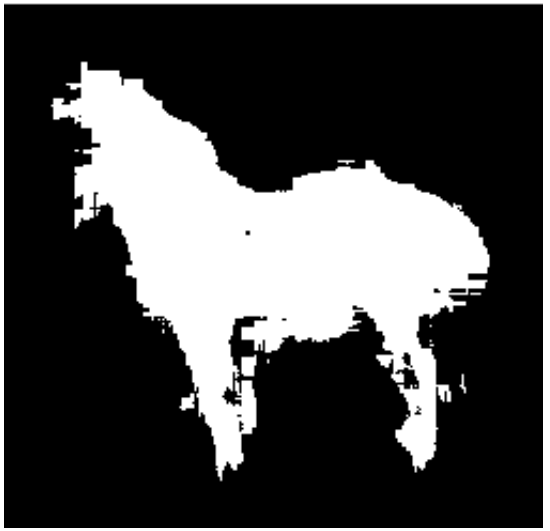
- For each test image
  - For each BU segment of the image
    - Compute the ratio of the pixels detected as foreground at the end of TDS.
    - If the ratio is greater than a threshold, label the entire BU segment as foreground.
    - Otherwise, label the entire BU segment as background.

input

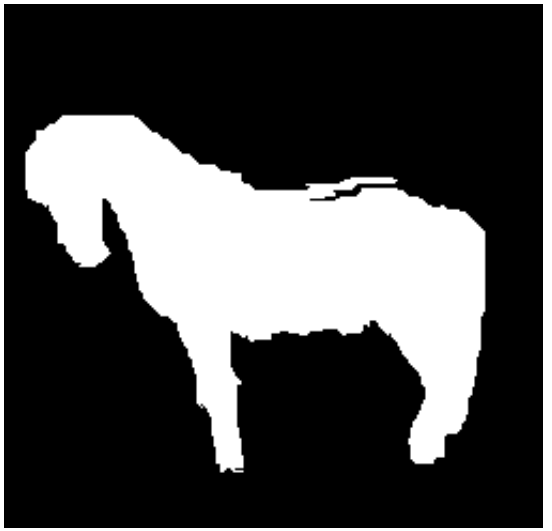


groundtruth

BUS output

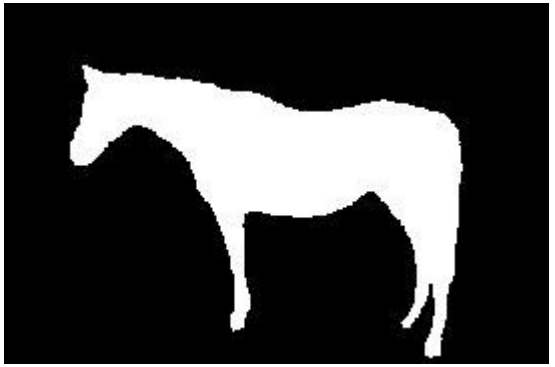


TDS output



Combined output

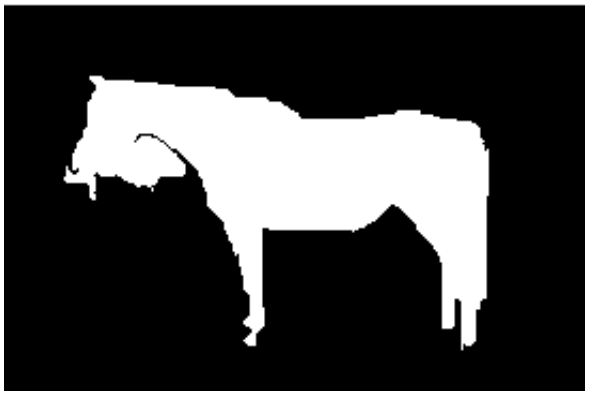
input



groundtruth

BUS output

TDS output



Combined output

Experimental Results

- Weizmann Horse Dataset
  - 328 images in total
  - Training: 30 images
  - Test: 298 images
- I generated a fragment bank for once using the criteria I explained before.
  - 300 fragments in total
- I didn't evaluate the BUS, because it doesn't end up with a figure-ground segmentation.
- I used BUS to generate 9 segmentations per test image. As we go from the 1st level to the 9th level, the number of segments decreases.

How to evaluate?

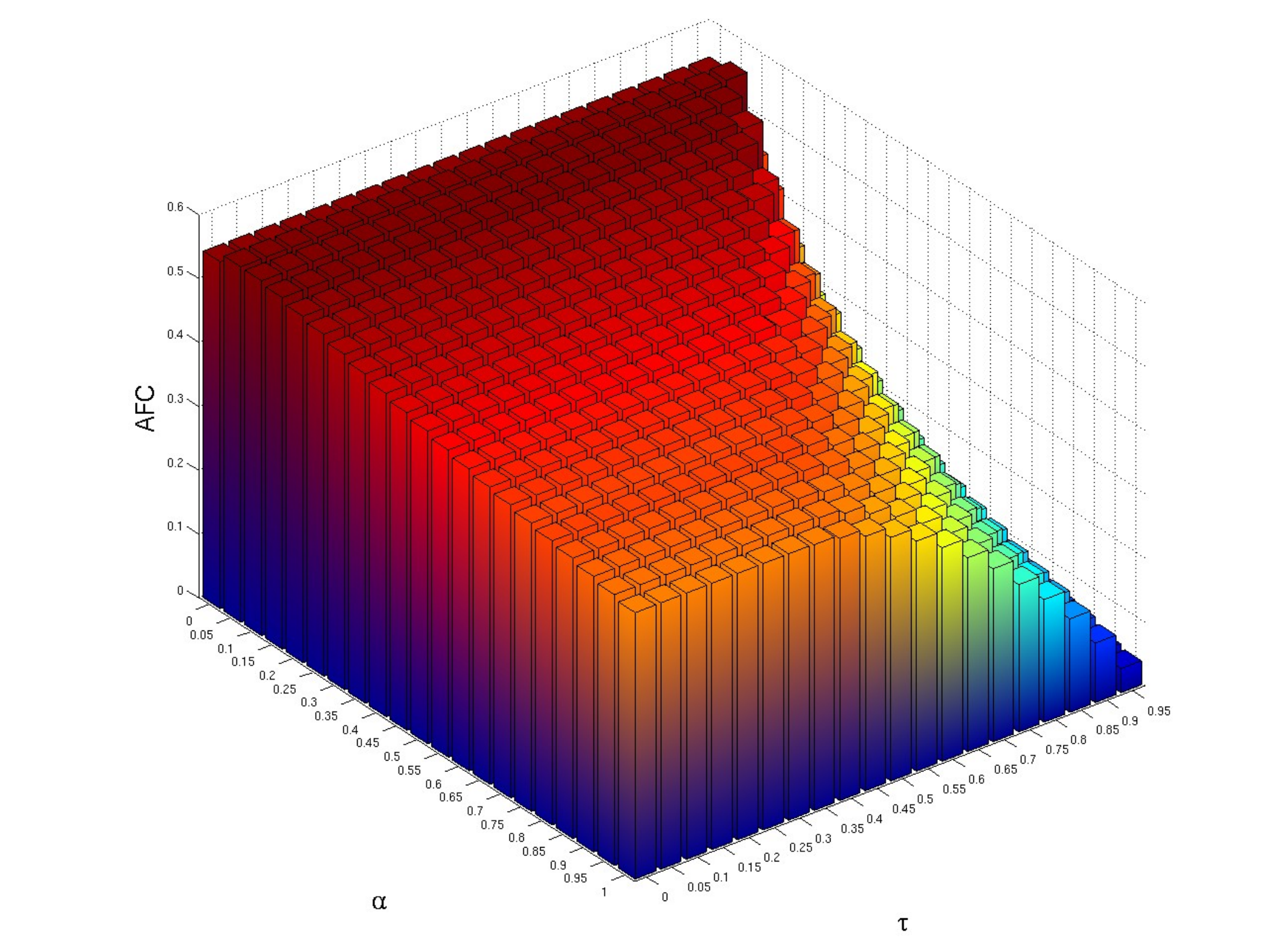
- I used two measures as the authors did in their paper:
  - Figure consistency

$$|F \cap S| / |F \cup S|$$

  - Overall consistency
    - {Total number of correctly classified foreground and background pixels} / {Total number of pixels}

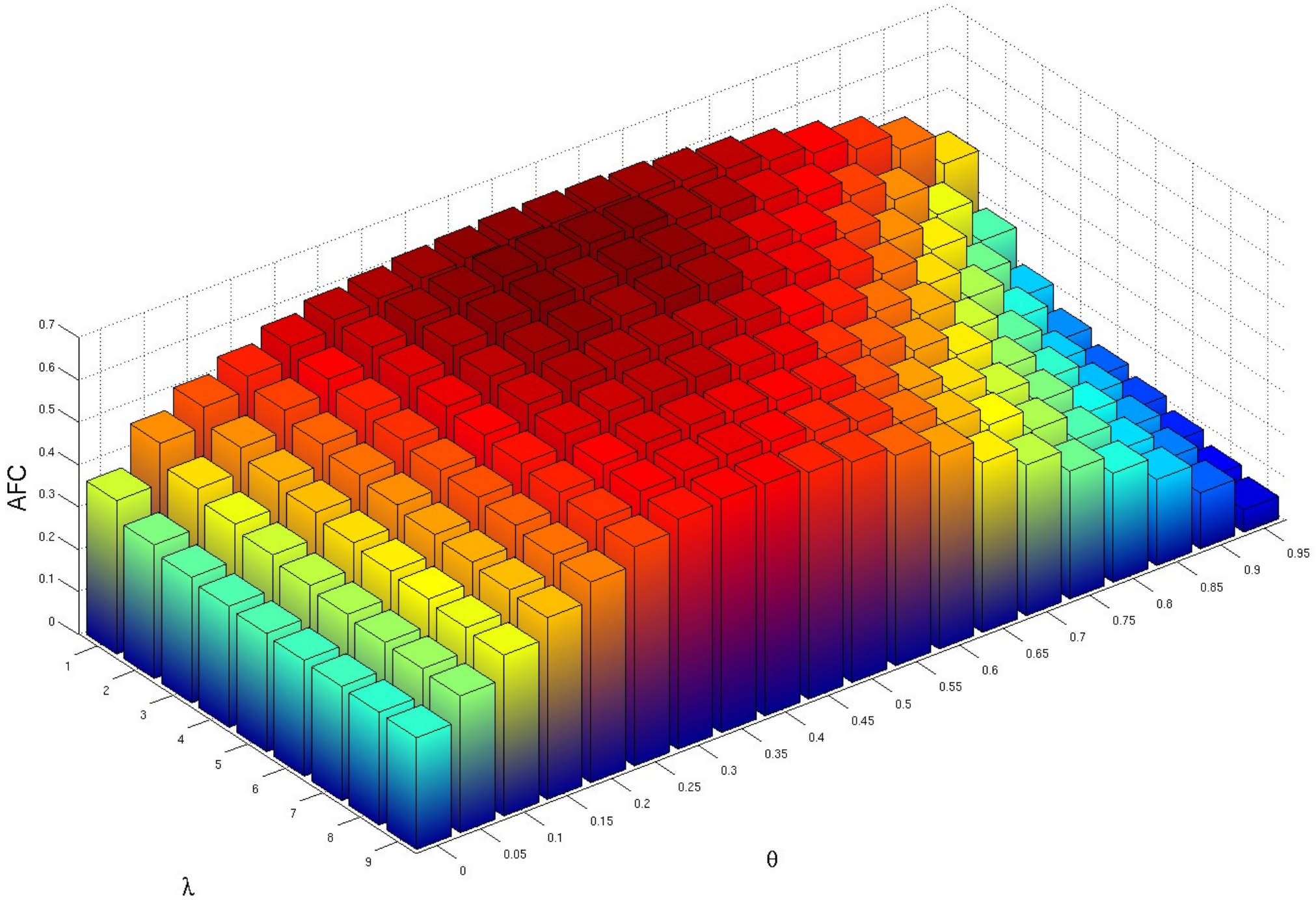
Experimental Results

- Top-down Segmentation
  - There are two parameters:
    - α: Weight representing the contributions of normalized cross correlation response and the spatial consistency of a fragment.
    - τ: Normalized cross correlation response threshold
  - Conducted experiments:
    - $\alpha \in \{0, 0.05, 0.10... 0.95, 1.00\}$
    - $\tau \in \{0, 0.05, 0.10... 0.95\}$
  - According to the average figure consistency
    - Best → $\alpha = 0.05$, $\tau = 0.95$, AFC = 55.68%
    - Worst → $\alpha = 1.00$, $\tau = 0.95$, AFC = 3.79%
  - According to the overall consistency
    - Best → $\alpha = 0.15$, $\tau = 0.90$, OC = 85.62%
    - Worst → $\alpha = 1.00$, $\tau = 0.90$, OC = 75.51%


AFC
0.6
0.5
0.4
0.3
0.2
0.1
0
α
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
1
τ
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95

Experimental Results

- Combined Segmentation (Using TDS with the highest AFC)
  - There are two parameters:
    - λ: BUS level
    - θ: Threshold to label an entire BU segment as foreground
  - Conducted experiments:
    - λ: 1 – 9
    - θ ∈{0, 0.05, 0.10... 0.95}
  - According to the average figure consistency
    - Best: λ = 2, θ = 0.45, AFC = 63.64%
    - Worst: λ = 9, θ = 0.95, AFC = 5.28%
  - According to the overall consistency
    - Best: λ = 2, θ = 0.55, OC = 88.79%
    - Worst: λ = 9, θ = 0.05, OC = 30.41%


AFC
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
λ
1
2
3
4
5
6
7
8
9
θ
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95

Experimental Results

- Comparison of my results to the ones in the paper:
  - Paper: TDS (AFC = **58**%, OC = 84%)
  - Mine: TDS (AFC = 55.68%, OC = **85.62**%)
  - Paper: Combined with Flat Segmentation

(AFC = **65**%, OC = 85%)

  - Mine: Combined with Flat Segmentation

(AFC = 63.64%, OC = **88.79**%)

Conclusions

- A simpler version of the TDS approach works comparably well as the one presented in the paper.
- Combining the output of the BU and TD processes increases the segmentation accuracy as expected.

Plan

- Apr 21 – Apr 27: Work on improving the TDS stage.
- Apr 28 – May 4 : Work on using the entire segmentation hierarchy instead of using a fixed level in the BUS stage.
- May 5 – May 11: Work on the components which still require improvements.
- May 12 – May 21: Evaluate the system using different datasets.