



Mining, and Intro to Categorization

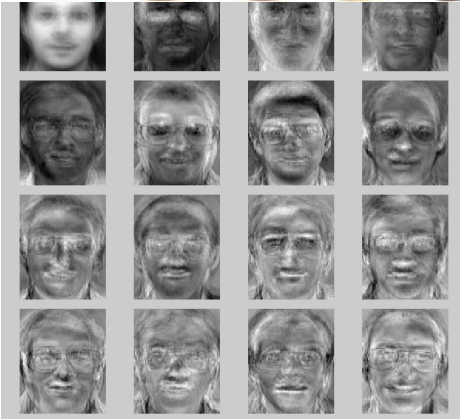
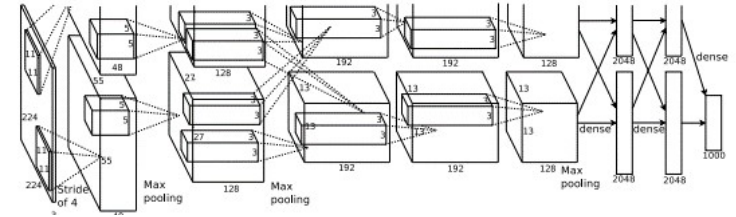
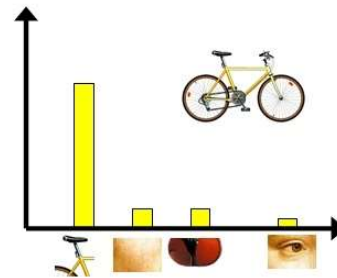
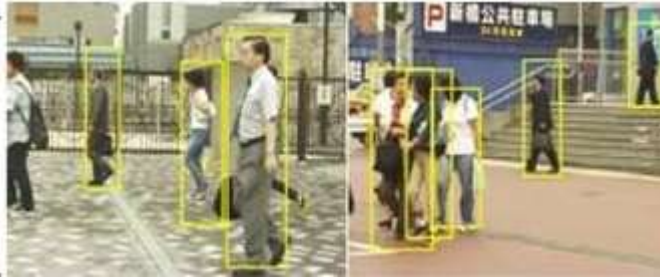
Tues April 10

Kristen Grauman
UT Austin

UT Austin, CS 376 Computer Vision - lecture 21

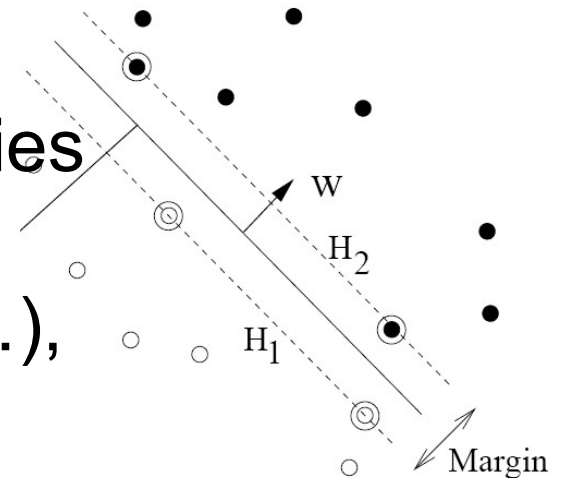


Recognition and learning



Recognizing categories
(objects, scenes,
activities, attributes...),
learning techniques

UT Austin, CS-576 Computer
Vision - lecture 21



Last time

- Instance recognition wrap up:
 - Spatial verification
 - Sky mapping example
 - Query expansion

Review questions

- Does an inverted file index sacrifice accuracy in bag-of-words image retrieval? Why or why not?
- Why does a single SIFT match cast a 4D vote for the Generalized Hough spatial verification model?
- What does a perfect precision recall curve look like?

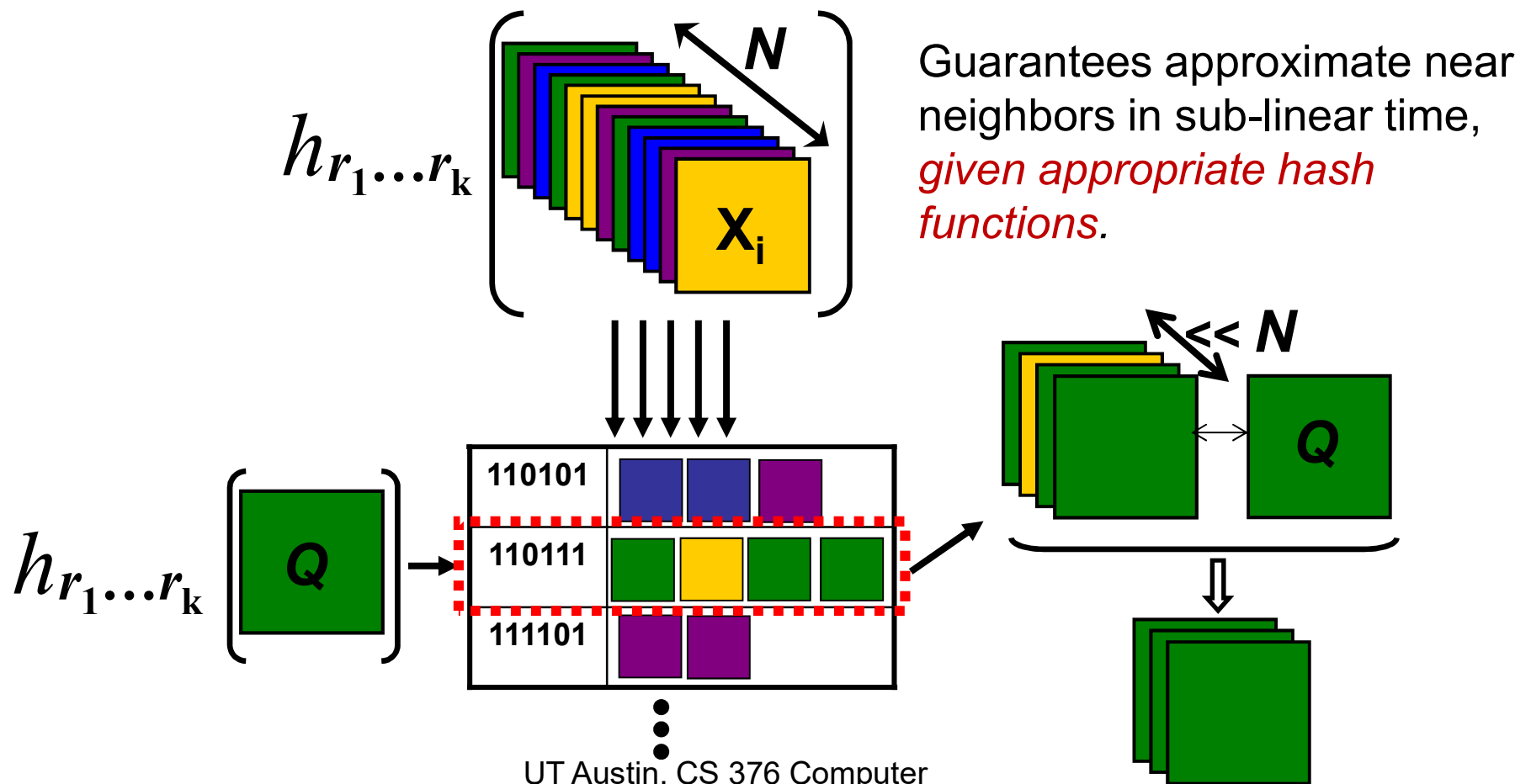
Today

- Discovering visual patterns
 - Randomized hashing algorithms
 - Mining large-scale image collections
- Introduction to visual categorization

Locality Sensitive Hashing (LSH)

[Indyk and Motwani '98, Gionis et al. '99, Charikar '02, Andoni et al. '04]

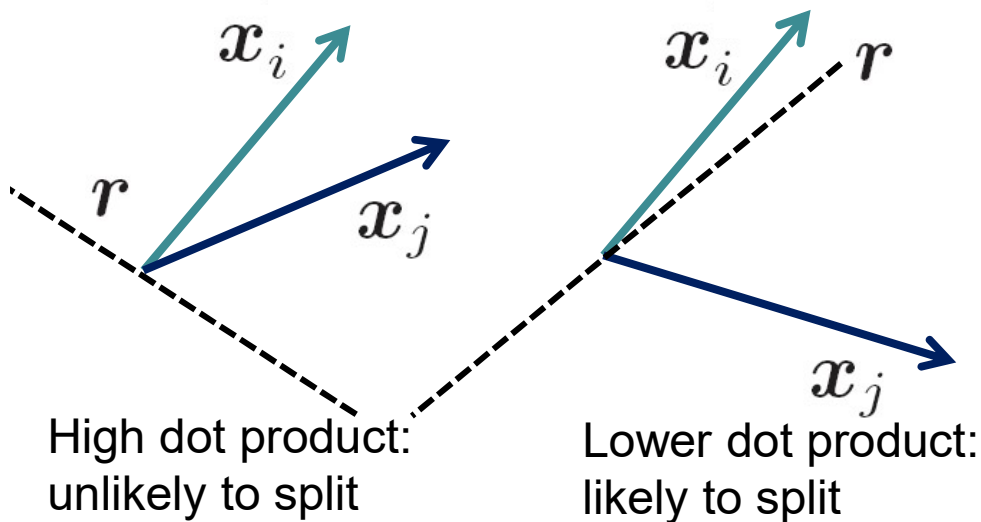
$$\Pr_{h \in \mathcal{F}} [h(x) = h(y)] = \text{sim}(x, y)$$



LSH function example: inner product similarity

The probability that a *random hyperplane* separates two unit vectors depends on the angle between them:

$$\Pr[\text{sign}(\mathbf{x}_i^T \mathbf{r}) = \text{sign}(\mathbf{x}_j^T \mathbf{r})] = 1 - \frac{1}{\pi} \cos^{-1}(\mathbf{x}_i^T \mathbf{x}_j)$$



Corresponding hash function:

$$h_{\mathbf{r}}(\mathbf{x}) = \begin{cases} 1, & \text{if } \mathbf{r}^T \mathbf{x} \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\text{for } \vec{r}_i \sim N(\mu = 0, \sigma^2 = 1)$$

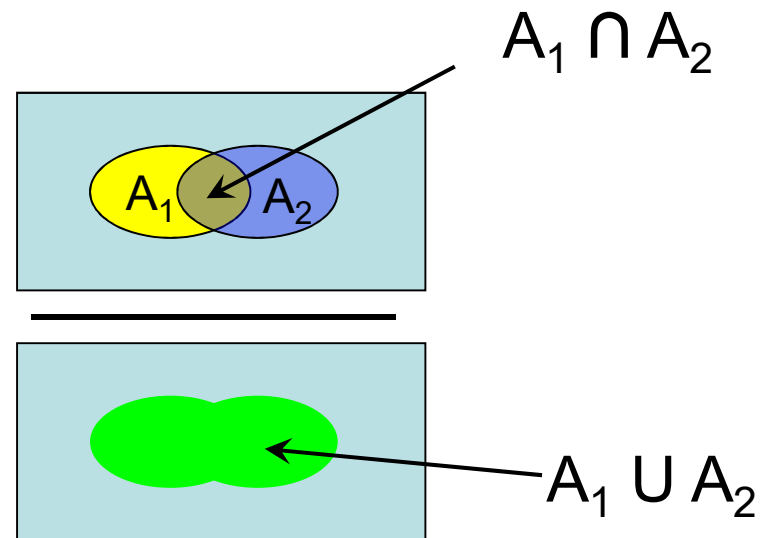
[Goemans and Williamson 1995, Charikar 2004]

LSH function example: Min-hash for set overlap similarity

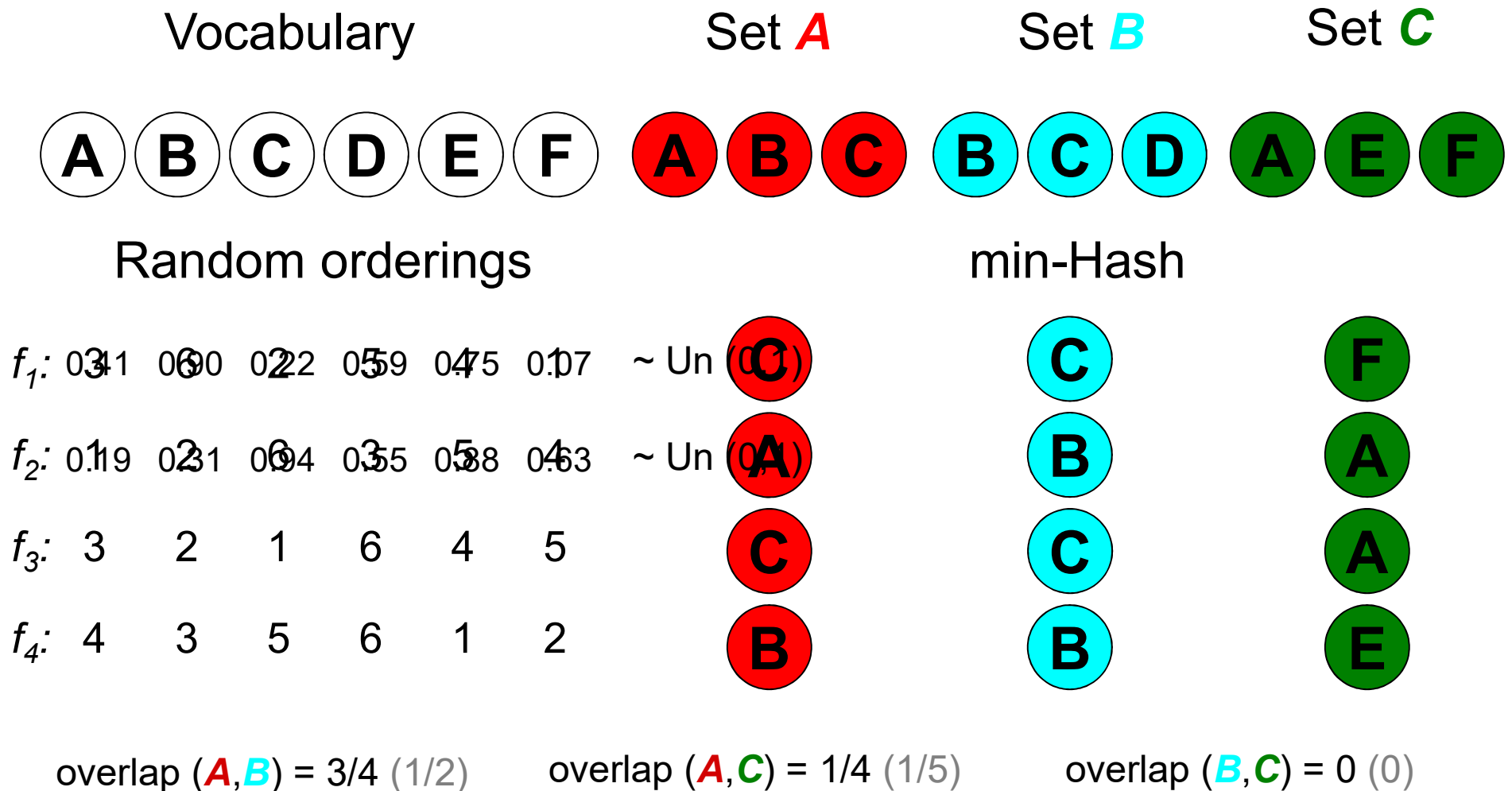
[Broder, 1999]

$$\Pr_{h \in \mathcal{F}} [h(x) = h(y)] = \text{sim}(x, y)$$

$$\text{sim}(\mathcal{A}_1, \mathcal{A}_2) = \frac{|\mathcal{A}_1 \cap \mathcal{A}_2|}{|\mathcal{A}_1 \cup \mathcal{A}_2|} \in \langle 0, 1 \rangle.$$



LSH function example: Min-hash for set overlap similarity



UT Austin, CS 376 Computer

[Broder, 1999]

LSH function example: Min-hash for set overlap similarity

A: A E J Q R V Y **B:** A C E Q V Z

Ordering by f_2

A \cup **B:** A C E J Q R V Y Z

↑










h1(**A**) A h1(**B**) A
h2(**A**) Q h2(**B**) C

$$P(h(\mathbf{A}) = h(\mathbf{B})) = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A} \cup \mathbf{B}|}$$

Multiple hash functions and tables

- Generate k such hash functions, concatenate outputs into hash key:

$$P(h_{1,\dots,k}(x) = h_{1,\dots,k}(y)) =$$

110101				
110111				
111101				

- To increase recall, search multiple independently generated hash tables
 - Search/rank the union of collisions in each table, or
 - Require that two examples in at least T of the tables to consider them similar.

TABLE 1



















110101				
110111				
111101				

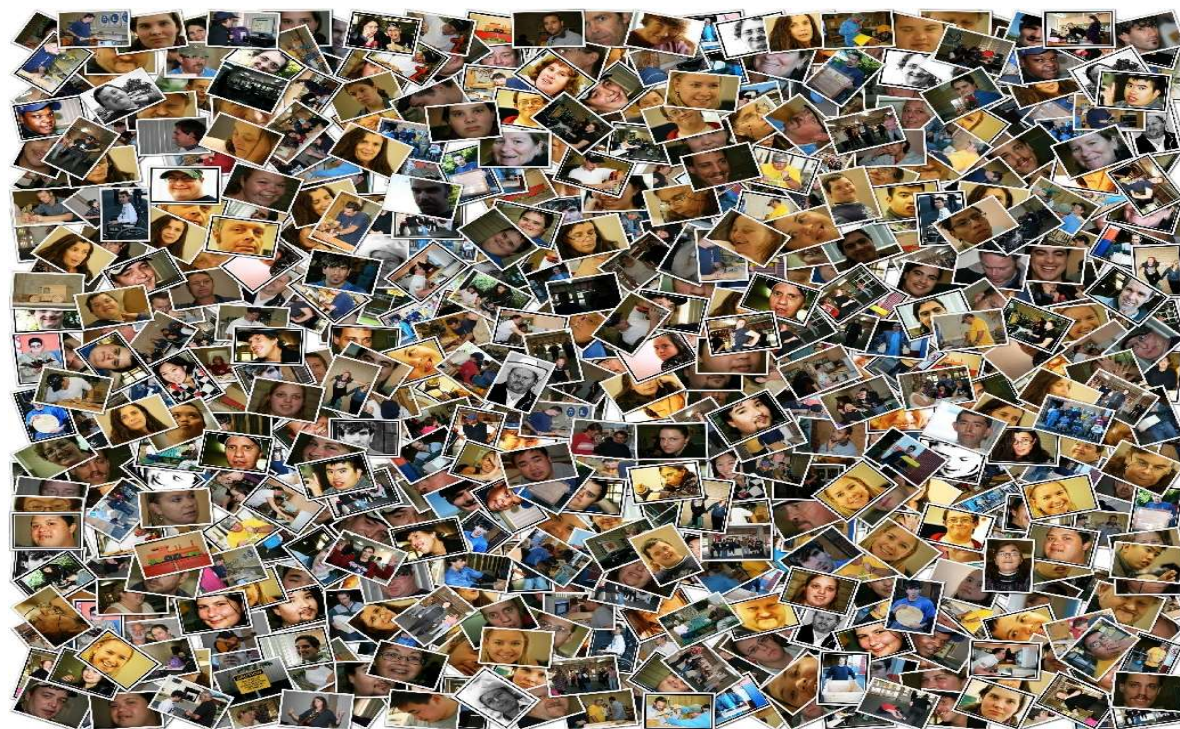
TABLE 2

110100				
111111				
111001				

Mining for common visual patterns

In addition to visual search, want to be able to **summarize, mine, and rank** the large collection as a whole.

- What is common?
- What is unusual?
- What co-occurs?
- Which exemplars are most representative?



UT Austin, CS 376 Computer
Vision - lecture 21

Mining for common visual patterns

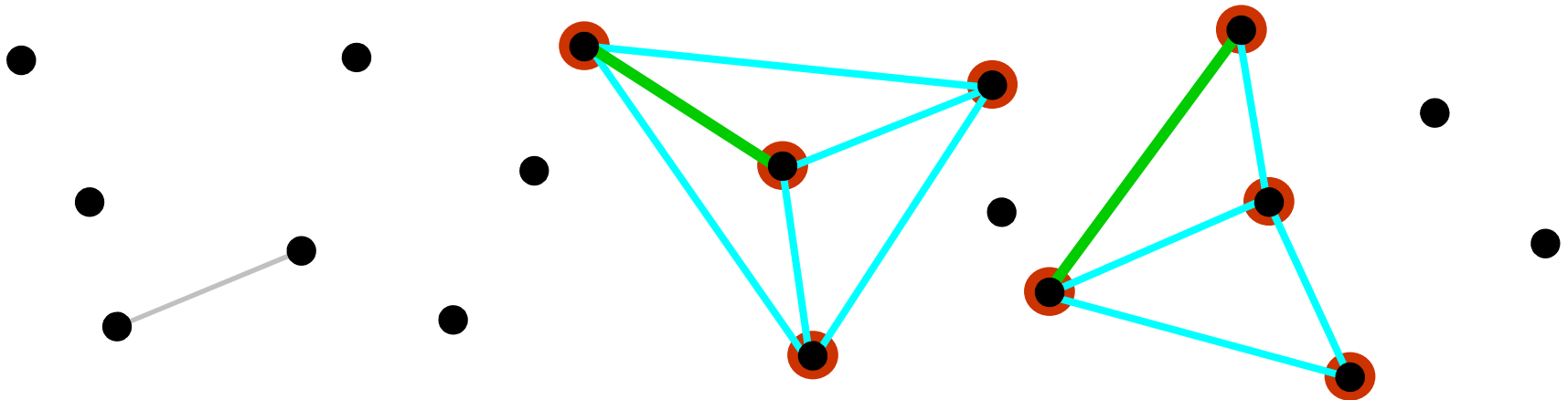
In addition to visual search, want to be able to **summarize, mine, and rank** the large collection as a whole.

We'll look at a few examples:

- **Connected component clustering** via hashing
 - [Geometric Min-hash, Chum et al. 2009]
- **Visual Rank** to choose “image authorities”
 - [Jing and Baluja, 2008]
- **Frequent item-set mining** with spatial patterns
 - [Quack et al., 2007]

Connected component clustering with hashing

1. Detect seed pairs via hash collisions
2. Hash to related images
3. Compute connected components of the graph



Contrast with frequently used quadratic-time clustering algorithms

Geometric Min-hash

[Chum, Perdoch, Matas, CVPR 2009]



- Main idea: build spatial relationships into the hash key construction:
 - Select first hash output according to min hash (“central word”)
 - Then append subsequent hash outputs from within its neighborhood

Results:

Geometric Min-hash clustering

[Chum, Perdoch, Matas, CVPR 2009]

All Soul's



Ashmolean



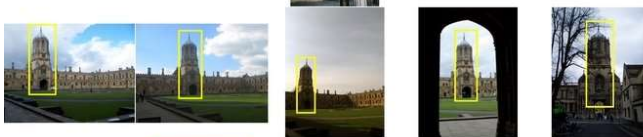
Balliol



Bodleian



Christ Church



Cornmarket



Hertford



Keble



Magdalen



Pitt Rivers



Radcliffe Camera



100 000 Images downloaded from FLICKR

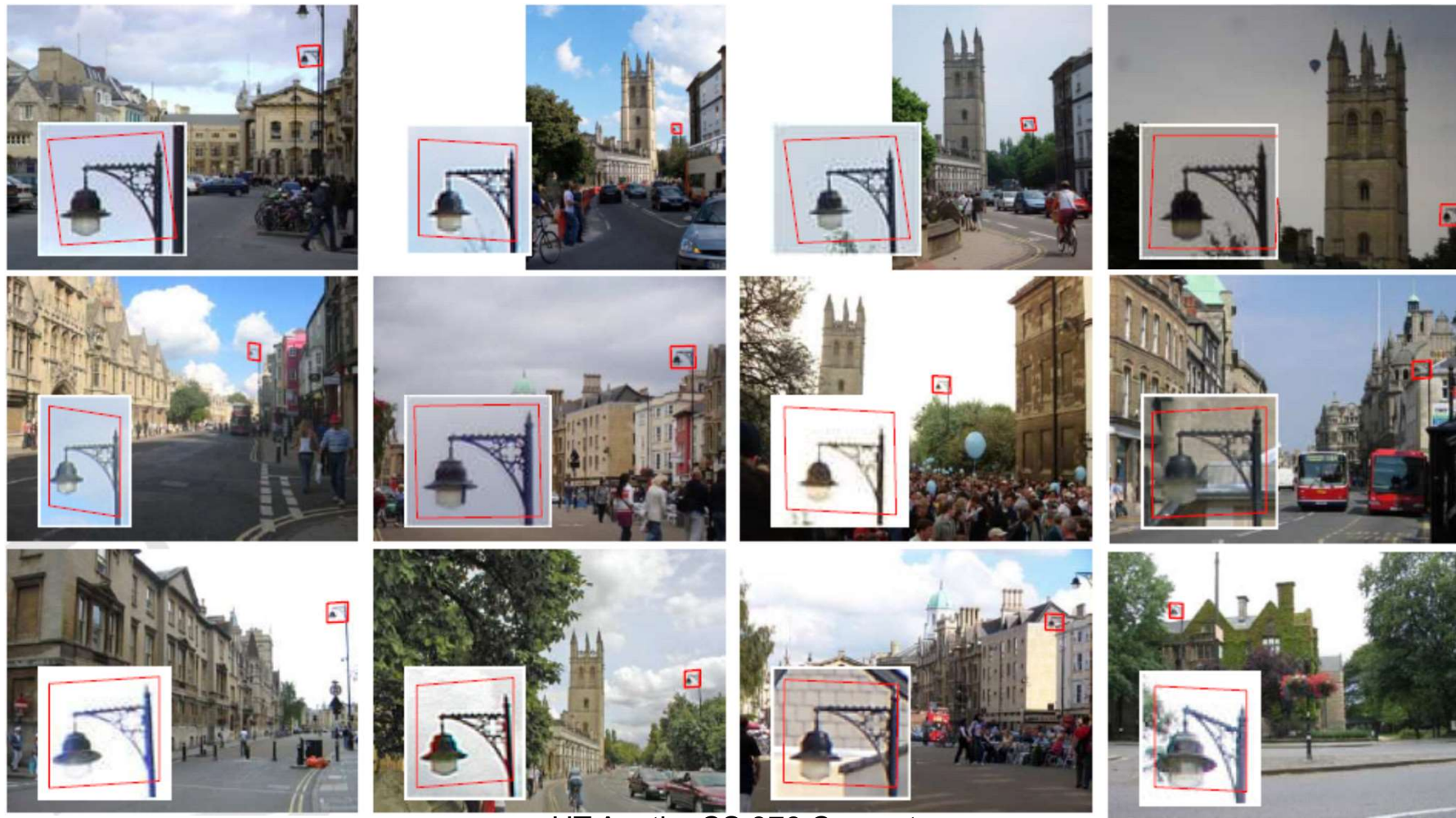
Includes 11 Oxford Landmarks with manually labeled ground truth

UT Austin, CS 376 Computer
Vision - lecture 21

Results:

Geometric Min-hash clustering

[Chum, Perdoch, Matas, CVPR 2009]



UT Austin, CS 376 Computer
Discovering small objects

Slide credit: Ondrej Chum

Results:

Geometric Min-hash clustering

[Chum, Perdoch, Matas, CVPR 2009]



UT Austin, CS 376 Computer
Discovering small objects

Slide credit: Ondrej Chum

Mining for common visual patterns

In addition to visual search, want to be able to **summarize, mine, and rank** the large collection as a whole.

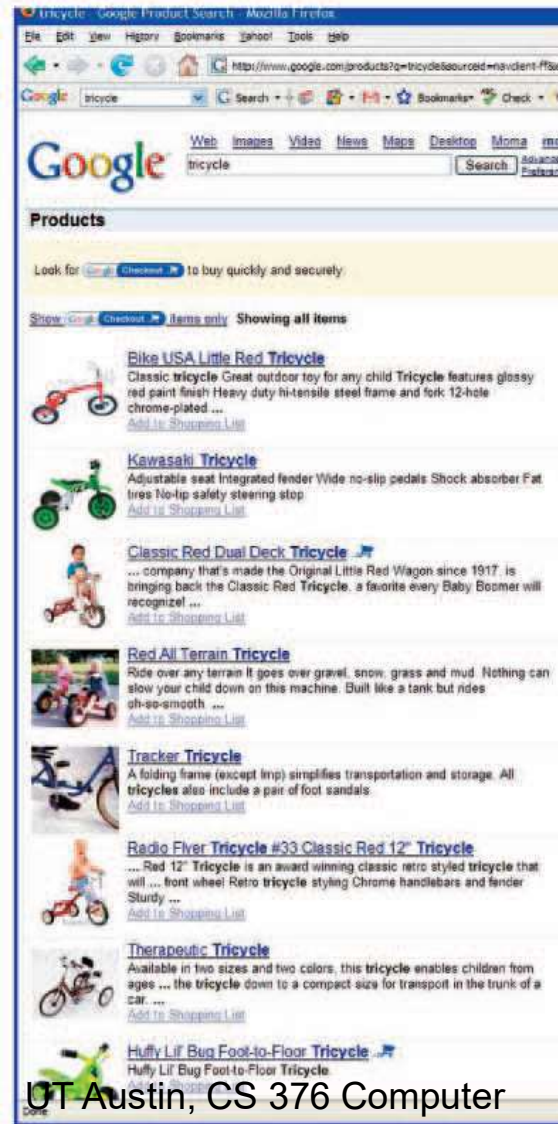
We'll look briefly at a few recent examples:

- **Connected component clustering** via hashing [Geometric Min-hash, Chum et al. 2009]
- **Visual Rank** to choose “image authorities” [Jing and Baluja, 2008]
- **Frequent item-set mining** with spatial patterns [Quack et al., 2007]

Visual Rank: motivation



Product search



Mixed-type search

- **Goal:** select small set of “best” images to display among millions of candidates

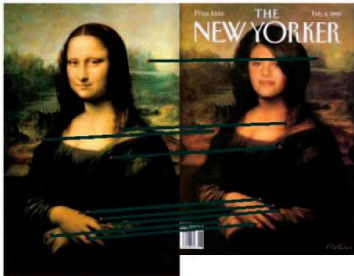
Visual Rank

[Jing and Baluja, PAMI 2008]

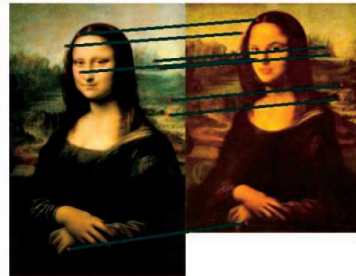
- Compute relative “authority” of an image based on random walk principle.
 - Application of PageRank to visual data
- **Main ideas:**
 - Graph weights = number of matched local features between two images
 - Exploit text search to narrow scope of each graph
 - Use LSH to make similarity computations efficient

Results: Visual Rank

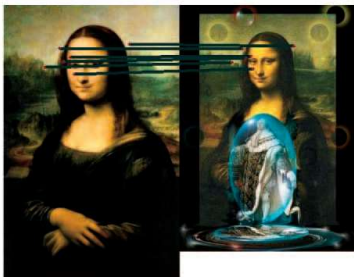
[Jing and Baluja, PAMI 2008]



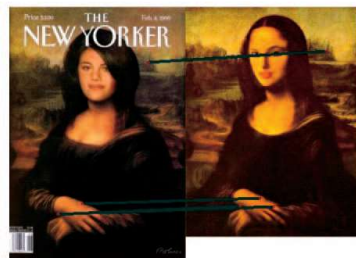
(a)



(b)



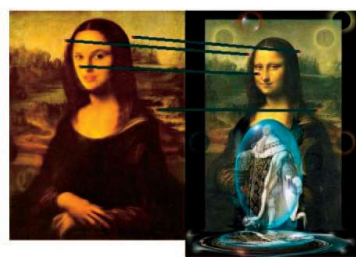
(c)



(d)

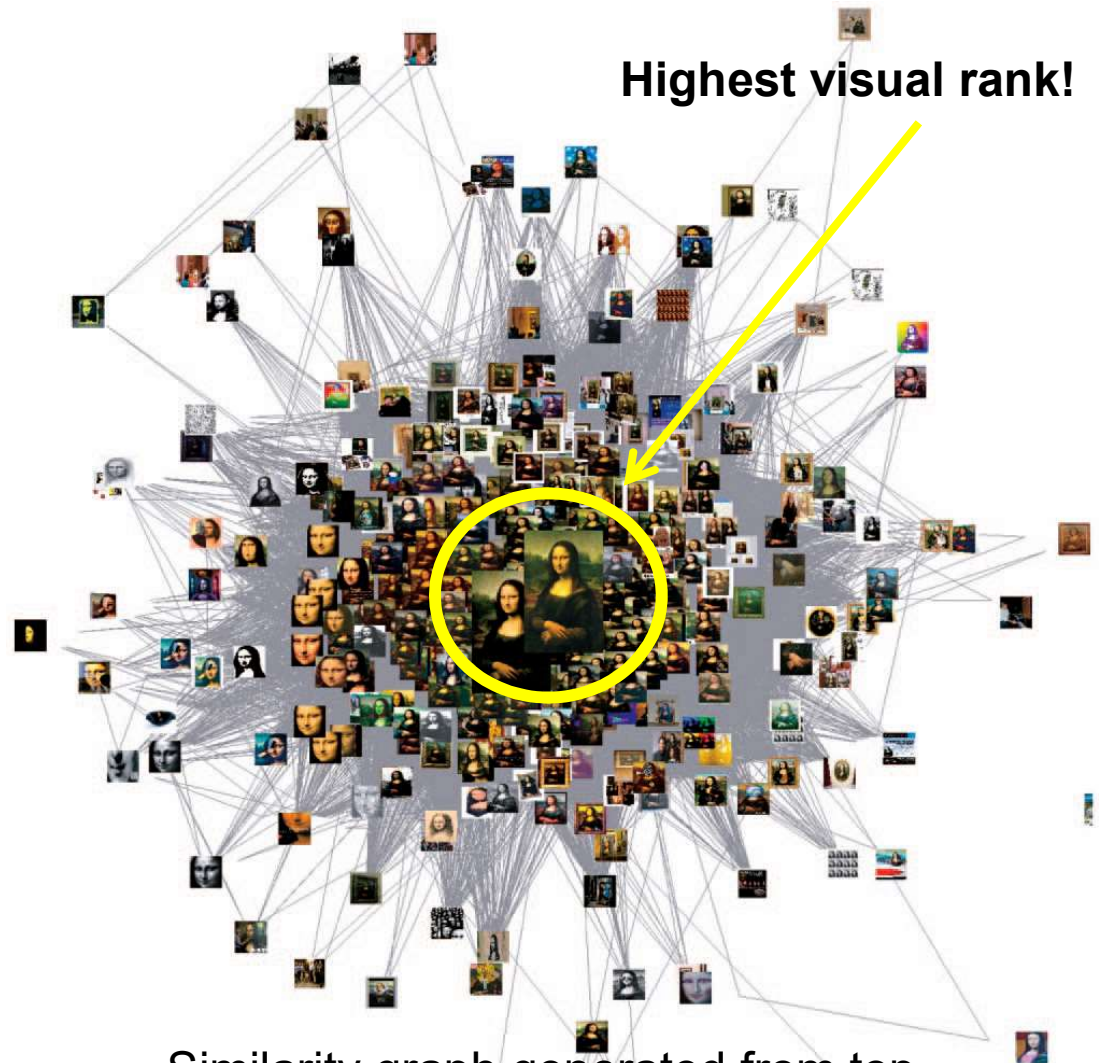


(e)



(f)

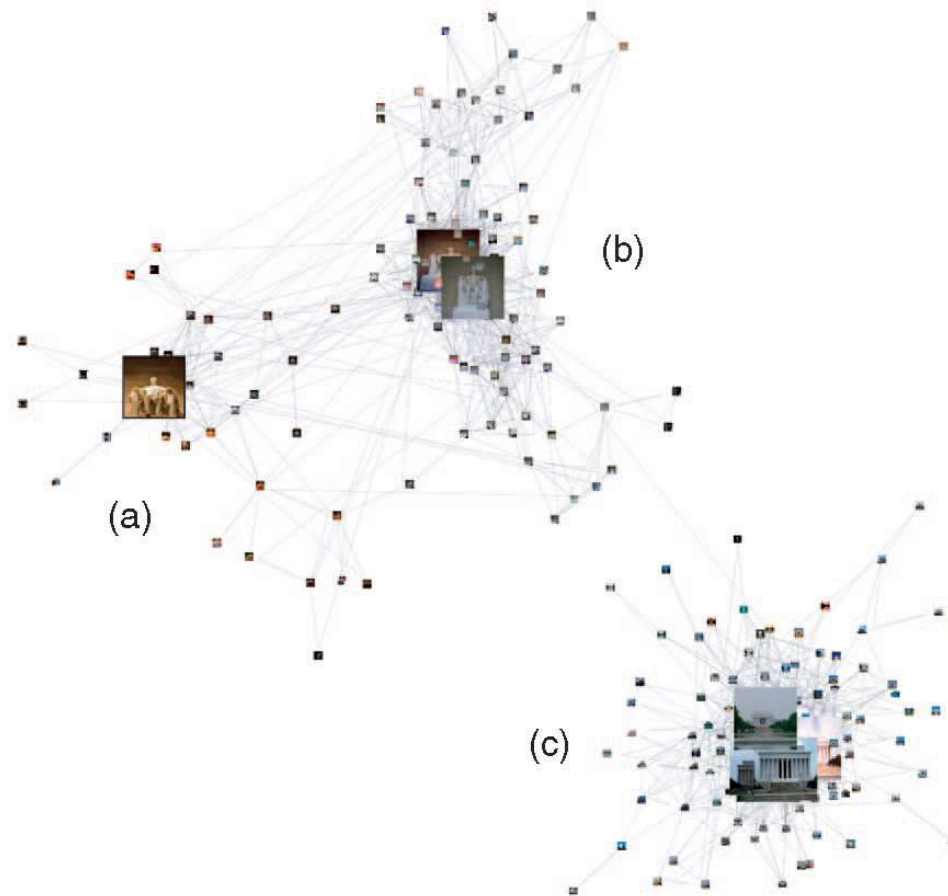
Original has more matches to rest
Kristen Grauman



Similarity graph generated from top
1,000 text search results of "Mona-Lisa"

Results: Visual Rank

[Jing and Baluja, PAMI 2008]



Similarity graph generated from top 1,000 text search results of “Lincoln Memorial”

Kristen Grauman Note the diversity of the high-ranked images.

Mining for common visual patterns

In addition to visual search, want to be able to **summarize, mine, and rank** the large collection as a whole.

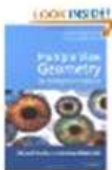
We'll look briefly at a few recent examples:

- **Connected component clustering** via hashing [Geometric Min-hash, Chum et al. 2009]
- **Visual Rank** to choose “image authorities” [Jing and Baluja, 2008]
- **Frequent item-set mining** with spatial patterns [Quack et al., 2007]

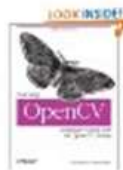
Frequent item-sets

Frequently Bought Together

Customers buy this book with [Learning OpenCV: Computer Vision with the OpenCV Library](#) by Gary Bradski



+



Price For Both: \$131.77

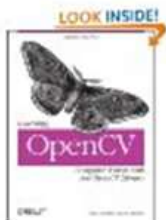


Add both to Cart

Add both to Wish List

[Show availability and shipping details](#)

Customers Who Bought This Item Also Bought



[Learning OpenCV: Computer Vision with the OpenCV...](#) by Gary Bradski



[Computer Vision: A Modern Approach](#) by David A. Forsyth



[Pattern Recognition and Machine Learning...](#) by Christopher M. Bishop

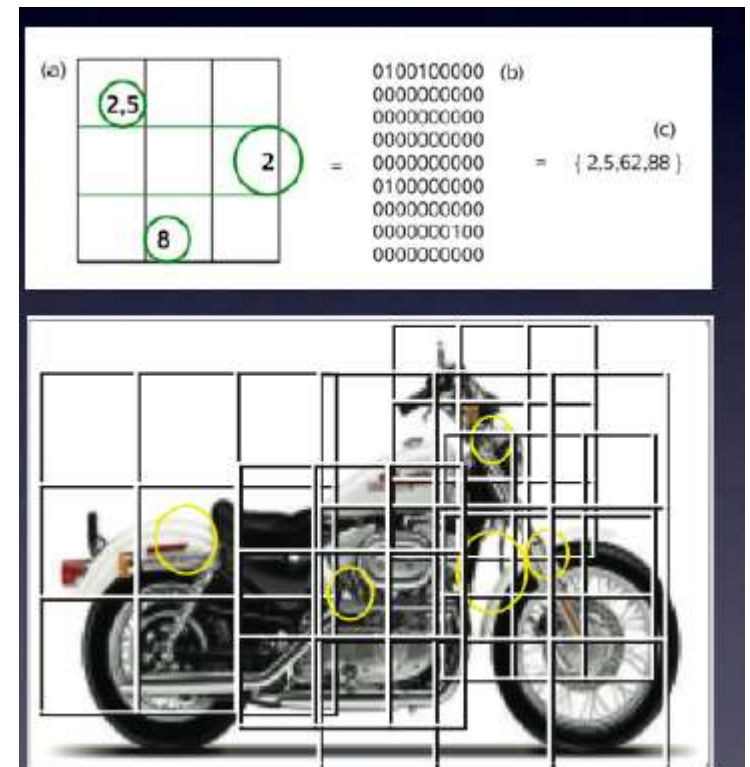


[Machine Vision, Third Edition: Theory, Algorithms...](#) by E. R. Davies

Frequent item-set mining for spatial visual patterns

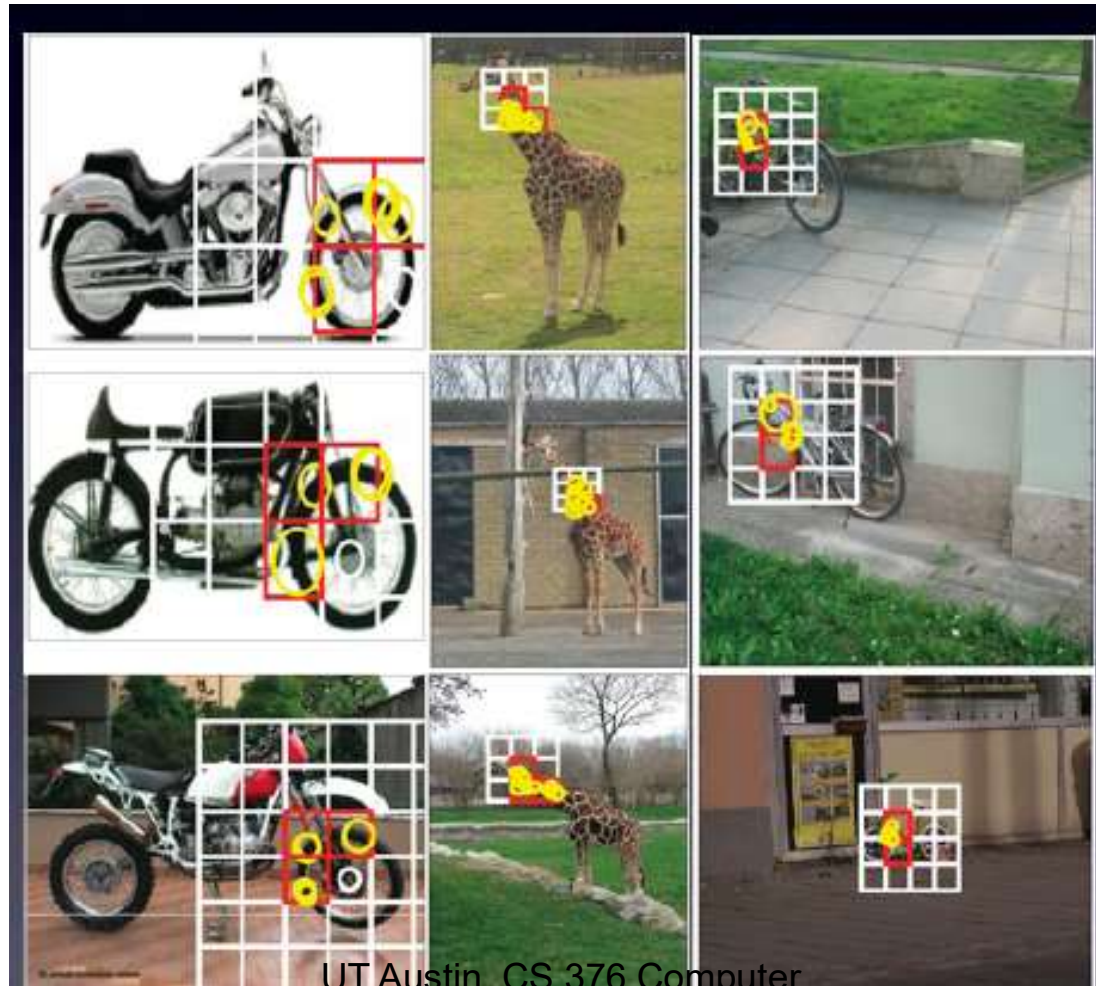
[Quack, Ferrari, Leibe, Van Gool, CIVR 2006, ICCV 2007]

- What configurations of local features frequently occur in large collection?
- **Main idea:** Identify *item-sets* (visual word layouts) that often occur in *transactions* (images)
- Efficient algorithms from data mining (e.g., Apriori algorithm, Agrawal 1993)



Frequent item-set mining for spatial visual patterns

[Quack, Ferrari, Leibe, Van Gool, CIVR 2006, ICCV 2007]



UT Austin, CS 376 Computer
Vision - lecture 21

Frequent item-set mining for spatial visual patterns

[Quack, Ferrari, Leibe, Van Gool, CIVR 2006, ICCV 2007]

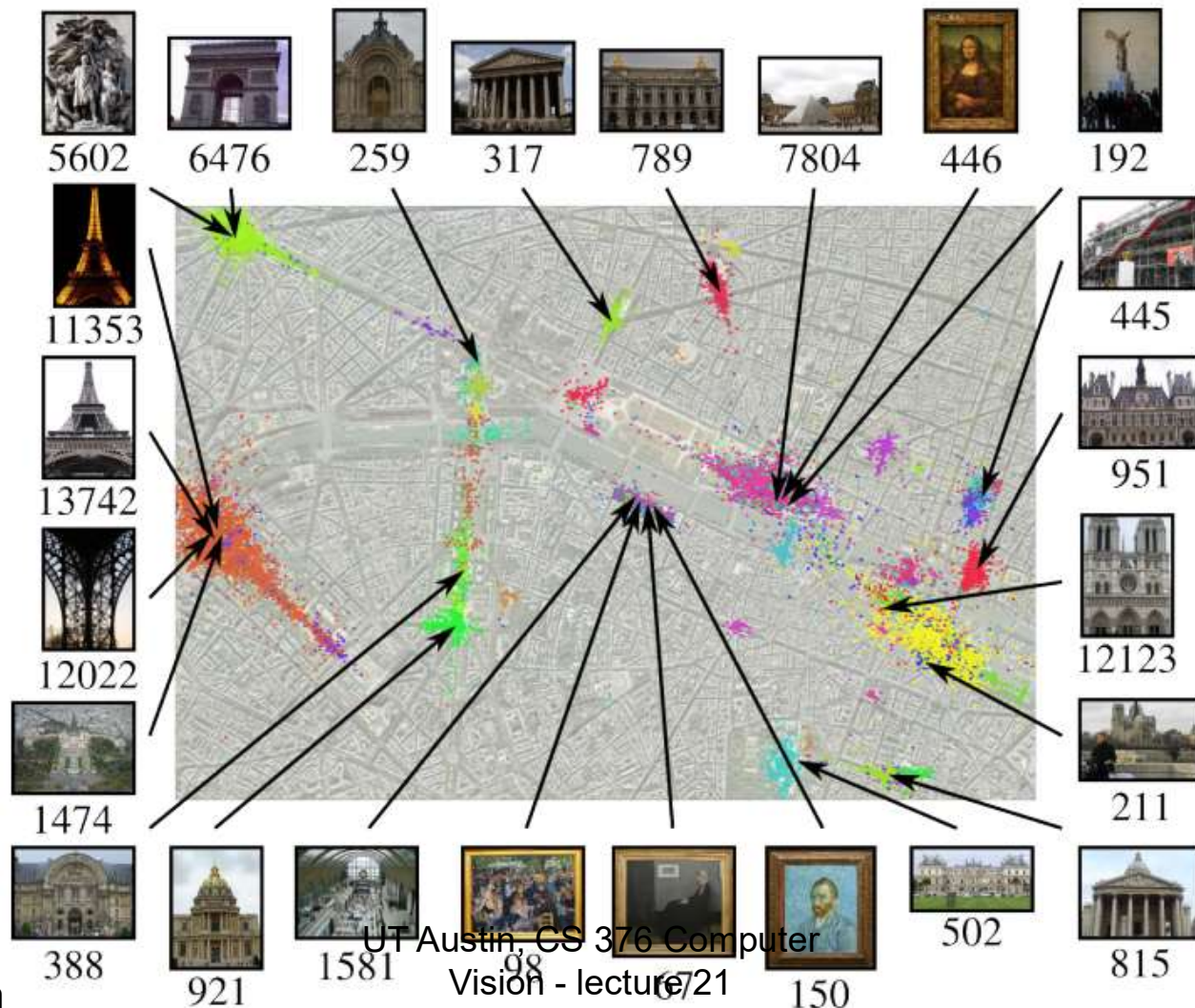


Two example itemset clusters

UT Austin, CS 376 Computer
Vision - lecture 21

Discovering favorite views

Discovering Favorite Views of Popular Places with Iconoid Shift. T. Weyand and B. Leibe. ICCV 2011.



Today

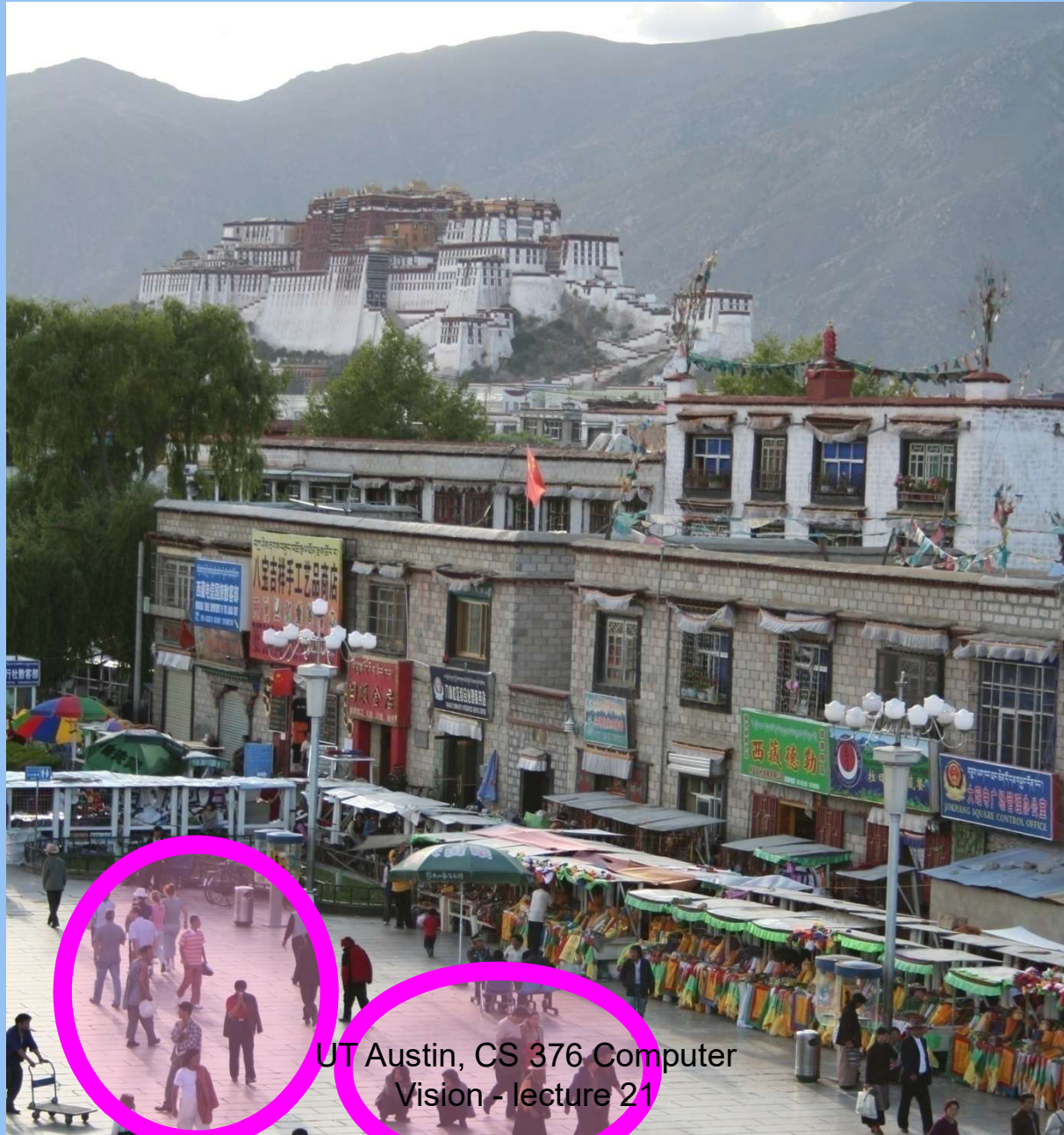
- Discovering visual patterns
 - Randomized hashing algorithms
 - Mining large-scale image collections
- **Introduction to visual categorization**

What does recognition involve?



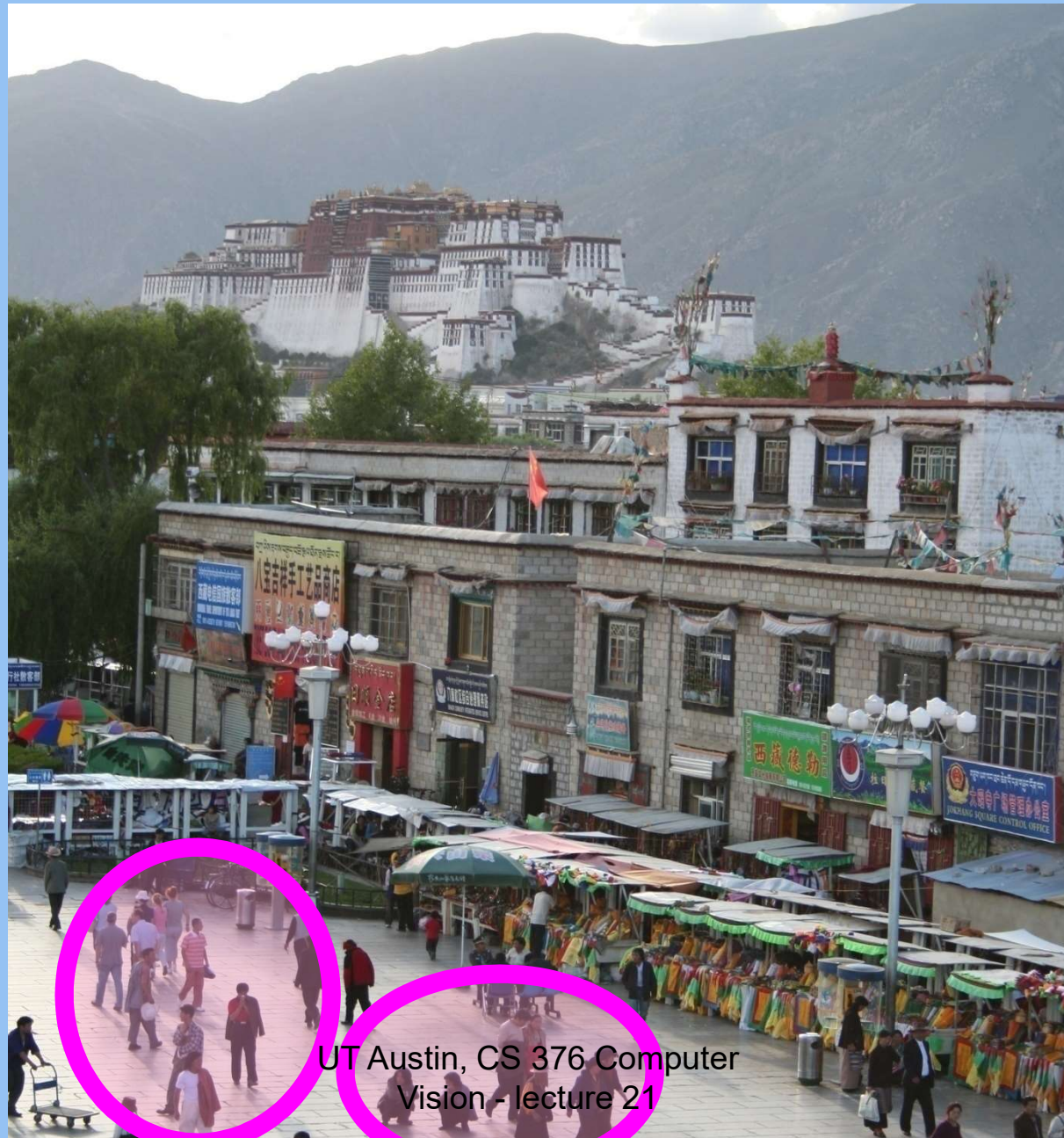
UT Austin, CS 376 Computer
Vision, Lecture 21

Detection: are there people?



UT Austin, CS 376 Computer
Vision - lecture 21

Activity: What are they doing?

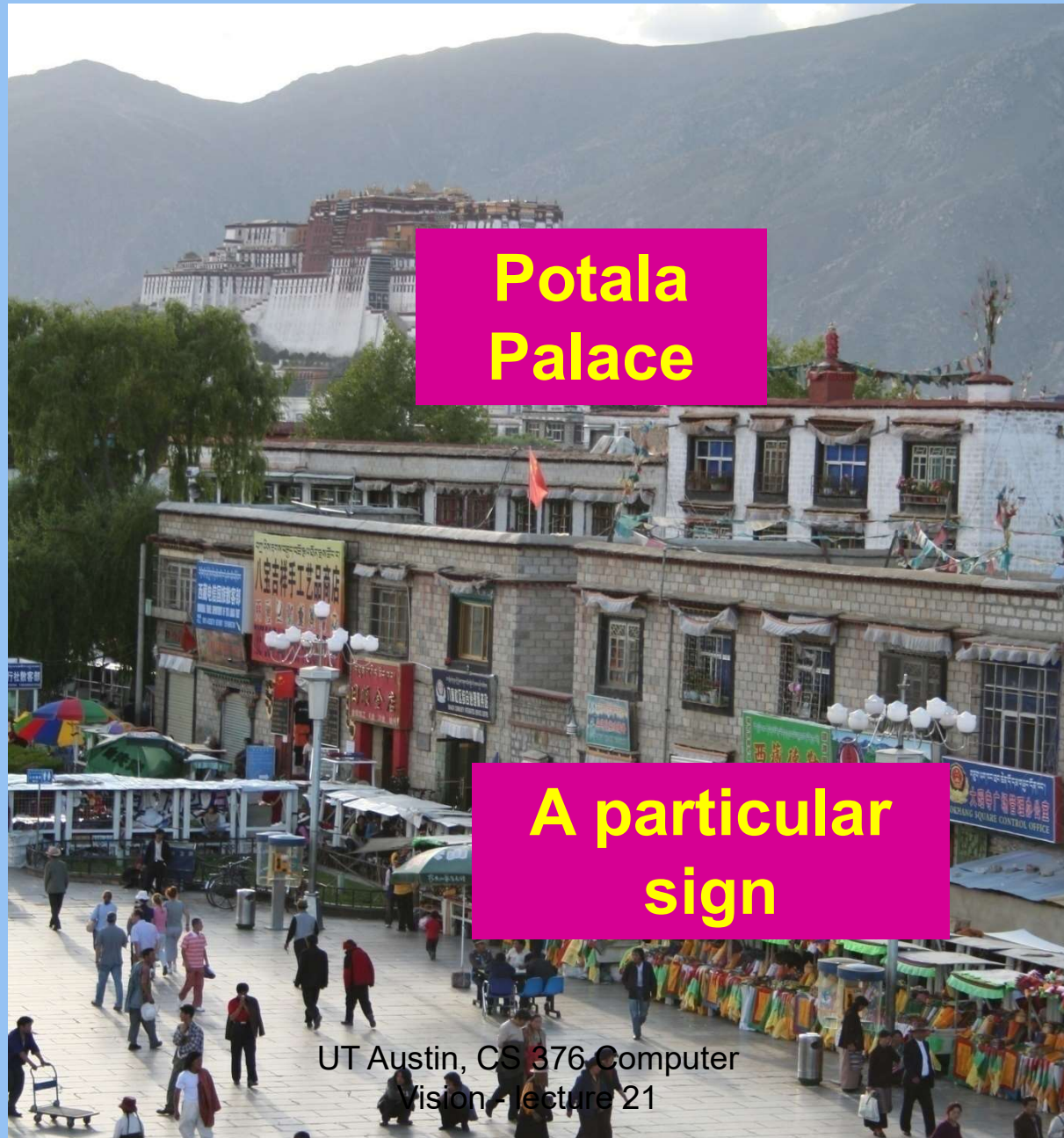


UT Austin, CS 376 Computer
Vision - lecture 21

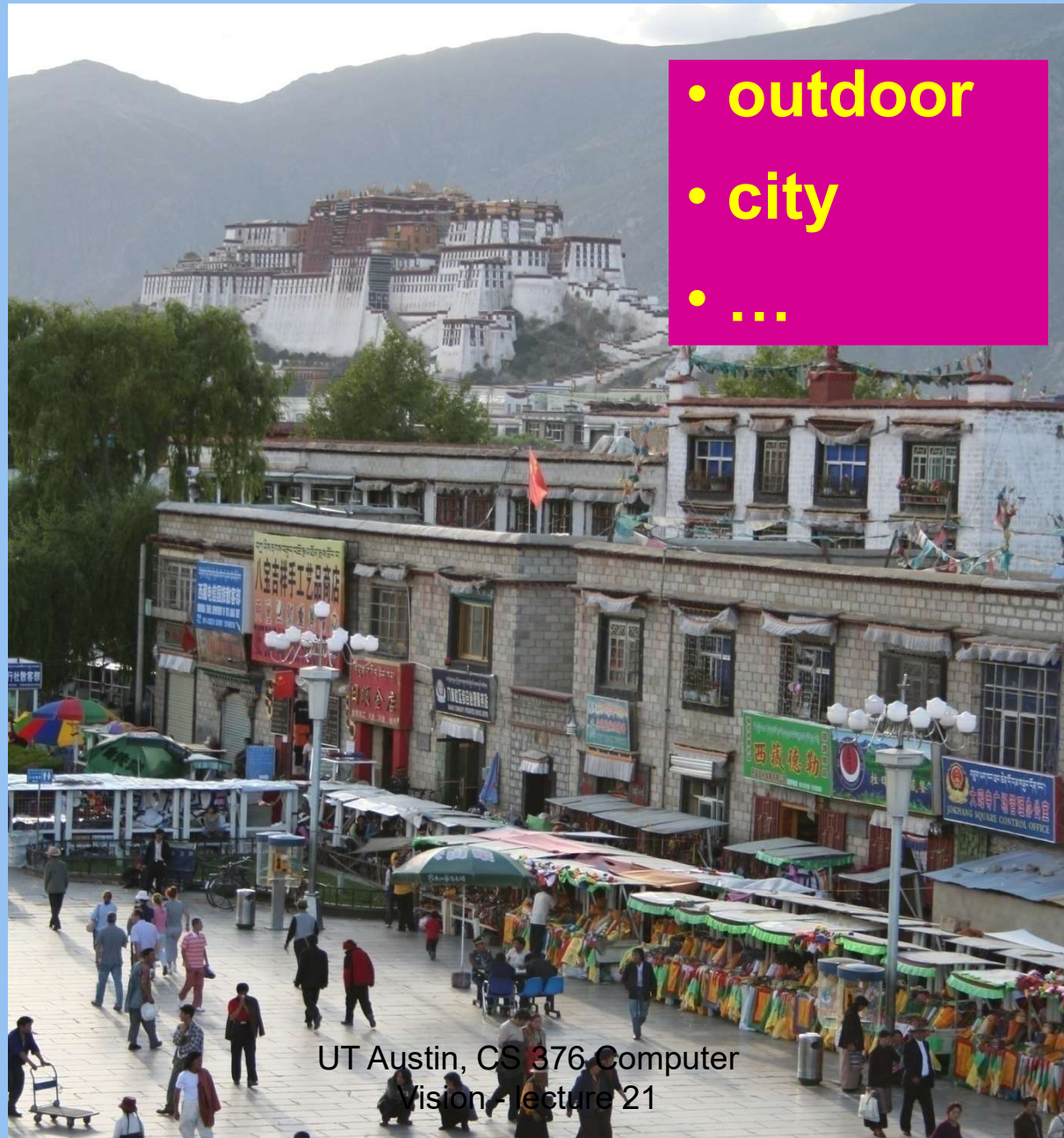
Object categorization



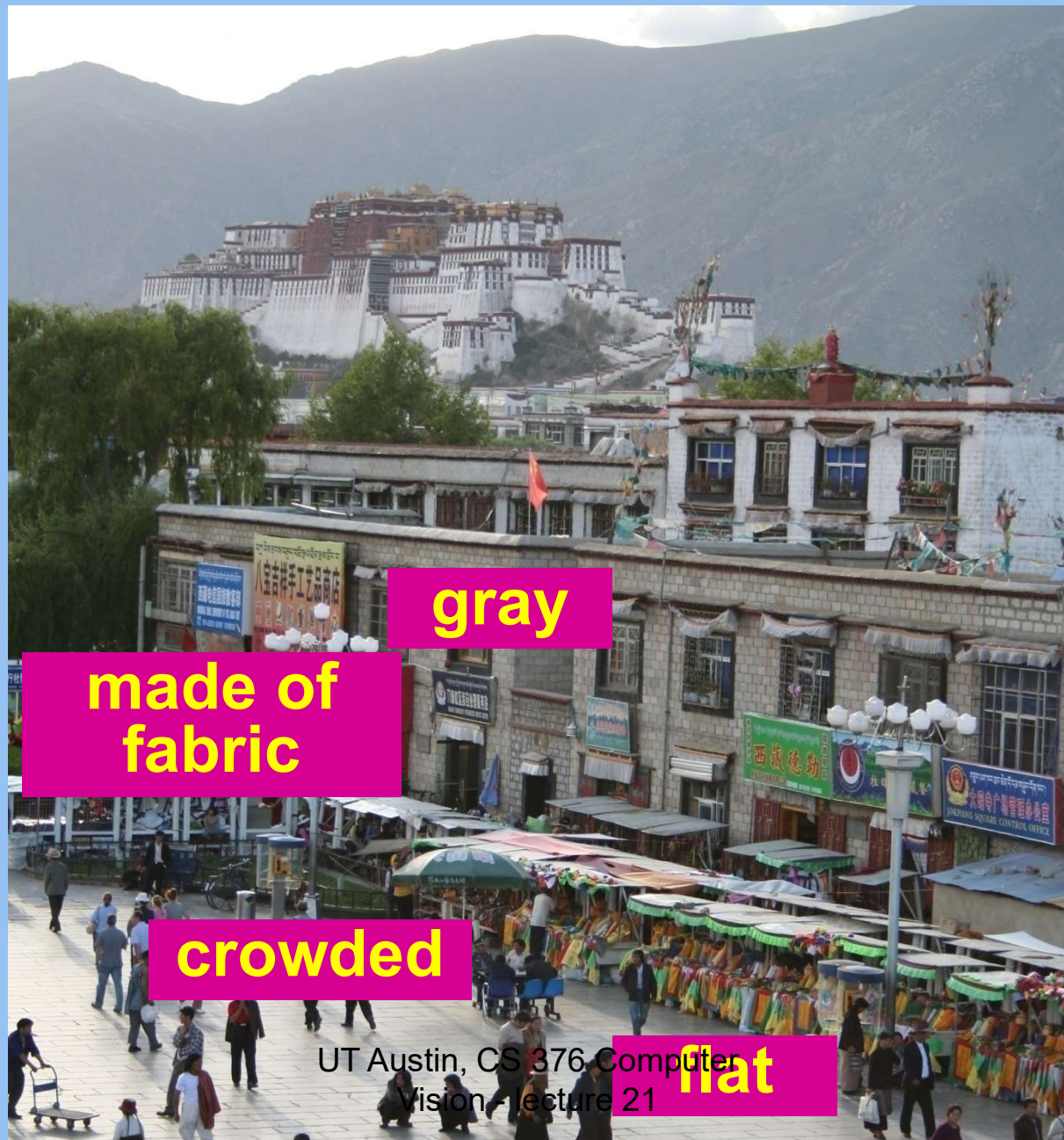
Instance recognition



Scene and context categorization



Attribute recognition



Object Categorization

- Task Description
 - “Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label.”
- Which categories are feasible visually?

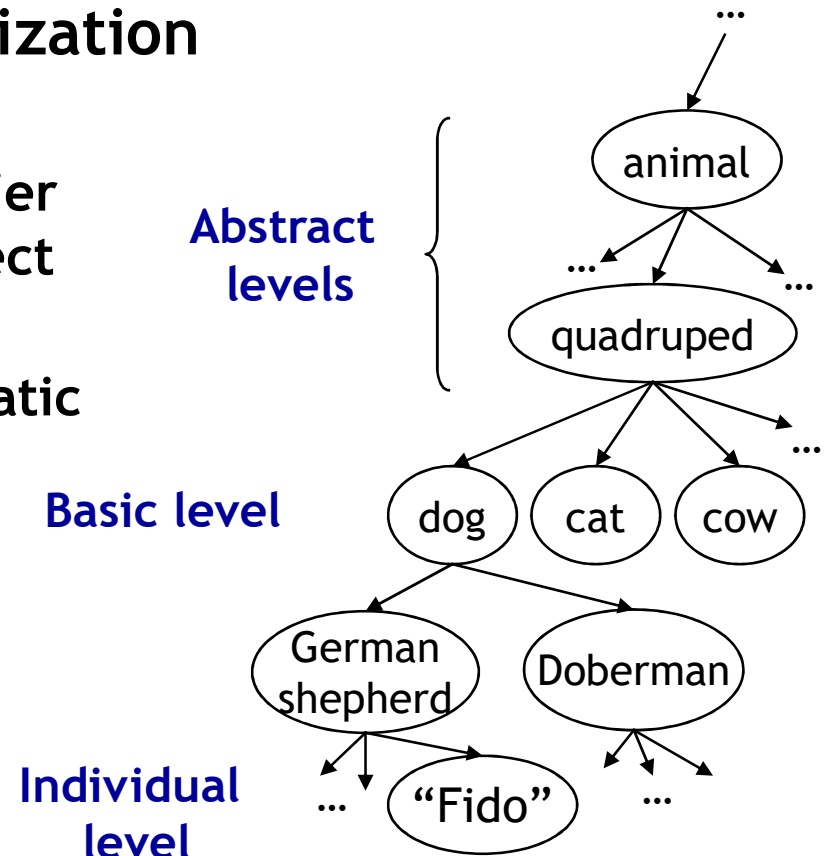


Visual Object Categories

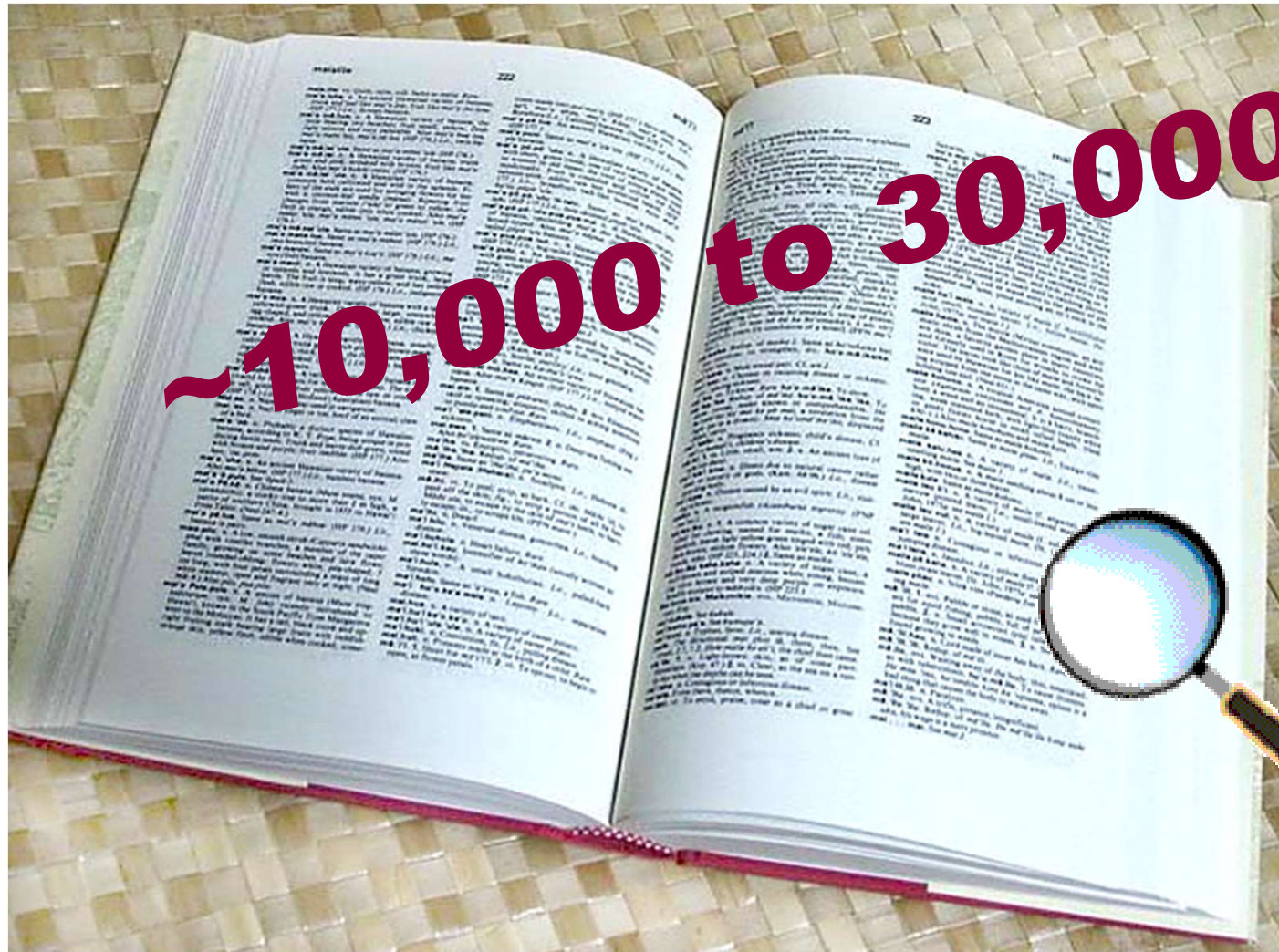
- **Basic Level Categories in human categorization**
[Rosch 76, Lakoff 87]
 - The highest level at which category members have similar perceived shape
 - The highest level at which a single mental image reflects the entire category
 - The level at which human subjects are usually fastest at identifying category members
 - The first level named and understood by children
 - The highest level at which a person uses similar motor actions for interaction with category members

Visual Object Categories

- Basic-level categories in humans seem to be defined predominantly visually.
- There is evidence that humans (usually) start with basic-level categorization *before* doing identification.
 - ⇒ Basic-level categorization is easier and faster for humans than object identification!
 - ⇒ How does this transfer to automatic classification algorithms?



How many object categories are there?



UT Austin, CS 376 Computer

Vision - lecture 21

Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Biederman 1987



Other Types of Categories

- Functional Categories
 - e.g. chairs = “*something you can sit on*”



Why recognition?

- Recognition a fundamental part of perception
 - e.g., robots, autonomous agents
- Organize and give access to visual content
 - Connect to information
 - Detect trends and themes

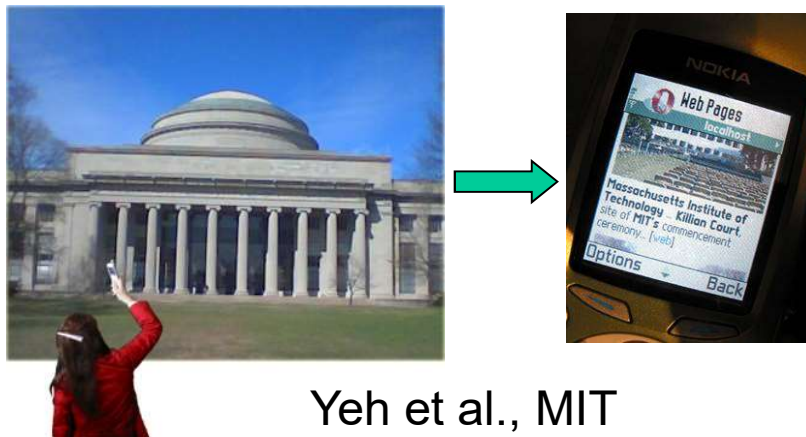
Autonomous agents able to detect objects



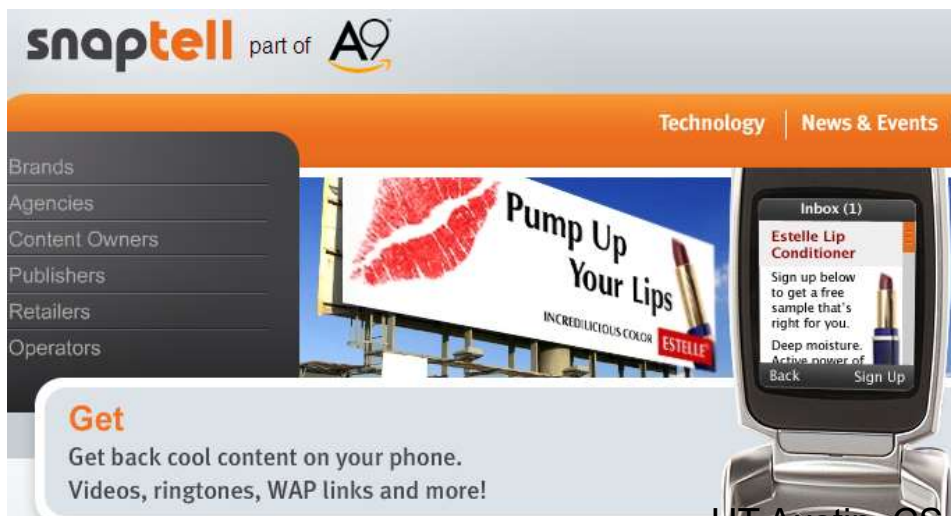
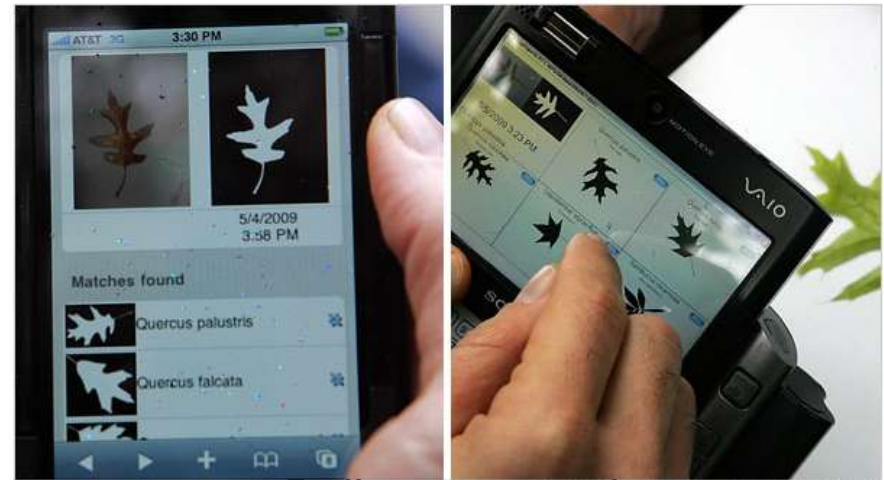
UT Austin, CS 376 Computer
Vision - lecture 21

<http://www.darpa.mil/grandchallenge/gallery.asp>


Posing visual queries



Digital Field Guides Eliminate the Guesswork



Finding visually similar objects


**like** visual shopping *alpha*


[My Like List](#) | [NewsLetter](#) | [Blog](#)


ALL | SHOES | BAGS | WOMEN'S APPAREL | MEN'S APPAREL | KIDS | ACCESSORIES | JEWELRY & WATCHES | HOLIDAY | FOR THE HOME


IN Women's Shoes

Refine by Style



[Pumps](#)



[Sandals](#)



[Flats](#)



[Patent](#)

Refine by Color



[crimson](#)



[taupe](#)


[scarlet](#)


[c](#)

Refine by Brand


[Clarks](#)



[Sofft](#)

Why is Like.com Different?


Like is a visual shopping engine that lets you find items by color, shape and pattern.

Click on [Likeness Search](#) to get started

Your Search Item



Which part of the image do you like?
Draw a box on the item to focus your search on that area.




Cole Haan - Carma OT Air Pump
\$278.95
[More Details](#) + [Save to LikeList](#)
[Shop at Zappos.com](#)


[All Products](#) > [Shoes](#) > [Women's Shoes](#) > [Cole Haan](#) > Cole Haan - Carma OT Air Pump

Search Results Results 1 - 20 of 140,207


Sort By LikenessSM Price Change Your View: ☐ ☐ ☐ ☐ 1 2 3 4 5 6 7 [NEXT >>](#)



Natural Comfort - LV58
a sexy classic pump with a pillow-like footbed to keep your feet happy. leather or patent leather upper. wrapped memory-foam footbed. covered heel. leather sole.
[Compare Prices](#) [More Details](#) [Save to LikeList](#)



Cole Haan 'Carma Air' Patent Leather Open Toe Pump
Open toe styles a sleek, cushioned pump with a wrapped heel and a mini platform. Color(s): black patent, dark chocolate suede, wine patent, black python, natural python, beige leather. Brand: Cole Haan.
[Compare Prices](#) [More Details](#) [Save to LikeList](#)



rsvp - Caitlyn
an easy on the eyes pump features craftsmanship to make it easy on your feet too. patent leather uppers. almond shaped toe. cushioned footbed. covered heel. leather outsole. made in brazil. 7 oz.
[Compare Prices](#) [More Details](#) [Save to LikeList](#)

\$99.95
[Shop at Zappos.com](#)
Free Shipping Available

Shop for more items like this:
[Likeness Search](#)

\$275.00
[Shop at NORDSTROM.com](#)

Shop for more items like this:
[Likeness Search](#)

\$89.95
[Shop at Zappos.com](#)
Free Shipping Available

Shop for more items like this:
[Likeness Search](#)

Exploring community photo collections



Snaveley et al.



UT Austin, CS 376 Computer Vision - lecture 21

Slide: Kristen Grauman

Simon & Seitz

Discovering visual patterns



Objects Sivic & Zisserman



Categories Lee & Grauman



Actions Wang et al.
UT Austin, CS 376 Computer
Vision - lecture 21

Auto-annotation



Figure 9. Results of automatic object-level annotation with bounding boxes. Groundtruth annotation is shown with dashed lines, correct detection with solid green lines, false detections with solid red lines. Auto-annotation with related Wikipedia articles is also shown. All results are also labeled with their GPS position and estimated tags (not shown here).

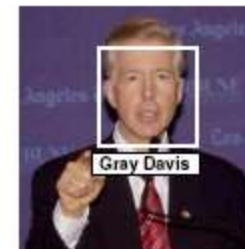
Gammeter et al.



President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters



British director Sam Mendes and his partner actress Kate Winslet arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The film stars Tom Hanks as a Chicago hit man who has a separate family life and co-stars Paul Newman and Jude Law. REUTERS/Dan Chung



Incumbent California Gov. Gray Davis (news - web sites) leads Republican challenger Bill Simon by 10 percentage points - although 17 percent of voters are still undecided, according to a poll released October 22, 2002 by the Public Policy Institute of California. Davis is shown speaking to reporters after his debate with Simon in Los Angeles, on Oct. 7. (Jim Ruymen/Reuters)

T. Berg et al.

UT Austin, CS 376 Computer Vision - lecture 21

Slide: Kristen Grauman

Challenges: robustness



Illumination



Object pose



Clutter



Occlusions



**Intra-class
appearance**



Viewpoint

Challenges: context and human experience



Context cues

Challenges: context and human experience



Context cues



Function



Dynamics

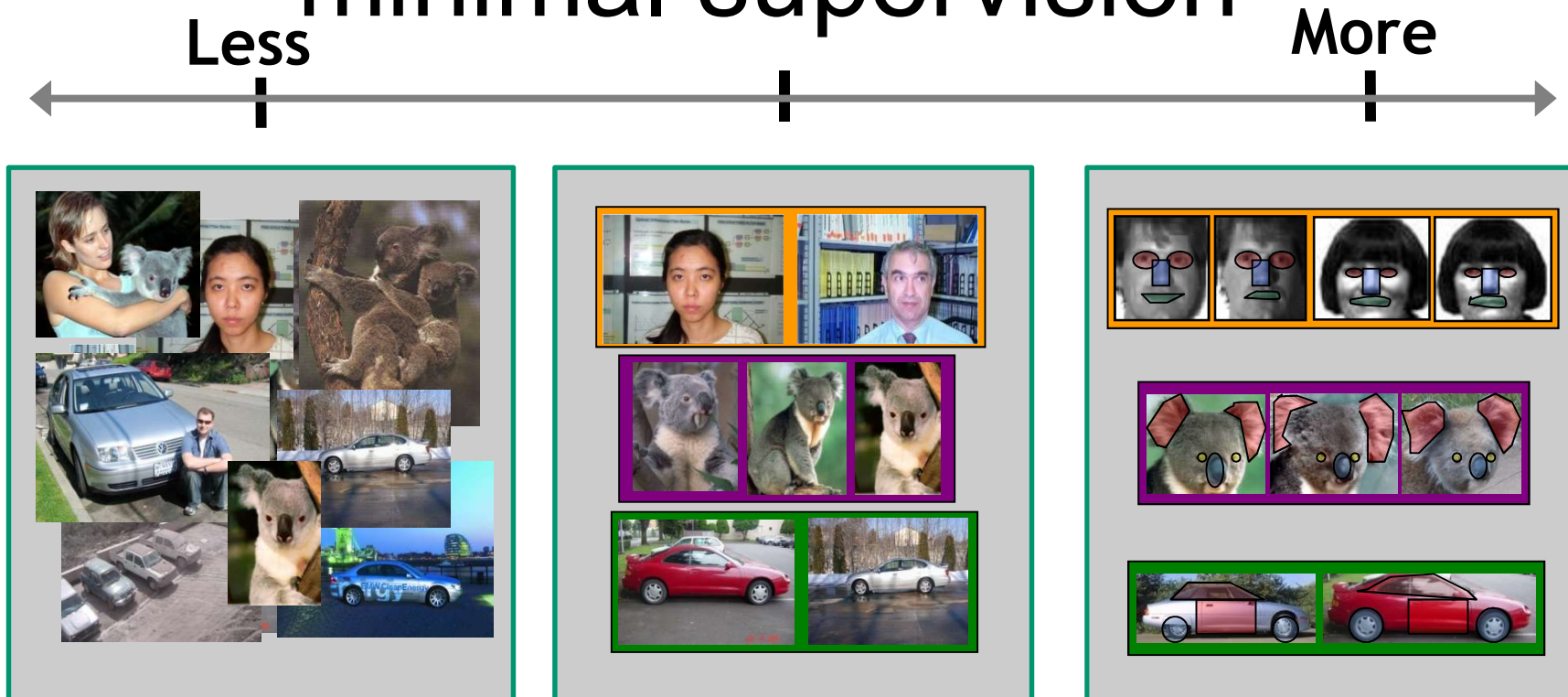
Challenges: complexity

- Millions of pixels in an image
- 30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images online
- 82 years to watch all videos uploaded to YouTube per day!

...

- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

Challenges: learning with minimal supervision



Unlabeled,
multiple objects

Classes labeled,
some clutter

Cropped to object,
parts and classes
labeled

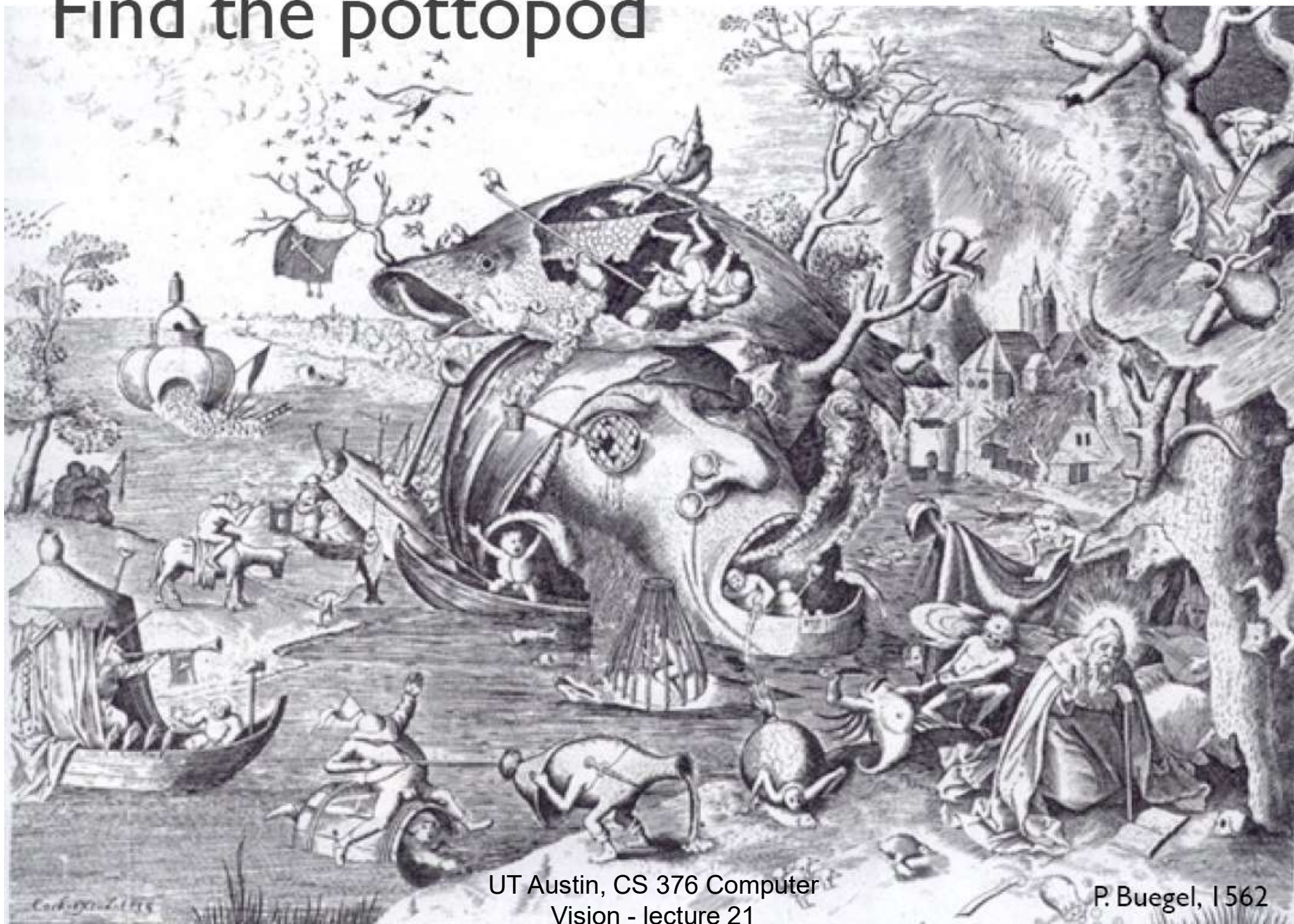
This is a
pottopod



UT Austin, CS 376 Computer
Vision - lecture 21

Slide from Pietro Perona, 2004 Object Recognition workshop

Find the pottopod



UT Austin, CS 376 Computer
Vision - lecture 21

P. Buegel, 1562

Slide from Pietro Perona, 2004 Object Recognition workshop

What kinds of things work best today?

3 6 8 1 7 9 6 6 9 1
6 7 5 7 8 6 3 4 8 5
2 1 7 9 7 1 2 8 4 5
4 8 1 9 0 1 8 8 9 4

Reading license plates,
zip codes, checks



Frontal face detection



Recognizing flat, textured
objects (like books, CD
covers, posters)



Fingerprint recognition

What kinds of things work best today?

clarifai

ABOUT

TECHNOLOGY

API

NEWS

BLOG

CAREERS

CONTACT

Paste a url here...

USE THE URL

CHOOSE A FILE INSTEAD

*By using the demo you agree to our terms of service



Predicted Tags

mammal livestock cattle
pasture agriculture bovine
farm nobody meadow grass

Similar Images



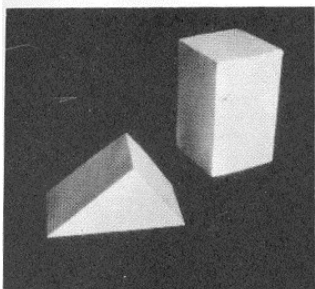
UT Austin, CS 376 Computer
Vision - lecture 21



Predicted Tags

Similar Images

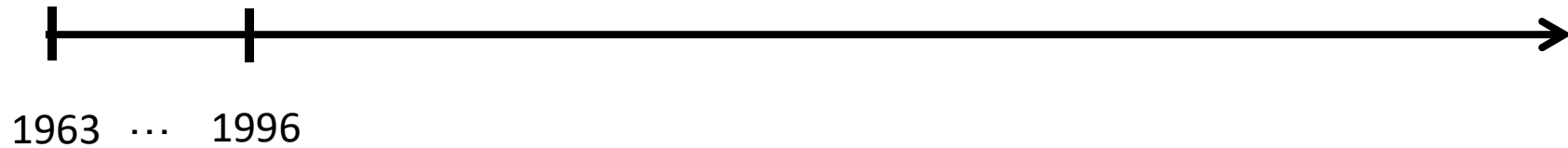
Progress charted by datasets



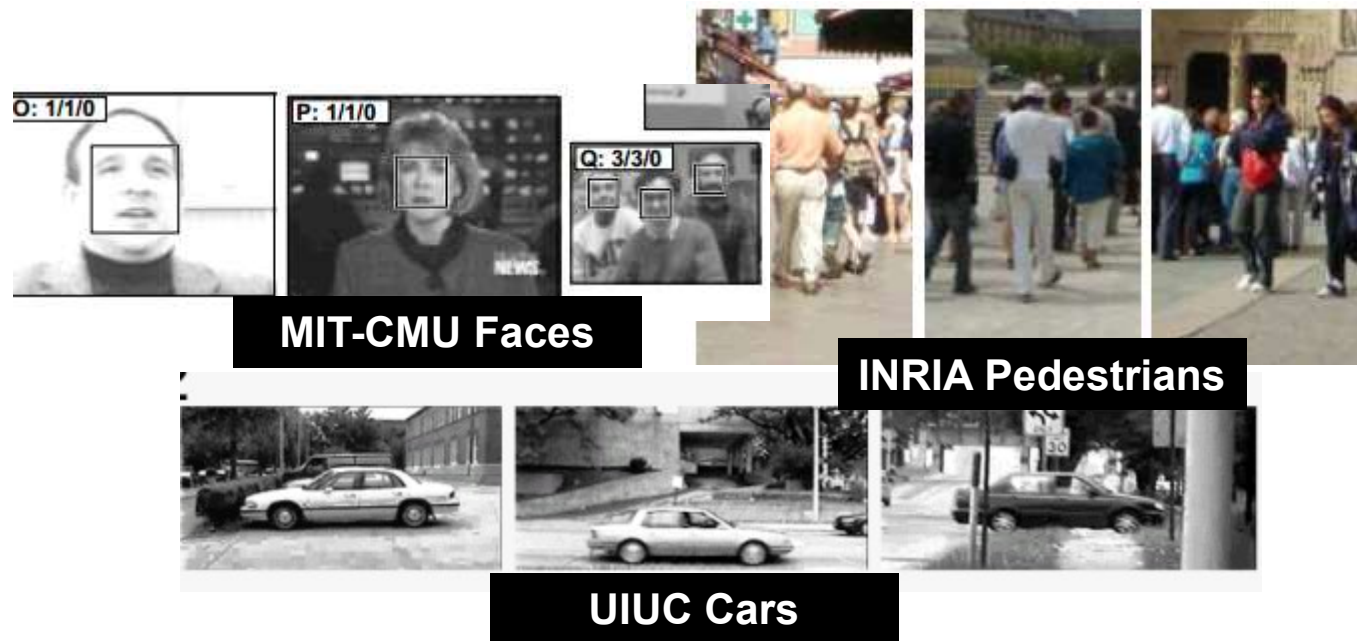
Roberts 1963



COIL



Progress charted by datasets



1963 ... 1996 2000



Slide: Kristen Grauman

UT Austin, CS 376 Computer Vision -
lecture 21

Progress charted by datasets



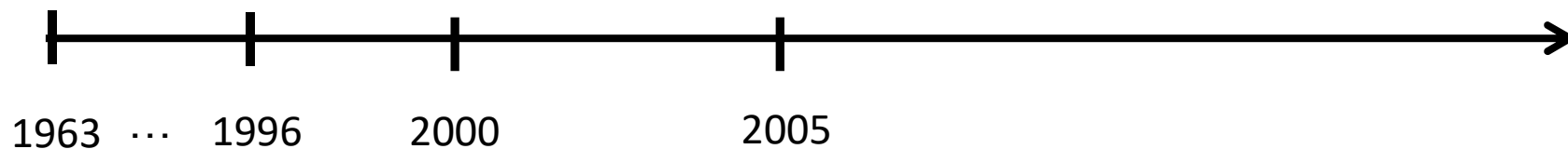
MSRC 21 Objects



Caltech-101



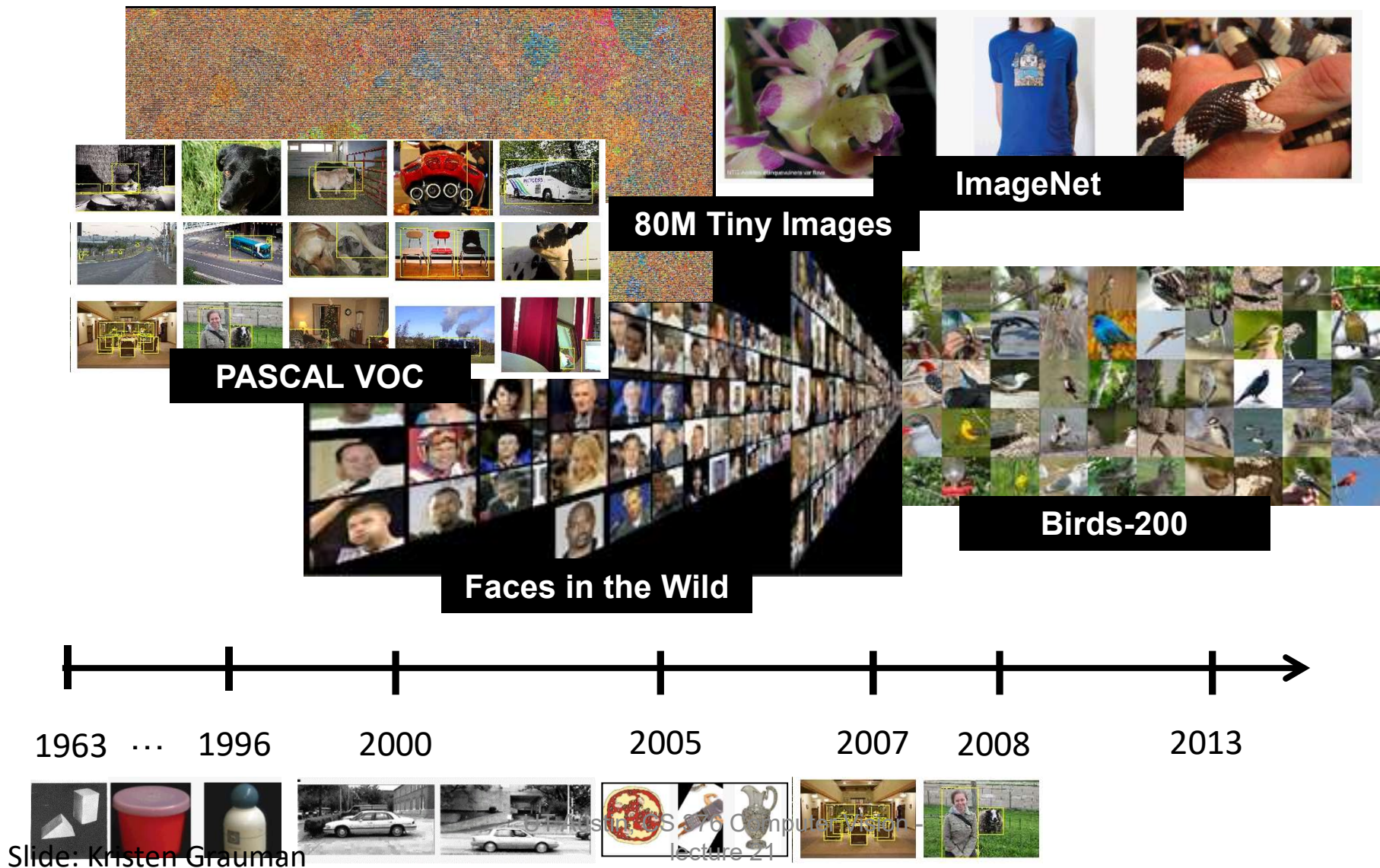
Caltech-256



Slide: Kristen Grauman

Justin, CS 376 Computer Vision -
lecture 21

Progress charted by datasets



Expanding horizons: large-scale recognition

[ABOUT](#)[TECHNOLOGY](#)[API](#) ▾[NEWS](#)[BLOG](#)[CAREERS](#)[CONTACT](#)[USE THE URL](#)[CHOOSE A FILE INSTEAD](#)

*By using the demo you agree to our terms of service



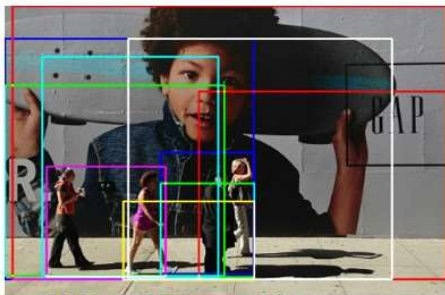
Predicted Tags

mammal livestock cattle
pasture agriculture bovine
farm nobody meadow grass

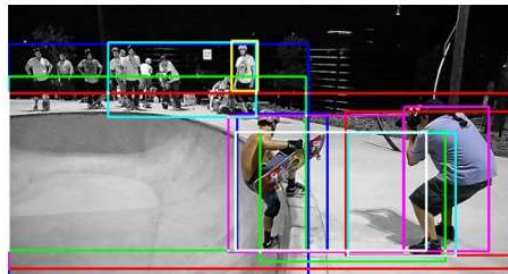
Similar Images



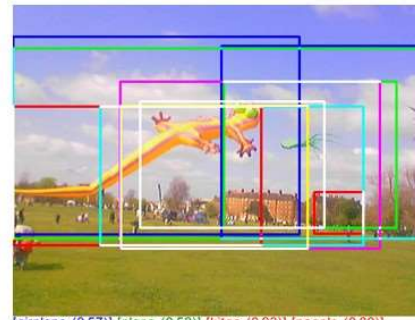
Expanding horizons: captioning



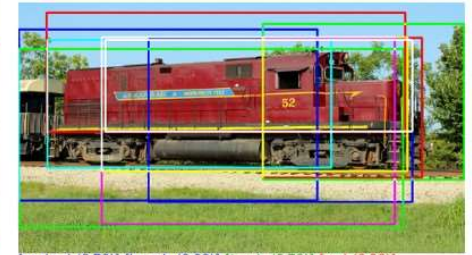
[men (0.59)] [group (0.66)] [woman (0.64)]
[people (0.89)] [holding (0.60)] [playing (0.61)] [tennis (0.69)]
[court (0.51)] [standing (0.59)] [skis (0.58)] [street (0.52)]
[man (0.77)] [skateboard (0.67)]
a group of people standing next to each other
people stand outside a large ad for gap featuring a young boy



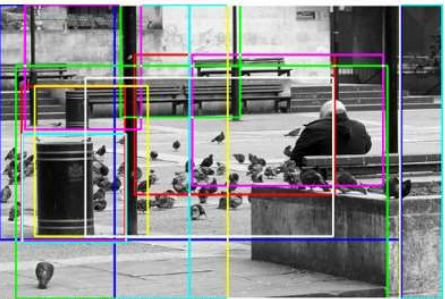
[person (0.55)] [street (0.53)] [holding (0.55)] [group (0.63)] [slope (0.51)]
[standing (0.62)] [snow (0.91)] [skis (0.74)] [player (0.54)]
[people (0.85)] [men (0.57)] [skiing (0.51)]
[skateboard (0.89)] [riding (0.75)] [tennis (0.74)] [trick (0.53)] [skate (0.52)]
[woman (0.52)] [man (0.86)] [down (0.61)]
a group of people riding skis down a snow covered slope
a guy on a skate board on the side of a ramp



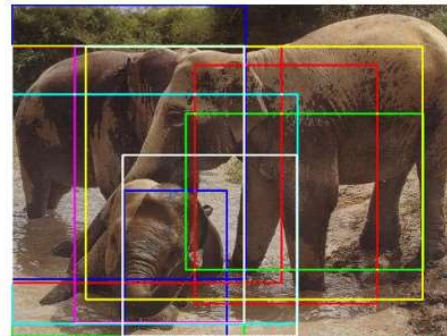
[airplane (0.57)] [plane (0.58)] [kites (0.93)] [people (0.80)]
[flying (0.93)] [man (0.57)] [beach (0.84)] [wave (0.61)]
[sky (0.61)] [kite (0.74)] [field (0.75)]
a couple of people flying kites in a field
people in a field flying different styles of kites



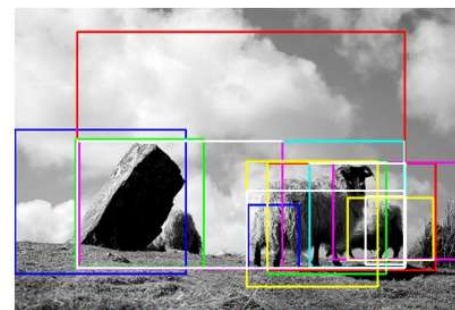
[parked (0.72)] [bench (0.63)] [truck (0.70)] [red (0.88)]
[train (1.00)] [sitting (0.73)] [cars (0.58)] [traveling (0.52)]
[grass (0.65)] [track (0.69)] [car (0.59)] [yellow (0.57)]
[field (0.80)] [engine (0.56)] [down (0.54)] [tracks (0.94)]
a train traveling down train tracks near a field



[umbrella (0.59)] [woman (0.52)]
[fire (0.96)] [hydrant (0.96)] [street (0.79)] [old (0.50)]
[bench (0.81)] [building (0.75)] [standing (0.57)] [baseball (0.55)]
[white (0.82)] [sitting (0.65)] [people (0.79)] [photo (0.53)]
[black (0.84)] [kitchen (0.54)] [man (0.72)] [water (0.56)]
a black and white photo of a fire hydrant
a courtyard full of poles pigeons and garbage cans also has benches on either side of it one of which shows the back of a large person facing in the direction of the pigeons



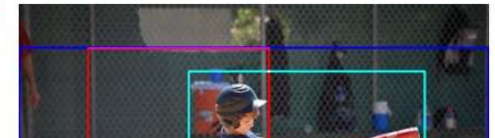
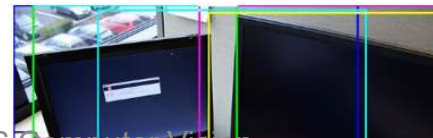
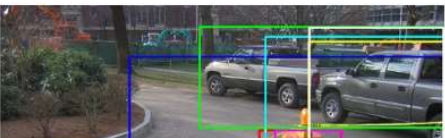
[horse (0.53)] [bear (0.71)] [elephant (0.99)] [elephants (0.95)]
[brown (0.68)] [baby (0.62)] [walking (0.57)] [laying (0.61)]
[man (0.57)] [standing (0.79)] [field (0.65)]
[water (0.83)] [large (0.71)] [dirt (0.65)] [river (0.58)]
a baby elephant standing next to each other on a field
elephants are playing together in a shallow watering hole



[man (0.59)] [beach (0.54)] [sky (0.53)] [bird (0.50)] [field (0.88)]
[snow (0.86)] [mountain (0.59)] [standing (0.81)] [white (0.64)]
[people (0.51)] [dog (0.60)] [cows (0.55)]
[sheep (0.97)] [black (0.84)] [grass (0.64)] [horse (0.60)]
[elephants (0.57)] [bear (0.81)]
a black bear standing on top of a grass covered field
a couple of sheep standing up on a small hill



[bus (0.56)] [car (0.79)] [black (0.57)] [truck (0.86)]
[street (0.57)] [bed (0.51)] [parked (0.55)] [dog (0.65)]
[sitting (0.55)] [man (0.53)] [cat (0.72)]
a dog sitting on top of a car
a cat is lying on the hood of a black car



<https://pdollar.wordpress.com/2015/01/21/image-captioning/>

UT Austin, CS 376 Computer Vision -

lec 10.1

Expanding horizons: visual question answering



What color are her eyes?
What is the mustache made of?



How many slices of pizza are there?
Is this a vegetarian pizza?



Is this person expecting company?
What is just under the tree?



Does it appear to be rainy?
Does this person have 20/20 vision?

Expanding horizons: vision for autonomous vehicles



Turn on your speakers!

Expanding horizons: interactive visual search

QuickTime Player File Edit View Window Help

Widl x

www.widl.it

widl What are you looking for? About

Find visually appealing shoes.

Our recommendations are based on appearance, not click-tracking.

womens girls mens boys

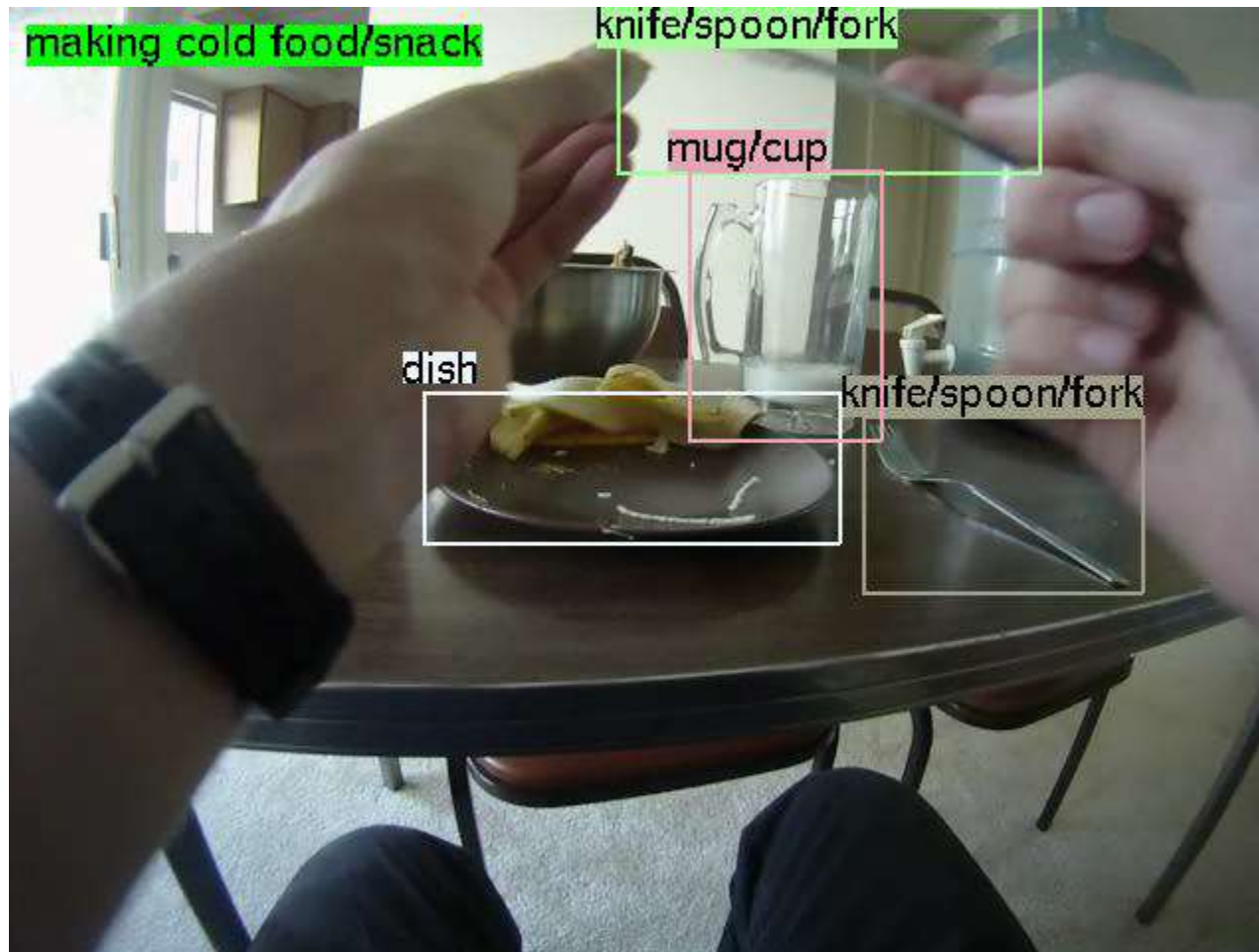
Slide: Kristen Grauman

UT Austin, CS 376 Computer Vision, Lecture 11


f t p g+

Slide navigation icons: back, forward, search, etc.

Expanding horizons: first-person vision



Evolution of methods

- Hand-crafted models
 - 3D geometry
 - Hypothesize and align
 - Hand-crafted features
 - Learned models
 - Data-driven
 - “End-to-end” learning of features and models*,**
- 

* Labeled data availability

** Architecture design decisions, parameters.

Next

- Sliding window object detection (Faces!)

