Fast Discriminative Visual Codebooks using Randomized Clusering Forests

Frank Moosmann, Bill Triggs, and Frederic Jurie

Presented by: Andrew F. Dreher
CS 395T - Spring 2007

Contributions

1) Creating visual “words” using classification trees

2) Small ensembles of randomized trees can outperform k-means clustering

Using stochasticity to improve accuracy

Trees as “Words”

Visual “Words”

1) High dimensional vectors; typically extracted features or clusters of features summarized at a point

2) Clusters forming is usually performed using k-means clustering

3) Used with “bag of words” methods derived from text processing

Trees as “words”

1) Trees are trained as classifiers

2) Leaves are used as “words”

   Represent a classified cluster of visual features

   Provides spacial information and intuition lacking in k-means

3) Classification is a separate stage (using SVM) over the leaves

Information Gain with Entropy

1) Useful with limited number of values

2) Often prefers "pure" nodes

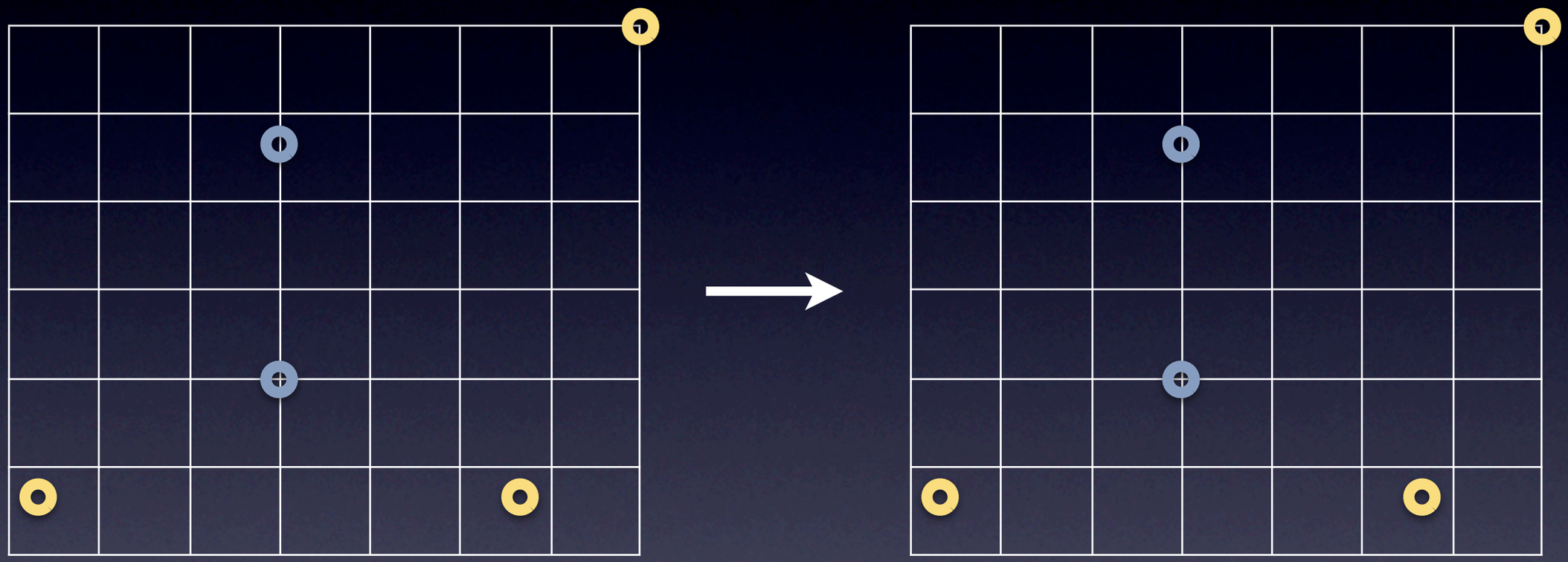
Randomization of thresholds helps create different splits and trees

Paper parameters S_{min} and T_{max}

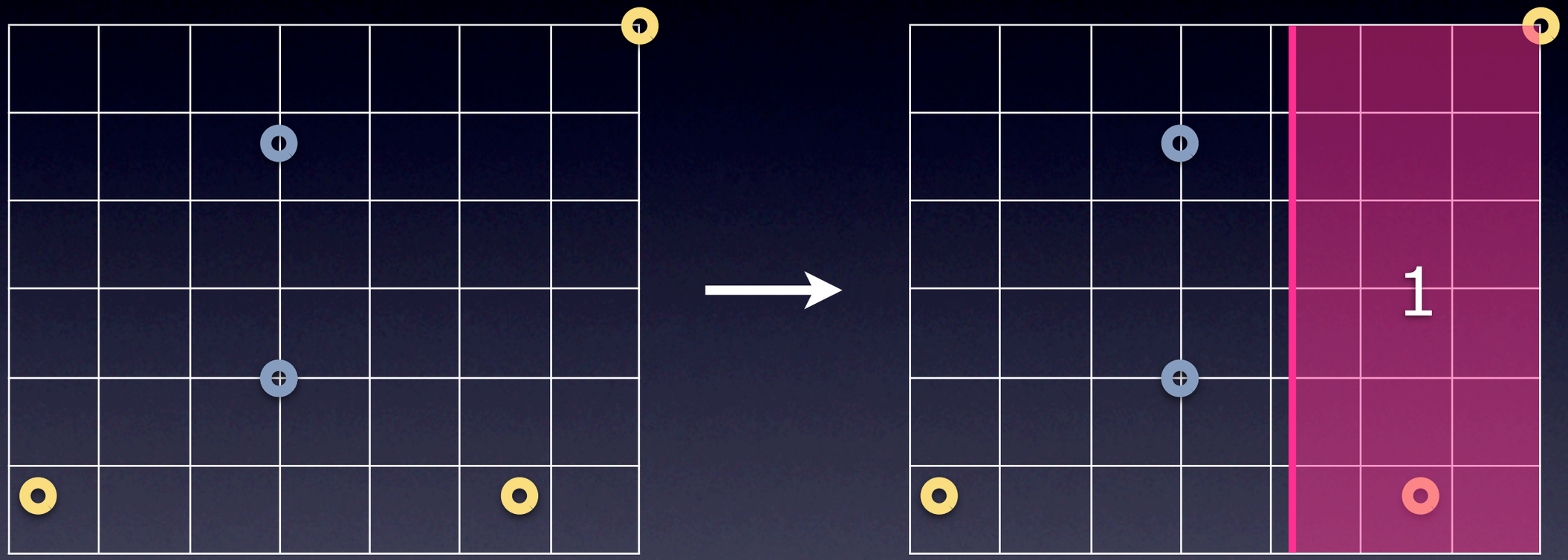
[0, 1] → Completely random trees

[1, D] → Discriminative trees (classic ID3)

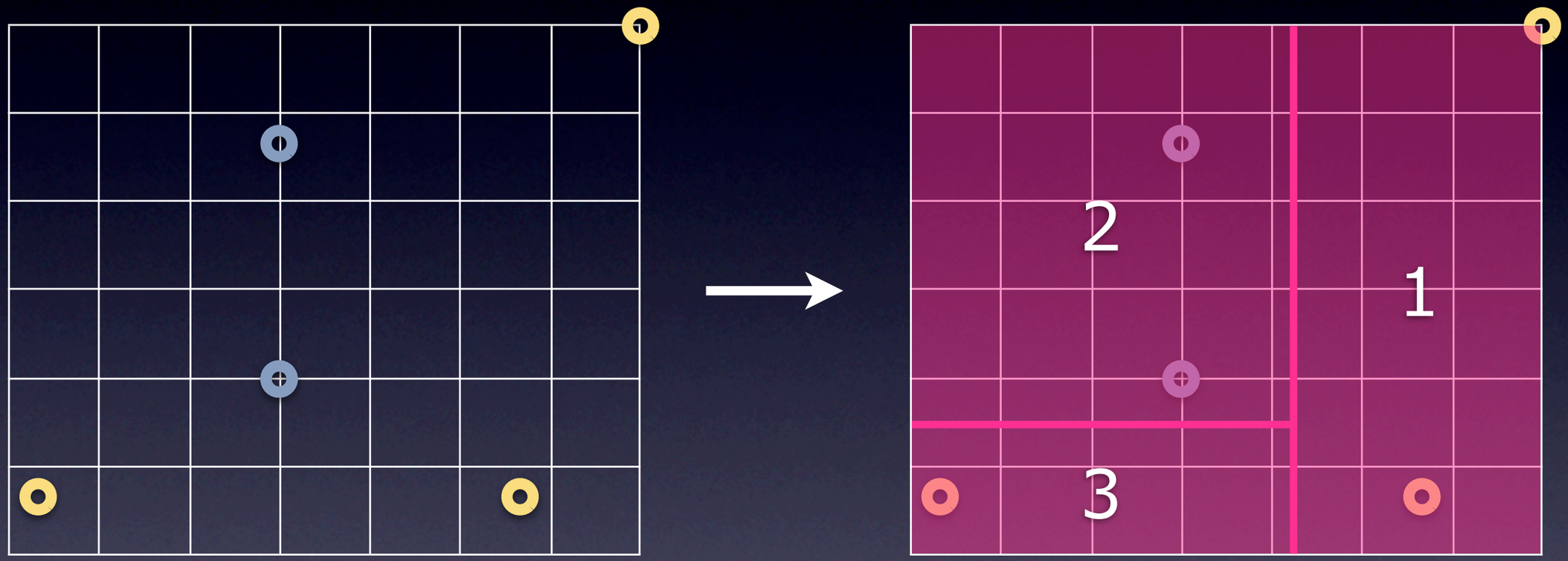
Basic Example of Entropy



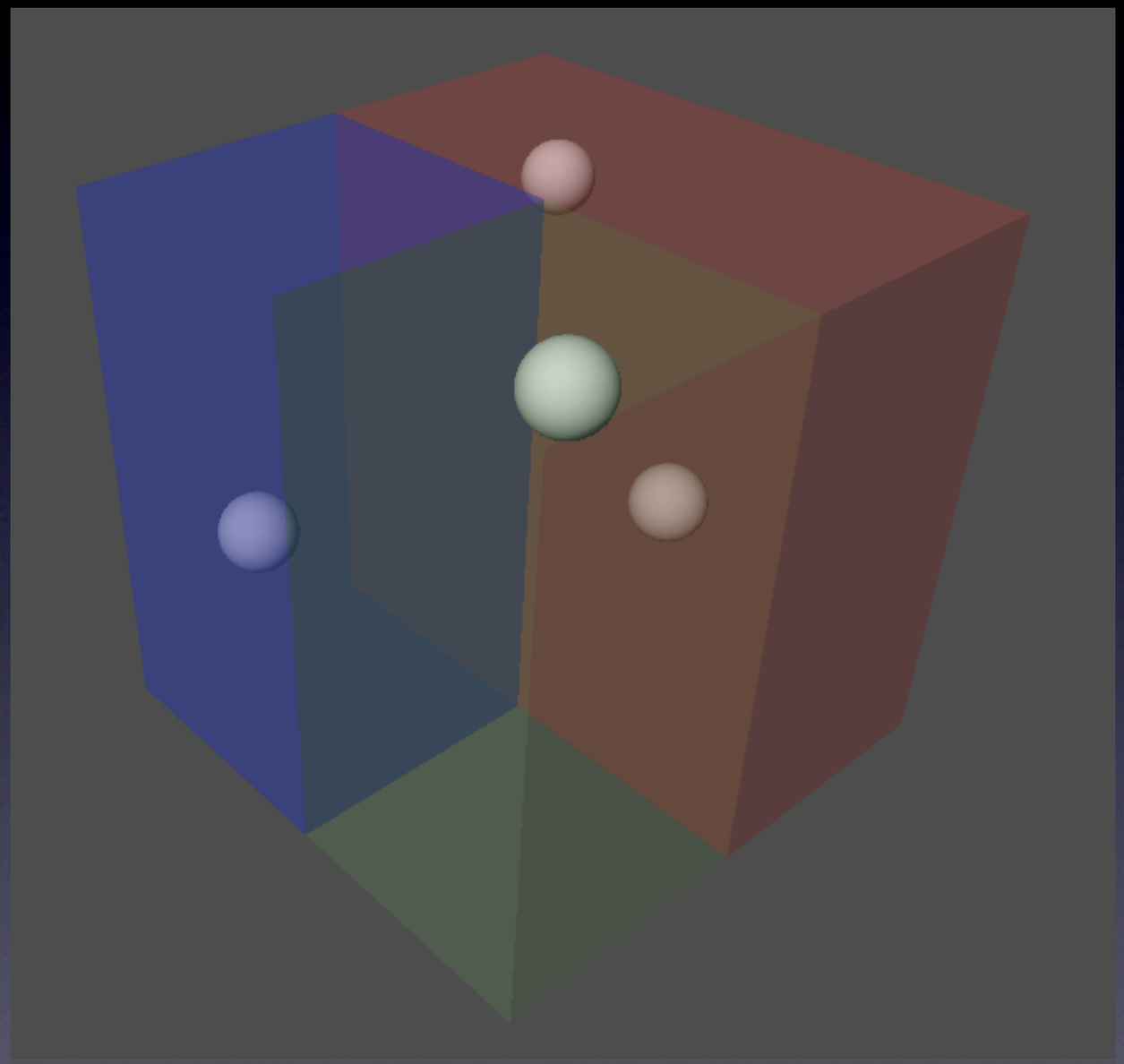
Basic Example of Entropy



Basic Example of Entropy



Basic Example of Entropy



Experiments

General Overview

1) Descriptors - Dataset dependent
    - HSV color (768-D vector)
    - Wavelet (768-D vector)
        - Created from HSV using Haar transform
    - SIFT (128-D vector)
2) Performance Metrics
    1) Receiver Operating Characteristic (ROC)
    2) Equal Error Rate (EER)

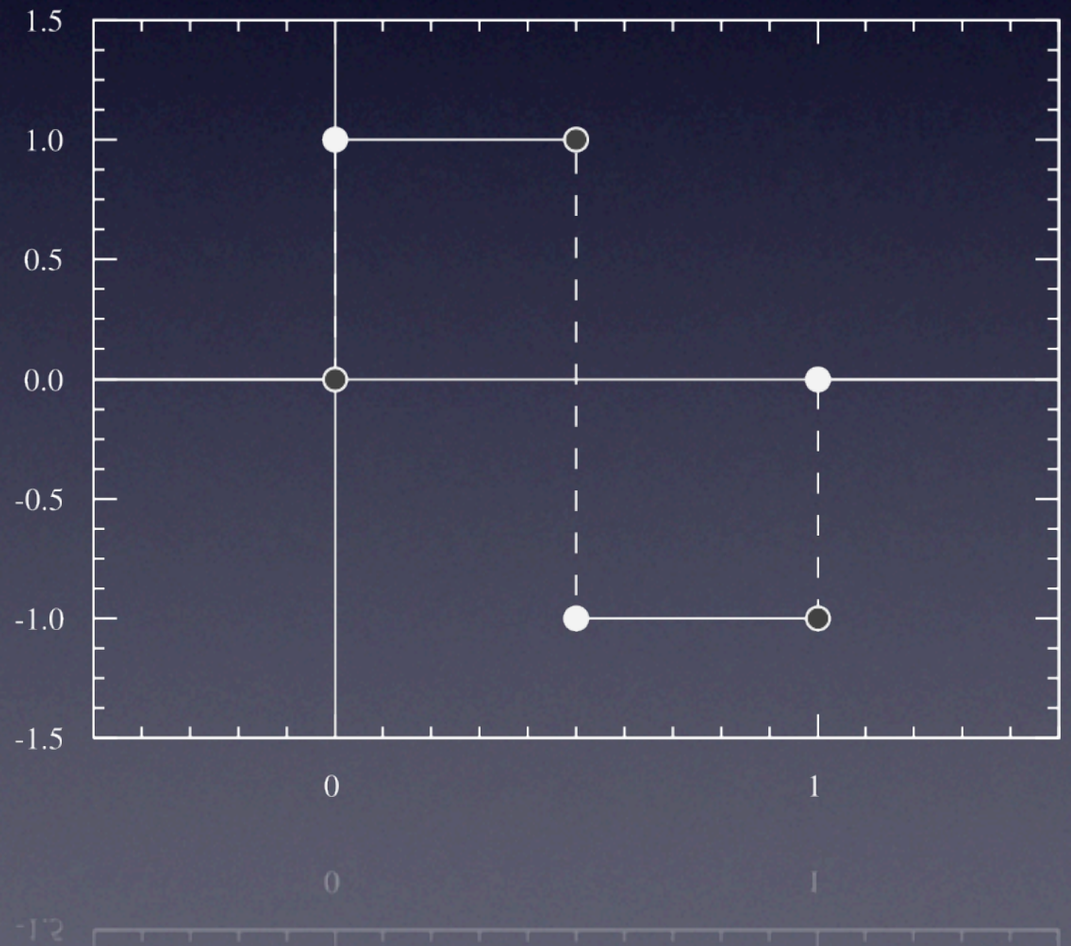
Haar Wavelet

1) First known wavelet

2) Not continuous or differentiable

3) Described as:

$$f(x) = \begin{cases} 1 & 0 \le x \le \frac{1}{2} \\ -1 & \frac{1}{2} \le x \le 1 \\ 1 & 0 \textit{ otherwise} \end{cases}$$



Source: Wikipedia (http://en.wikipedia.org/wiki/Haar_wavelet)

Specific Parameters

1) Descriptors: Color Wavelet

2) Tree parameters: S_{min} = 0.5; T_{max} ≈ 50

3) Dataset: GRAZ-02

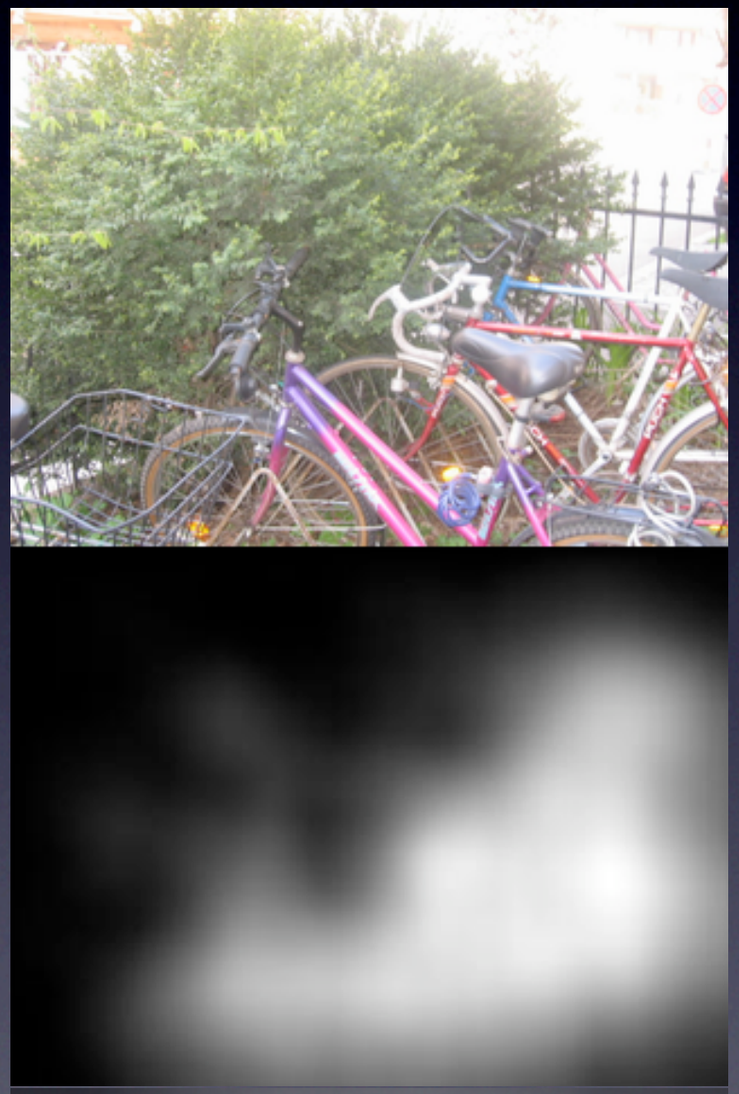
Three categories

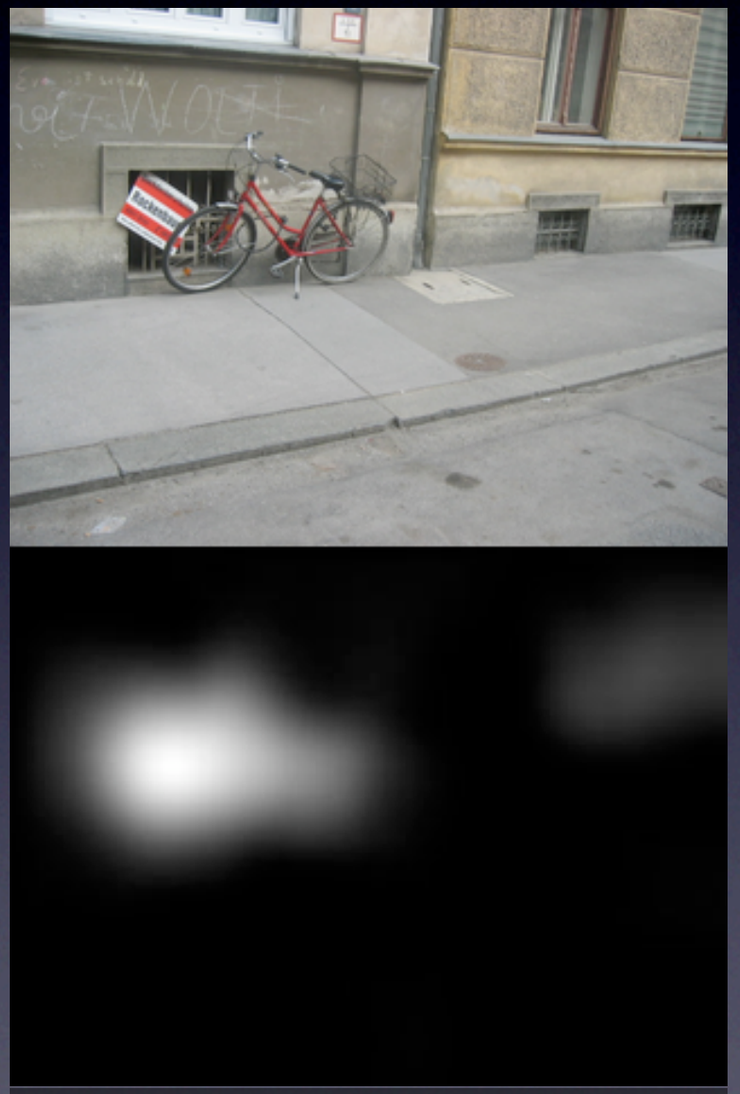
300 Images from each category

½ for training; ½ for testing

Spacial Results

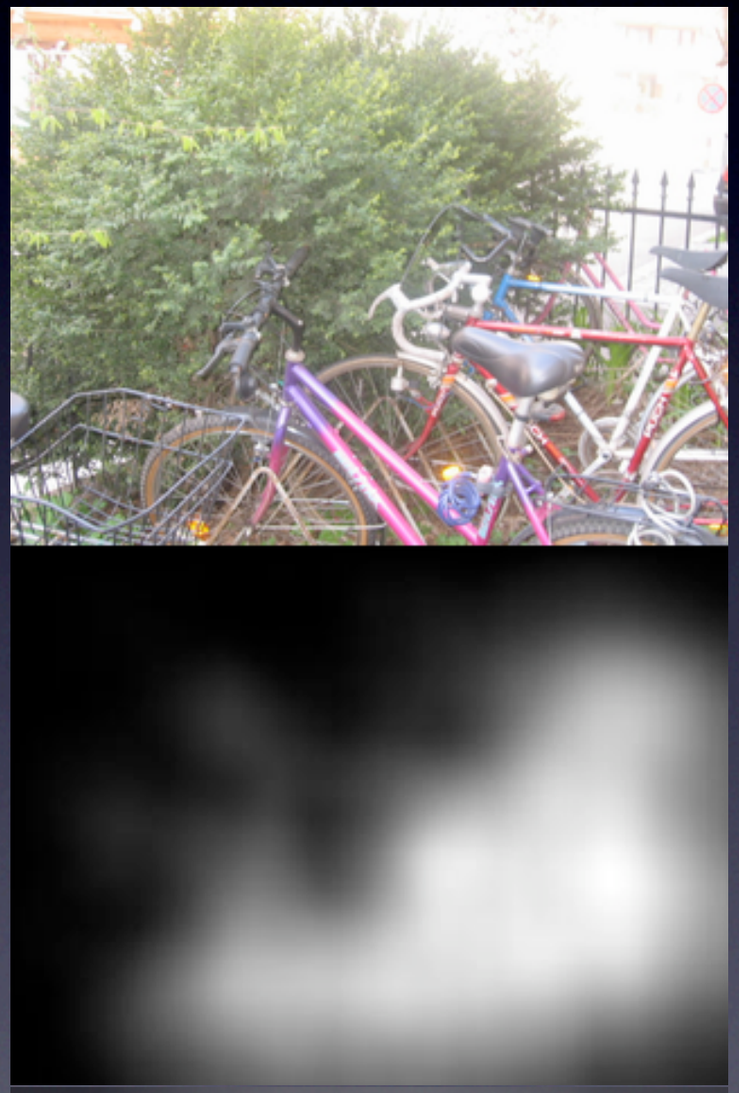
Posterior probabilities at a given position to be labeled “bike”

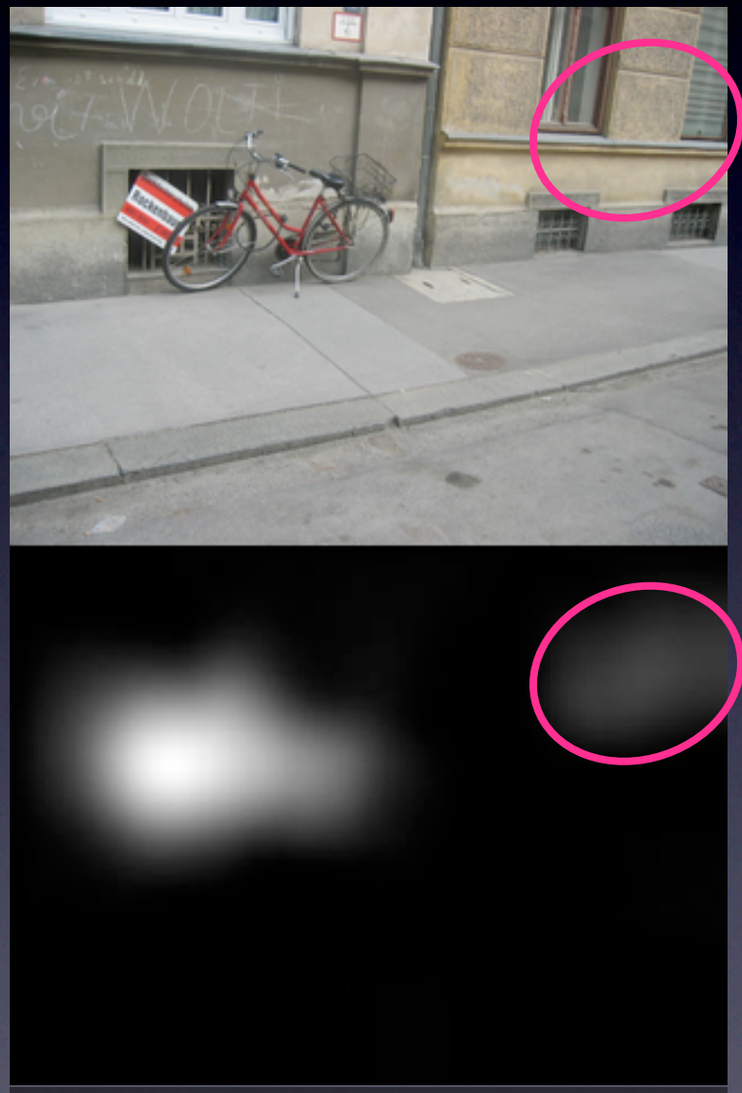




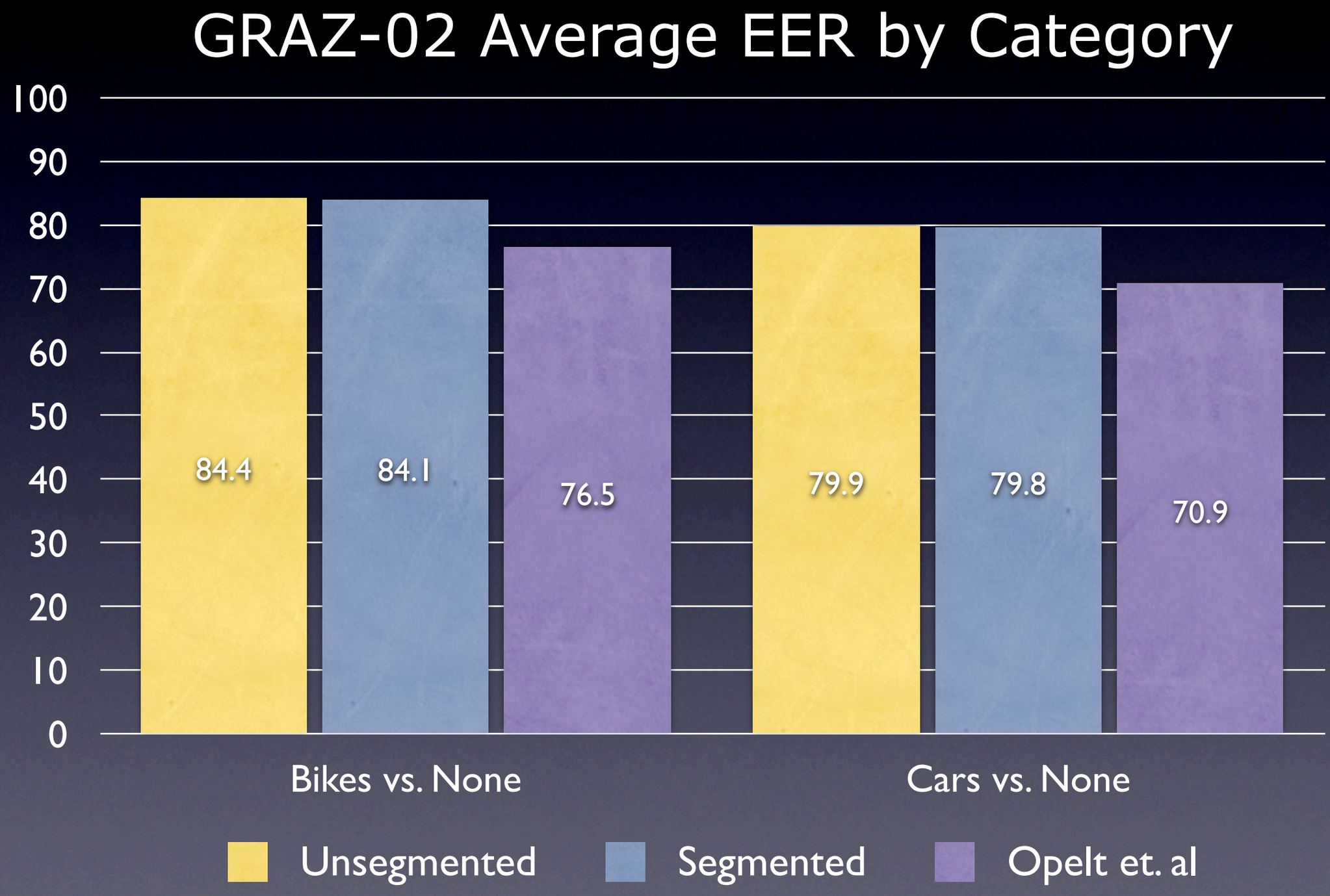
Spacial Results

Posterior probabilities at a given position to be labeled “bike”





Category vs. Negative



Parameters for ERC-Forest vs. K-Means

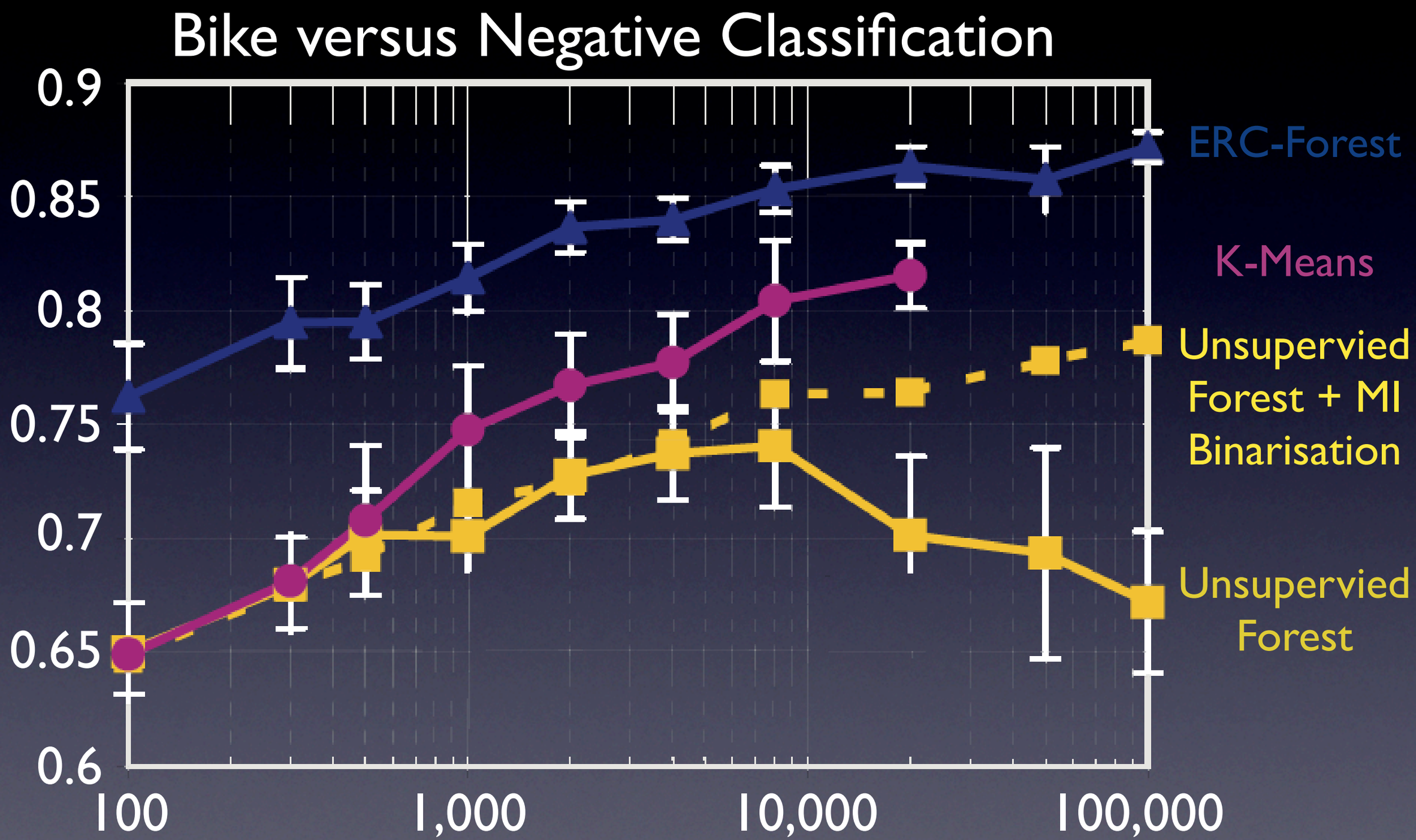
1) 20,000 total features (only 67 per image)

2) 1000 spacial bins per tree; 5 trees

3) 8000 sampled patches to create global histogram

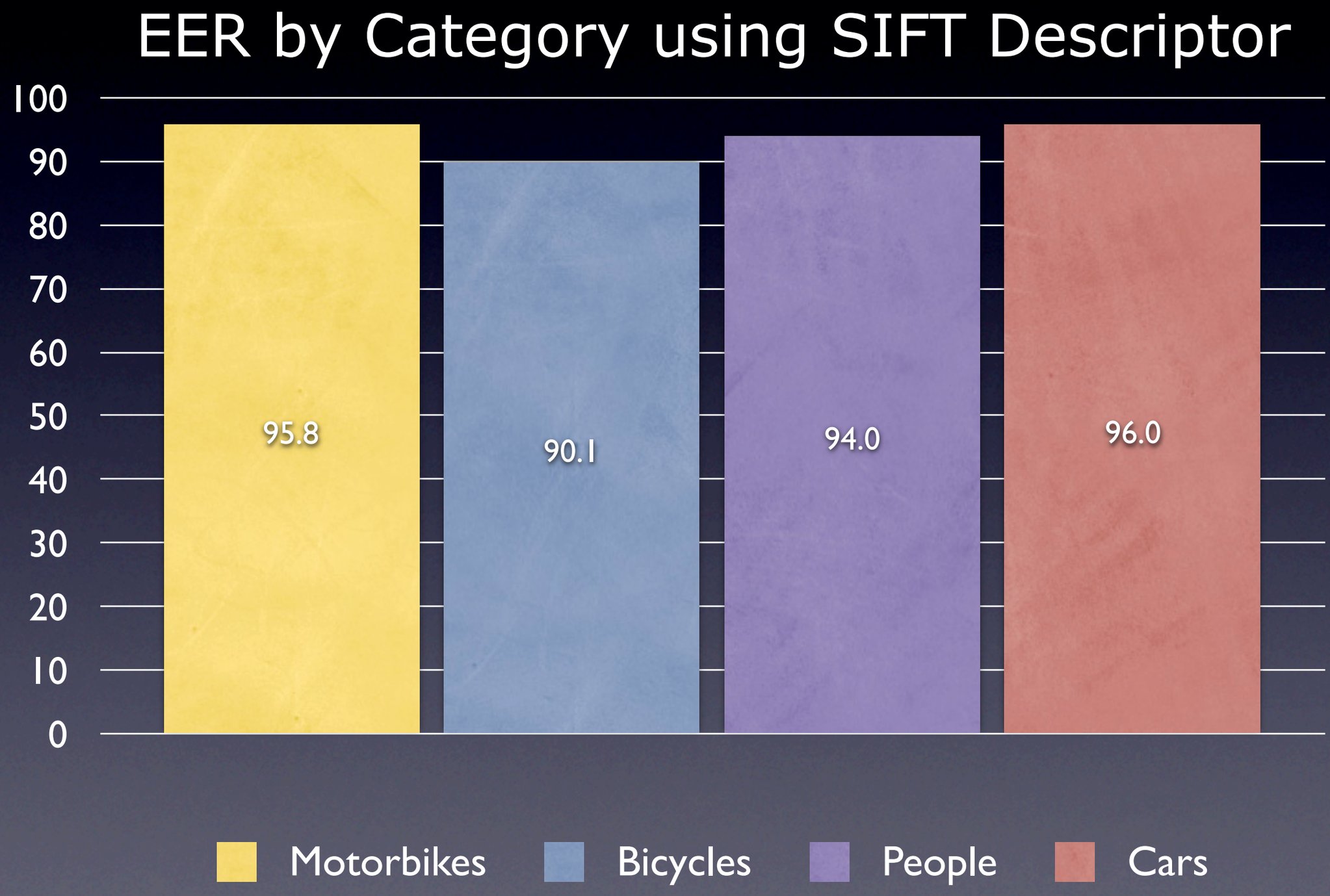
4) 20,000 windows per image for k-means

ERC-Forest vs. K-Means



Other Results

Pascal Challenge Dataset



Pascal Horses Dataset

1) Highly variable images

2) SIFT Descriptors

3) 100 Patches per image for training

4) 10,000 patches per image for testing

5) Average EER: 85.3%

Conclusion

1) Method uses forest of randomized classification trees to create a vocabulary

   Good classification, reasonable training

2) Uses two (2) stage processing

   Use forest to obtain descriptive “word”

   Classify “word” using another method

3) Stochasticity improves accuracy

Thank You