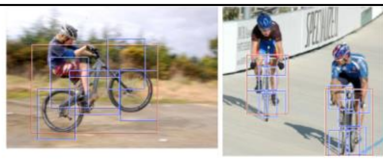


Object detection

Wed Feb 24
Kristen Grauman
UT Austin

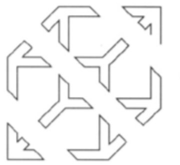
Announcements

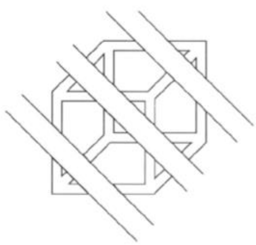
- Reminder: Assignment 2 is due Mar 9 and Mar 10
  - Be ready to run your code again on a new test set on Mar 10
- Vision talk next Tuesday 11 am:
  - Distinguished Lecture
  - Prof. Jim Rehg, Georgia Tech
  - "Understanding Behavior through First Person Vision"

Last time: Mid-level cues

Tokens beyond pixels and filter responses but before object/scene categories

- Edges, contours
- Texture
- Regions
- Surfaces





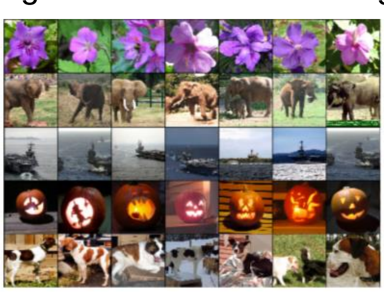
Continuity, explanation by occlusion

http://entertainthis.usatoday.com/2015/09/09/how-tom-hardys-legend-poster-hid-this-hilariously-bad-review/

Today

- Overview of object detection challenges
- Global scene context
  - Torralba's GIST for contextual priming
- Part-based models
  - Deformable part models (brief)
  - Implicit shape models
  - Hough forests
- Evaluating a detector
  - Precision recall
  - Visualizing mistakes

Image classification challenge



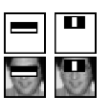
ImageNet

Object detection challenge

PASCAL VOC

Recall: Window-based representations

Four landmark case studies

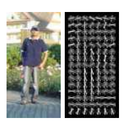


Boosting + face detection

Viola & Jones

NN + scene Gist classification

e.g., Hays & Efros



SVM + person detection

e.g., Dalal & Triggs

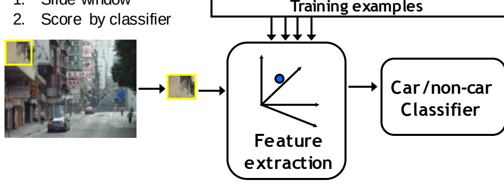
CNNs + image classification

e.g., Krizhevsky et al.

Recall: Window-based object detection

Training:
1. Obtain training data
2. Define features
3. Define classifier

Given new image:
1. Slide window
2. Score by classifier



Kristen Grauman

- What are the pros and cons of sliding window-based object detection?

Window-based detection: strengths

- Sliding window detection and global appearance descriptors:
  - Simple detection protocol to implement
  - Good feature choices critical
  - Past successes for certain classes

Kristen Grauman

Visual Object Recognition Tutorial

Window-based detection: Limitations

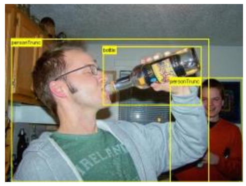
- High computational complexity
  - For example: 250,000 locations x 30 orientations x 4 scales = 30,000,000 evaluations!
  - If training binary detectors independently, means cost increases linearly with number of classes
- With so many windows, false positive rate better be low

Kristen Grauman

Visual Object Recognition Tutorial

Limitations (continued)

- Not all objects are "box" shaped

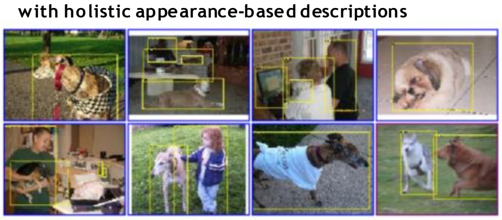


Kristen Grauman

Visual Object Recognition Tutorial

Limitations (continued)

- Non-rigid, deformable objects not captured well with representations assuming a fixed 2d structure; or must assume fixed viewpoint
- Objects with less-regular textures not captured well with holistic appearance-based descriptions



Kristen Grauman

Visual Object Recognition Tutorial

Limitations (continued)

- If considering windows in isolation, context is lost

Sliding window

Detector's view

Figure credit: Derek Hoiem

Kristen Grauman

Limitations (continued)

- In practice, often entails large, cropped training set (expensive)
- Requiring good match to a global appearance description can lead to sensitivity to partial occlusions

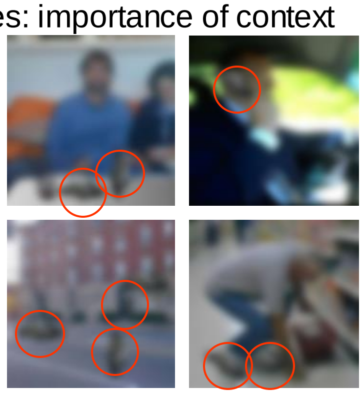
Image credit: Adam, Rivlin, & Shimshoni

Kristen Grauman

Beyond image classification: Issues in object detection

- How to perform localization?
- How to perform efficient search?
- How to represent non-box-like objects? non-texture-based objects? occluded objects?
- How to jointly detect multiple objects in a scene?
- How to handle annotation costs and quality control for localized, cropped instances?
- How to model scene context?

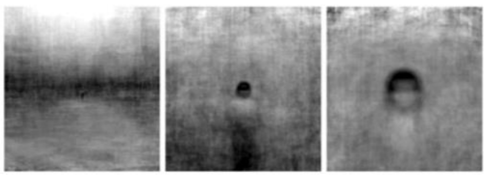
Challenges: importance of context



slide credit Fei-Fei, Fergus & Torralba

Global scene context

Strong relationship between the background and the objects that can be found inside of it



- Contextual Priming for Object Detection. Antonio Torralba. IJCV 2003.

Global scene context

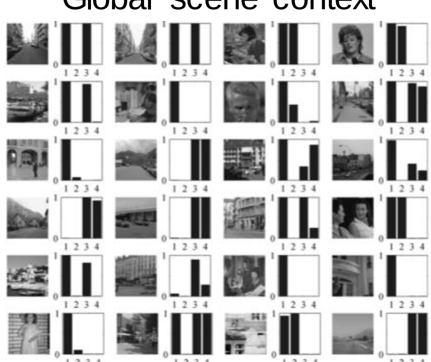
Strong relationship between the background and the objects that can be found inside of it

Given GIST descriptor, represent probability of
- Object being present
- Object being present at a given location/scale

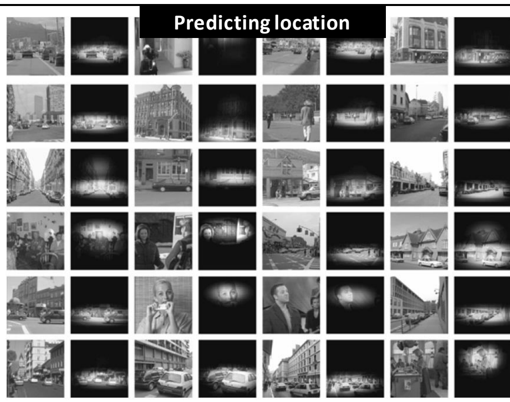
Provides a prior to detector that may help speed or accuracy

- Contextual Priming for Object Detection. Antonio Torralba. IJCV 2003.

Global scene context



Predicting location



- Contextual Priming for Object Detection. Antonio Torralba. IJCV 2003.

Predicting scale

- Contextual Priming for Object Detection. Antonio Torralba. IJCV 2003.

- Video

Today

- Overview of object detection challenges
- Global scene context
  - Torralba's GIST for contextual priming
- Part-based models
  - Deformable part models (brief)
  - Implicit shape models
  - Hough forests
- Evaluating a detector
  - Precision recall
  - Visualizing mistakes

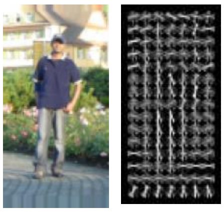
Beyond image classification: Issues in object detection

- How to perform localization?
- How to perform efficient search?
- How to represent non-box-like objects? non-texture-based objects? occluded objects?
- How to jointly detect multiple objects in a scene?
- How to handle annotation costs and quality control for localized, cropped instances?
- How to model scene context?

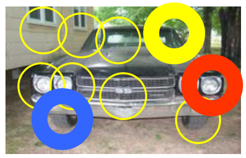
Beyond "window-based" object categories?

Kristen Grauman

Generic category recognition: representation choice



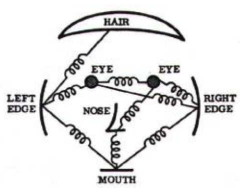
Window-based



Part-based

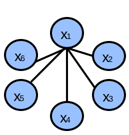
Part-based models

- Origins in Fischler & Elschlager 1973
- Model has two components
  - parts (2D image fragments)
  - structure (configuration of parts)



Shape/structure representation in part-based models

"Star" shape model



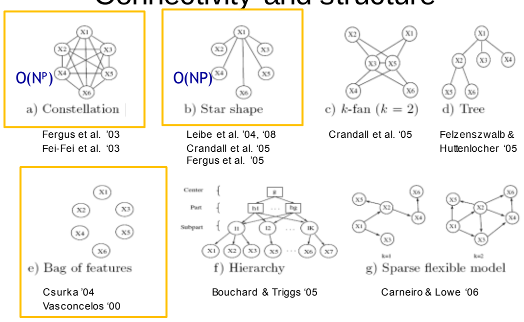
- Parts mutually independent

- Deformable parts model
  - [Felzenszwalb et al.]
- Implicit shape model
  - [Leibe et al.]
- Hough forest
  - [Gall et al.]

N image features, P parts in the model

Kristen Grauman

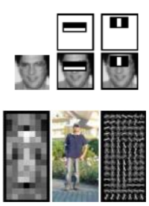
Spatial models: Connectivity and structure



from [Carneiro & Lowe, ECCV'06]

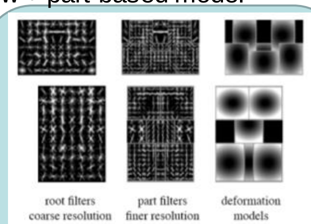
Deformable part model
Felzenszwalb et al. 2008

- A hybrid window + part-based model



Viola & Jones
Dalal & Triggs

vs

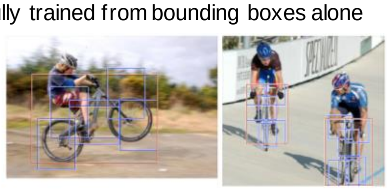


Felzenszwalb et al.

Main idea: Global template ("root filter") plus deformable parts whose placements relative to root are latent variables

Deformable part model
Felzenszwalb et al. 2008

- Mixture of deformable part models
- Each component has global template + deformable parts
- Fully trained from bounding boxes alone



Adapted from Felzenszwalb's slides at http://people.cs.uchicago.edu/~pff/talks/

Beyond image classification: Issues in object detection

- How to perform localization?
- How to perform efficient search?
- How to represent non-box-like objects? non-texture-based objects? occluded objects?
- How to jointly detect multiple objects in a scene?
- How to handle annotation costs and quality control for localized, cropped instances?
- How to model scene context?

Voting algorithms

- It's not feasible to check all combinations of features by fitting a model to each possible subset.
- **Voting** is a general technique where we let the features *vote for all models that are compatible with it*.
  - Cycle through features, cast votes for model parameters.
  - Look for model parameters that receive a lot of votes.
- Noise & clutter features will cast votes too, *but* typically their votes should be inconsistent with the majority of "good" features.

Kristen Grauman

Recall: Hough transform for line fitting

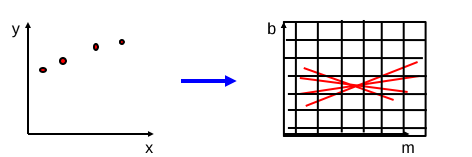


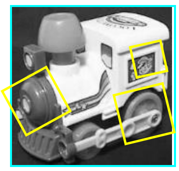
image space Hough (parameter) space

How can we use this to find the most likely parameters (m,b) for the most prominent line in the image space?

- Let each edge point in image space *vote* for a set of possible parameters in Hough space
- Accumulate votes in discrete set of bins; parameters with the most votes indicate line in image space.

Recall: Generalized Hough transform

- A hypothesis generated by a single match may be unreliable,
- So let each match **vote** for a hypothesis in Hough space

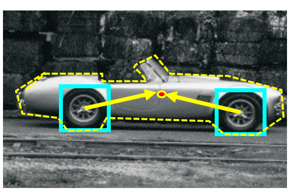


Model

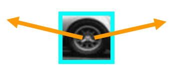
Novel image

Implicit shape models

- Visual vocabulary is used to index votes for object position [a visual word = "part"]



training image annotated with object localization info



visual codeword with displacement vectors

B. Leibe, A. Leonardis, and B. Schiele, Combined Object Categorization and Segmentation with an Implicit Shape Model, ECCV Workshop on Statistical Learning in Computer Vision 2004

Implicit shape models

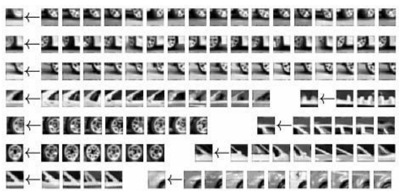
- Visual vocabulary is used to index votes for object position [a visual word = "part"]

test image

B. Leibe, A. Leonardis, and B. Schiele, Combined Object Categorization and Segmentation with an Implicit Shape Model, ECCV Workshop on Statistical Learning in Computer Vision 2004

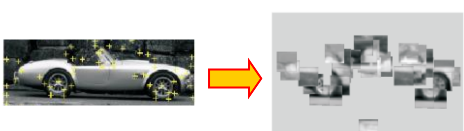
Implicit shape models: Training

1. Build vocabulary of patches around extracted interest points using clustering



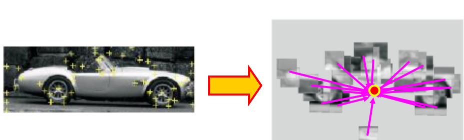
Implicit shape models: Training

1. Build vocabulary of patches around extracted interest points using clustering
2. Map the patch around each interest point to closest word



Implicit shape models: Training

1. Build vocabulary of patches around extracted interest points using clustering
2. Map the patch around each interest point to closest word
3. For each word, store all positions it was found, relative to object center

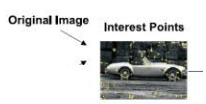


Implicit shape models: Testing

1. Given new test image, extract patches, match to vocabulary words
2. Cast votes for possible positions of object center
3. Search for maxima in voting space
4. (Extract weighted segmentation mask based on stored masks for the codebook occurrences)

What is the dimension of the Hough space?

Implicit shape models: Testing



Example: Results on Cows

Original image

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

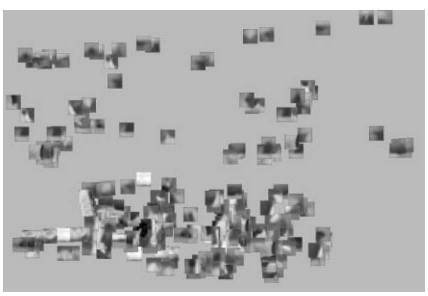
Example: Results on Cows

Interest points

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

Example: Results on Cows

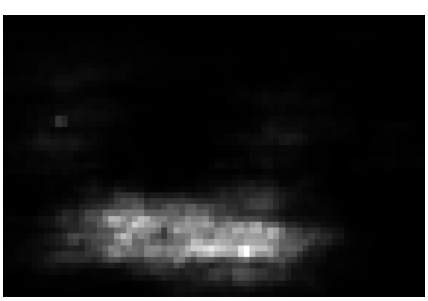


Matched patches

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

Example: Results on Cows



Votes

K. Grauman, B. Leibe

51

Visual Object Recognition Tutorial

Example: Results on Cows

1st hypothesis

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

52

Example: Results on Cows

2nd hypothesis

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

53

Example: Results on Cows

3rd hypothesis

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

54

Detection Results

- Qualitative Performance
  - Recognizes different kinds of objects
  - Robust to clutter, occlusion, noise, low contrast

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

55

Today

- Overview of object detection challenges
- Global scene context
  - Torralba's GIST for contextual priming
- Part-based models
  - Deformable part models (brief)
  - Implicit shape models
  - Hough forests
- Evaluating a detector
  - Precision recall
  - Visualizing mistakes

Class-Specific Hough Forests for Object Detection

Juergen Gall[1] and Victor Lempitsky[2]

[1]BIWI, ETH Zurich
[1]Max-Planck-Institute for Informatics
[2]Microsoft Research Cambridge

Motivation: Hough Forests for object detection

- Parts of an object provide useful spatial information
- Classification of object parts (foreground/background)
- Combine spatial information and class information during learning

Random Forest

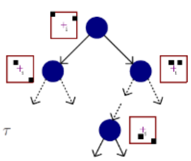
- Image patch:

$$\mathcal{I}_i = (I_i^1, I_i^2, \dots I_i^C)$$

- Binary tests:

$$t_{a,p,q,r,s,\tau}(\mathcal{I}) = \begin{cases} 0, & \text{if } I^a(p,q) < I^a(r,s) + \tau \\ 1, & \text{otherwise.} \end{cases}$$

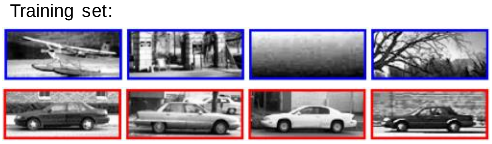
- Binary tests are selected during training from a random subset of all binary tests



Leaf nodes: contain training patches and displacement vectors

Training

- Training set:



$$A = \{\mathcal{P}_i = (\mathcal{I}_i, c_i, \mathbf{d}_i)\}$$

- Class information: c_i (class label)
- Spatial information: $\mathbf{d}_i$ (relative position to object center)

Binary Tests Selection

- Test with optimal split:

$$\underset{k}{\operatorname{argmin}} \Big(U_\star(\{p_i | \, t^k(\mathcal{I}_i){=}0\}) + U_\star(\{p_i | \, t^k(\mathcal{I}_i){=}1\}) \Big)$$

- Class-label uncertainty:

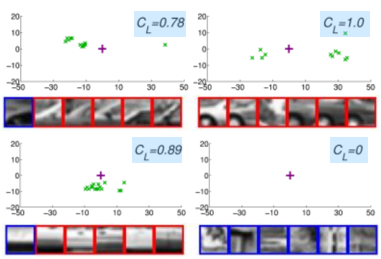
$$U_1(A) = |A| \cdot Entropy(\{c_i\})$$

- Offset uncertainty:

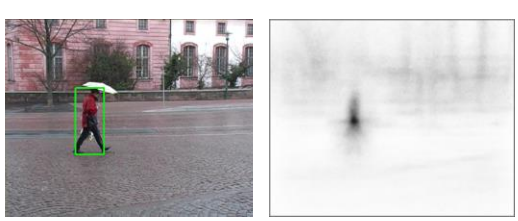
$$U_2(A) = \sum_{i:c_i=1} (\mathbf{d}_i - \mathbf{d}_A)^2$$

- Interleaved: Type of uncertainty is randomly selected for each node

Leaves

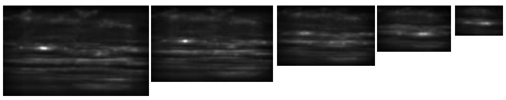


Detection



Multi-Scale and Multi-Ratio

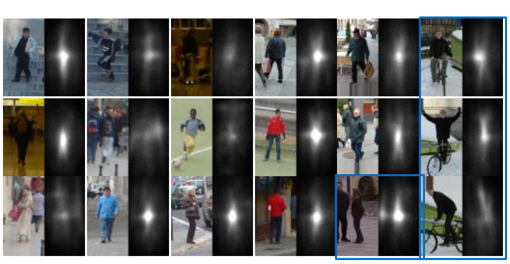
- Multi Scale: 3D Votes (x, y, scale)



Comparison

Methods	UIUC-Single	UIUC-Multi
Hough-based methods		
Implicit Shape Model [10]	91%	–
ISM+verification [10]	97.5%	95%
Boundary Shape Model [17]	85%	–
Random forest based method		
LayoutCRF [27]	93%	–
State-of-the-art		
Mutch and Lowe CVPR'06 [15]	99.9%	90.6%
Lampert et al. CVPR'08 [9]	98.5%	98.6%
Our approach		
Hough Forest	98.5%	98.6%
HF - Weaker supervision	94.4%	–

Pedestrians (INRIA)



Pedestrians (TUD)

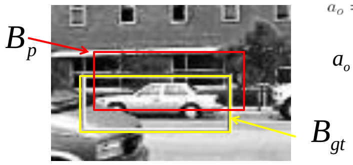
Today

- Overview of object detection challenges
- Global scene context
  - Torralba's GIST for contextual priming
- Part-based models
  - Deformable part models (brief)
  - Implicit shape models
  - Hough forests
- Evaluating a detector
  - Precision recall
  - Visualizing mistakes

Evaluating object detectors

- How accurately is the detector performing?
- What has the detector learned?

Scoring a sliding window detector



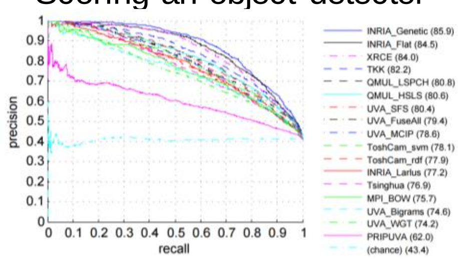
$$a_o = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})}$$

$$a_o > 0.5 \Rightarrow correct$$

We'll say the detection is correct (a "true positive") if the intersection of the bounding boxes, divided by their union, is > 50%.

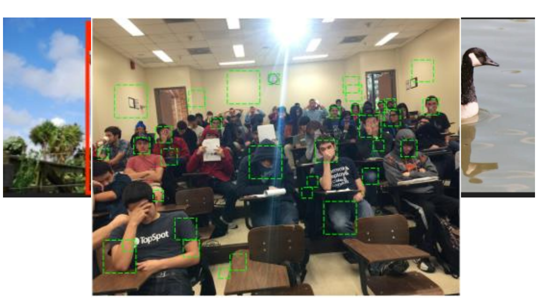
Kristen Grauman

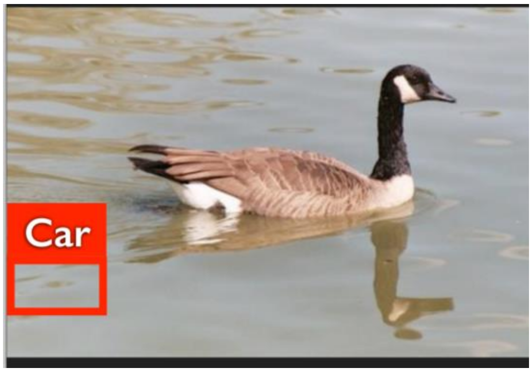
Scoring an object detector



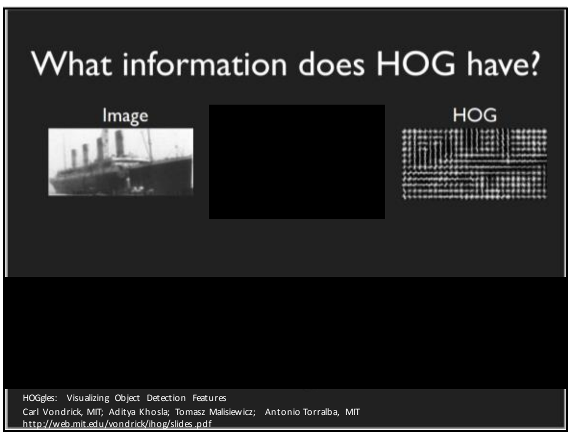
- If the detector can produce a *confidence score* on the detections, then we can plot its precision vs. recall as a threshold on the confidence is varied.
- **Average Precision (AP)**: mean precision across recall levels.

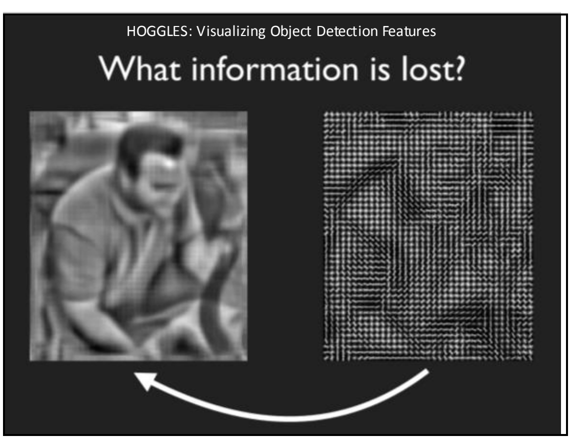
Understanding classifier mistakes





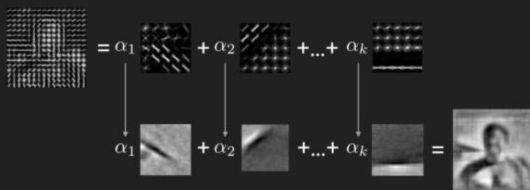
Carl Vondrick http://web.mit.edu/vondrick/ihog/slides.pdf





HOGGLES: Visualizing Object Detection Features

Method: Paired Dictionary



HOGgles: Visualizing Object Detection Features
Carl Vondrick, MIT; Aditya Khosla; Tomasz Malisiewicz; Antonio Torralba, MIT
http://web.mit.edu/vondrick/ihog/slides.pdf

HOGGLES: Visualizing Object Detection Features

A microscope to view HOG

HOGgles: Visualizing Object Detection Features;
Carl Vondrick, MIT; Aditya Khosla; Tomasz Malisiewicz;
Antonio Torralba, MIT
http://web.mit.edu/vondrick/ihog/slides.pdf

HOGGLES: Visualizing Object Detection Features

Human Vision HOG Vision

HOGGLES: Visualizing Object Detection Features

HOGgles: Visualizing Object Detection Features; ICCV 2013
Carl Vondrick, MIT; Aditya Khosla; Tomasz Malisiewicz; Antonio Torralba, MIT
http://web.mit.edu/vondrick/ihog/slides.pdf

Announcements

- Reminder: Assignment 2 is due Mar 9 and Mar 10
  - Be ready to run your code again on a new test set on Mar 10
- Vision talk next Tuesday 11 am:
  - Distinguished Lecture
  - Prof. Jim Rehg, Georgia Tech
  - "Understanding Behavior through First Person Vision"