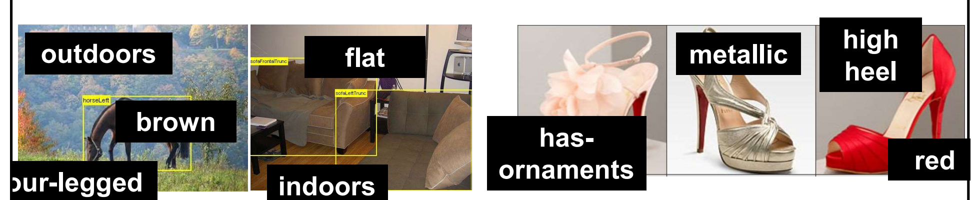
Attributes

Sept 28, 2016
Kristen Grauman
UT Austin

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

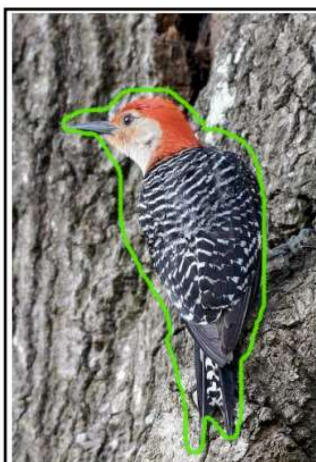
zebra

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

Lampert et al. 2009

Examples: Binary Attributes

Animal properties

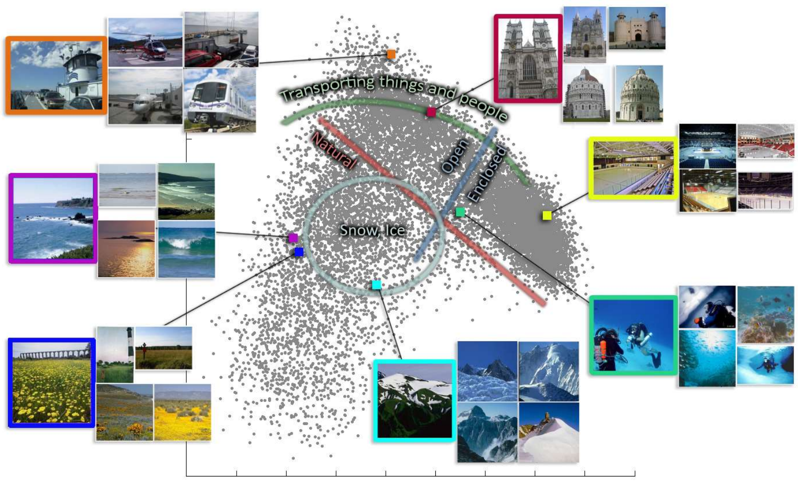


forehead_color	**red**	red	**red**
breast_pattern	multi-colored	solid	solid
breast_color	white	white/red	white
head_pattern	**capped**	capped	**capped**
back_color	**white/ black**	**white/ black**	**white/ black**
wing_color	**white/ black**	white/black	white/black
leg_color	buff	black	black
size	**small**	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

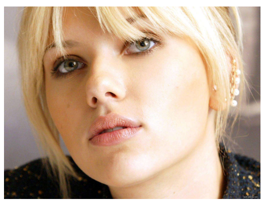
Shopping descriptors

Berg et al. 2010

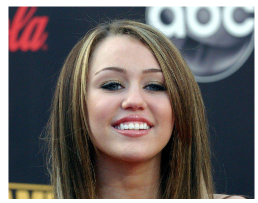
Examples: Relative Attributes

Comparative properties

>
more
natural



<
less
smiling



Parikh and Grauman 2011

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?

I want elegant silver sandals with high heels

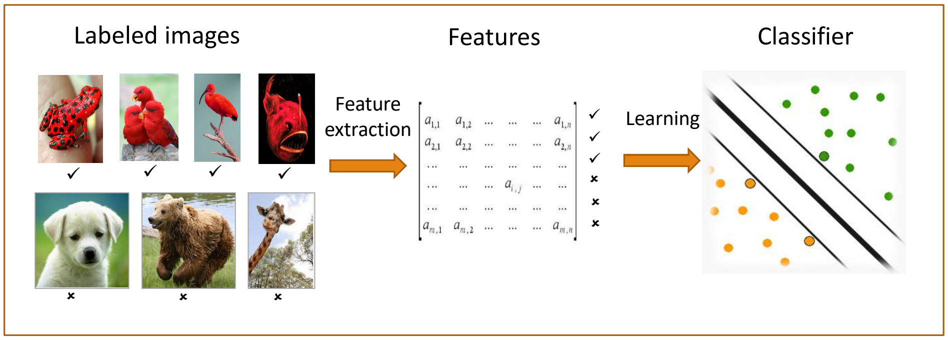
Image search

Zebras have stripes.

Semantic "teaching"

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

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

Attributes

A mule…

Is furry

Has four legs

Has a tail

Binary attributes

A mule…

Is furry

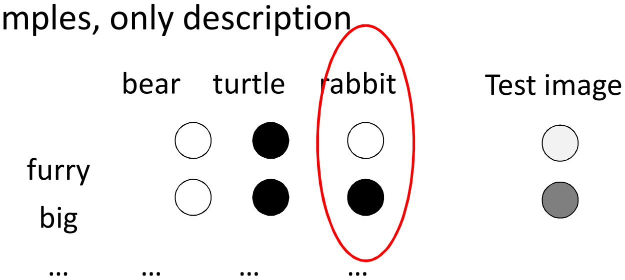
Has four legs

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

18

Relative attributes

A mule…

Is furry

Has four legs

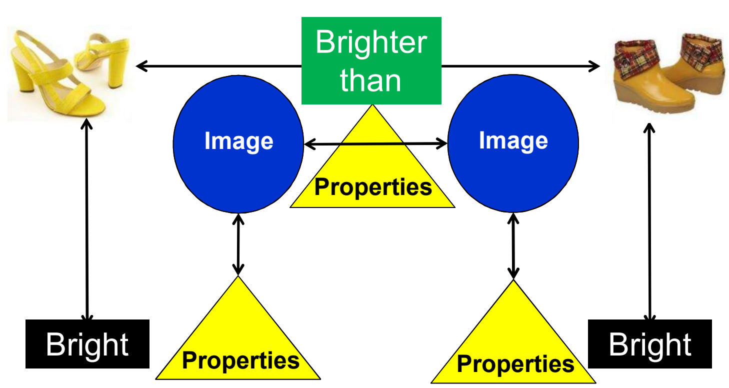
Has a tail

Legs **shorter** than horses'

Tail **longer** than donkeys'

Relative attributes

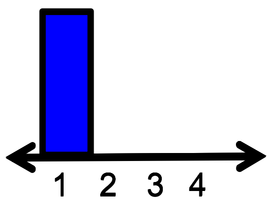
Idea: represent ***visual comparisons*** between classes, images, and their properties.



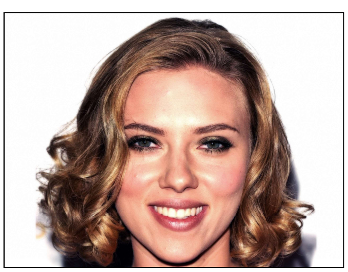
[Parikh & Grauman, ICCV 2011]

How to teach relative visual concepts?

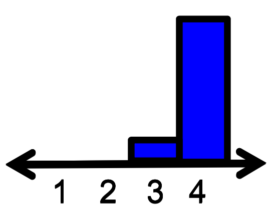
How much is the person *smiling?*



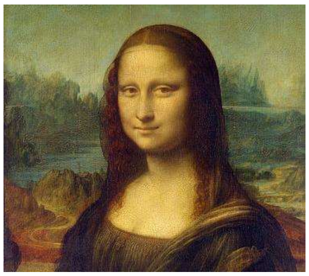
How to teach relative visual concepts?



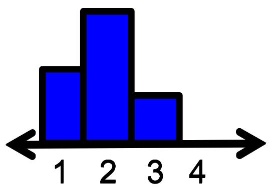
How much is the person *smiling?*



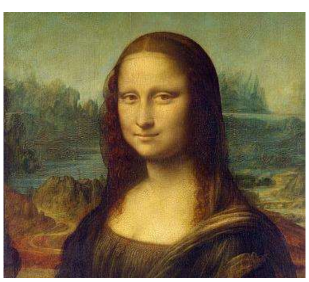
How to teach relative visual concepts?



How much is the person *smiling?*

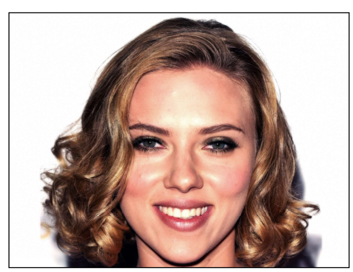


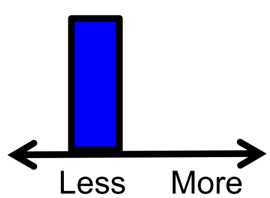
How to teach relative visual concepts?



>

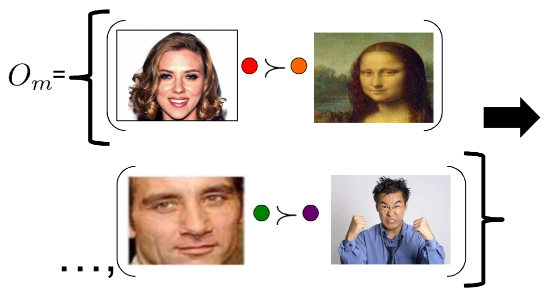
?

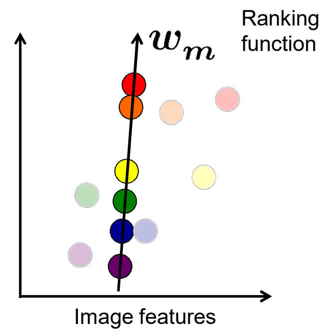




Learning relative attributes

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





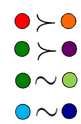
$$\boldsymbol{w_m^T x_i} > \boldsymbol{w_m^T x_j}$$

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

[Parikh & Grauman, ICCV 2011; Joachims 2002]

A13

Learning relative attributes



Max-margin learning to rank formulation

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

$$s.t. \quad \boldsymbol{w_m^T}(\boldsymbol{x_i} - \boldsymbol{x_j}) \geq 1 - \xi_{ij}$$

$$|\boldsymbol{w_m^T}(\boldsymbol{x_i} - \boldsymbol{x_j})| \leq \gamma_{ij}$$

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

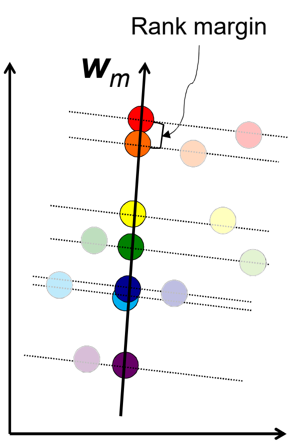


Image → Relative attribute score

Joachims, KDD 2002

Slide credit: Devi Parikh

Slide 26

A13 image space - GIST, color
Adriana, 5/20/2013

Relating images

Rather than simply label images with their properties,

Relating images

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

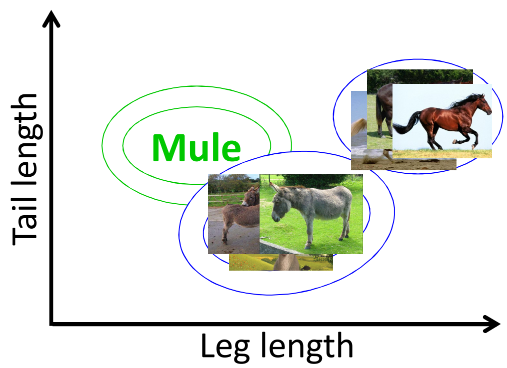
Relative zero-shot learning

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

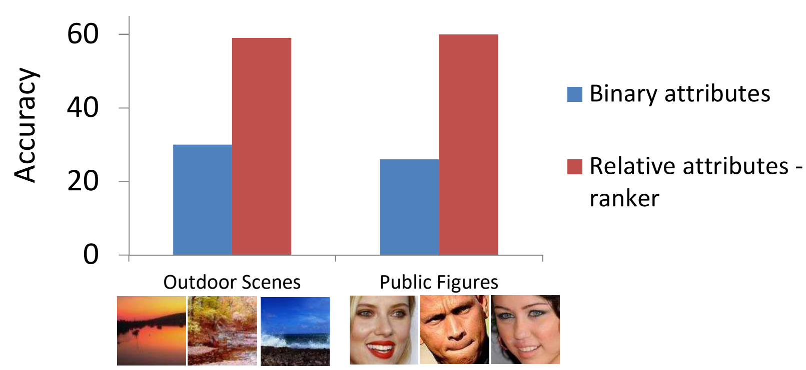
Leg length: Horse $\succ$ Mule

Tail length: Mule $\succ$ Donkey

⋮



Relative zero-shot learning



Comparative descriptions are more discriminative than categorical definitions.

Attributes for search and recognition

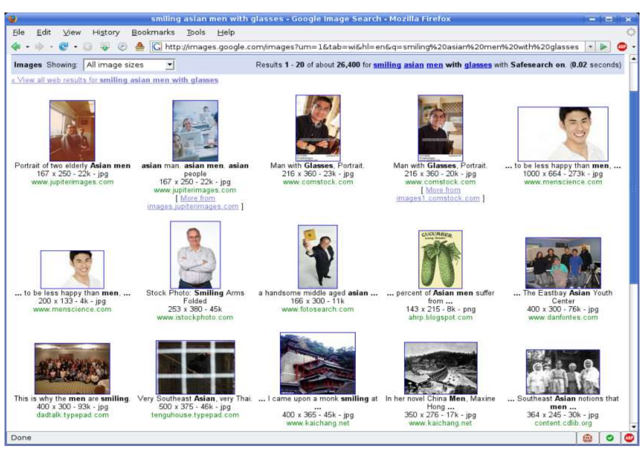
Attributes give human user way to

- 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?

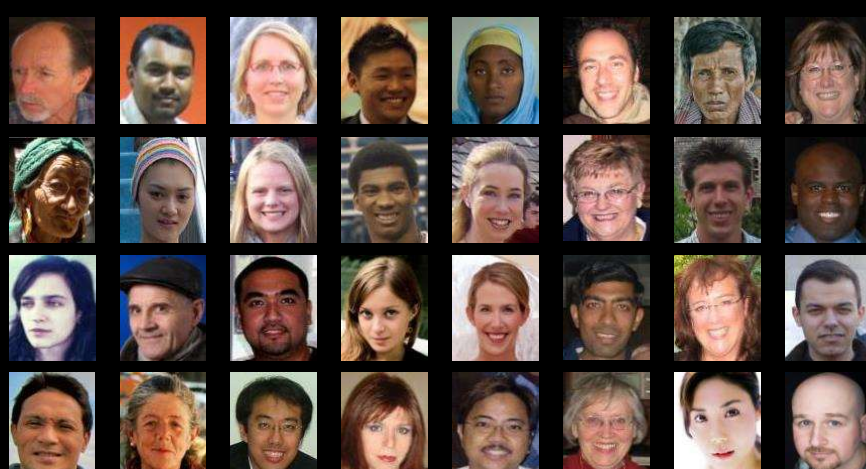
- Human understandable
- Support familiar keyword-based queries

- 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

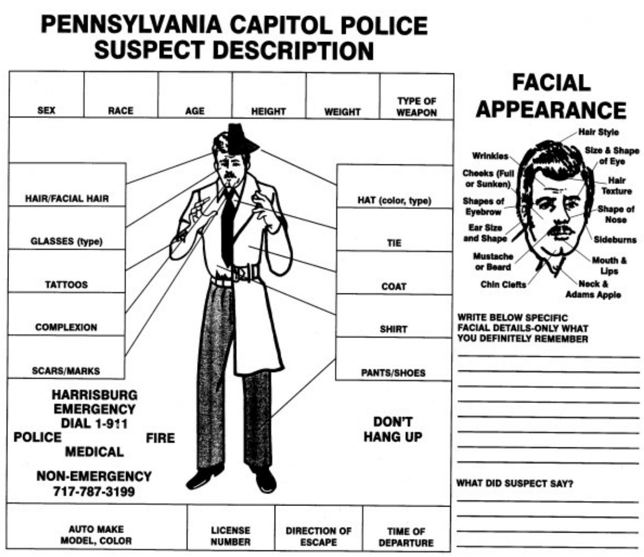
Attributes efficiently divide the space of images

Female Caucasian Eyeglasses Older

k attributes can distinguish 2^k categories

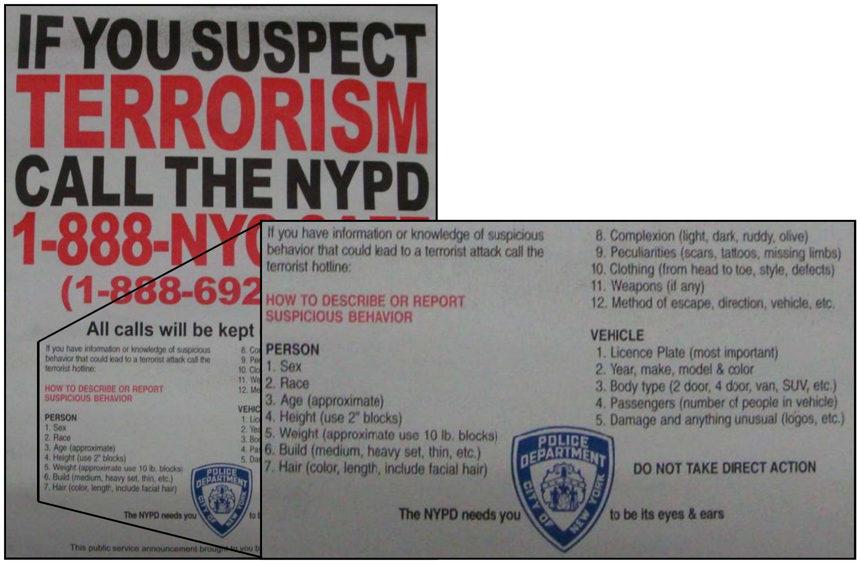
Slide credit: Neeraj Kumar

Search applications: finding people



Slide credit: Rogerio Feris

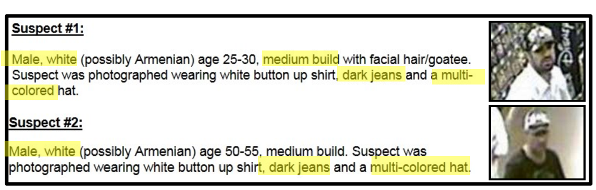
Search applications: finding people



Slide credit: Rogerio Feris

Search applications: finding people

Search surveillance feeds for suspects

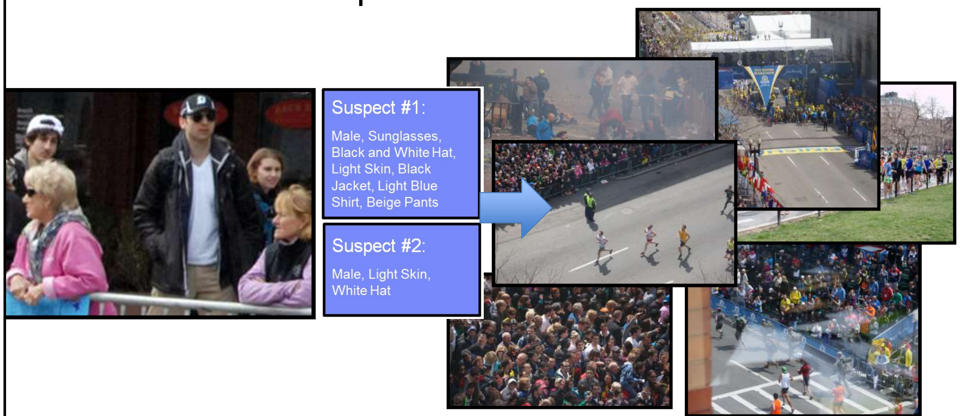


http://lacrimestoppers.com/wanteds.aspx

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

Search applications: products

Query: "I want a bright, open shoe that is short on the leg."

Slide credit: Kristen Grauman

Search applications: graphic design

Query: "I want an outdoor scene that looks uncrowded and calm

Slide credit: Kristen Grauman

Face Search with Attributes

FaceTracer: A Search Engine for Large Collections of Images with Faces, Neeraj Kumar, Peter N. Belhumeur, Shree K. Nayar, ECCV 2008.

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

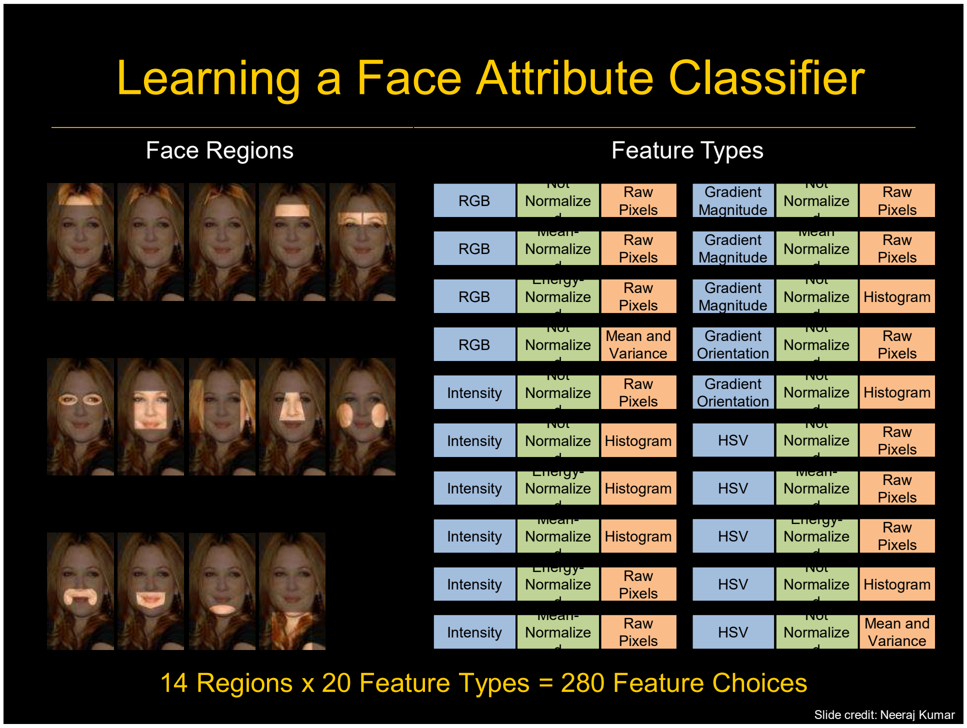
Facial Attributes

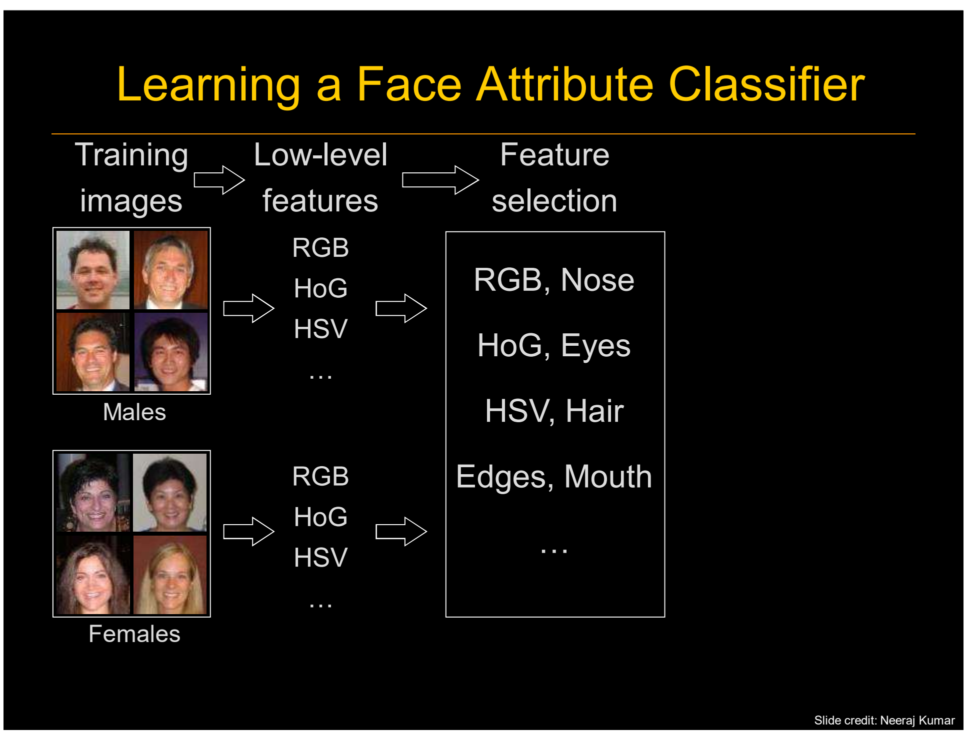
- Various properties of interest to search
- Many are spatially localized within face

Gender
Male
Female
Age
Baby
Child
Youth
Middle Aged
Senior
Hair Color
Black
Blond

Smiling
True
False
Mustache
True
False
Blurry
True
False
Lighting
Flash
Harsh

Race
White
Black
Asian
Eye Wear
None
Eyeglasses
Sunglasses
Environment
Outdoor
Indoor


Learning a Face Attribute Classifier
Face Regions
Feature Types
RGB | Normalize | Raw Pixels | Gradient Magnitude | Normalize | Raw Pixels
RGB | Normalize | Raw Pixels | Gradient Magnitude | Normalize | Raw Pixels
RGB | Normalize | Raw Pixels | Gradient Magnitude | Normalize | Histogram
RGB | Normalize | Mean and Variance | Gradient Orientation | Normalize | Raw Pixels
Intensity | Normalize | Raw Pixels | Gradient Orientation | Normalize | Histogram
Intensity | Normalize | Histogram | HSV | Normalize | Raw Pixels
Intensity | Normalize | Histogram | HSV | Normalize | Raw Pixels
Intensity | Normalize | Histogram | HSV | Normalize | Raw Pixels
Intensity | Normalize | Raw Pixels | HSV | Normalize | Histogram
Intensity | Normalize | Raw Pixels | HSV | Normalize | Mean and Variance
14 Regions x 20 Feature Types = 280 Feature Choices
Slide credit: Neeraj Kumar


Learning a Face Attribute Classifier
Training images
Low-level features
Feature selection
RGB
HoG
HSV
…
Males
RGB, Nose
HoG, Eyes
HSV, Hair
Edges, Mouth
…
RGB
HoG
HSV
…
Females
Slide credit: Neeraj Kumar

Learning a Face Attribute Classifier

Training images ⇨ Low-level features ⇨ Feature selection ⇨ Train classifier

Males

RGB
HoG
HSV
…

Females

RGB
HoG
HSV
…

RGB, Nose
HoG, Eyes
HSV, Hair
Edges, Mouth
…

⇨ Gender classifier

Slide credit: Neeraj Kumar

Attribute Classifier Accuracies

Binary facial attributes in Columbia Face Database

Typically 80%-90% accuracy

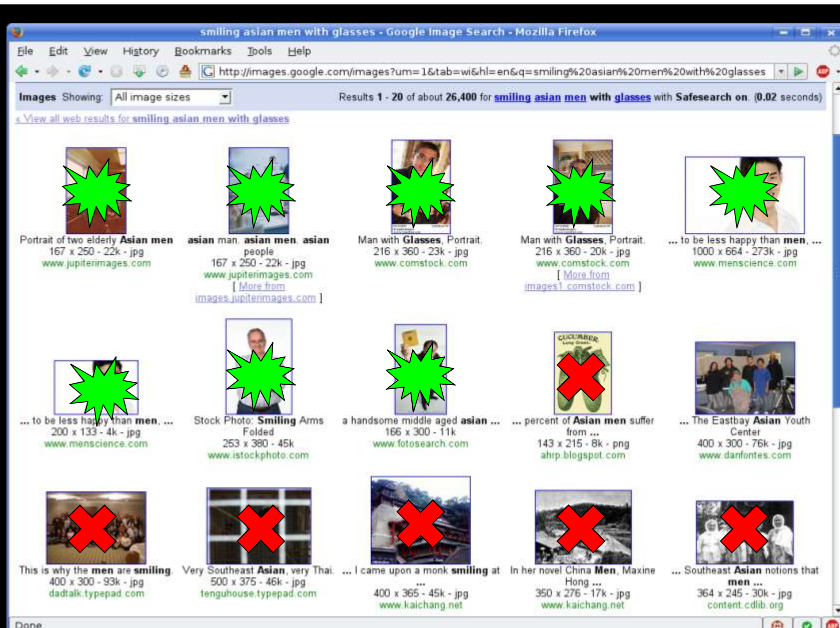
gender	85.78	hair color: black	90.82	flushed face	88.85
age: young	87.72	hair color: blond	88.39	chubby	81.16
age: middle aged	84.93	hair color: brown	74.88	forehead: fully visible	89.31
age: senior	92.04	hair color: gray	89.86	forehead: partially visible	76.96
race: Asian	92.32	hair color: bald	90.39	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

FaceTracer: A Search Engine for Large Collections of Images with Faces, Neeraj Kumar, Peter N. Belhumeur, Shree K. Nayar, ECCV 2008.

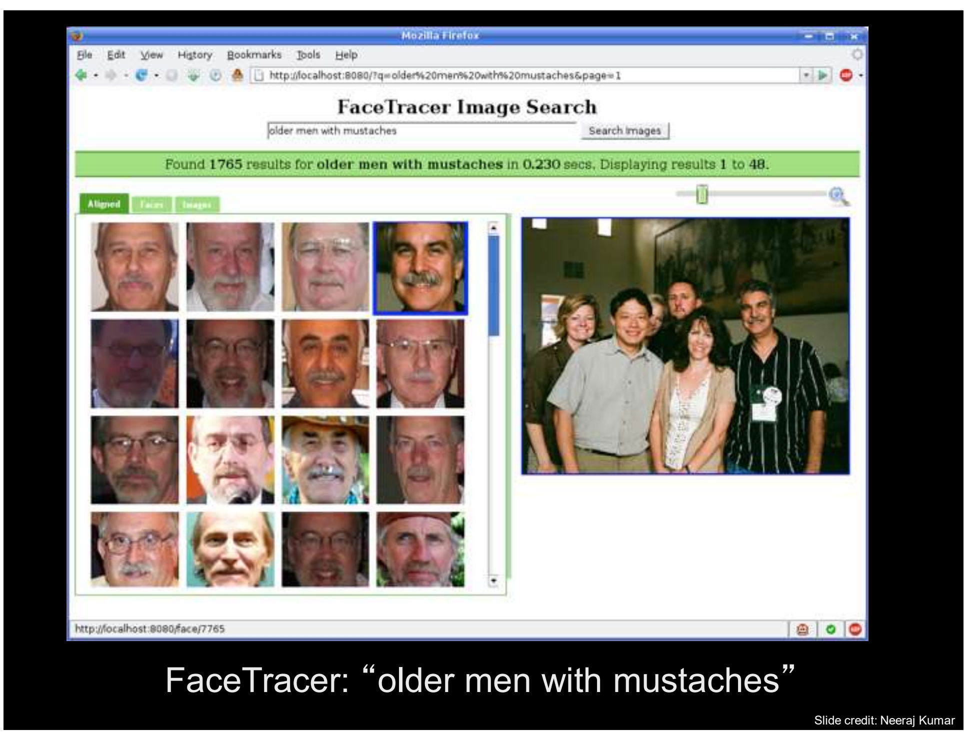
- Offline:
  - Apply attribute classifiers to database images
  - Map classifier outputs to probabilities
- Online:
  - Convey available attribute names to user
  - Given query attributes, rank database images by confidence (e.g., product of probabilities)



Google: “smiling asian men with glasses” July 2008

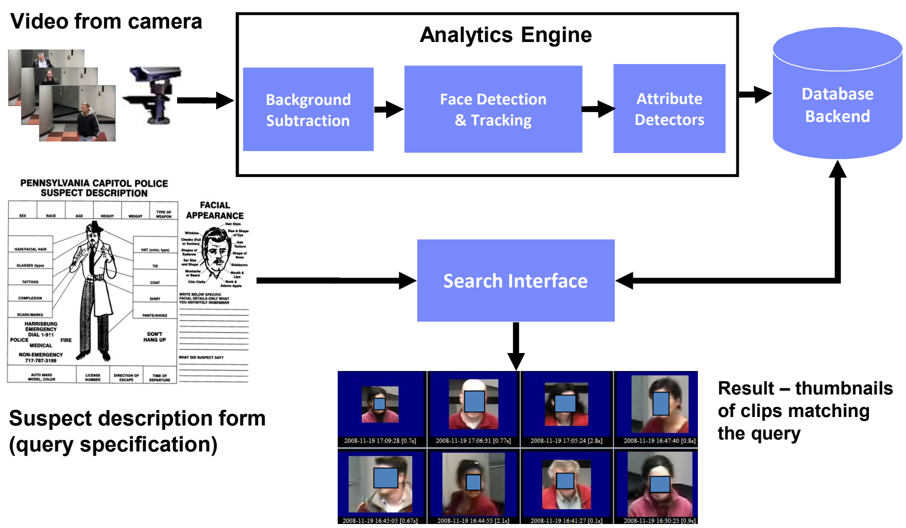
Slide credit: Neeraj Kumar


Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:8080/?q=smiling%20asian%20men%20with%20glasses&page=1
FaceTracer Image Search
smiling asian men with glasses
Search Images
Found 1344 results for smiling asian men with glasses in 0.220 secs. Displaying results 1 to 48.
Aligned
Faces
Images
javascript:void(0)
FaceTracer: “smiling asian men with glasses”
Slide credit: Neeraj Kumar


Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:8080/?q=older%20men%20with%20mustaches&page=1
FaceTracer Image Search
older men with mustaches
Search Images
Found 1765 results for older men with mustaches in 0.230 secs. Displaying results 1 to 48.
Aligned
Faces
Images
http://localhost:8080/face/7765
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.* WACV 2009.



Slide credit: Rogerio Feris

Video Demo: Attribute-based People Search

Suspect #1

Vaquero, Feris, Tran, Brown, Hampapur and Turk. *Attribute-Based People Search in Surveillance Environments.* WACV 2009.

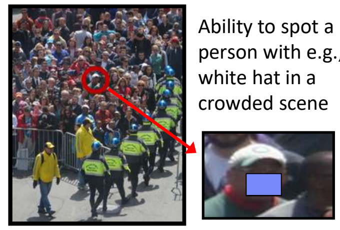
Example query: Boston bombing scenario

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

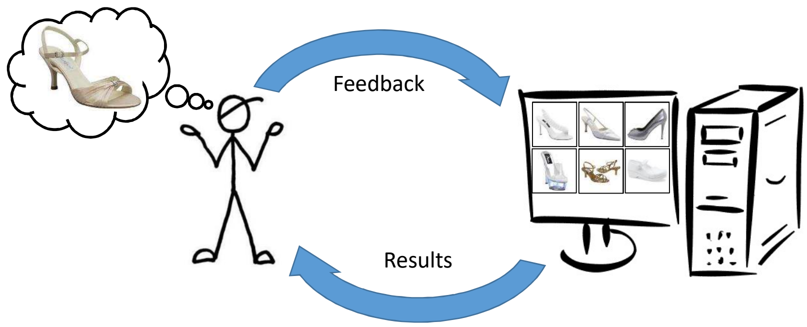
Slide credit: Rogerio Feris

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

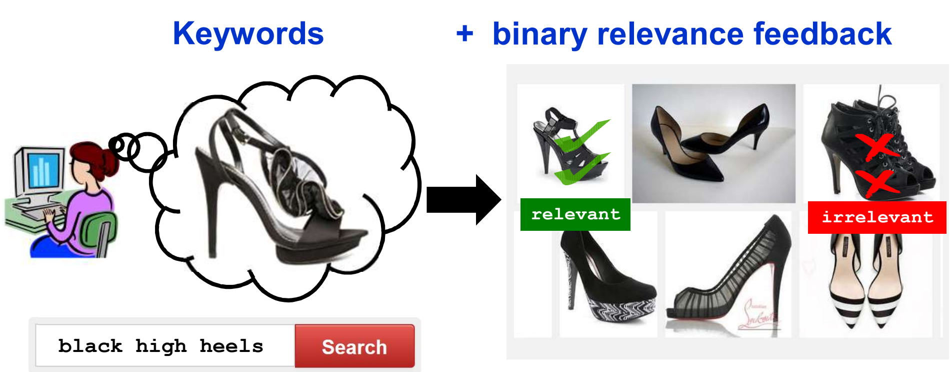
Interactive visual search



- 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?



- 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

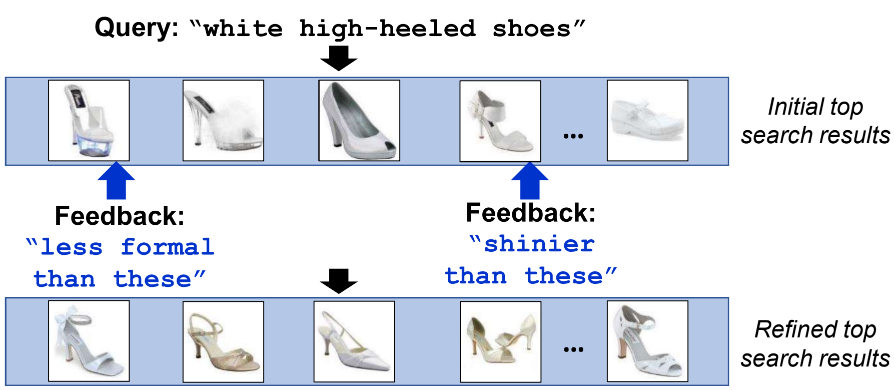
[Kovashka et al., CVPR 2012]

"Like this… but *more ornate*"

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

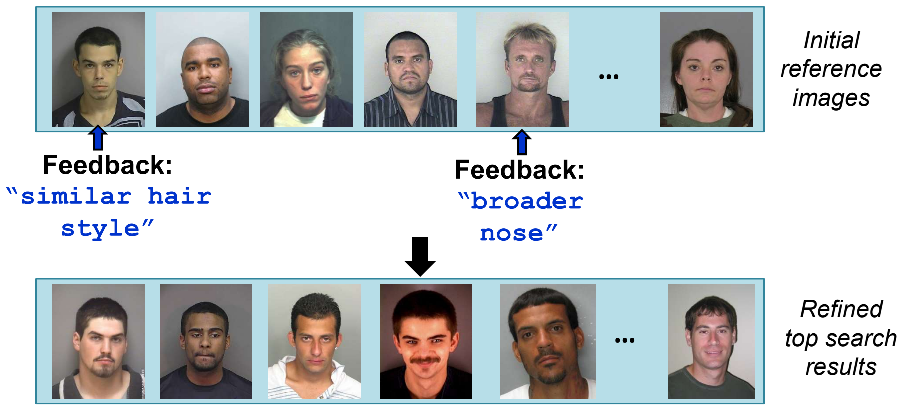
WhittleSearch: Relative attribute feedback

[Kovashka et al., CVPR 2012]



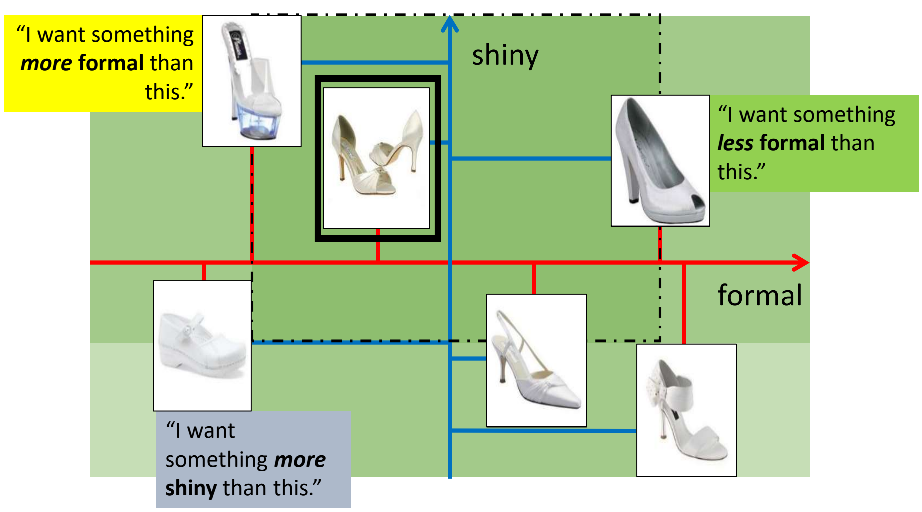
WhittleSearch: Relative attribute feedback

[Kovashka et al., CVPR 2012]



WhittleSearch: Relative attribute feedback

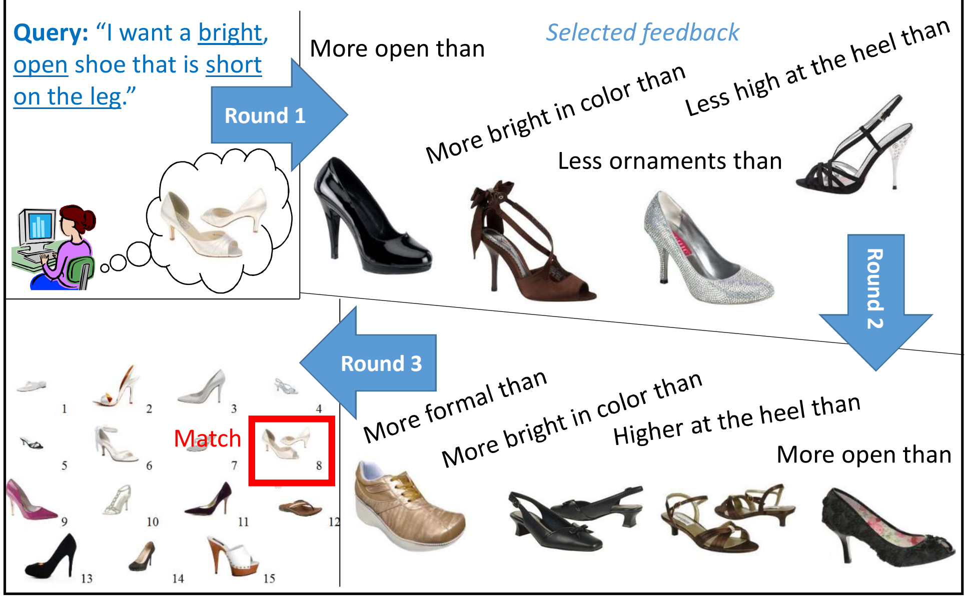
[Kovashka et al., CVPR 2012]



[Kovashka, Parikh, and Grauman, CVPR 2012]

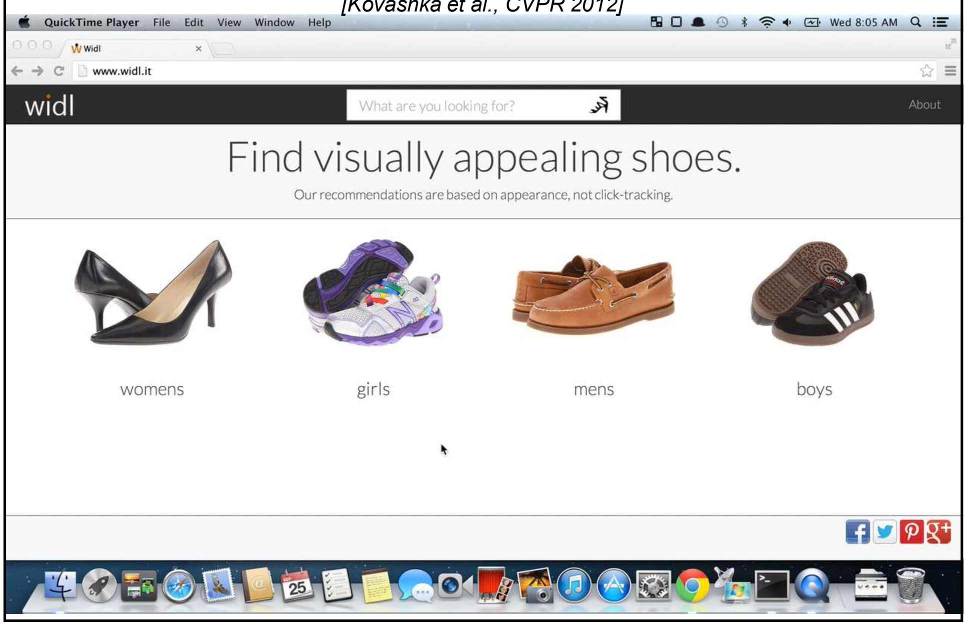
Example WhittleSearch

[Kovashka et al., CVPR 2012]



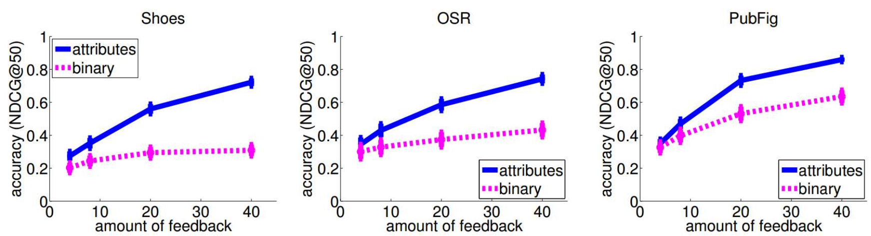
WhittleSearch: Relative attribute feedback

[Kovashka et al., CVPR 2012]



WhittleSearch results

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



We more rapidly converge on the envisioned visual content.

[Kovashka et al., CVPR 2012]

Attributes for search and recognition

Attributes give human user way to

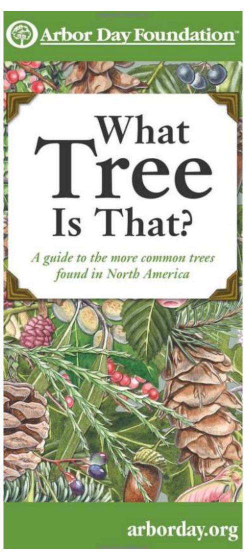
- Teach novel categories with description
- Communicate search queries
- Give feedback in interactive search
- $\boxed{\text{Assist in interactive recognition}}$

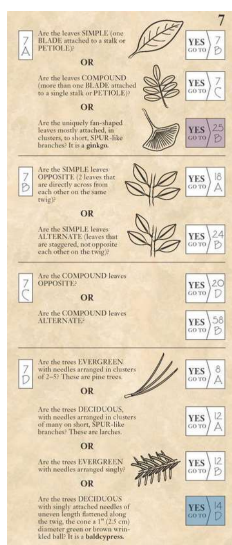
Slide credit: Kristen Grauman

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

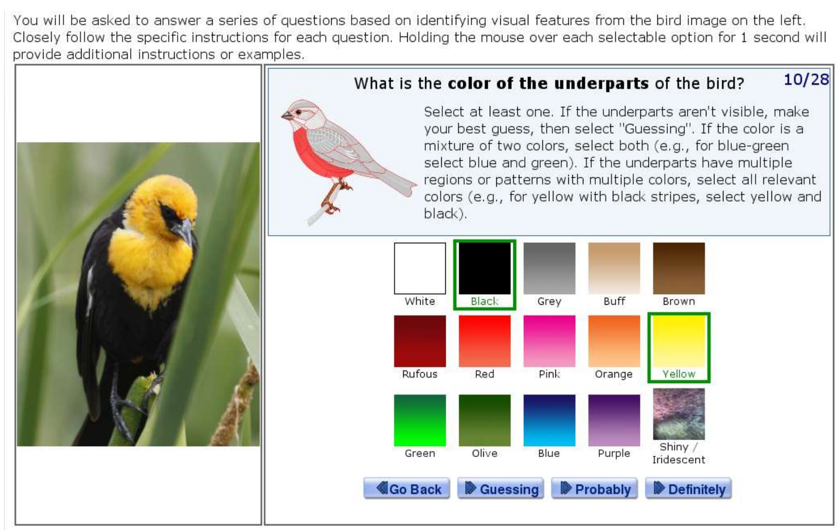
Slide credit: Steve Branson

Example Questions: Localize

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

Slide credit: Steve Branson

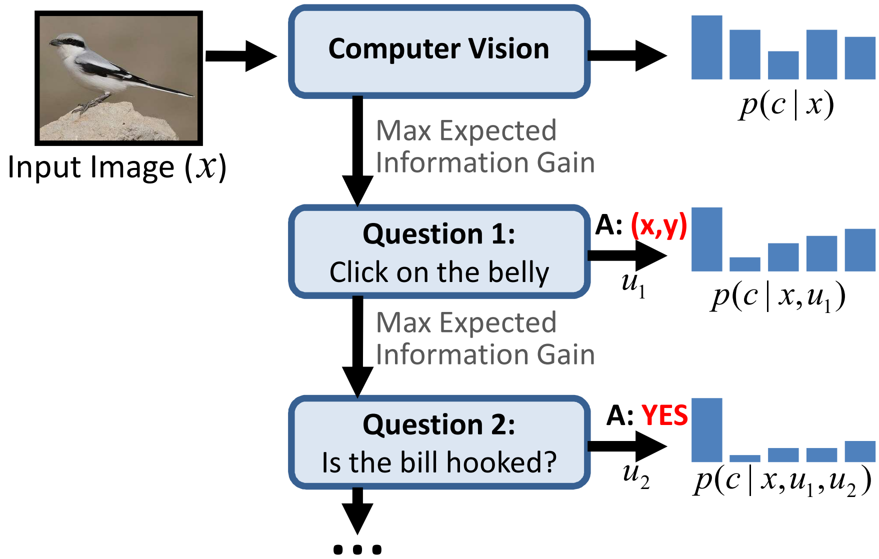
Example Questions: Name attributes



Wah et al., ICCV 2011

Slide credit: Steve Branson

Basic Algorithm



Wah et al., ICCV 2011

Slide credit: Steve Branson

Basic Algorithm

Select the next question that maximizes expected information gain:

- Involves estimating probabilities of the form:

$$p(\underset{\text{Object Class}}{c} \mid \underset{\text{Image}}{x}, \underbrace{u_1, u_2 \ldots u_t}_{\text{Sequence of user responses}})$$

Wah et al., ICCV 2011

Slide credit: Steve Branson

Basic Algorithm

Integrate over all possible locations of the parts:

$$p(c \mid x, u_1, u_2 \ldots u_t)$$

$$\propto \int_{\Theta} \underbrace{p(u_1, u_2 \ldots u_t \mid c, \Theta, x)}_{\text{Model of user responses}} \underbrace{p(c \mid \Theta, x)}_{\text{Localized attribute estimator}} \underbrace{p(\Theta \mid x)}_{\text{Part Localization}}$$

Wah et al., ICCV 2011

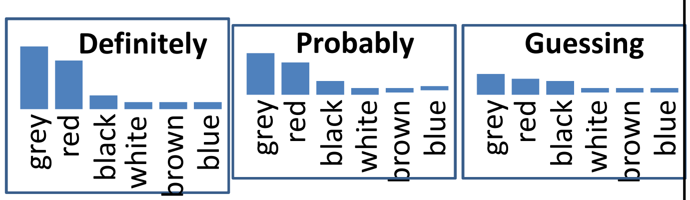
Slide credit: Steve Branson

Modeling User Responses: Attribute Questions

- Assume: $p(u_1, u_2 \ldots u_t \mid c, \Theta) \approx \prod_{i=1\ldots t} p(u_i \mid c, \Theta)$
- Estimate $p(u_i \mid c, \Theta)$ using Mechanical Turk

Pine Grosbeak

What is the color of the belly?



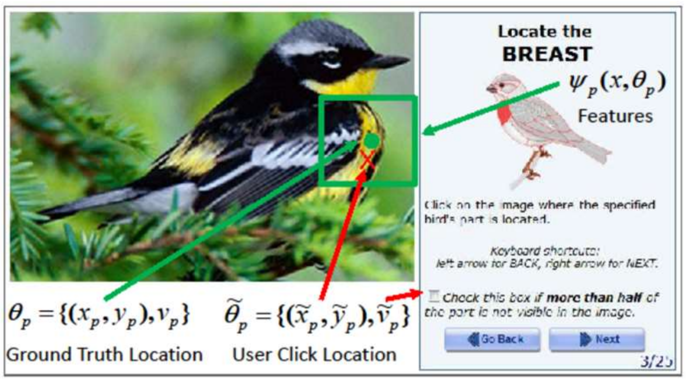
Wah et al., ICCV 2011

Slide credit: Steve Branson

Modeling User Responses: Click Questions

- Assume: $p(u_1, u_2 \ldots u_t \mid c, \Theta) \approx \prod_{i=1\ldots t} p(u_i \mid c, \Theta)$
- Estimate $p(u_i \mid c, \Theta)$ using Mechanical Turk

Click on the breast



Wah et al., ICCV 2011

Slide credit: Steve Branson

CUB-200-2011 Dataset


Black-footed Albatross

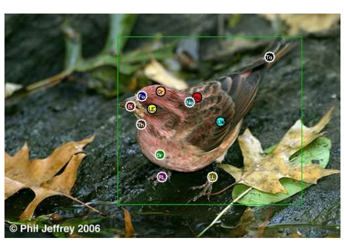

Groove-Billed Ani


Parakeet Auklet

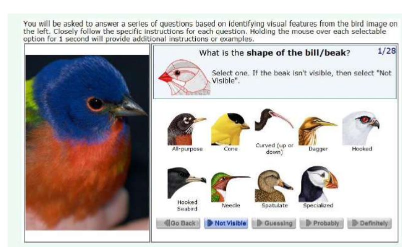

Field Sparrow


Vesper Sparrow

11,877 images, 200 bird species



13 part locations

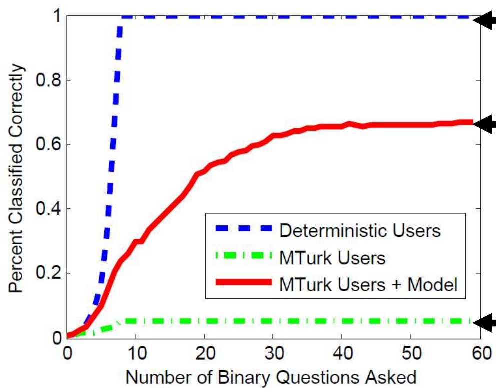


288 binary attributes

Wah et al., ICCV 2011

Slide credit: Steve Branson

Results: Without Computer Vision



Perfect Users, Field Guide Attributes

100% accuracy in 8≈log_2(200) questions if users agree with field guides…

Real Users, Probabilistic User Model

Tolerate ambiguous responses, user error

Real Users, Field Guide Attributes

MTurkers don't always agree with field guides…

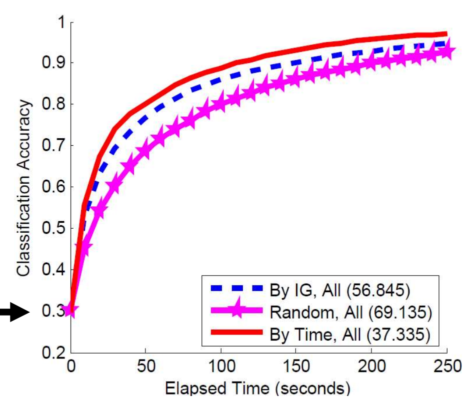
Branson et al., ECCV 2010

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

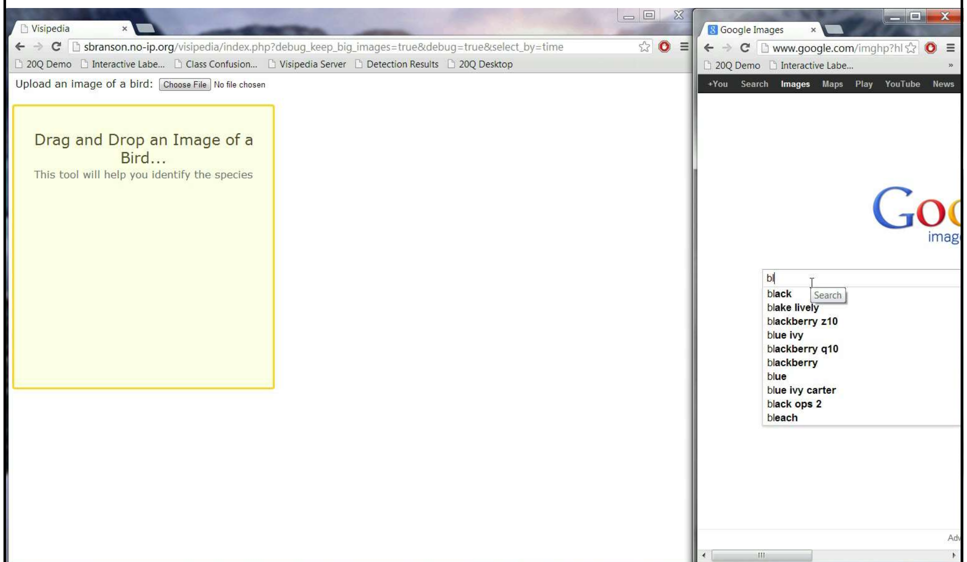
Base Computer Vision performance (30%)



Wah et al., ICCV 2011

Slide credit: Steve Branson

Demo



Wah, Branson, Perona, Belongie, Multi-class Recognition and Part Localization with Humans in the Loop, ICCV 2011

Slide credit: Steve Branson

Summary: 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

Ongoing challenges (1)

- Accuracy of attribute models crucial to success
- Human perception of attributes can vary
- When is the attribute vocabulary expressive enough?
- If large attribute vocabulary is available, how to convey it to the search user?
- Practical issues in calibration and fusion
- Localized vs. global properties

Slide credit: Kristen Grauman

Ongoing challenges (2)

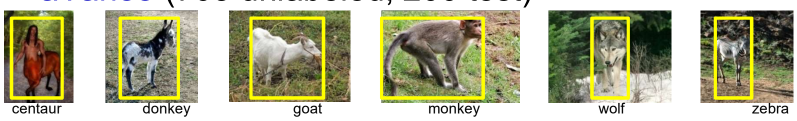
- What attributes should be in the vocabulary?
- How to align user's attribute language with the visual attribute models?
- Integrated treatment of binary and relative attributes?
- Joint learning of multiple attributes?
- Class-specific attributes?
- How do we make sure we're learning the "right" thing?

Kristen Grauman, UT-Austin

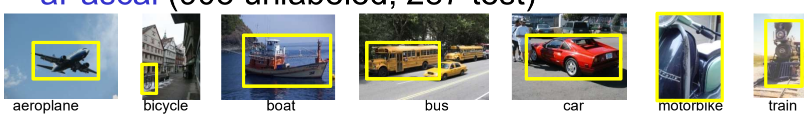
- Animals with Attributes – 1 (1003 unlabeled, 732 test)

- Animals with Attributes – 2 (1002 unlabeled, 993 test)

- aYahoo (703 unlabeled, 200 test)



- aPascal (903 unlabeled, 287 test)



Slide credit: Adriana Kovashka