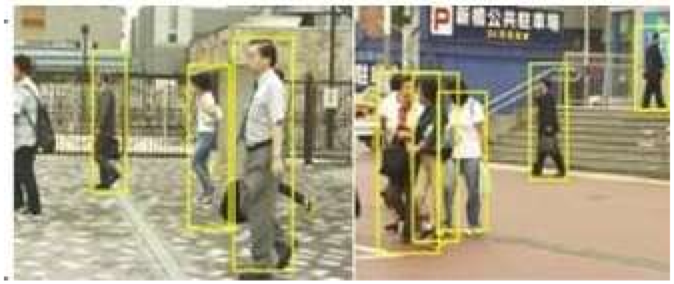
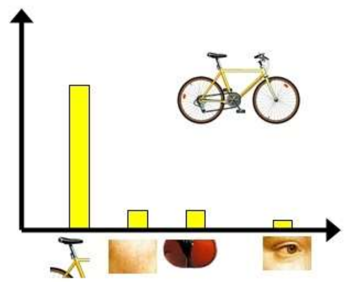
Mining, and Intro to Categorization

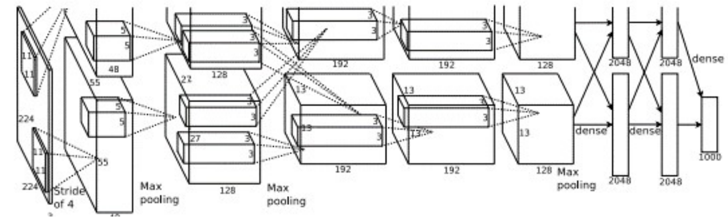
Tues April 10

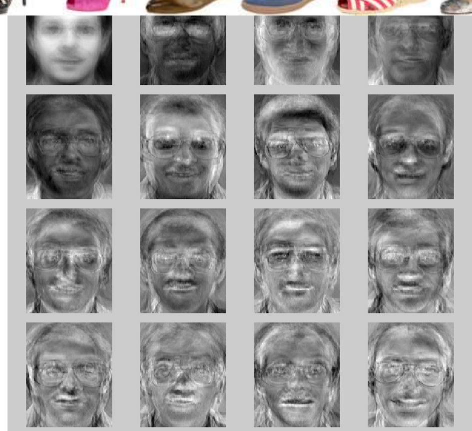
Kristen Grauman
UT Austin

Recognition and learning

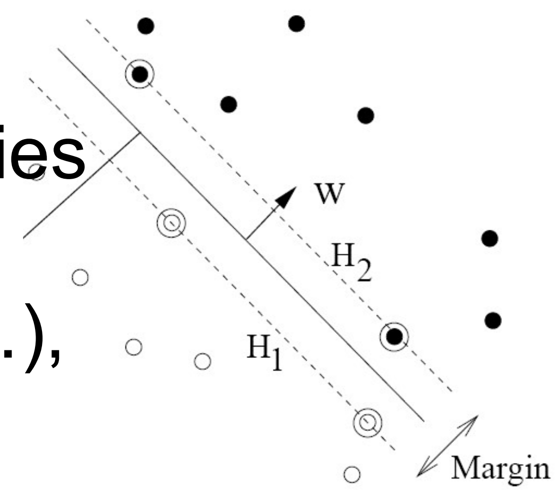








Recognizing categories (objects, scenes, activities, attributes…), learning techniques



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

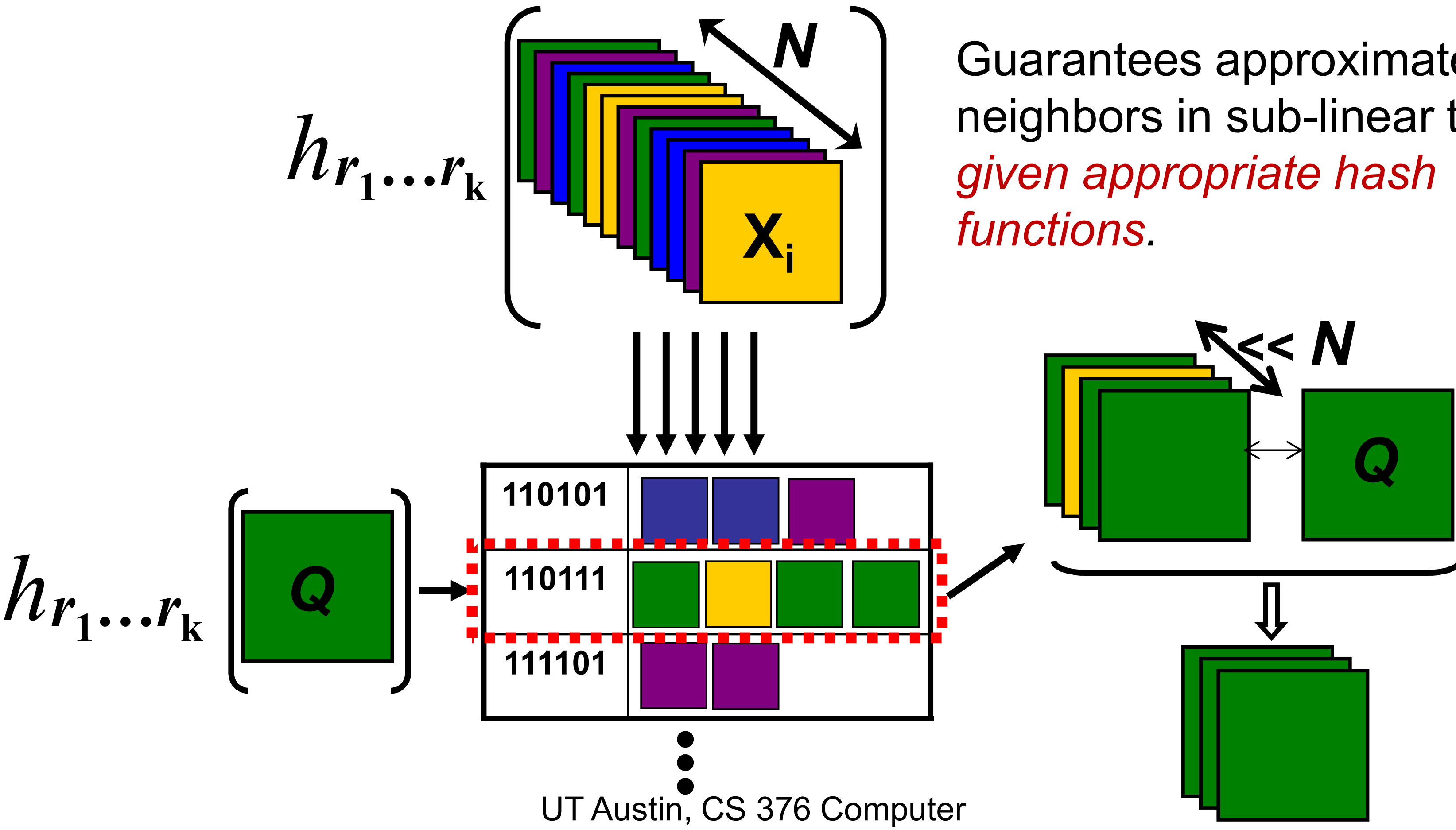
- Introduction to visual categorization

Locality Sensitive Hashing (LSH)

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

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

Guarantees approximate near neighbors in sub-linear time, *given appropriate hash functions.*

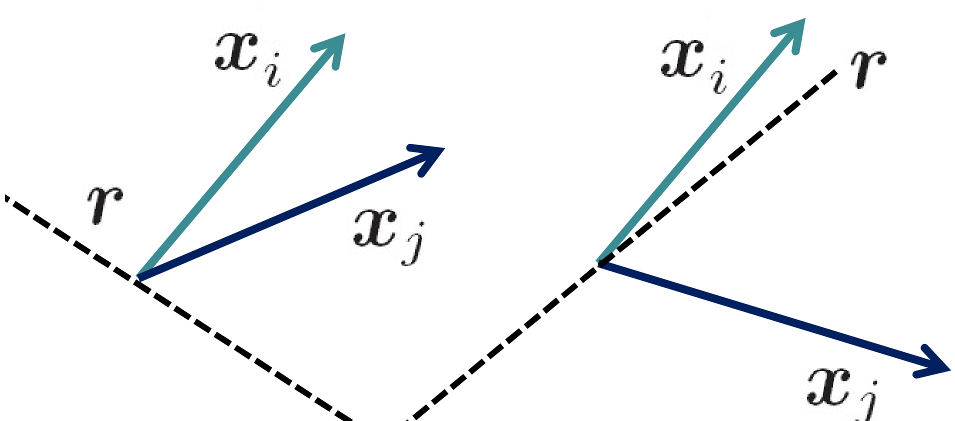


Kristen Grauman

LSH function example: inner product similarity

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

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



High dot product: unlikely to split

Lower dot product: likely to split

Corresponding hash function:

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

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

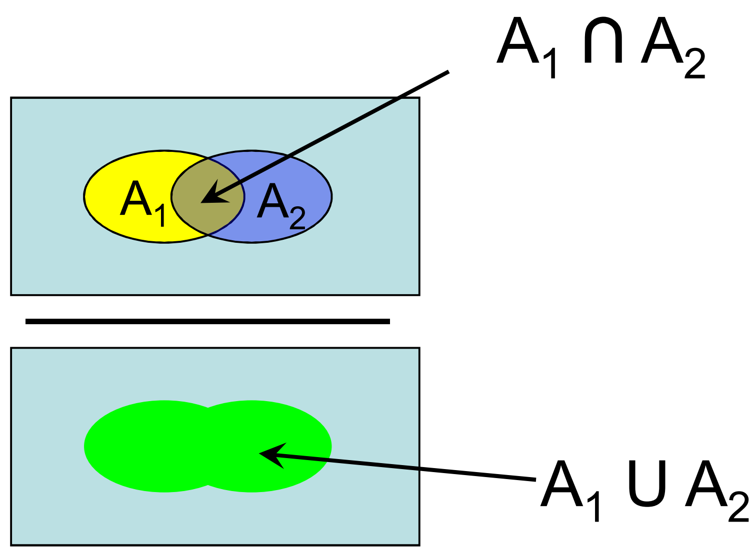
[Goemans and Williamson 1995, Charikar 2004]

LSH function example: Min-hash for set overlap similarity

[Broder, 1999]

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

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



Kristen Grauman

LSH function example: Min-hash for set overlap similarity

Vocabulary	Set ***A***	Set ***B***	Set ***C***
A B C D E F	A B C	B C D	A E F

Random orderings / **min-Hash**

								min-Hash A	min-Hash B	min-Hash C
f_1:	0.41 / 3	0.90 / 6	0.22 / 2	0.59 / 5	0.75 / 4	0.07 / 1	~ Un (0,1)	C	C	F
f_2:	0.19 / 1	0.31 / 2	0.94 / 6	0.55 / 3	0.88 / 5	0.63 / 4	~ Un (0,1)	A	B	A
f_3:	3	2	1	6	4	5		C	C	A
f_4:	4	3	5	6	1	2		B	B	E

overlap (***A***,***B***) = 3/4 (1/2) overlap (***A***,***C***) = 1/4 (1/5) overlap (***B***,***C***) = 0 (0)

[Broder, 1999]

Slide credit: Ondrej Chum

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 ∪ *B*: A C E J Q R V Y Z

h1(*A*) A h1(*B*) A

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

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

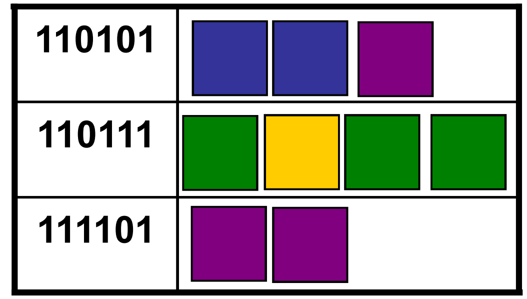
[Broder, 1999]

Slide credit: Ondrej Chum

Multiple hash functions and tables

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

$$P\left(h_{1,\ldots,k}(x) = h_{1,\ldots,k}(y)\right) =$$



- 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

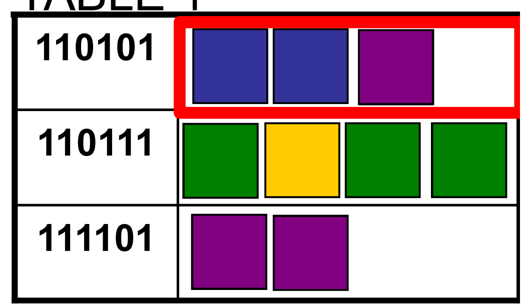
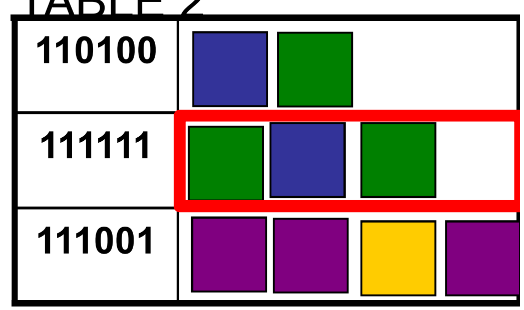


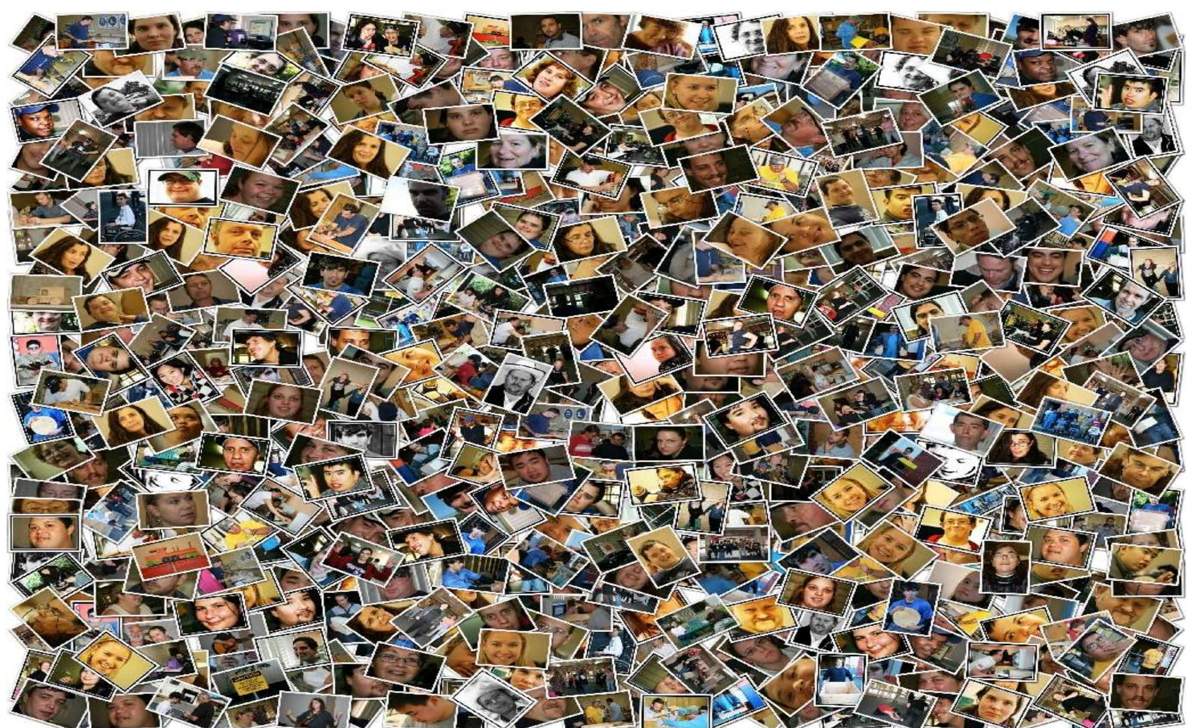
TABLE 2



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?



Kristen Grauman

Mining for common visual patterns

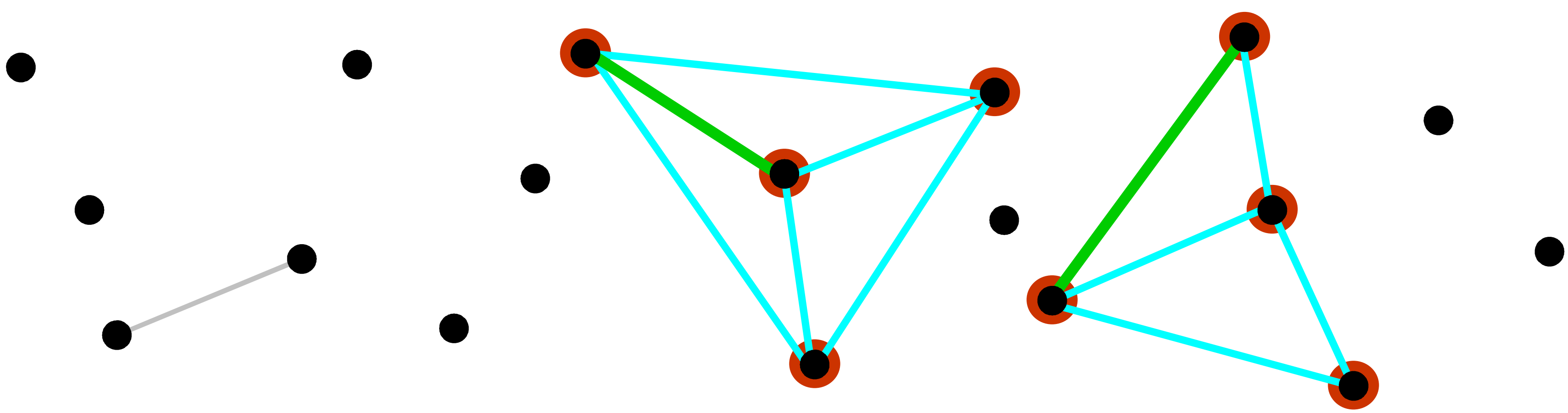
In addition to visual search, want to be able to **summarize, mine, and rank** the large collection as a whole.

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

Slide credit: Ondrej Chum

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

Figure from Ondrej Chum

Results:
Geometric Min-hash clustering

[Chum, Perdoch, Matas, CVPR 2009]

All Soul's

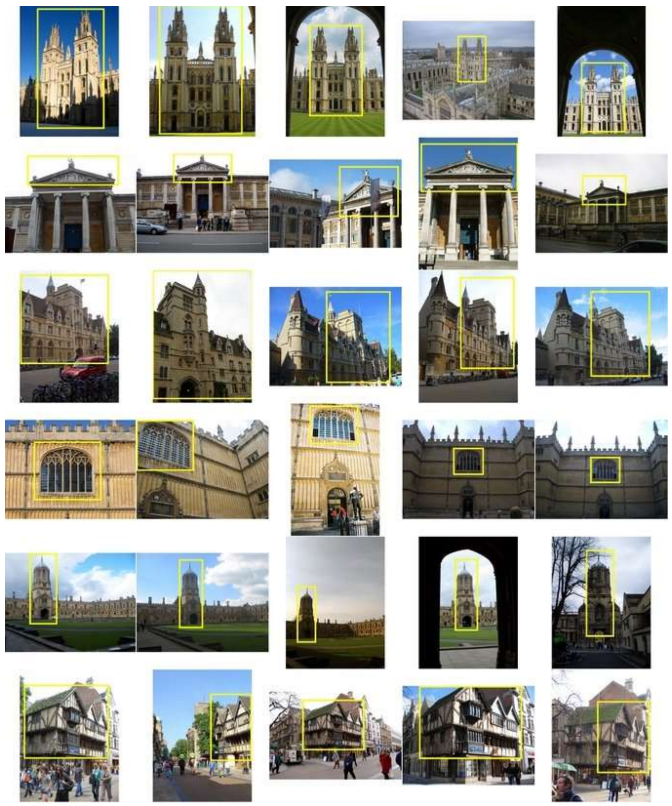
Ashmolean

Balliol

Bodleian

Christ Church

Cornmarket



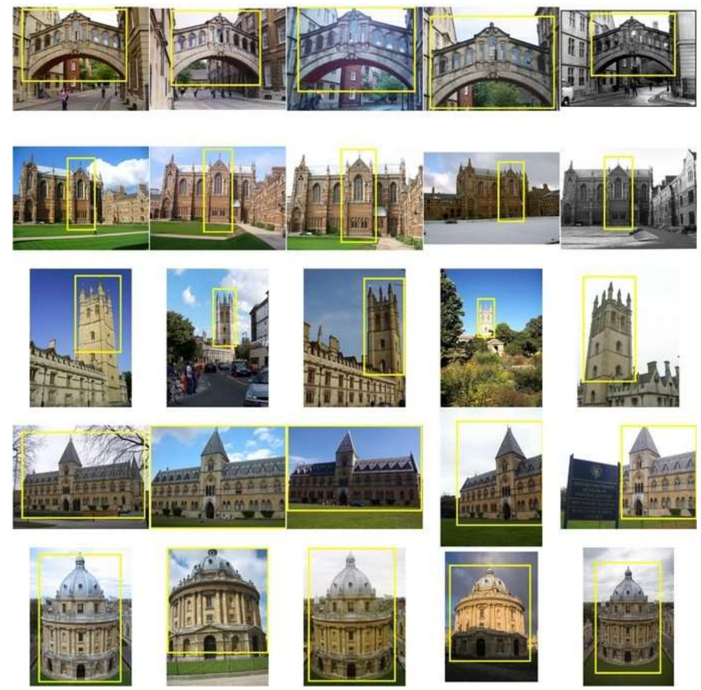
Hertford

Keble

Magdalen

Pitt Rivers

Radcliffe Camera



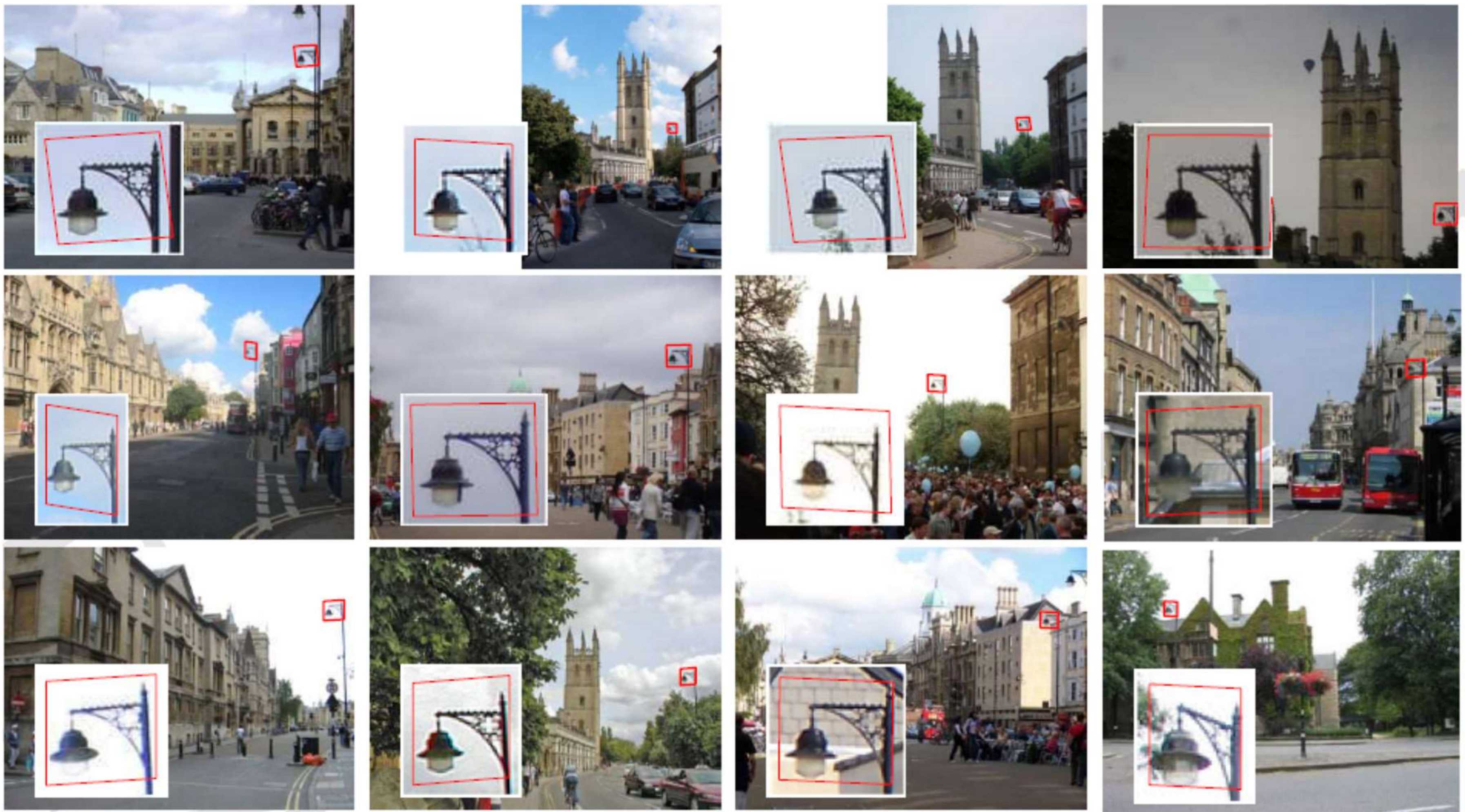
100 000 Images downloaded from FLICKR

Includes 11 Oxford Landmarks with manually labeled ground truth

Slide credit: Ondrej Chum

Results: Geometric Min-hash clustering

[Chum, Perdoch, Matas, CVPR 2009]

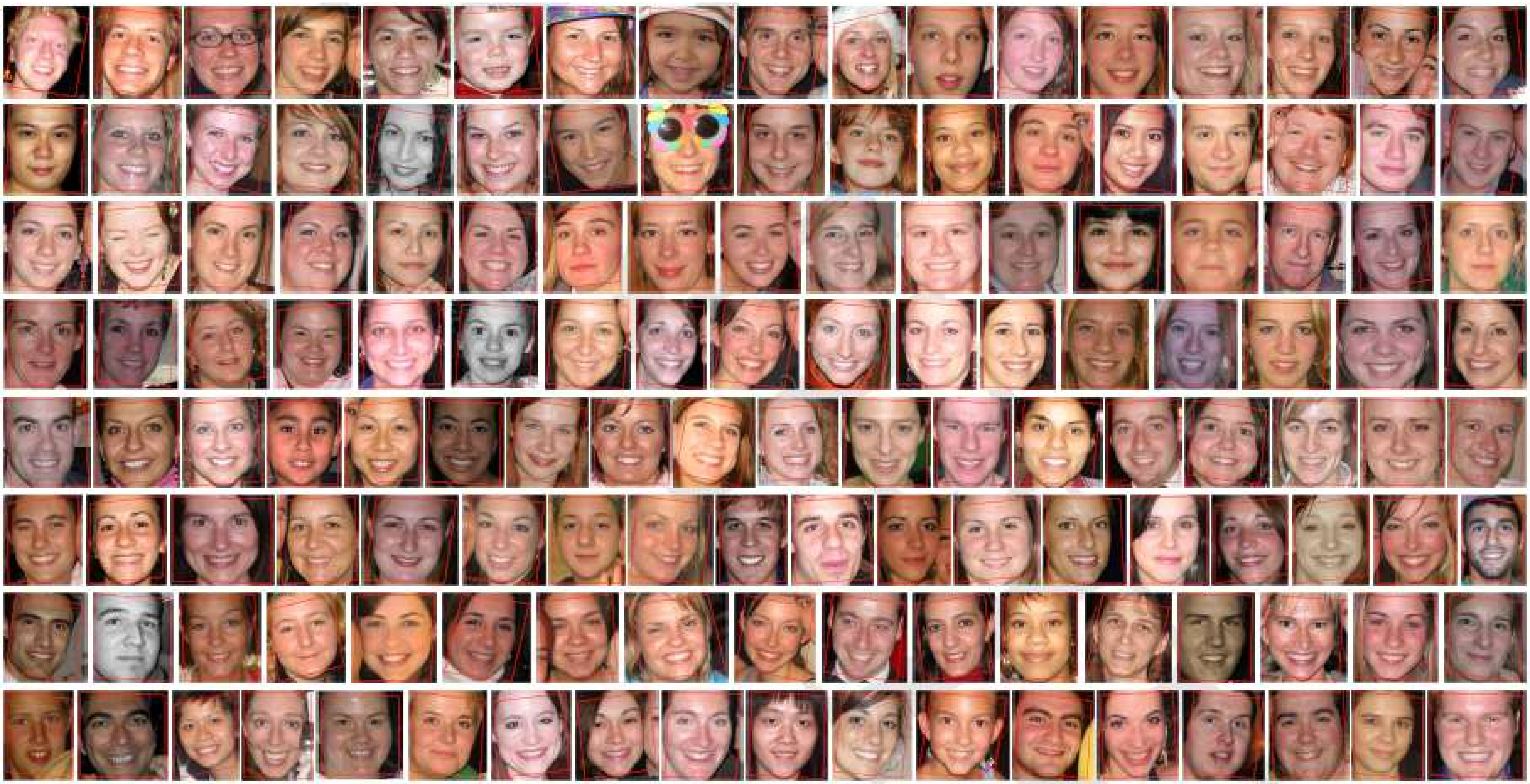


Discovering small objects

Slide credit: Ondrej Chum

Results: Geometric Min-hash clustering

[Chum, Perdoch, Matas, CVPR 2009]



Discovering small objects

Slide credit: Ondrej Chum

Mining for common visual patterns

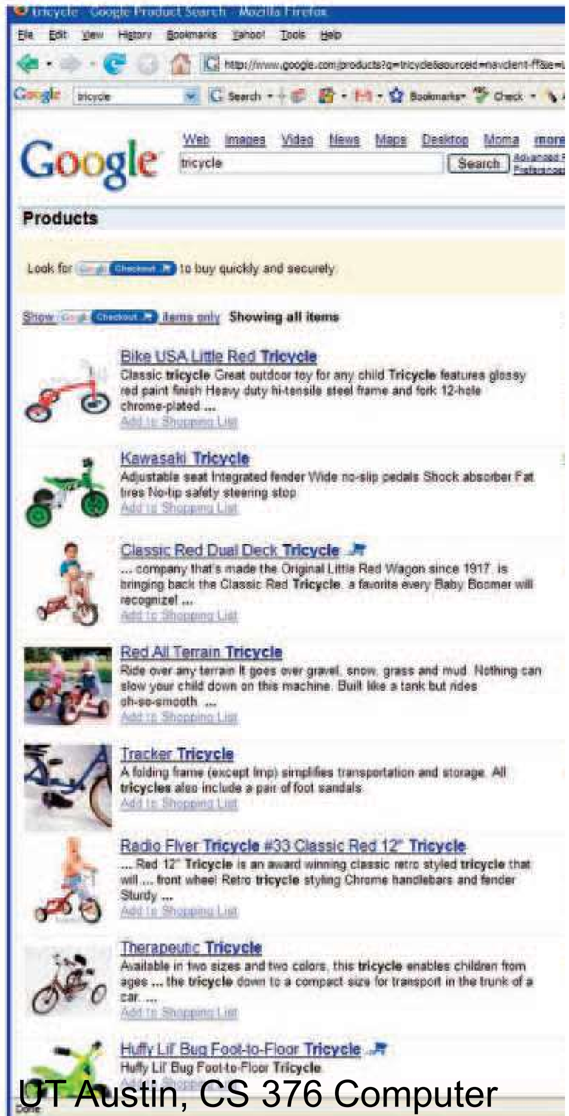
In addition to visual search, want to be able to **summarize, mine, and rank** the large collection as a whole.

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

Kristen Grauman

Visual Rank

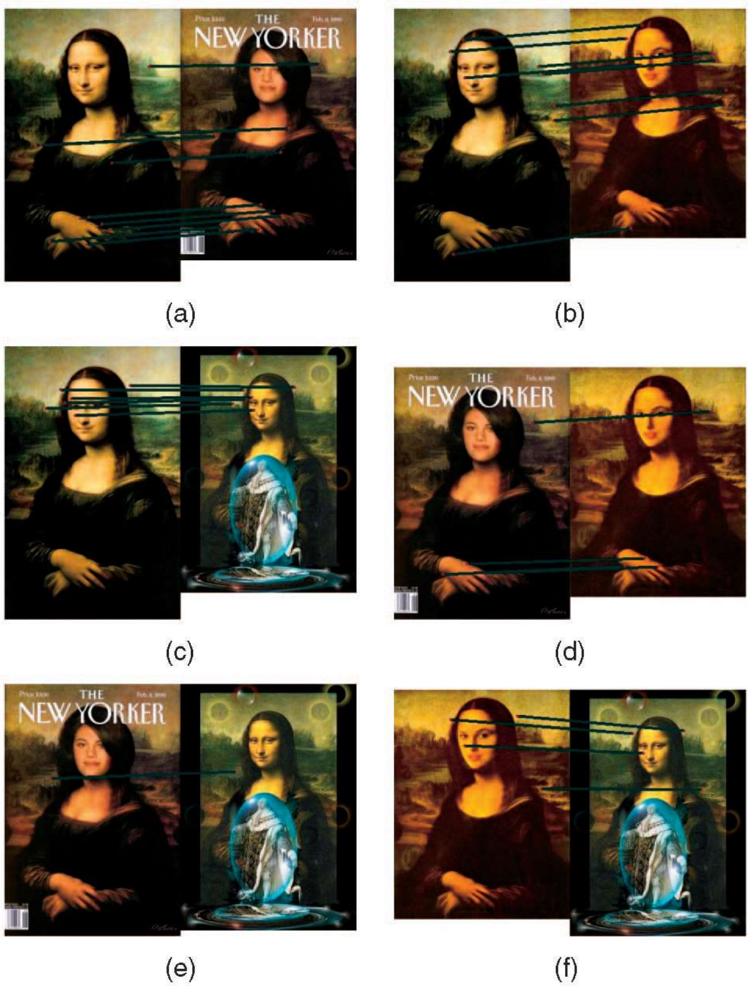
[Jing and Baluja, PAMI 2008]

- Compute relative “authority” of an image based on random walk principle.
  - Application of PageRank to visual data

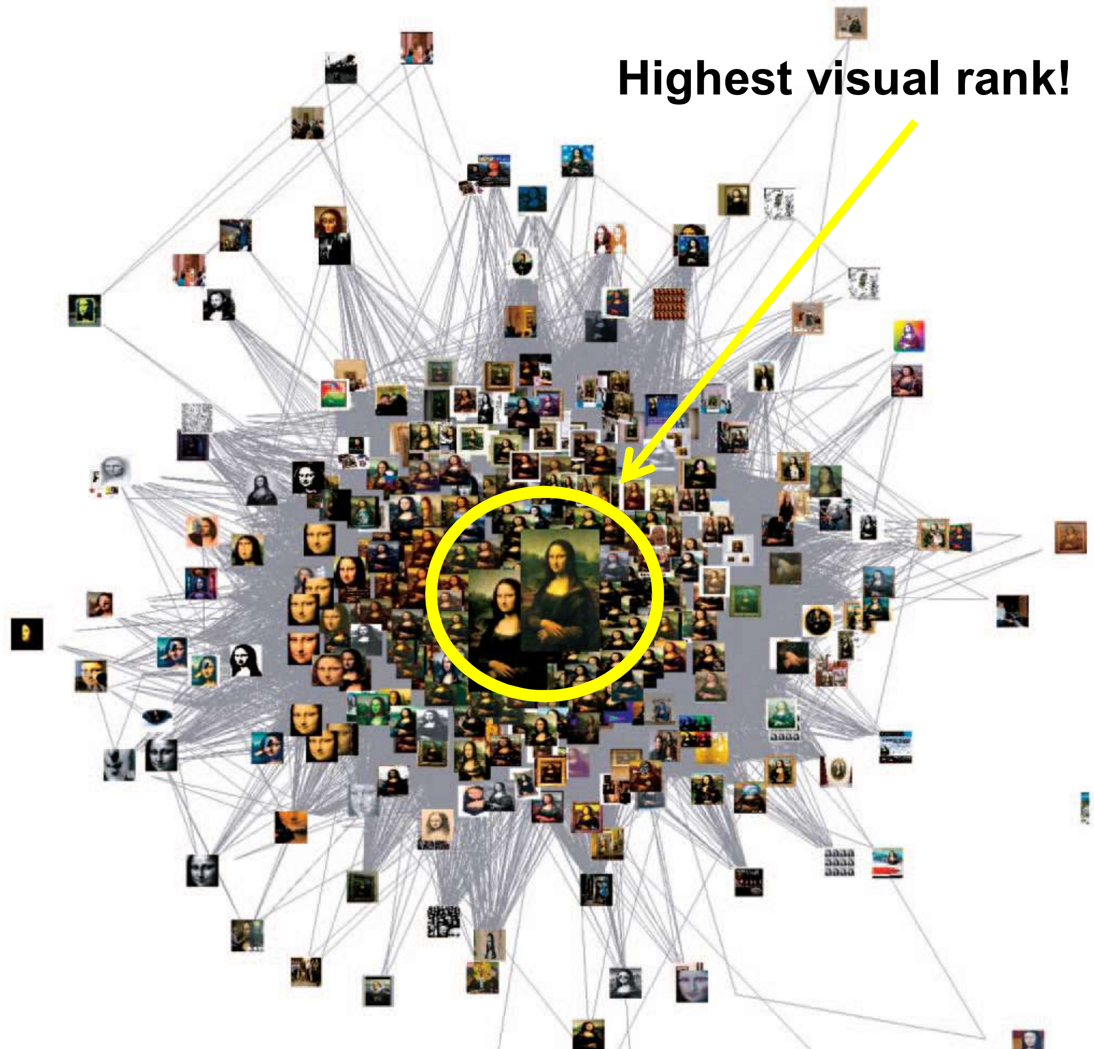
- **Main ideas**:
  - Graph weights = number of matched local features between two images
  - Exploit text search to narrow scope of each graph
  - Use LSH to make similarity computations efficient

Results: Visual Rank

[Jing and Baluja, PAMI 2008]



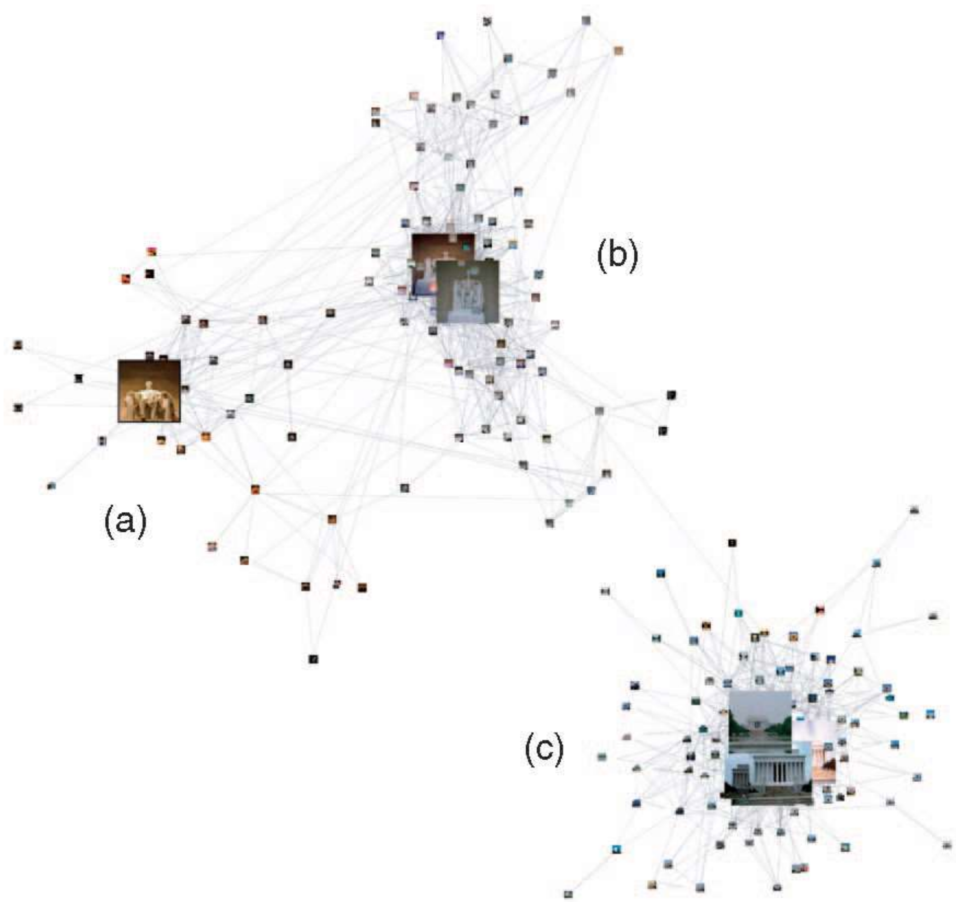
Original has more matches to rest



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".
Note the diversity of the high-ranked images.

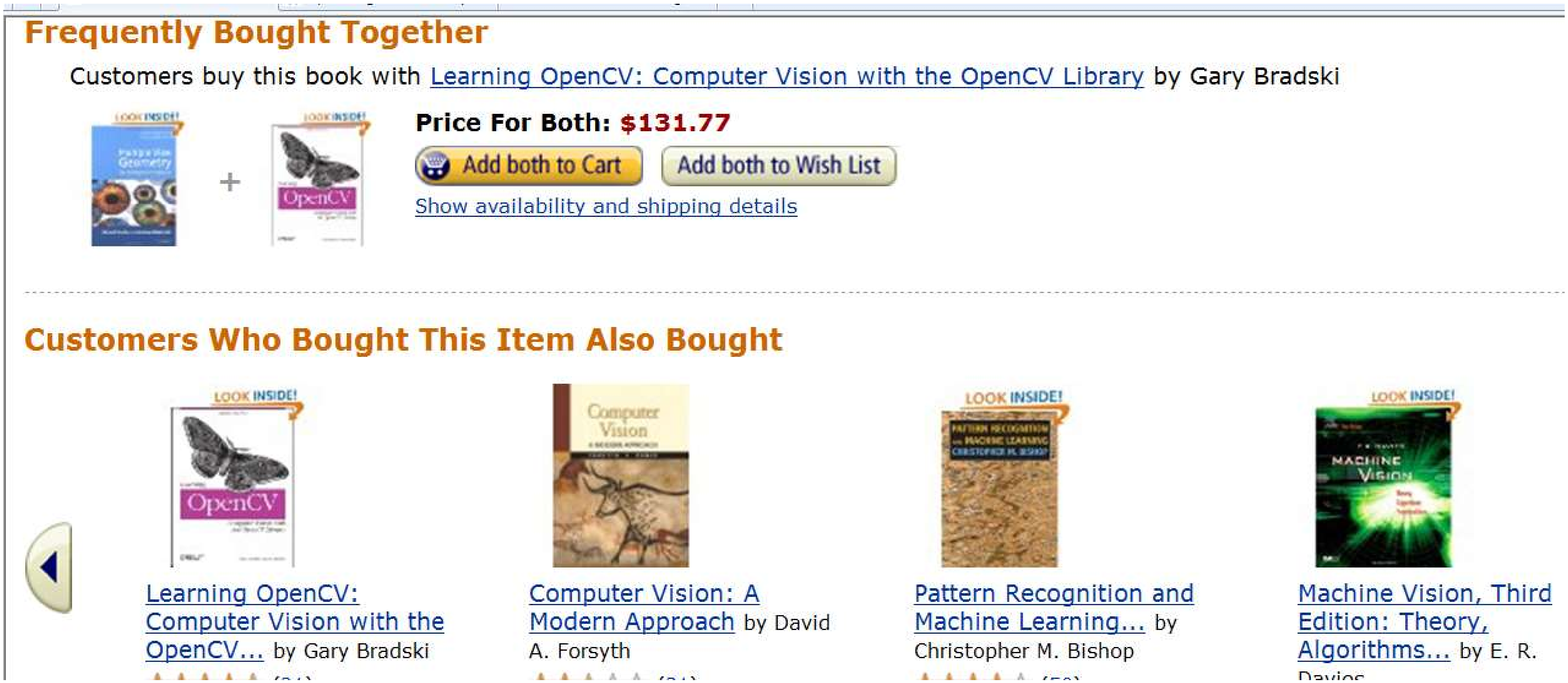
Mining for common visual patterns

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]

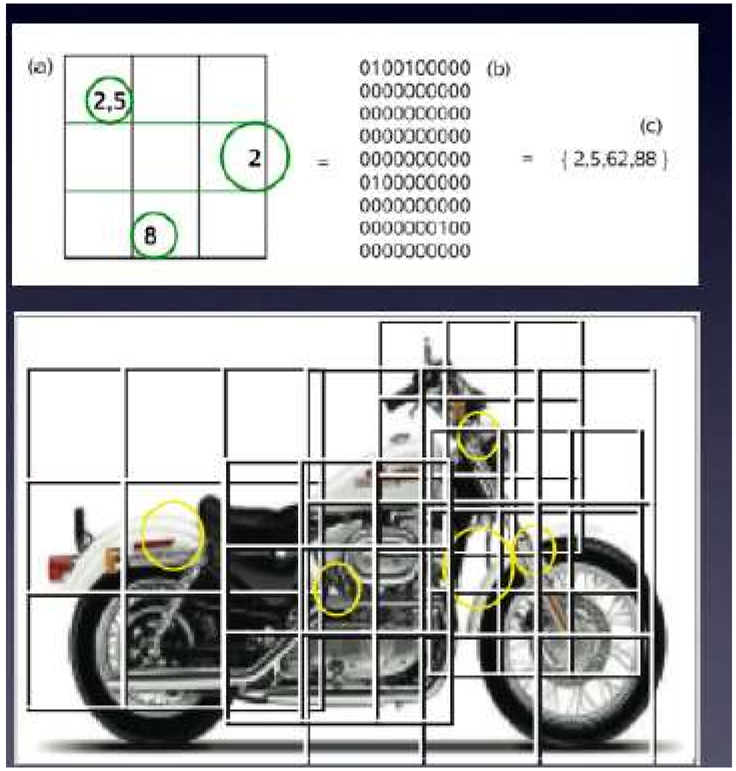
Frequent item-sets



Frequent item-set mining for spatial visual patterns

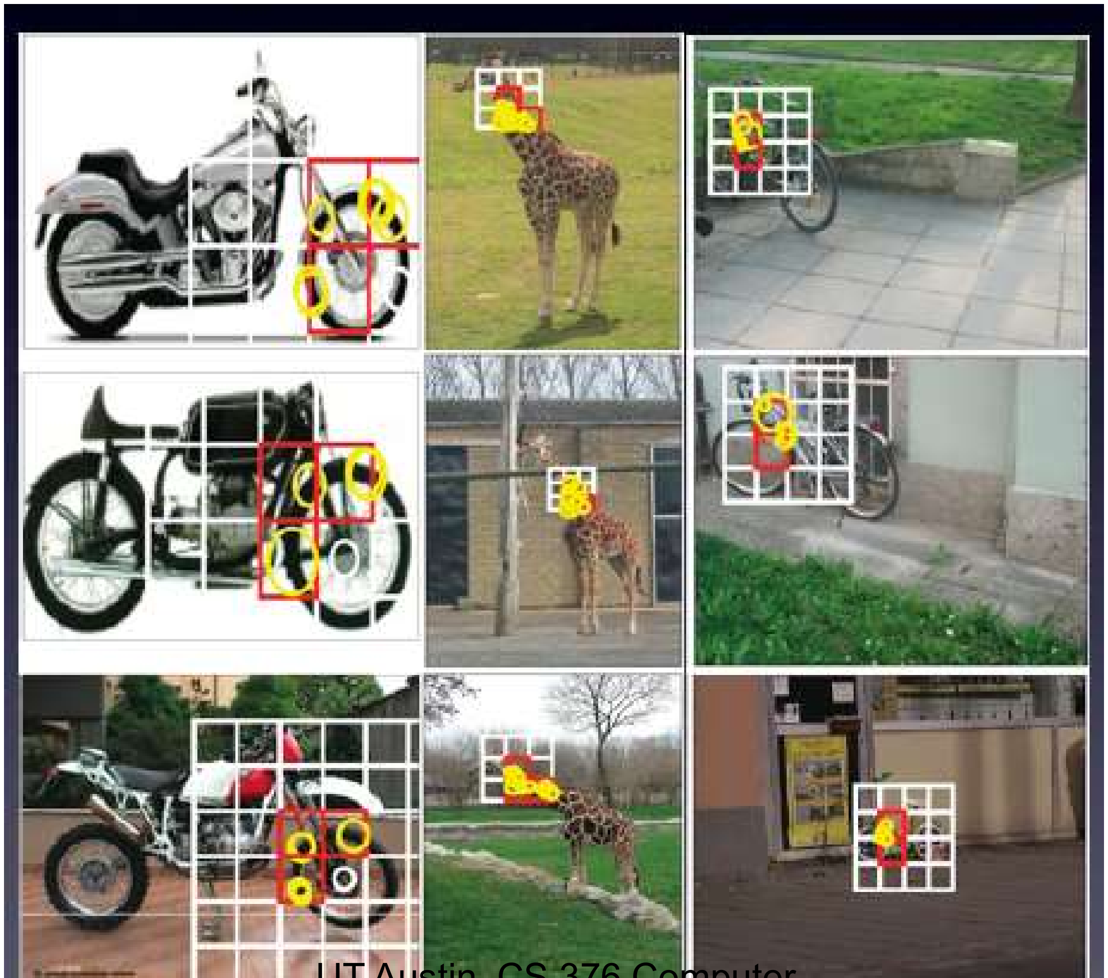
[Quack, Ferrari, Leibe, Van Gool, CIVR 2006, ICCV 2007]

- 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)



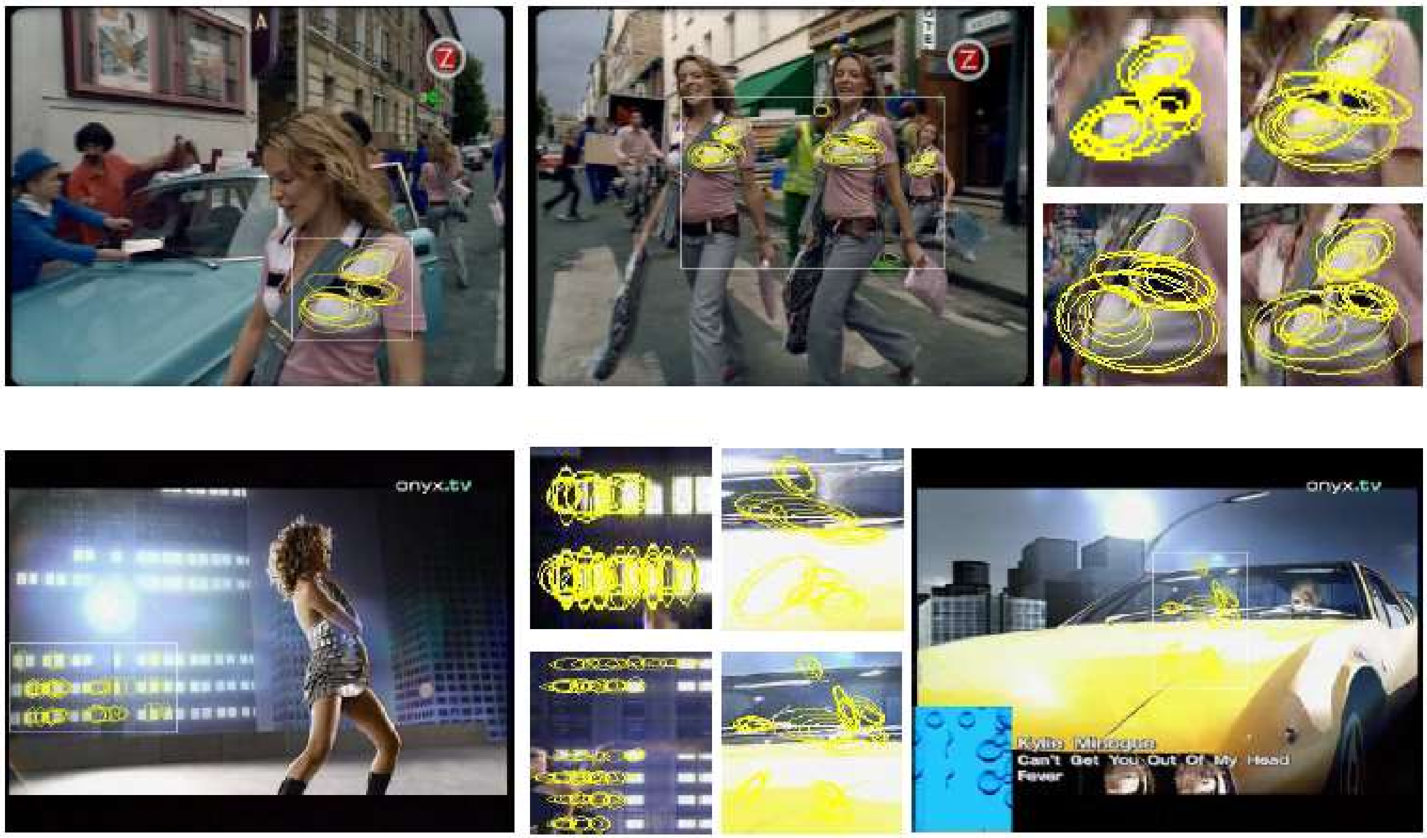
Frequent item-set mining for spatial visual patterns

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



Frequent item-set mining for spatial visual patterns

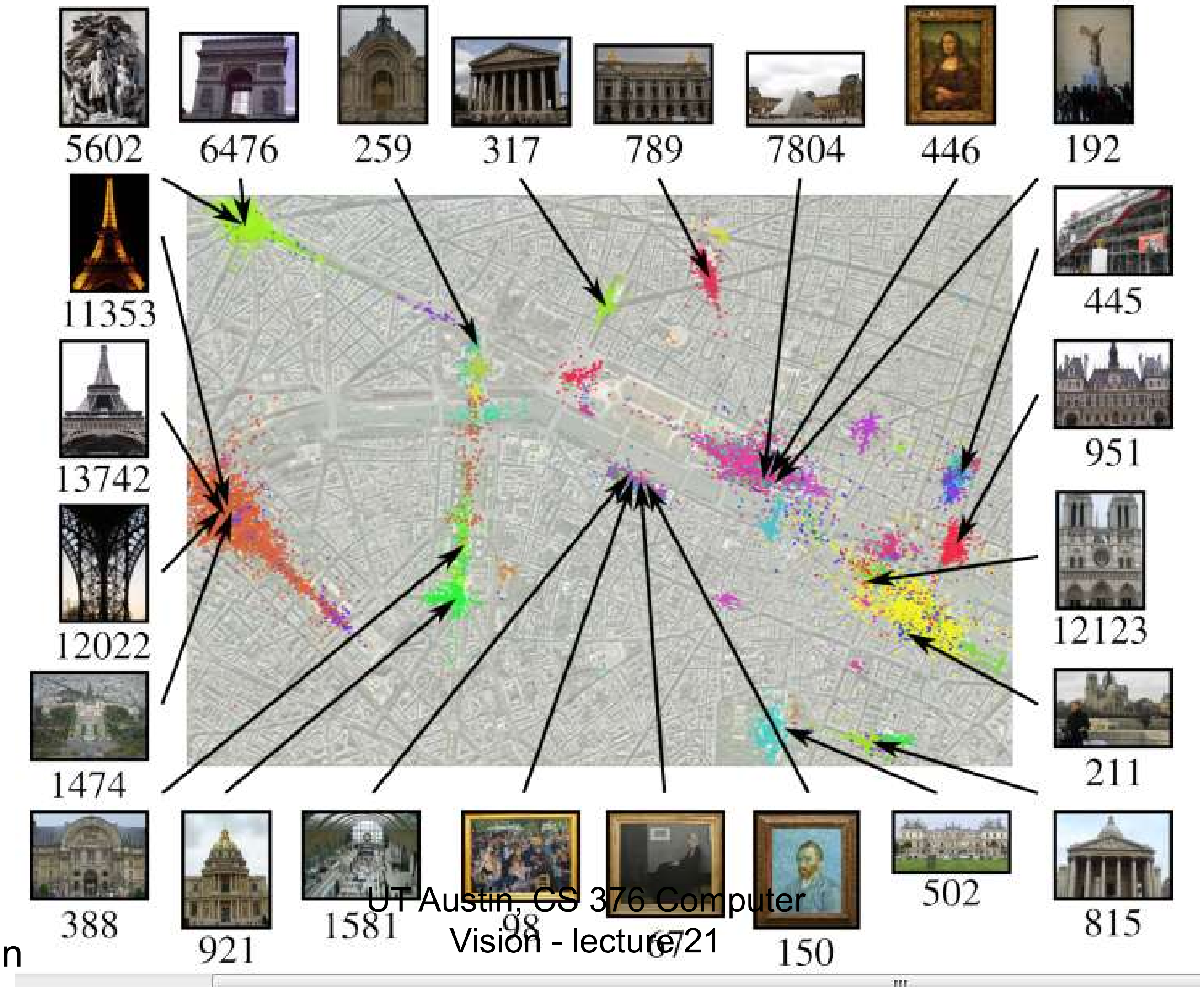
[Quack, Ferrari, Leibe, Van Gool, CIVR 2006, ICCV 2007]



Two example itemset clusters

Discovering favorite views

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



Kristen Grauman

Today

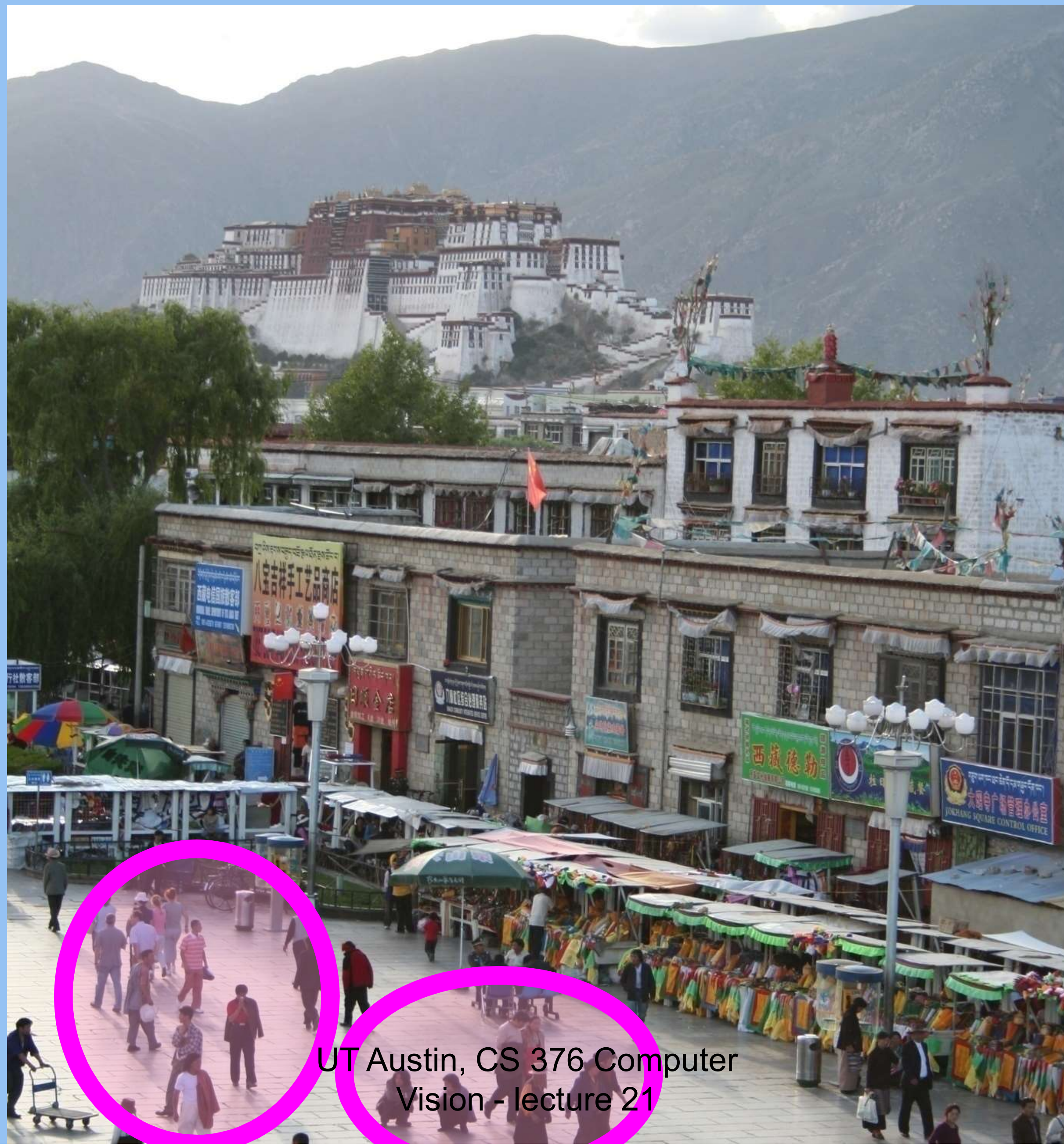
- Discovering visual patterns
  - Randomized hashing algorithms
  - Mining large-scale image collections

- **Introduction to visual categorization**

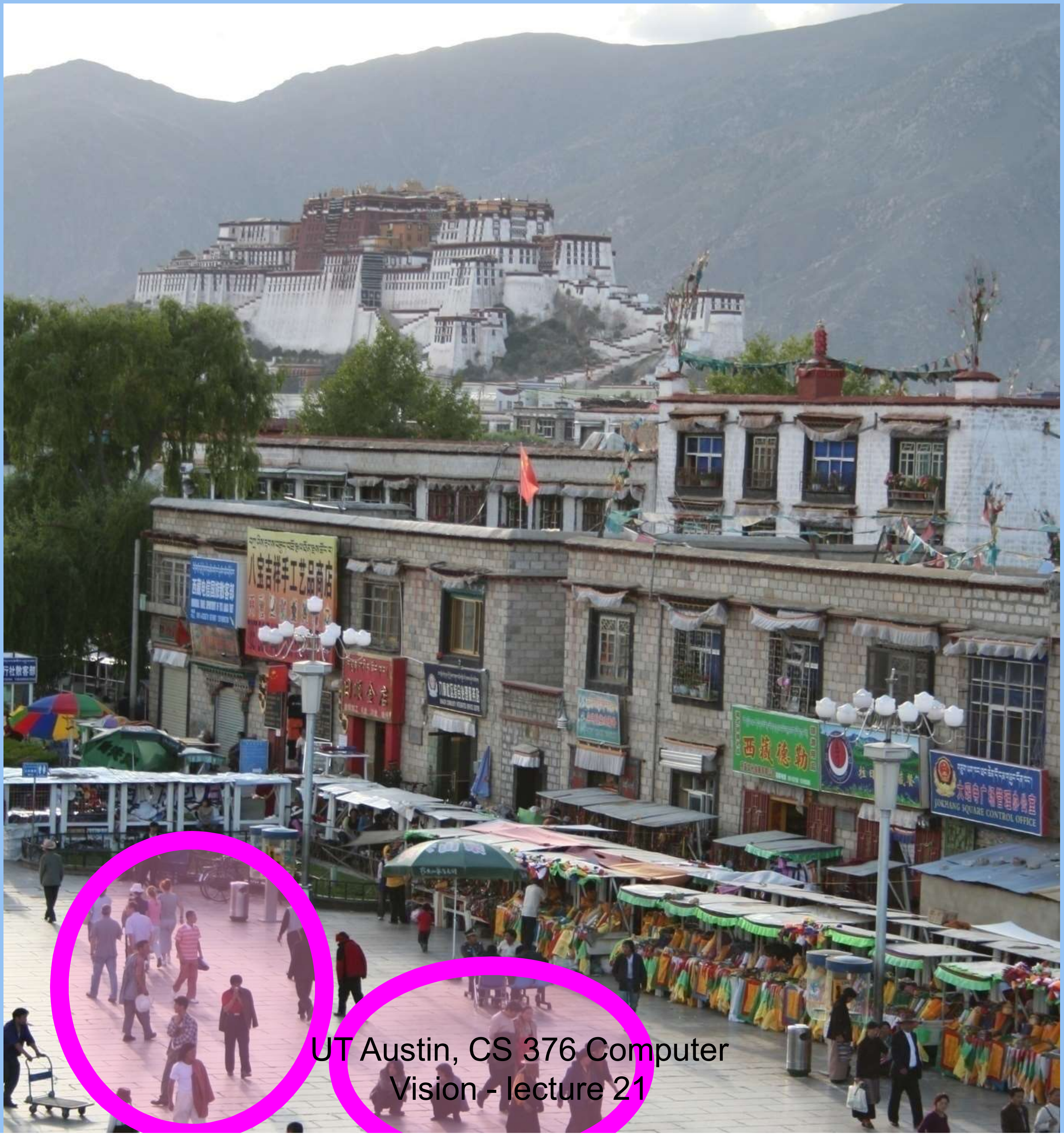
What does recognition involve?

Fei-Fei Li

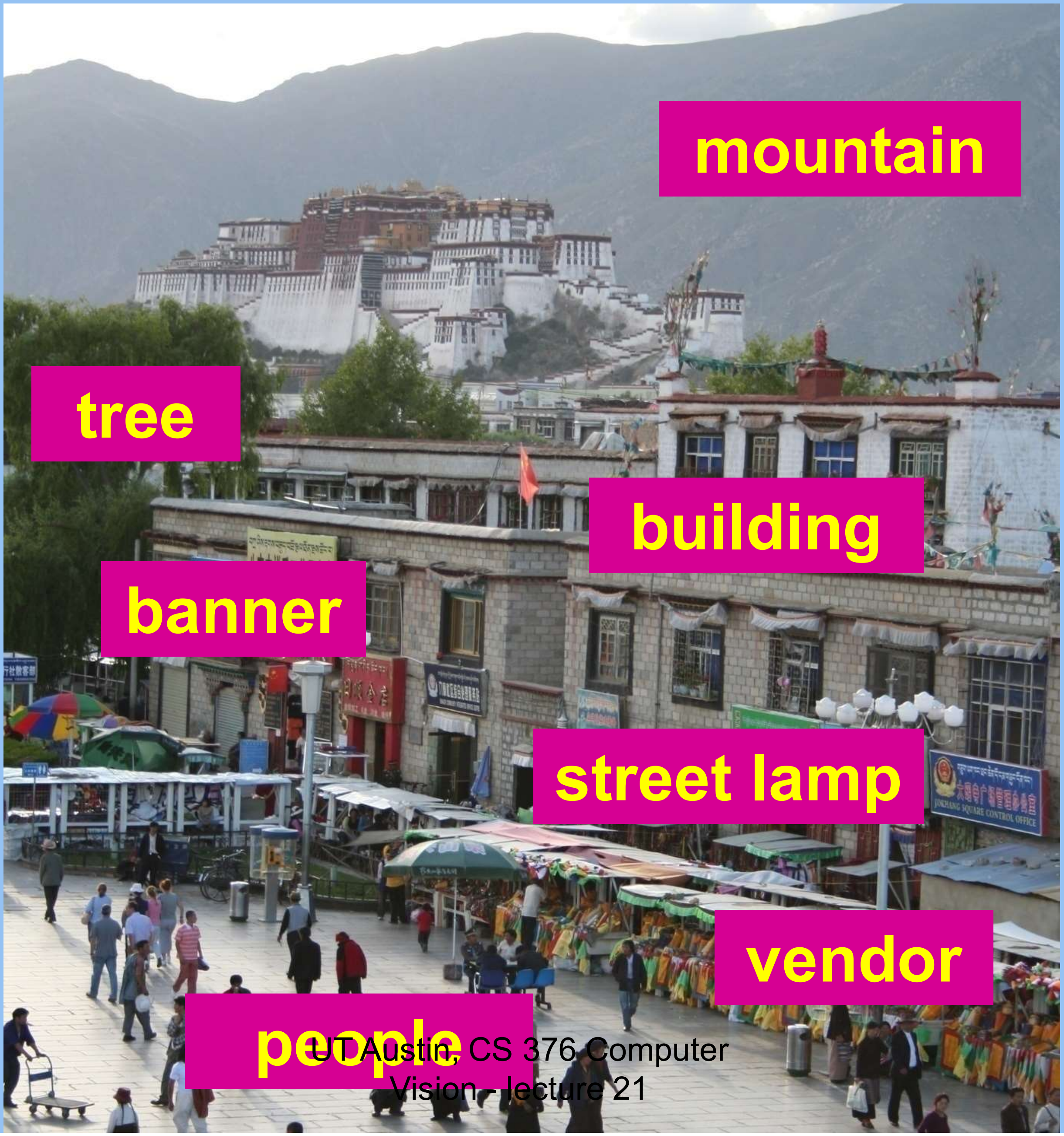
Detection: are there people?



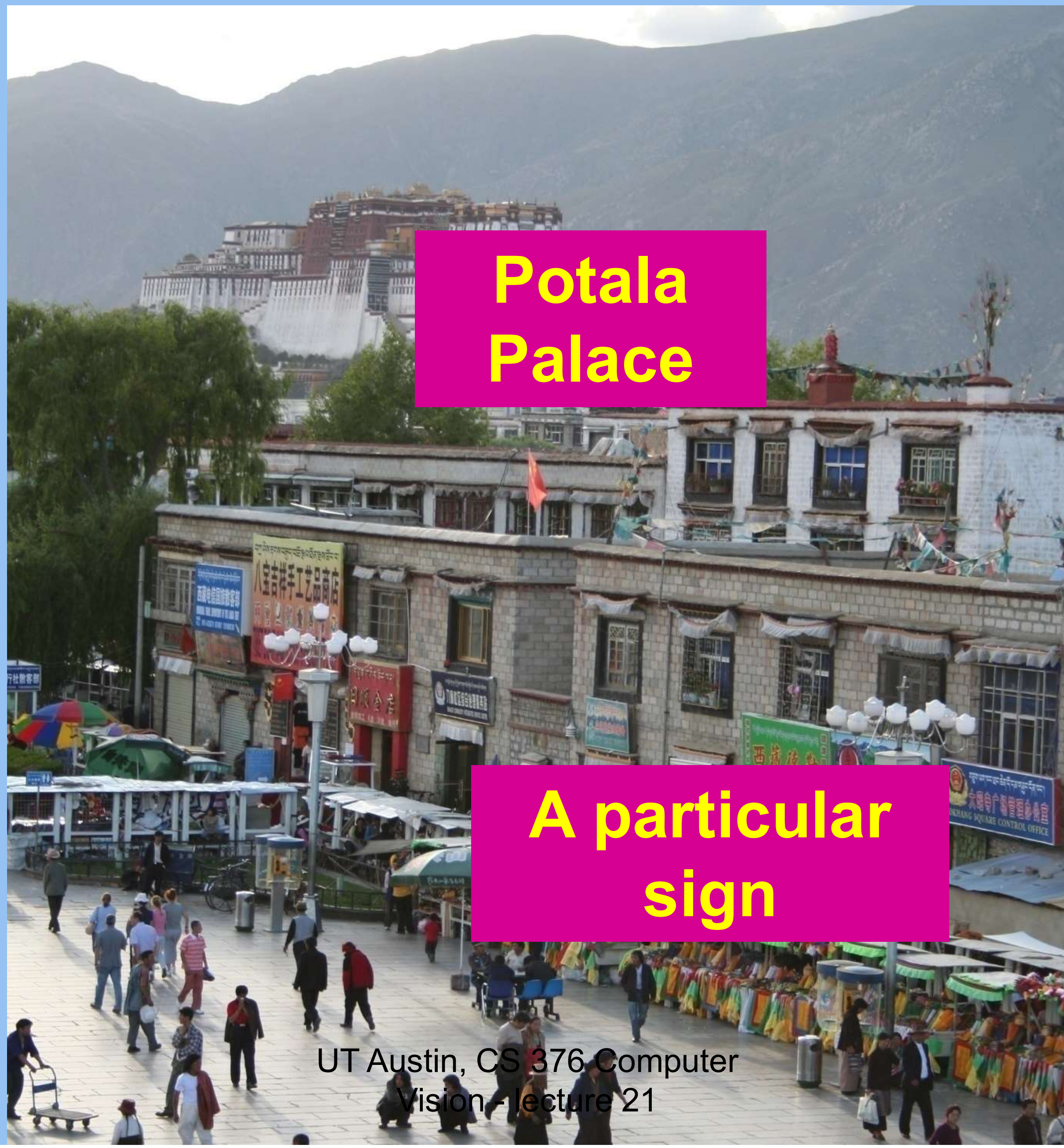
Activity: What are they doing?



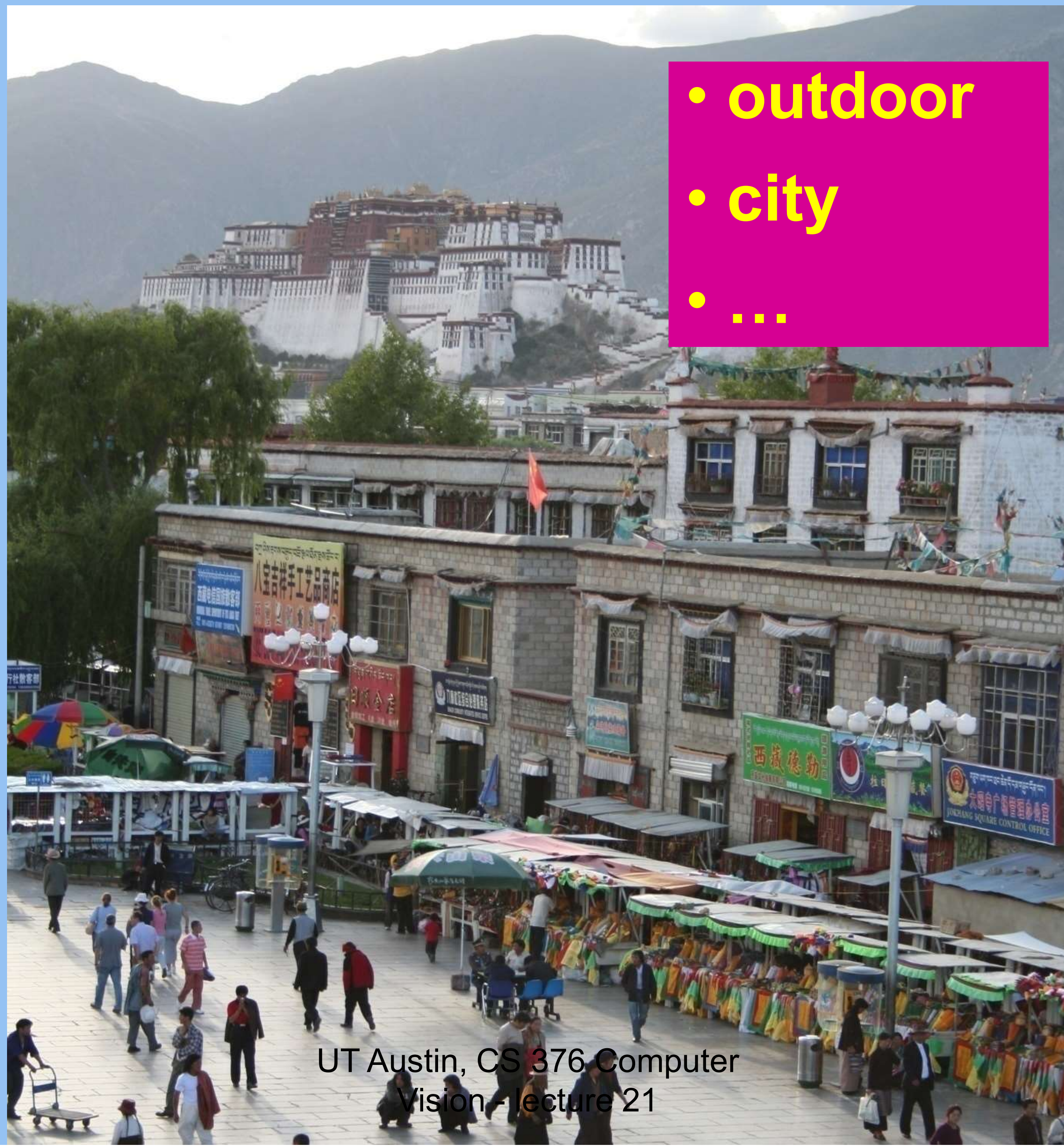
Object categorization



Instance recognition

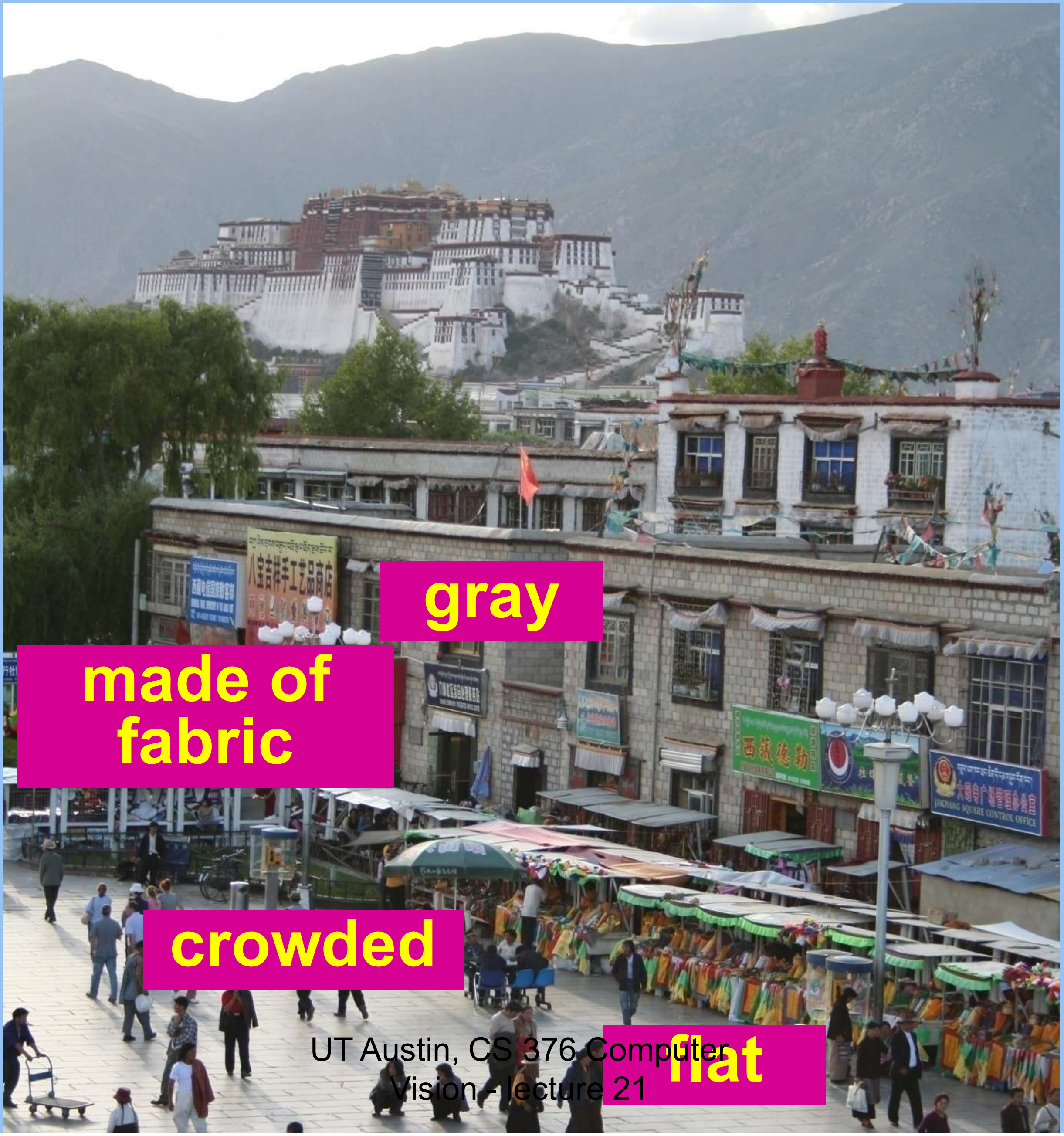


Scene and context categorization



- **outdoor**
- **city**
- **…**

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?**

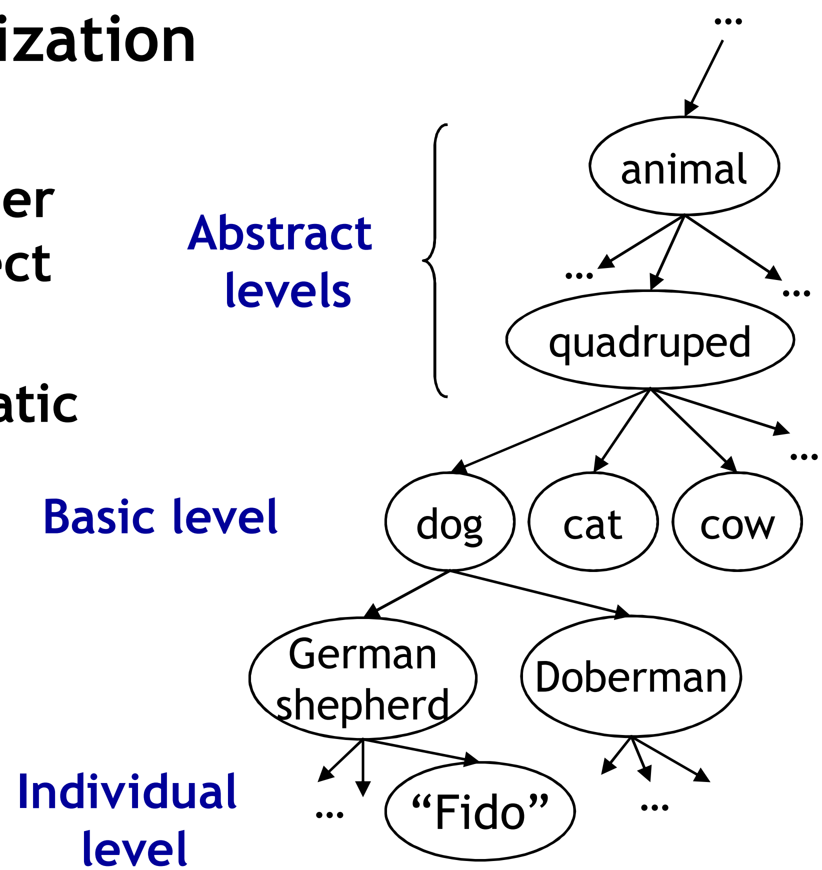
"Fido" → German shepherd → dog → animal → living being

Visual Object Categories

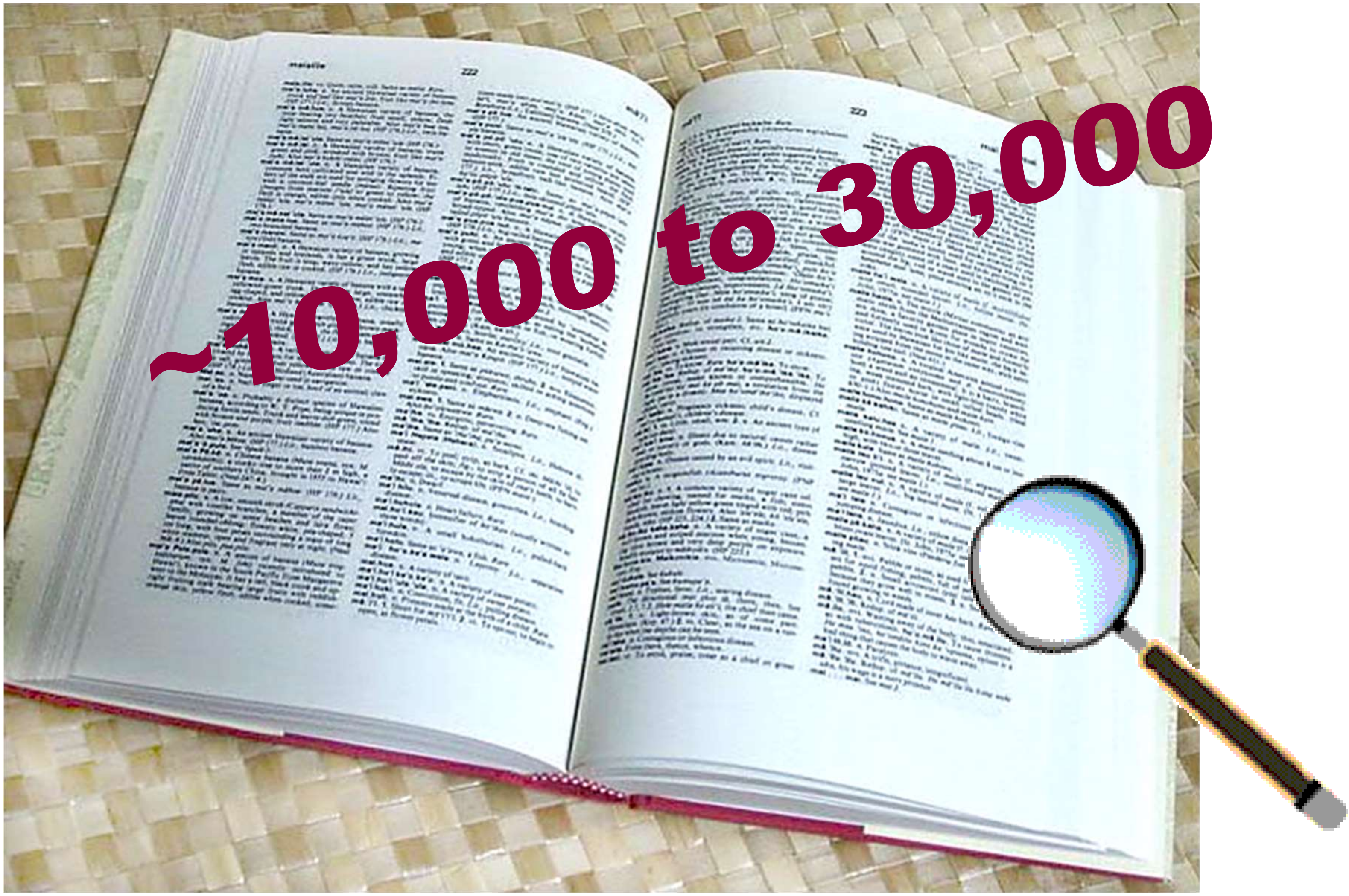
- **Basic Level Categories in human categorization [Rosch 76, Lakoff 87]**
  - The highest level at which category members have similar perceived shape
  - The highest level at which a single mental image reflects the entire category
  - The level at which human subjects are usually fastest at identifying 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?



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

Biederman 1987


~10,000 to 30,000

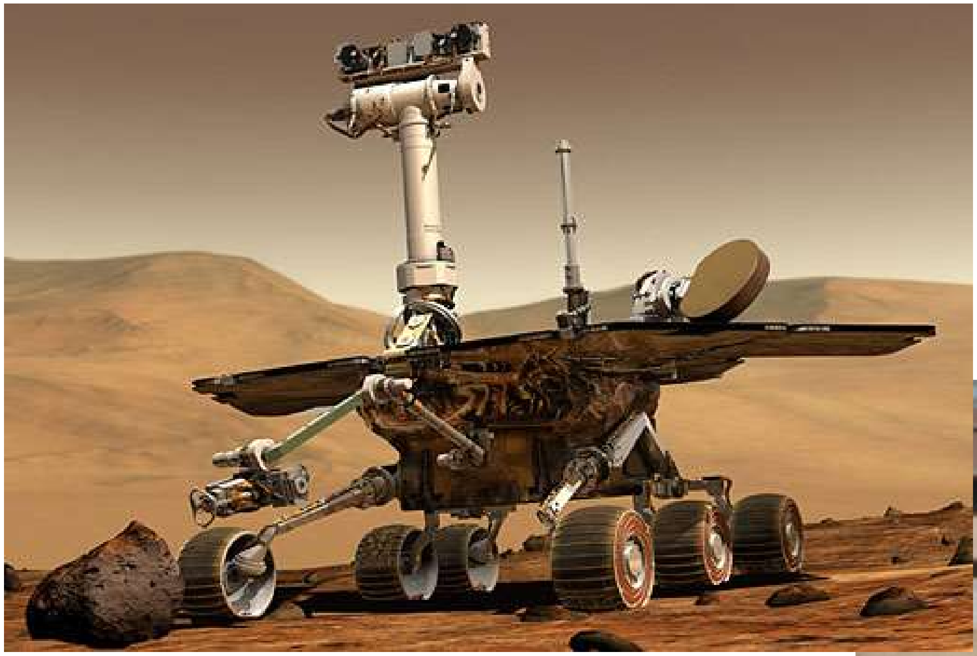
Other Types of Categories

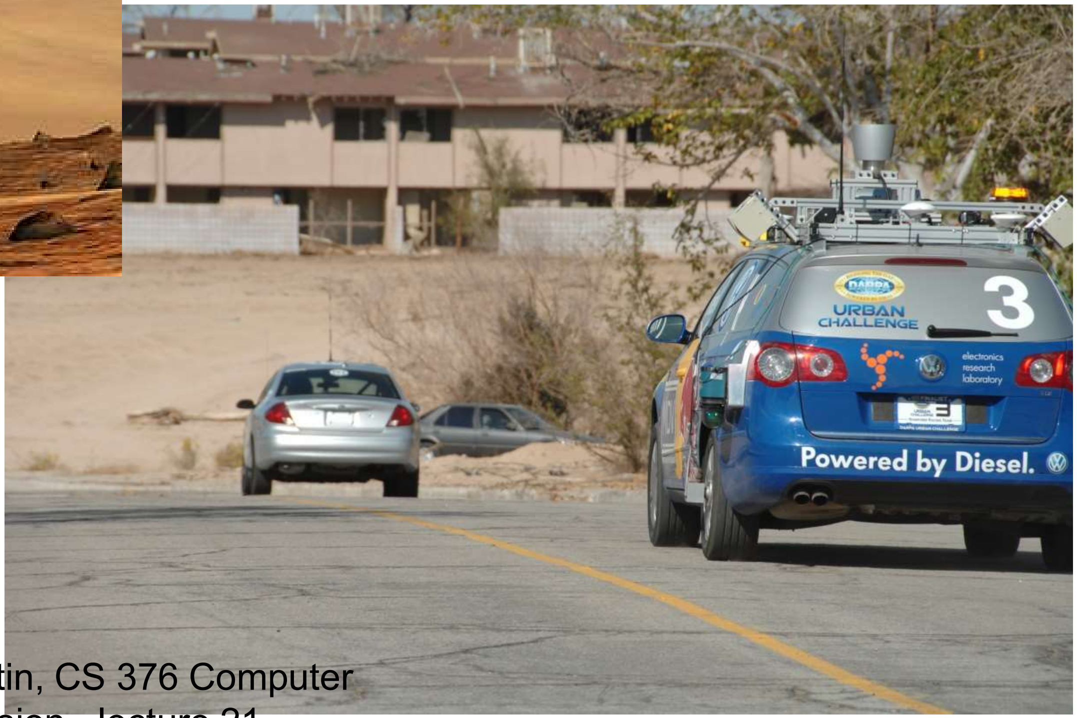
- **Functional Categories**
  - **e.g. chairs = *"something you can sit on"***

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

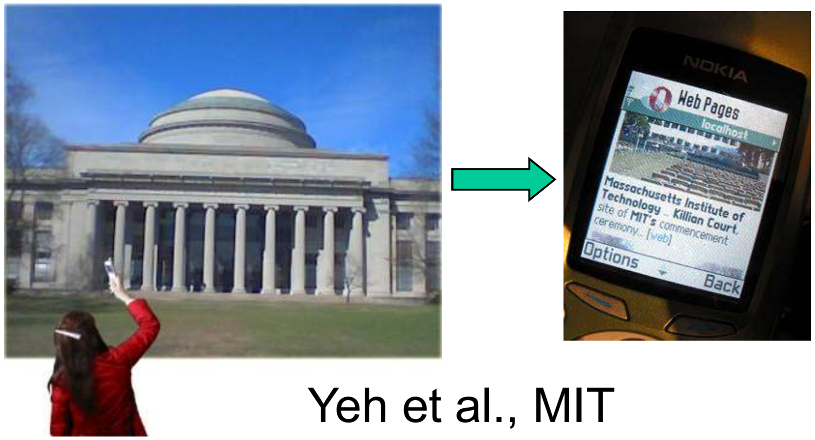
Autonomous agents able to detect objects



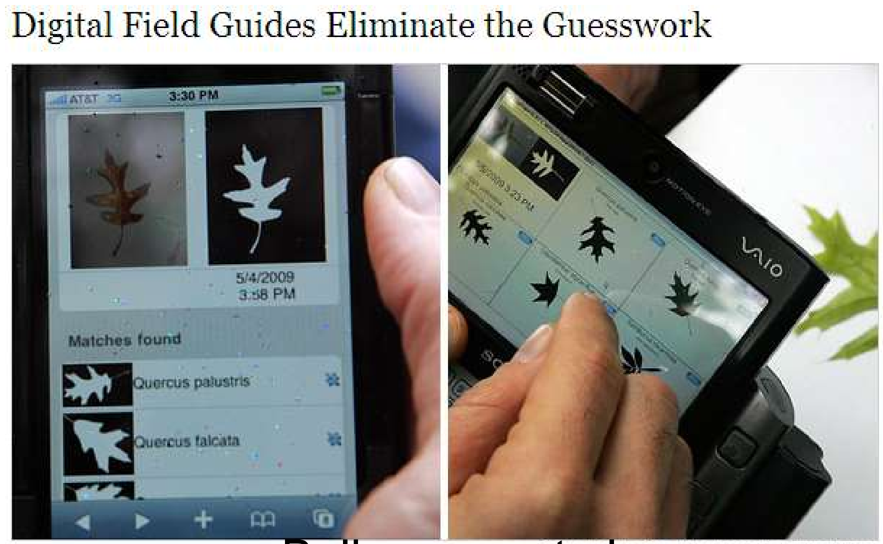


http://www.darpa.mil/grandchallenge/gallery.asp

Posing visual queries



Yeh et al., MIT

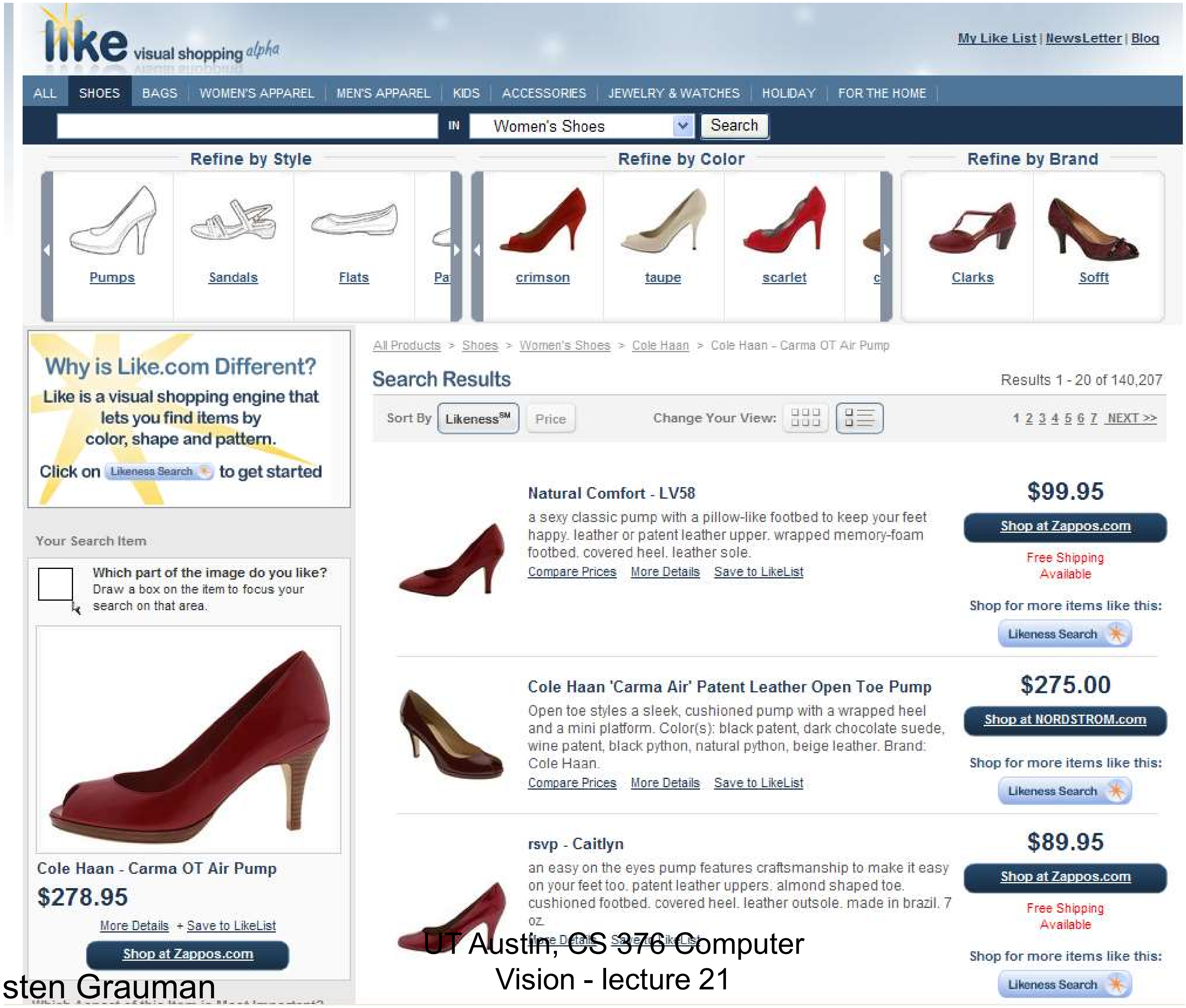


Belhumeur et al.

Kooaba, Bay & Quack et al.

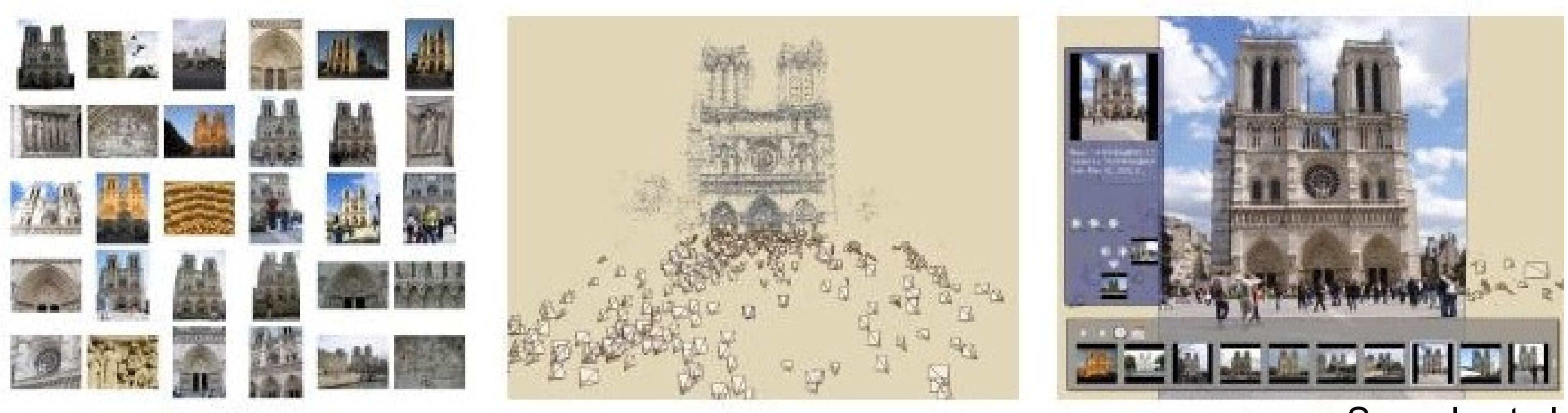
Slide: Kristen Grauman

Finding visually similar objects



Slide: Kristen Grauman

Exploring community photo collections



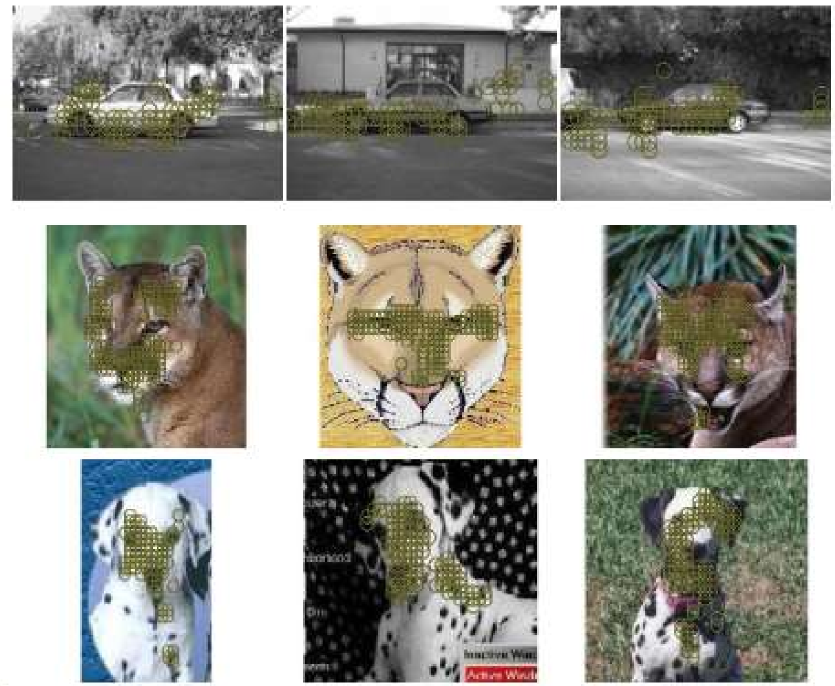
Snavely et al.

Slide: Kristen Grauman

Simon & Seitz

Discovering visual patterns

Objects Sivic & Zisserman



Lee & Grauman
Categories

Wang et al.
Actions

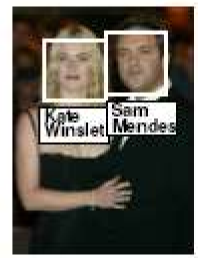
Slide: Kristen Grauman

Auto-annotation

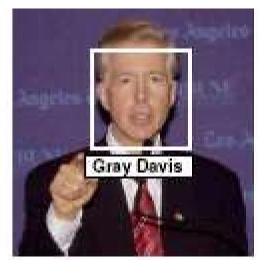
Figure 9. Results of automatic object-level annotation with bounding boxes. Groundtruth annotation is shown with dashed lines, correct detection with solid green lines, false detections with solid red lines. Auto-annotation with related Wikipedia articles is also shown. All results are also labeled with 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 films 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.

Challenges: robustness

Illumination

Object pose

Clutter

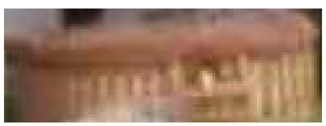
Occlusions

Intra-class appearance

Viewpoint

Slide: Kristen Grauman

Challenges:
context and human experience



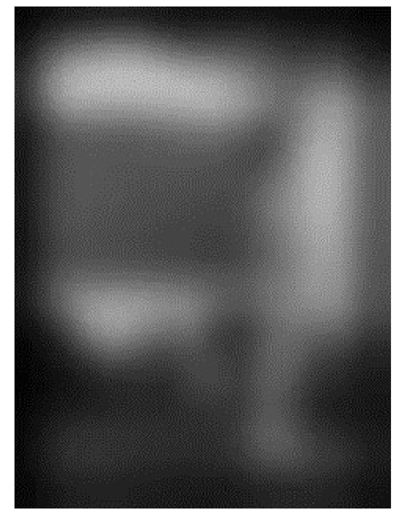
Context cues

Slide: Kristen Grauman

Challenges: context and human experience

Context cues

Function



Dynamics

Slide: Kristen Grauman

Video credit: J. Davis

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]

Slide: Kristen Grauman

Challenges: learning with minimal supervision

Unlabeled, multiple objects

Classes labeled, some clutter

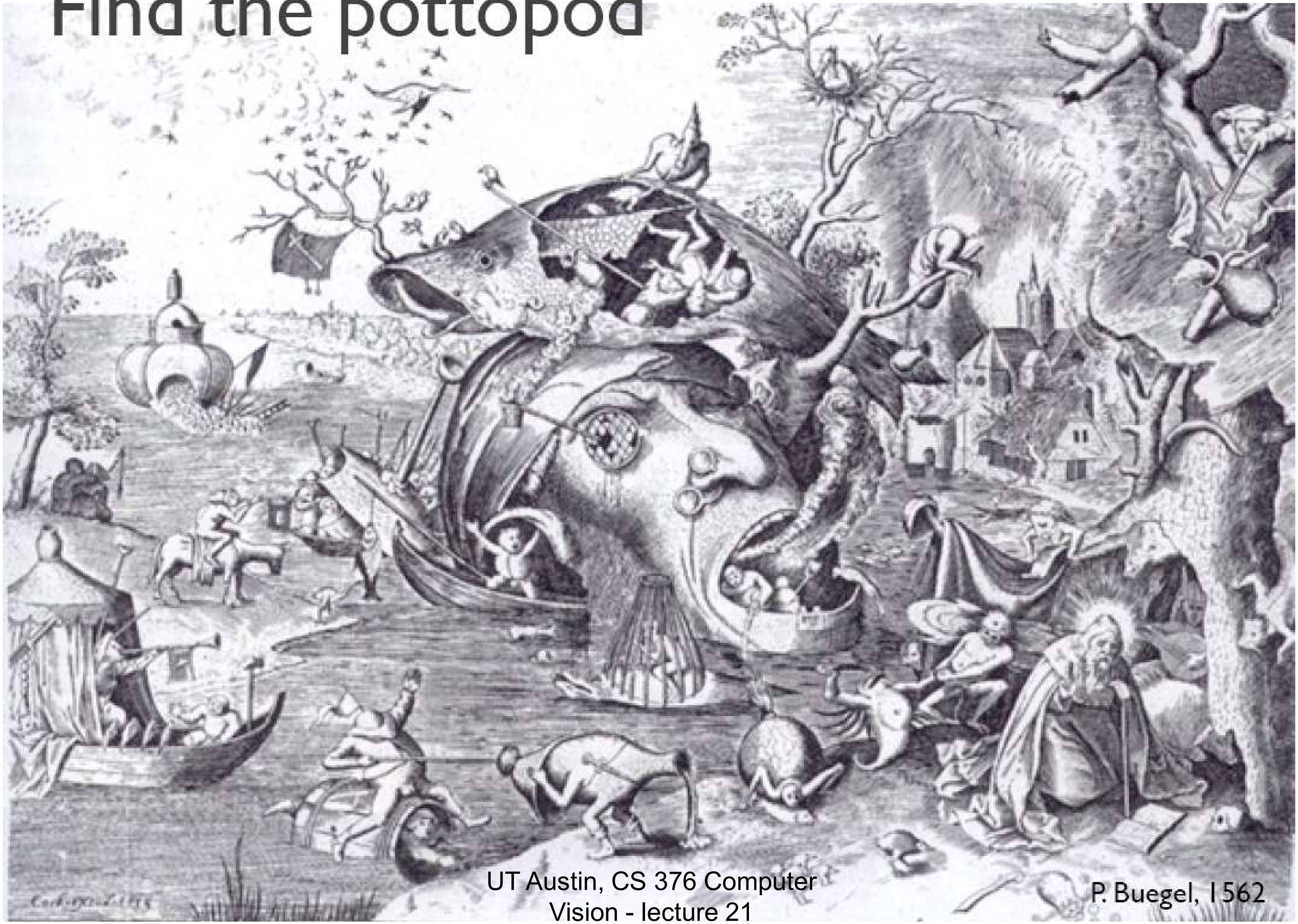
Cropped to object, parts and classes labeled

Slide: Kristen Grauman

This is a pottopod

Slide from Pietro Perona, 2004 Object Recognition workshop

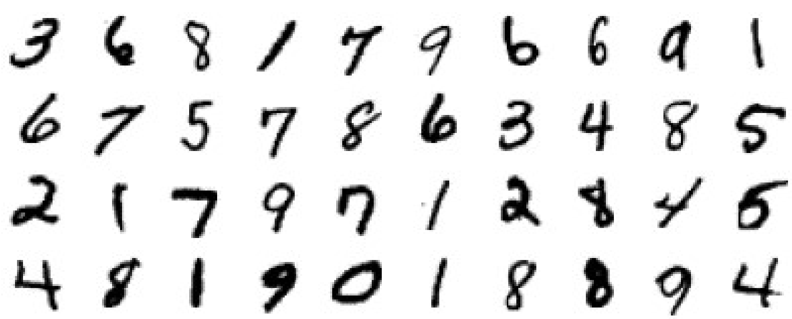
Find the pottopod



P. Buegel, 1562

Slide from Pietro Perona, 2004 Object Recognition workshop

What kinds of things work best today?



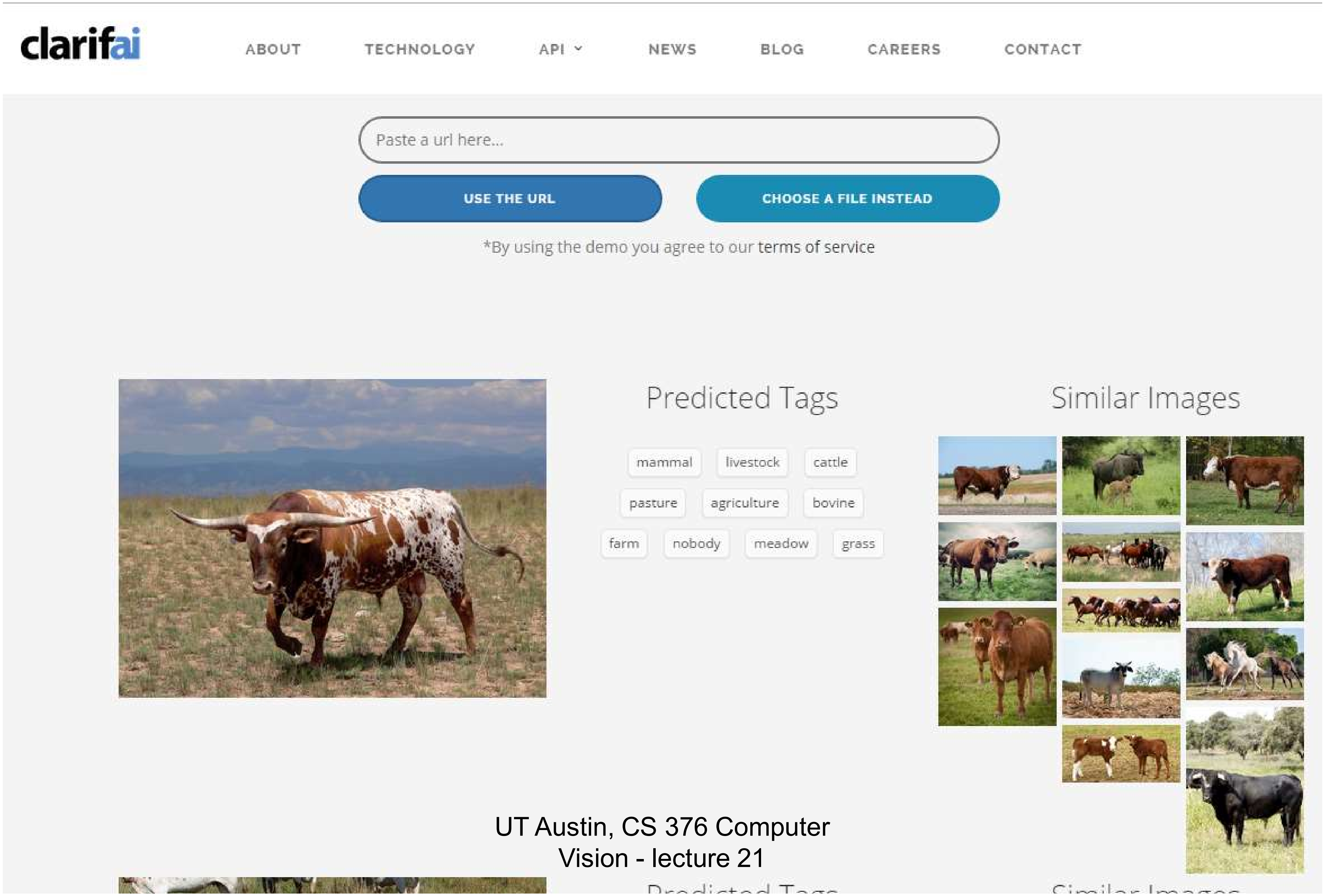
Reading license plates, zip codes, checks

Frontal face detection

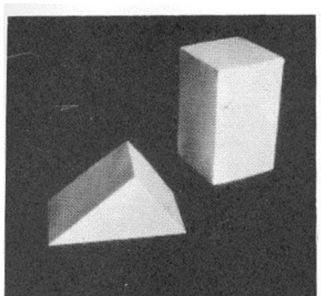
Recognizing flat, textured objects (like books, CD covers, posters)

Fingerprint recognition

What kinds of things work best today?

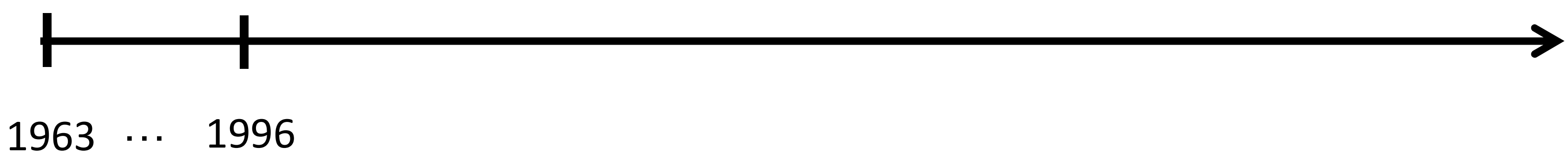


Progress charted by datasets



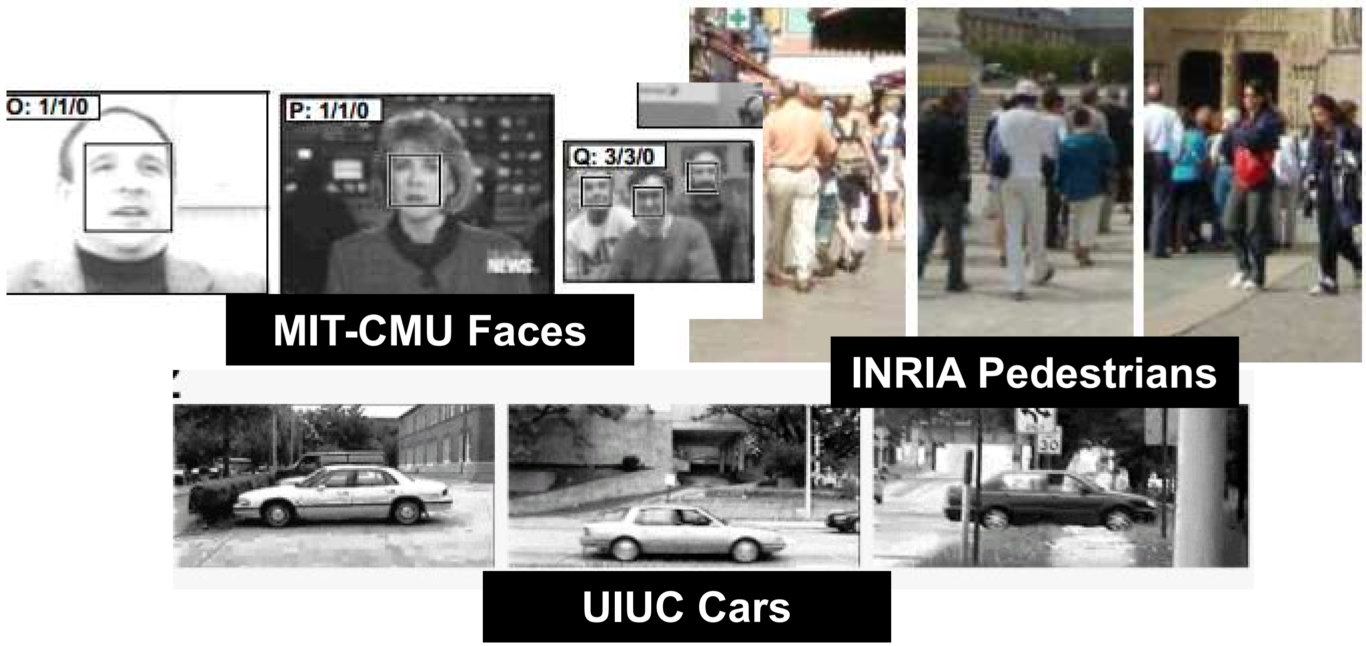
Roberts 1963

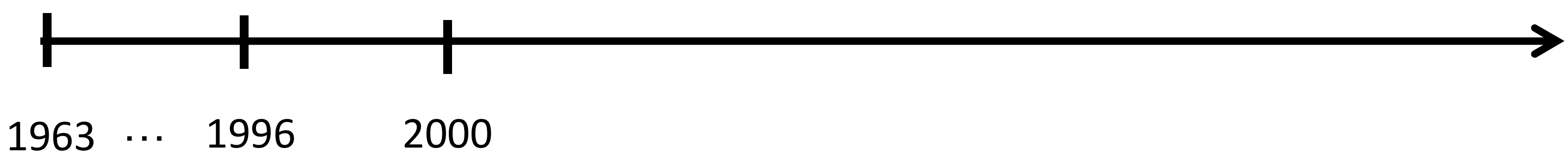
COIL



Slide: Kristen Grauman

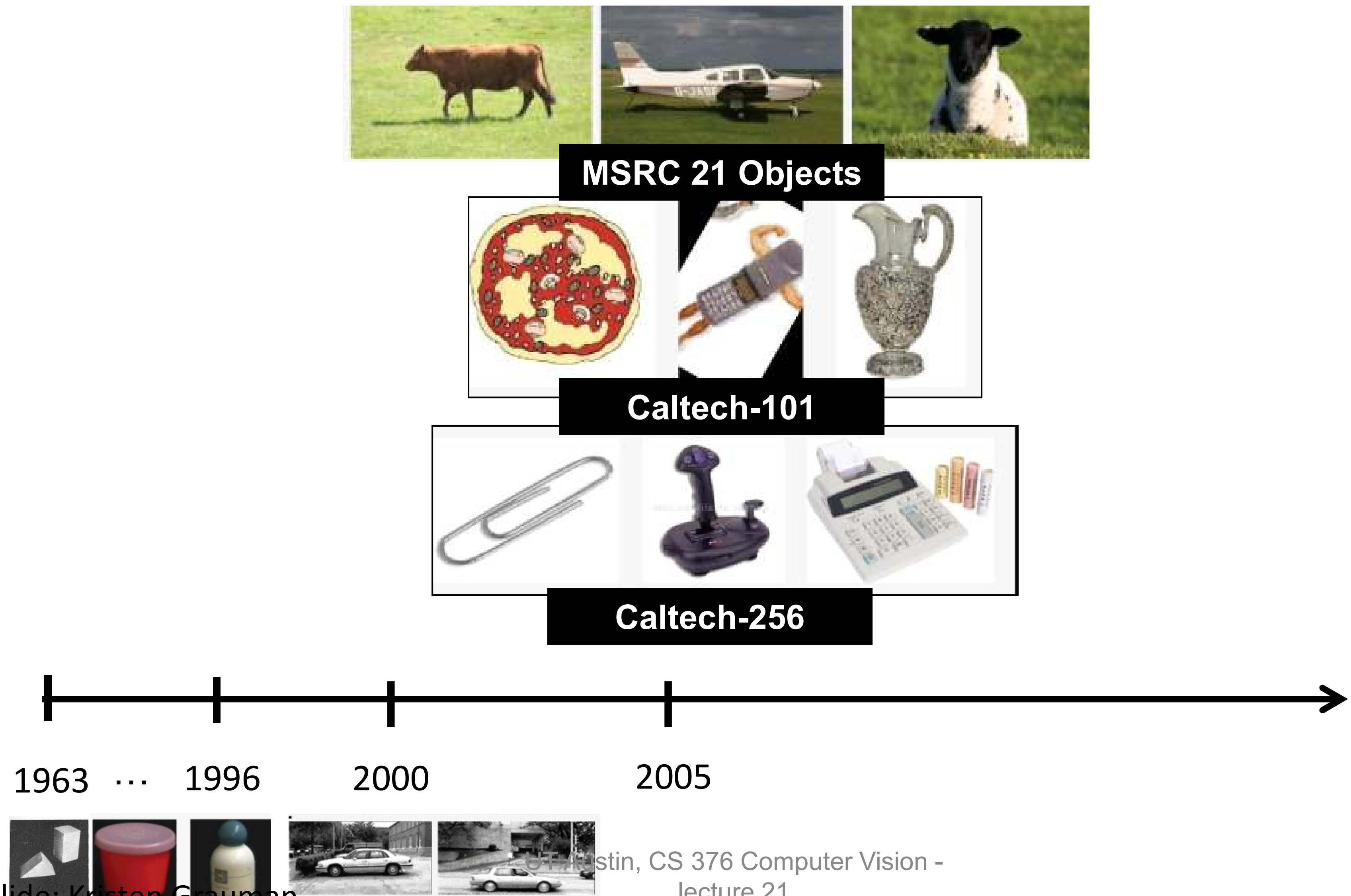
Progress charted by datasets





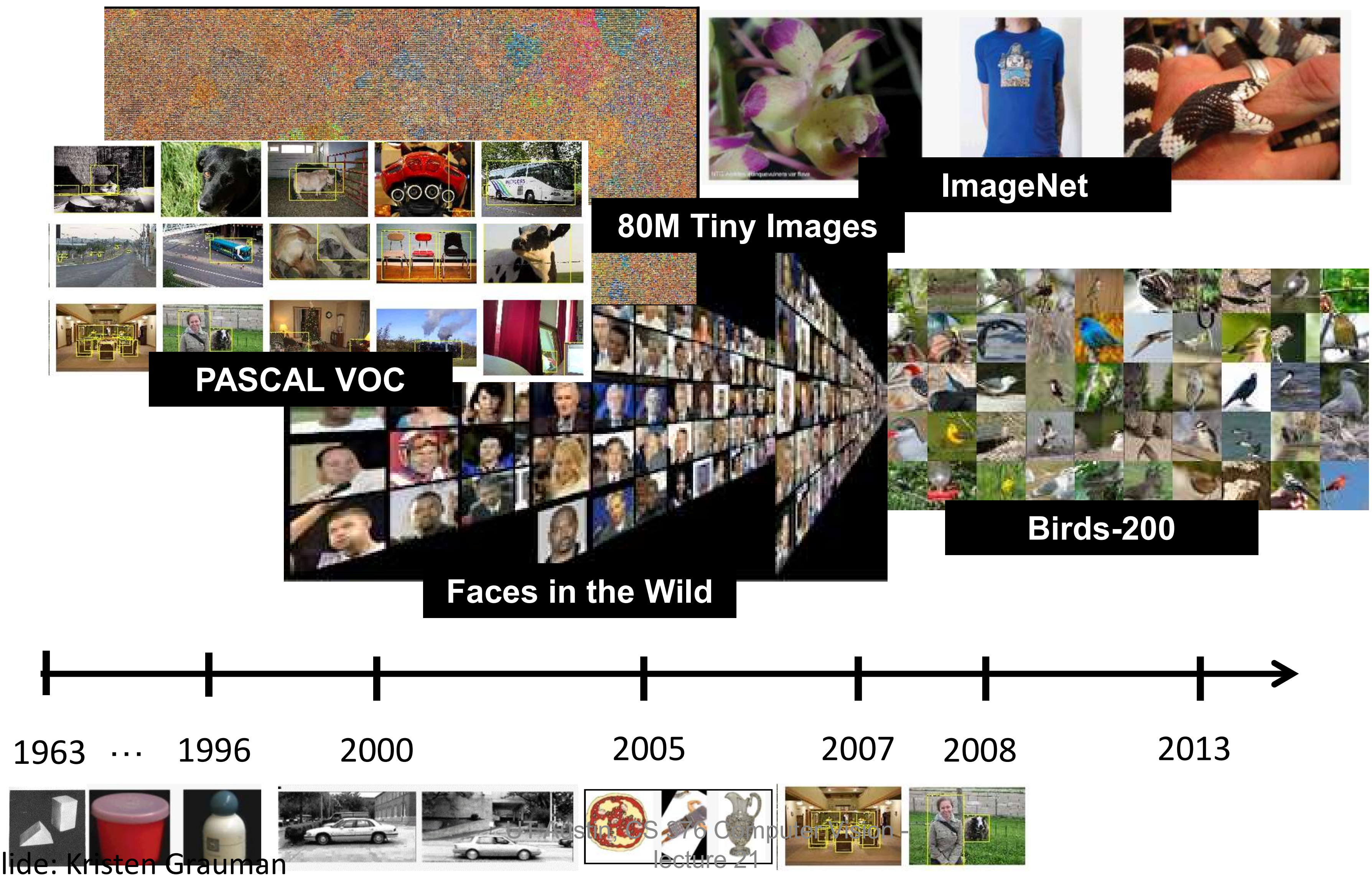
Slide: Kristen Grauman

Progress charted by datasets



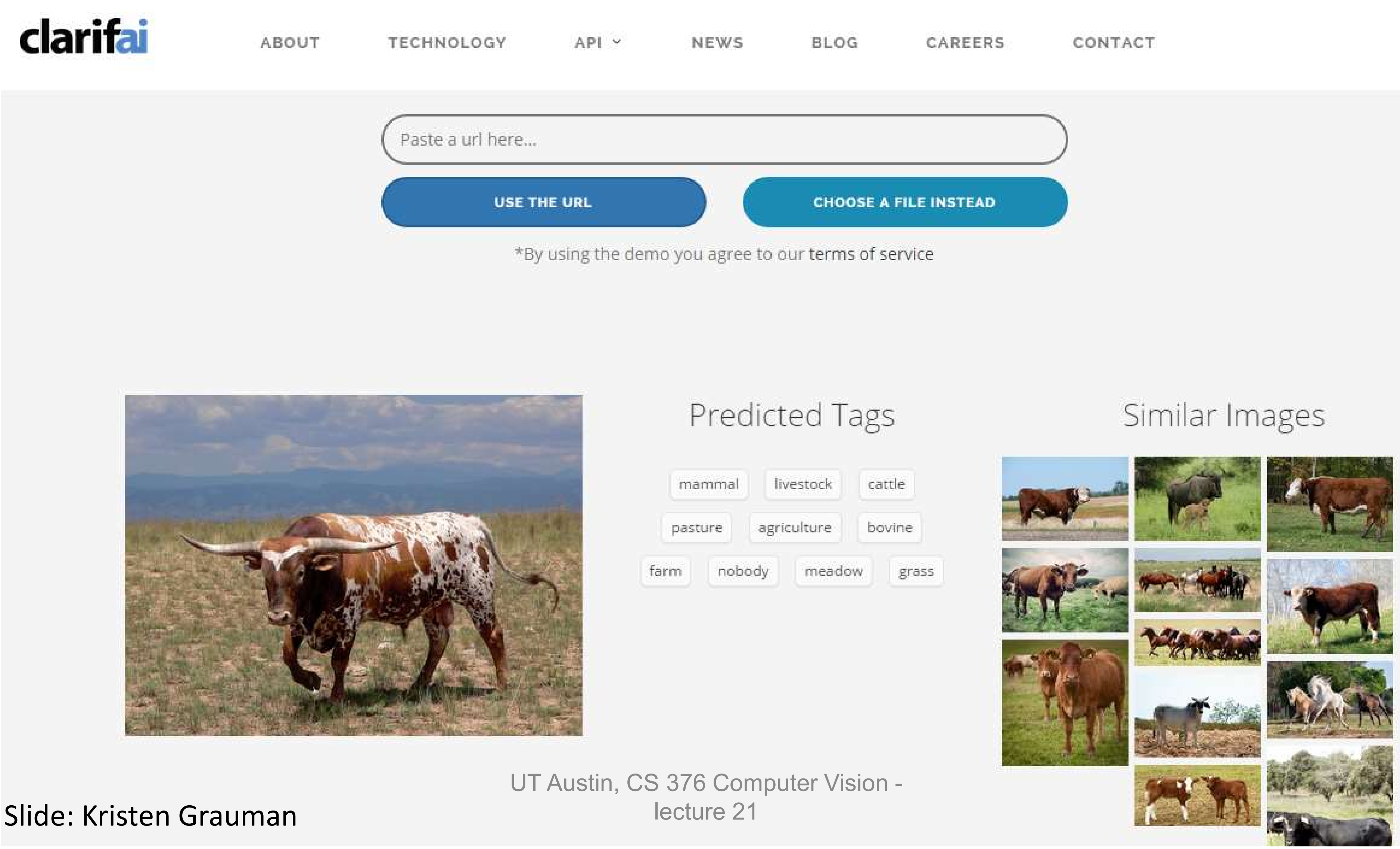
Slide: Kristen Grauman

Progress charted by datasets



Slide: Kristen Grauman

Expanding horizons: large-scale recognition



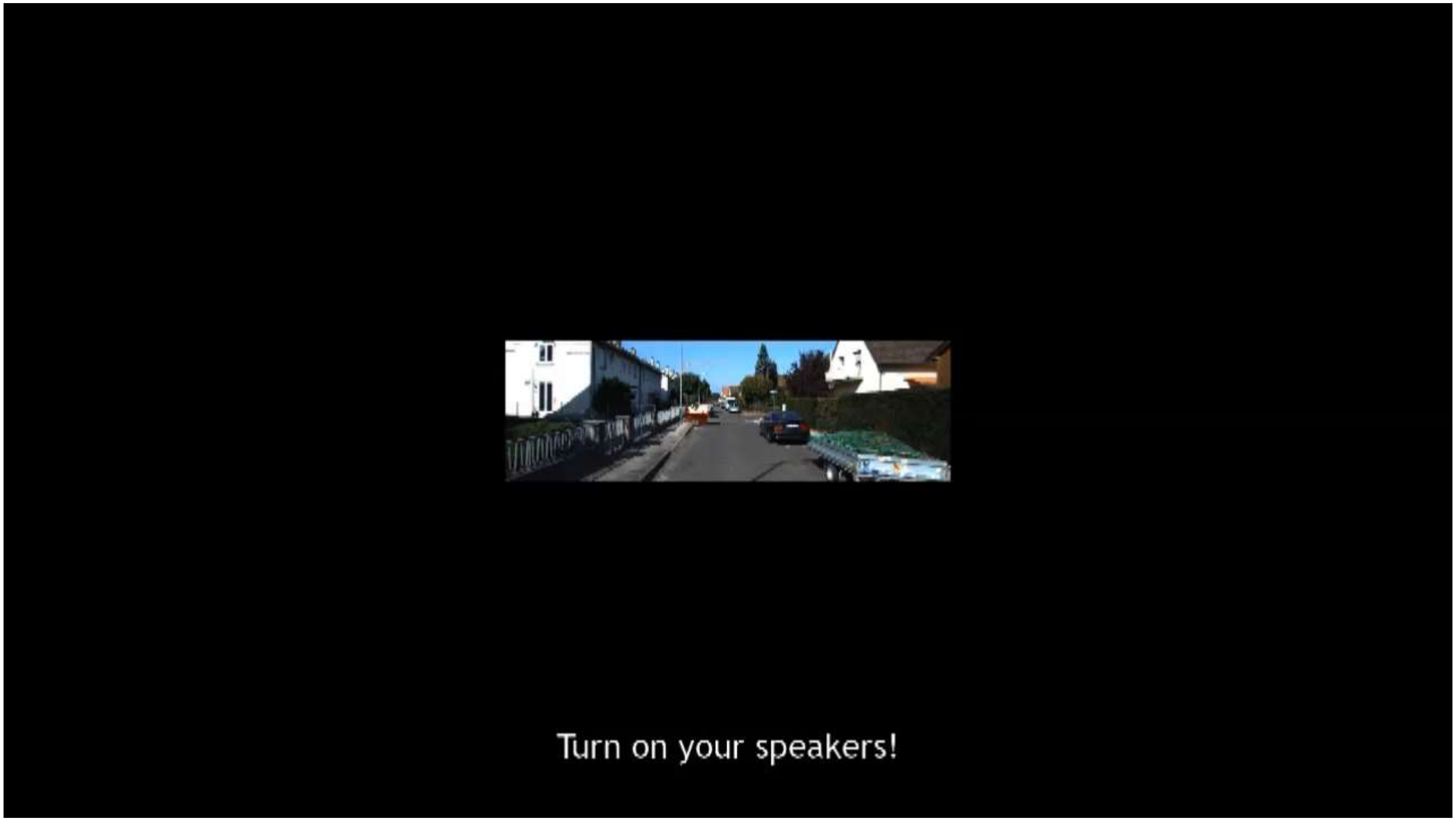
Slide: Kristen Grauman

Expanding horizons: captioning

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

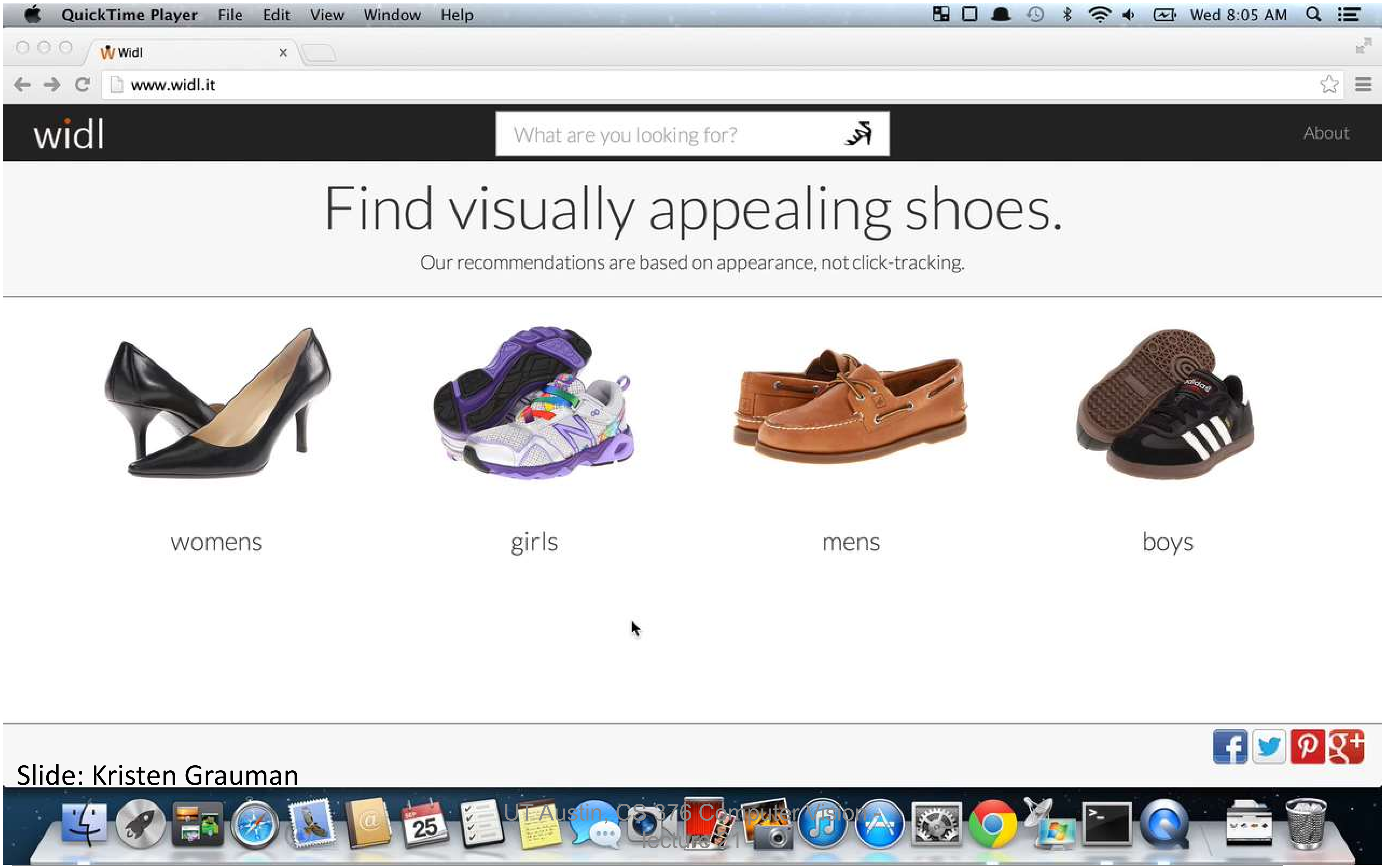
Expanding horizons: visual question answering

Expanding horizons: vision for autonomous vehicles



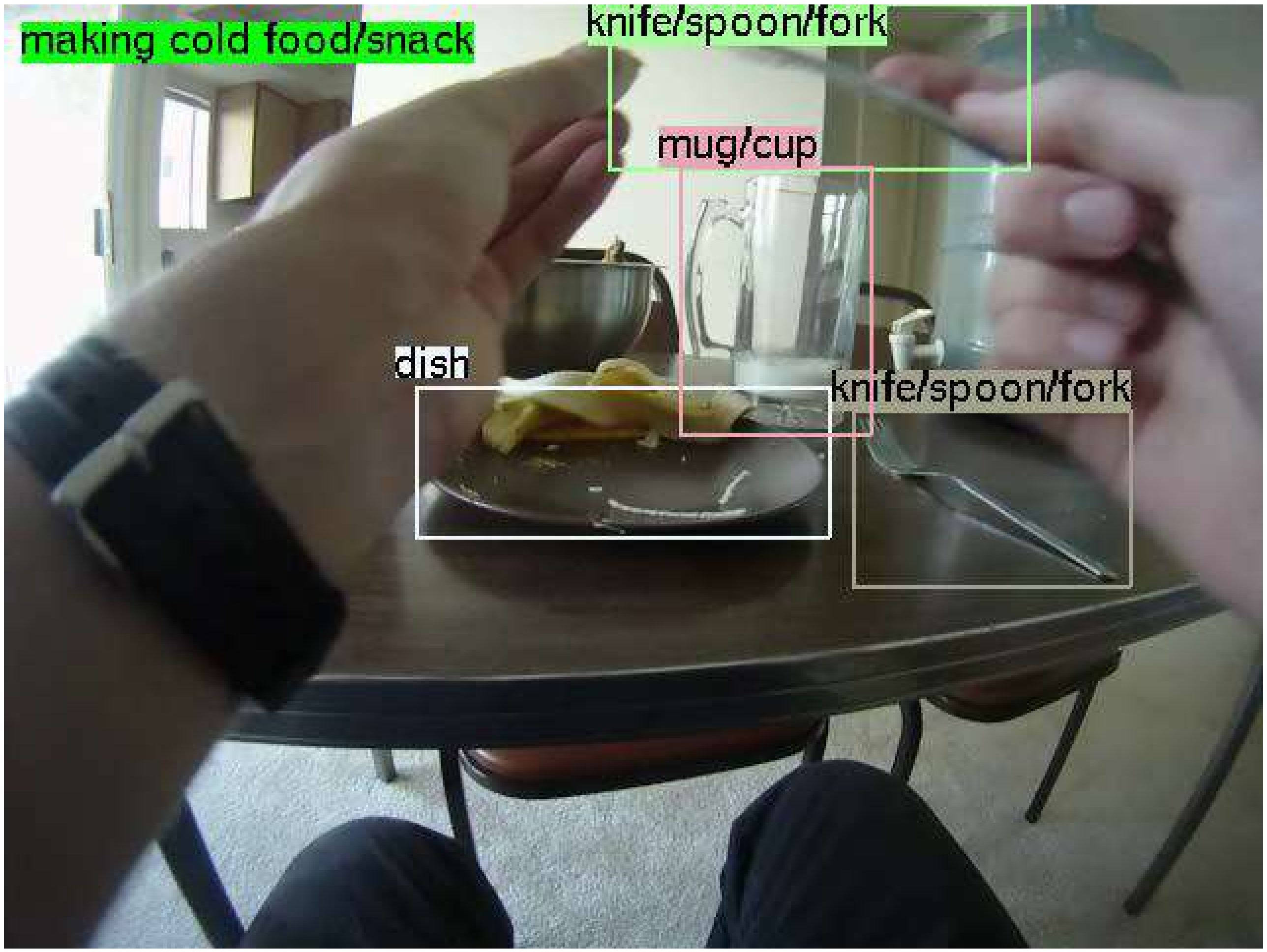
KITTI dataset – Andreas Geiger et al.

Expanding horizons: interactive visual search



Slide: Kristen Grauman

Expanding horizons: first-person vision



Activities of Daily Living – Hamed Pirsiavash et al.

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!)