# Attributes

Tues May 2
Kristen Grauman
UT Austin

## A5 end game

- Deadline extended to Friday EXCEPT for extra credit
- Two leaderboards will be posted
  - Tuesday, Friday
  - Extra credit for top 5 performing submissions

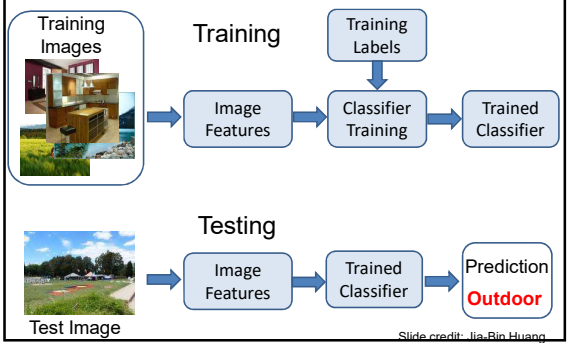
## Final exam

- Tues May 16, 9-12 noon in GDC 1.304
- Comprehensive
- Closed book
- Two pages of notes allowed

## Last time

- Neural networks / multi-layer perceptrons
  - View of neural networks as learning hierarchy of features
- Convolutional neural networks
  - Architecture of network accounts for image structure: local connections, shared weights.
  - "End-to-end" recognition from pixels
  - Together with big (labeled) data and lots of computation → major success on benchmarks, image classification and beyond

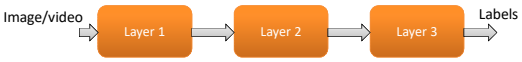
## Recall Traditional Image Categorization




Slide credit: Jia-Bin Huang

## Recall: Learning a Hierarchy of Feature Extractors

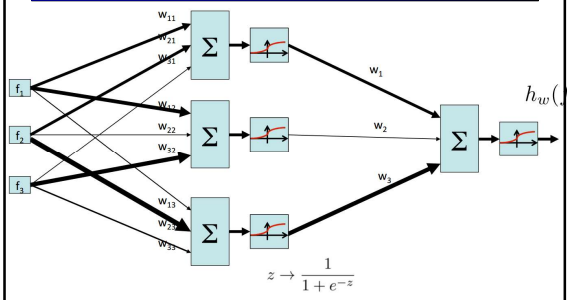
- Each layer of hierarchy extracts features from output of previous layer
- All the way from pixels → classifier
- Layers have the (nearly) same structure




- Train all layers jointly

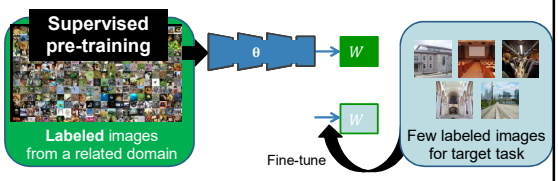
Slide: Rob Fergus

## Recall: Two-layer neural network




Slide credit: Pieter Abeel and Dan Klein

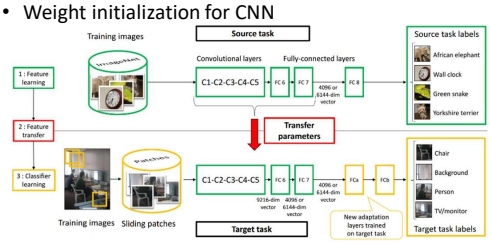
## Pre-training a representation




Slide credit: Kristen Grauman

## Transfer Learning with CNNs

- Improvement of learning in a **new** task through the *transfer of knowledge* from a **related** task that has already been learned.
- Weight initialization for CNN




Slide credit: Jia-Bin Huang

Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks [Oquab et al. CVPR 2014]

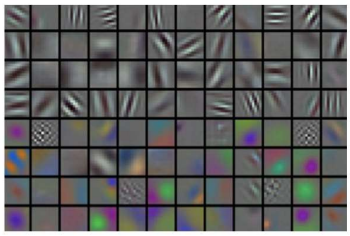
## Understanding and Visualizing CNN

- Find images that maximize some class scores
- Individual neuron activation
- Visualize input pattern using deconvnet



## Recall: visualizing what was learned

- What do the learned filters look like?



Typical first layer filters

## Individual Neuron Activation

 RCNN [Girshick et al. CVPR 2014]

## Individual Neuron Activation

Jia-Bin Huang and Derek Hoiem, UIUC RCNN [Girshick et al. CVPR 2014]

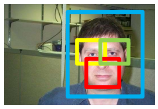
## Individual Neuron Activation

Jia-Bin Huang and Derek Hoiem, UIUC RCNN [Girshick et al. CVPR 2014]

## Recall: Learning Feature Hierarchy

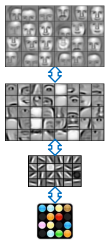
Goal: Learn **useful higher-level features** from images

Input data



Lee et al., ICML 2009; CACM 2011

Feature representation



3rd layer "Objects"

2nd layer "Object parts"

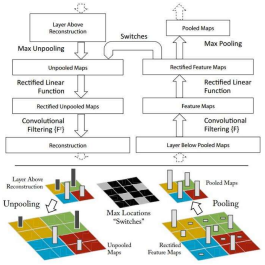
1st layer "Edges"

Pixels

Slide: Rob Fergus

## Map activation back to the input pixel space

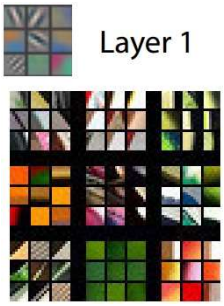
- What input pattern originally caused a given activation in the feature maps?




Visualizing and Understanding Convolutional Networks [Zeiler and Fergus, ECCV 2014]



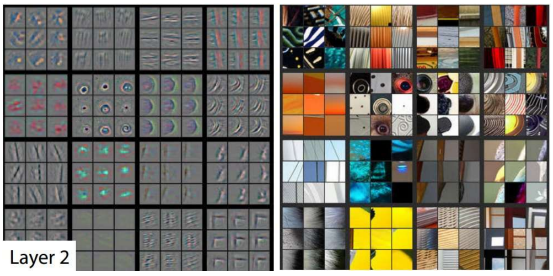
## Layer 1




Visualizing and Understanding Convolutional Networks [Zeiler and Fergus, ECCV 2014]

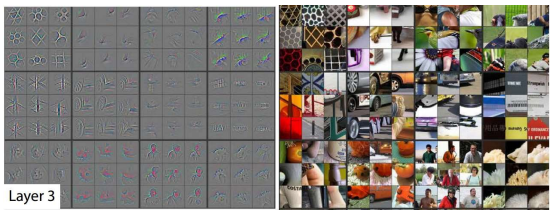


## Layer 2




Visualizing and Understanding Convolutional Networks [Zeiler and Fergus, ECCV 2014]

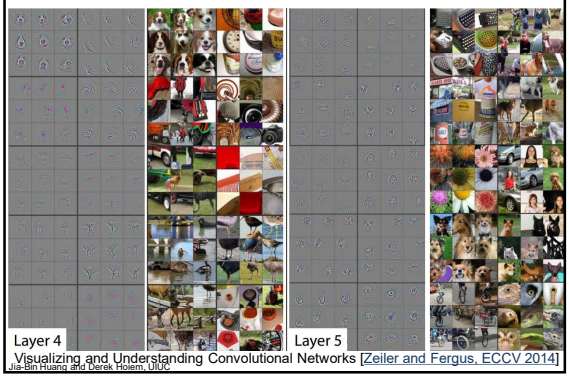
## Layer 3




Visualizing and Understanding Convolutional Networks [Zeiler and Fergus, ECCV 2014]

Jia-Bin Huang and Derek Hoiem, UIUC

## Layer 4 and 5




Visualizing and Understanding Convolutional Networks [Zeiler and Fergus, ECCV 2014]

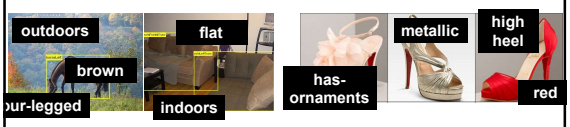
Jia-Bin Huang and Derek Hoiem, UIUC

# Attributes

and learning to rank
and local learning

## What are visual attributes?

- Mid-level semantic properties shared by objects
- Human-understandable *and* machine-detectable




- Material, Appearance, Function/affordance, Parts…
- Adjectives
- Statements *about* visual concepts

*[Oliva et al. 2001, Ferrari & Zisserman 2007, Kumar et al. 2008, Farhadi et al. 2009, Lampert et al. 2009, Endres et al. 2010, Wang & Mori 2010, Berg et al. 2010, Branson et al. 2010, Parikh & Grauman 2011, …]*

## Examples: Binary Attributes

### Facial properties

"Smiling Asian Men With Glasses"


Kumar et al. 2008

## Examples: Binary Attributes

### Object parts and shapes


Farhadi et al. 2009

# Examples: Binary Attributes

## Animal properties

```
otter
black:     yes
white:     no
brown:     yes
stripes:   no
water:     yes
eats fish: yes

polar bear
black:     no
white:     yes
brown:     no
stripes:   no
water:     yes
eats fish: yes

zebra
black:     yes
white:     yes
brown:     no
stripes:   yes
water:     no
eats fish: no
```

Lampert et al. 2009

# Examples: Binary Attributes

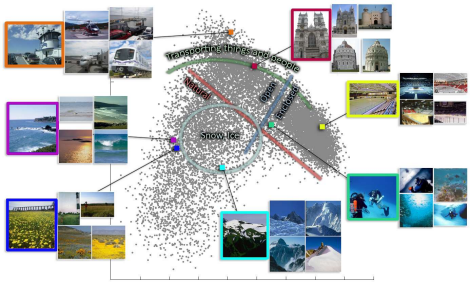
## Animal properties

| | | | |
|---|---|---|---|
| forehead_color | **red** | red | **red** |
| breast_pattern | multi-colored | solid | solid |
| breast_color | white | white/red | white |
| head_pattern | **capped** | capped | **capped** |
| back_color | **white/ black** | **white/ black** | **white/ black** |
| wing_color | **white/ black** | white/black | white/black |
| leg_color | buff | black | black |
| size | **small** | medium | medium |
| bill_shape | **all-purpose** | **dagger** | **all-purpose** |
| wing_shape | **pointed** | tapered | pointed |
| ... | ... | ... | ... |
| primary_color | black, red | white, black | **white, black** |

Welinder et al. 2010

# Examples: Binary Attributes

## Scene properties




Patterson and Hays 2011

## Examples: Binary Attributes

Shopping descriptors


Berg et al. 2010

## Why attributes?

- Why would a robot need to recognize a scene?

Can I walk around here? Is this walkable?

Slide credit: Devi Parikh

## Why attributes?

- Why would a robot need to recognize an object?

How hard should I grip this? Is it brittle?

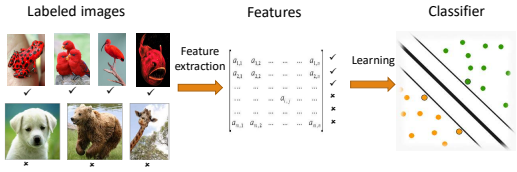
Slide credit: Devi Parikh

## Why attributes?

- How do people naturally describe visual concepts?


Slide credit: Devi Parikh

## Training attribute classifiers




Farhadi et al., CVPR 2009
Kumar et al. , ECCV 2008

Kovashka et al, CVPR 2012
Lampert et al, CVPR 2009

Kumar et al, ECCV 2008
Yu et al, CVPR 2013

Slide credit: Dinesh Jayaraman

## Attributes

*A mule…*

Is furry

Has four legs

Has a tail

## Binary attributes

*A mule…*

Is furry

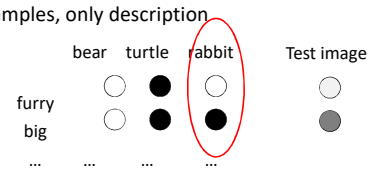
Has four legs

Has a tail

*[Ferrari & Zisserman 2007, Kumar et al. 2008, Farhadi et al. 2009, Lampert et al. 2009, Endres et al. 2010, Wang & Mori 2010, Berg et al. 2010, Branson et al. 2010, …]*

## Zero-shot Learning

- Seen categories with labeled images
  - Train attribute predictors
- Unseen categories
  - No examples, only description




Farhadi et al. 2009, Lampert et al. 2009

36

## Relative attributes

*A mule…*

Is furry

Has four legs

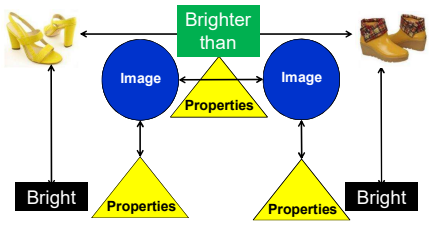
Has a tail

Legs **shorter** than horses'

Tail **longer** than donkeys'

## Relative attributes

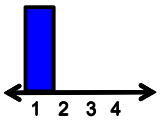
**Idea**: represent ***visual comparisons*** between classes, images, and their properties.




*[Parikh & Grauman, ICCV 2011]*

## How to teach relative visual concepts?

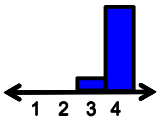
*How much is the person smiling?*




Slide credit: Kristen Grauman

## How to teach relative visual concepts?

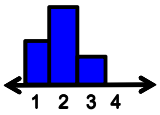
*How much is the person smiling?*




Slide credit: Kristen Graumar

## How to teach relative visual concepts?

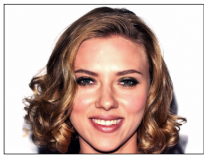
*How much is the person smiling?*

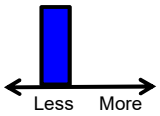



Slide credit: Kristen Graumar

## How to teach relative visual concepts?

> ?

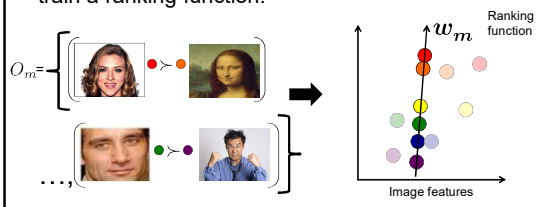





Slide credit: Kristen Graumar

## Learning relative attributes

For each attribute, use ordered image pairs to train a ranking function:




$$\boldsymbol{w}_m^T \boldsymbol{x}_i > \boldsymbol{w}_m^T \boldsymbol{x}_j$$

$$\forall (i,j) \in O_m$$

[Parikh & Grauman, ICCV 2011; Joachims 2002]

Slide credit: Kristen Grauman

## Learning relative attributes

Max-margin learning to rank formulation

$$\min \quad \left(\frac{1}{2}||\boldsymbol{w}_m^T||_2^2 + C\left(\sum \xi_{ij}^2 + \sum \gamma_{ij}^2\right)\right)$$

$$s.t. \quad \boldsymbol{w}_m^T(\boldsymbol{x}_i - \boldsymbol{x}_j) \geq 1 - \xi_{ij}$$

$$|\boldsymbol{w}_m^T(\boldsymbol{x}_i - \boldsymbol{x}_j)| \leq \gamma_{ij}$$

$$\xi_{ij} \geq 0; \gamma_{ij} \geq 0$$

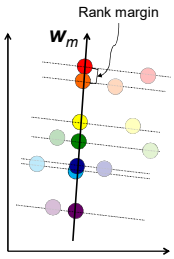



Image → Relative attribute score

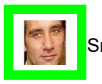
Joachims, KDD 2002

Slide credit: Devi Parikh

## Relating images

Rather than simply label images with their properties,

Not bright



Smiling

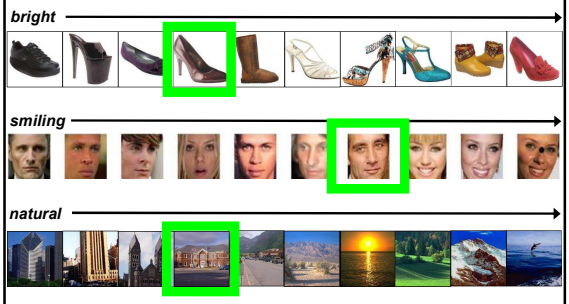
Not natural

Slide credit: Kristen Grauman

[Parikh & Grauman, ICCV 2011]

## Relating images

Now we can compare images by attribute's "strength"




Slide credit: Kristen Grauman

[Parikh & Grauman, ICCV 2011]

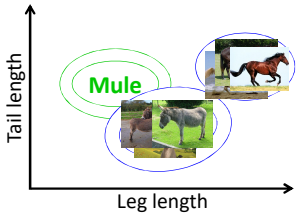
## Relative zero-shot learning

Predict new classes based on their **relationships** to existing classes – even without training images.

Leg length: **Horse** $\succ$ **Mule**

Tail length: **Mule** $\succ$ **Donkey**

⋮

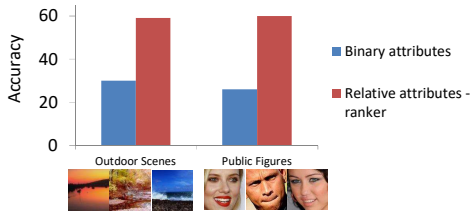



$$\mu_{ijm}^{(q)} \sim \mathcal{N}(\mu_i^{(s)}, \Sigma_i^{(s)})$$

Slide credit: Kristen Grauman

[Parikh & Grauman, ICCV 2011]

## Relative zero-shot learning




Comparative descriptions are more discriminative than categorical definitions.

Slide credit: Kristen Grauman

[Parikh & Grauman, ICCV 2011]

## Attributes for search and recognition

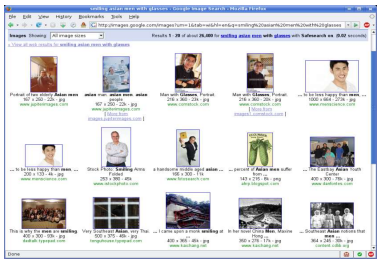
Attributes give human user way to

- Teach novel categories with description
- Communicate search queries
- Give feedback in interactive search
- Assist in interactive recognition

Slide credit: Kristen Grauman

## Image search

- Meta-data commonly used, but insufficient



**Keyword query**: "smiling asian men with glasses"

Slide credit: Kristen Grauman

## Why are attributes relevant to image search?

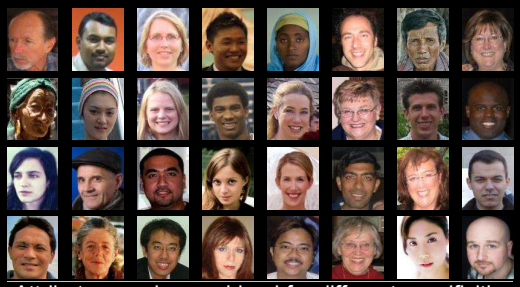
- Human understandable
- Support familiar keyword-based queries

- Composable for different specificities
- Efficiently divide space of images

Slide credit: Kristen Grauman

## Attributes are composable

Caucasian Teeth showing Outside Tilted head



Attributes can be combined for different specificities

Slide credit: Neeraj Kumar

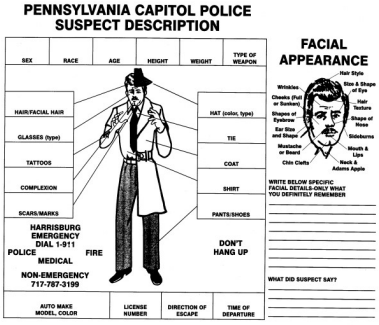
## Attributes efficiently divide the space of images

Female Caucasian Eyeglasses Older

k attributes can distinguish $2^k$ categories

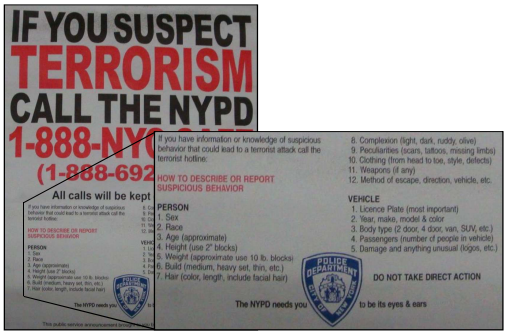
Slide credit: Neeraj Kumar

## Search applications: finding people




Slide credit: Rogerio Feris

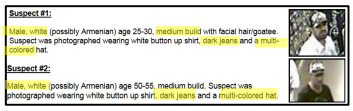
## Search applications: finding people




Slide credit: Rogerio Feris

## Search applications: finding people

Search surveillance feeds for suspects

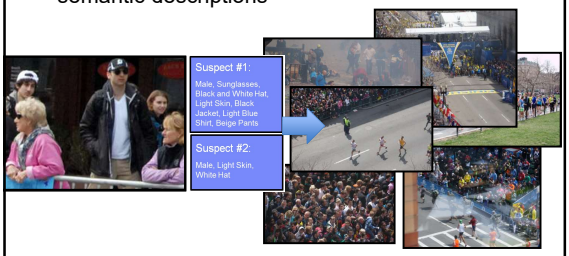




http://lacrimestoppers.com/wanteds.aspx

Slide credit: Rogerio Feris

## Search applications: finding people

Search images from ad hoc cameras using semantic descriptions




Adapted from: Rogerio Feris

## Search applications: finding people

What actress looks like a young Hillary Clinton?


Slide credit: Kristen Grauman

## Face Search with Attributes

Describable Visual Attributes for Face Verification and Image Search, Neeraj Kumar, Alexander C. Berg, Peter N. Belhumeur, Shree K. Nayar, PAMI 2011.

## Attribute Classifier Accuracies

Binary facial attributes in Columbia Face Database

Typically 80%-90% accuracy

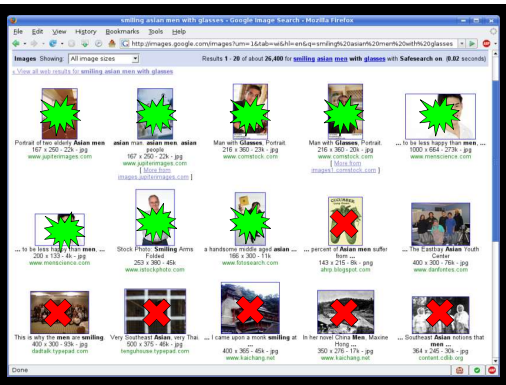
| | | | | | |
|---|---|---|---|---|---|
| gender | 85.78 | hair color: black | 90.82 | flushed face | 88.85 |
| age: young | 87.72 | hair color: blond | 88.39 | chubby | 81.16 |
| age: middle aged | 84.93 | hair color: brown | 74.88 | forehead: fully visible | 89.31 |
| age: senior | 92.04 | hair color: gray | 89.86 | forehead: partially visible | 76.96 |
| race: Asian | 92.32 | hair color: bald | 90.39 | forehead: obstructed | 81.24 |
| race: white | 91.50 | bangs | 91.54 | blurry | 93.42 |
| race: black | 88.65 | receding hairline | 86.83 | color / b&w | 97.88 |
| race: indian | 86.47 | attractive woman | 82.56 | photo type | 71.89 |
| face_shape: oval | 73.30 | attractive man | 74.16 | lighting: soft | 68.46 |
| face_shape: square | 78.60 | eye wear: eyeglasses | 93.32 | lighting: harsh | 77.01 |
| face_shape: round | 75.4 | eye wear: sunglasses | 96.50 | lighting: flash | 73.36 |
| hair_texture: curly | 70.07 | eye wear: none | 93.32 | environment | 85.27 |
| hair_texture: wavy | 66.58 | wearing hat | 89.12 | expression: smiling | 95.91 |
| hair texture: straight | 78.38 | pale skin | 89.36 | expression: frowning | 95.28 |
| heavy makeup | 89.01 | shiny skin | 84.25 | | |

Slide credit: Neeraj Kumar

## FaceTracer: Searching for faces with attributes

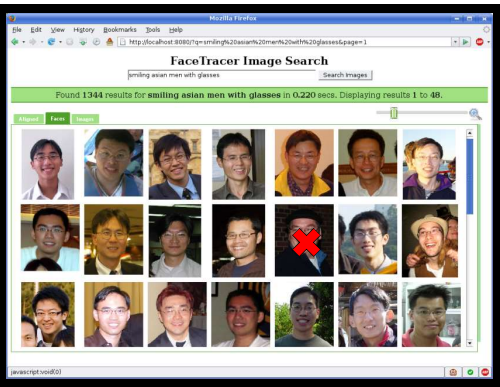
- Offline:
  - Apply attribute classifiers to database images
  - Map classifier outputs to probabilities
- Online:
  - Convey available attribute names to user
  - Given query attributes, rank database images by confidence (e.g., product of probabilities)

Kumar et al.



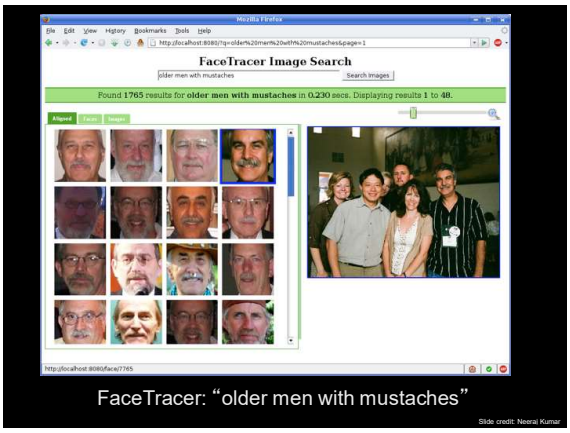
Google: “smiling asian men with glasses”

Slide credit: Neeraj Kumar




FaceTracer: “smiling asian men with glasses”

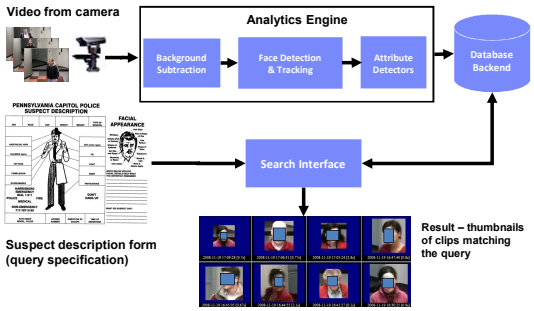
Slide credit: Neeraj Kumar




FaceTracer: "older men with mustaches"

Slide credit: Neeraj Kumar

## Attribute-based person search in video




Vaquero, Feris, Tran, Brown, Hampapur and Turk. *Attribute-Based People Search in Surveillance Environments.*

Slide credit: Rogerio Feris

## Example query: Boston bombing scenario

Rogerio Feris et al., IBM Research

- 1071 detected faces from 50 high-res Boston images (all from Flickr)

Suspect #1: Male, Sunglasses, Black and White Hat, Light Skin, Black Jacket, Light Blue Shirt, Beige Pants

Suspect #2: Male, Light Skin, White Hat

Ability to spot a person with e.g., a white hat in a crowded scene



Suspect #1 found in 4 images in top 8 results

Suspect #2 found in 3 images in top page

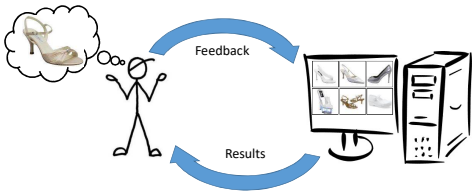
Slide credit: Rogerio Feris

## Problem with one-shot visual search

- Keywords (even attributes) can be insufficient to capture query in one shot.
- Complete "indicator vector" over attributes need not adequately capture envisioned target.

Slide credit: Kristen Grauman

## Interactive visual search




- Iteratively refine the set of retrieved images based on user feedback on results so far
- Potential to communicate more precisely the desired visual content

Slide credit: Adriana Kovashka

## How is interactive search done today?

**Keywords** **+ binary relevance feedback**


- Traditional binary feedback is imprecise
- Coarse communication between user and system

*(Rui et al. 1998, Zhou et al. 2003, Tong & Chang 2001, Cox et al. 2000, Ferecatu & Geman 2007, …)*

## Idea: Search via comparisons

*[Kovashka et al., CVPR 2012]*

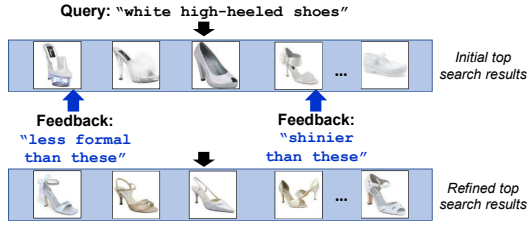
"Like this… but *more* *ornate*"

- Whittle away irrelevant images via comparative feedback on properties of results

Slide credit: Kristen Grauman

## WhittleSearch: Relative attribute feedback

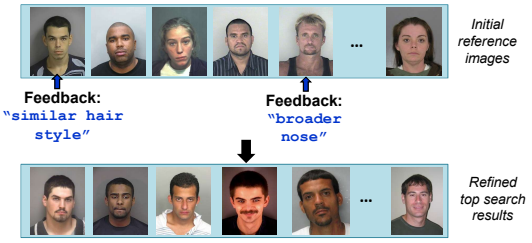
*[Kovashka et al., CVPR 2012]*




Slide credit: Kristen Grauman

## WhittleSearch: Relative attribute feedback

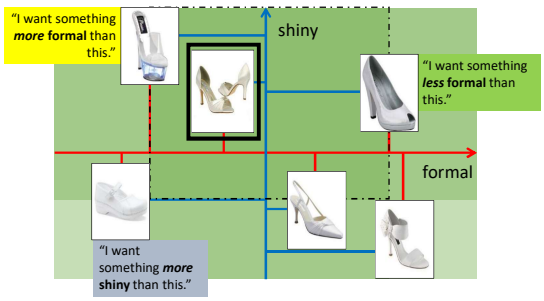
*[Kovashka et al., CVPR 2012]*




Slide credit: Kristen Grauman

## WhittleSearch: Relative attribute feedback

*[Kovashka et al., CVPR 2012]*

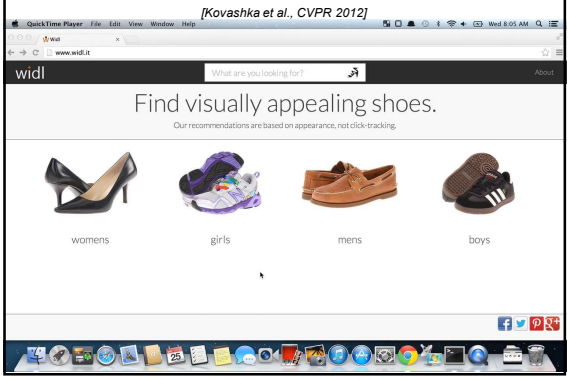



Slide credit: Adriana Kovashka

*[Kovashka, Parikh, and Grauman, CVPR 2012]*

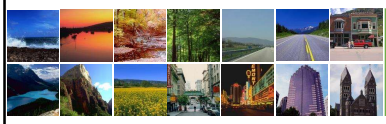
## WhittleSearch: Relative attribute feedback

*[Kovashka et al., CVPR 2012]*




## Datasets

**Shoes:**
14,658 shoe images;
10 attributes:
"pointy", "bright", "high-heeled", "feminine" etc.

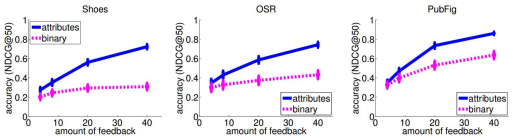


**OSR:**
2,688 scene images;
6 attributes:
"natural", "perspective", "open-air", "close-depth" etc.

**PubFig/LFW:**
772-2000 face images;
11 attributes:
"masculine", "young", "smiling", "round-face", etc.

## WhittleSearch results


Binary relevance feedback **vs.** Relative attribute feedback




More rapidly converge on the envisioned visual content.

*[Kovashka et al., CVPR 2012]*

## **Problem**: Fine-grained attribute comparisons

Coarse

Which is ***more comfortable***?

## **Idea**: Local learning for fine-grained relative attributes

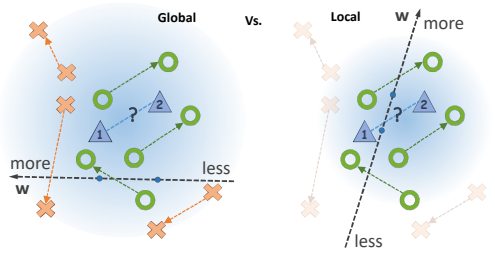
- Lazy: Train query-specific model on the fly.
- Local: Use only pairs that are similar/relevant to test case.

Test comparison

Relevant nearby training pairs

*Yu & Grauman, CVPR 2014*

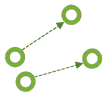
## **Idea**: Local learning for fine-grained relative attributes




*Yu & Grauman, CVPR 2014*

## Learning attribute-specific metrics

Determine neighbor pairs based on learned distance

- E.g., Information-Theoretic Metric Learning [Davis et al. ICML '07]

## UT Zappos50K Dataset

Large shoe dataset, consisting of 50,025 catalog images from Zappos.com

- 4 relative attributes
- High quality pairwise labels from mTurk workers
- 6,751 ordered labels + 4,612 "equal" labels
- 4,334 twice-labeled fine-grained labels (no "equal" option)

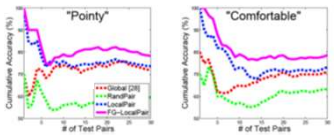

*Yu & Grauman, CVPR 2014*

## Results: Fine-grained attributes

Accuracy of comparisons – all attributes

| | Zap50K-1 | Zap50K-2 | OSR | PubFig |
|---|---|---|---|---|
| RelTree [2] | – | – | 90.41 | 83.37 |
| Global [3] | 89.57 | 61.62 | 88.80 | 80.56 |
| RandPair | 84.34 | 57.98 | 86.93 | 72.46 |
| FG-LocalPair | **91.64** | **66.43** | **92.37** | **89.72** |

Accuracy on the 30 hardest test pairs




**RelTree**: Li et al., ACCV, 2012
**Global**: Parikh & Grauman, ICCV 2011

## Results: Fine-grained attributes

Local model succeeds, global model fails:

More sporty than

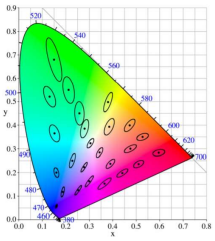
Local model failure cases:

## Just noticeable differences

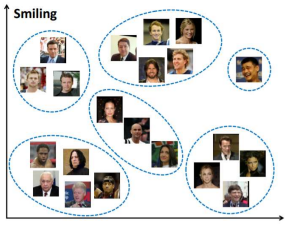
At what point is the relative strength of an attribute in two images indistinguishable?

## Just noticeable differences

Non-trivial: relative attribute space is non-uniform

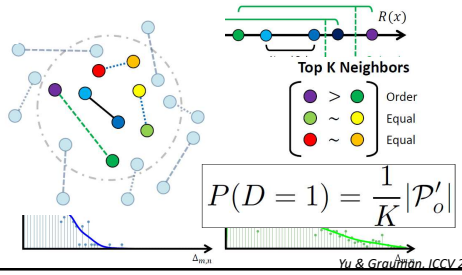


MacAdam ellipses




## Approach: Just noticeable differences

We propose Bayesian local learning strategy to predict whether images $x_m$, $x_n$ are distinguishable:

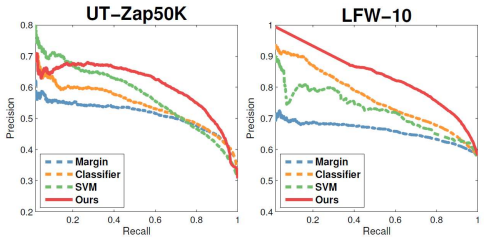



$$P(D=1) = \frac{1}{K}|\mathcal{P}'_o|$$

*Yu & Grauman, ICCV 2015*

## Results: just noticeable differences

Proposed model pinpoints those pairs that are not distinguishable




*Yu & Grauman, ICCV 2015*

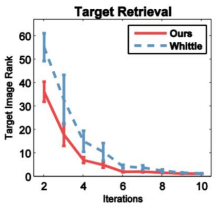
## Visualizing learned JND for an attribute

t-SNE embedding for “pointy”

## Example “indistinguishable” predictions


## Results: just noticeable differences

Positive impact on WhittleSearch: user can now express “equal” constraints




“Something *similarly streamlined* like this!”

*Yu & Grauman, ICCV 2015*

## Attributes for search and recognition

Attributes give human user way to

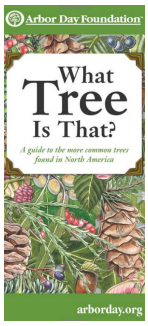
- Teach novel categories with description
- Communicate search queries
- Give feedback in interactive search
- Assist in interactive recognition

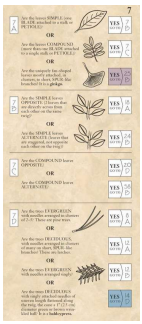
Slide credit: Kristen Grauman

## What Plant Species is This?

Slide credit: Neeraj Kumar

## Let's Use a Field Guide



Slide credit: Neeraj Kumar

## Categories of Recognition

Basic-Level

Airplane? Chair? Bottle? …

| Humans |
| --- |
| Easy |
| **Computers** |
| Some Success |

Slide credit: Steve Branson

## Recognition With Humans in the Loop

*Wah et al., Multi-class Recognition and Part Localization with Humans in the Loop, ICCV 2011*

Computer Vision → Cone-shaped Beak? yes → Computer Vision → American Goldfinch? yes

- Computers: reduce number of required questions
- Humans: drive up accuracy of vision algorithms

Slide credit: Steve Branson

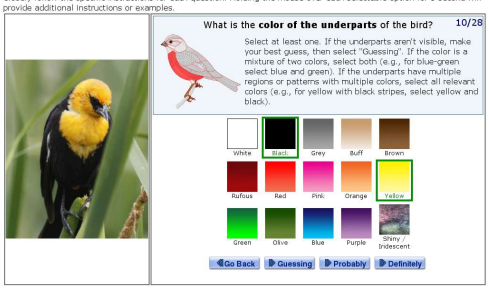
## Example Questions: Localize


*Wah et al., Multi-class Recognition and Part Localization with Humans in the Loop, ICCV 2011*

Slide credit: Steve Branson

## Example Questions: Name attributes

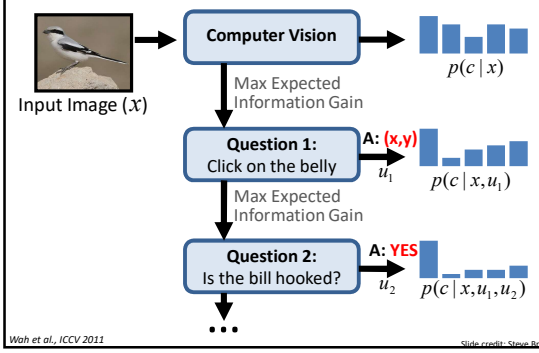
You will be asked to answer a series of questions based on identifying visual features from the bird image on the left. Closely follow the specific instructions for each question. Holding the mouse over each selectable option for 1 second will provide additional instructions or examples.




*Wah et al., ICCV 2011*

Slide credit: Steve Branson

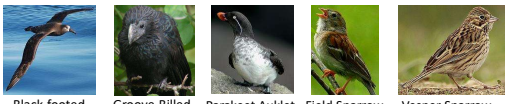
## Basic Algorithm




*Wah et al., ICCV 2011*

Slide credit: Steve Branson

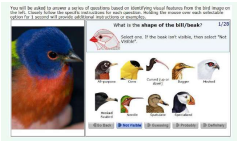
## CUB-200-2011 Dataset



Black-footed Albatross | Groove-Billed Ani | Parakeet Auklet | Field Sparrow | Vesper Sparrow

11,877 images, 200 bird species

13 part locations

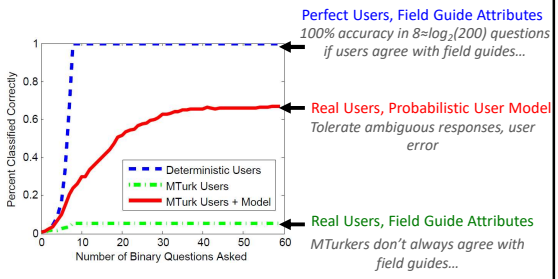


288 binary attributes

*Wah et al., ICCV 2011*

Slide credit: Steve Branson

## Results: Without Computer Vision



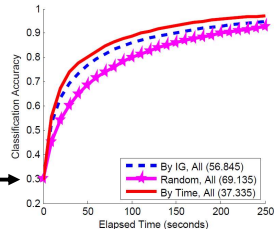

*Branson et al., ECCV 2010*

Slide credit: Steve Branson

## Results: With Computer Vision

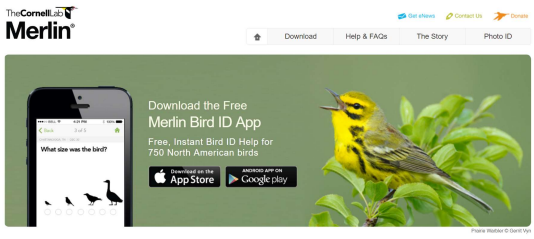
- Incorporating computer vision reduces ave time to identify true species from 109 sec to 37 sec
- Intelligently selecting questions reduces ave time from 69 sec to 37 sec

Base Computer Vision performance (30%)




*Wah et al., ICCV 2011*

Slide credit: Steve Branson




What's Merlin? Watch now!

## Coming up

- Last lecture on Thursday
- Course wrap-up
- Applications and frontiers of computer vision