Computer Vision: The Modern Approach

David Forsyth

Professor

University of Illinois, Urbana Champaign

February 20-22, 2013 | Orlando World Marriott Center | Orlando, Florida USA

What do we do in Computer Vision?

- Two major intertwining themes
  - Reconstruction
    - Build me a model of it
  - Recognition
    - What is this like

- Wildly successful field
  - 20 years ago:
    - eccentric preoccupation of few
  - Now:
    - massive impact, including numerous applications

Reconstruction - SOA

- ## Multiple cameras
  - Incredible results at huge geometric scales
    - centimeters of error from multiple building quadrangles
  - essentially automatic
  - video is better than scattered single images, but...
- ## Single views
  - much more difficult, but some progress
  - cues include: symmetry, stylized shapes, contour, texture, shading

M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, Visual modeling with a hand-held camera, International Journal of Computer Vision 59(3), 207-232, 2004

Why is visual object recognition useful?

- If you want to act, you must draw distinctions
- For robotics
  - recognition can predict the future
    - is the ground soggy?
    - is that person doing something dangerous?
    - does it matter if I run that over?
    - which end is dangerous?
- For information systems
  - recognition can unlock value in pictures
    - for search, clustering, ordering, inference, ...
- General engineering
  - recognition can tell what people are doing
- If you have vision, you have some recognition system

Words and Pictures Affect One Another

Marc by Marc Jacobs
Adorable peep-toe pumps, great for any occasion.
Available in an array of uppers. Metallic fabric trim
and bow detail. Metallic leather lined footbed.
Lined printed design.
Leather sole. 3 3/4" heel.

Zappos.com

soft and glassy patent calfskin trimmed with
natural vachetta cowhide, open top satchel for
daytime and weekends, interior double slide
pockets and zip pocket, seersucker stripe cotton
twill lining, kate spade leather license plate logo,
imported
2.8" drop length
14"h x 14.2"w x 6.9"d

Katespade.com

It's the perfect party dress. With distinctly feminine
details such as a wide sash bow around an empire
waist and a deep scoopneck, this linen dress will
keep you comfortable and feeling elegant all
evening long. Measures 38" from center back, hits
at the knee.
  * Scoopneck, full skirt.
  * Hidden side zip, fully lined.
  * 100% Linen. Dry clean.

bananarepublic.com

ROBOTIC INDUSTRIES ASSOCIATION

advancing VISION+IMAGING

MCA
MOTION CONTROL ASSOCIATION

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
- Crucial open questions
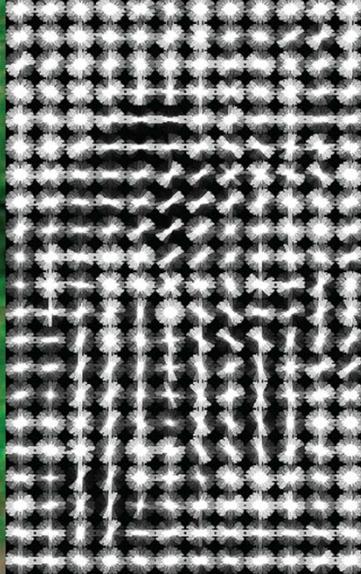 - dataset bias
  - links to utility
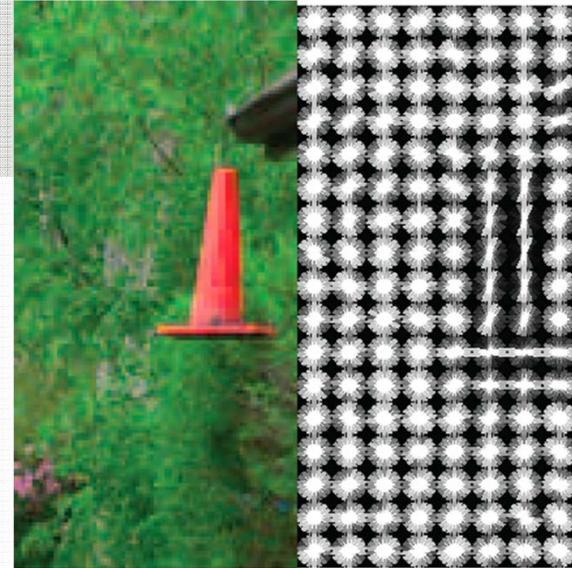
A Belief Space About Recognition

- Object categories are fixed and known
  - Each instance belongs to one category of k

- Good training data for categories is available
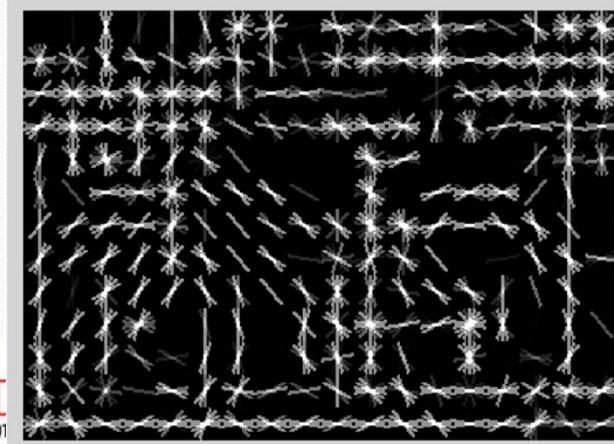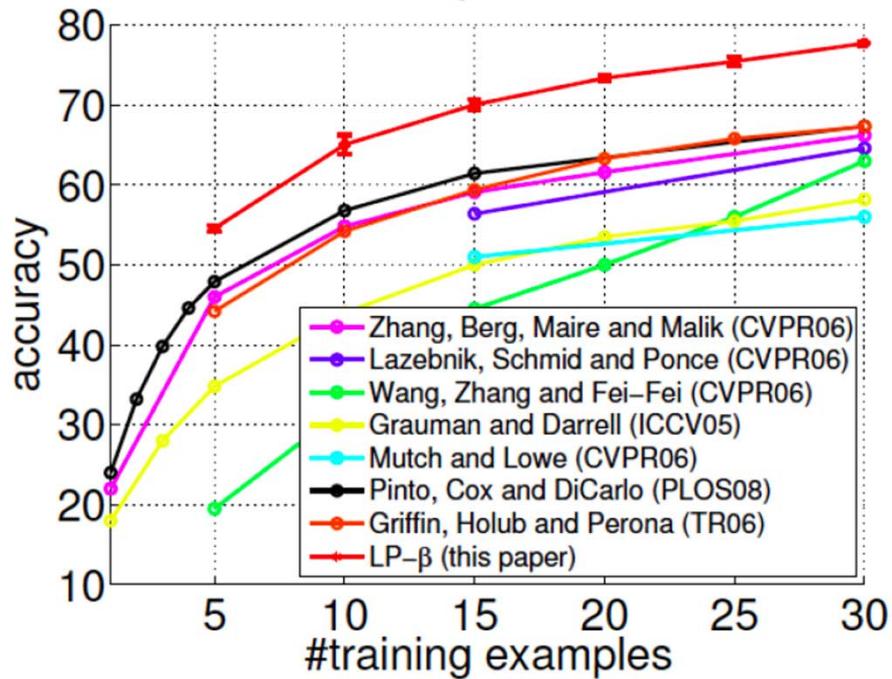
- Object recognition=k-way classification

- Detection = lots of classification

Features

- Principles
  - illumination invariant (robust) -> gradient orientation features
  - windows always slightly misaligned -> local histograms
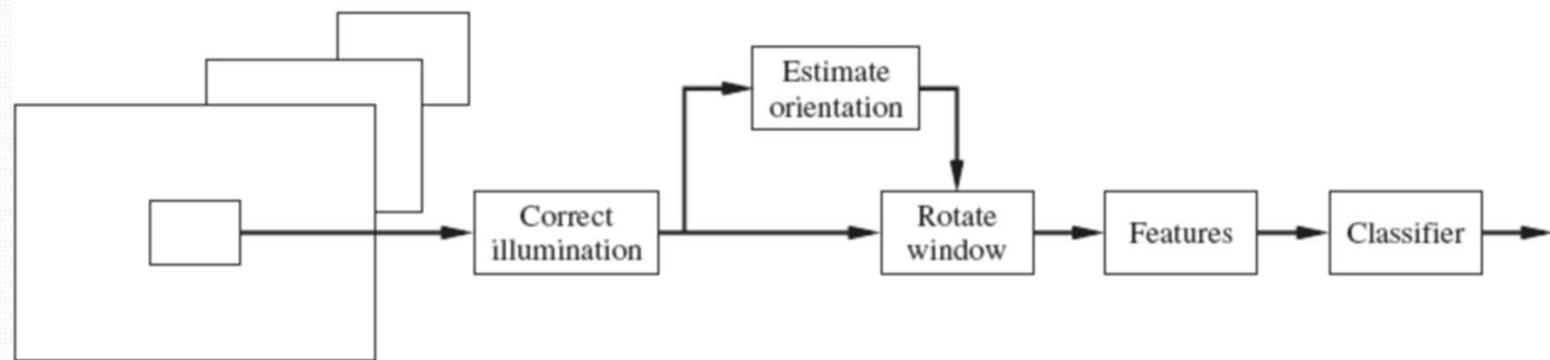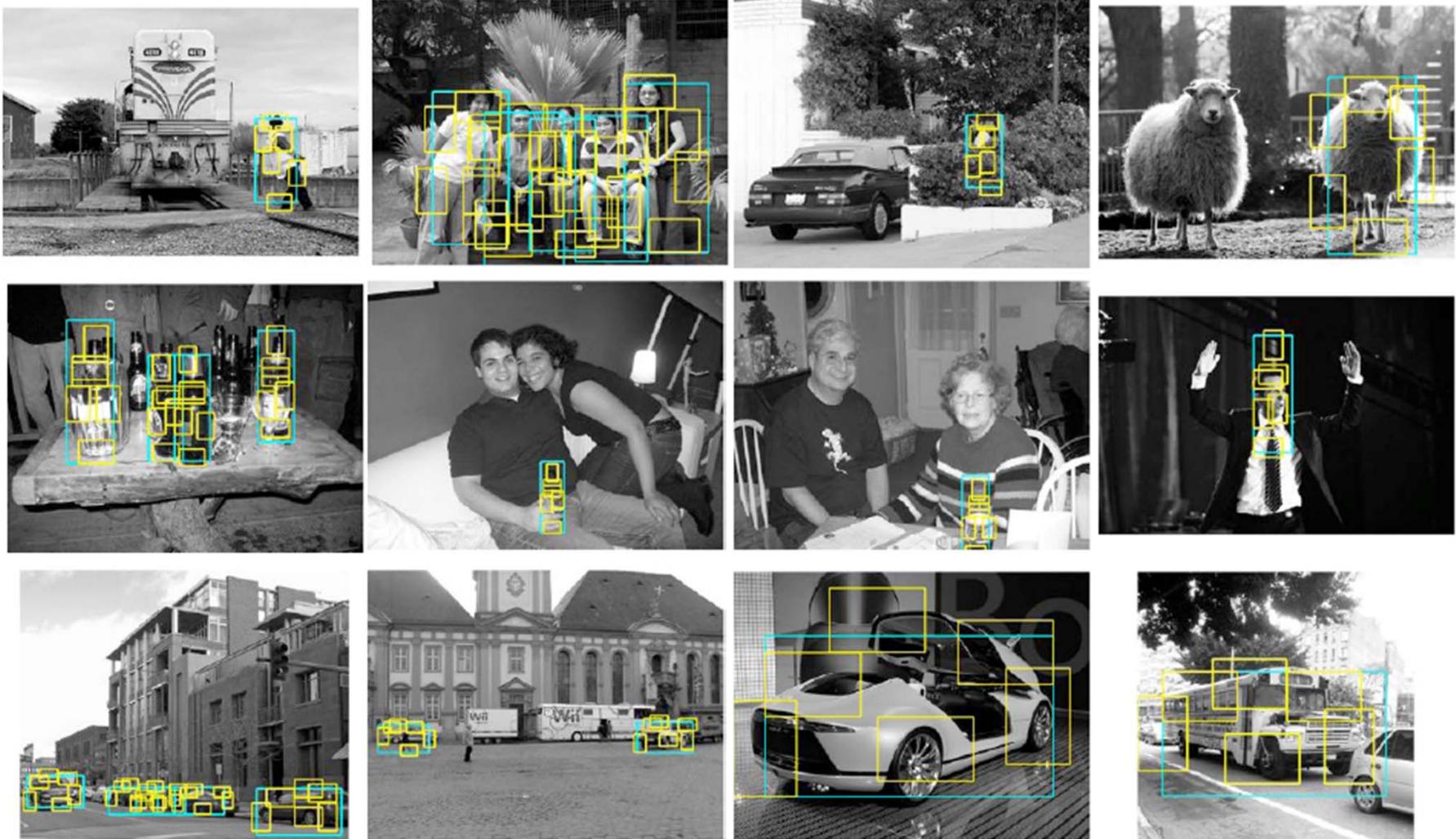- HOG, SIFT features (Lowe, 04; Dalal+Triggs 05)
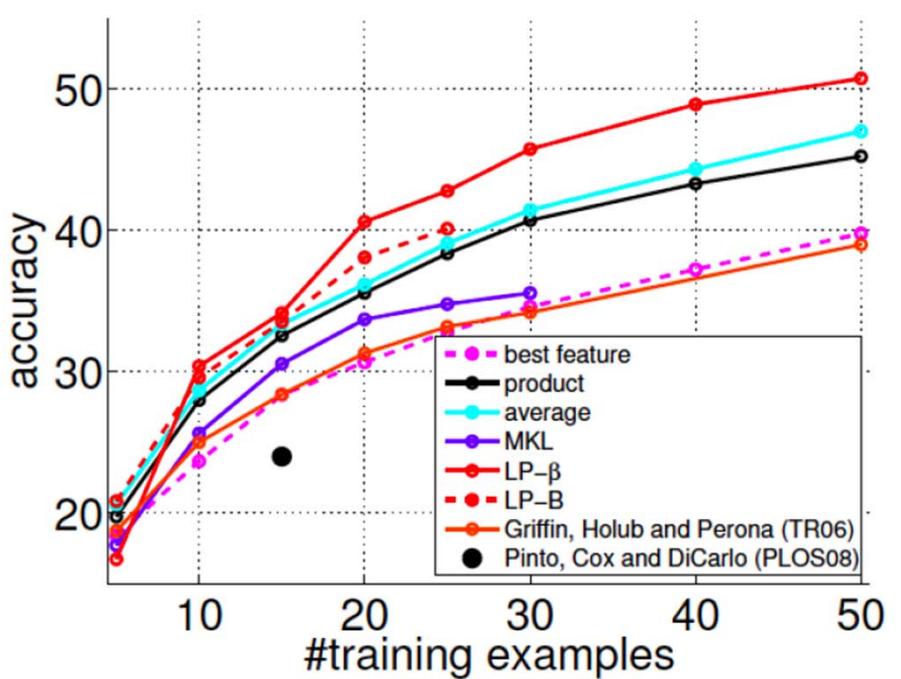
Classification Works Well

Caltech101 comparison to literature

Zhang, Berg, Maire and Malik (CVPR06)
Lazebnik, Schmid and Ponce (CVPR06)
Wang, Zhang and Fei−Fei (CVPR06)
Grauman and Darrell (ICCV05)
Mutch and Lowe (CVPR06)
Pinto, Cox and DiCarlo (PLOS08)
Griffin, Holub and Perona (TR06)
LP−β (this paper)

Caltech−256 (39 kernels)

best feature
product
average
MKL
LP−β
LP−B
Griffin, Holub and Perona (TR06)
Pinto, Cox and DiCarlo (PLOS08)

Detection with a Classifier

P. Felzenszwalb, D. McAllester, D. Ramanan. "A Discriminatively Trained, Multiscale, Deformable Part Model" CVPR 2008.

A Belief Space About Eecognition

- Object categories are fixed and known
  - Each instance belongs to one category of k   Obvious nonsense

- Good training data for categories is available Obvious nonsense

- Object recognition=k-way classification   Obvious nonsense

- Detection = lots of classification

ROBOTIC INDUSTRIES ASSOCIATION

aia advancing VISION+IMAGING

mca MOTION CONTROL ASSOCIATION

Are these monkeys?

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - sentences
- Crucial open questions
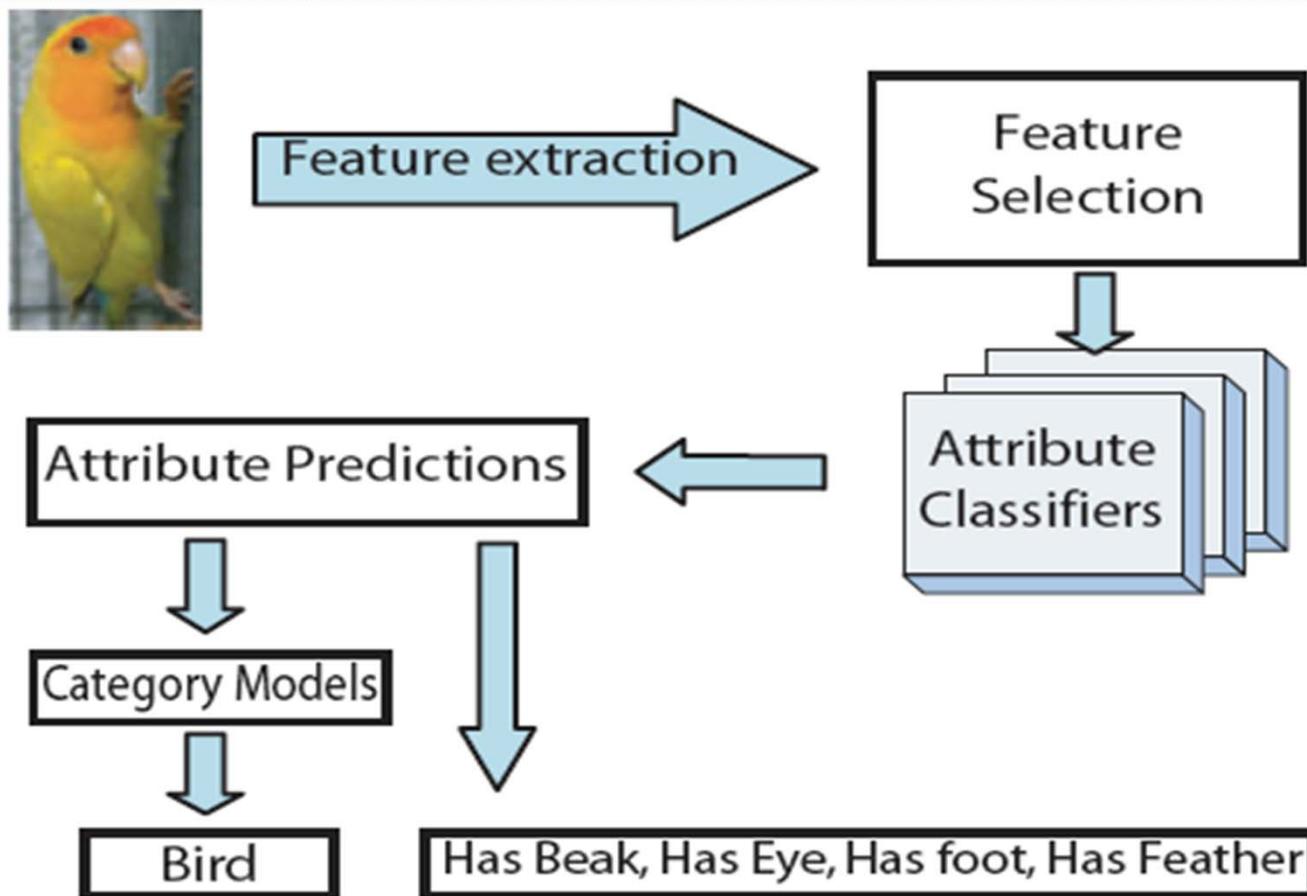 - dataset bias
  - links to utility

The Unfamiliar

The Unfamiliar

General Architecture

Farhadi et al 09; cf Lampert et al 09

Attribute Predictions for Unknown Objects

Farhadi et al 09; cf Lampert et al 09

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility

ROBOTIC INDUSTRIES ASSOCIATION

aia advancing VISION+IMAGING

mca MOTION CONTROL ASSOCIATION
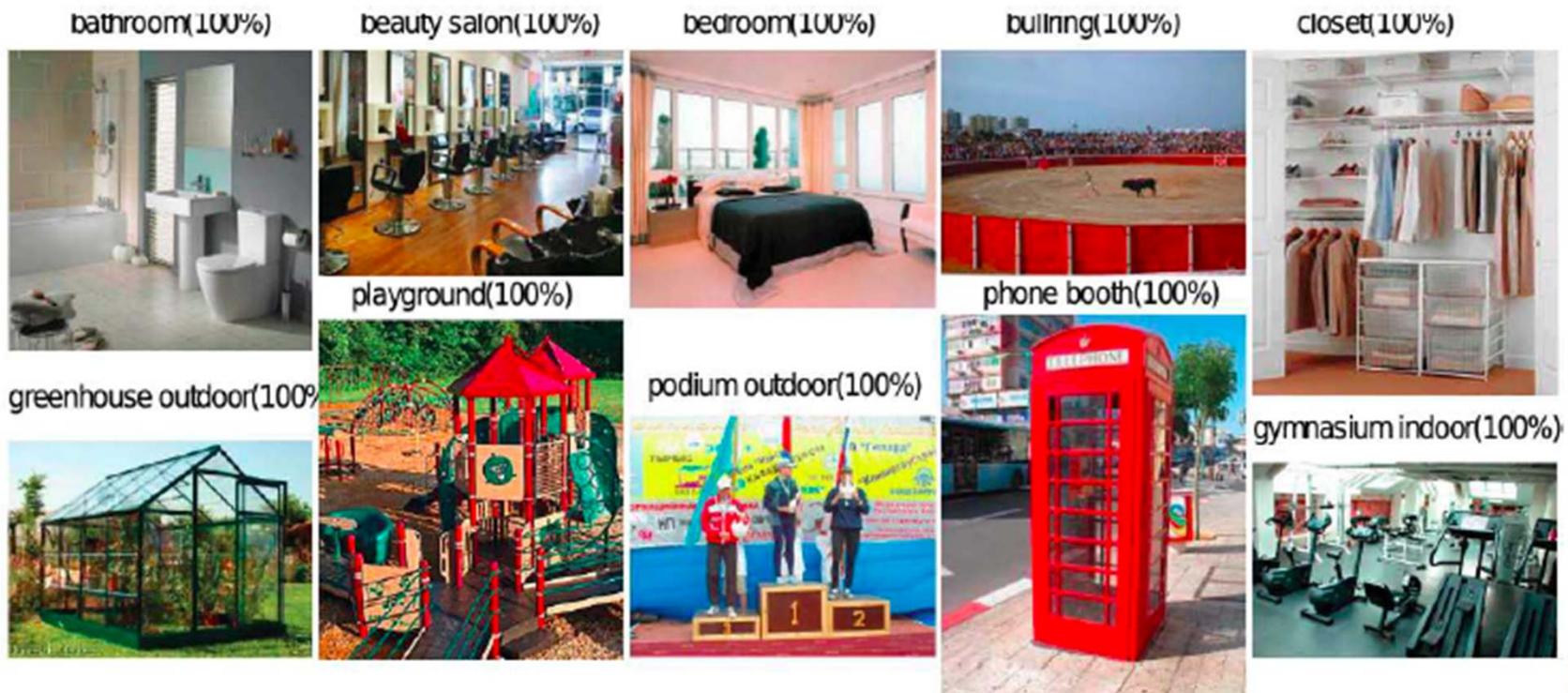
Meaning Comes in Clumps

"Sledder"
Is this one thing?
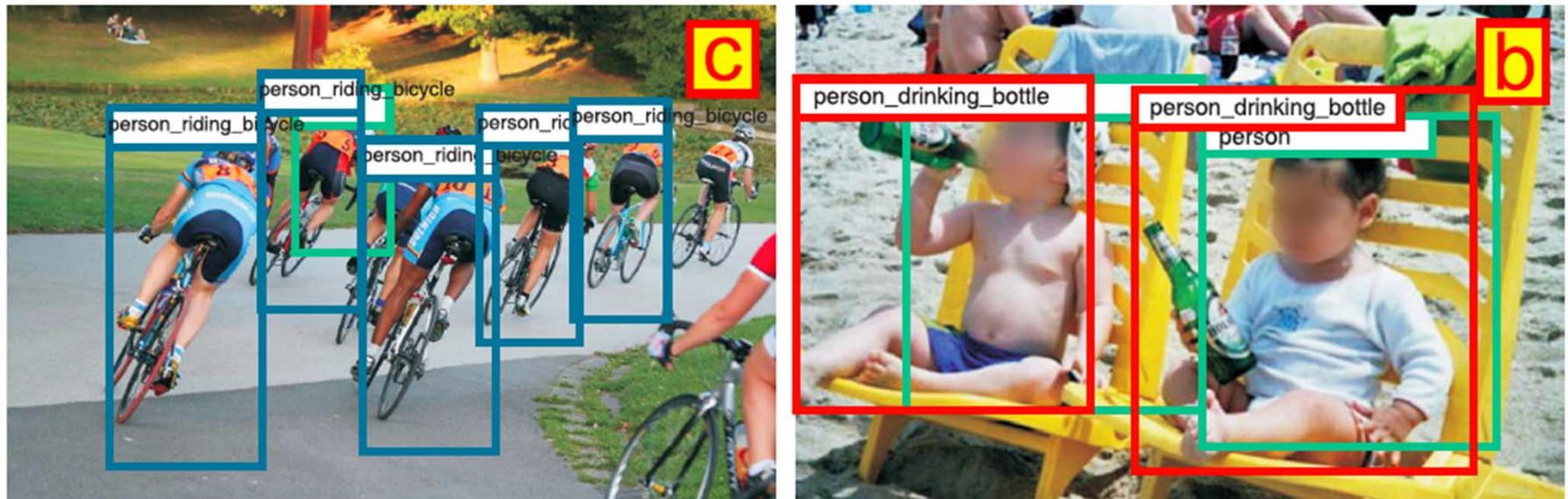Should we cut her off her sled?

Scenes

- Likely stages for
  - Particular types of object
  - Particular types of activity

Xiao et al 10

Scenes > Visual phrases > Objects

- Composites
  - easier to recognize than their components
  - because appearance is simpler

Farhadi + Sadeghi 11
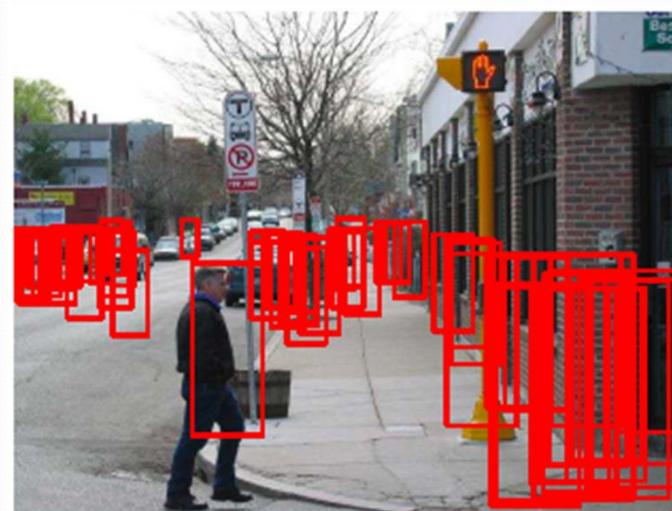
Decoding

Decoding Helps

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
- Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
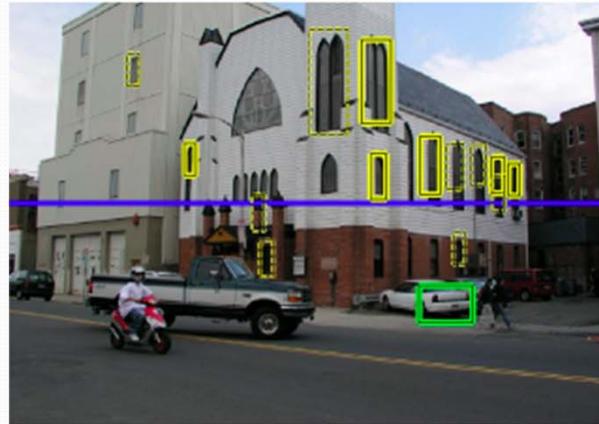- Crucial open questions
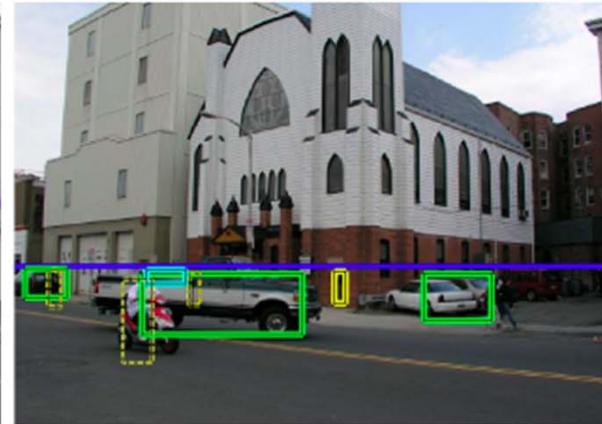 - dataset bias
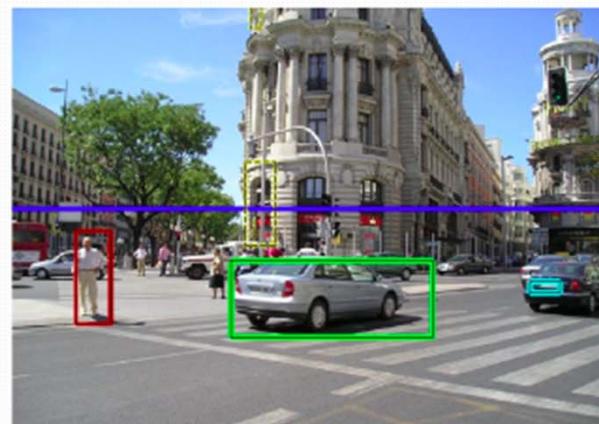 - links to utility

Environmental Knowledge

$2.2 < y_c < 2.8$

Hoiem et al 06

Environmental Knowledge is Powerful

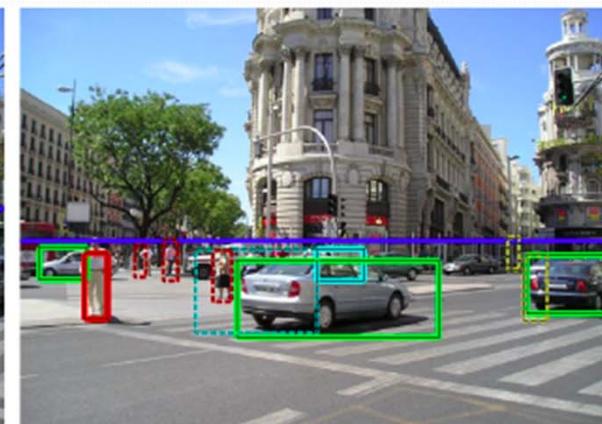(b) Local Detection (b) Full Model Detection

(a) Local Detection (a) Full Model Detection

Hoiem et al 06

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - selection
- Crucial open questions
  - dataset bias
  - links to utility

Selection: What is worth saying?

Two girls take a break to sit and talk .

Two women are sitting , and one of them is holding something .

Two women chatting while sitting outside

Two women sitting on a bench talking .

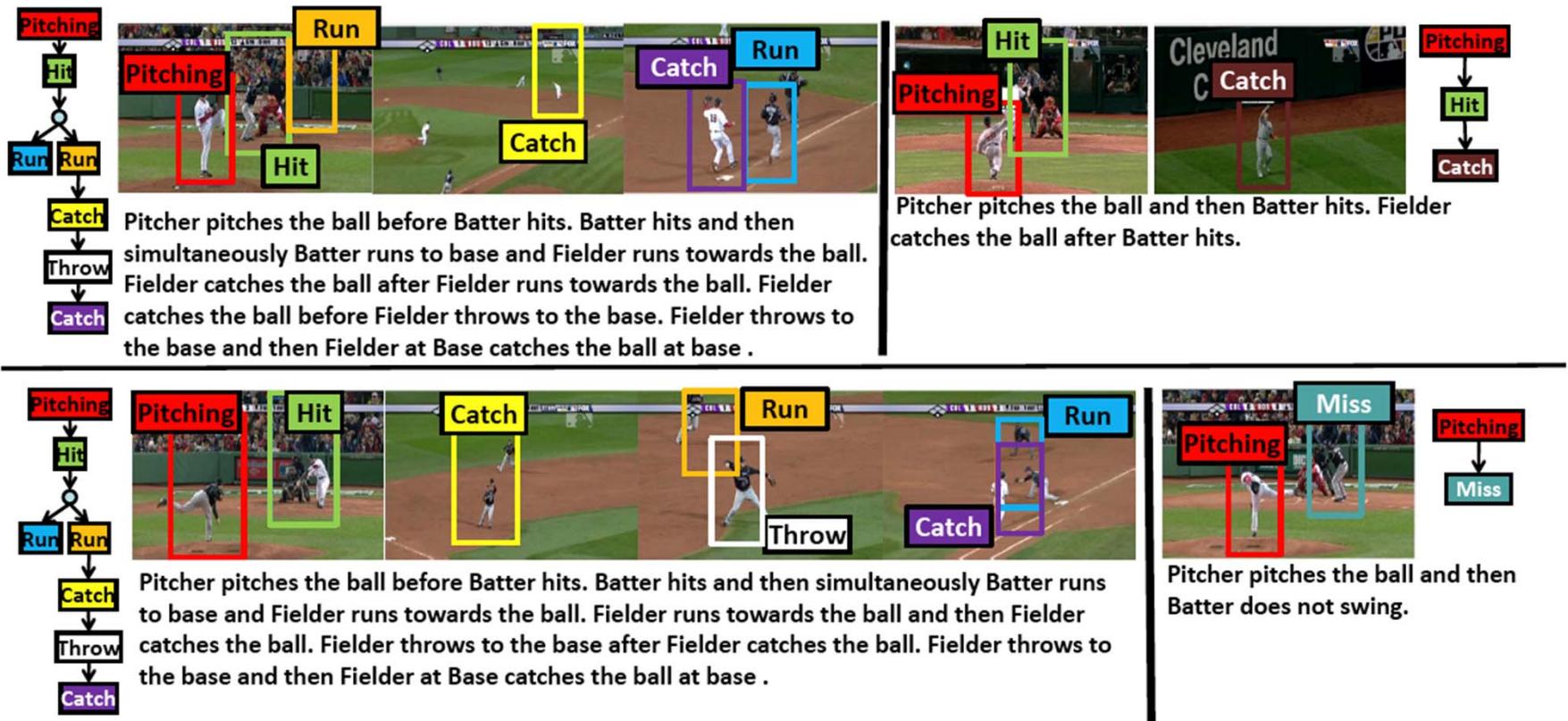Two women wearing jeans , one with a blue scarf around her head , sit and talk .

Sentences from Julia Hockenmaier's work

For language people: Pragmatics - what is worth saying?

Rashtchian ea 10

Describing Structured Events

Pitcher pitches the ball before Batter hits. Batter hits and then simultaneously Batter runs to base and Fielder runs towards the ball. Fielder catches the ball after Fielder runs towards the ball. Fielder catches the ball before Fielder throws to the base. Fielder throws to the base and then Fielder at Base catches the ball at base .

Pitcher pitches the ball and then Batter hits. Fielder catches the ball after Batter hits.

Pitcher pitches the ball before Batter hits. Batter hits and then simultaneously Batter runs to base and Fielder runs towards the ball. Fielder runs towards the ball and then Fielder catches the ball. Fielder throws to the base after Fielder catches the ball. Fielder throws to the base and then Fielder at Base catches the ball at base .

Pitcher pitches the ball and then Batter does not swing.

Gupta ea 09

Adding Attributes and Prepositions

This is a photograph of one sky, one road and one bus. The blue sky is above the gray road. The gray road is near the shiny bus. The shiny bus is near the blue sky.
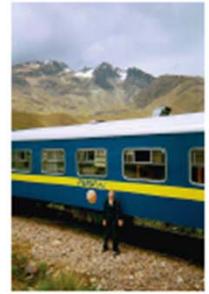

There are two aeroplanes. The first shiny aeroplane is near the second shiny aeroplane.

There are one cow and one sky. The golden cow is by the blue sky.

There are one dining table, one chair and two windows. The wooden dining table is by the wooden chair, and against the first window, and against the second white window. The wooden chair is by the first window, and by the second white window. The first window is by the second white window.

Here we see one person and one train. The black person is by the train.

This is a picture of one sky, one road and one sheep. The gray sky is over the gray road. The gray sheep is by the gray road.

Here we see one road, one sky and one bicycle. The road is near the blue sky, and near the colorful bicycle. The colorful bicycle is within the blue sky.

Here we see two persons, one sky and one aeroplane. The first black person is by the blue sky. The blue sky is near the shiny aeroplane. The second black person is by the blue sky. The shiny aeroplane is by the first black person, and by the second black person.

This is a picture of two dogs. The first dog is near the second furry dog.

This is a photograph of two buses. The first rectangular bus is near the second rectangular bus.

Conclusion

- Recognition is subtle
  - strong basic methods based on classifiers
  - many meanings, useful in different contexts
- Important recognition technologies coming
  - attributes
  - phrases
  - geometry
  - sentences
- Crucial open questions
  - dataset bias
  - links to utility

Bias is Pervasive
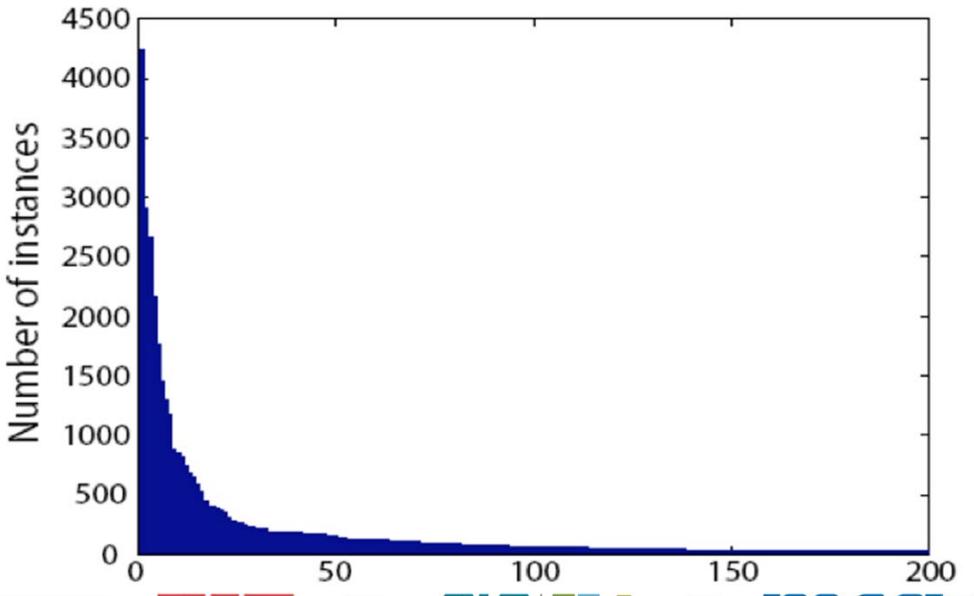
Object Recognition is Biased

- The world can't be like the dataset because
  - many things are rare
  - this exaggerates bias

Instance number of the top 200 object categories

Defenses Against Bias

- Appropriate feature representations
  - e.g. illumination invariance

- Appropriate intermediate representations
  - which could have less biased behavior
  - perhaps attributes? scenes? visual phrases?

- Appropriate representations of knowledge
  - e.g. geometry --- pedestrian example

What should we say about visual data?

- Most important question in vision
  - What does the output of a recognition system consist of?

- A useful representation of reasonable size
  - dubious answer
    - Useful in what way?
    - How do we make the size reasonable?

Object Categories Depend on Utility

Monkey or Plastic toy or both or irrelevant

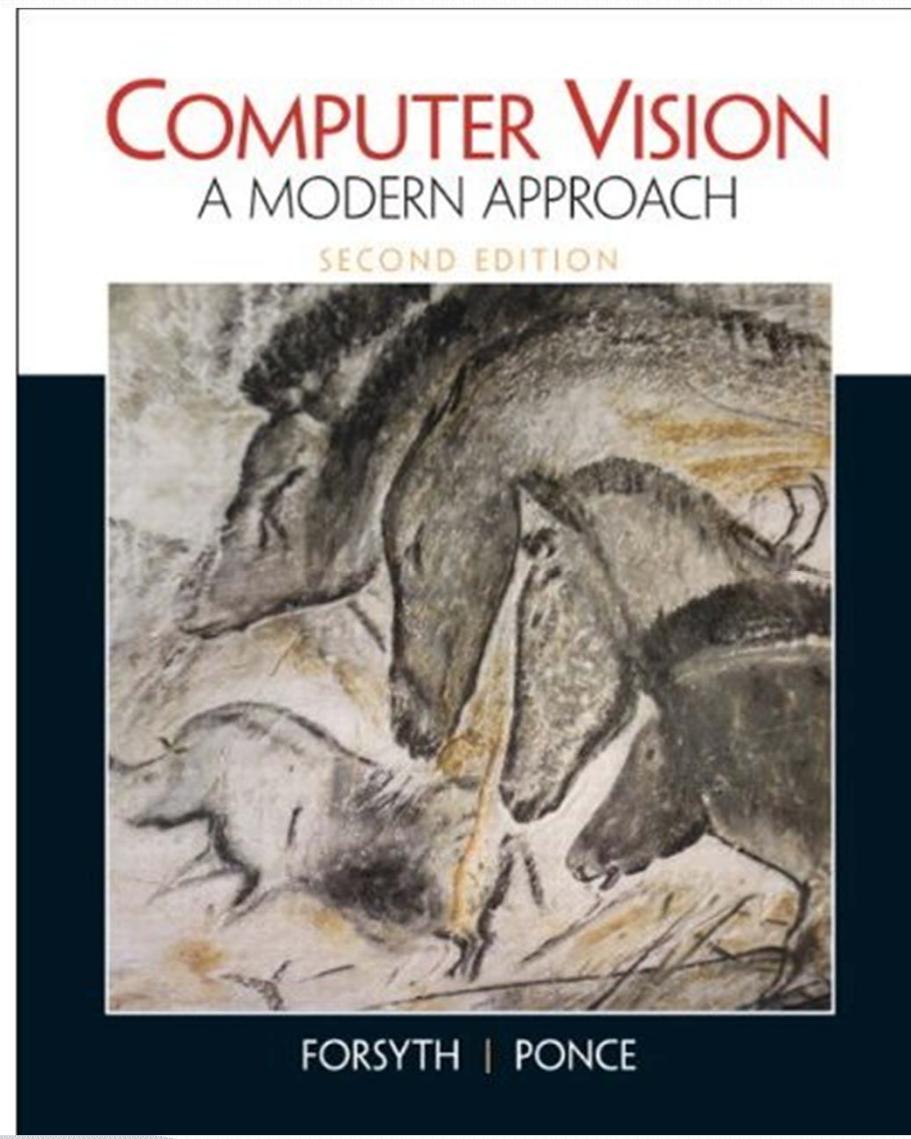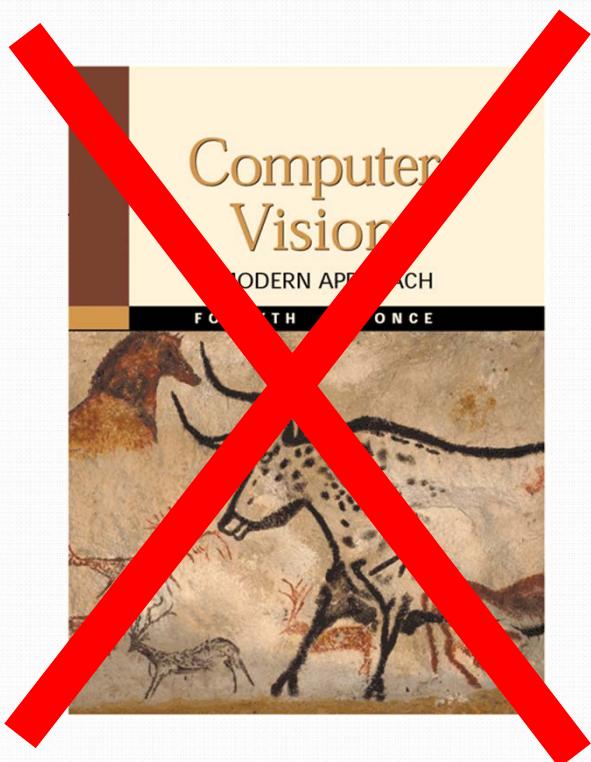Some of this depends on what you're trying to do, in ways we don't understan

Person or child or beer drinker or beer-drinking child or tourist or holidaymaker or obstacle or potential arrest or irrelevant or...

Conclusion

- ## Recognition is subtle
  - strong basic methods based on classifiers
- ## Important recognition technologies coming
  - the unfamiliar
  - phrases
  - geometry
  - selection
- ## Crucial open questions
  - dataset bias
  - links to utility

More Information

Contact Information

David Forsyth

Professor

University of Illinois

Computer Science Department

Urbana-Champaign, Illinois 61801

USA

Email: daf@uiuc.edu

Web: http://luthuli.cs.uiuc.edu/~daf