

Computer Vision (with a special emphasis on people)

D.A. Forsyth

Computer Vision

A MODERN APPROACH

FORSYTH ■ PONCE



Topics

- Monday: Overview, Cameras, Light, Edges
- Tuesday: HOG/SIFT, Classifiers, Detectors, NDD
- Wednesday: Reconstruction and Tracking
- Thursday: People: finding, tracking, activity, faces
- Friday: Applications: shadow removal, shading recovery, texture synthesis, words and pictures
- What's missing
 - structure from motion
 - most segmentation
 - much detail
 - exercises, homeworks, etc.

Conclusions

- 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

Example problems

- **Obstacle avoidance**
 - A cricketer avoids being hit in the head (->) (<-)
 - the gannet pulls its wings in in time, by measuring time to contact
- **Reconstructing representations of the 3D world**
 - from multiple views
 - from shading
 - from structural models, etc
- **Recognition**
 - draw distinctions between what is seen
 - is it soggy?
 - will it eat me?
 - can I eat it?
 - is it a cat?
 - is it my cat?

Reconstruction

- Build me a model
 - from pictures
 - from video
 - containing
 - geometric information
 - surface texture information
 - where was the camera
- Core ideas
 - describing and matching points
 - camera geometry

Matching

- **Problem:**
 - what matches what
- **Solution**
 - detailed local descriptions of points (SIFT)

Matching points is important



M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003

Matching points

- A description of tiny gradients near point is distinctive
 - Lowe's SIFT feature

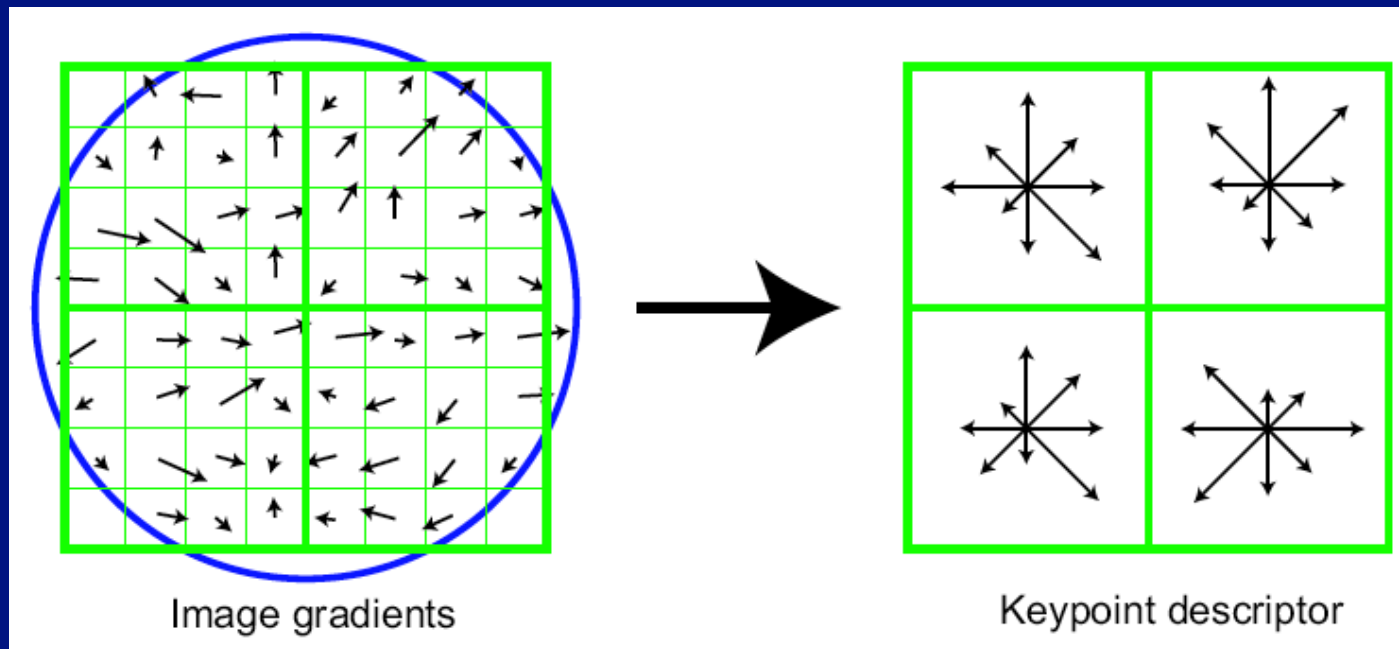
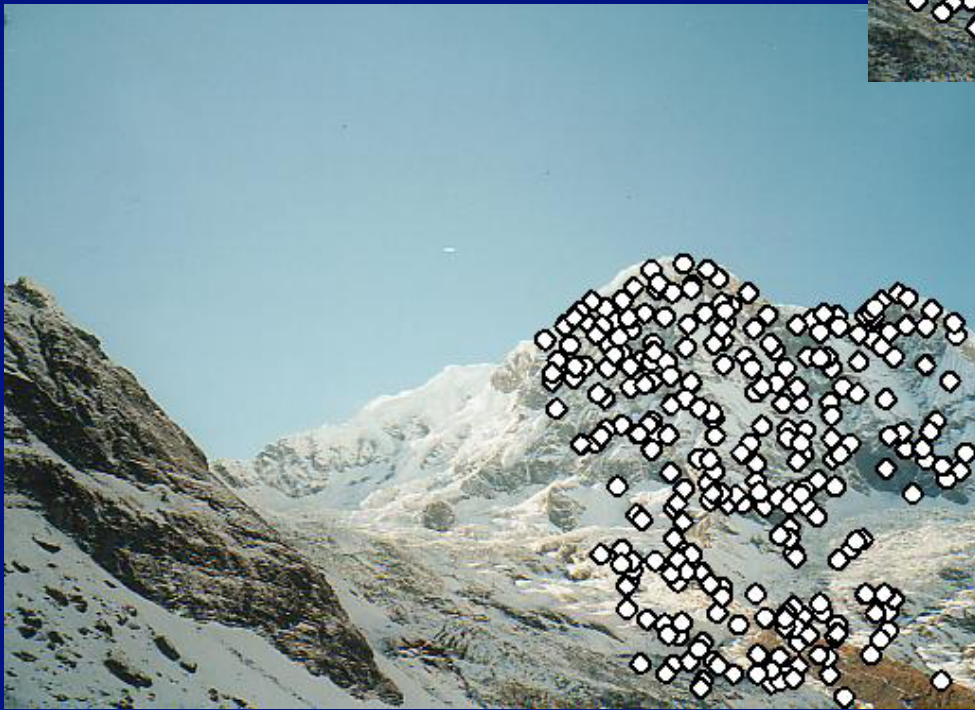
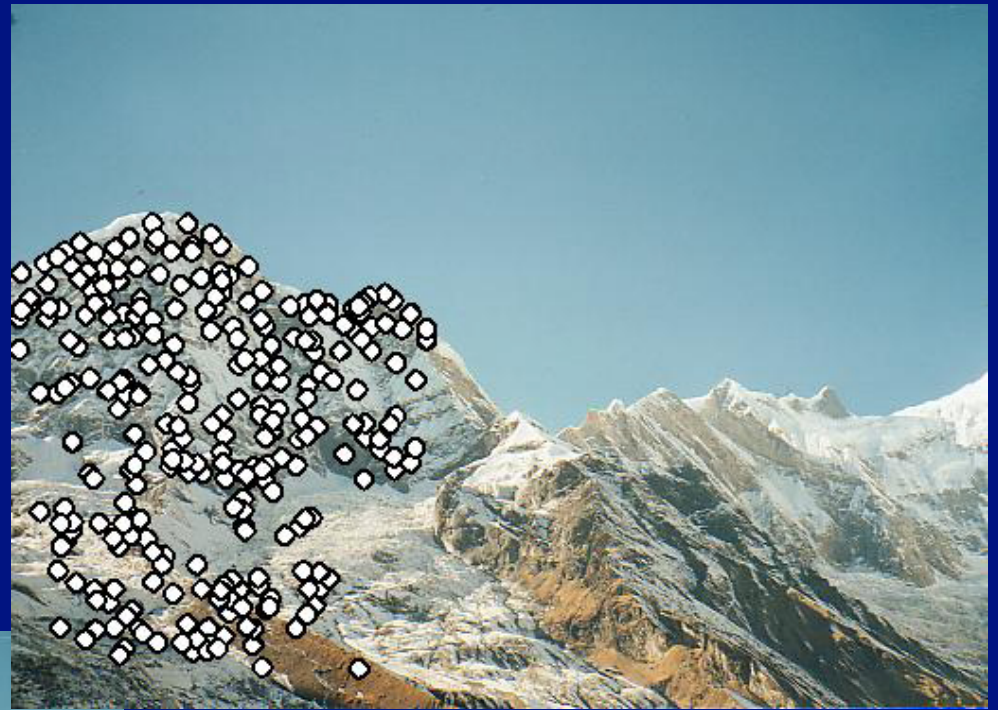


Fig 7 from:
Distinctive image features from scale-invariant keypoints
David G. Lowe, *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.



M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003



Translation isn't enough to align the images - we need to use a homography



M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003



M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003



M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003



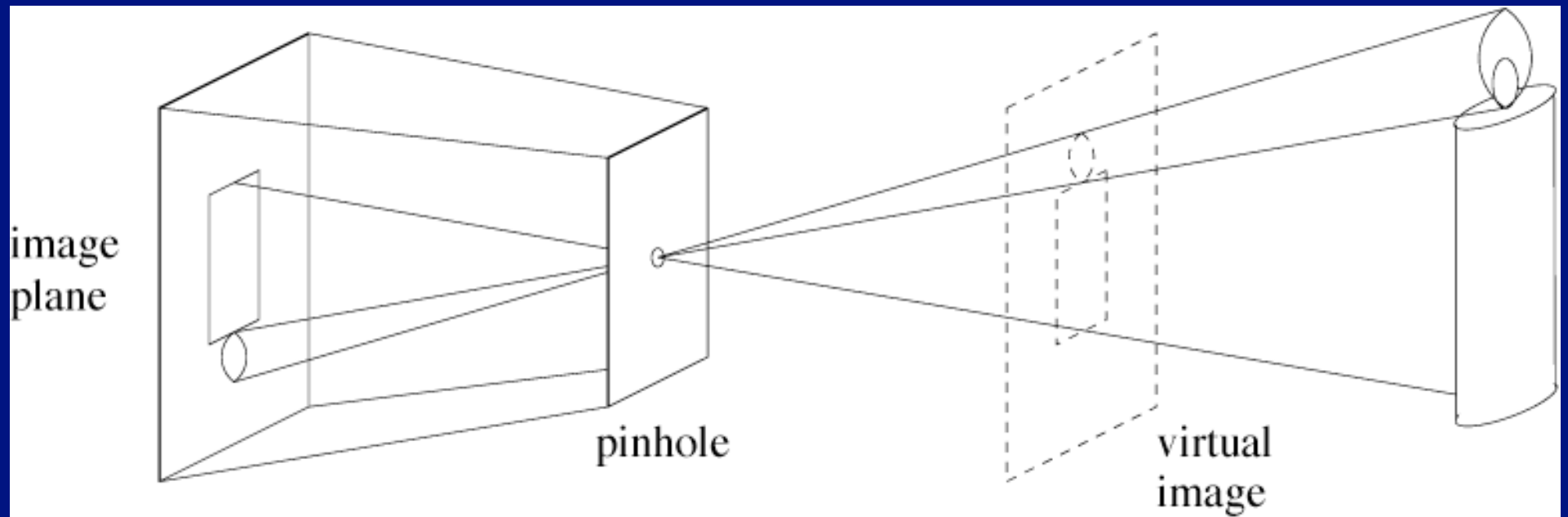
M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003



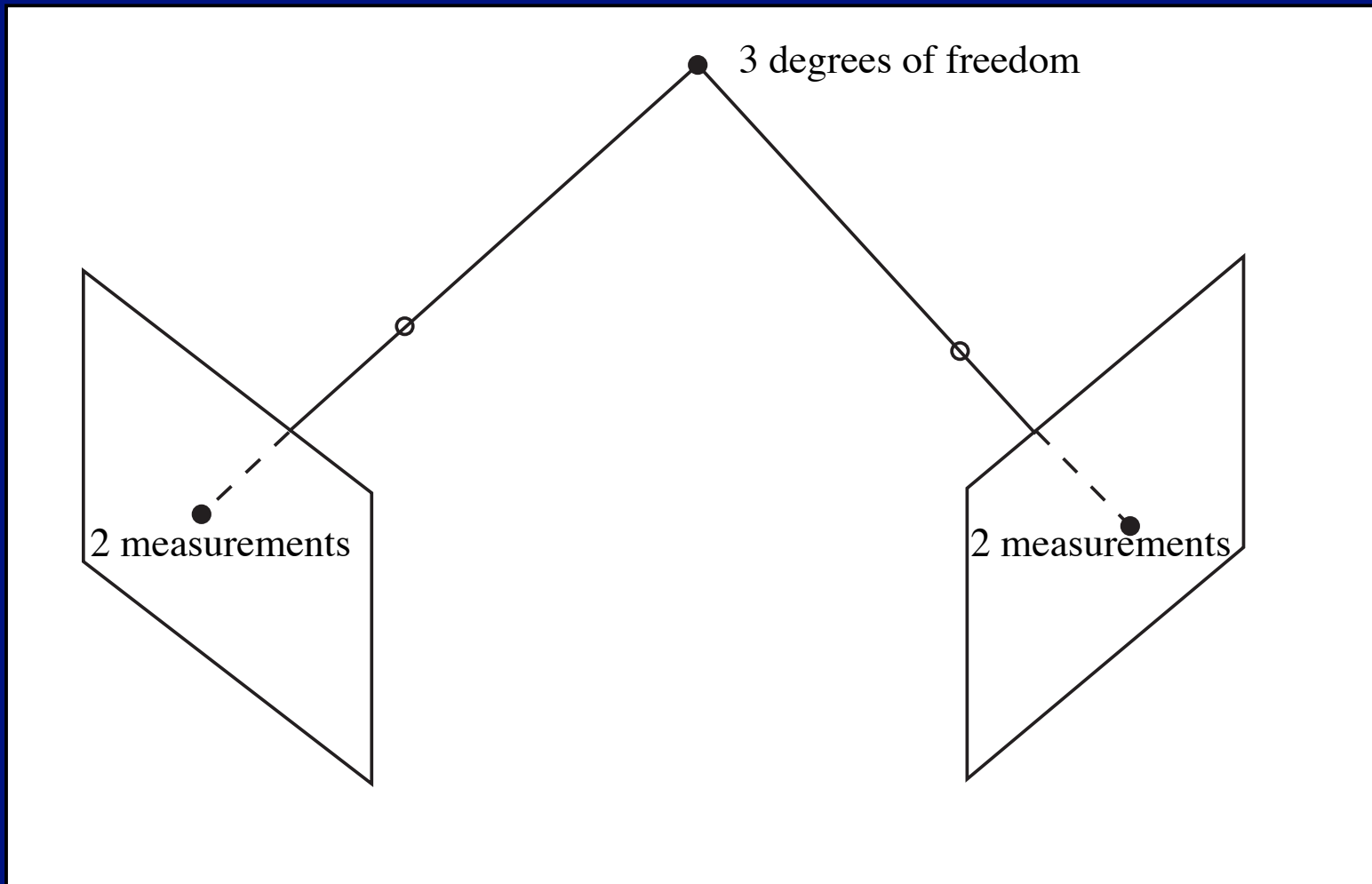
M. Brown and D. Lowe, "Recognising Panoramas", ICCV 2003

How cameras work

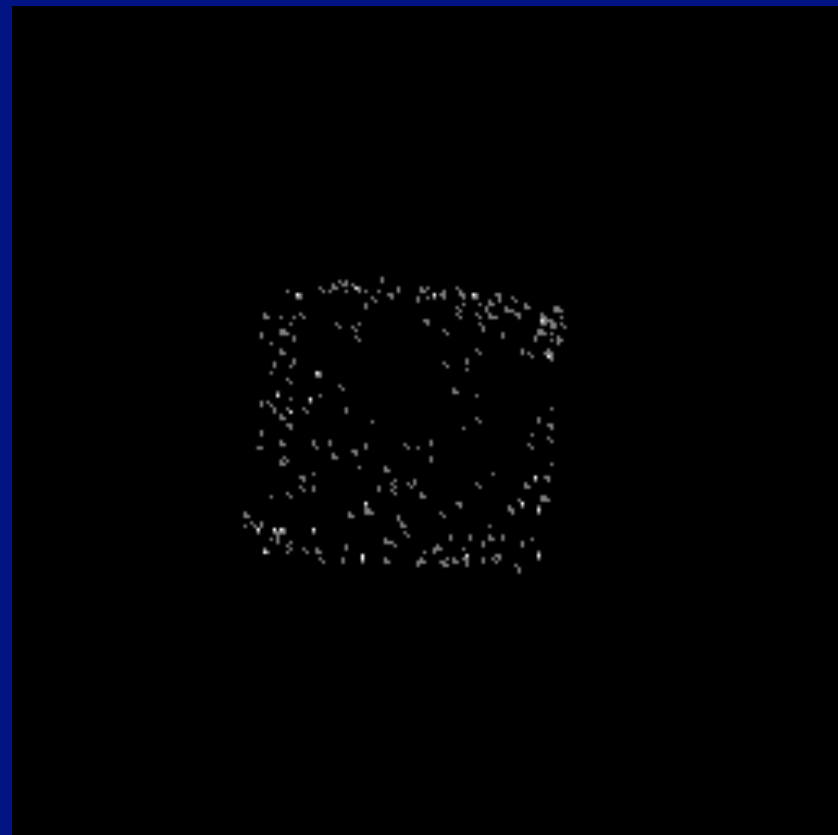
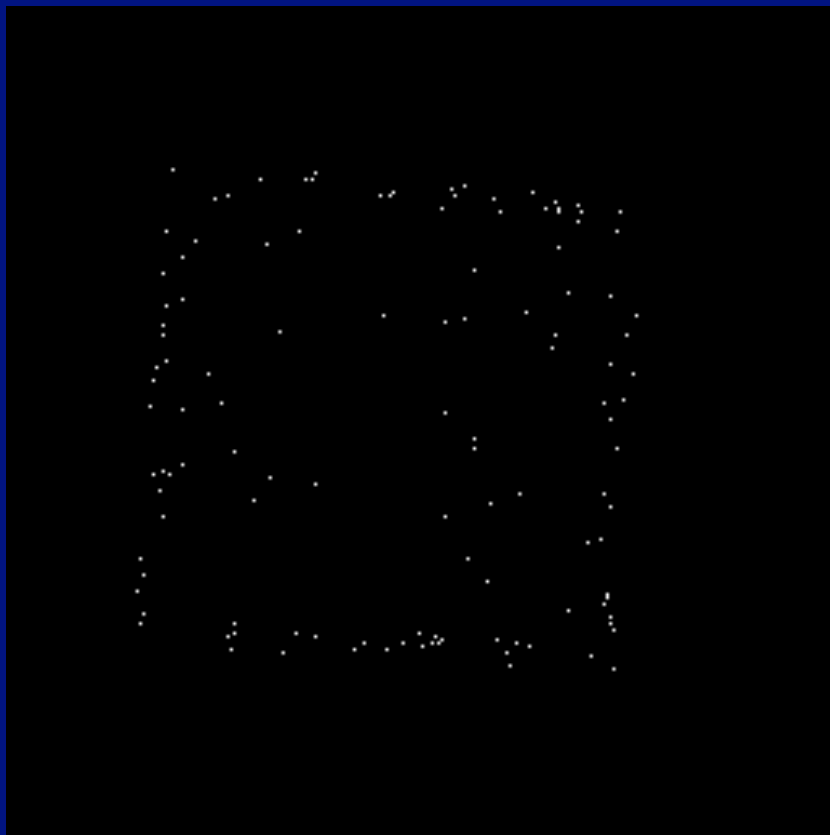
Pinhole camera - an effective abstraction



What happens in two views



Structure from motion

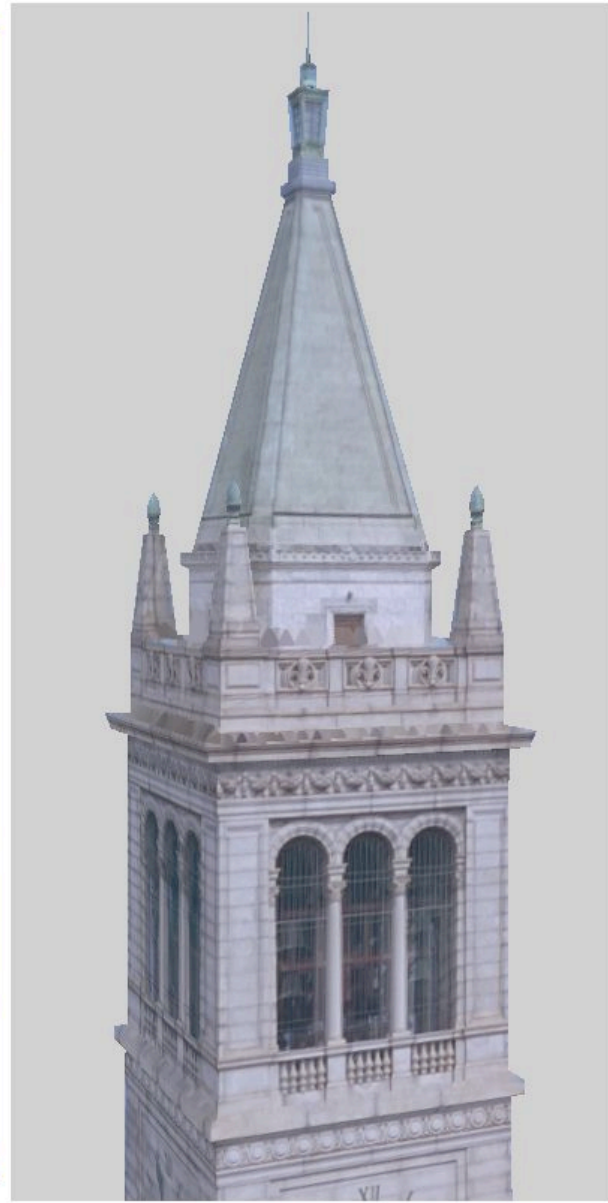


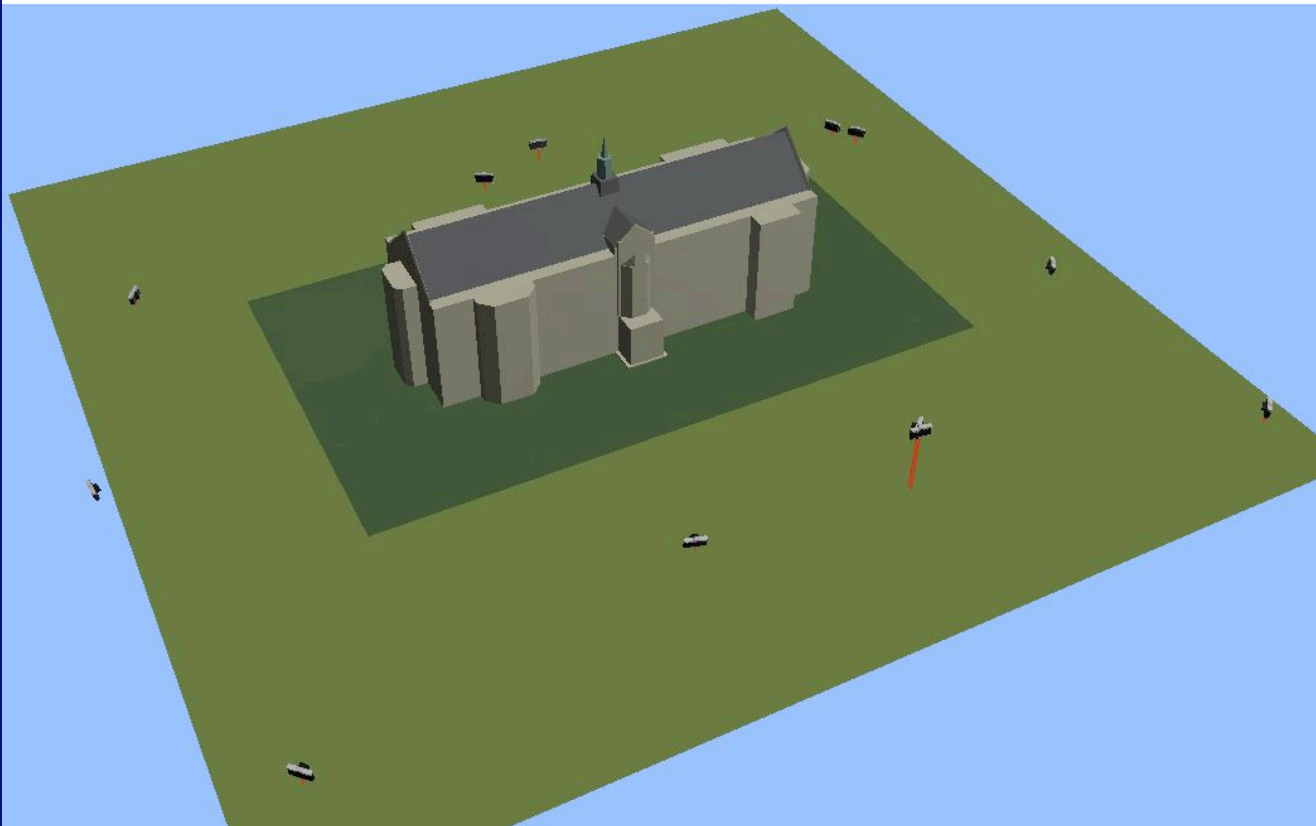
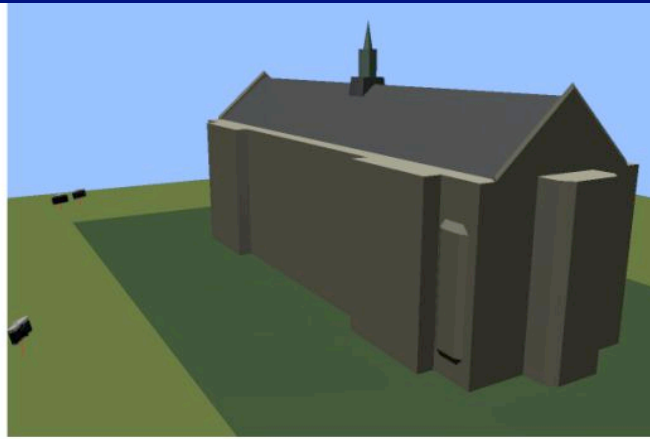
All of Camera Geometry

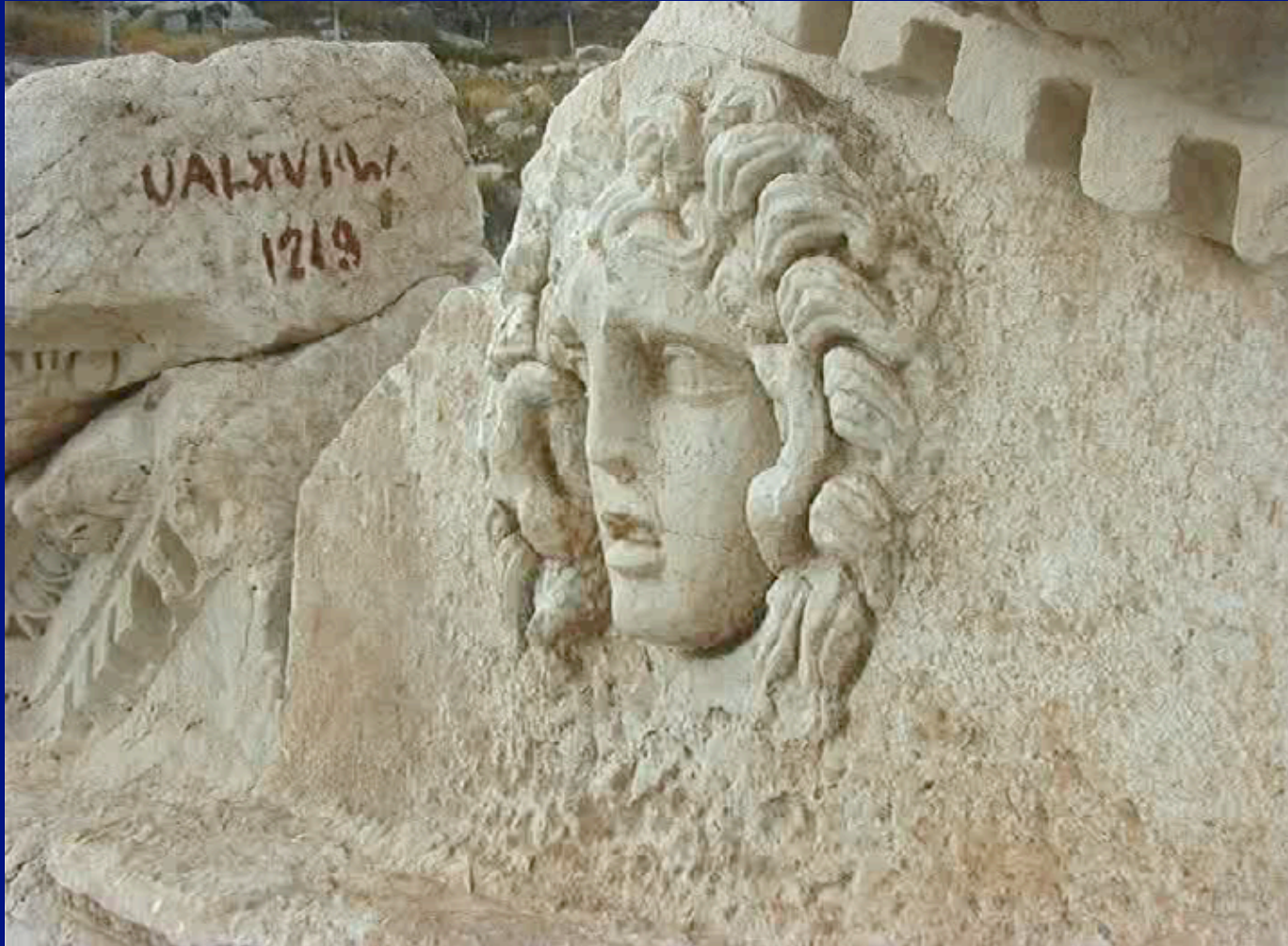
- From the picture
 - two views of a point give four measurements of three DOF
 - this means
 - correspondence is constrained
 - if we have enough points and enough pix we can recover
 - points
 - cameras



Work by Paul Debevec and Jitendra Malik



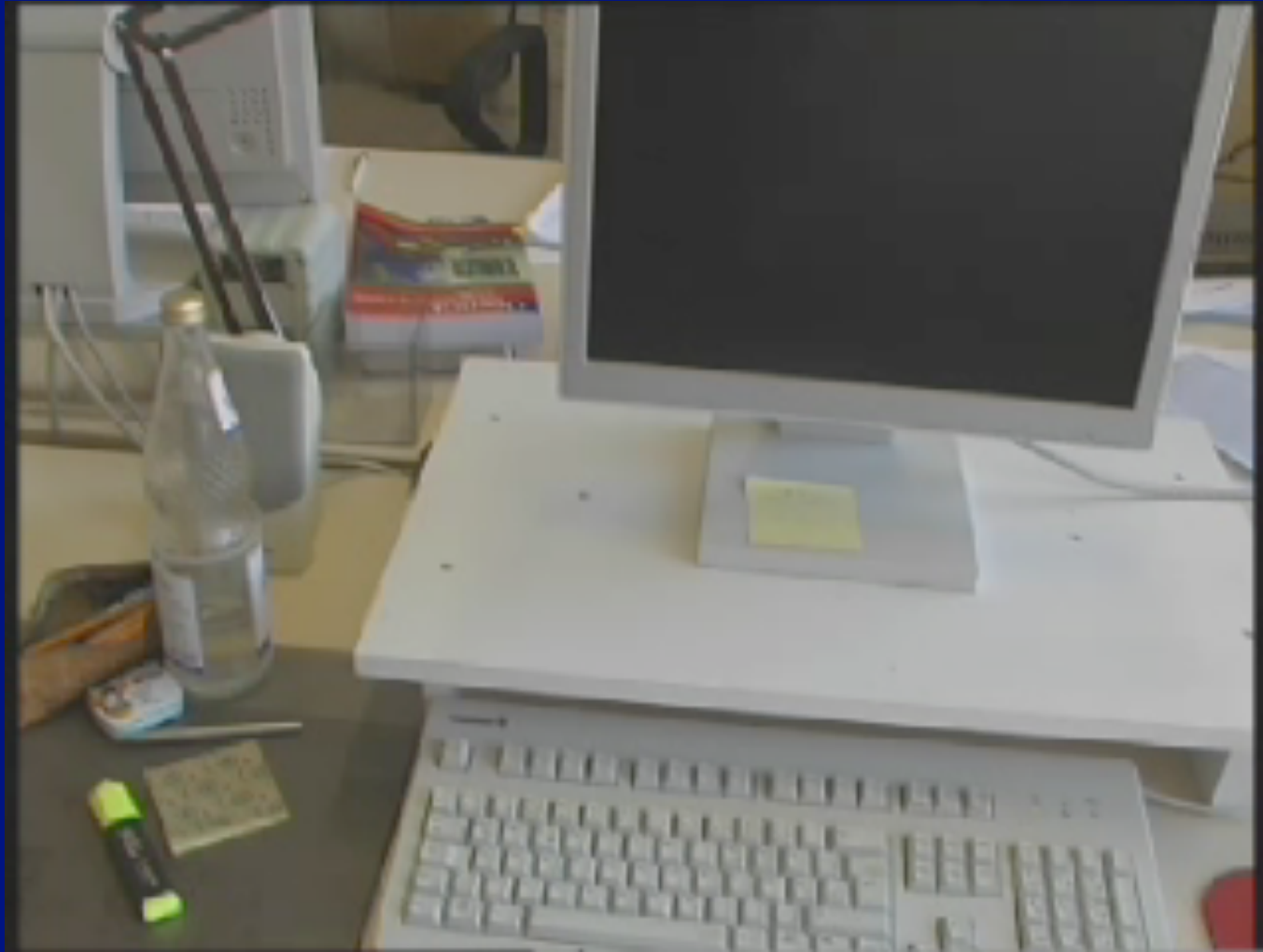


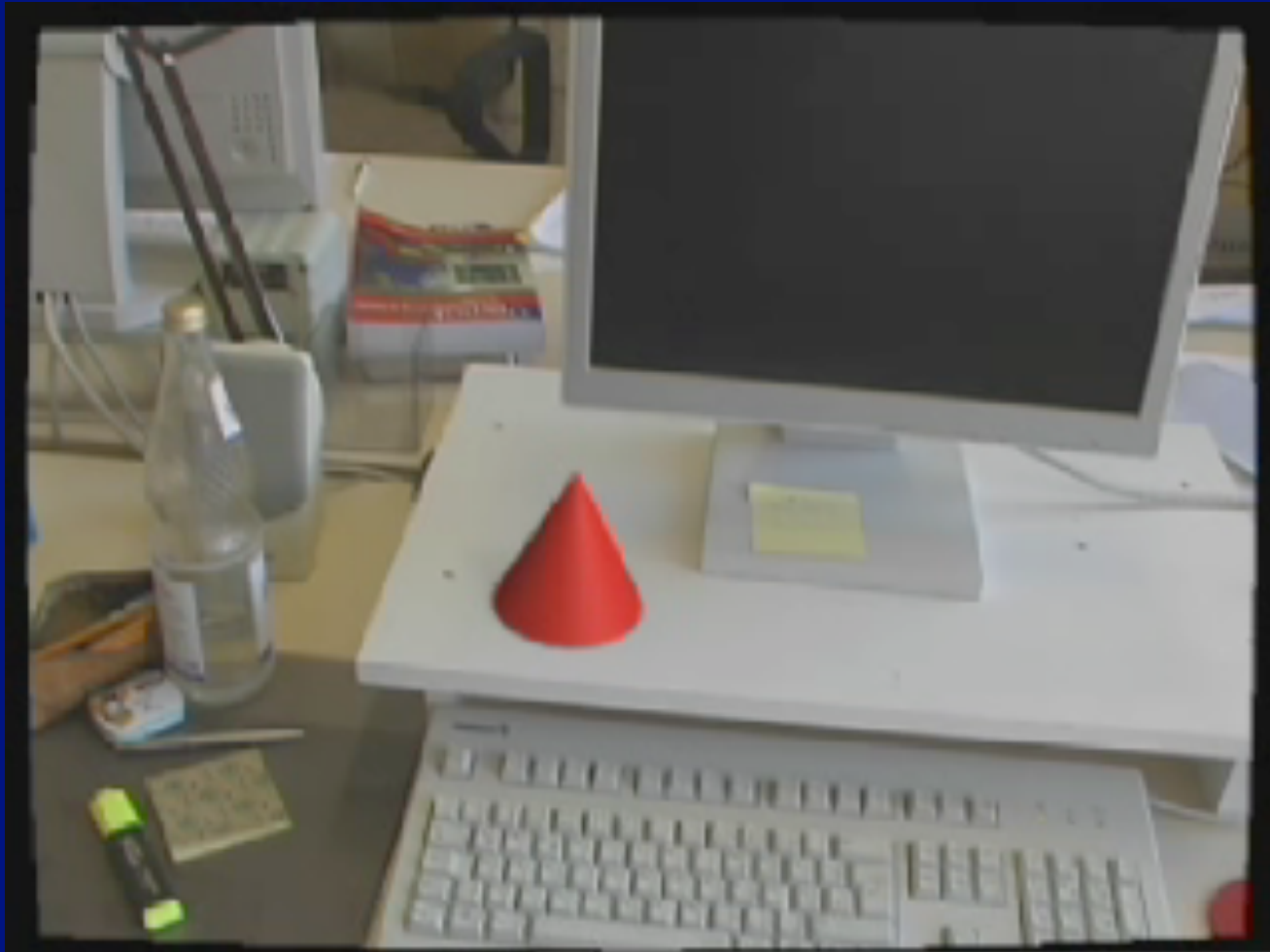


M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, Visual modeling with a hand-held camera, *International Journal of Computer Vision* 59(3), 207-232, 2004

Match moves

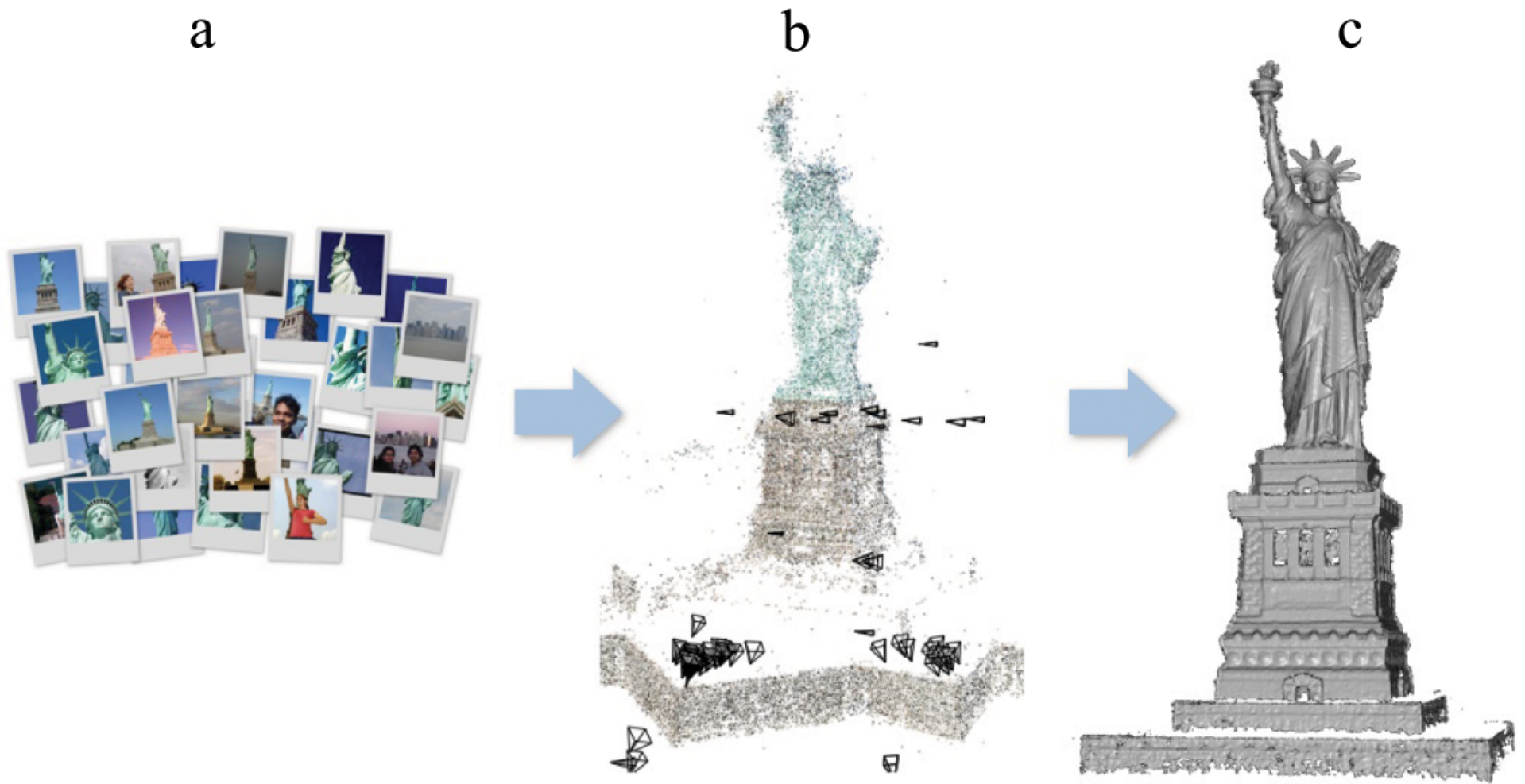
- If you know where the camera is for a sequence you can
 - attach a virtual sequence to the end smoothly
 - insert virtual objects realistically
- Commercial products
 - boujou, MatchMover, 3DEqualizer
- Free research software
 - voodoo







Capturing and animating occluded cloth - R White, K Crane, DA Forsyth SIGGRAPH 2007



M Goesele, N Snavely, B Curless, H Hoppe, "Multi-view stereo for community photo collections",
ICCV 2007

Photo Tourism

Exploring photo collections in 3D

Noah Snavely Steven M. Seitz Richard Szeliski
University of Washington *Microsoft Research*

SIGGRAPH 2006

Noah Snavely, Steven M. Seitz, Richard Szeliski, “Photo tourism: Exploring photo collections in 3D,” ACM Transactions on Graphics (SIGGRAPH Proceedings), 25(3), 2006, 835-846.

Reconstructions from one view

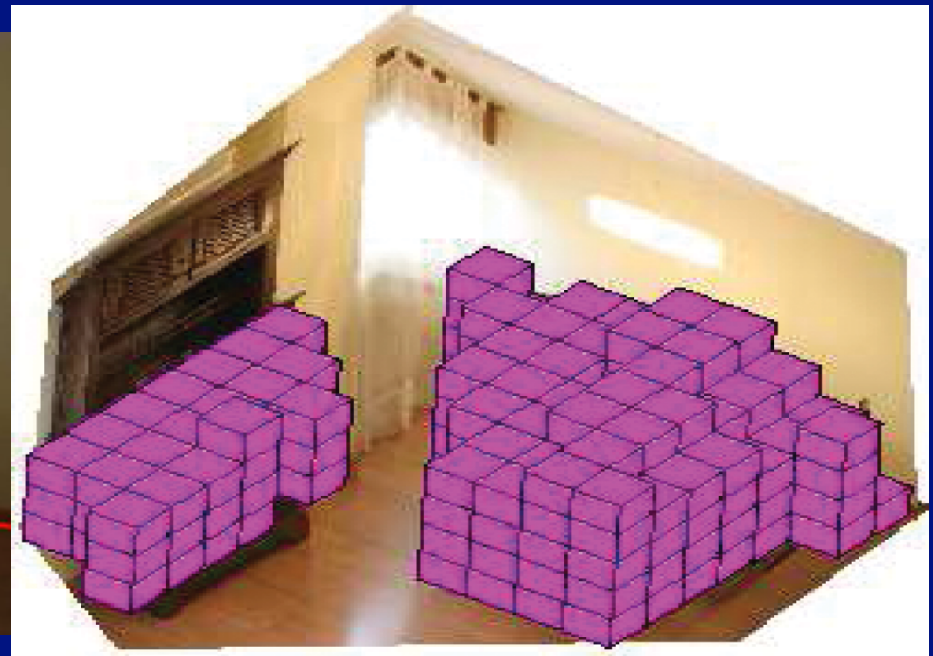
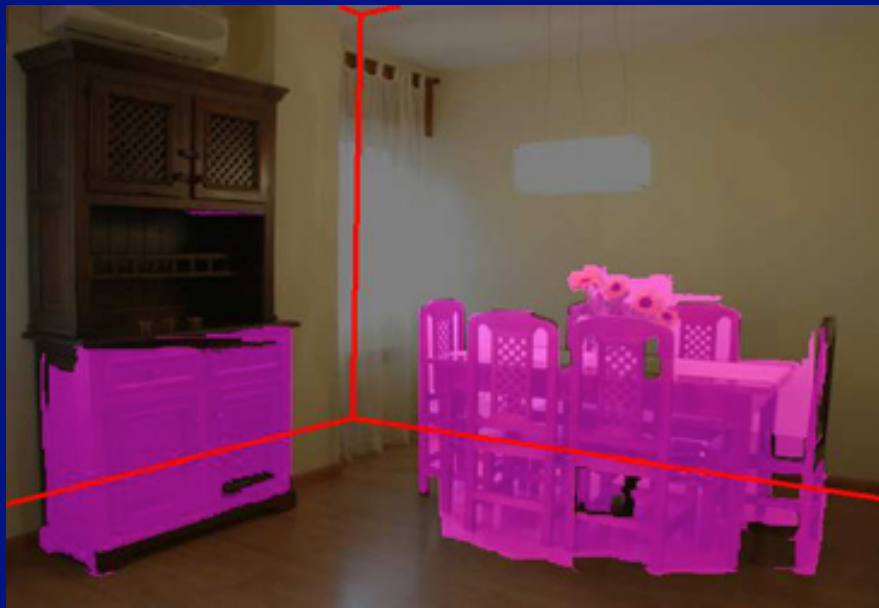
- Symmetry
- Horizontal / vertical planes and contours
- Special geometries
- Texture
- Shading

Automatic Photo Pop-up

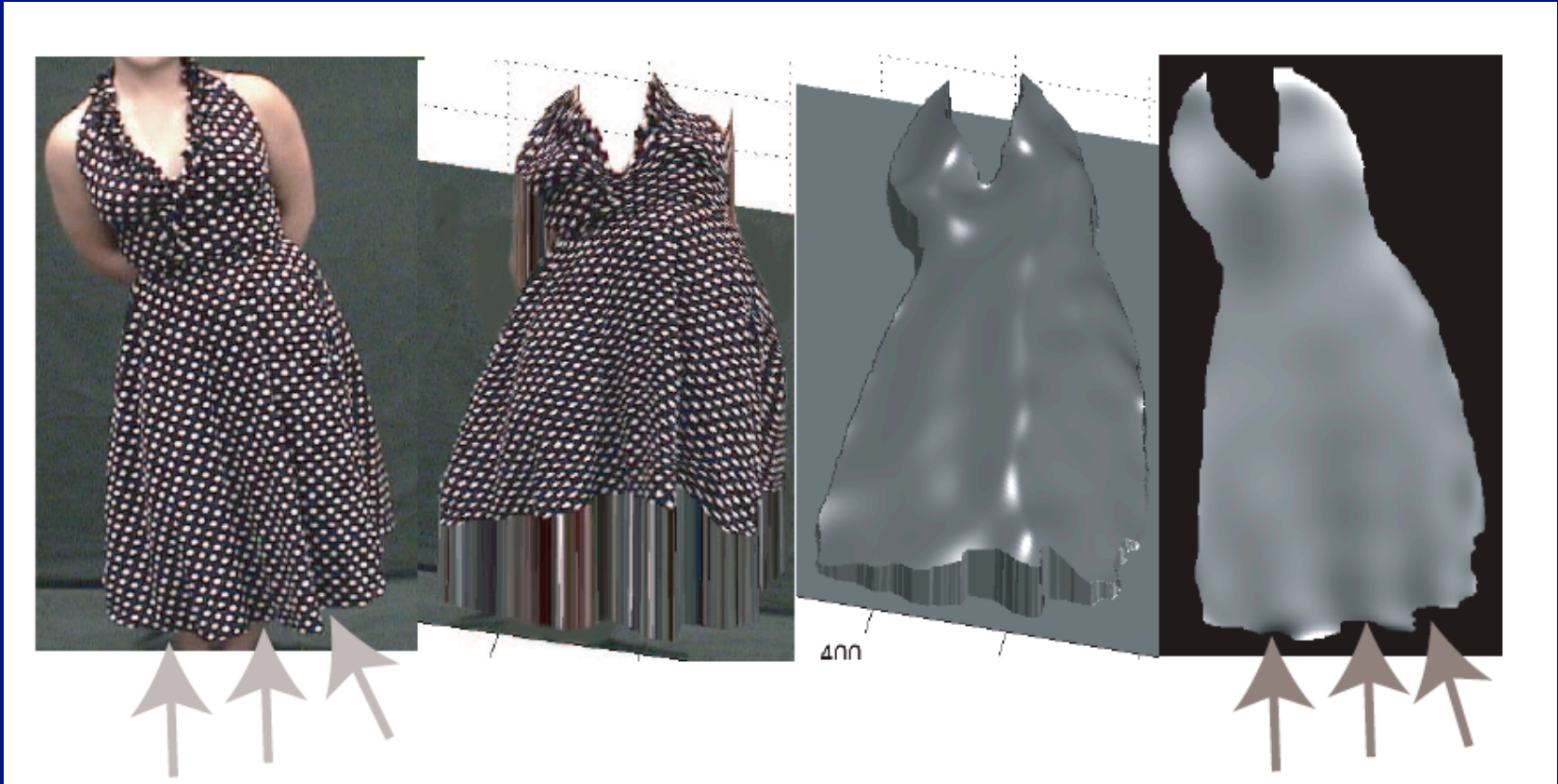
D. Hoiem A.A. Efros M. Hebert
Carnegie Mellon University



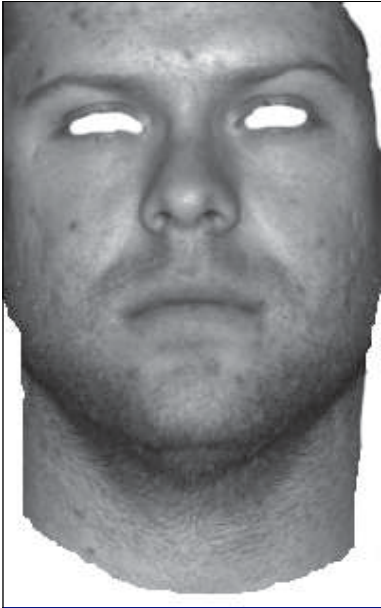
V. Hedau, D. Hoiem, D.A. Forsyth, "Recovering the layout of cluttered rooms", ICCV 2009



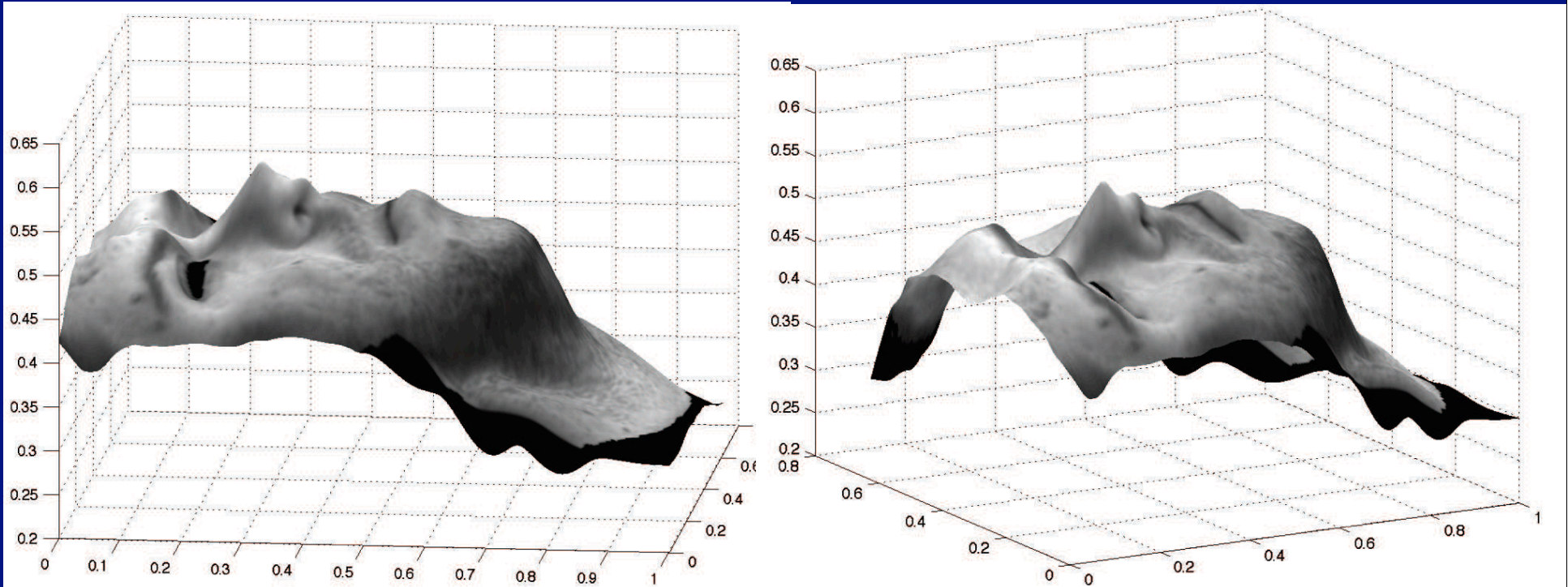
V. Hedau, D. Hoiem, D.A. Forsyth, "Recovering the layout of cluttered rooms", ICCV 2009



A. Lobay and D.A. Forsyth, "Recovering shape and irradiance maps from rich dense texture fields",
CVPR 2004



D.A. Forsyth, "Variable source shading analysis," In review

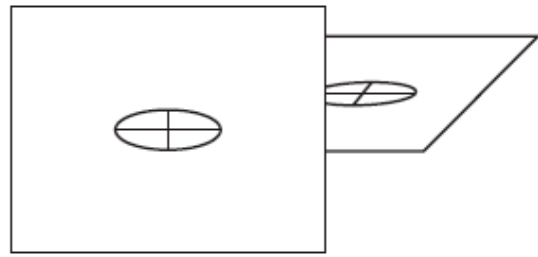




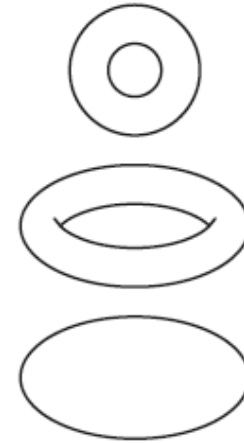
M.K. Johnson and E.H. Adelson “Retrographic sensing for the measurement of surface texture and shape”
CVPR, 2009

Recognition

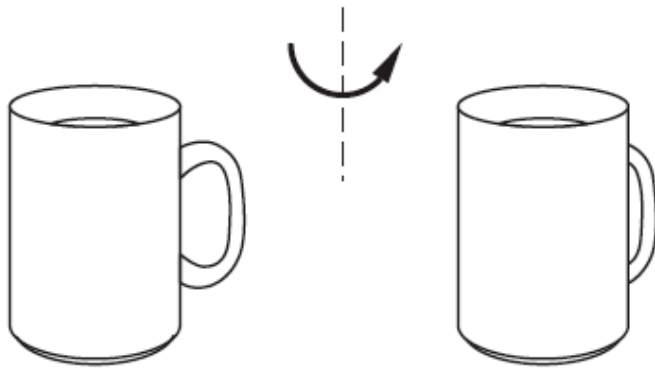
- Problem is somewhat vague
 - what is this
 - what is this like
- Major technologies
 - Classifier
 - stick in a feature, and it says yes/no
 - Rich local feature descriptors (like SIFT)



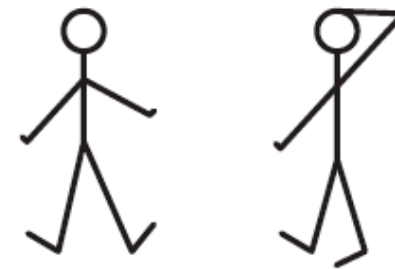
Foreshortening



Aspect



Occlusion

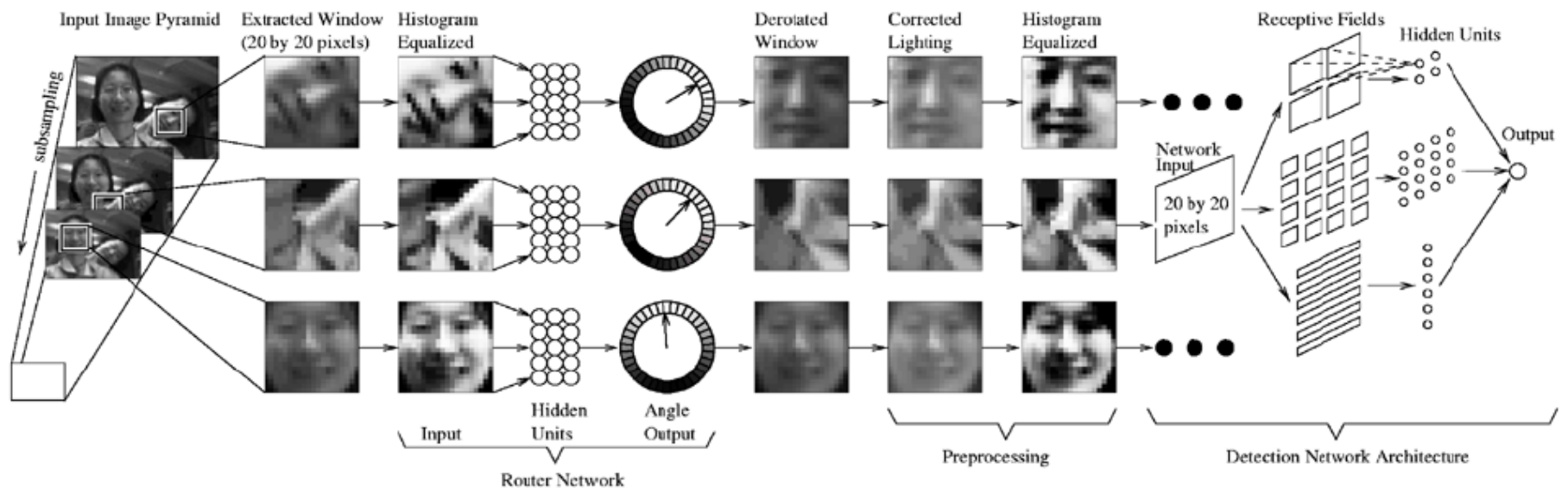


Deformation

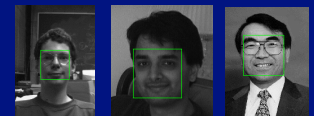
Finding faces

- Faces “look like” templates (at least when they’re frontal).
- General strategy:
 - search image windows at a range of scales
 - Correct for illumination
 - Present corrected window to classifier
- Issues
 - How corrected?
 - What features?
 - What classifier?
 - what about lateral views?

Rowley-Baluja-Kanade face finder (1)

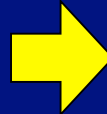


“Rotation invariant neural-network based face detection,”
H.A. Rowley, S. Baluja and T. Kanade, CVPR 1998



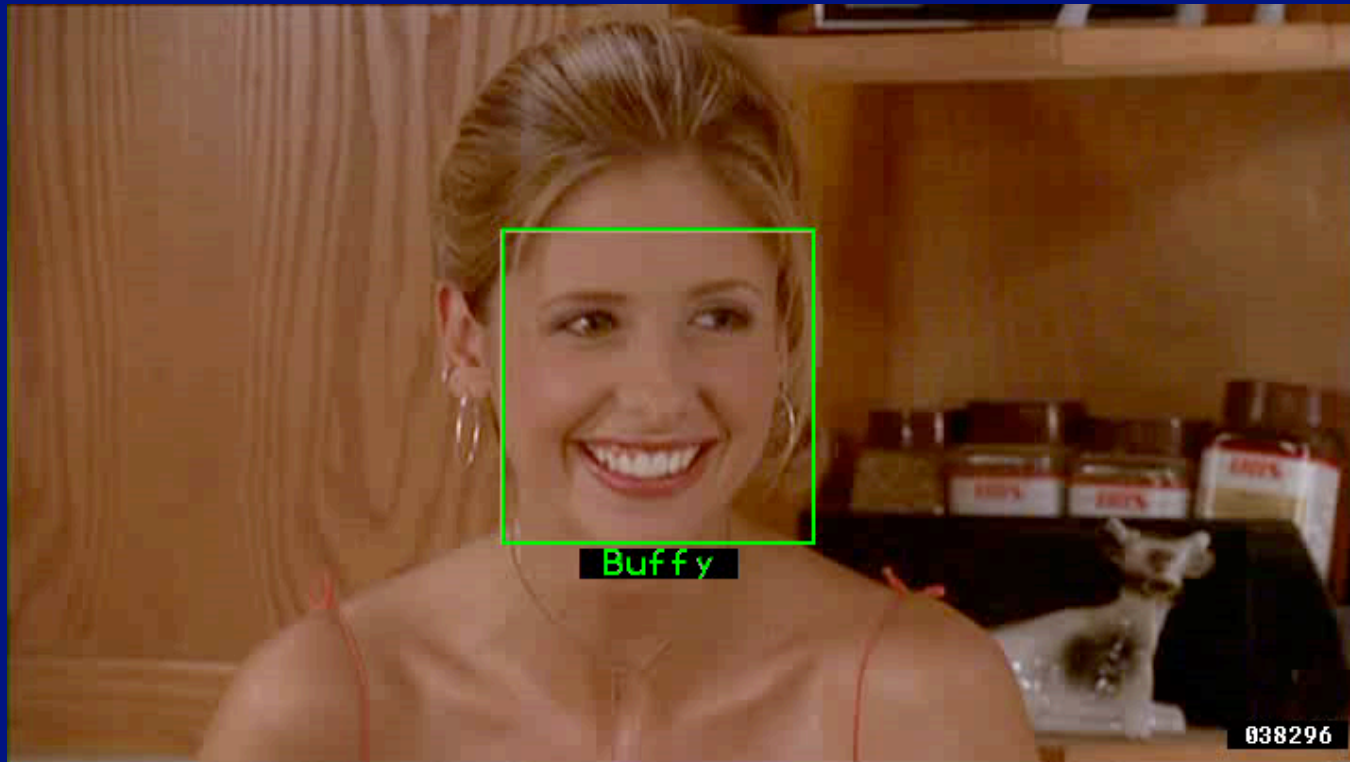


“Rotation invariant neural-network based face detection,”
 H.A. Rowley, S. Baluja and T. Kanade, CVPR 1998

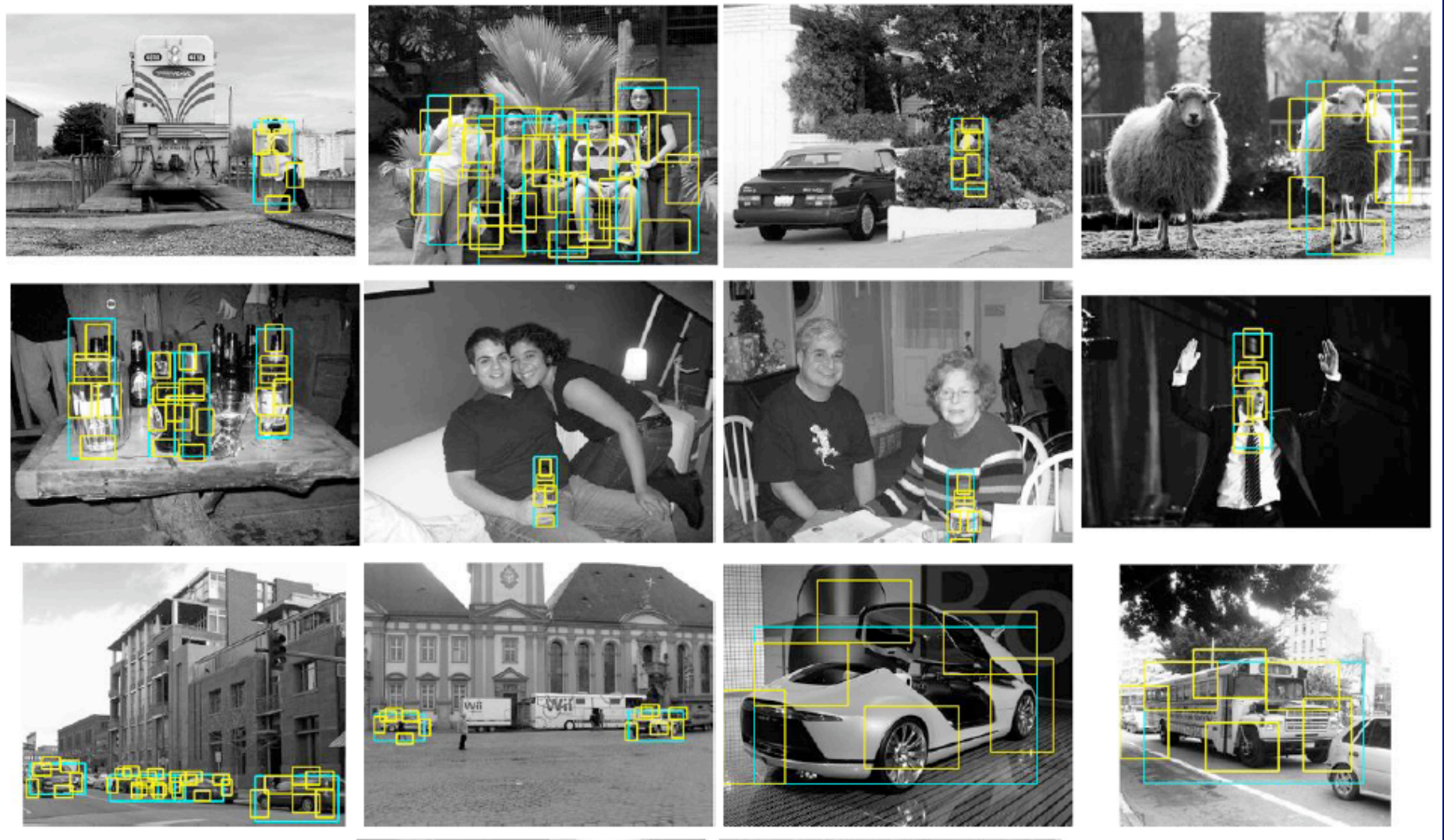


President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters

Names and faces in the news - T. Berg et al CVPR 2004



Everingham, M., Sivic, J. and Zisserman, A.
“Hello! My name is... Buffy” - Automatic naming of characters in TV video
BMVC 2006



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

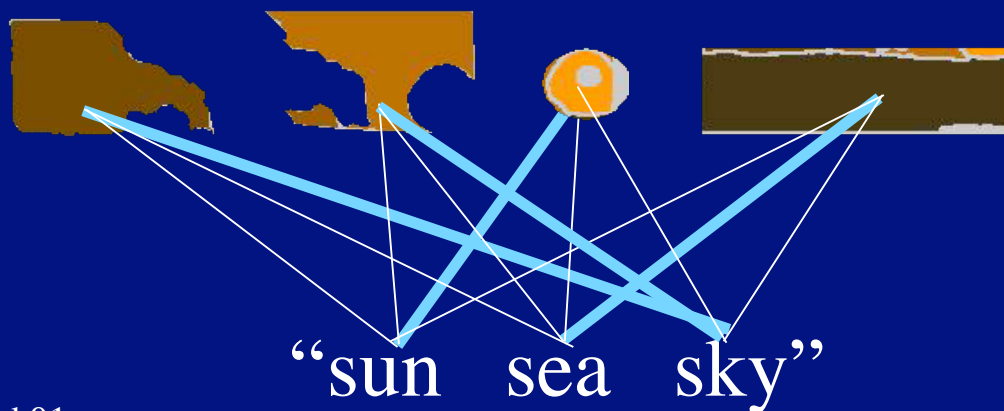
Linking words and pictures

- In its simplest form, missing variable problem
- Caveats
 - might take a lot of data; symmetries, biases in data create issues

“the beautiful sun”



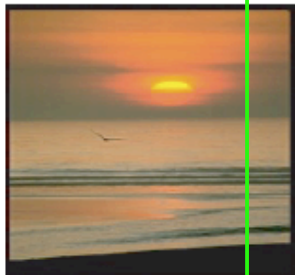
“le soleil beau”



Brown, Della Pietra, Della Pietra & Mercer 93; Melamed 01

Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary -
P Duygulu, K Barnard, JFG de Freitas, DA Forsyth ECCV 2002

It was there and we didn't



sky, sun, clouds, sea, waves, birds, water



tree, people, sand, road, stone, statue, temple, sculpture, pillar



tree, birds, snow, fly



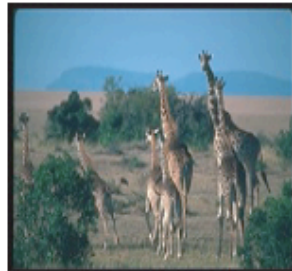
sky, water, tree, plane, elephant, herd



mountain, sky, water, clouds, tree



sky, sun, jet, plane



mountain, sky, water, tree, grass, plane, ground, giraffe



water, people, pool, swimmers



tree, people, shadows, road, stone, statue, sculpture, pillar

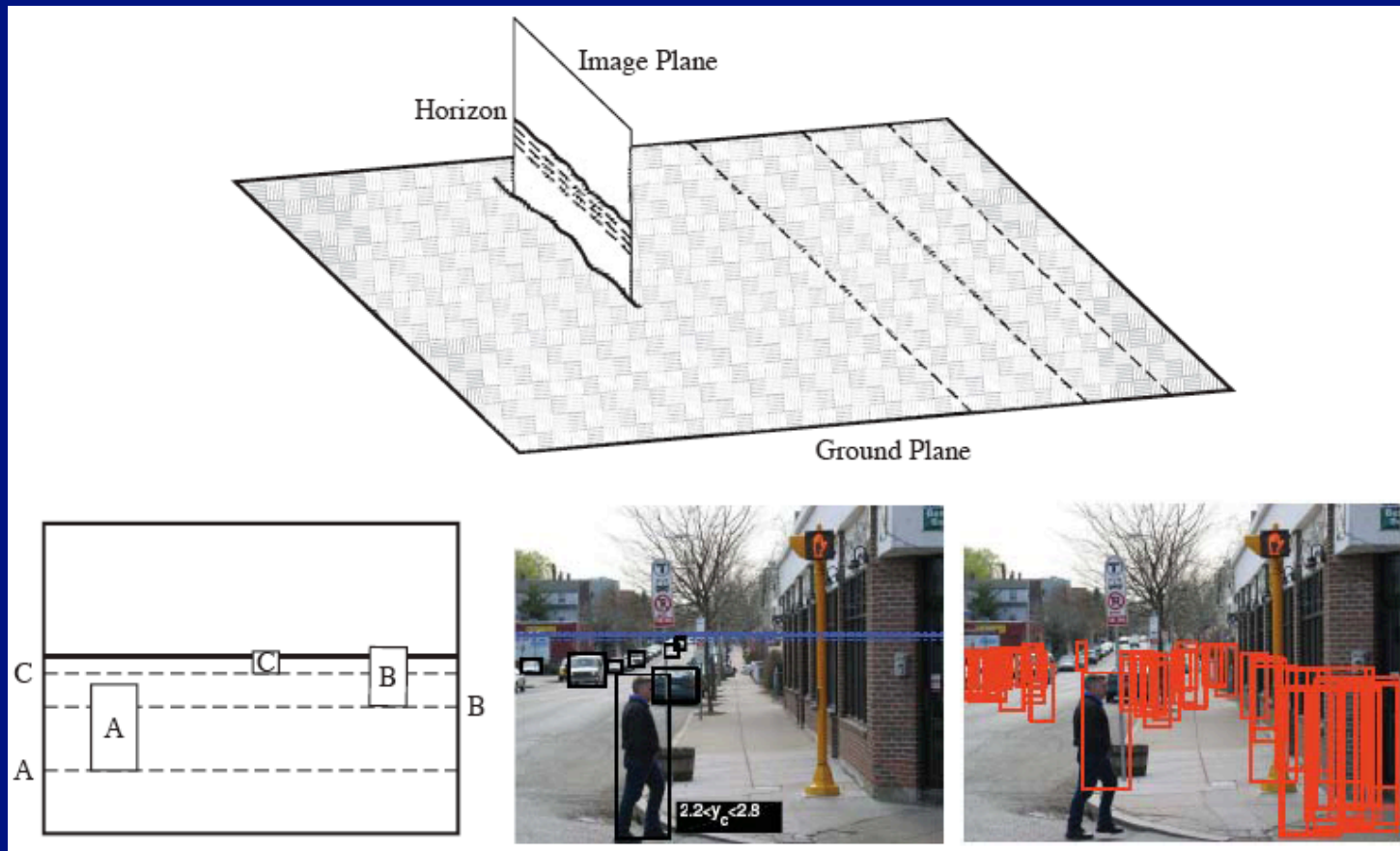


people, buildings, stone, temple, sculpture, pillar, mosque

It was there and we predicted it

It wasn't and we did

Scene Discovery by Matrix Factorization, N Loeff, A Farhadi, ECCV 2008



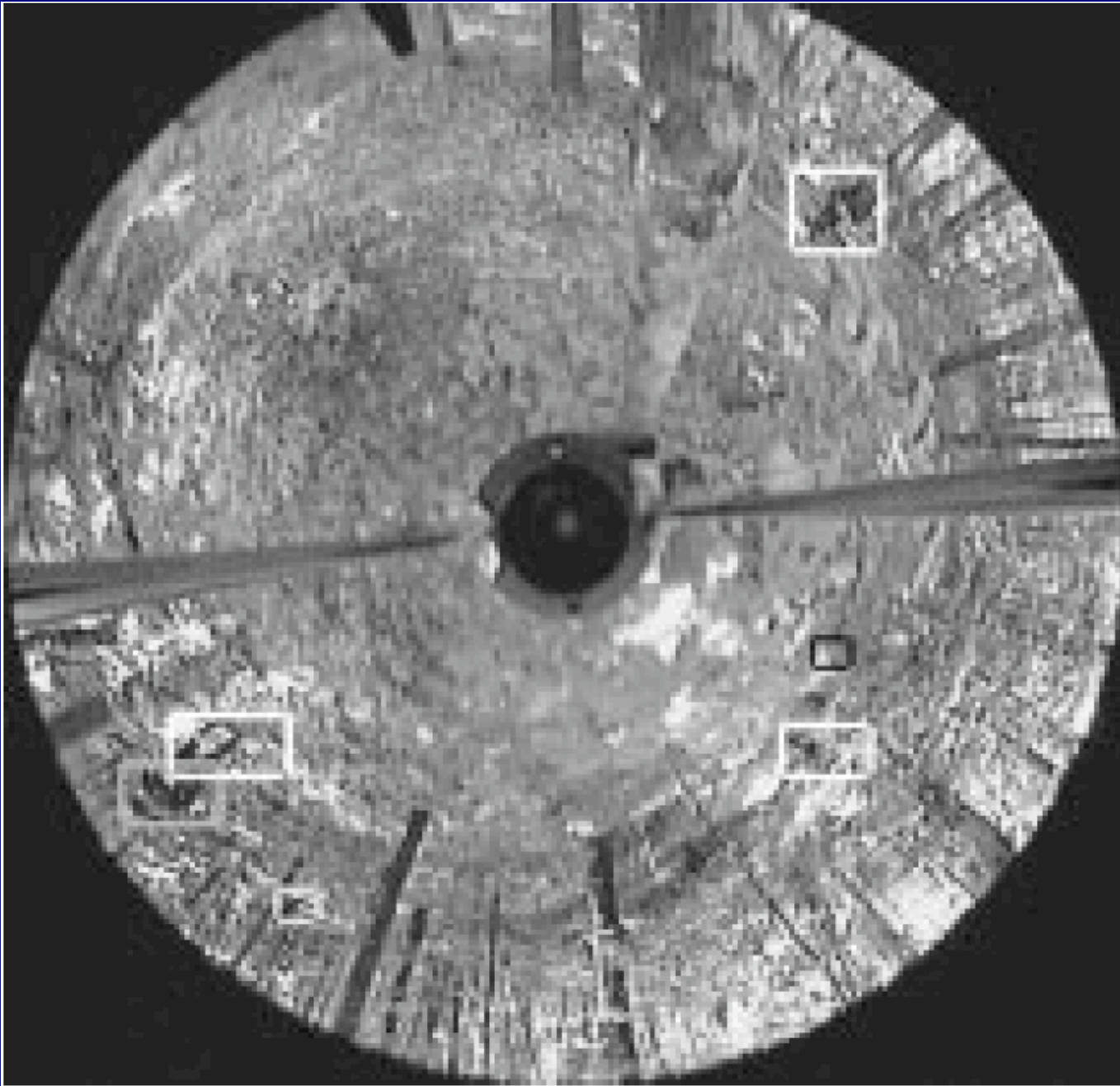
Putting Objects in Perspective, D. Hoiem, A.A. Efros, and M. Hebert, in Proc IEEE Conf. Computer Vision and Pattern Recognition, 2006

Looking at People

- What are people doing?
- Hard because
 - hard to extract people from video
 - but quite good trackers are now becoming available
 - hard to know what form the answer takes
 - “walking” vs “running”?
 - “not much of interest” vs “might be a problem”?
 - goals/intentions?

Why are humans important?

- **Surveillance**
 - prosecution; intelligence gathering; crime prevention
 - HCI; architecture;
- **Synthesis**
 - games; movies;
- **Safety applications**
 - pedestrian detection
- **People are interesting**
 - movies; news



Where you are can suggest
you are doing something
you shouldn't be
Boult 2001



Bill Freeman flies a magic carpet.

Orientation histograms detect body configuration to control bank, raised arm to fire magic spell.

Freeman et al, 98.



9 An example of a user playing a Decathlon event, the javelin throw. The computer's timing of the set and release for the javelin is based on when the integrated downward and upward motion exceeds predetermined thresholds.

Motion fields set javelin timing
Freeman et al 98

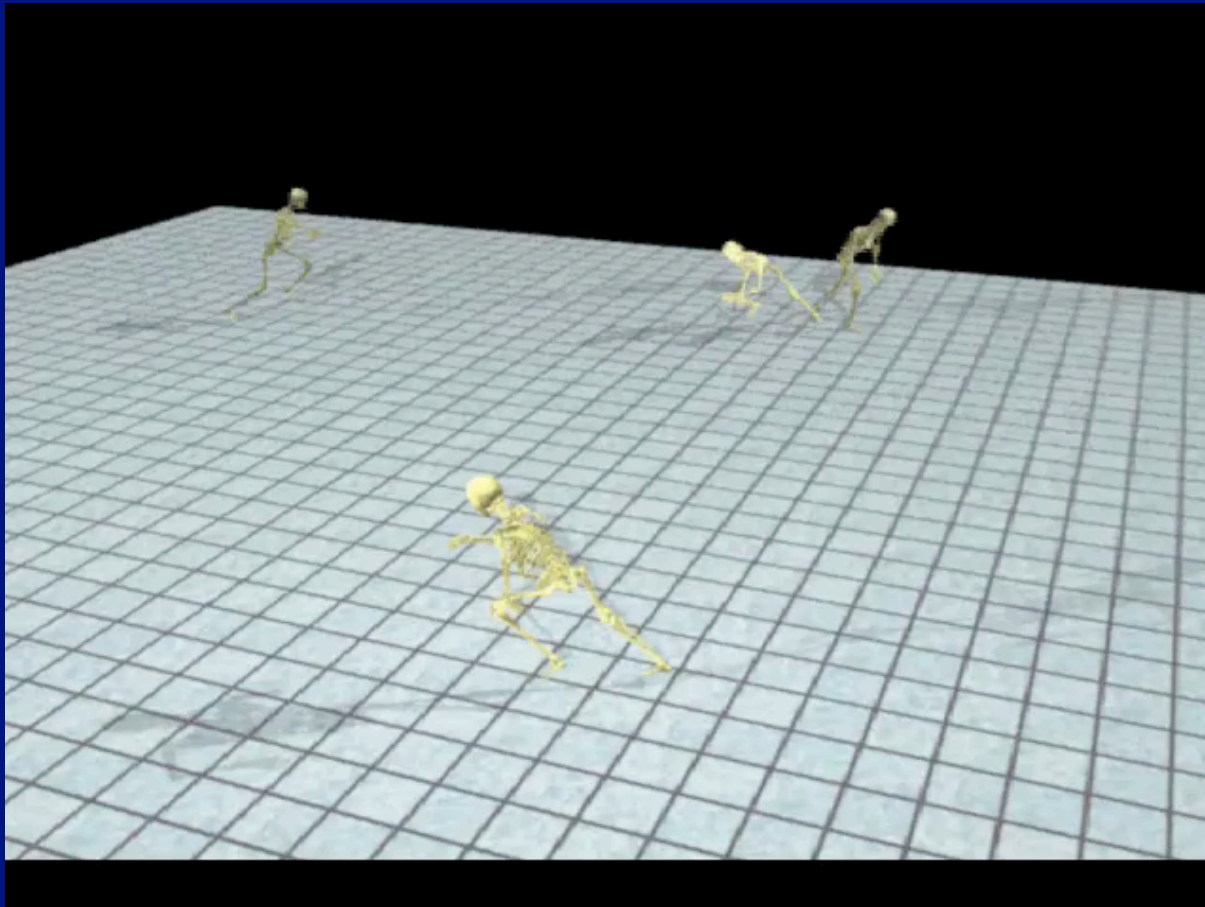


Sony's eyetoy estimates motion fields,
links these to game inputs.
Huge hit in EU, well received in US



Why are humans important?

- Surveillance
 - prosecution; intelligence gathering; crime prevention
 - HCI; architecture;
- **Synthesis**
 - games; movies;
- Safety applications
 - pedestrian detection
- People are interesting
 - movies; news

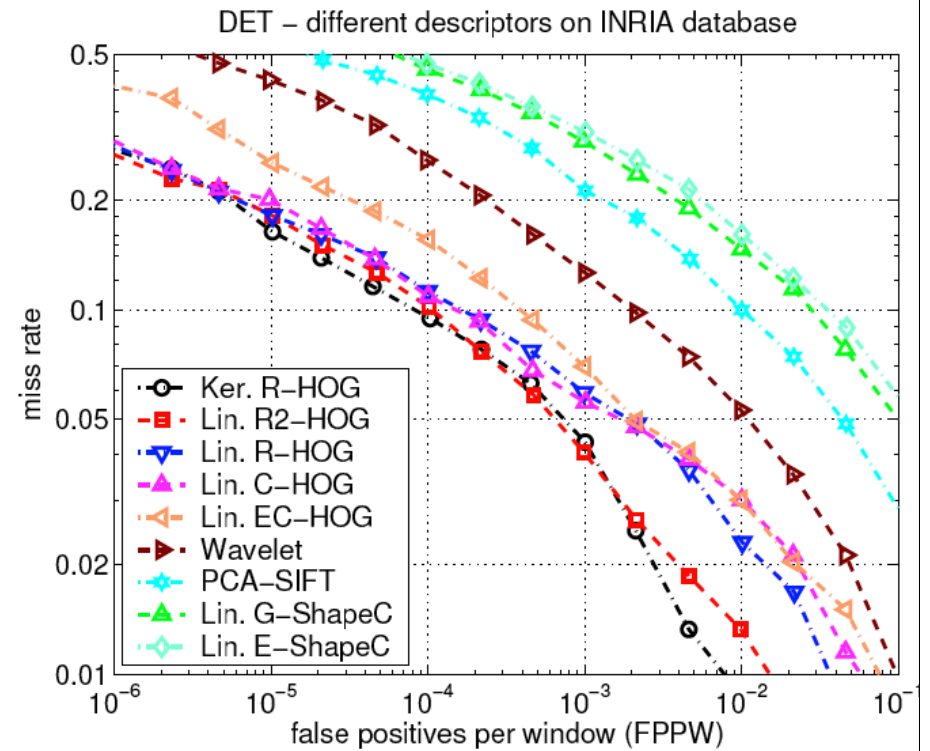
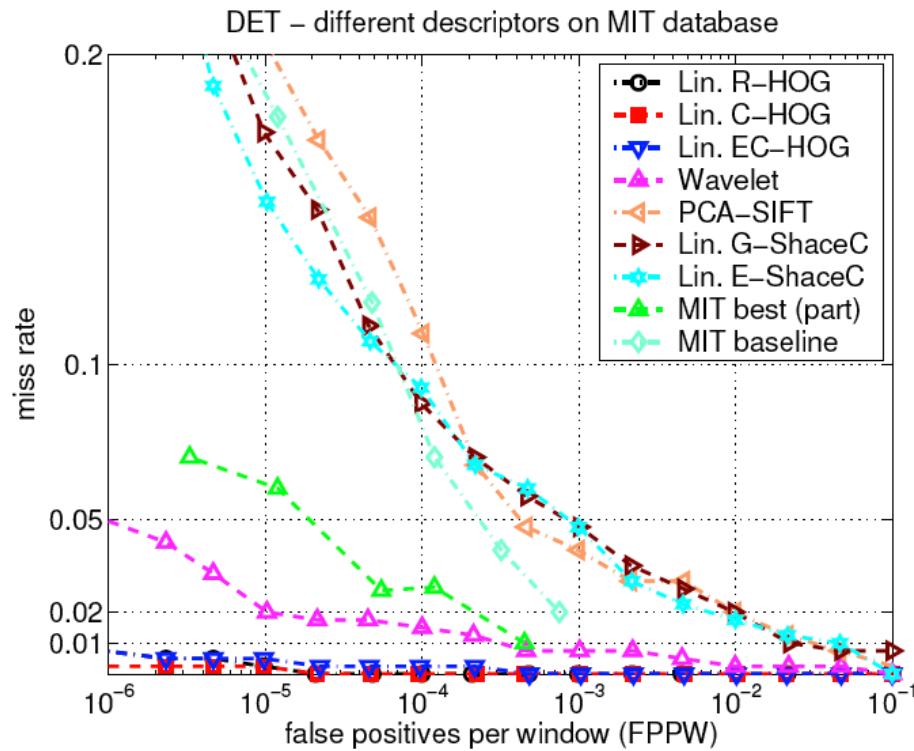


Why are humans important?

- Surveillance
 - prosecution; intelligence gathering; crime prevention
 - HCI; architecture;
- Synthesis
 - games; movies;
- **Safety applications**
 - pedestrian detection
- People are interesting
 - movies; news



From Dalal+Triggs, 05



Why are humans important?

- Surveillance
 - prosecution; intelligence gathering; crime prevention
 - HCI; architecture;
- Synthesis
 - games; movies;
- Safety applications
 - pedestrian detection
- **People are interesting**
 - movies; news

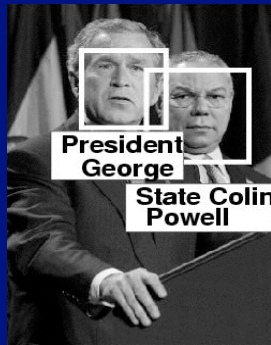
News Faces

- 5e5 captioned news images
- Mainly people “in the wild”
- Correspondence problem
 - some images have many (resp. few) faces, few (resp. many) names (cf. Srihari 95)
- Process
 - Extract proper names
 - Detect faces (Vogelhuber Schmid 00) 44773 big face responses
 - Rectify faces 34623 properly rectified
 - Kernel PCA rectified faces
 - Estimate linear discriminants
 - Now have (face vector; name_1, ..., name_k)
27742 for $k \leq 4$
- Apply a form of modified k-means



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





US President George W. Bush (L) makes remarks while Secretary of State Colin Powell (R) listens before signing the US Leadership Against HIV /AIDS , Tuberculosis and Malaria Act of 2003 at the Department of State in Washington, DC. The five-year plan is designed to help prevent and treat AIDS, especially in more than a dozen African and Caribbean nations(AFP/ Luke Frazza)



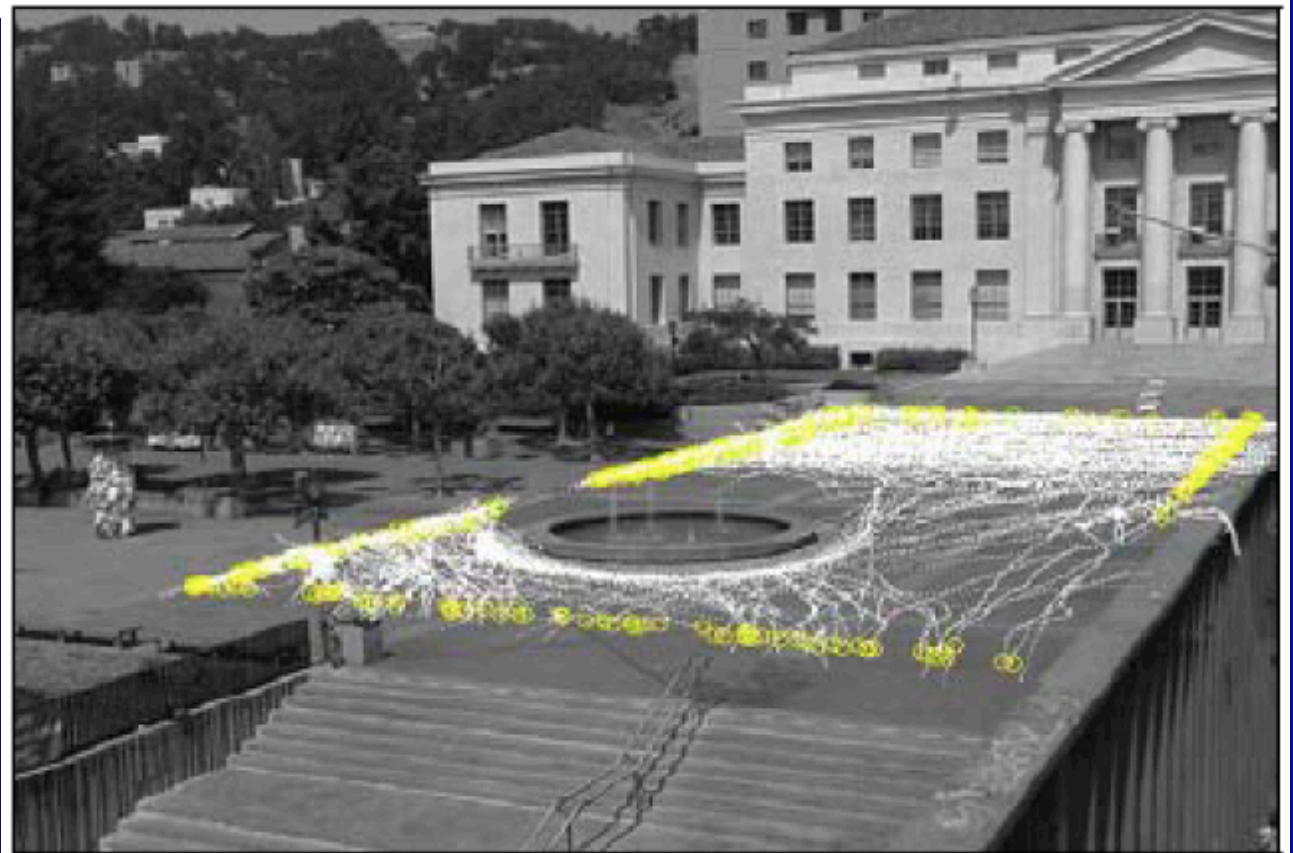
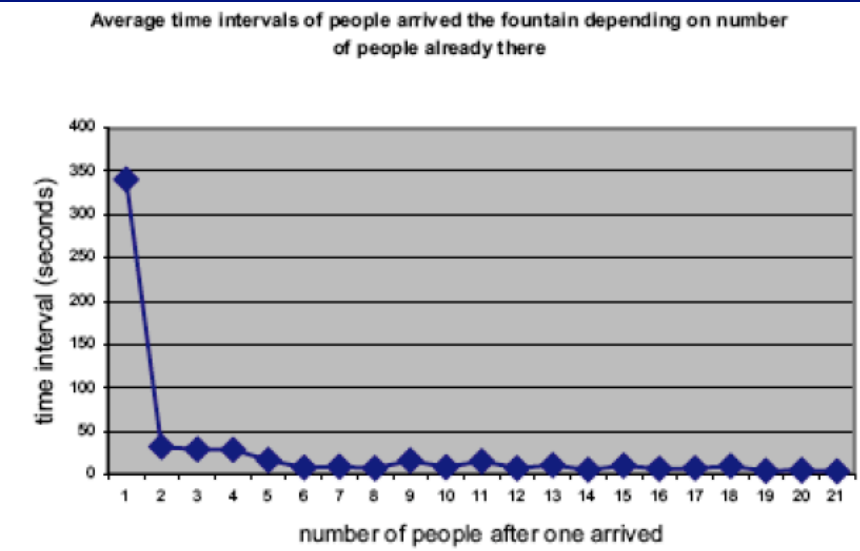
German supermodel Claudia Schiffer gave birth to a baby boy by Caesarian section January 30, 2003, her spokeswoman said. The baby is the first child for both Schiffer, 32, and her husband, British film producer Matthew Vaughn, who was at her side for the birth. Schiffer is seen on the German television show 'Bet It...?!' ('Wetten Dass...?!') in Braunschweig, on January 26, 2002. (Alexandra Winkler/Reuters)

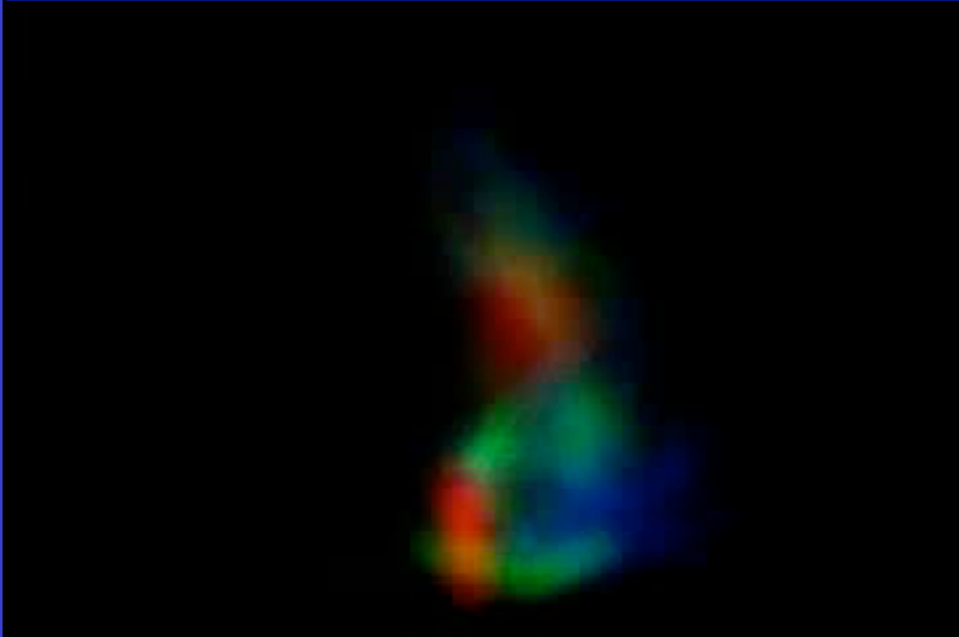


British director Sam Mendes and his partner actress Kate Winslet arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The films stars Tom Hanks as a Chicago hit man who has a separate family life and co-stars Paul Newman and Jude Law. REUTERS/Dan Chung

Curious phenomena in public spaces

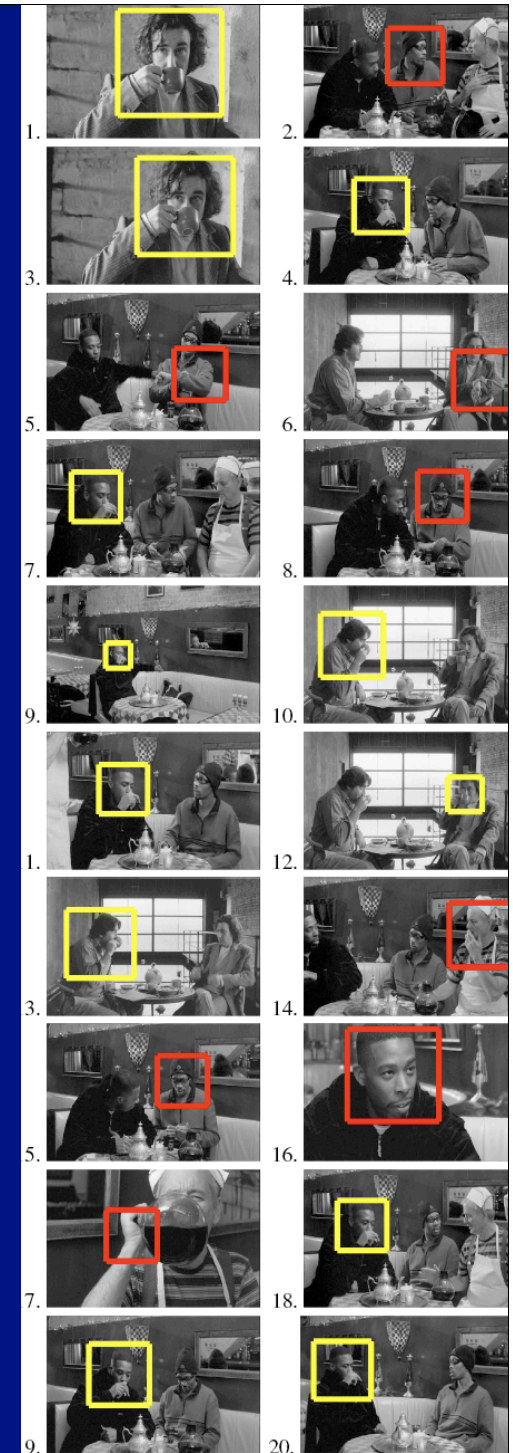
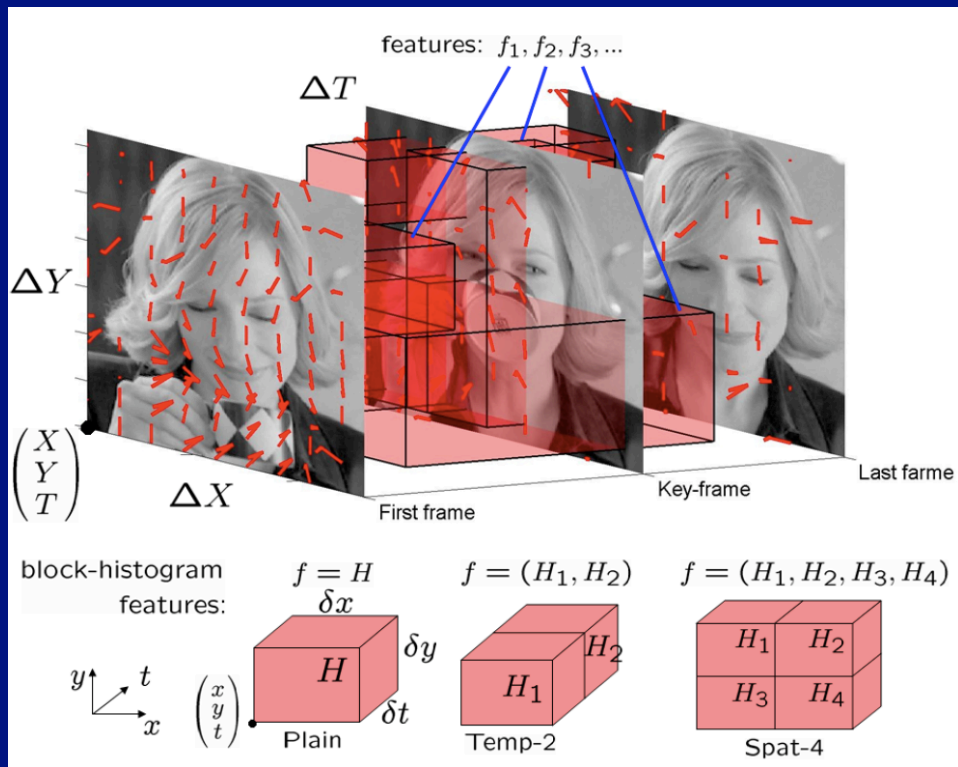
Yan+Forsyth, 04





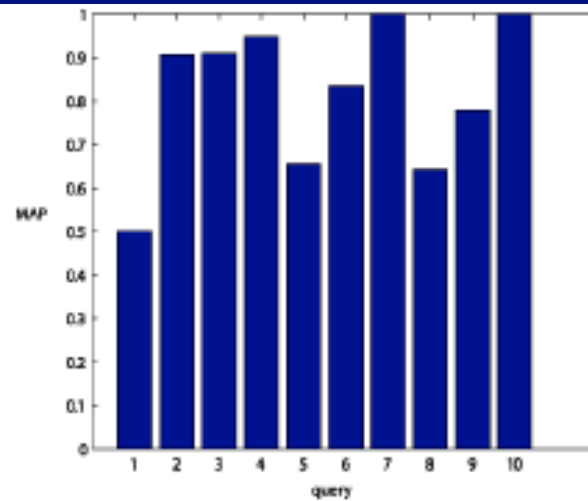
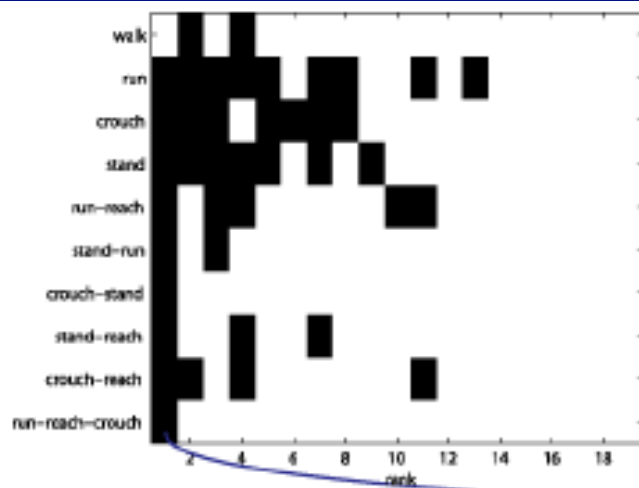


Ramanan, Forsyth and Zisserman CVPR05



Retrieving actions in movies

I Laptev, P Perez - International Conference on Computer Vision, 2007

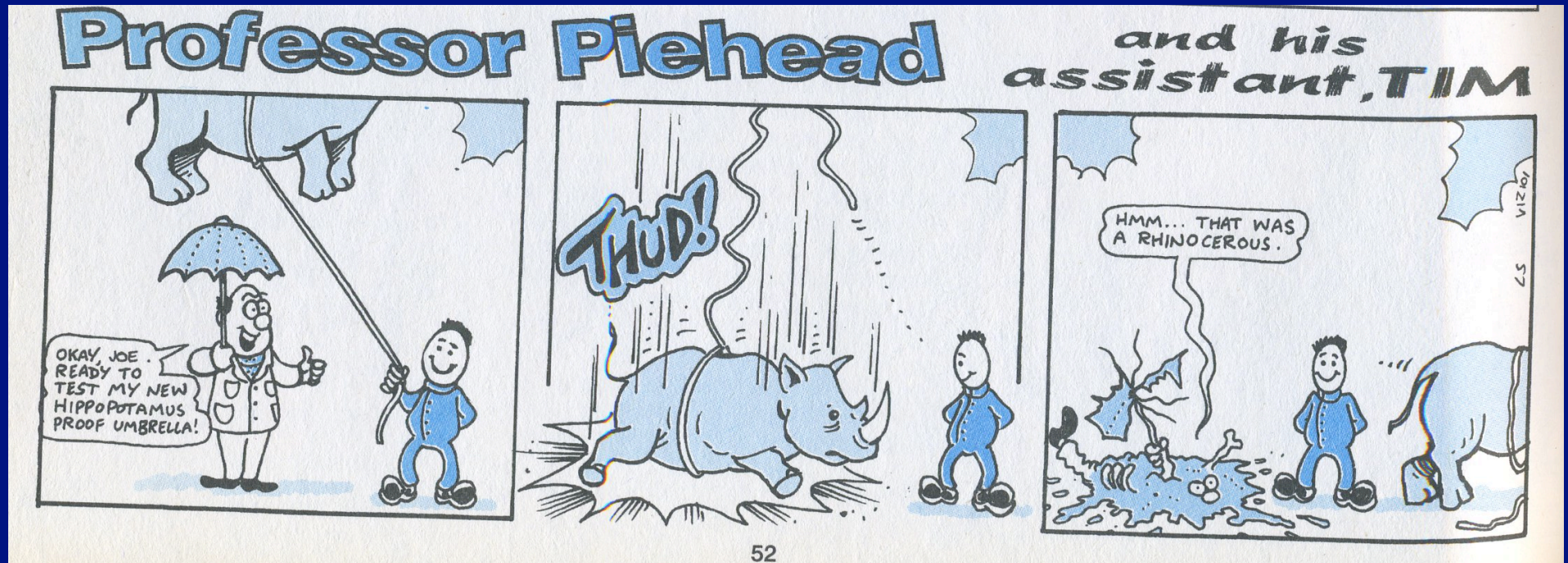


the first video retrieved for query "run-reach-couch"

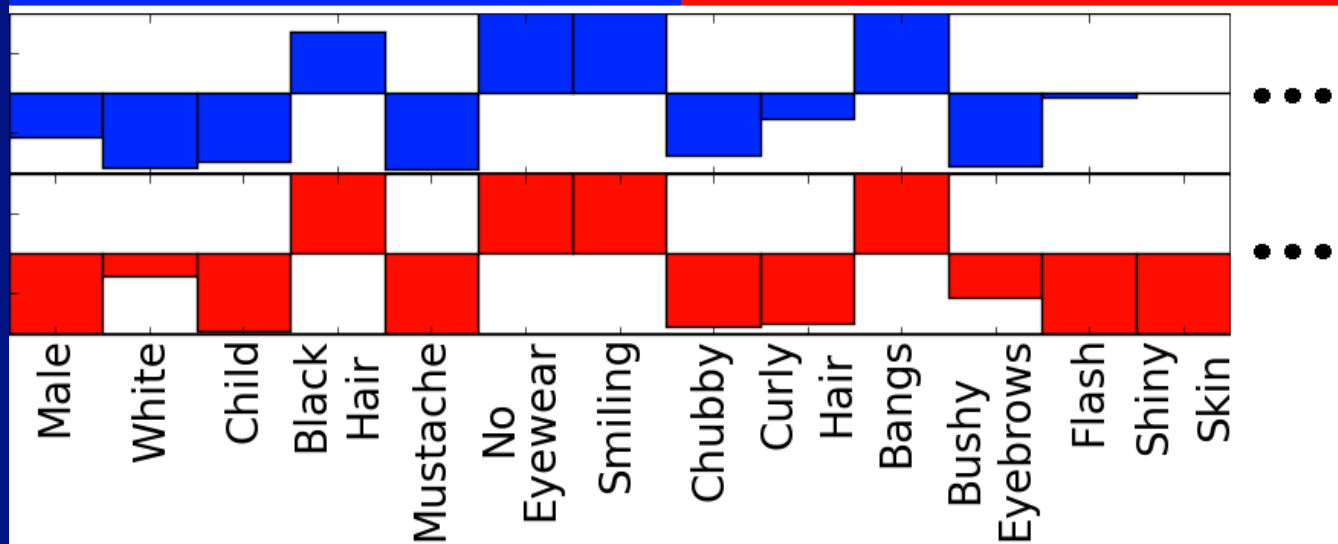


Searching for complex human activities with no visual examples N İkizler, DA Forsyth - IJCV, 2008

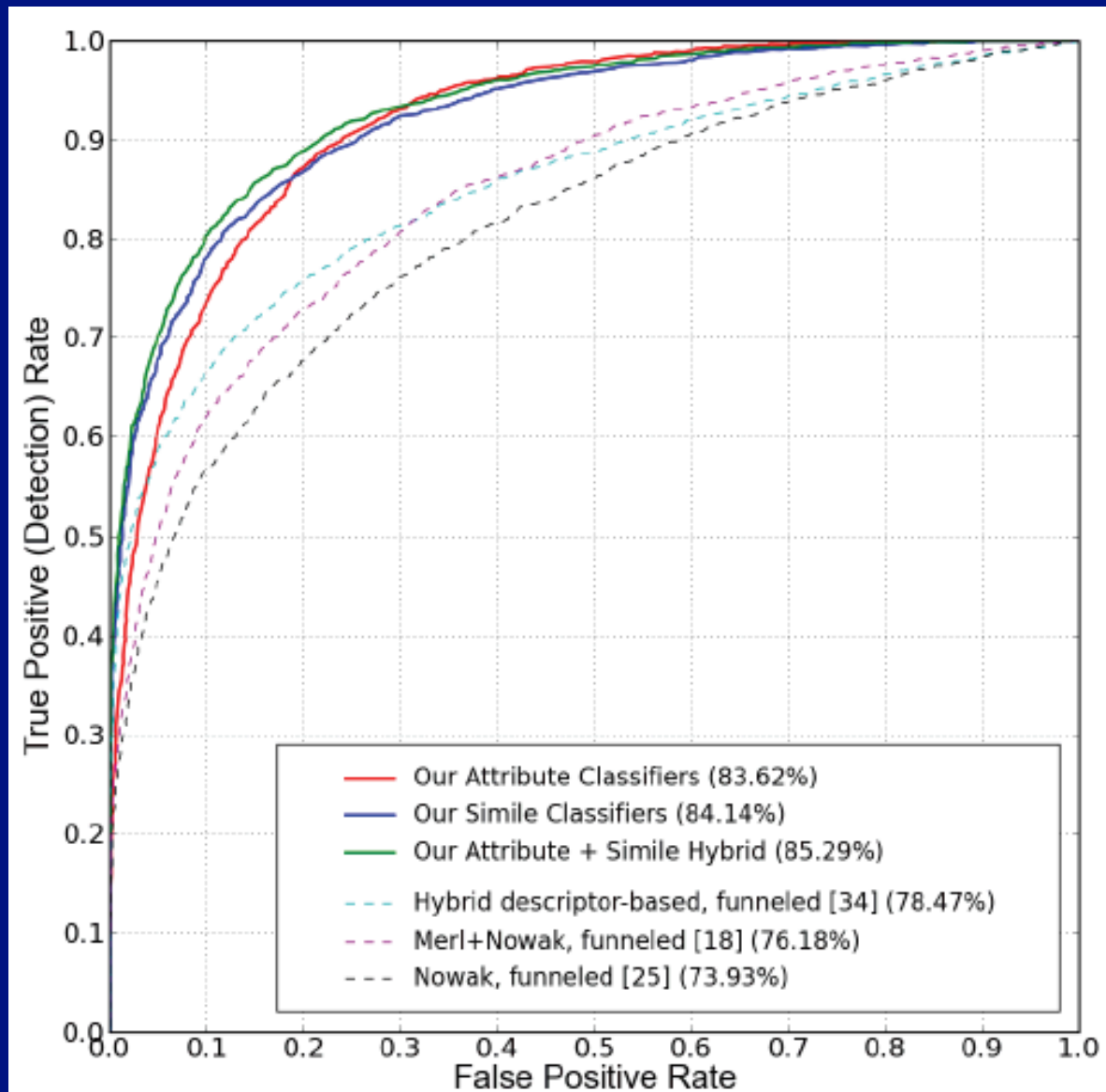
What is an object like?



Viz comic, issue 101

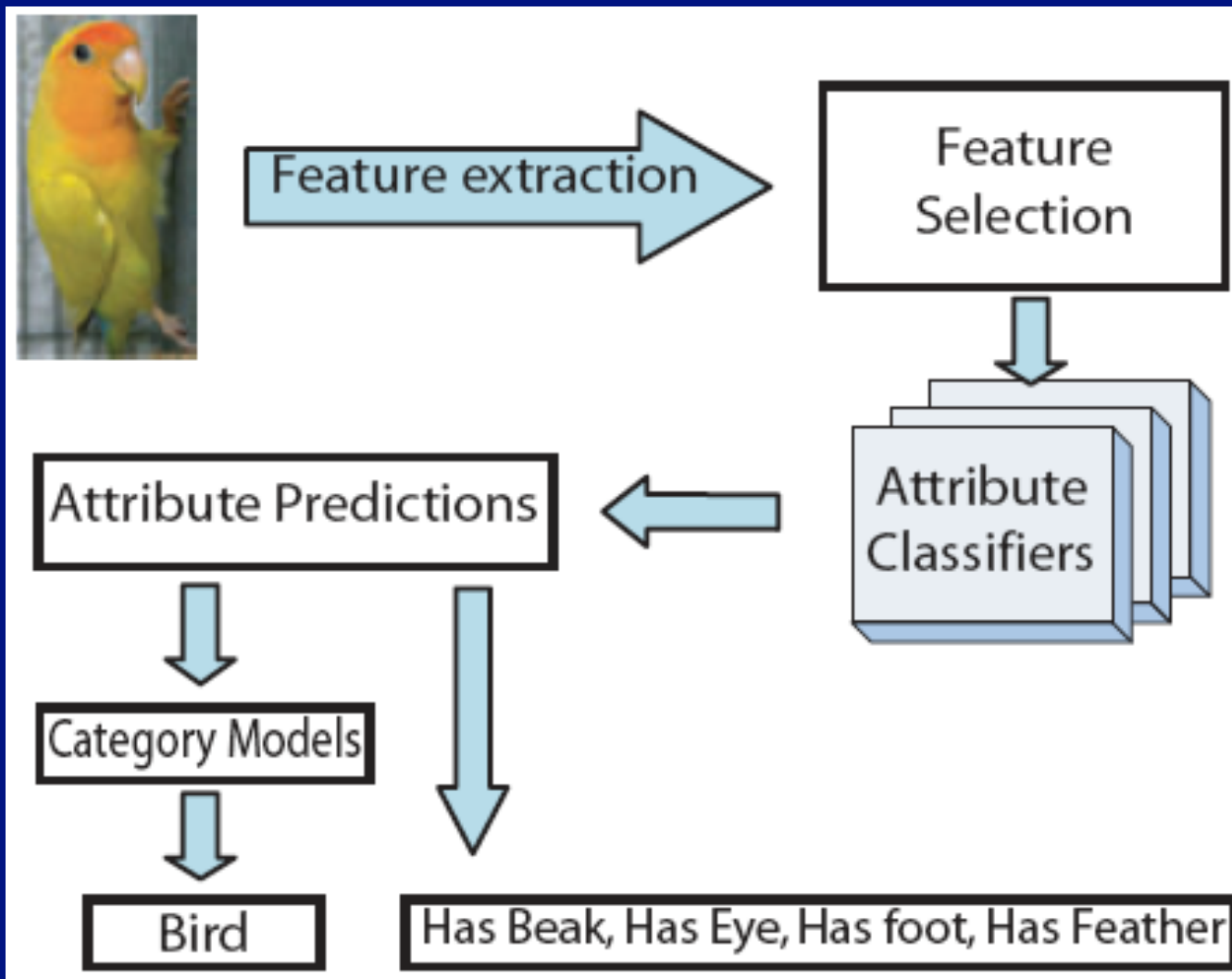


“Attribute and Simile Classifiers for Face Verification,” ICCV 2009. (N. Kumar, A. Berg, P. Belhumeur, S. K. Nayar)



“Attribute and Simile Classifiers for Face Verification,” ICCV 2009. (N. Kumar, A. Berg, P. Belhumeur, S. K. Nayar)

General architecture

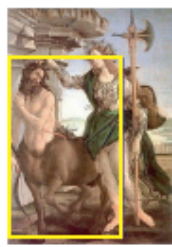




'is 3D Boxy'
 'is Vert Cylinder'
 'has Window'
 'has Row Wind'
~~'has Headlight'~~



'has Hand'
 'has Arm'
~~'has Screen'~~
 'has Plastic'
 'is Shiny'



'has Head'
 'has Hair'
 'has Face'
~~'has Saddle'~~
 'has Skin'



'has Head'
 'has Torso'
 'has Arm'
 'has Leg'
~~'has Wood'~~



'has Head'
 'has Ear'
 'has Snout'
 'has Nose'
 'has Mouth'



'has Head'
 'has Ear'
 'has Snout'
 'has Mouth'
 'has Leg'



~~'has Furniture Back'~~
~~'has Horn'~~
~~'s Screen'~~
 'has Plastic'
 'is Shiny'



'is 3D Boxy'
 'has Wheel'
 'has Window'
 'is Round'
 'has Torso'



'has Tail'
 'has Snout'
 'has Leg'
~~'has Text'~~
~~'has Plastic'~~



'has Head'
 'has Ear'
 'has Snout'
 'has Leg'
 'has Cloth'

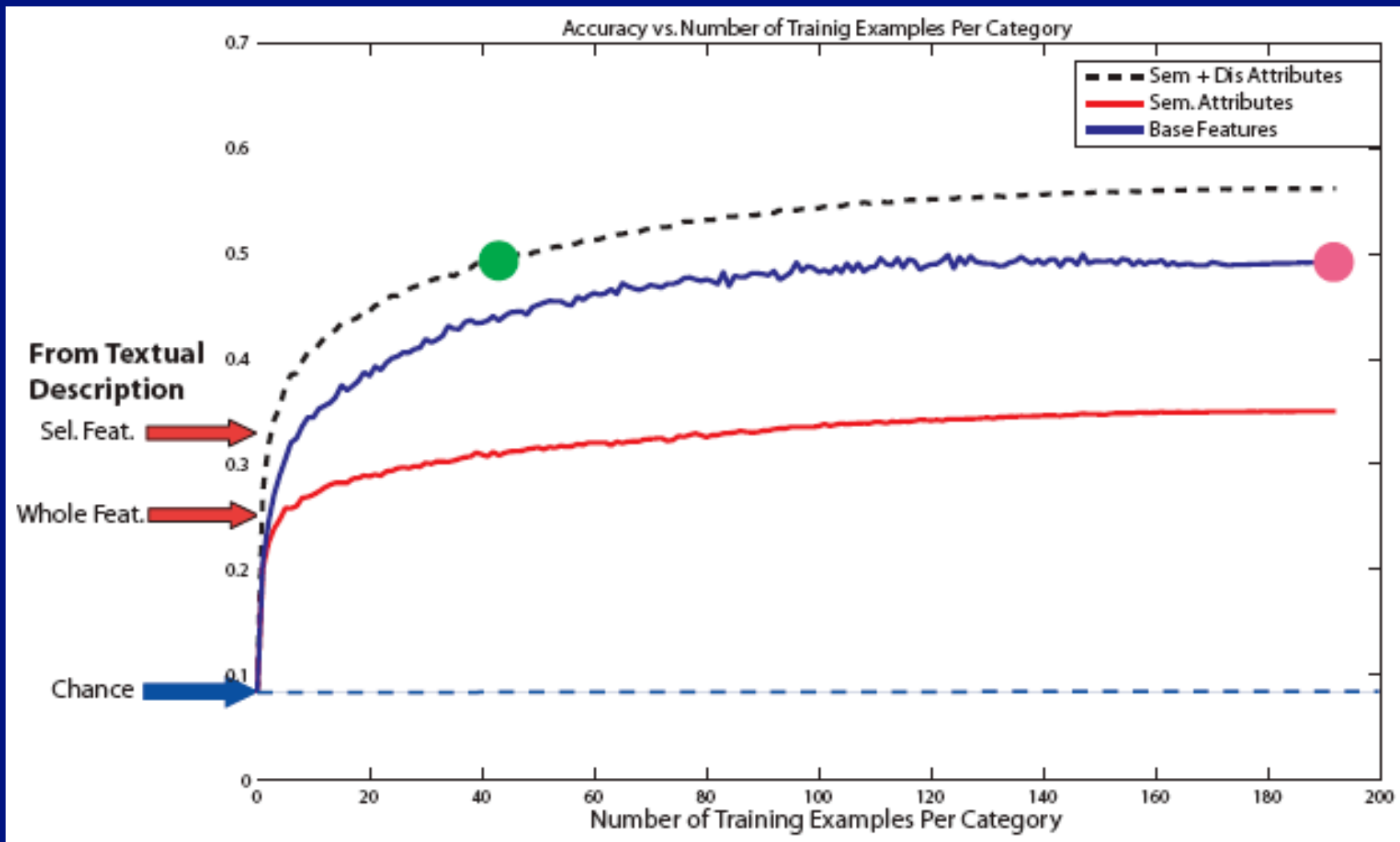


'is Horizontal Cylinder'
~~'has Beak'~~
~~'has Wing'~~
~~'has Side mirror'~~
 'has Metal'



'has Head'
 'has Snout'
 'has Horn'
 'has Torso'
~~'has Arm'~~

Few Examples



A. Farhadi, I. Endres, D. Hoiem, D.A. Forsyth, "Describing objects by their attributes", CVPR 2009

Motivating problems

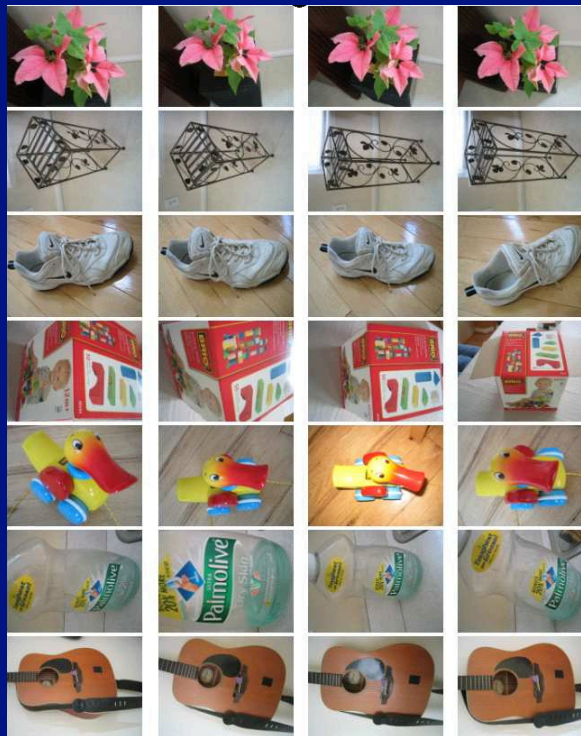
- Near duplicate image detection
 - find pictures in very large collections that are near duplicates of a query
 - trademarks, copyright, storage, matching
- Detecting objects
 - find objects in images using examples
- Detecting and tracking people
 - because they appear in lots of images/movies
- Point matching reconstruction
 - as in previous slides

Near duplicate image detection

- Strategy
 - if two images have many small patches that are similar, they might match
 - particularly if the arrangements are consistent



Query



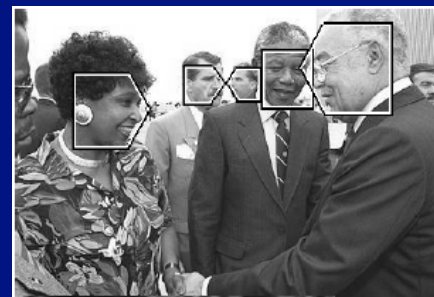
Collection



Response

Sliding window detection

- Detect instances of objects in images



Sliding window detection



Windows



What the detector sees

Detector: take a description of a window, say yes or no based on examples, statistical test

Sliding window detection

- We use fixed size windows
 - find small objects
 - run windows over big version of picture
 - find big objects
 - run windows over smaller version of picture

