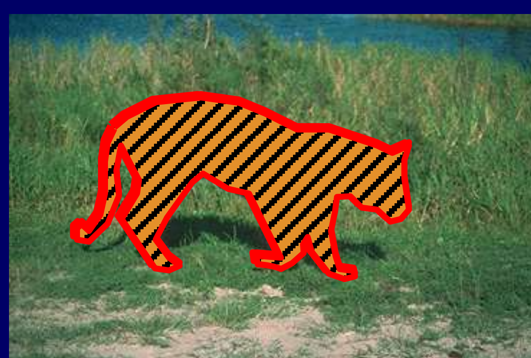
Visual Recognition: Prospects for Image & Video Analytics

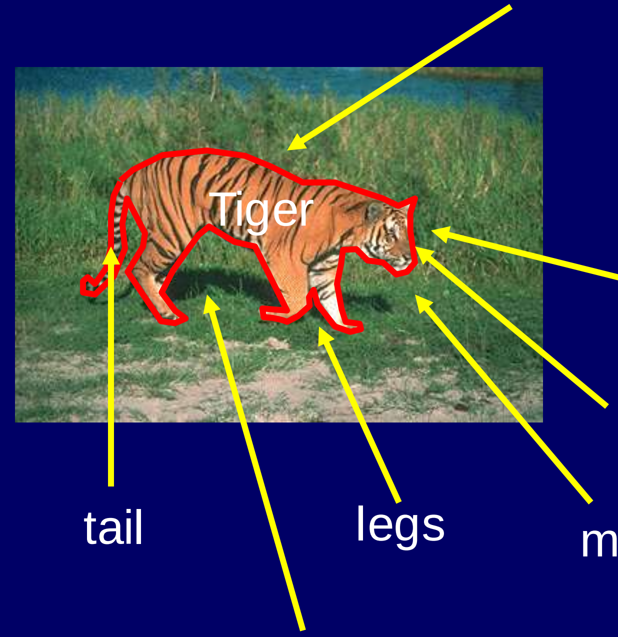
Jitendra Malik

University of California at Berkeley

Classification & Segmentation

outdoor
wildlife

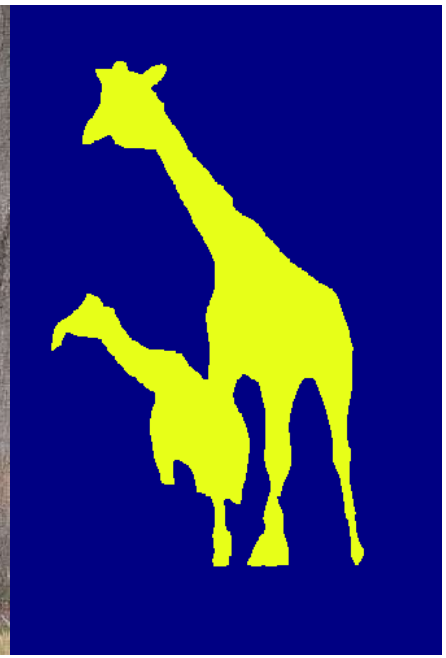


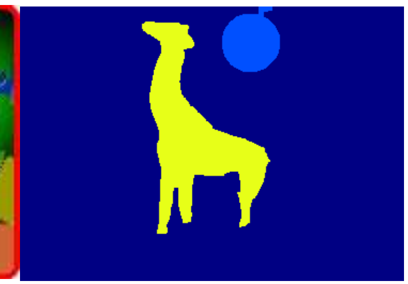


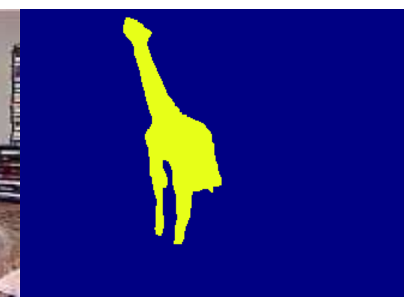
PASCAL Visual Object Challenge

We want to locate the object

Orig. Image | Segmentation | Orig. Image | Segmentation





Fifty years of computer vision 1963-2013

- 1960s: Beginnings in artificial intelligence, image processing and pattern recognition
- 1970s: Foundational work on image formation: Horn, Koenderink, Longuet-Higgins …
- 1980s: Vision as applied mathematics: geometry, multi-scale analysis, probabilistic modeling, control theory, optimization
- 1990s: Geometric analysis largely completed, vision meets graphics, statistical learning approaches resurface
- 2000s: Significant advances in visual recognition, range of practical applications

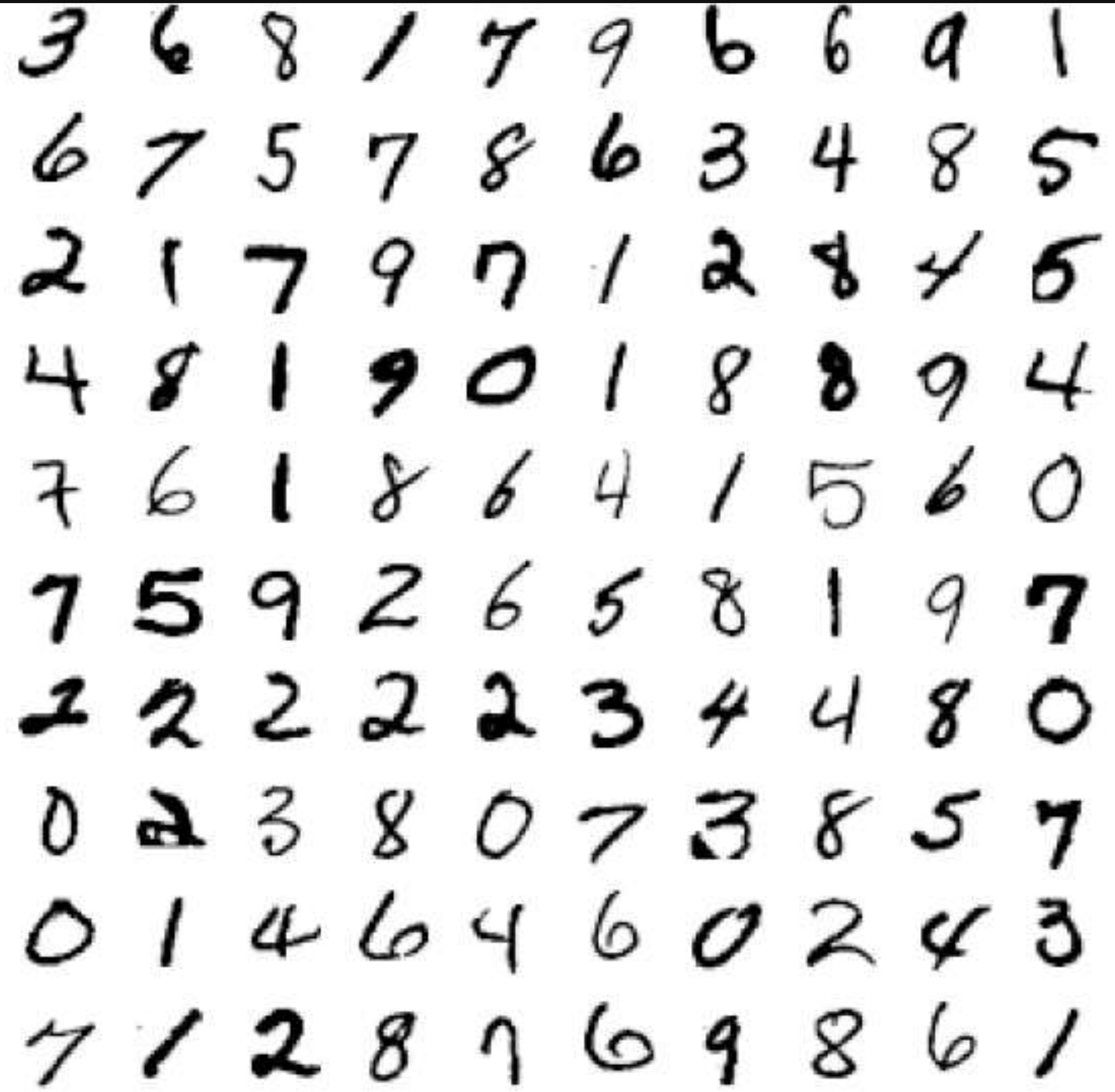


Fig. 4. Size-normalized examples from the MNIST database.

Handwritten digit recognition (MNIST,USPS)

- LeCun's Convolutional Neural Networks variations (0.8%, 0.6% and 0.4% on MNIST)
- Tangent Distance(Simard, LeCun & Denker: 2.5% on USPS)
- Randomized Decision Trees (Amit, Geman & Wilder, 0.8%)
- K-NN based Shape context/TPS matching (Belongie, Malik & Puzicha: 0.6% on MNIST)

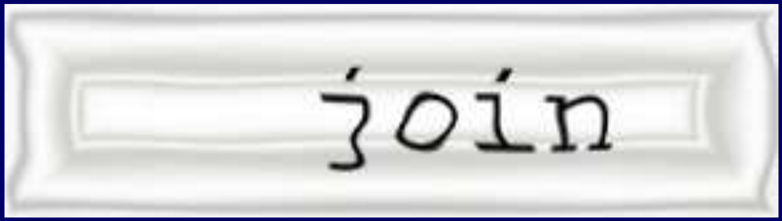
EZ-Gimpy Results (Mori & Malik, 2003)

- 171 of 192 images correctly identified: 92 %

horse

spade

smile



join

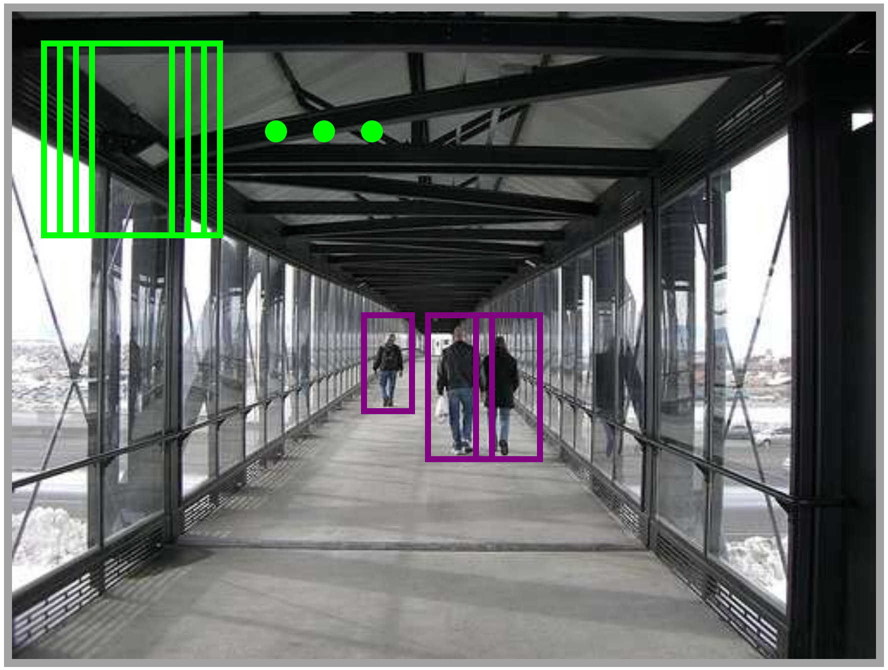
canvas

here

Face Detection

Carnegie Mellon University

Multiscale sliding window



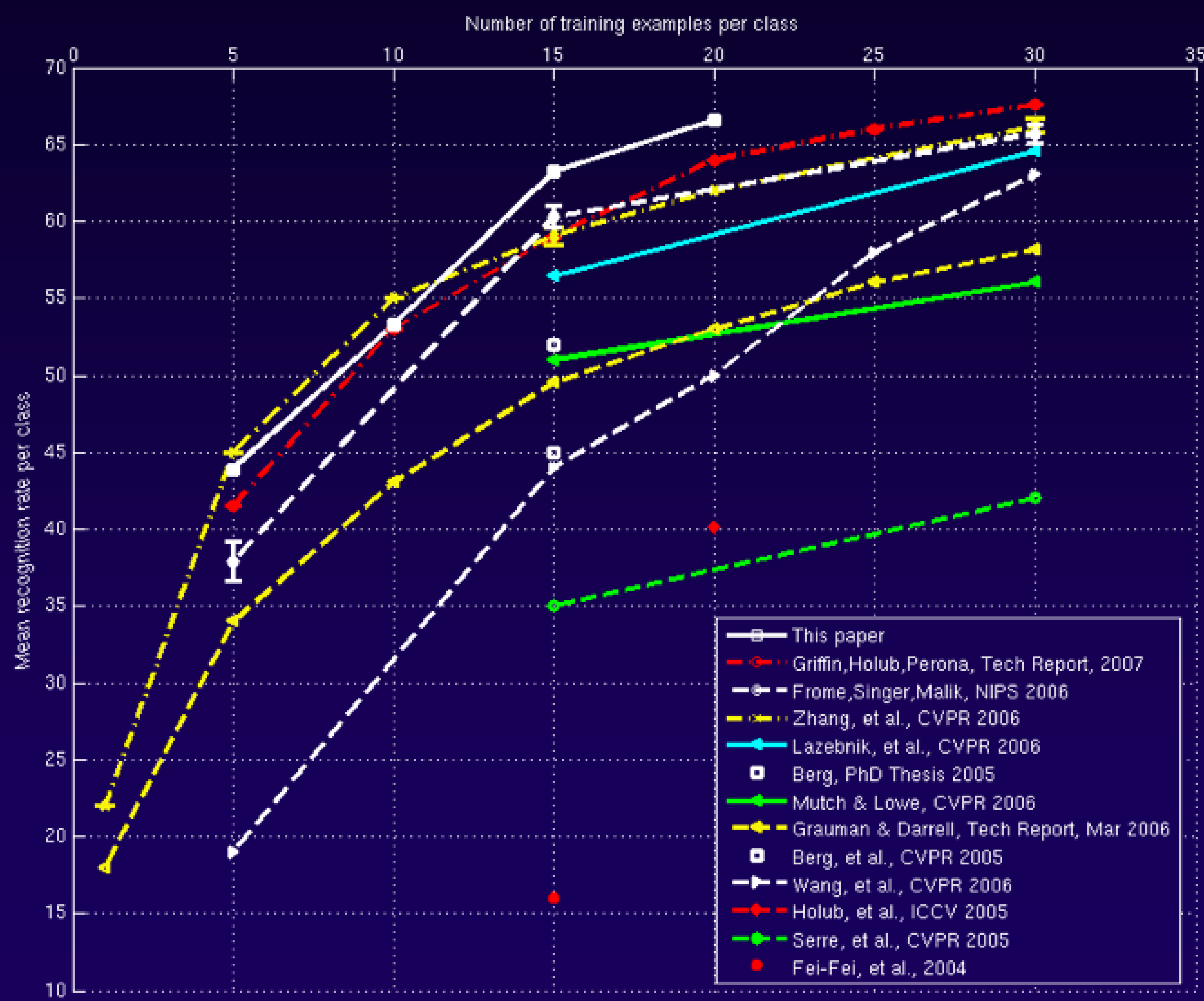
Paradigm introduced by Rowley, Baluja & Kanade 96 for face detection
Viola & Jones 01, Dalal & Triggs 05, Felzenszwalb, McAllester, Ramanan 08

Caltech-101 [Fei-Fei et al. 04]

- 102 classes, 31-300 images/class

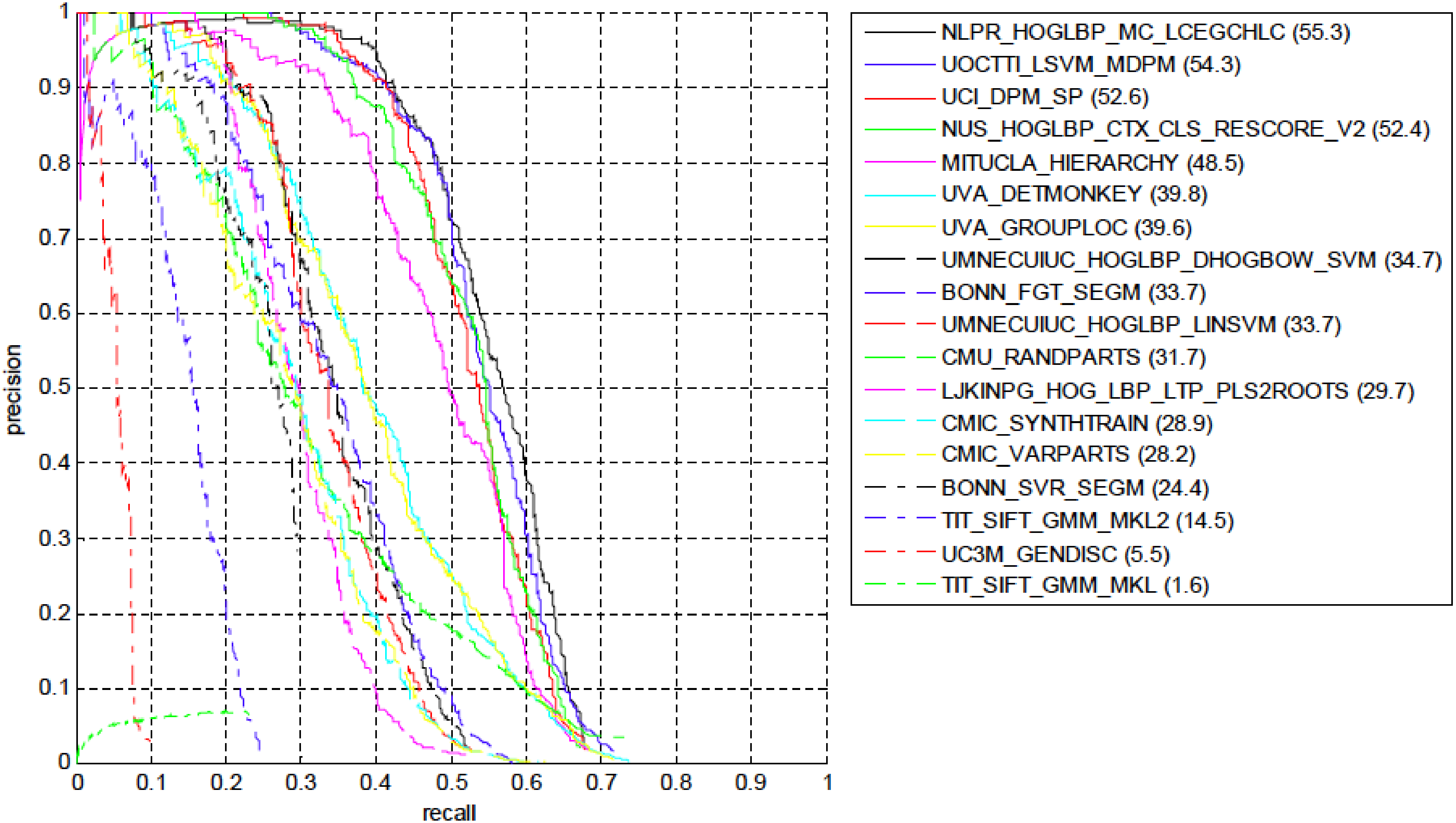
Caltech 101 classification results

(even better by combining cues..)

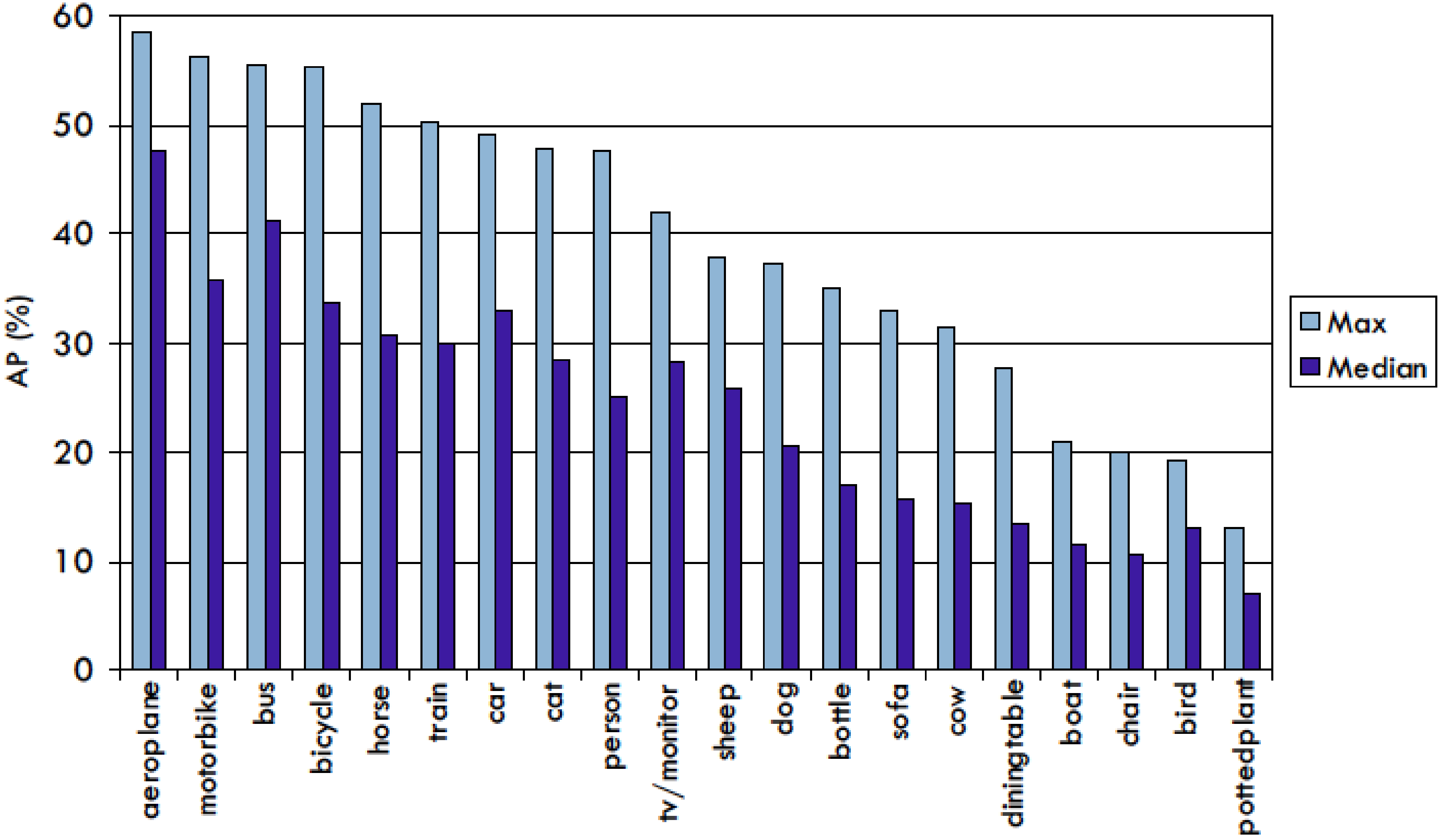


PASCAL Visual Object Challenge

Precision/Recall - Bicycle

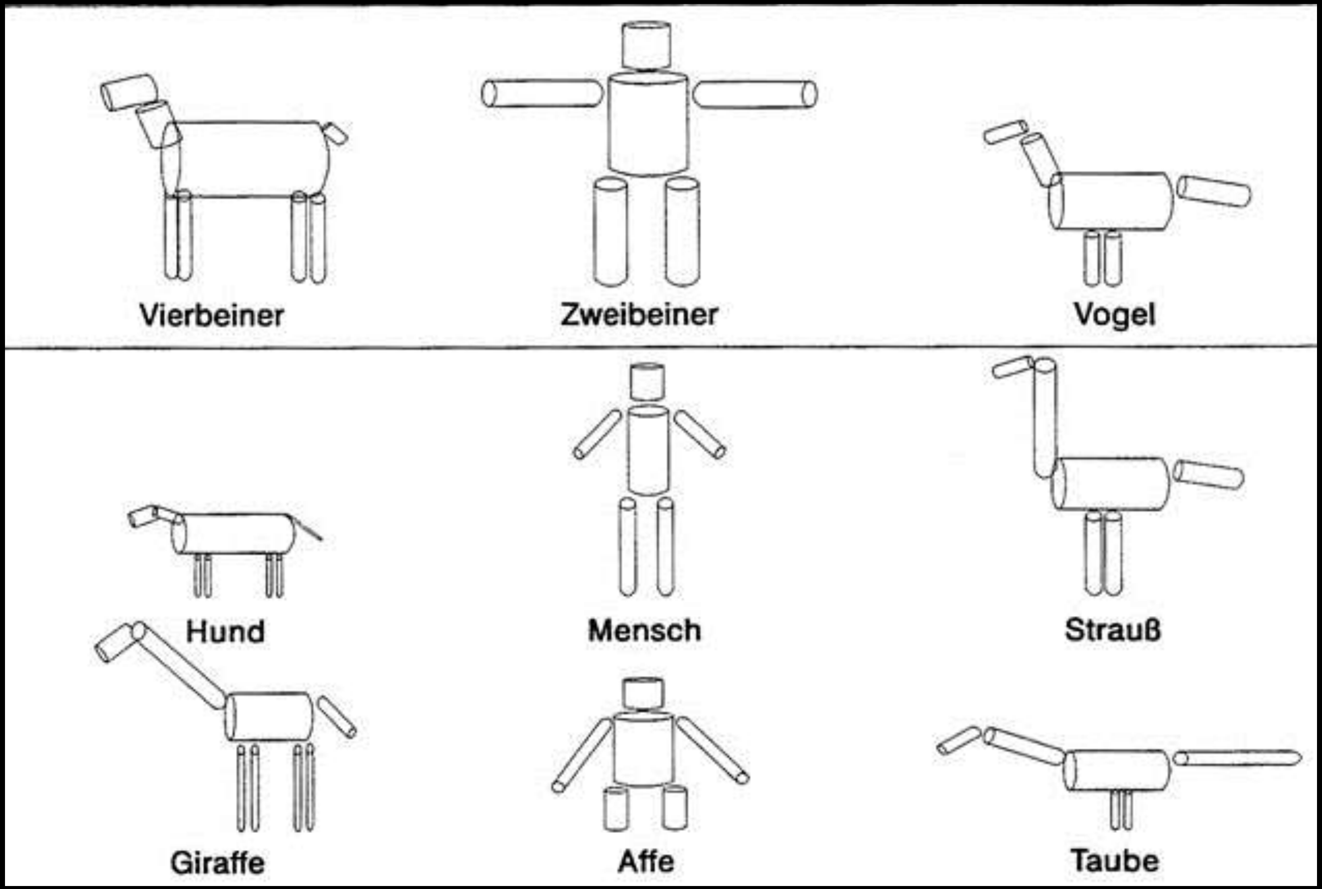


AP by Class



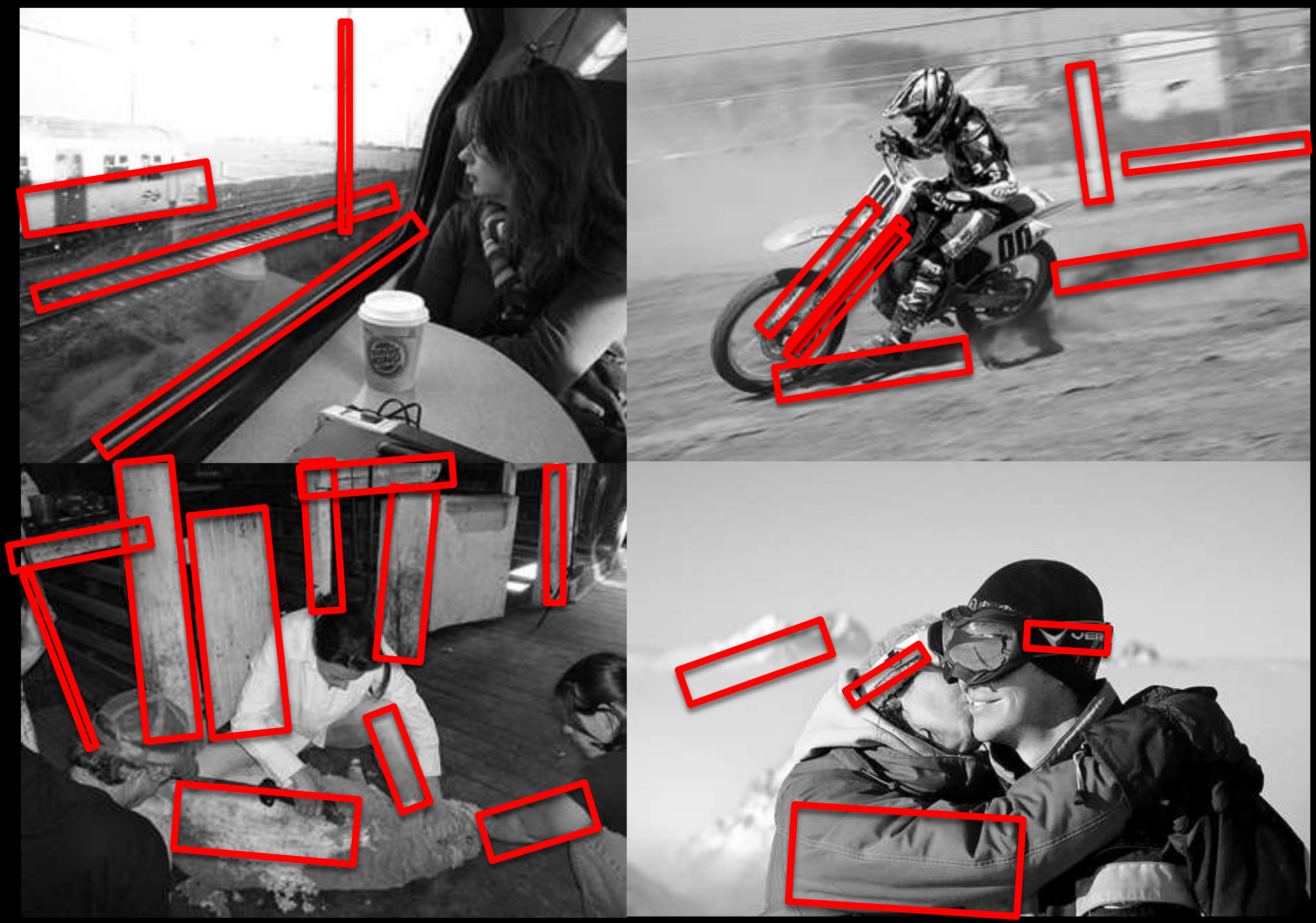
- Max AP: 58.4% (aeroplane) ... 13.0% (potted plant)

Trying to find stick figures is hard (and unnecessary!)



Generalized Cylinders (Binford, Marr & Nishihara)
Geons (Biederman)

Person detection is challenging



Can we build upon the success of faces and pedestrians?

Rowley, Baluja, Kanade CVPR96
Viola and Jones, IJCV01
…

Dalal and Triggs, CVPR05
…

- Pattern matching
- Capture patterns that are common and visually characteristic
- Are these the only two common and characteristic patterns?

Poselets

We will train classifiers for these different visual patterns

Segmenting people

Best person segmentation on PASCAL 2010 dataset

[Bourdev, Maji, Brox and Malik, ECCV10]

Describing people

"A man with short hair, glasses, short sleeves and shorts"

"A man with short hair and long sleeves"

"A woman with long hair, glasses and *long pants*"(??)

"A person with long pants"

Male or female?

Gender classifier per poselet is much easier to train

Is male

Has long hair

Wears long pants

Wears a hat

Wears long sleeves

Wears glasses

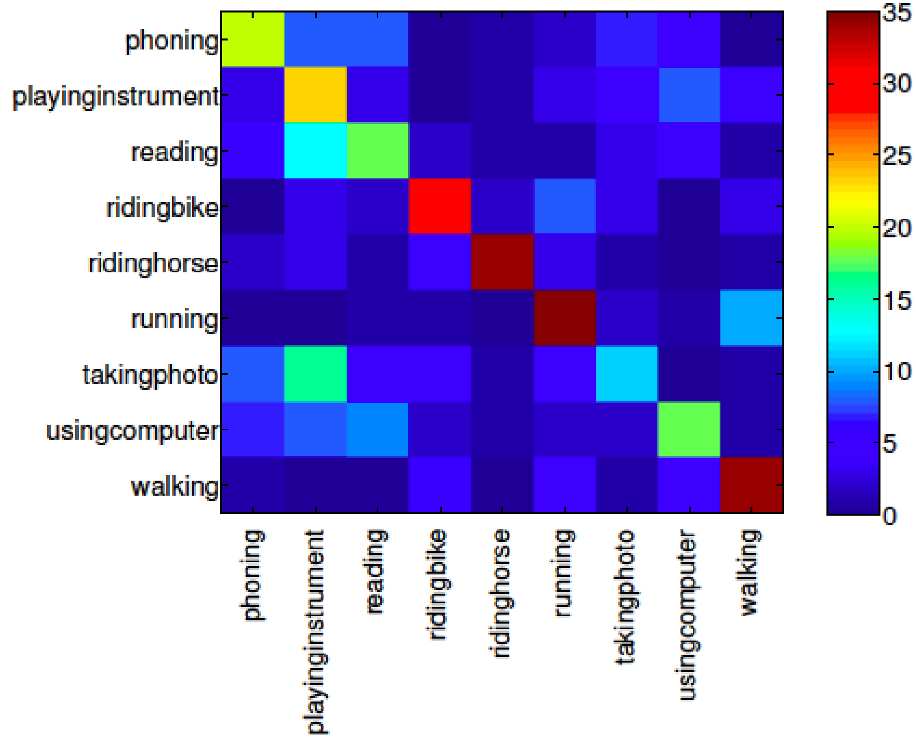
Actions in still images …

- have characteristic :
  - pose and appearance
  - interaction with objects and agents

Some discriminative poselets

Problem: Human Activity Recognition

Approach: Learn pose and appearance specific for an action



Mean Performance: 59.7% correct

Results : Top Confusions

Low-Cost Automated Tuberculosis Diagnostics Using Mobile Microscopy

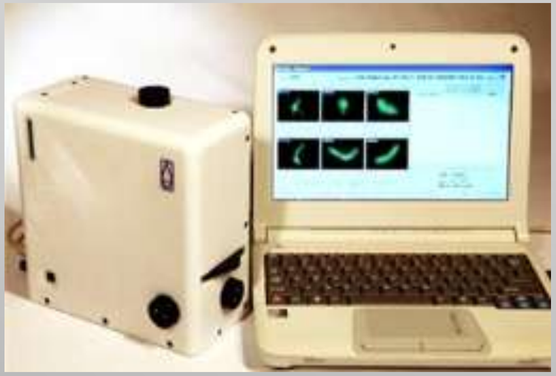
Jeannette Chang[1], Pablo Arbelaez[1], Neil Switz[2], Clay Reber[2], Asa Tapley[2,3]

Lucian Davis[3], Adithya Cattamanchi[3], Daniel Fletcher[2], and Jitendra Malik[1]

Department of Electrical Engineering and Computer Science, UC Berkeley[1]

Department of Bioengineering, UC Berkeley[2]

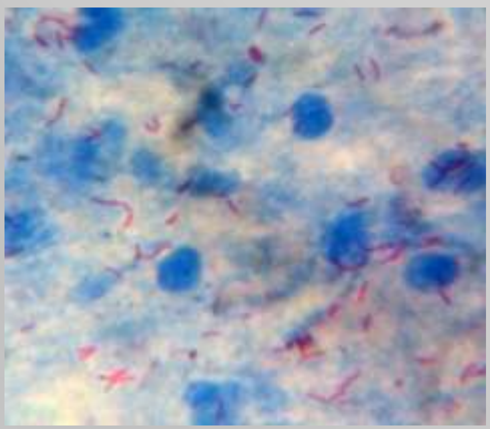
Medical School and San Francisco General Hospital, UC San Francisco[3]

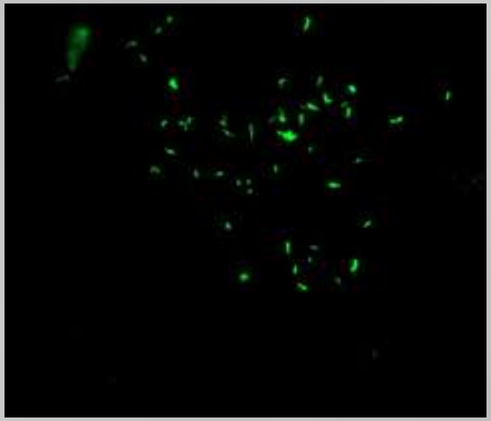


Why Tuberculosis?

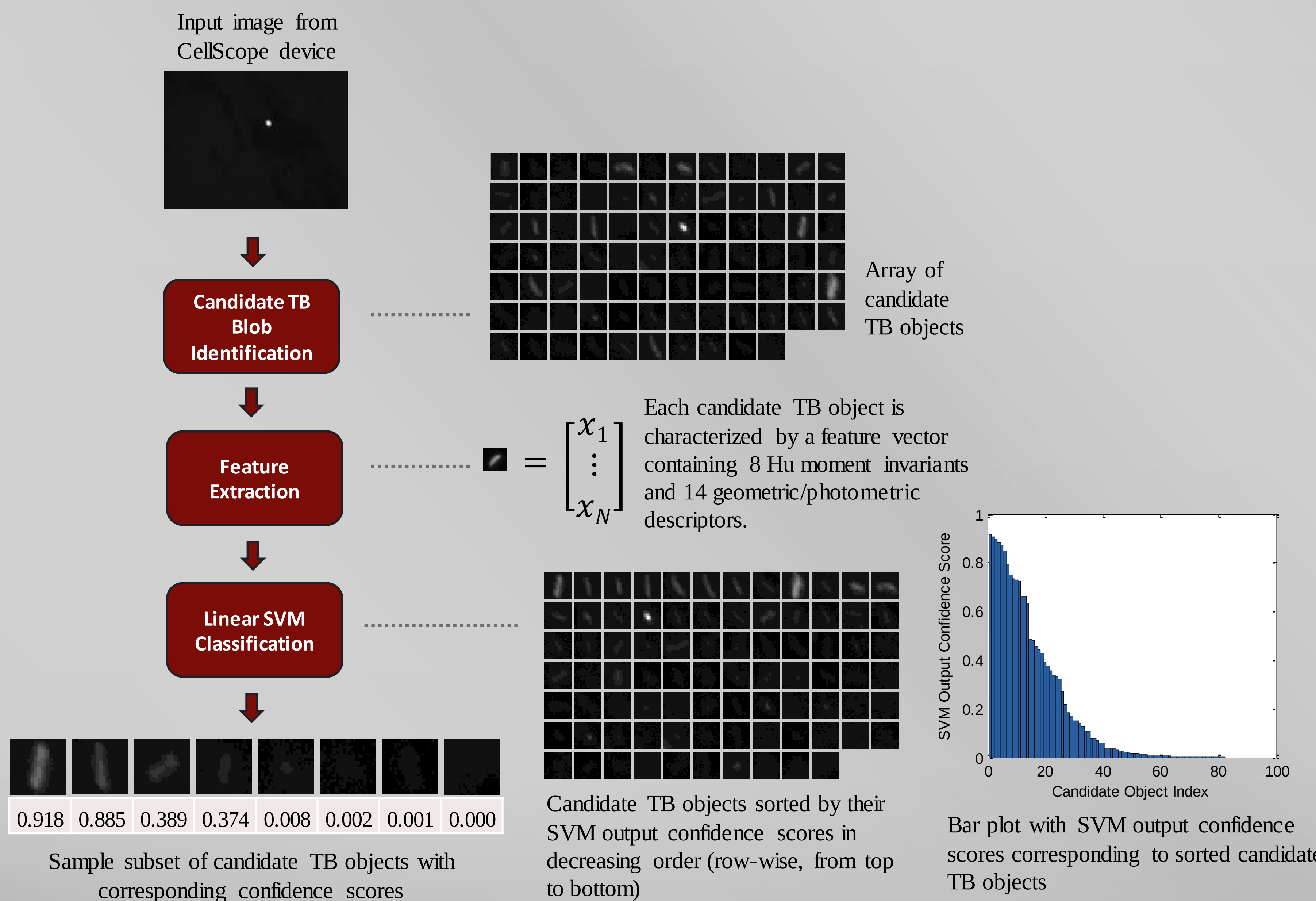
- Mortality and Treatment[1]
  - TB is second leading cause of deaths from infectious disease worldwide (after HIV/AIDS)
  - Highly effective antibiotic treatment
- Current Diagnostics
  - Technicians screen microscopic images of sputum smears manually
  - Other methods include culture and PCR
  - Tremendous potential benefit from automated processing or classification

1. http://www.who.int/tb/publications/global_report/2011/gtbr11_full.pdf
2. http://www.thehindu.com/health/rx/article21138.ece



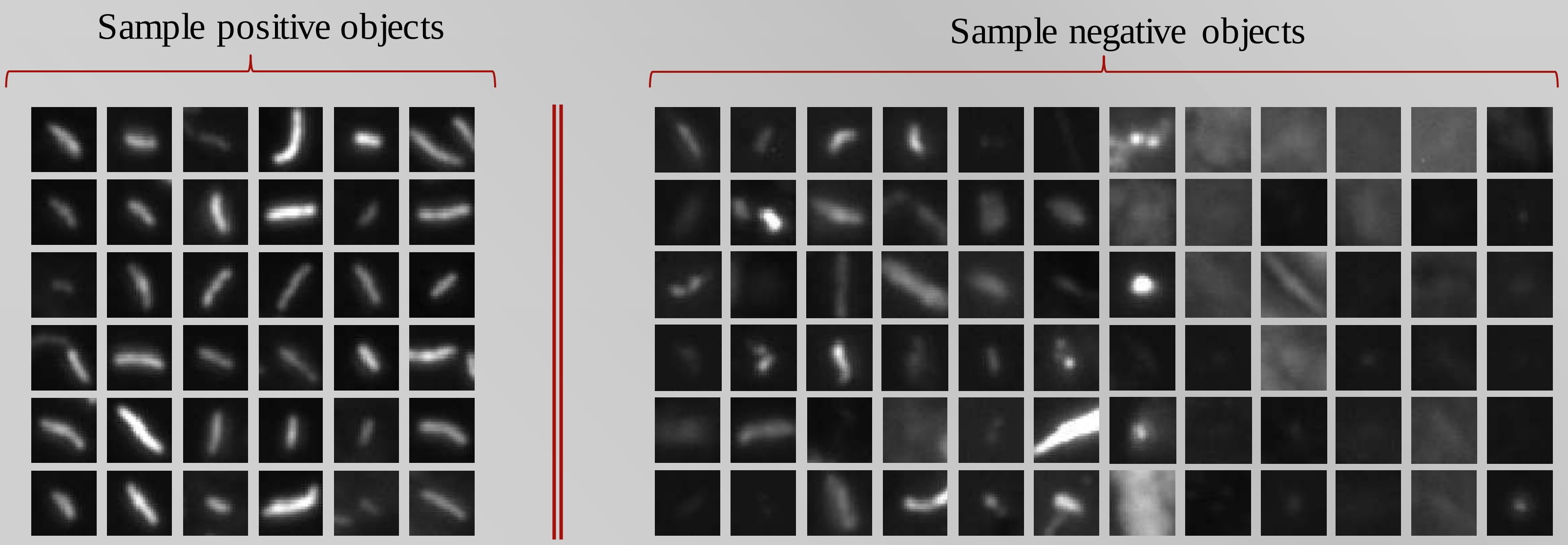


Examples of sputum smears with TB bacteria. Brightfield (top) and fluorescent (bottom) microscopy.[2]

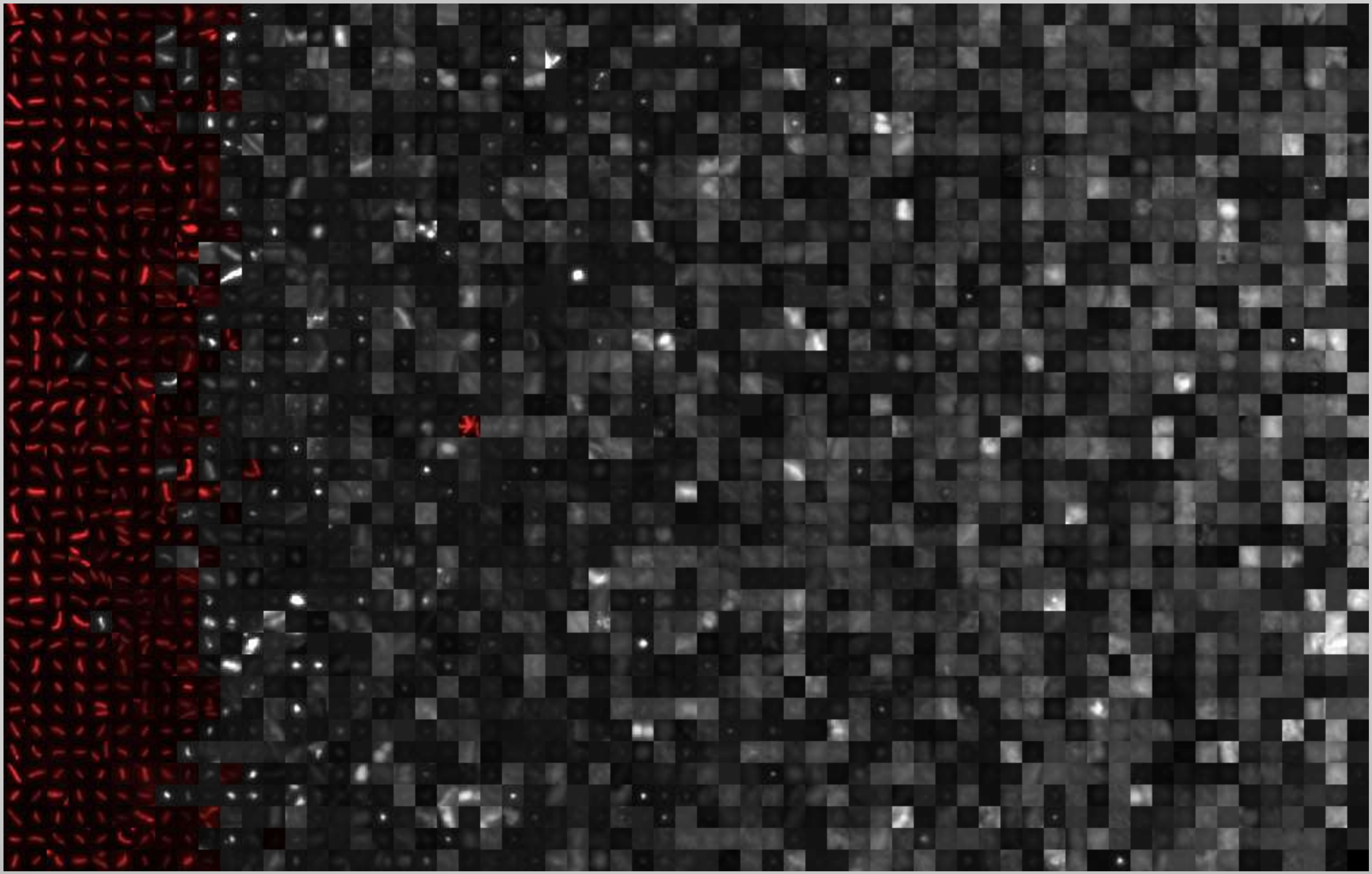


Input image from CellScope device
Candidate TB Blob Identification
Array of candidate TB objects
Feature Extraction
$= \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$
Each candidate TB object is characterized by a feature vector containing 8 Hu moment invariants and 14 geometric/photometric descriptors.
Linear SVM Classification
0.918
0.885
0.389
0.374
0.008
0.002
0.001
0.000
Sample subset of candidate TB objects with corresponding confidence scores
Candidate TB objects sorted by their SVM output confidence scores in decreasing order (row-wise, from top to bottom)
SVM Output Confidence Score
Candidate Object Index
0
0.2
0.4
0.6
0.8
1
0
20
40
60
80
100
Bar plot with SVM output confidence scores corresponding to sorted candidate TB objects

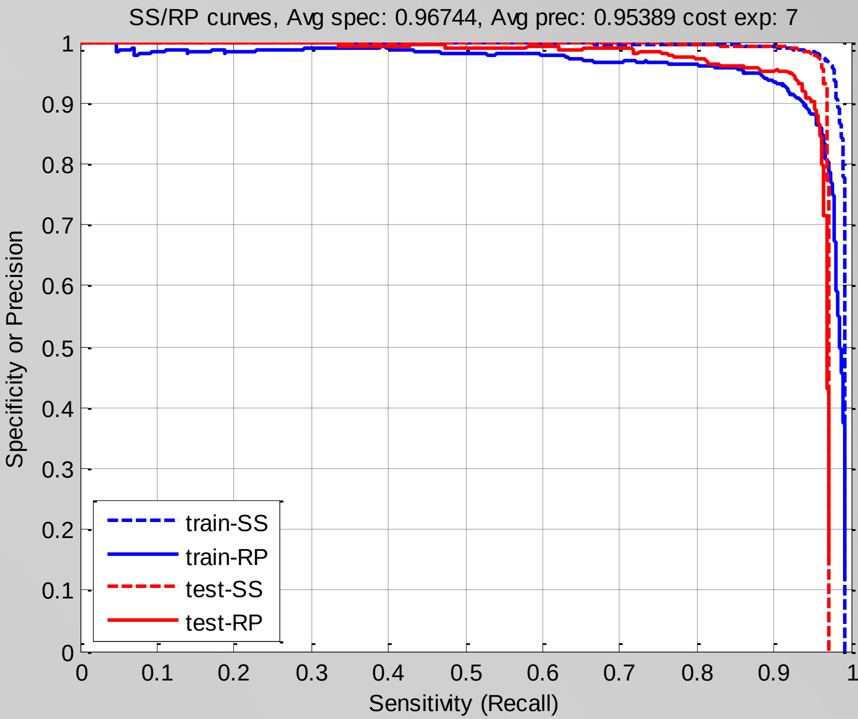
Sample Candidate Objects



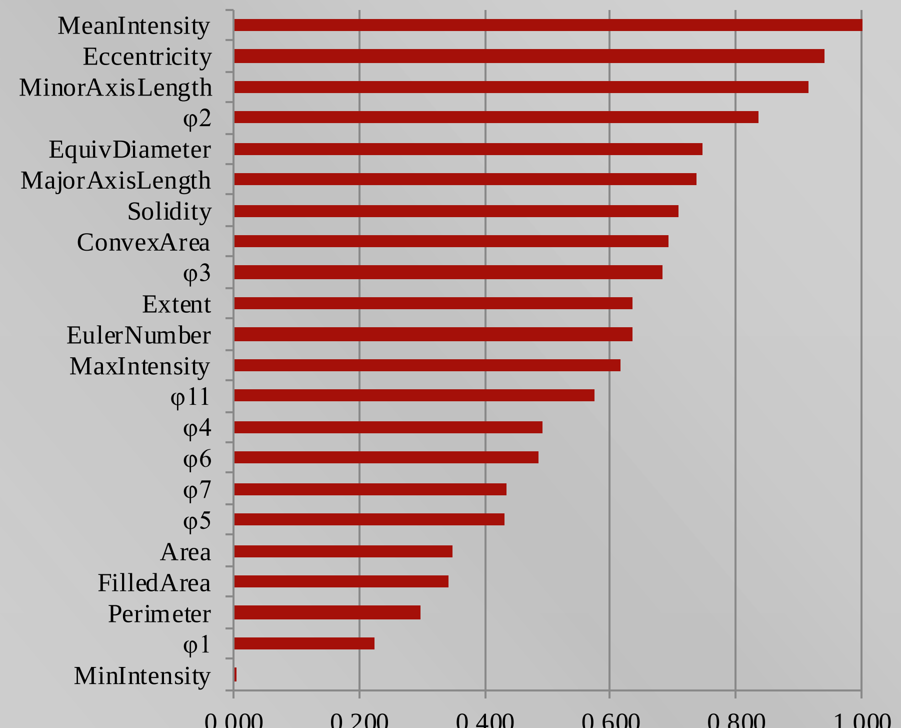
Patches in Descending Order of Confidence



Object-Level Performance (Uganda Data)

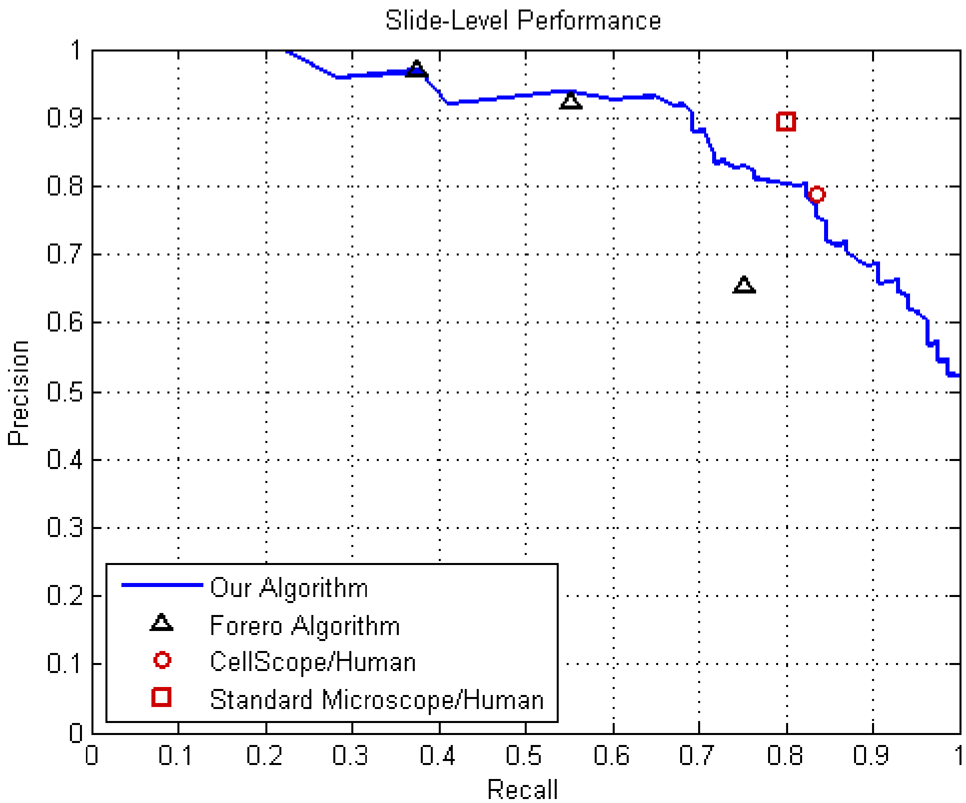


Data Entity	Positive	Negative
Slides (Sputum Smears)	42	45
Images	61	67
TB Bacilli Objects	1574	–



Features listed in descending order of normalized SVM weights.

Slide-Level Performance (Uganda Data)



Data Entity	Positive	Negative
Slides (Sputum Smears)	85	77
Images	235	198