Computer Vision Crash Course

Jia-Bin Huang

University of Illinois, Urbana-Champaign

www.jiabinhuang.com

Jan 12, 2016

Overview

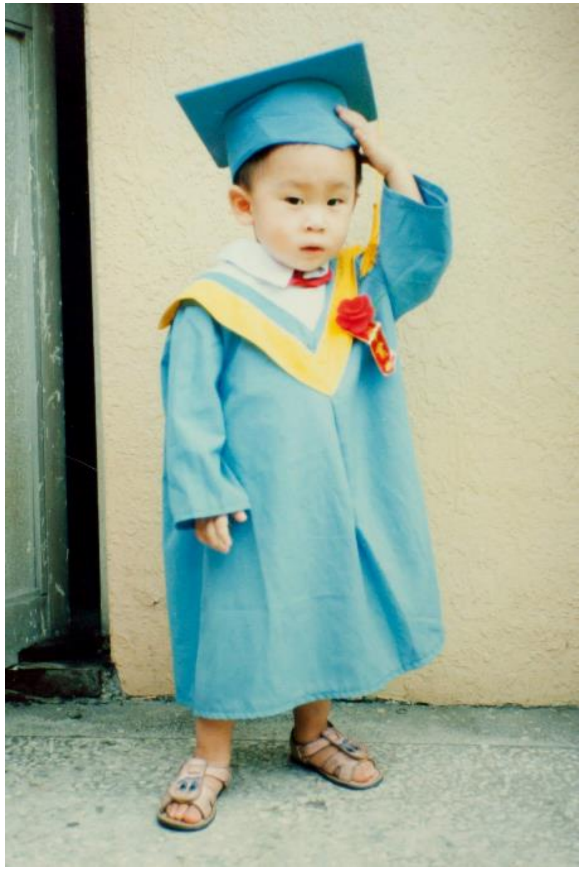
- A little about me
- Introduction to Computer Vision

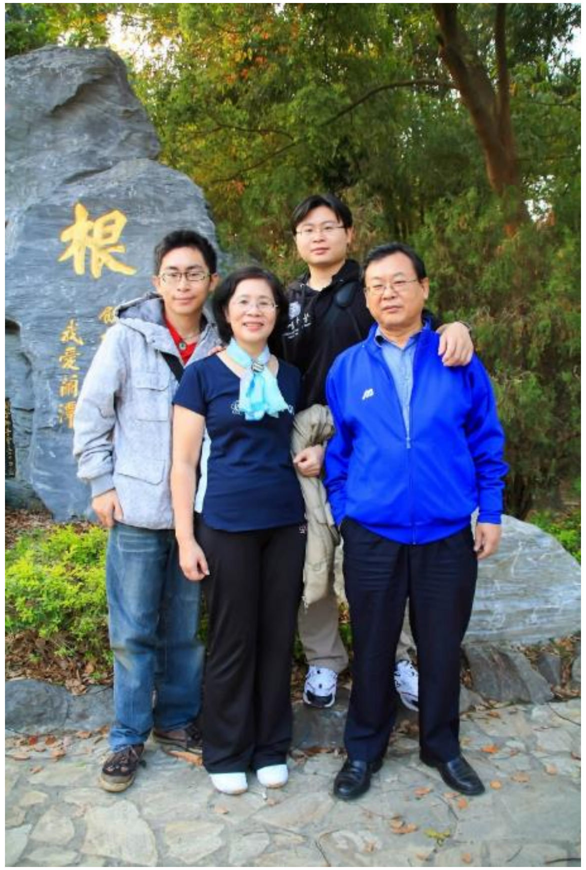
=== Intermission ===

- Fundamentals and Applications
- Resources

About me

- Born in Kaohsiung, raised in Taipei





About me

National Chiao-Tung University
B.S. in EE

Microsoft Research
Research Intern 2012, 2013

IIS, Academia Sinica
Research Assistant

Disney Research
Research Intern 2014

UC, Merced
Visiting Student

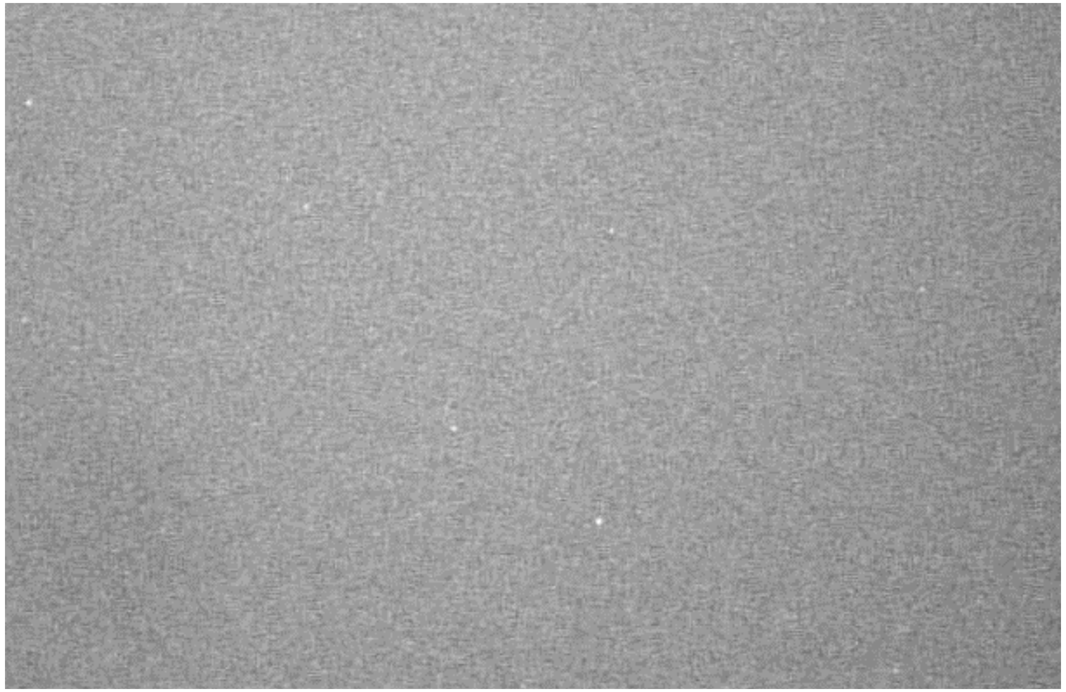
UIUC
Ph.D. in ECE
(Expected summer 2016)

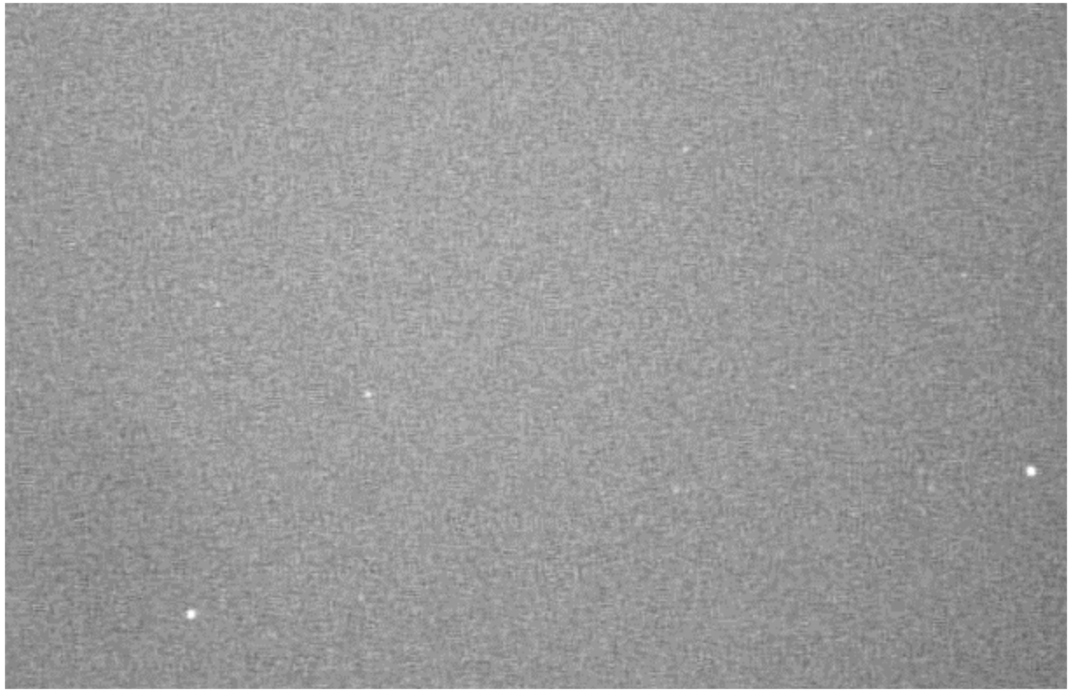
My research: Computational Photography

Image Completion

Image super-resolution

My Research: Video-based Bird Migration Monitoring





My Research: Visual Tracking

Object Tracking

Multi-face Tracking

What is Computer Vision?

- Make computers understand images and videos.

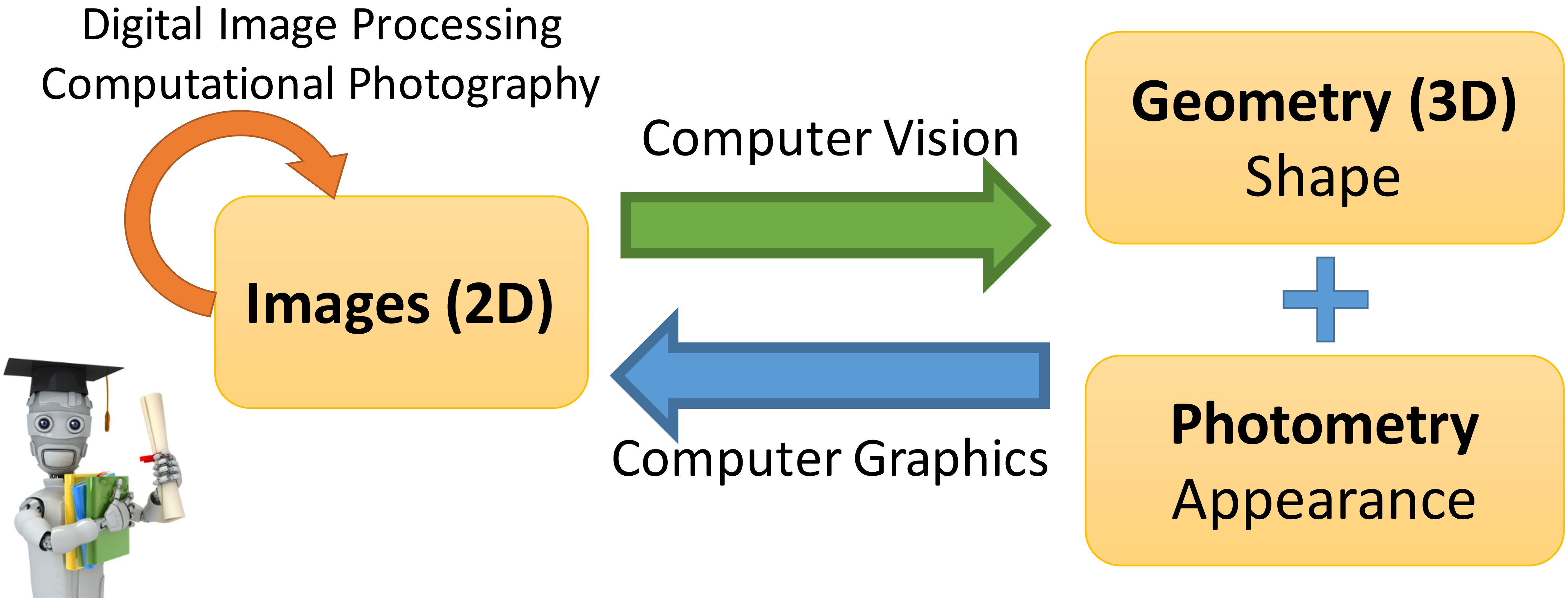
- What kind of scene?
- Where are the cars?
- How far is the building?

What is Computer Vision?

- Make computers understand images and videos.

- What are they doing?
- Why is this happening?
- What is important?
- What will I see?

Computer Vision and Nearby Fields



Machine learning:
Vision = Machine learning applied to visual data

Visual data on the Internet

- Flickr
  - 10+ billion photographs
  - 60 million images uploaded a month
- Facebook
  - 250 billion+
  - 300 million a day
- Instagram
  - 55 million a day
- YouTube
  - 100 hours uploaded every minute

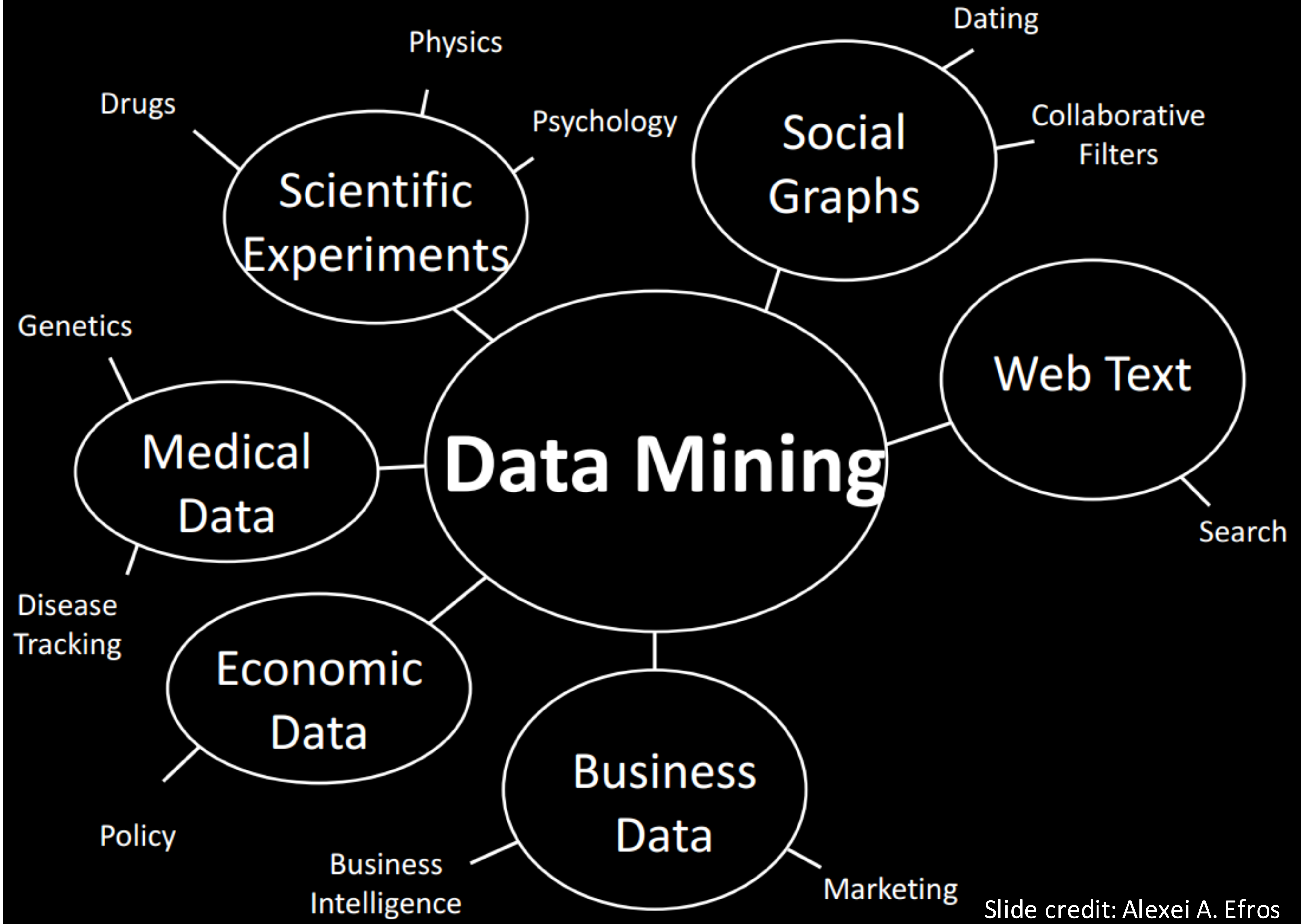
CISCO **90%** of net traffic will be visual!

Mostly about cats

Too big for humans

http://www.petittube.com/

- Need automatic tools to access and analyze visual data!



Slide credit: Alexei A. Efros

Vision is Really Hard

- Vision is an amazing feature of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - More human brain devoted to vision than anything else

Why is Computer Vision Hard?

Why is Computer Vision Hard?

What did you see?

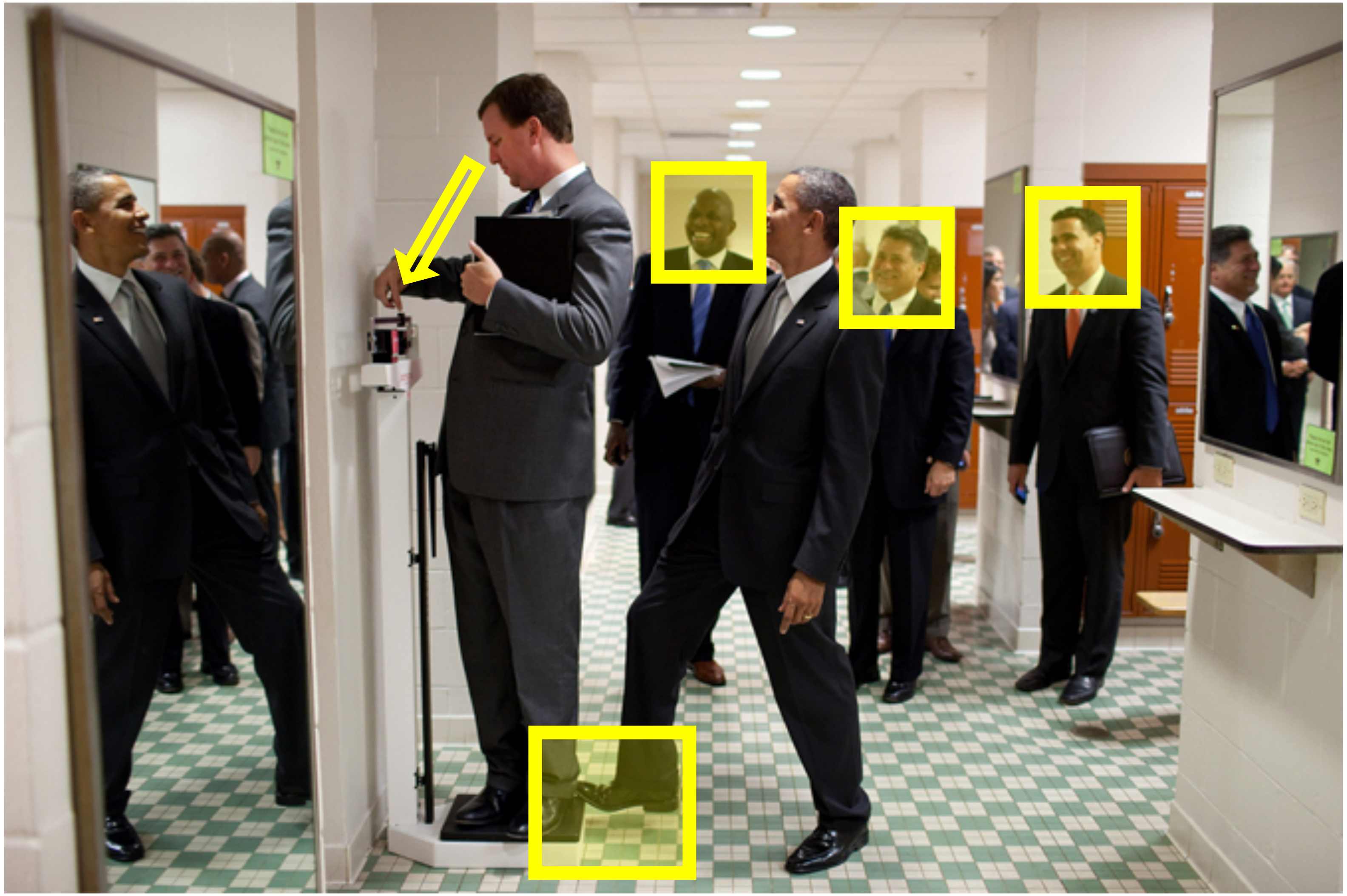
- Where this picture was taken?
- How many people are there?
- What are they doing?
- What object the person on the left standing on?
- Why this is a funny picture?

Why is Computer Vision Hard?

Why is Computer Vision Hard?

Why is Computer Vision Hard?

Why is Computer Vision Hard?



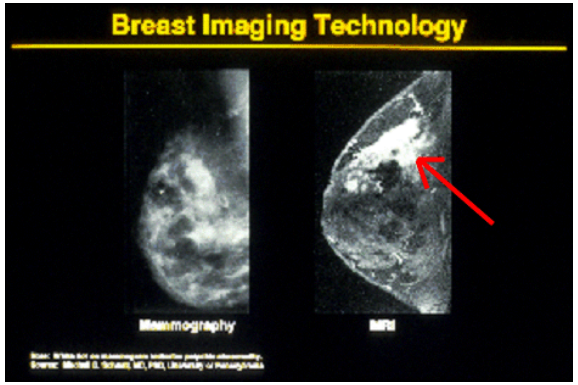
Why is Computer Vision Hard?

Why is Computer Vision Hard?

Computer: okay, it’s a funny picture

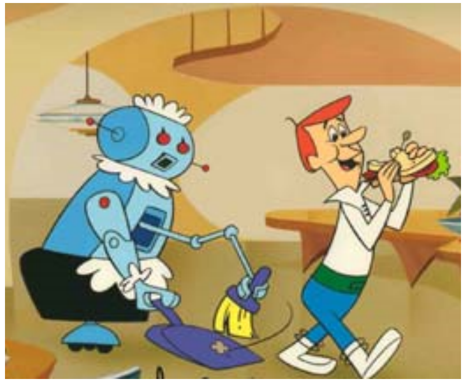
Computer Vision Matters

Safety



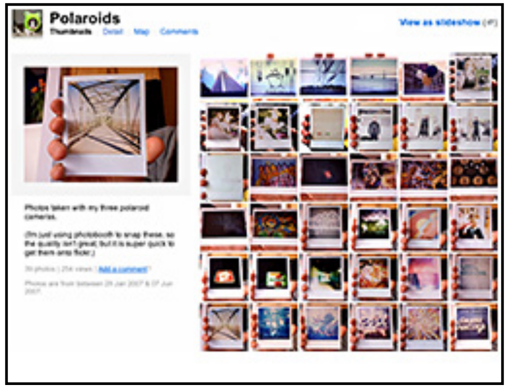
Health

Security



Comfort

Fun



Access

History of Computer Vision

Marvin Minsky, MIT
Turing award, 1969

"In 1966, Minsky hired a first-year undergraduate student and assigned him a problem to solve over the summer:

connect a camera to a computer and get the machine to describe what it sees."

Crevier 1993, pg. 88

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

Half a century later,
we're still working on it.

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

History of Computer Vision

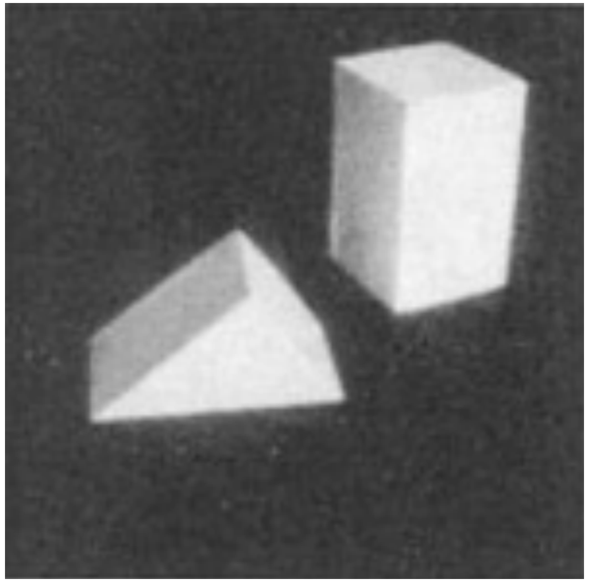
Marvin Minsky, MIT
Turing award, 1969

Gerald Sussman, MIT

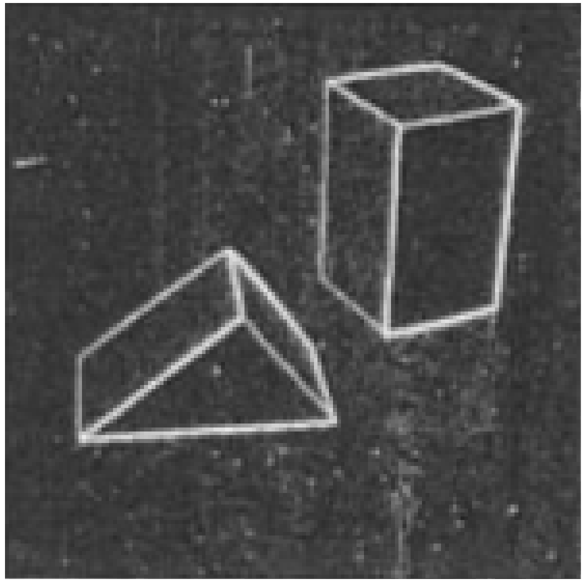
"You'll notice that Sussman never worked in vision again!" – Berthold Horn

1960's: interpretation of synthetic worlds

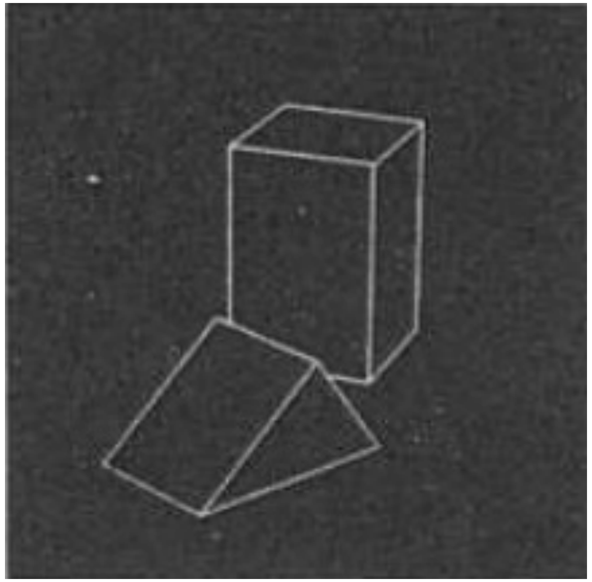
Larry Roberts
"Father of Computer Vision"



Input image



2x2 gradient operator

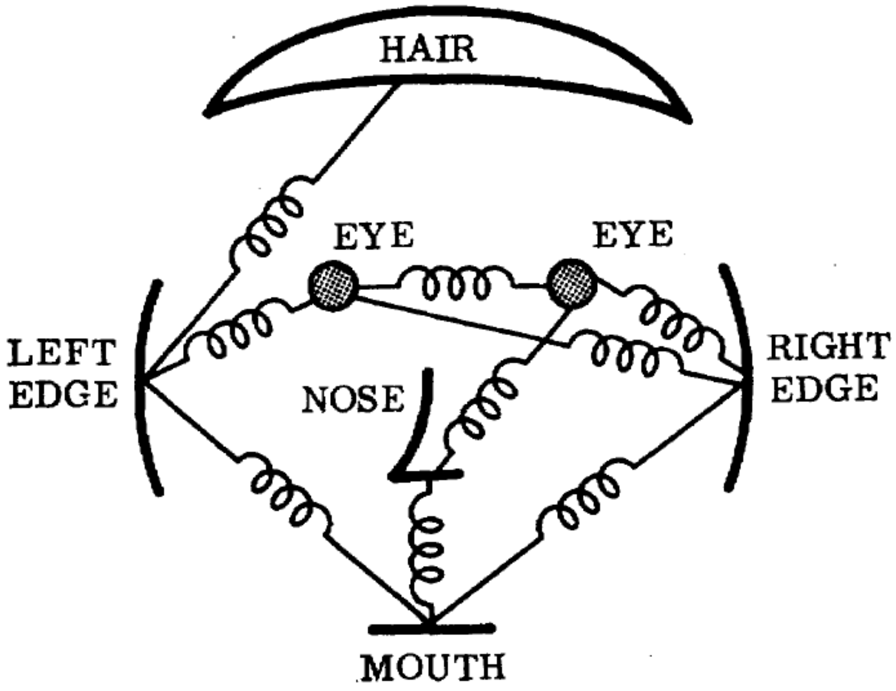


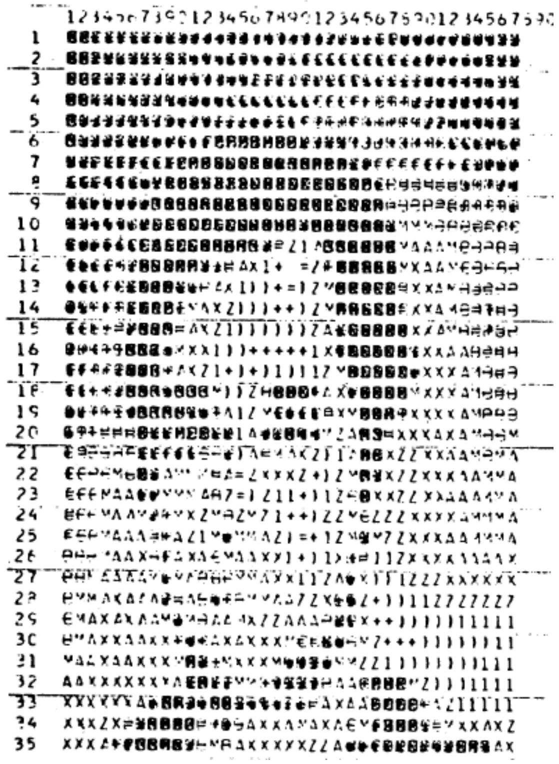
computed 3D model
rendered from new viewpoint

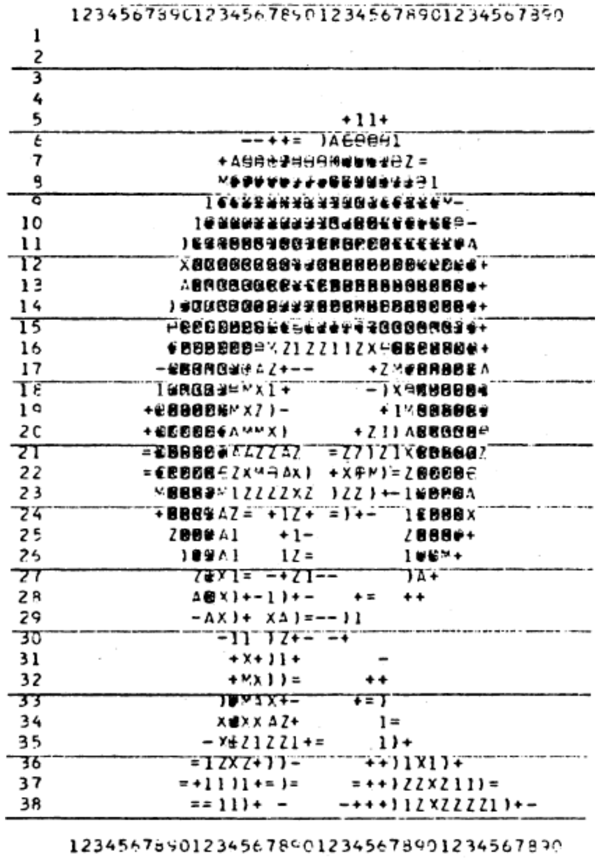
Larry Roberts PhD Thesis, MIT, 1963,
Machine Perception of Three-Dimensional Solids

Slide credit: Steve Seitz

1970’s: some progress on interpreting selected images

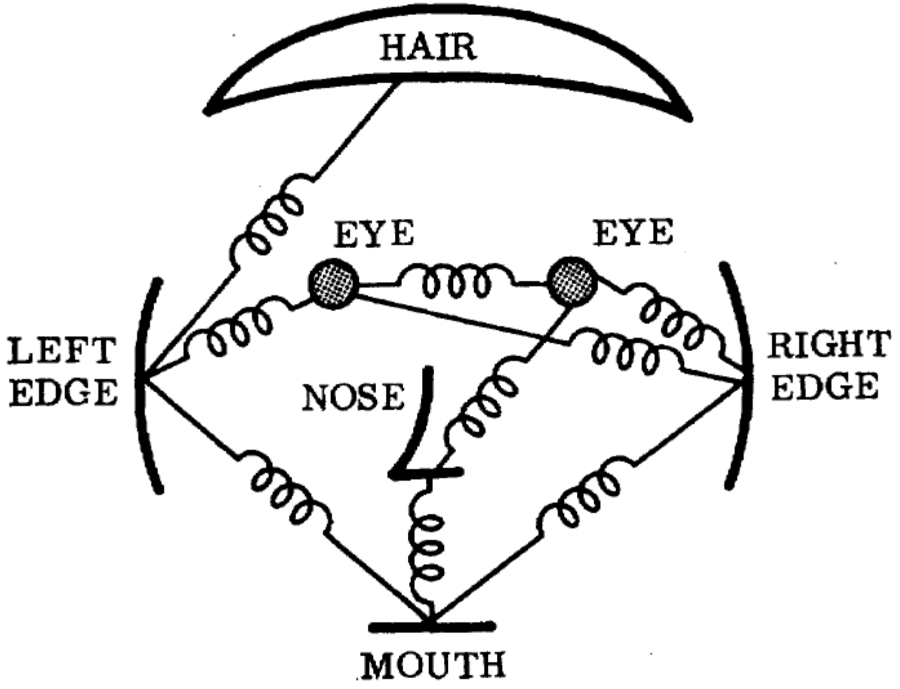


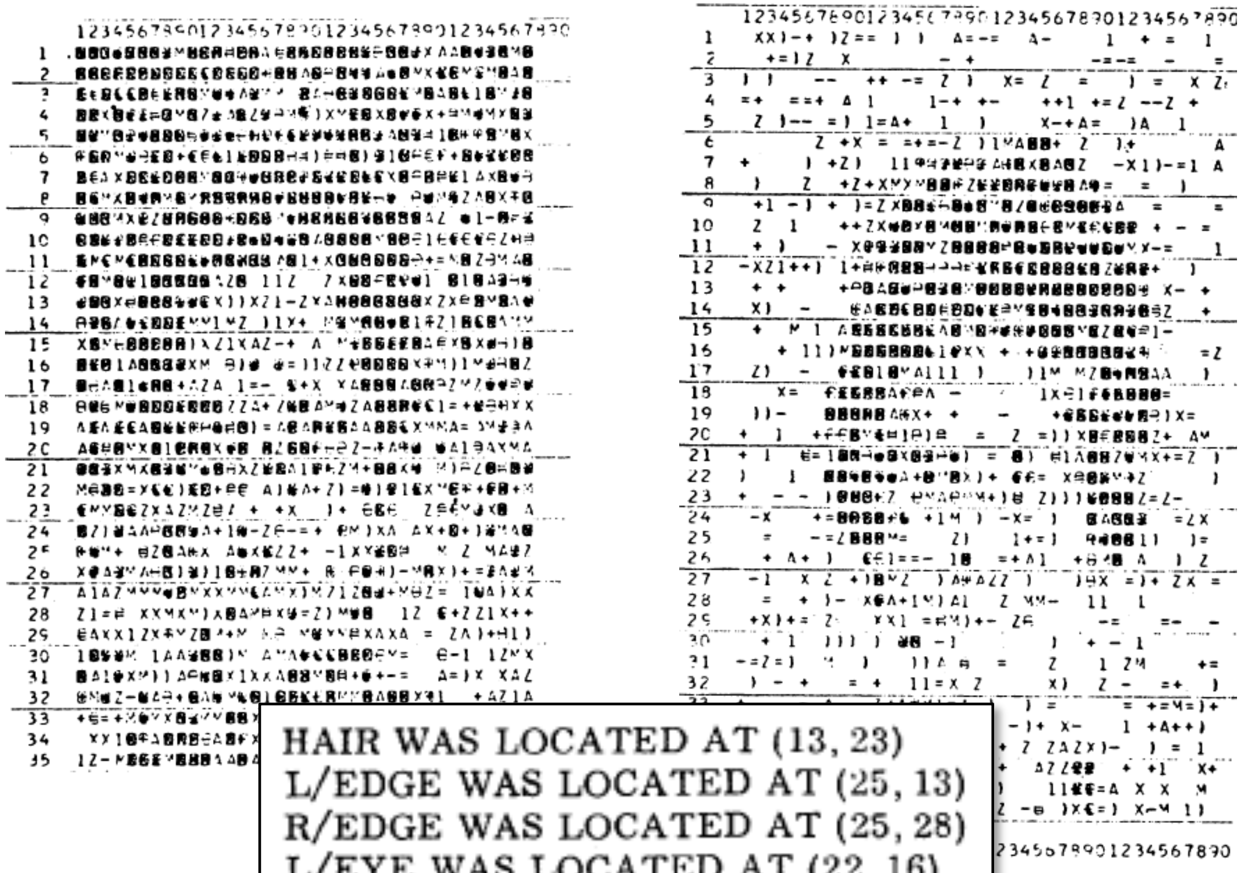




The representation and matching of pictorial structures
Fischler and Elschlager, 1973

1970’s: some progress on interpreting selected images



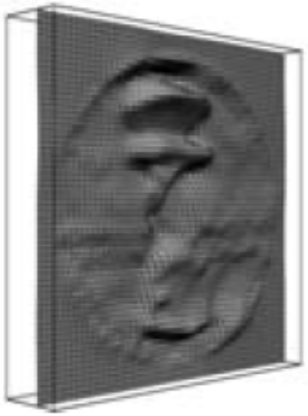


HAIR WAS LOCATED AT (13, 23)
L/EDGE WAS LOCATED AT (25, 13)
R/EDGE WAS LOCATED AT (25, 28)
L/EYE WAS LOCATED AT (22, 16)
R/EYE WAS LOCATED AT (22, 23)
NOSE WAS LOCATED AT (27, 20)
MOUTH WAS LOCATED AT (29, 19)

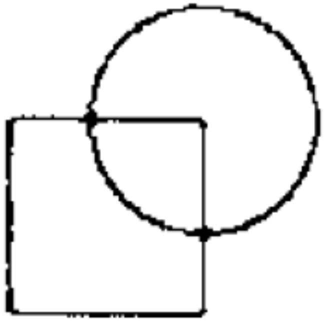
The representation and matching of pictorial structures
Fischler and Elschlager, 1973

1980's: ANNs come and go; shift toward geometry and increased mathematical rigor

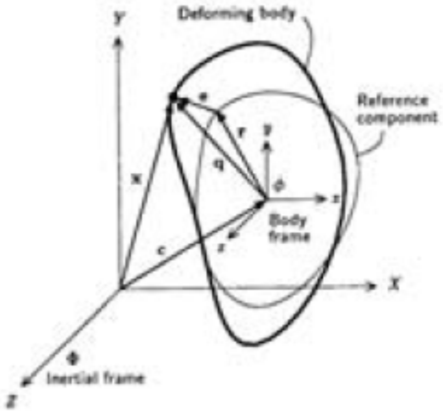
(a)



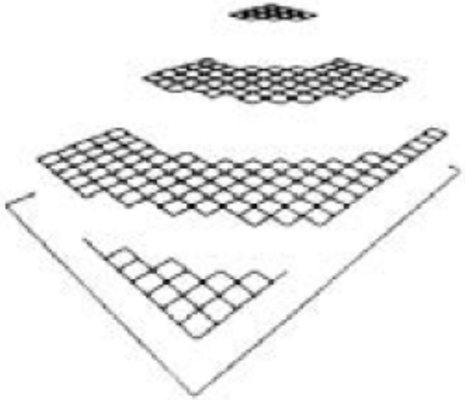
(b)



(c)



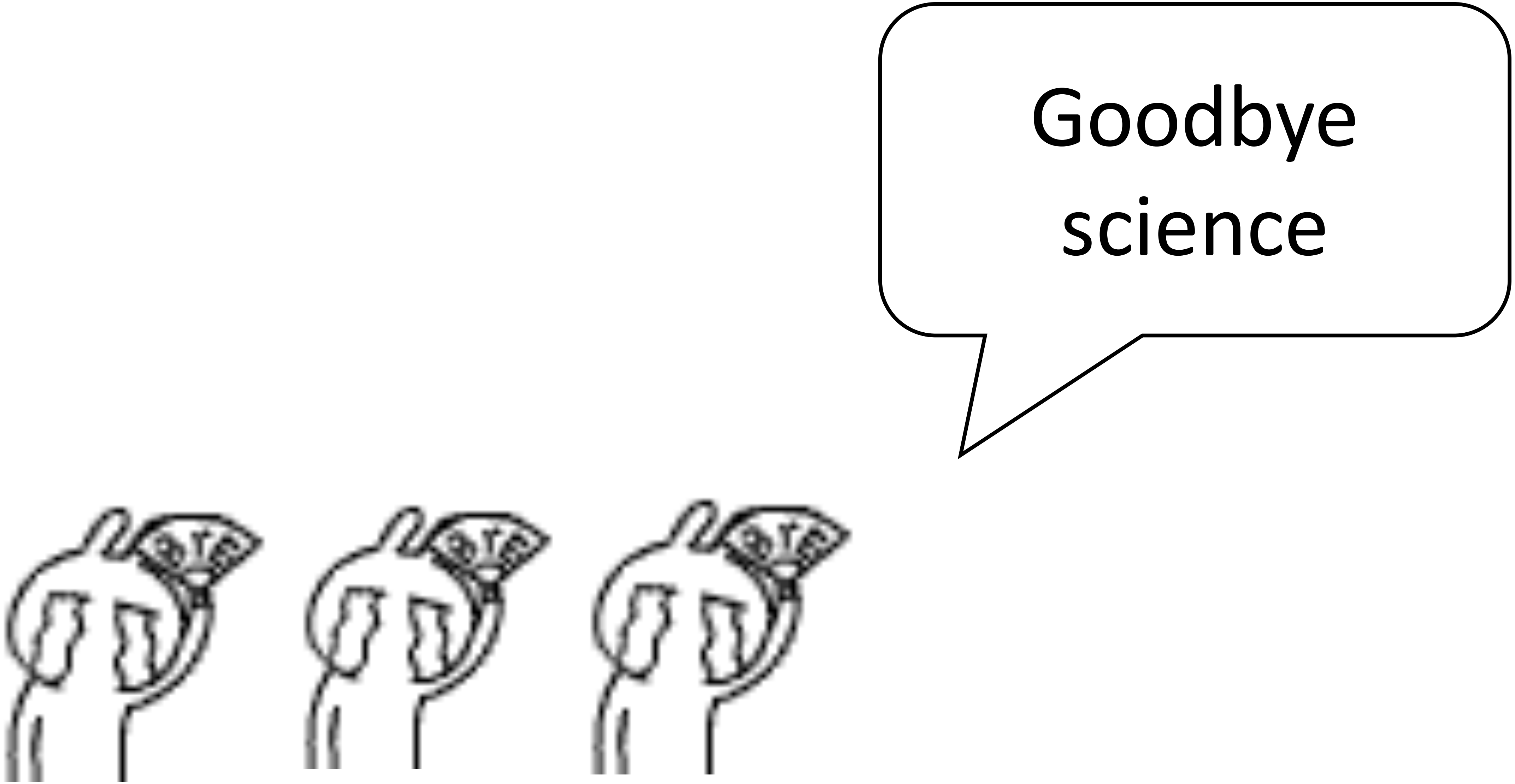
(d)



(e)

(f)

Image credit: Rick Szeliski


Goodbye science

1990’s: face recognition; statistical analysis in vogue

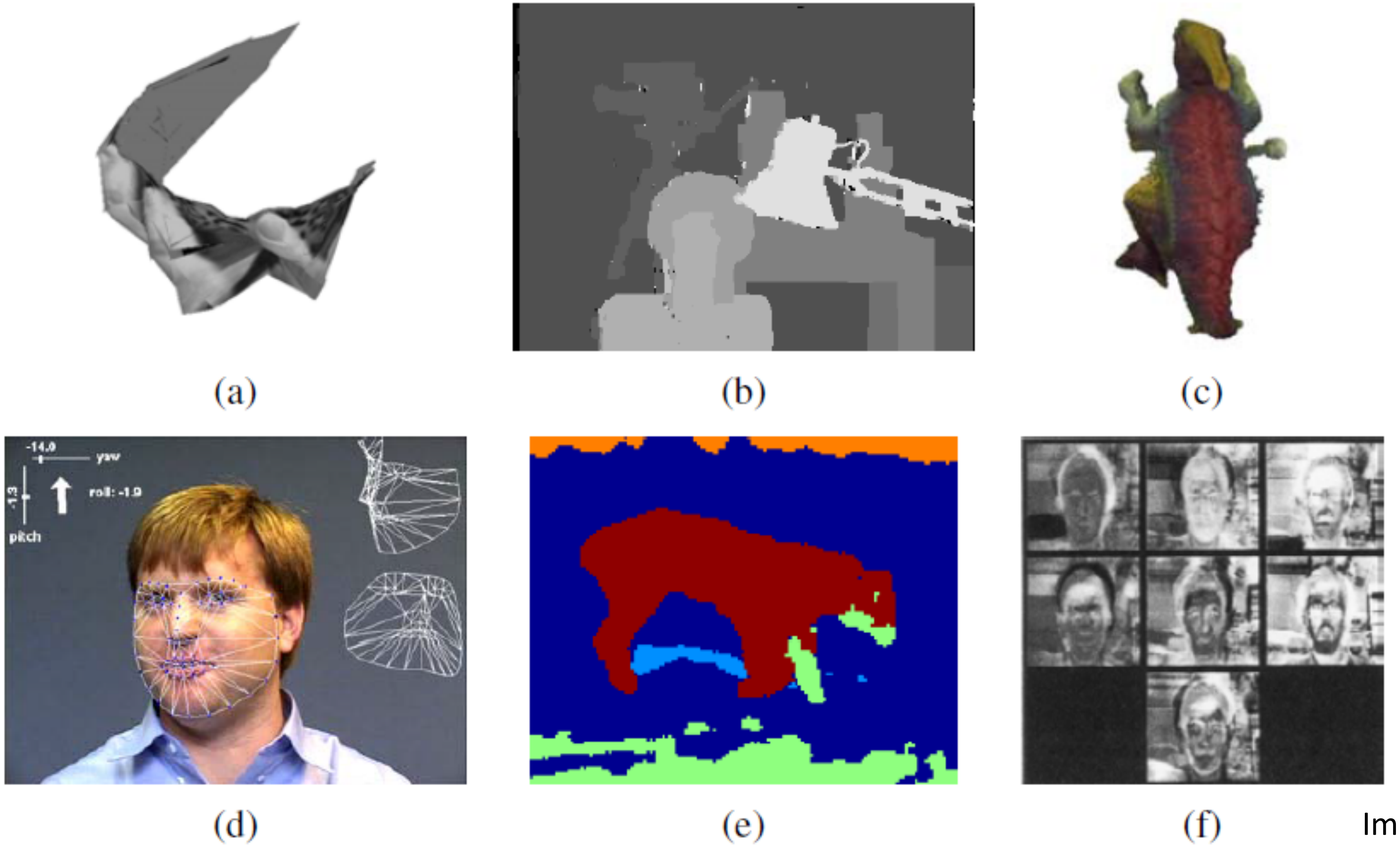
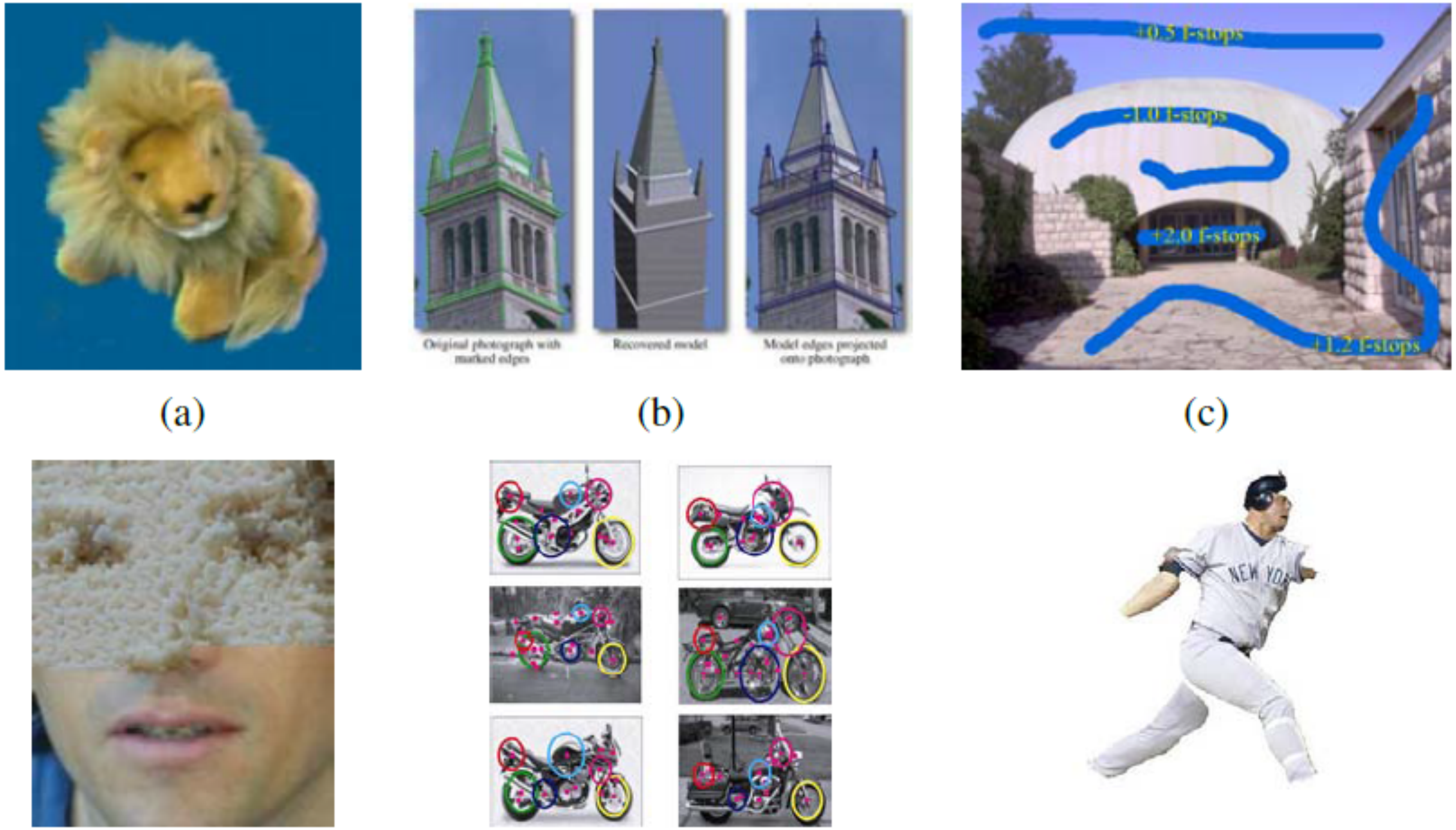


Image credit: Rick Szeliski

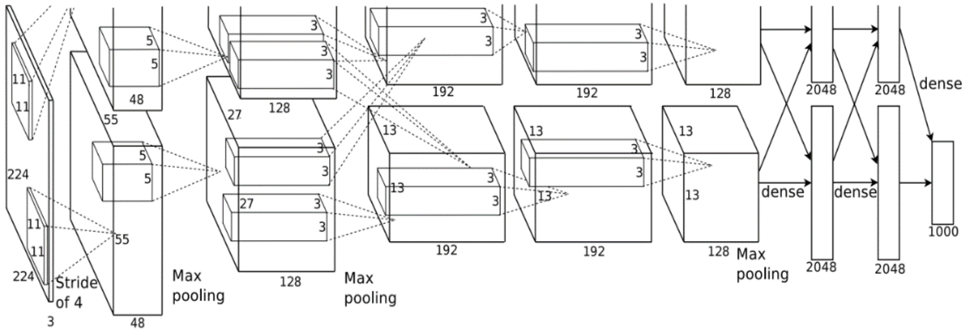
2000's: broader recognition; large annotated datasets available; video processing starts



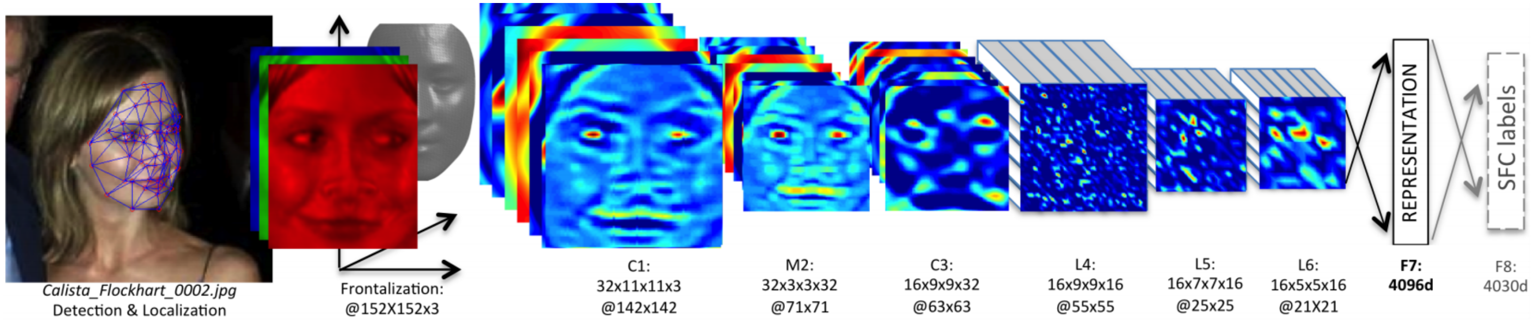
(a) (b) (c) (d) (e) (f)

Image credit: Rick Szeliski

2010's: resurgence of deep learning

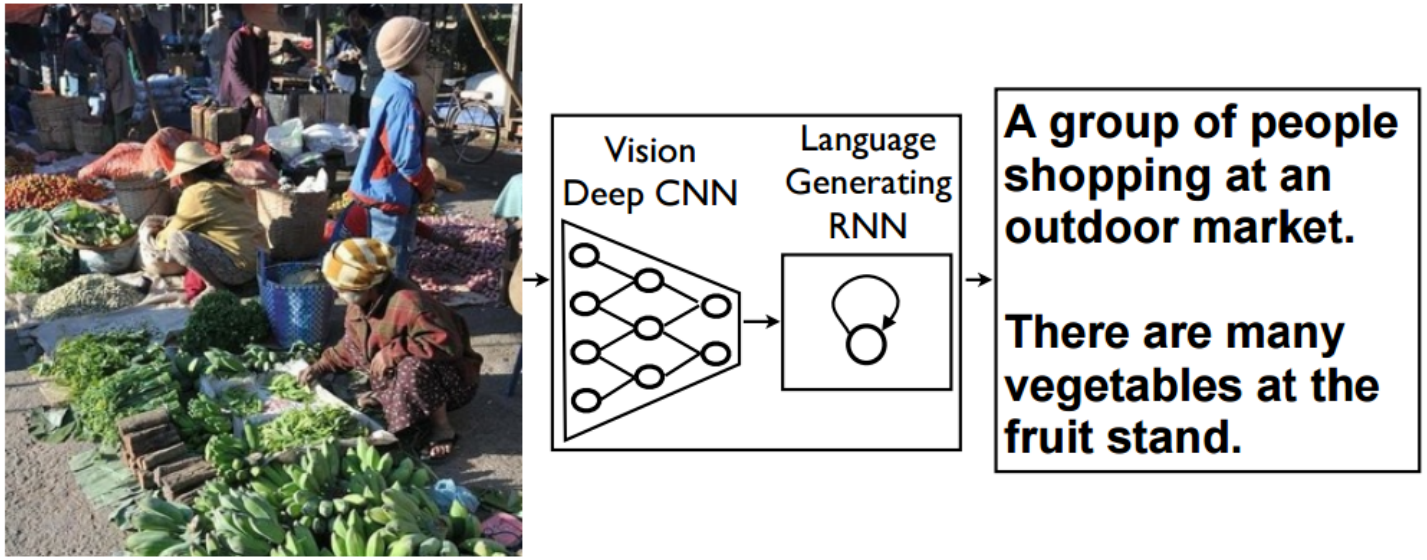


[AlexNet NIPS 2012]



[DeepFace CVPR 2014]

[DeepPose CVPR 2014]

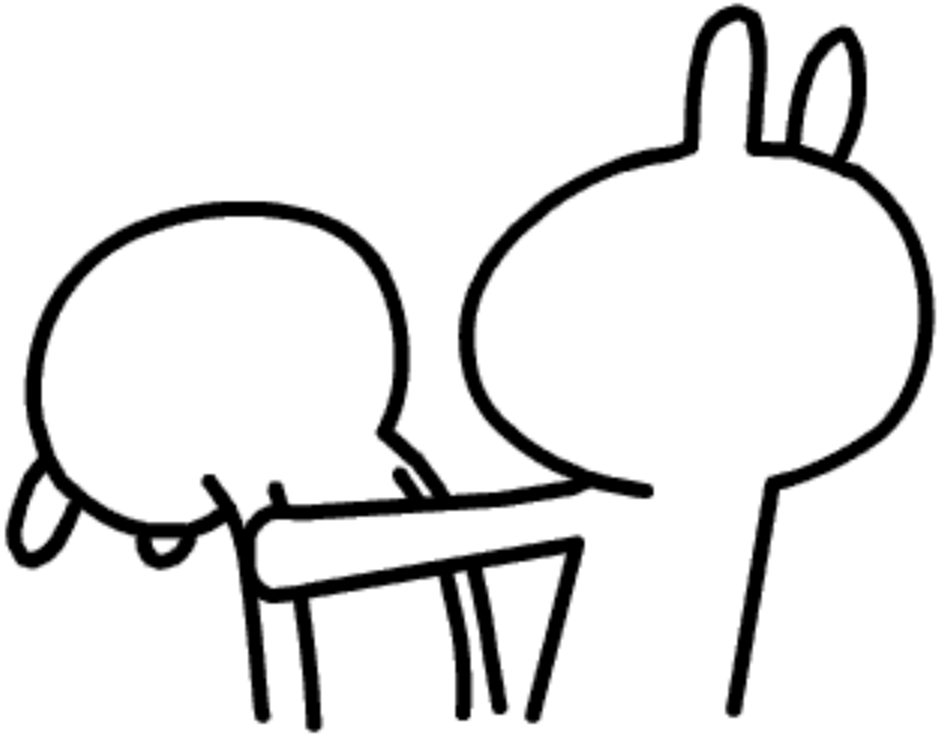


[Show, Attend and Tell ICML 2015]

2020’s: autonomous vehicles

2030's: robot uprising?

5-mins break



Examples of Computer Vision Applications

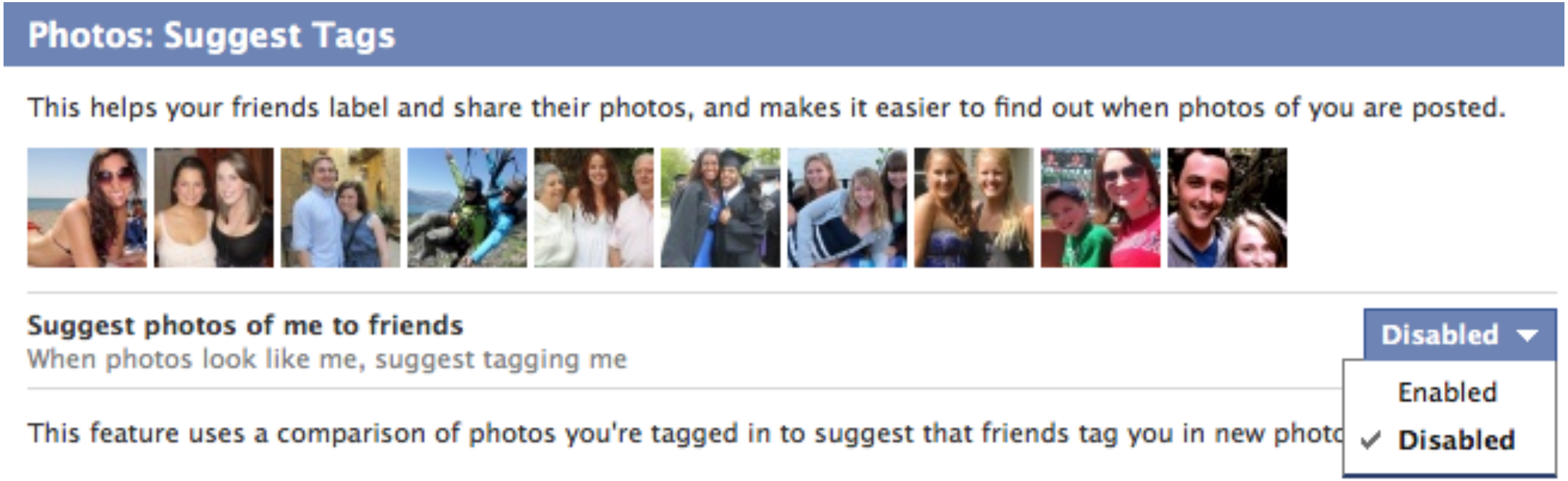
- How is computer vision used today?

Face detection

- Most digital cameras and smart phones detect faces (and more)
  - Canon, Sony, Fuji, …
- For smart focus, exposure compensation, and cropping

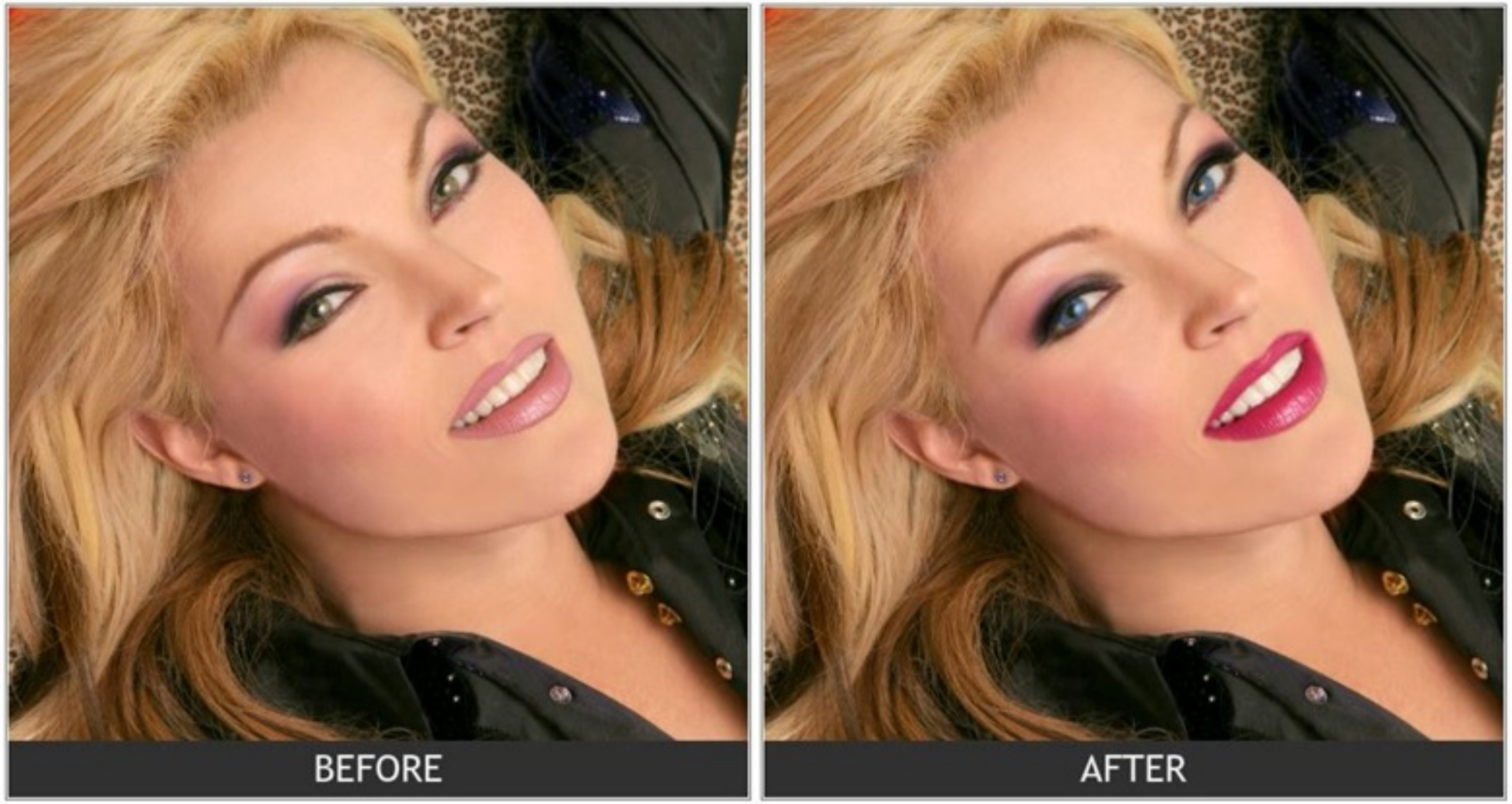
Slide credit: Steve Seitz

Face recognition

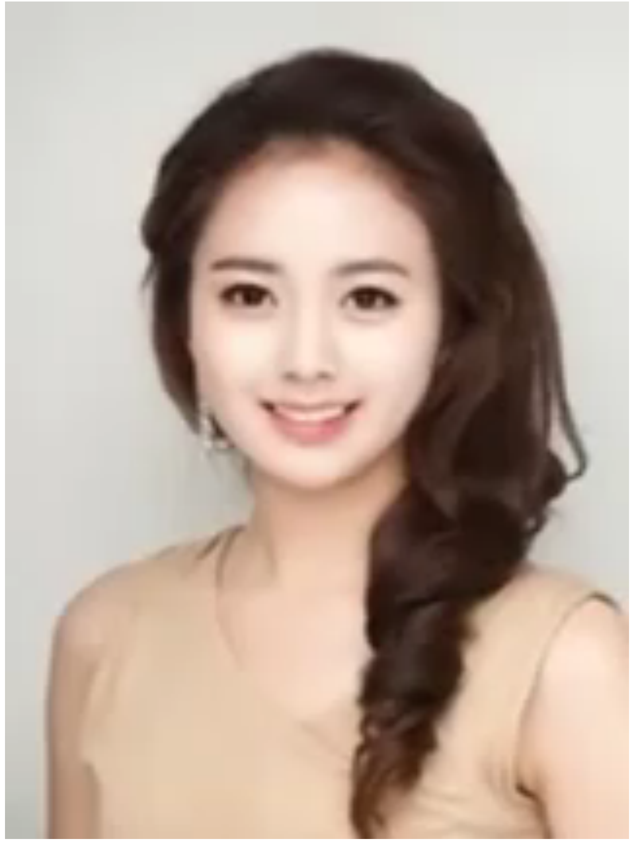


Facebook face auto-tagging

Face Landmark Alignment



TAAZ.com: Virtual makeover

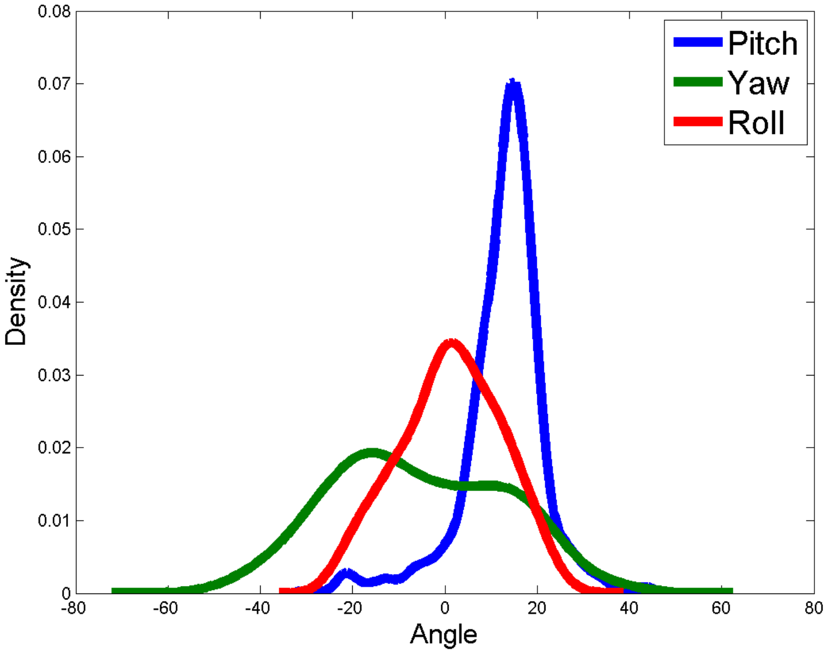


Face morphing

Jia-Bin Huang

Miss Daegu 2013 Contestants Face Morphing

Face Landmark Alignment- Pose Estimation



Pitch: ~ 15 degree

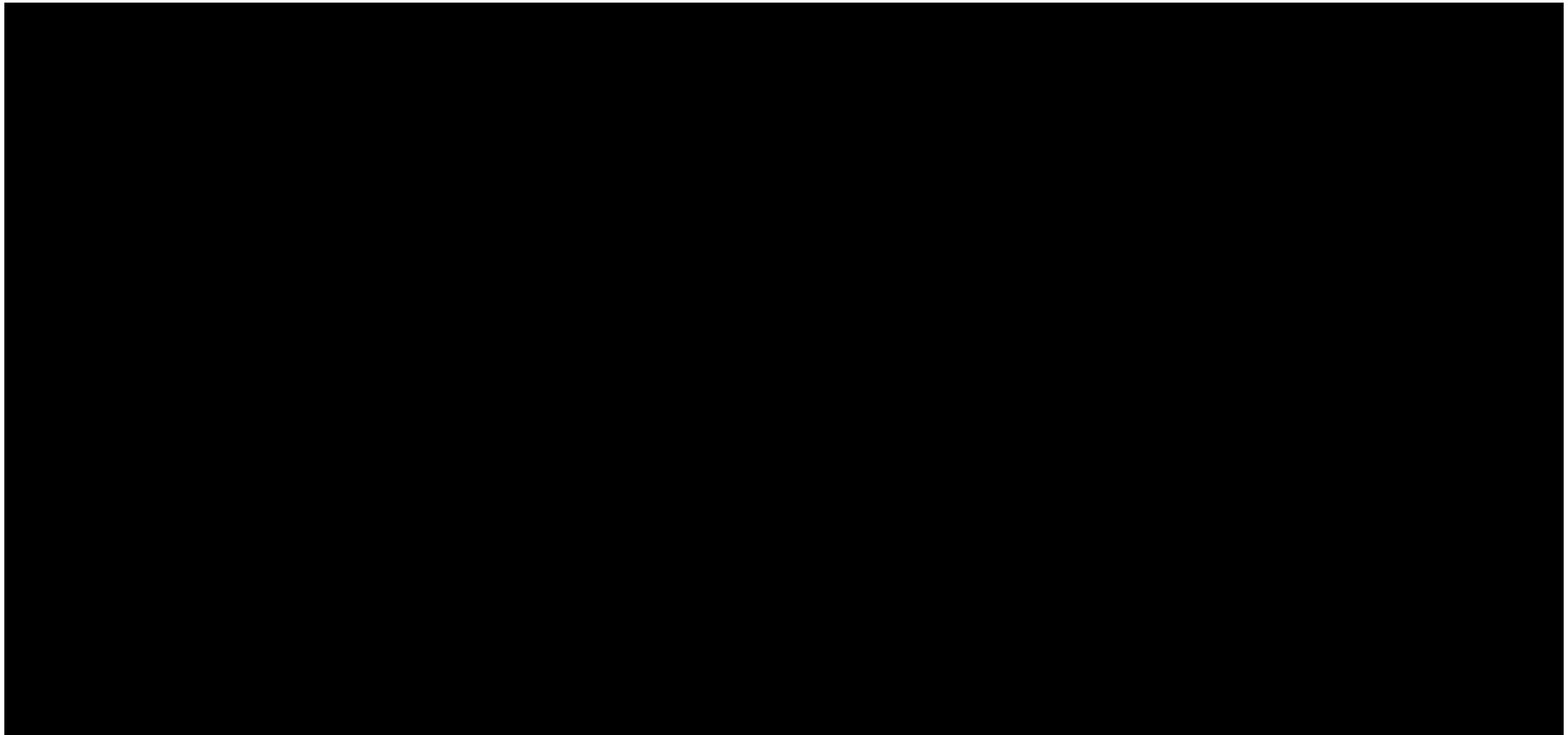
Yaw: ~ +20, -20 degree

Jia-Bin Huang [What's the Best Pose for a Selfie?]

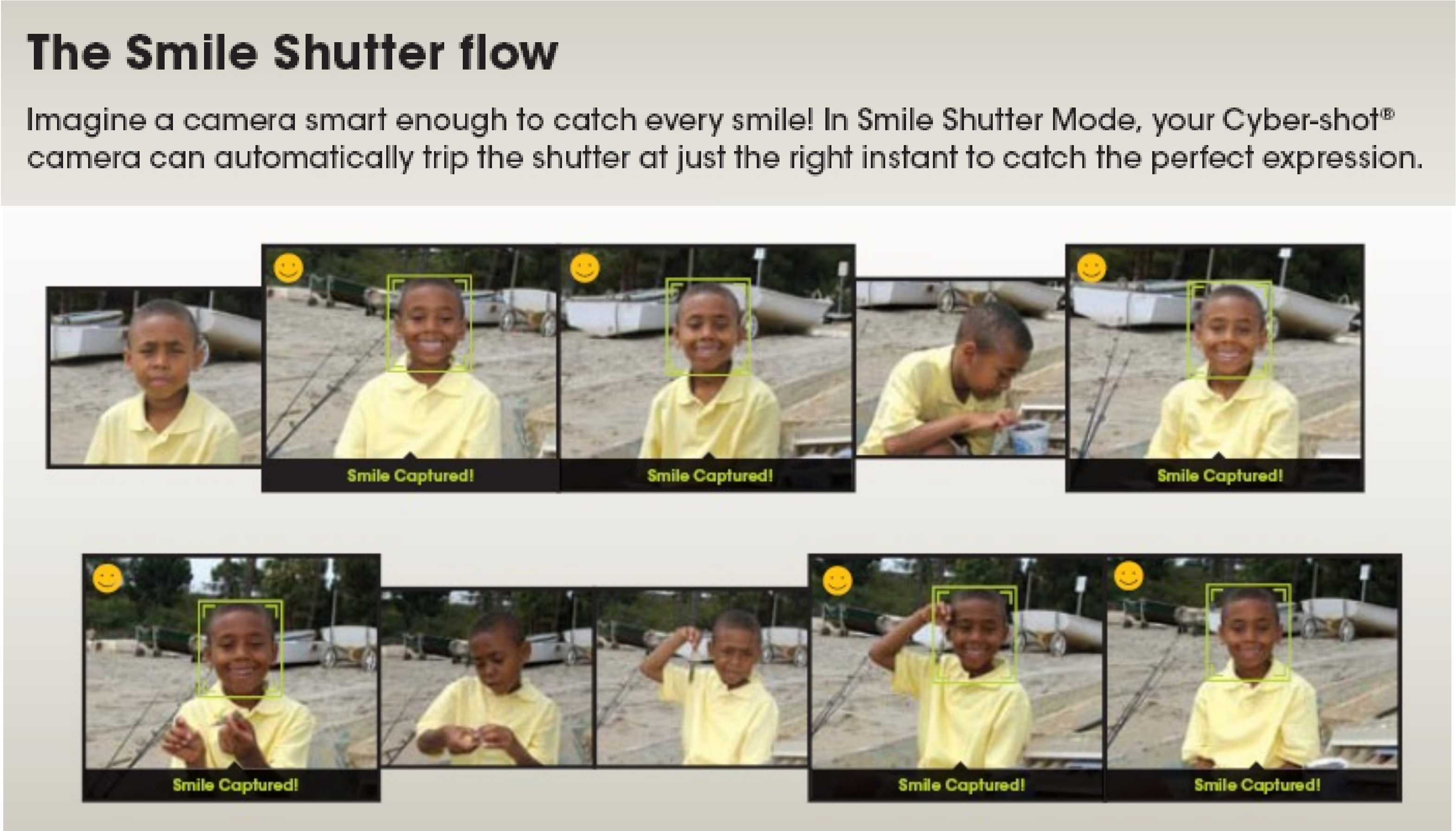
Face Landmark Alignment – 3D Persona

What Makes Tom Hanks Look Like Tom Hanks ICCV 2015

Video-based face replacement



Smile Detection



[Sony Cyber-shot® T70 Digital Still Camera]

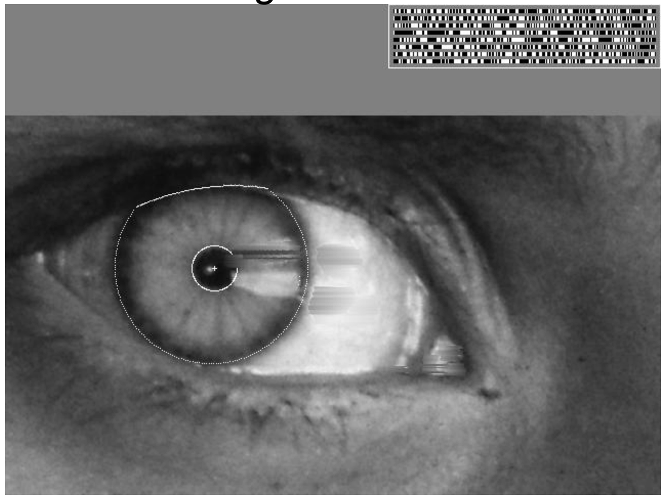
Slide credit: Steve Seitz

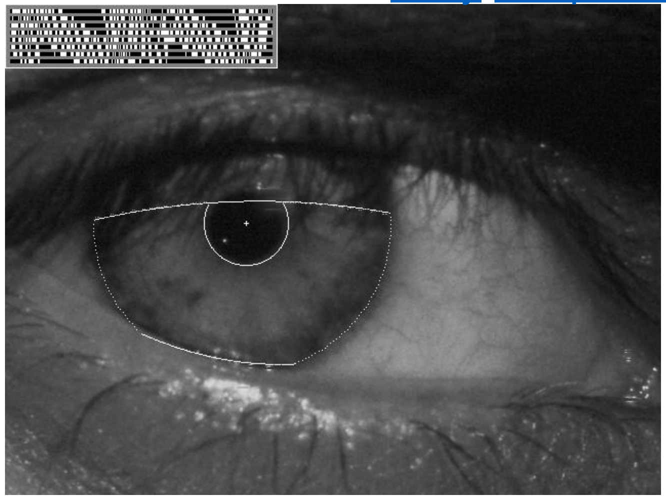
Eye contact detection

[Gaze locking, UIST 2013]

Vision-based Biometrics

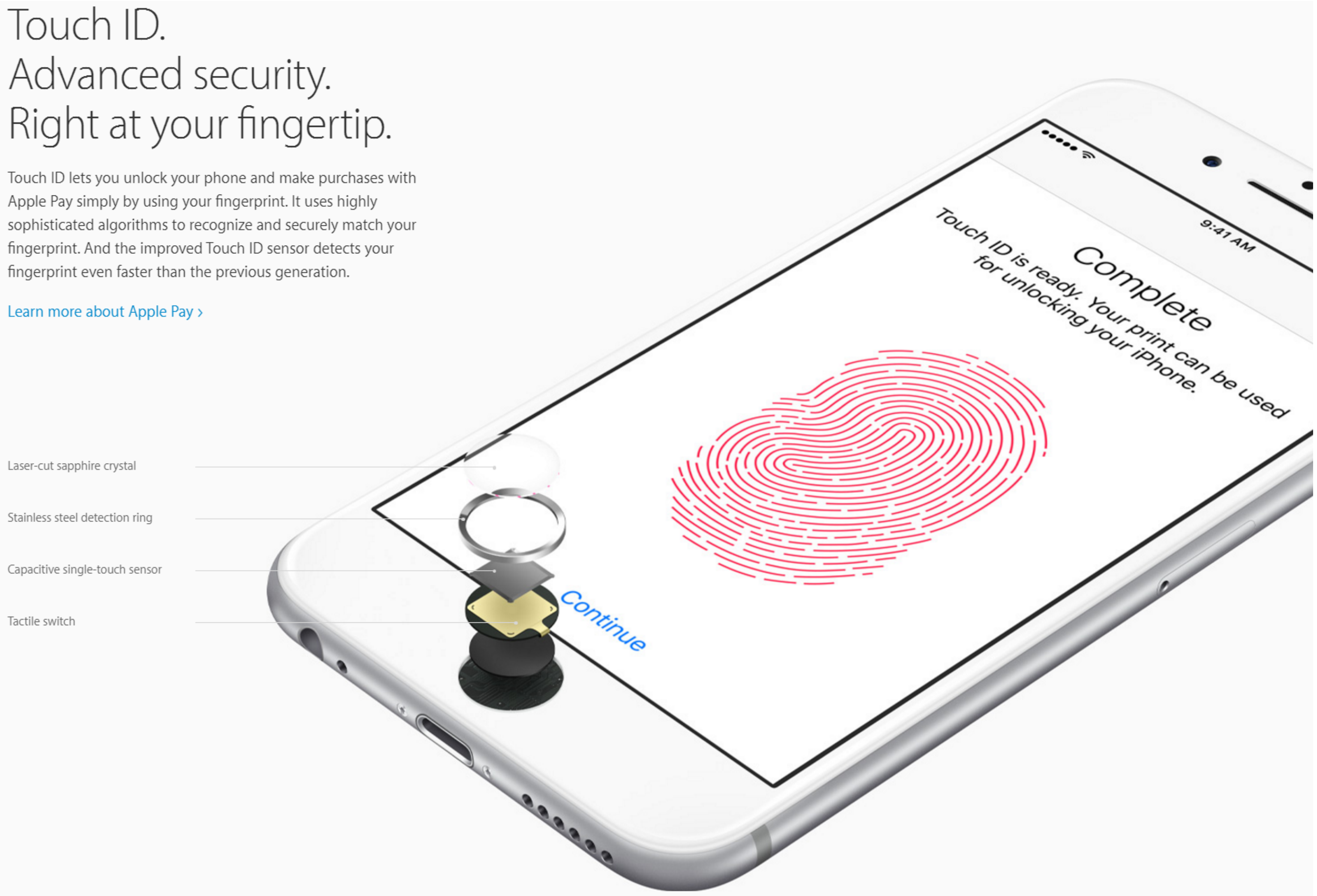
"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story wikipedia





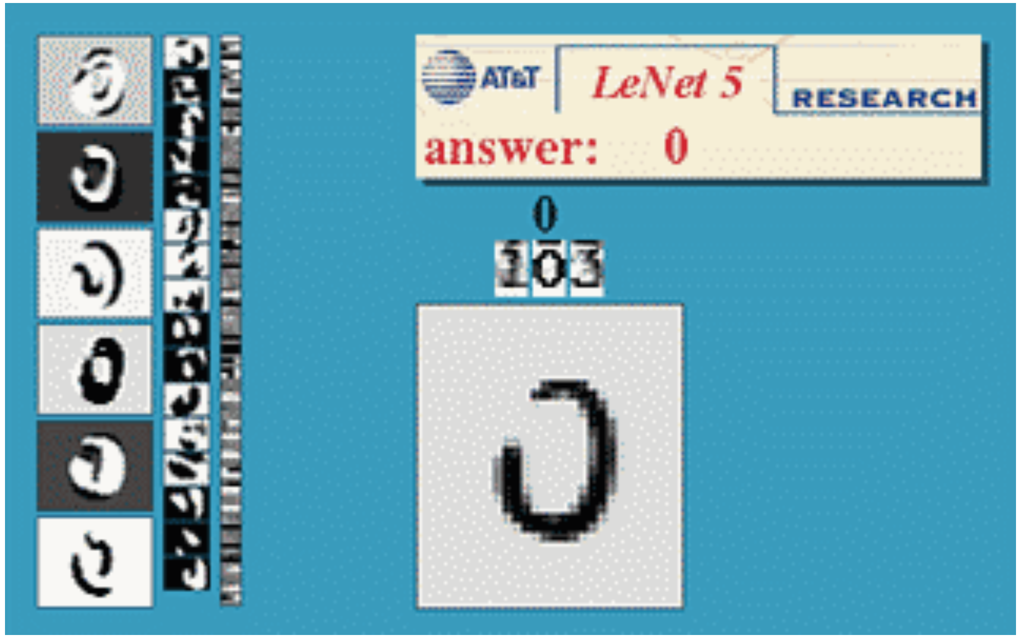
Slide credit: Steve Seitz

Vision-based Biometrics



Optical Character Recognition (OCR)

- Technology to convert scanned docs to text
  - If you have a scanner, it probably came with OCR software

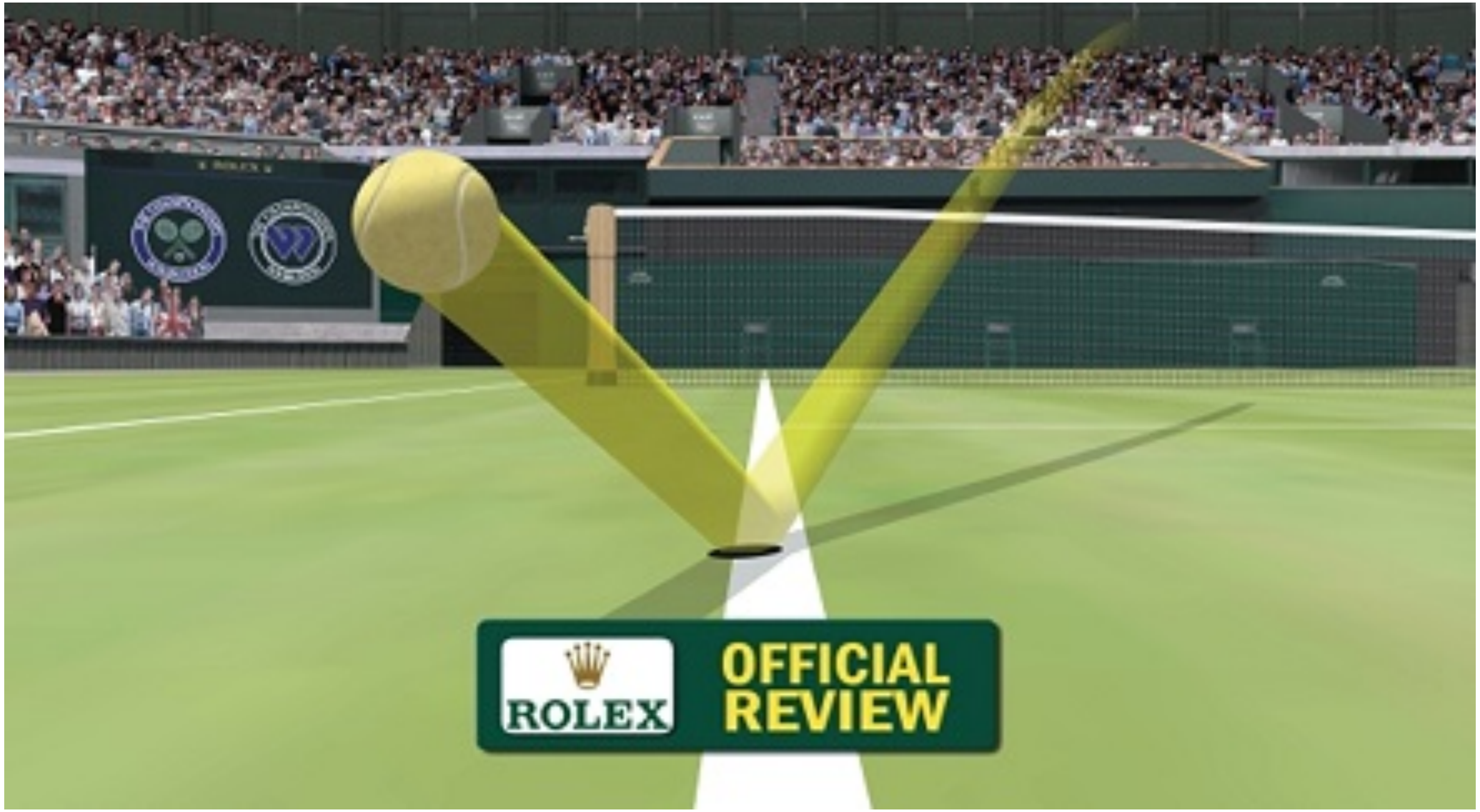


Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Slide credit: Steve Seitz

Computer vision in sports



Hawk-Eye: helping/improving referee decisions

Computer vision in sports

[SportVision](): improving viewer experiences

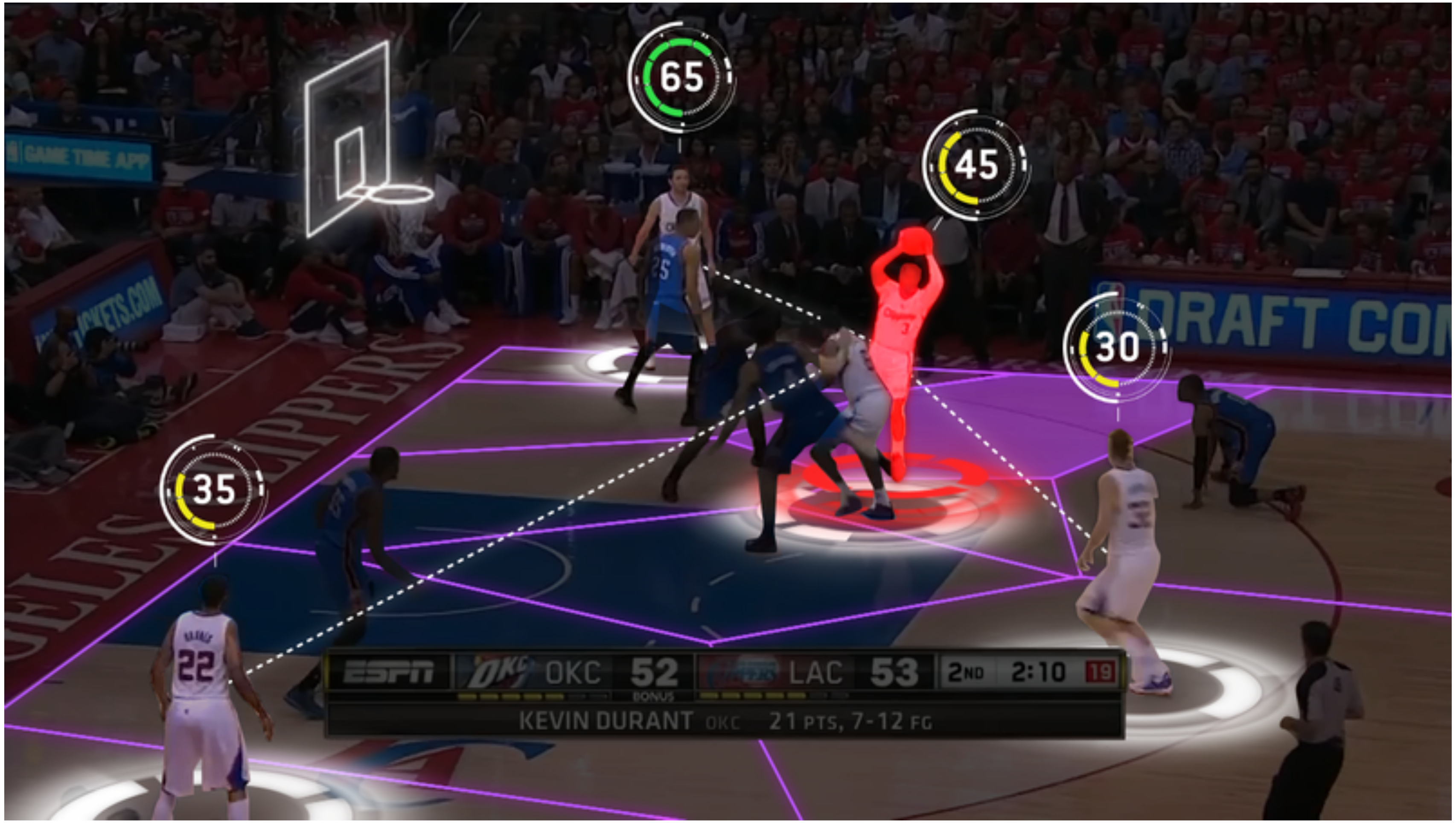
Computer vision in sports

Replay Technologies: improving viewer experiences

Computer vision in sports

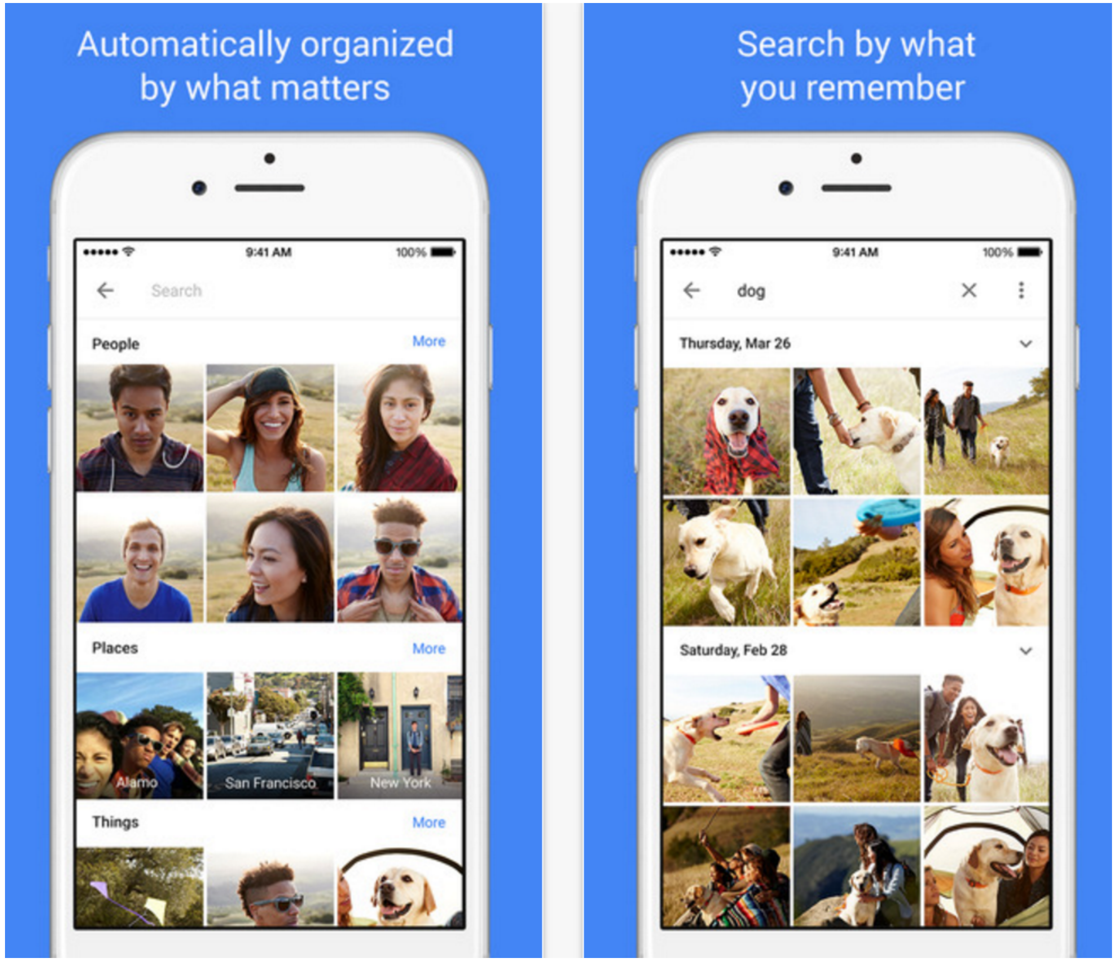
Play tracking

Computer vision in sports

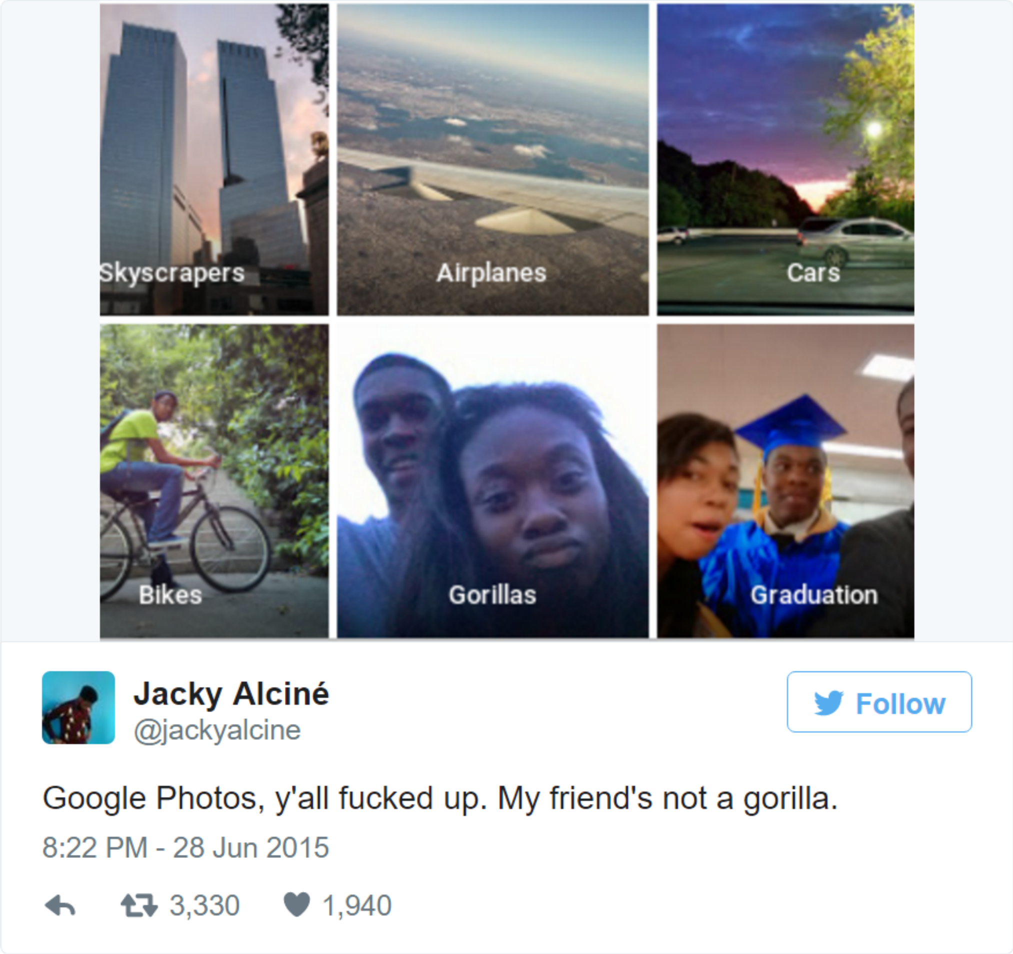


[Second Spectrum](): visual analytics

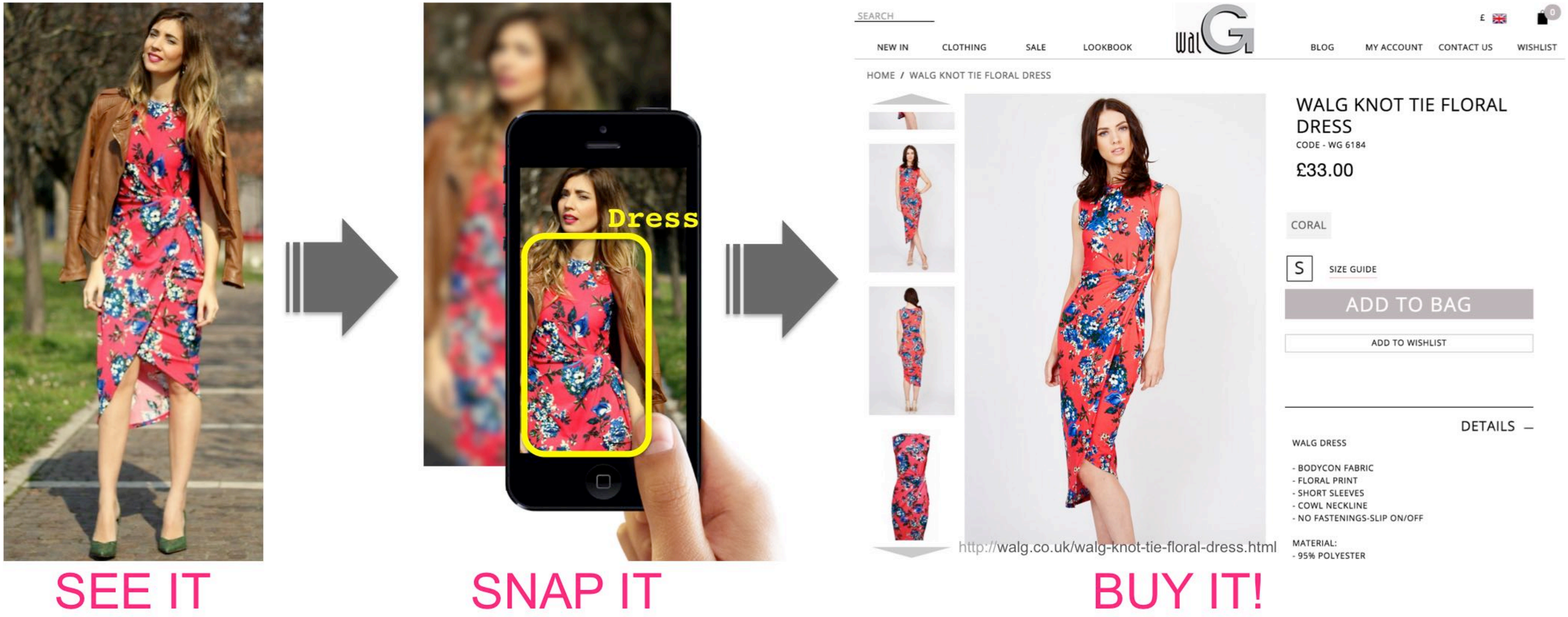
Visual recognition for photo organization



Google photo

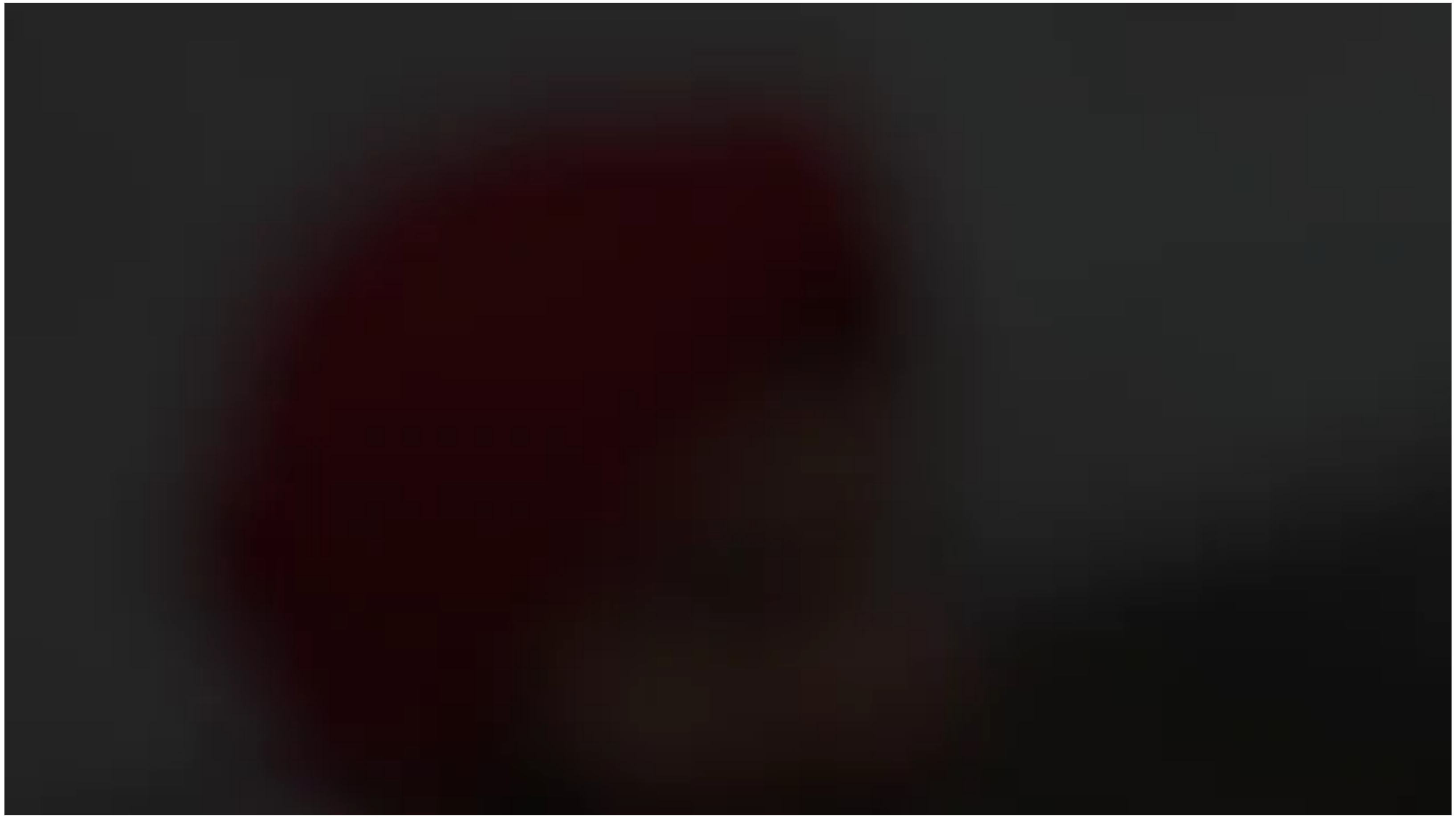


Matching street clothing photos in online Shops



Street2shop, ICCV 2015

3D scanner from a mobile camera



MobileFusion

Earth viewers (3D modeling)

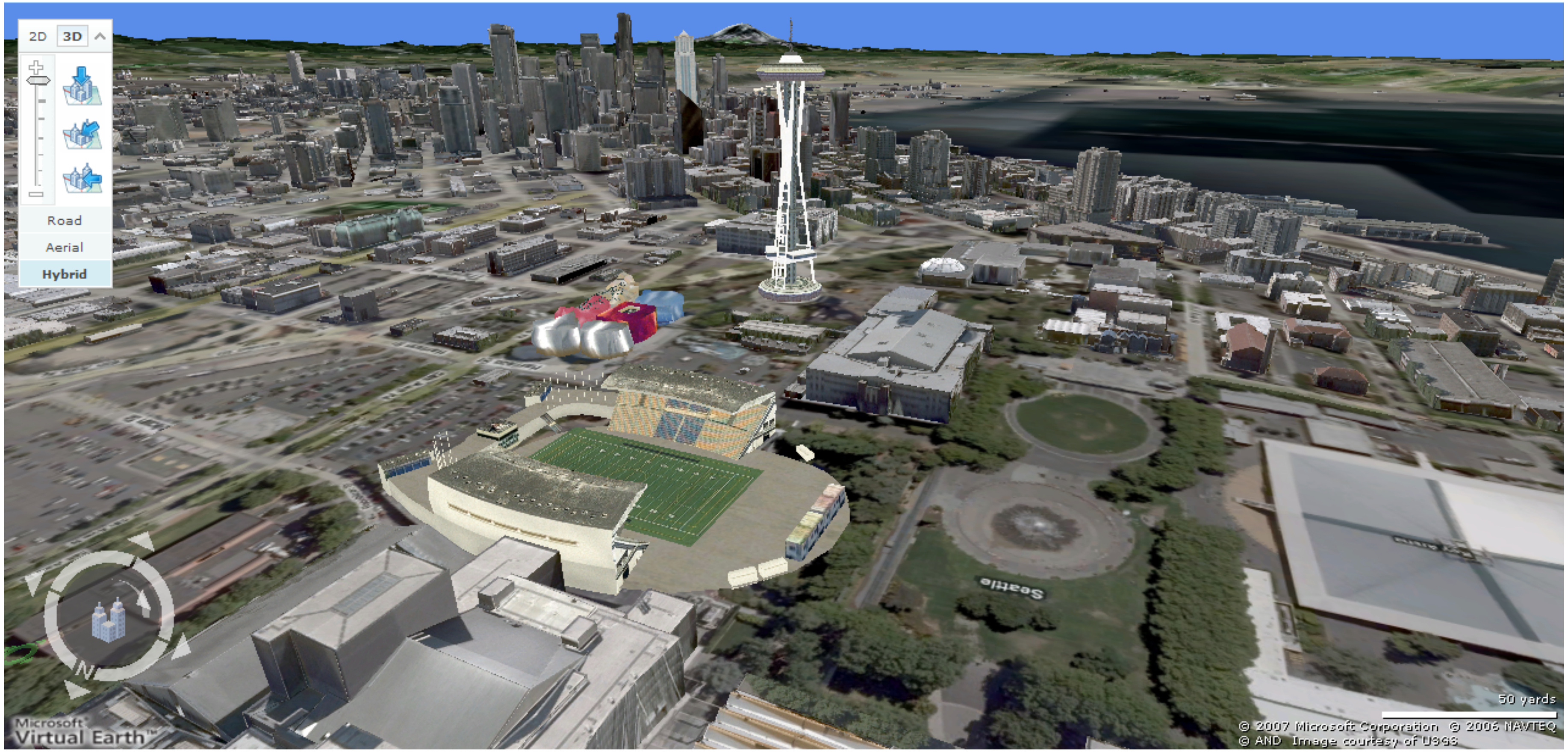


Image from Microsoft's [Virtual Earth](#)
(see also: [Google Earth](#))

Slide credit: Steve Seitz

3D from thousands of images

Building Rome in a Day: Agarwal et al. 2009

3D from thousands of images

[Furukawa et al. CVPR 2010]

Microsoft PhotoSynth: Photo Tourism

MS PhotoSynth in CSI

Indoor Scene Reconstruction

[Structured Indoor Modeling, ICCV2015]

Tour into picture

Adobe Affer Effect

First-person Hyperlapse Videos

[Kopf et al. SIGGRAPH 2014]

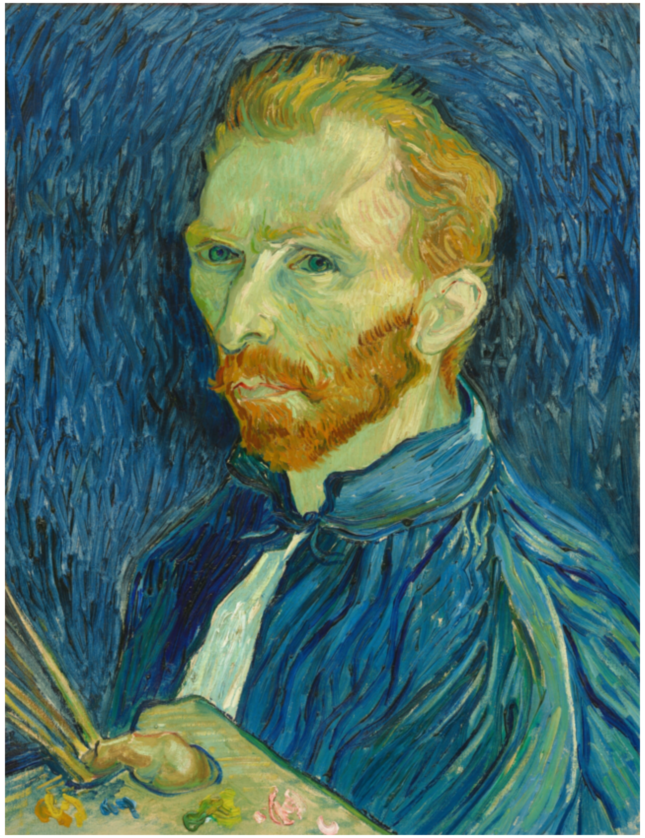
3D Time-lapse from Internet Photos

[3D Time-lapse from Internet Photos, ICCV 2015]

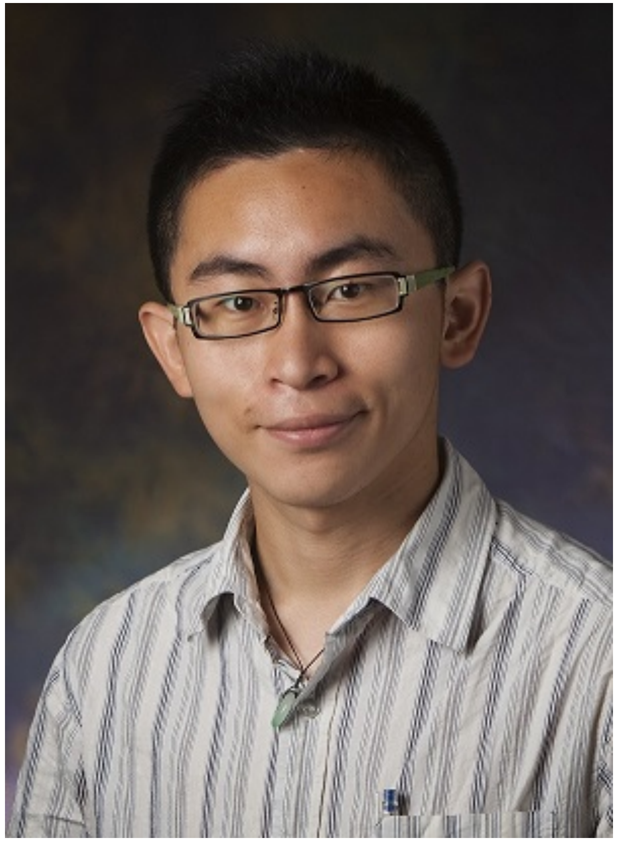
Special effects: matting and composition

[Kylie Minogue - Come Into My World]

Style transfer



Source image (**Style**)



Target image (**Content**)

Output (deepart)

A Neural Algorithm of Artistic Style [Gatys et al. 2015]

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Slide credit: Steve Seitz

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Slide credit: Steve Seitz

Google cars

[Google in talks with Ford, Toyota and Volkswagen to realise driverless cars]

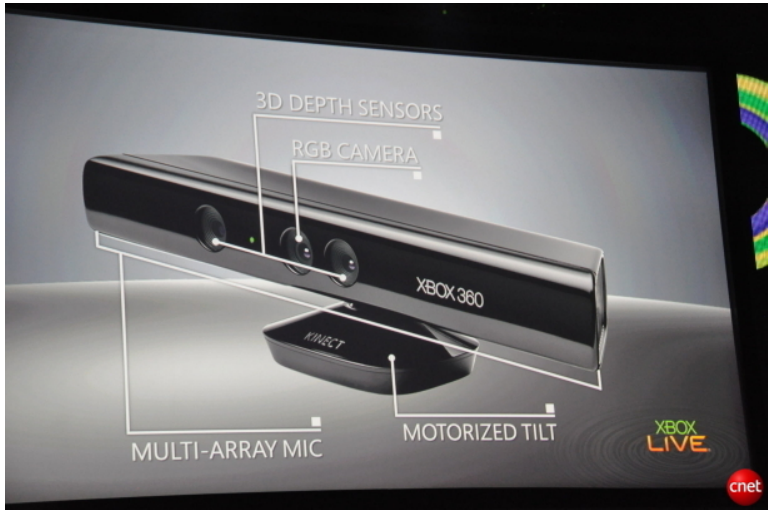
http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871/

Google cars

Oct 9, 2010. "Google Cars Drive Themselves, in Traffic". *The New York Times*. John Markoff
June 24, 2011. "Nevada state law paves the way for driverless cars". *Financial Post*. Christine Dobby
Aug 9, 2011, "Human error blamed after Google's driverless car sparks five-vehicle crash". *The Star* (Toronto)

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks

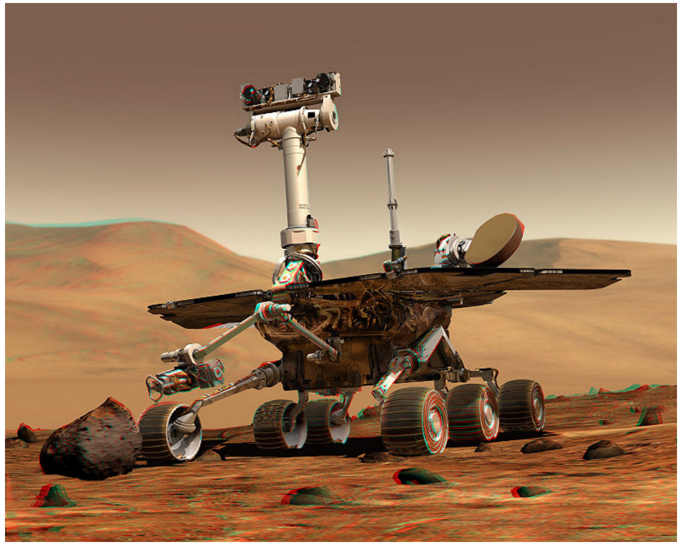
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Industrial robots

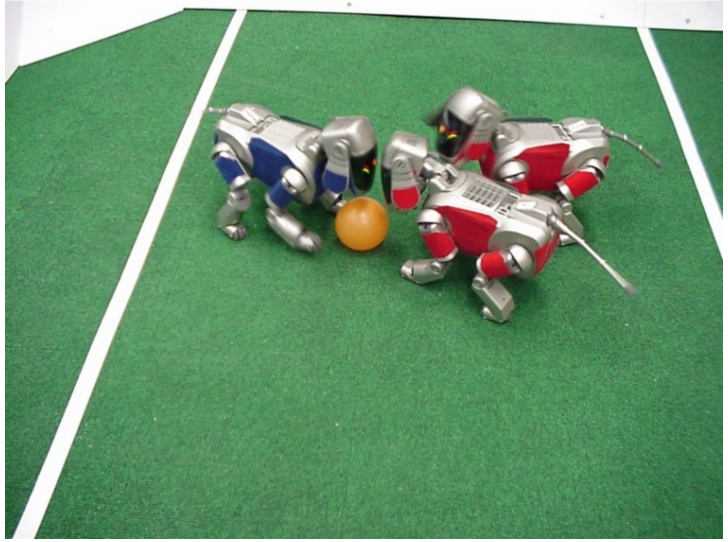
Vision-guided robots position nut runners on wheels

http://www.automationworld.com/computer-vision-opportunity-or-threat

Mobile robots



NASA's Mars Spirit Rover

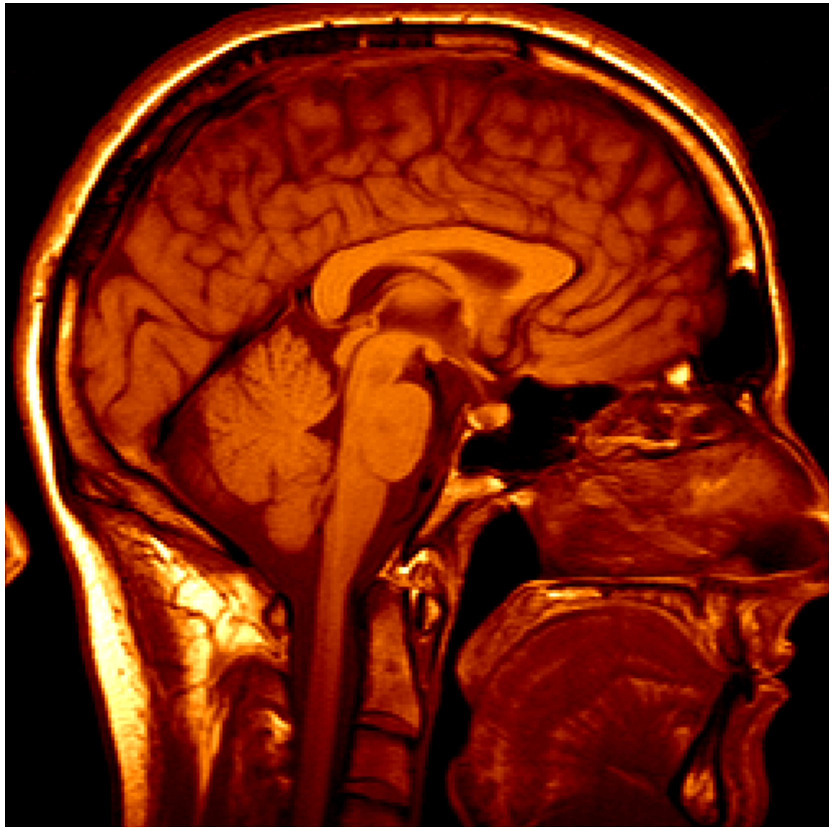


http://www.robocup.org/

Saxena et al. 2008
STAIR at Stanford

http://www.youtube.com/watch?v=DF39Ygp53mQ

Medical imaging



3D imaging
MRI, CT

Image guided surgery
[Grimson et al., MIT]

Computer vision for healthcare

BiliCam, Ubicomp 2014

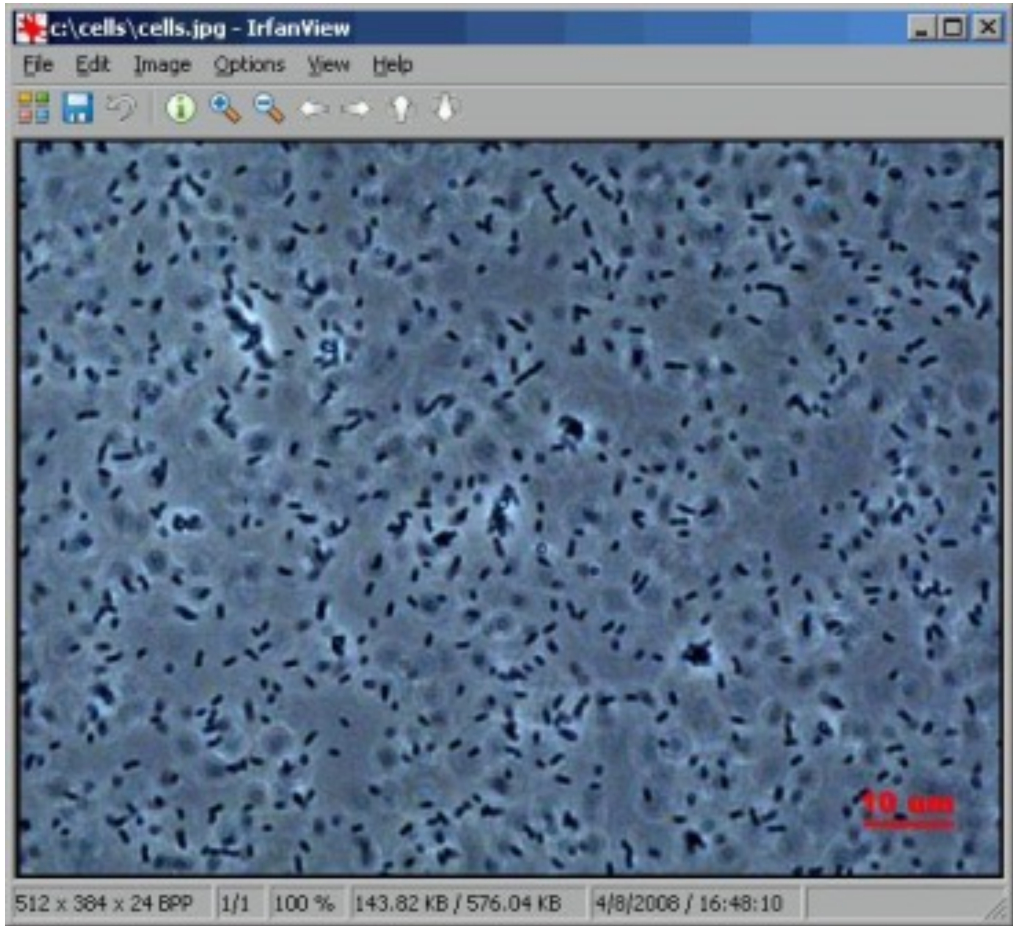
Computer vision for healthcare

[Autism Screening]

Computer vision for healthcare

[Video magnification]

Computer vision for the mass



Counting cells

Analyzing social effects of green space

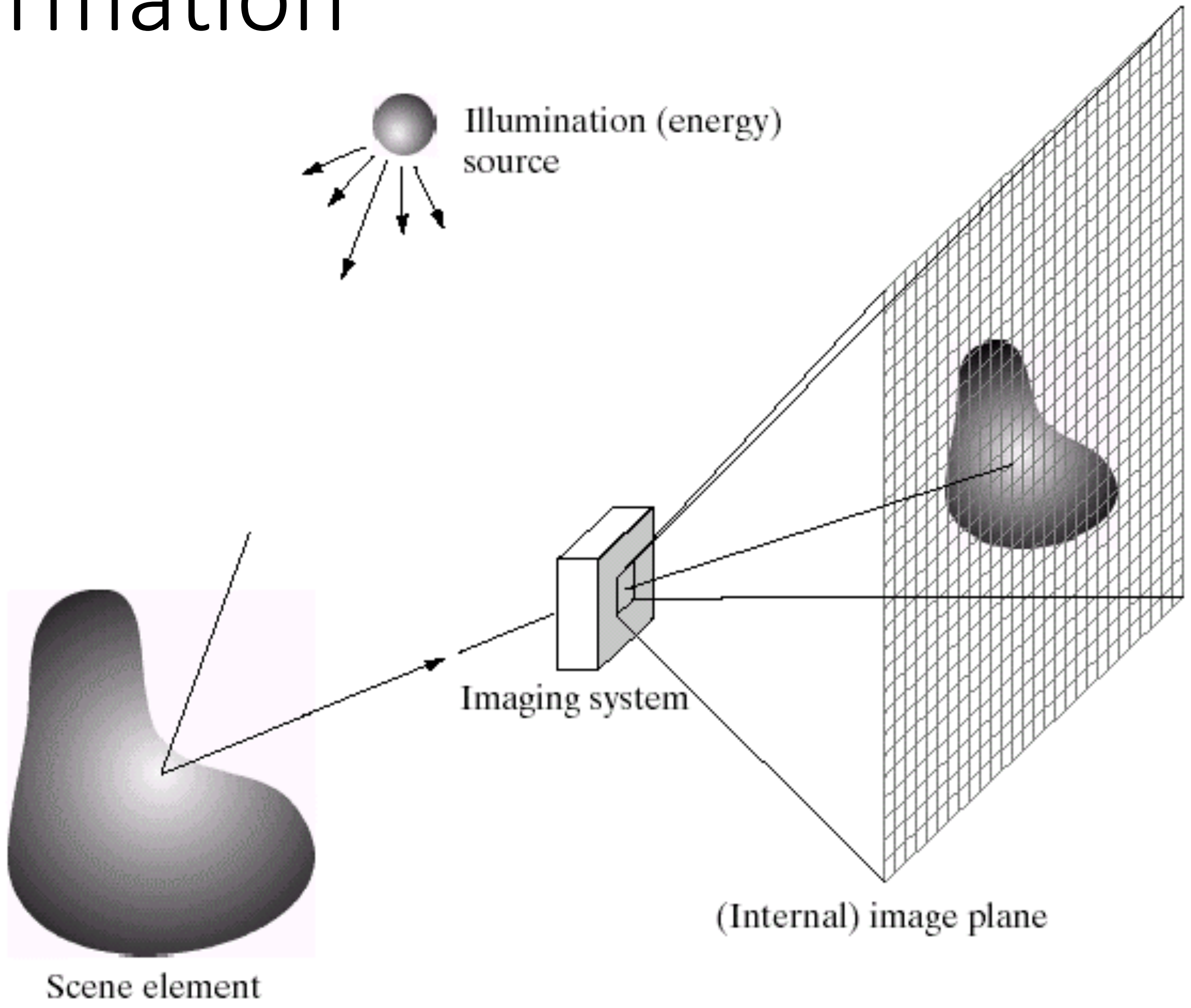
20-mins coffee break

Fundamentals of Computer Vision

- Light
  - What an image records
- Matching
  - How to measure the similarity of two regions
- Alignment
  - How to align points/patches
  - How to recover transformation parameters based on matched points
- Geometry
  - How to relate world coordinates and image coordinates
- Grouping
  - What points/regions/lines belong together?
- Categorization
  - What similarities are important?

Slide credit: Derek Hoiem

Image Formation



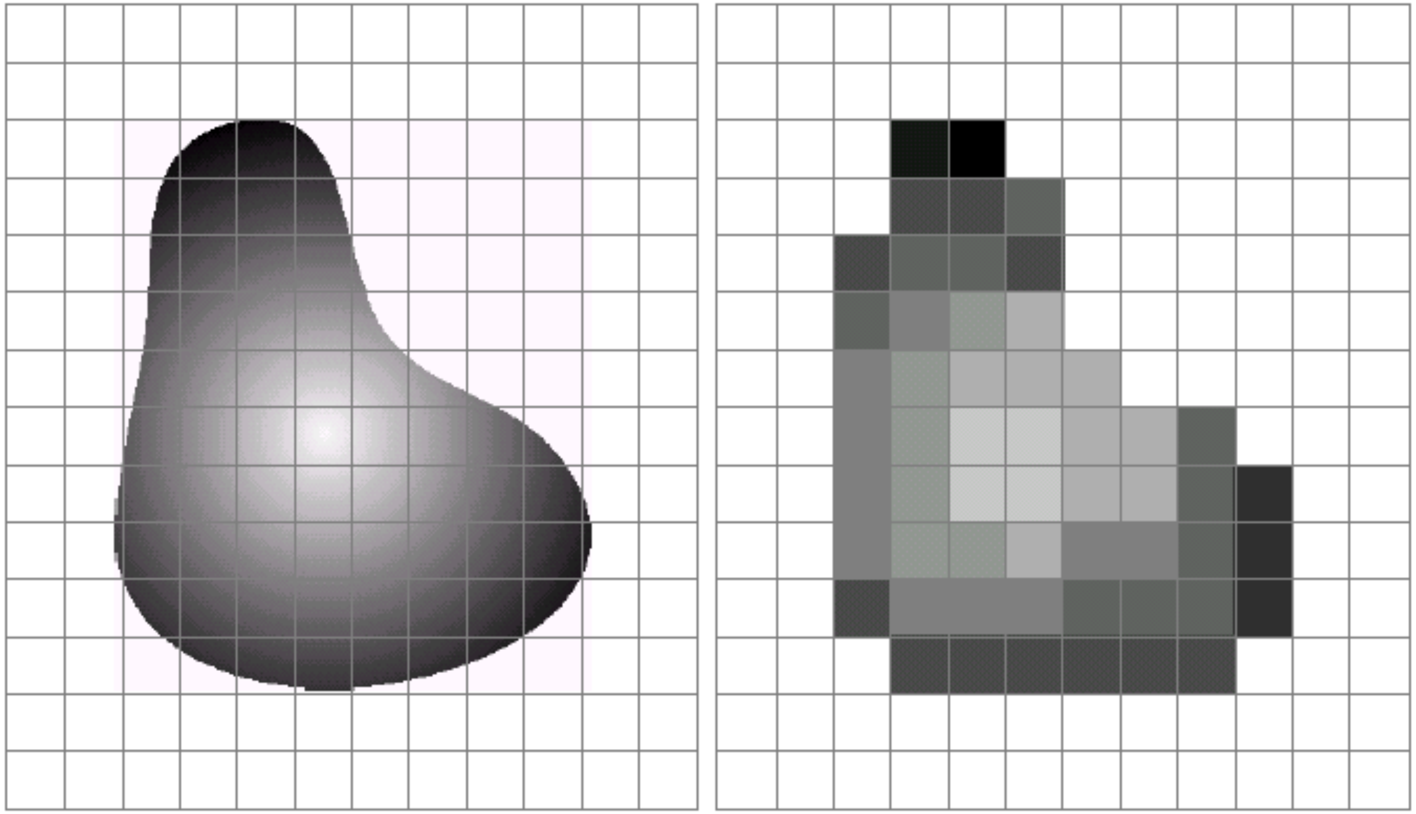
Digital camera

A digital camera replaces film with a sensor array

- Each cell in the array is light-sensitive diode that converts photons to electrons
- Two common types:
  - Charge Coupled Device (CCD): larger yet slower, better quality
  - Complementary Metal Oxide Semiconductor (CMOS): high bandwidth, lower quality
- http://electronics.howstuffworks.com/digital-camera.htm

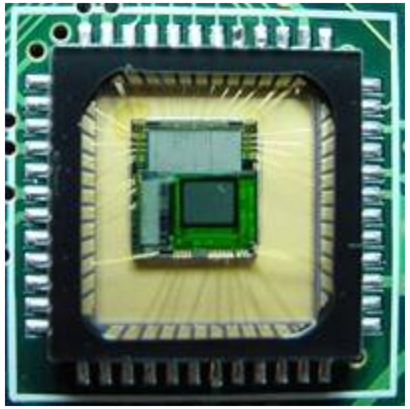
Slide by Steve Seitz

Sensor Array

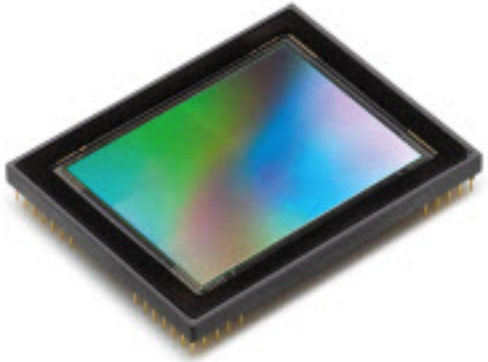


a b

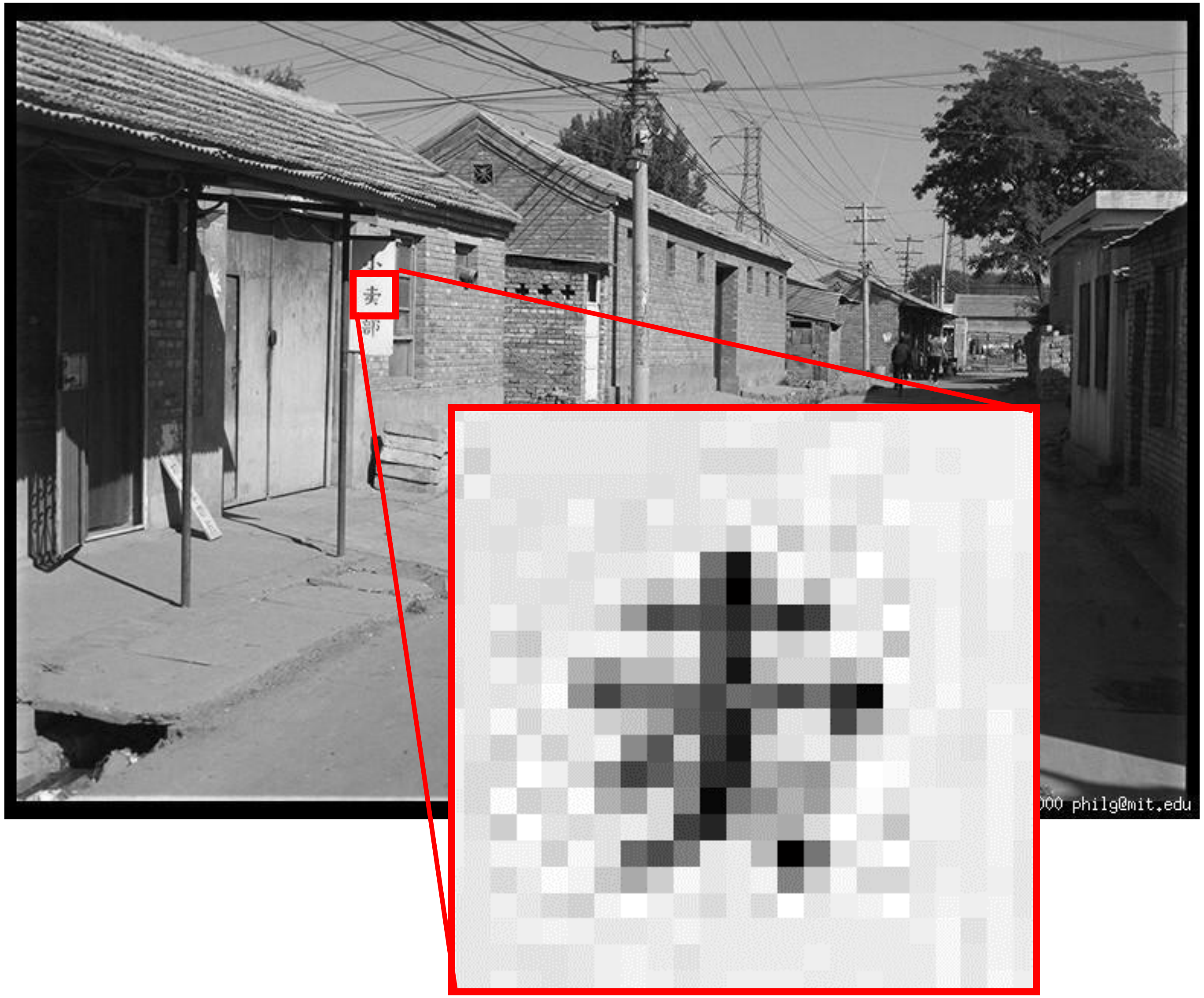
FIGURE 2.17 (a) Continuos image projected onto a sensor array. (b) Result of image sampling and quantization.

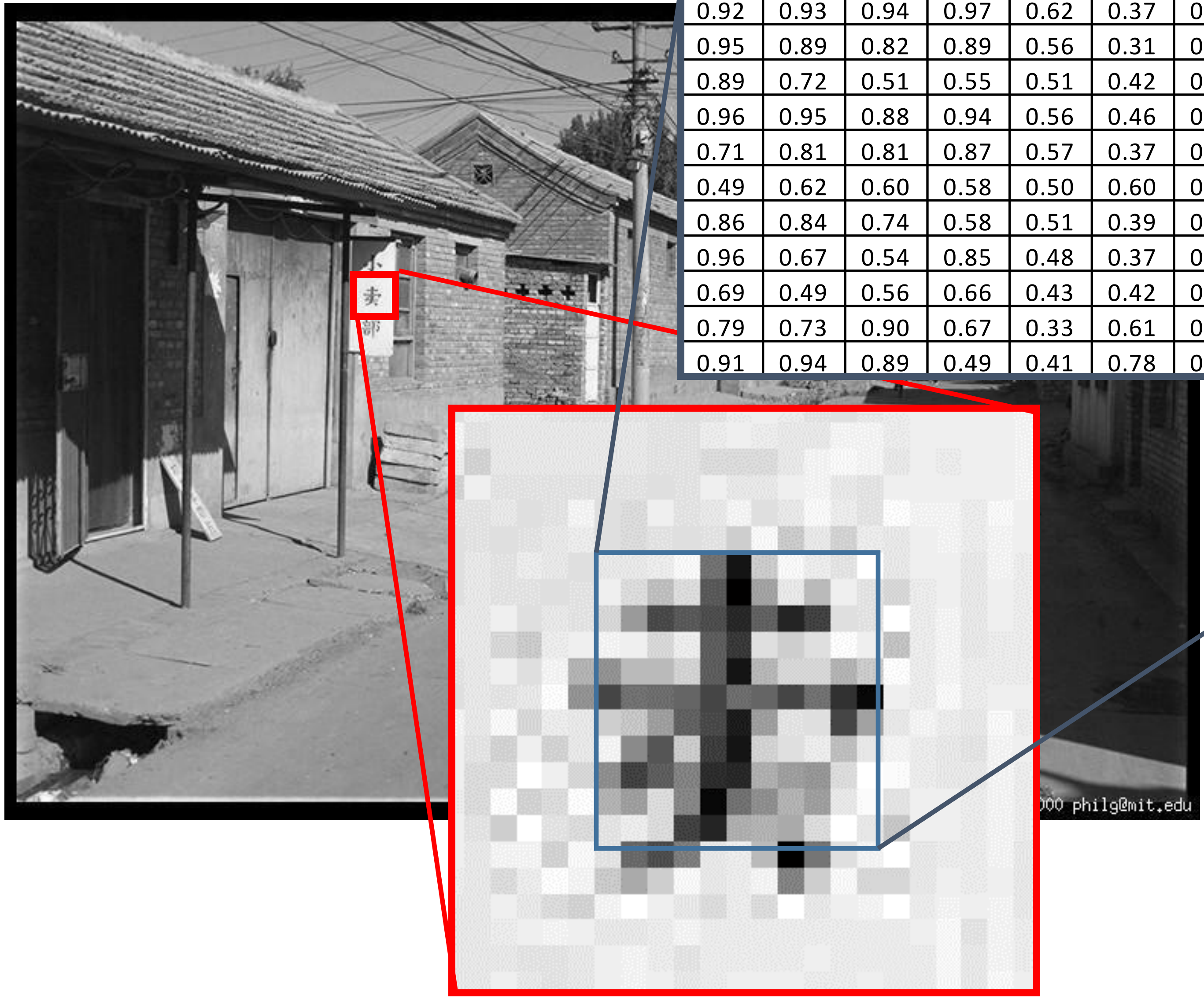


CMOS sensor



CCD sensor

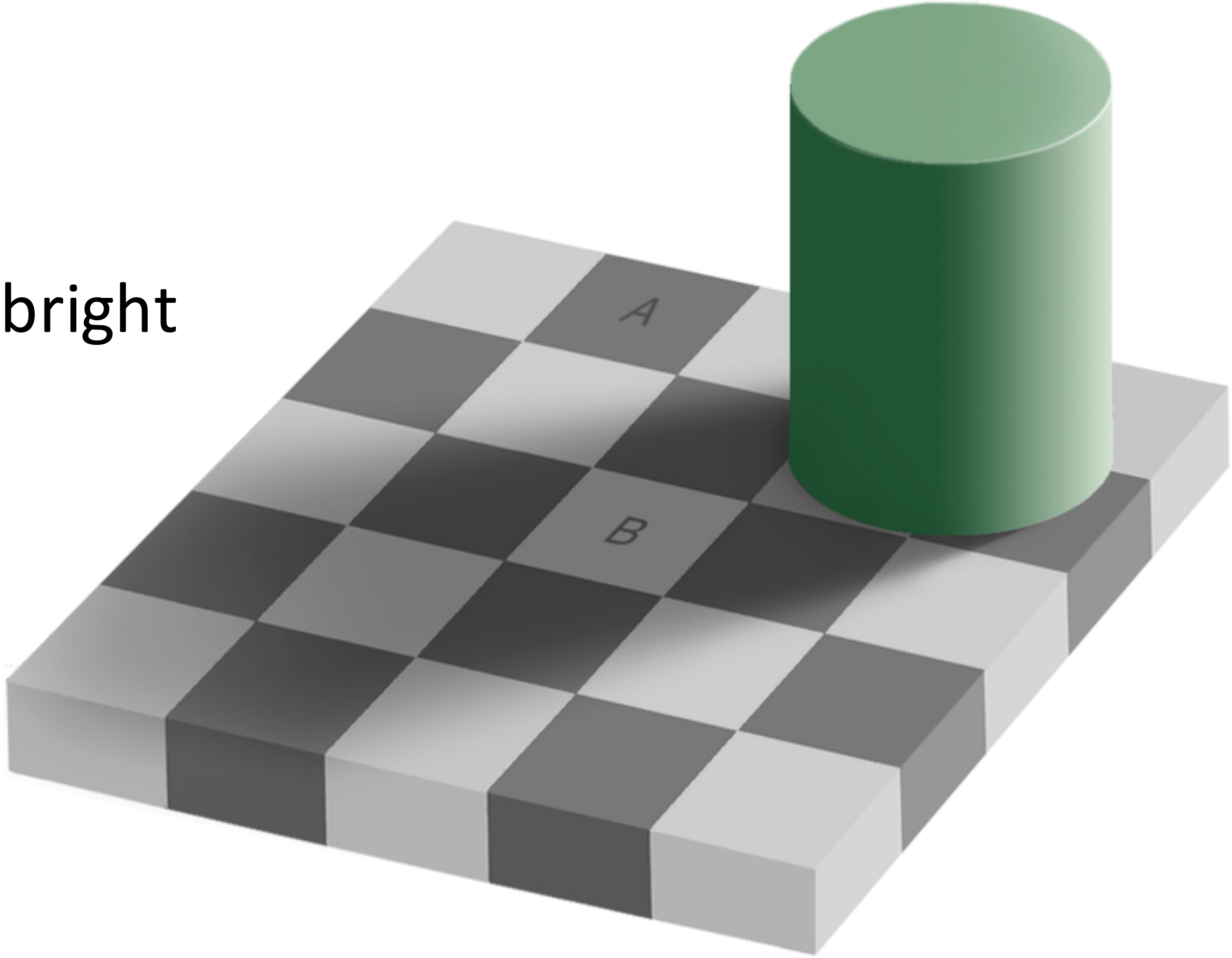

philg@mit.edu



0.92	0.93	0.94	0.97	0.62	0.37	0.85	0.97	0.93	0.92	0.99
0.95	0.89	0.82	0.89	0.56	0.31	0.75	0.92	0.81	0.95	0.91
0.89	0.72	0.51	0.55	0.51	0.42	0.57	0.41	0.49	0.91	0.92
0.96	0.95	0.88	0.94	0.56	0.46	0.91	0.87	0.90	0.97	0.95
0.71	0.81	0.81	0.87	0.57	0.37	0.80	0.88	0.89	0.79	0.85
0.49	0.62	0.60	0.58	0.50	0.60	0.58	0.50	0.61	0.45	0.33
0.86	0.84	0.74	0.58	0.51	0.39	0.73	0.92	0.91	0.49	0.74
0.96	0.67	0.54	0.85	0.48	0.37	0.88	0.90	0.94	0.82	0.93
0.69	0.49	0.56	0.66	0.43	0.42	0.77	0.73	0.71	0.90	0.99
0.79	0.73	0.90	0.67	0.33	0.61	0.69	0.79	0.73	0.93	0.97
0.91	0.94	0.89	0.49	0.41	0.78	0.78	0.77	0.89	0.99	0.93

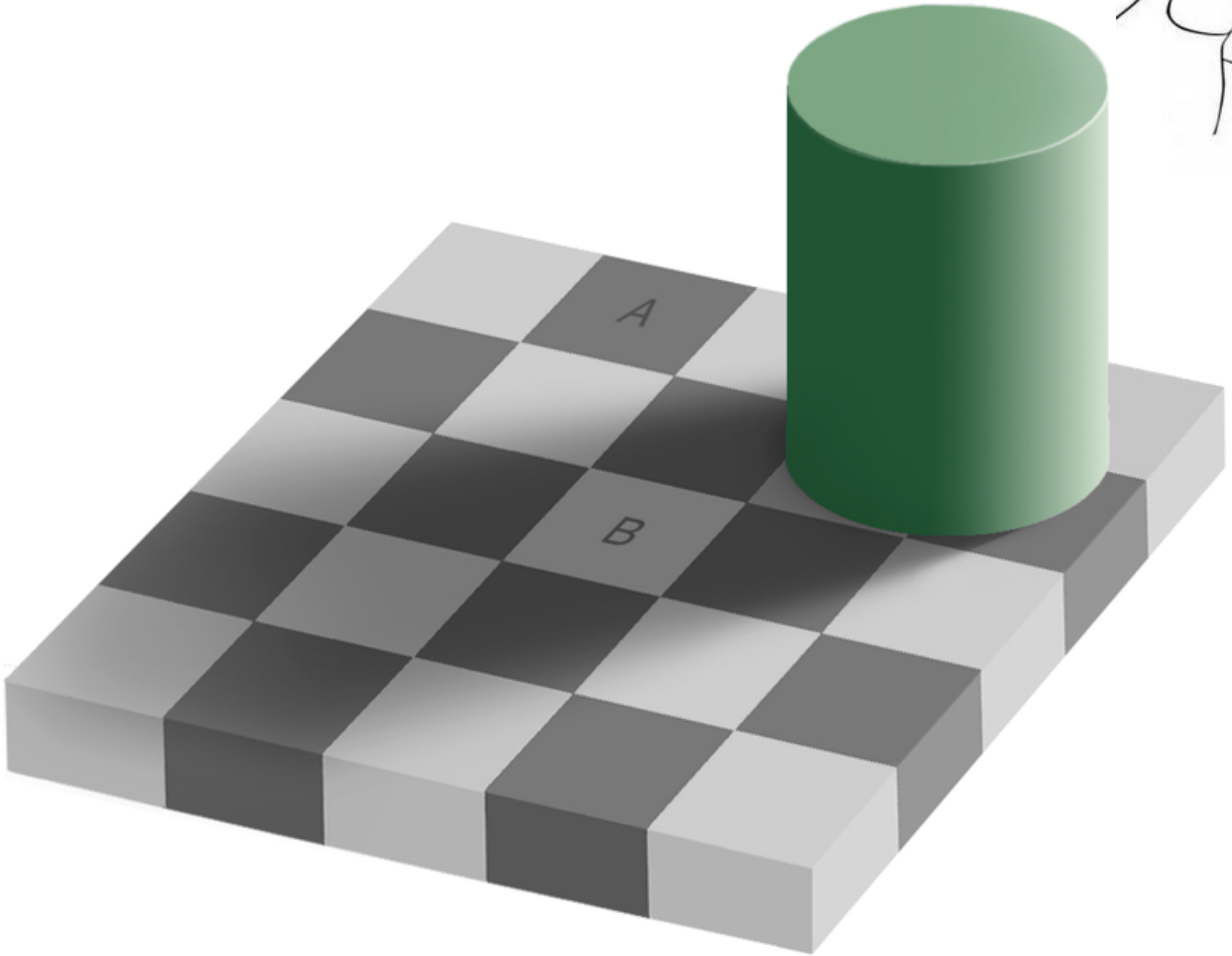
Perception of Intensity

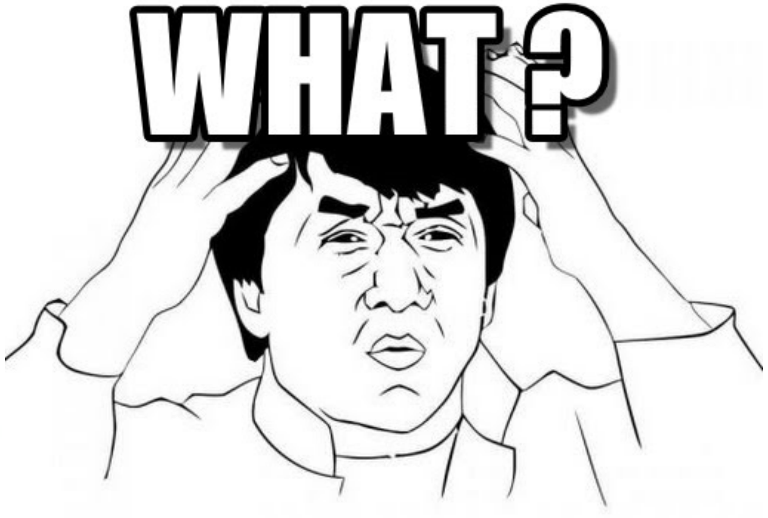
- A is brighter?
- B is brighter?
- They are equally bright



from Ted Adelson

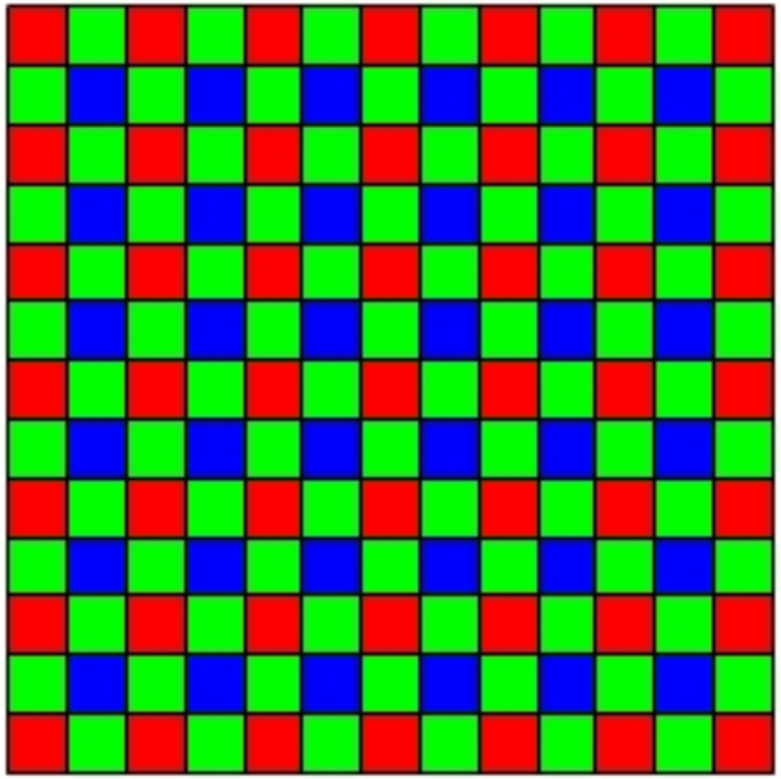
Perception of Intensity





from Ted Adelson

Digital Color Images


Bayer filter

Color Image

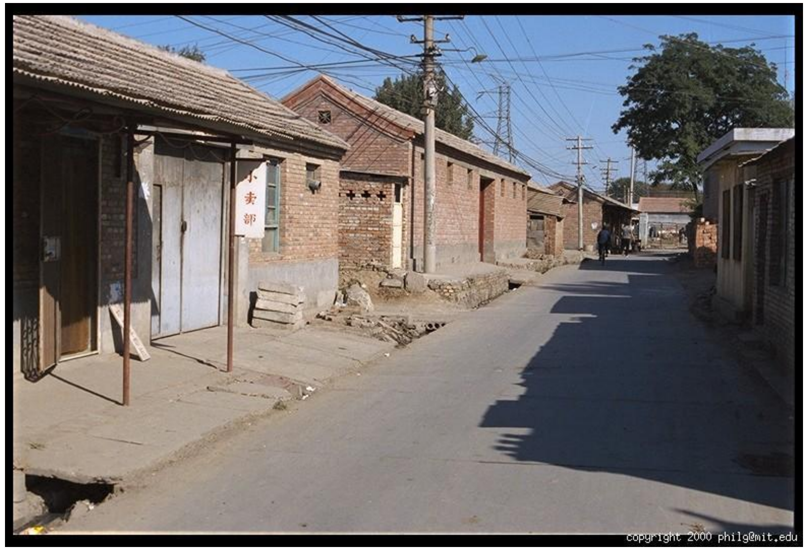


Image filtering

- Linear filtering: function is a weighted sum/difference of pixel values

- Really important!
  - Enhance images
    - Denoise, smooth, increase contrast, etc.
  - Extract information from images
    - Texture, edges, distinctive points, etc.
  - Detect patterns
    - Template matching

Example: box filter

$$g[\cdot\,,\cdot\,]$$

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Slide credit: David Lowe (UBC)

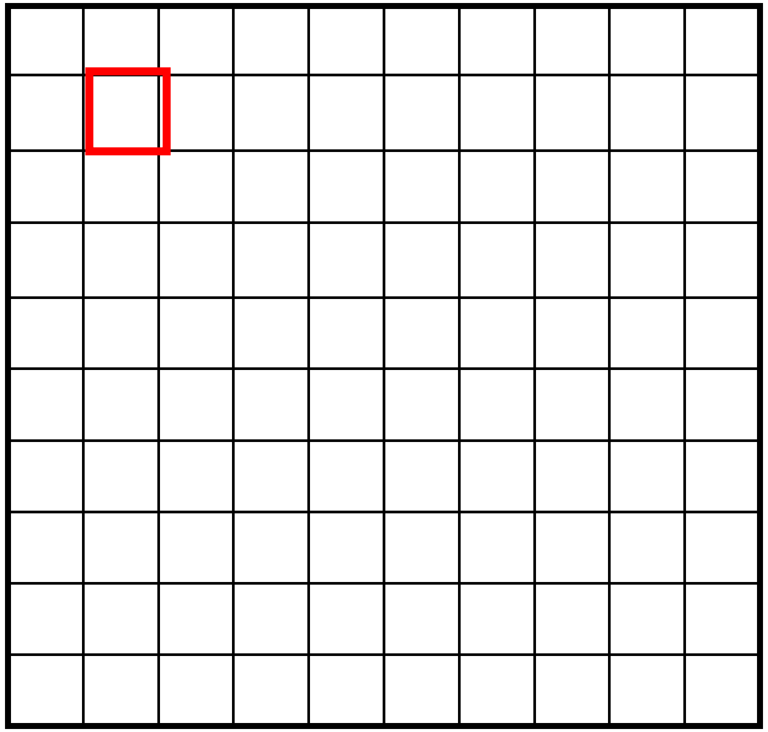
Image filtering

$$g[\cdot,\cdot]\ \frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$$

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$



$$h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$$

Credit: S. Seitz

Image filtering

$g[\cdot,\cdot]\ \frac{1}{9}$

1	1	1
1	1	1
1	1	1

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10							

$$h[m,n] = \sum_{k,l} g[k,l]\, f[m+k, n+l]$$

Credit: S. Seitz

Image filtering

$$g[\cdot,\cdot]\ \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10	20						

$$h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$$

Credit: S. Seitz

Image filtering

$g[\cdot,\cdot]\ \frac{1}{9}$

1	1	1
1	1	1
1	1	1

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10	20	30					

$$h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$$

Credit: S. Seitz

Image filtering

$g[\cdot,\cdot]\ \frac{1}{9}$

1	1	1
1	1	1
1	1	1

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10	20	30	30				

$$h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$$

Credit: S. Seitz

Image filtering

$$g[\cdot,\cdot]\ \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10	20	30	30				
				?					

$$h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$$

Credit: S. Seitz

Image filtering

$$g[\cdot,\cdot]\ \frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$$

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10	20	30	30				
						?			
				50					

$$h[m,n]=\sum_{k,l} g[k,l]\, f[m+k,n+l]$$

Credit: S. Seitz

Image filtering

$g[\cdot,\cdot]$ $\frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}$

$f[.,.]$

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	90	0	90	90	90	0	0
0	0	0	90	90	90	90	90	0	0
0	0	0	0	0	0	0	0	0	0
0	0	90	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

$h[.,.]$

	0	10	20	30	30	30	20	10	
	0	20	40	60	60	60	40	20	
	0	30	60	90	90	90	60	30	
	0	30	50	80	80	90	60	30	
	0	30	50	80	80	90	60	30	
	0	20	30	50	50	60	40	20	
	10	20	30	30	30	30	20	10	
	10	10	10	0	0	0	0	0	

$$h[m,n] = \sum_{k,l} g[k,l]\, f[m+k, n+l]$$

Credit: S. Seitz

Box Filter

What does it do?

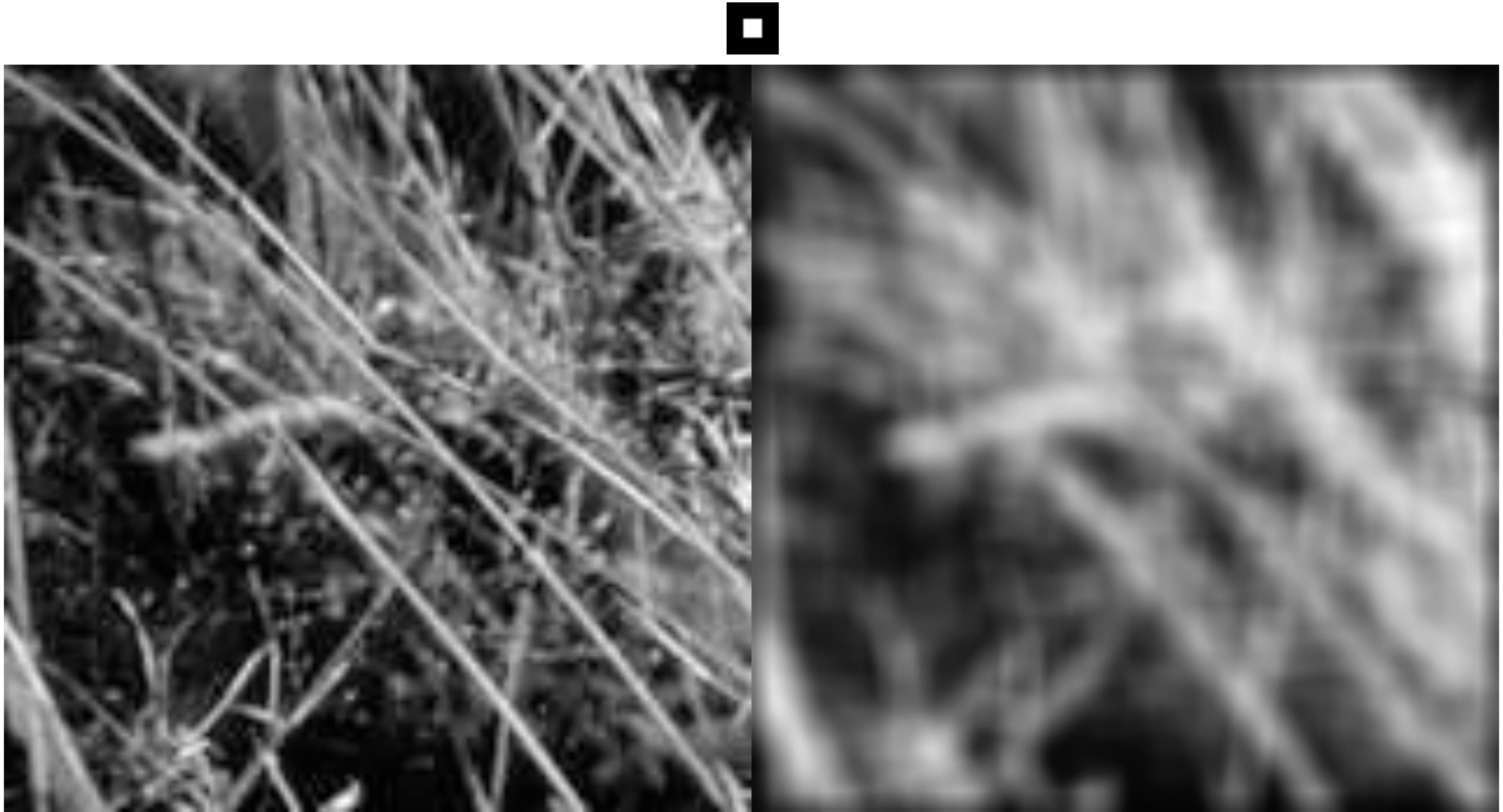
- Replaces each pixel with an average of its neighborhood
- Achieve smoothing effect (remove sharp features)

$$g[\cdot\,,\cdot\,]$$

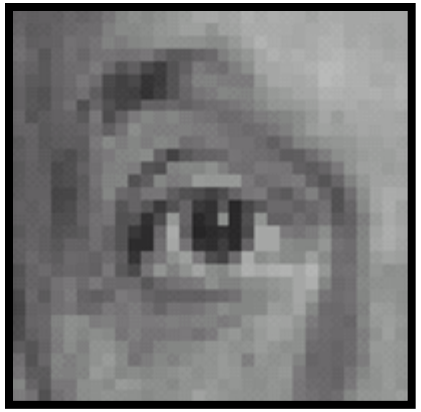
$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Slide credit: David Lowe (UBC)

Smoothing with box filter



Practice with linear filters



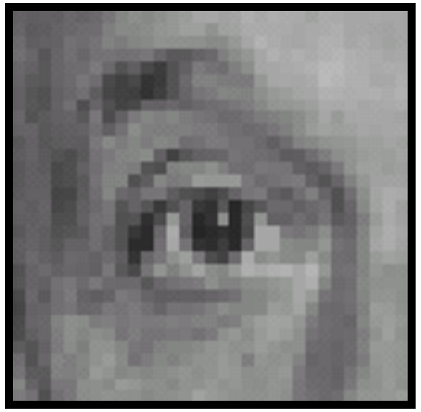
Original

0	0	0
0	1	0
0	0	0

?

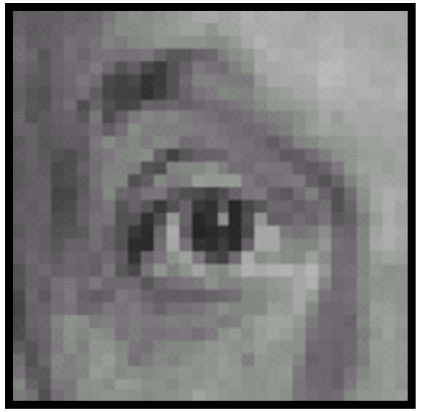
Source: D. Lowe

Practice with linear filters



Original

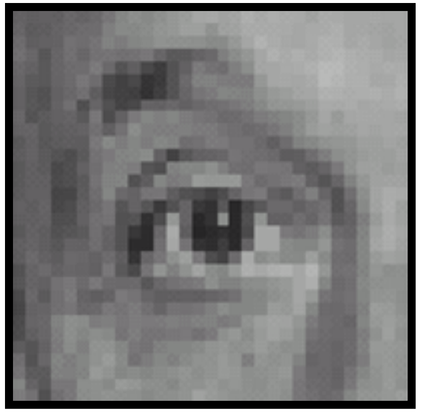
0	0	0
0	1	0
0	0	0



Filtered
(no change)

Source: D. Lowe

Practice with linear filters



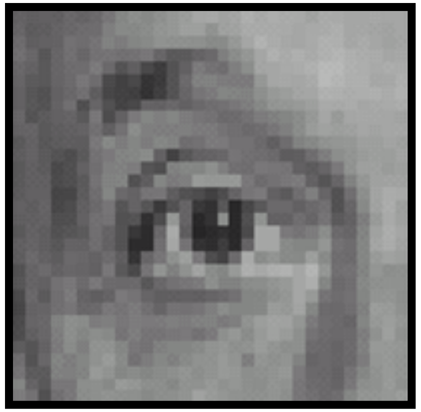
Original

0	0	0
0	0	1
0	0	0

?

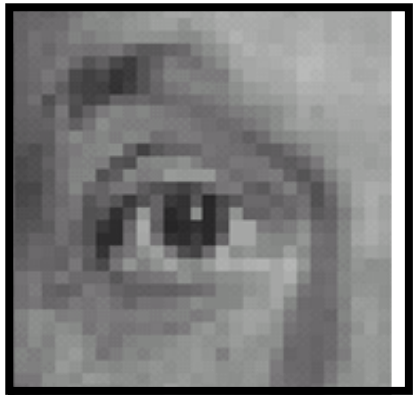
Source: D. Lowe

Practice with linear filters



Original

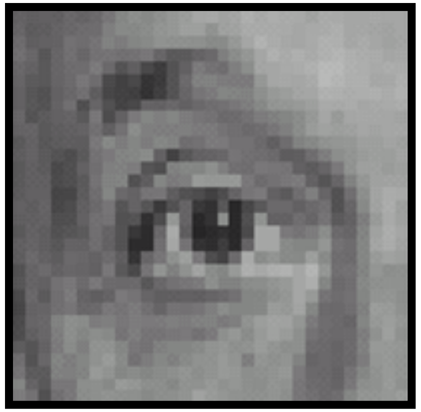
0	0	0
0	0	1
0	0	0



Shifted left
By 1 pixel

Source: D. Lowe

Practice with linear filters



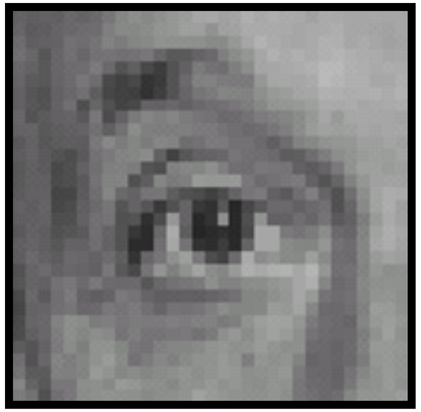
Original

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad ?$$

(Note that filter sums to 1)

Source: D. Lowe

Practice with linear filters



Original

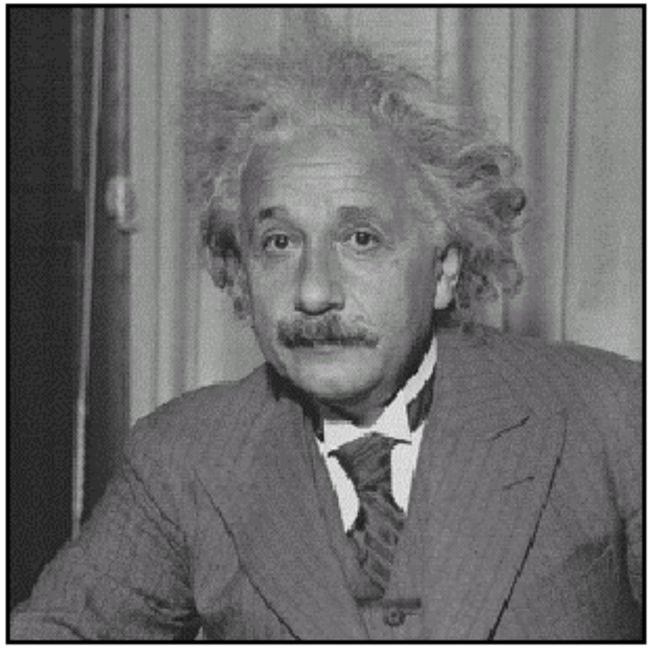
$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} - \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Sharpening filter

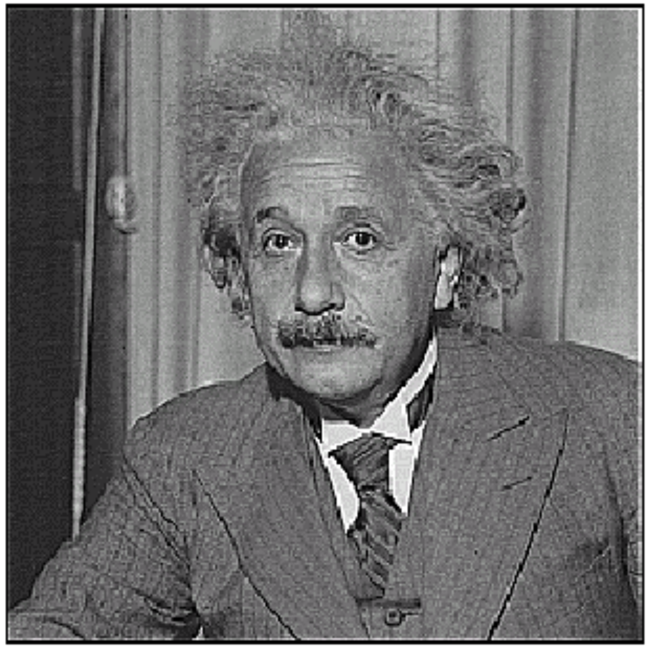
- Accentuates differences with local average

Source: D. Lowe

Sharpening



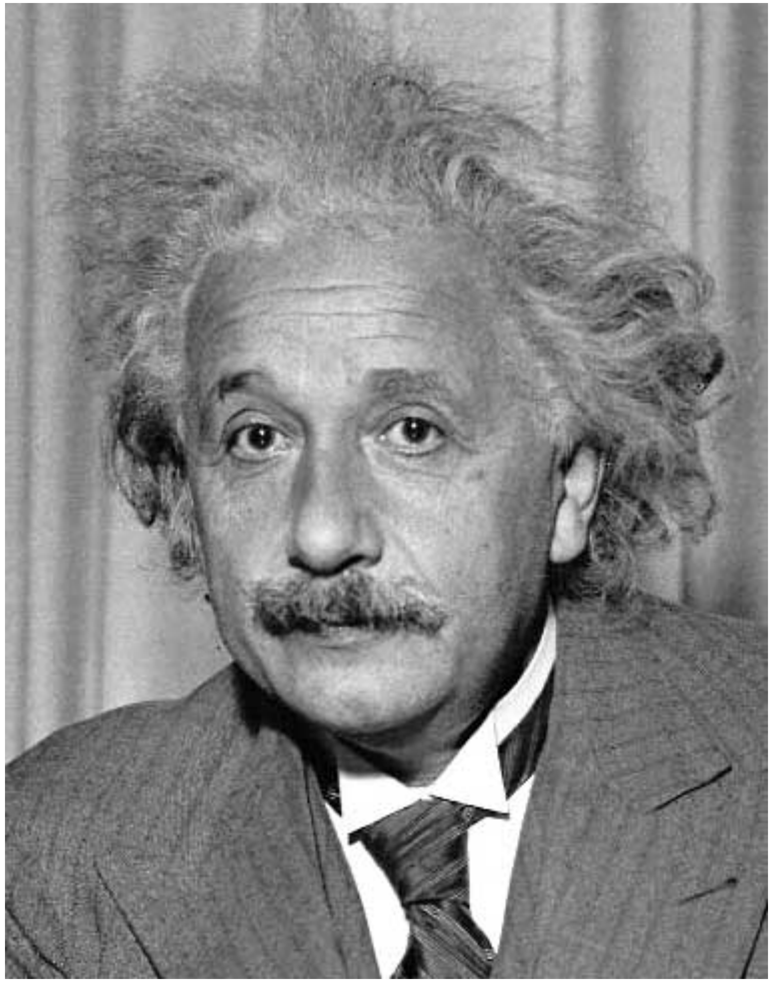
before



after

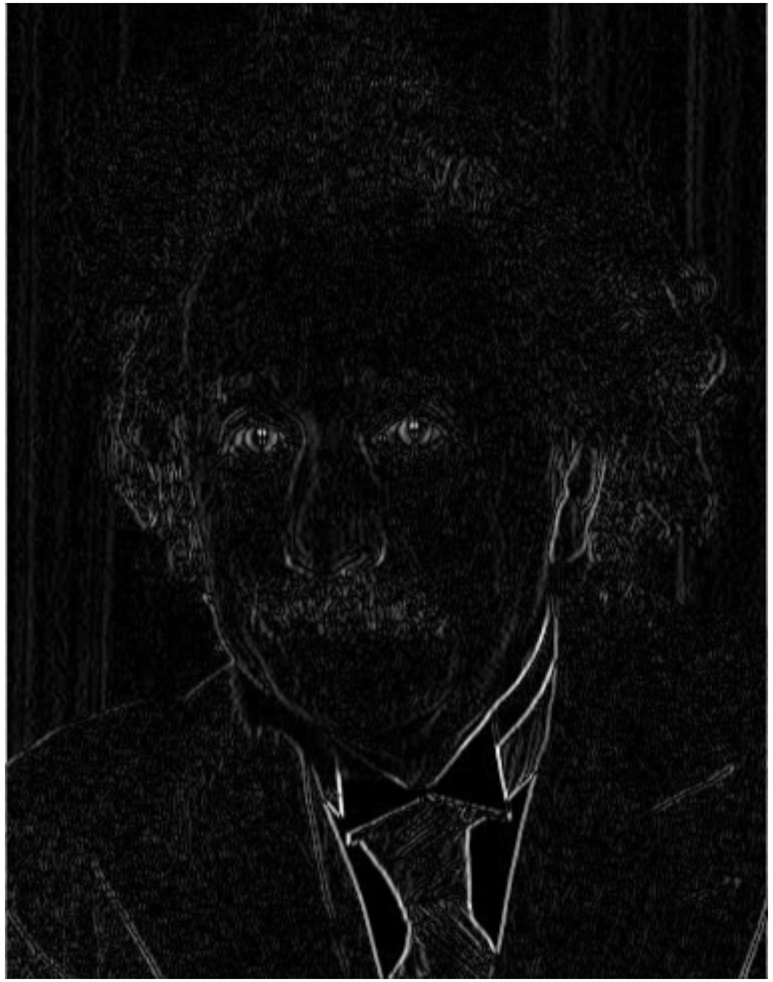
Source: D. Lowe

Other filters



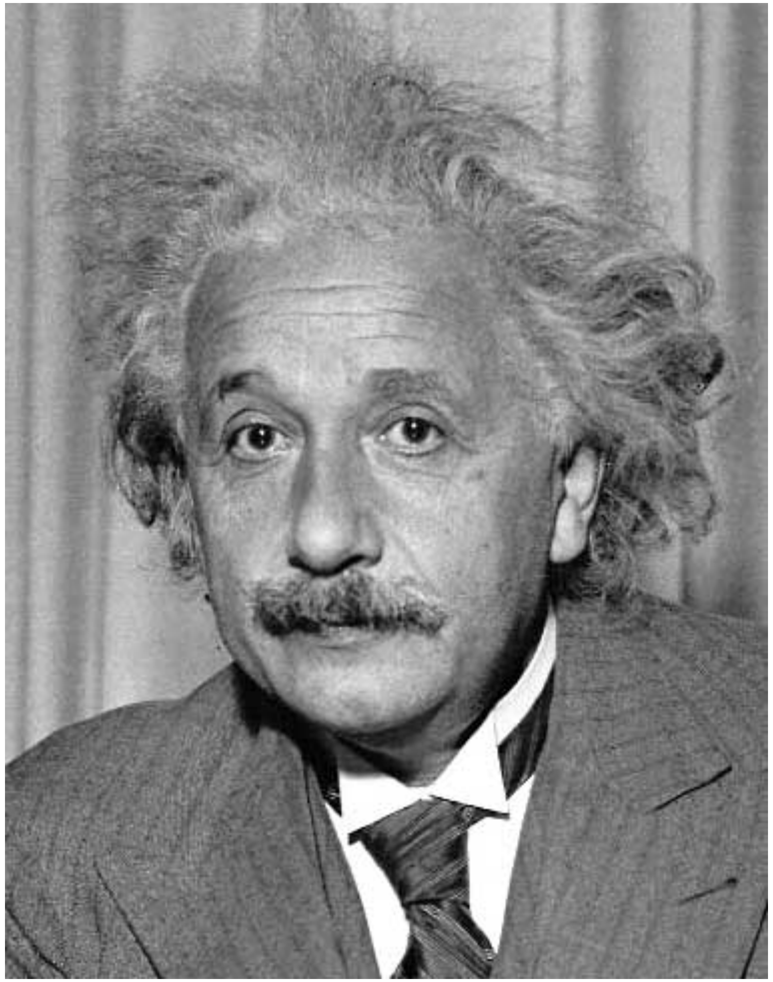
1	0	-1
2	0	-2
1	0	-1

Sobel



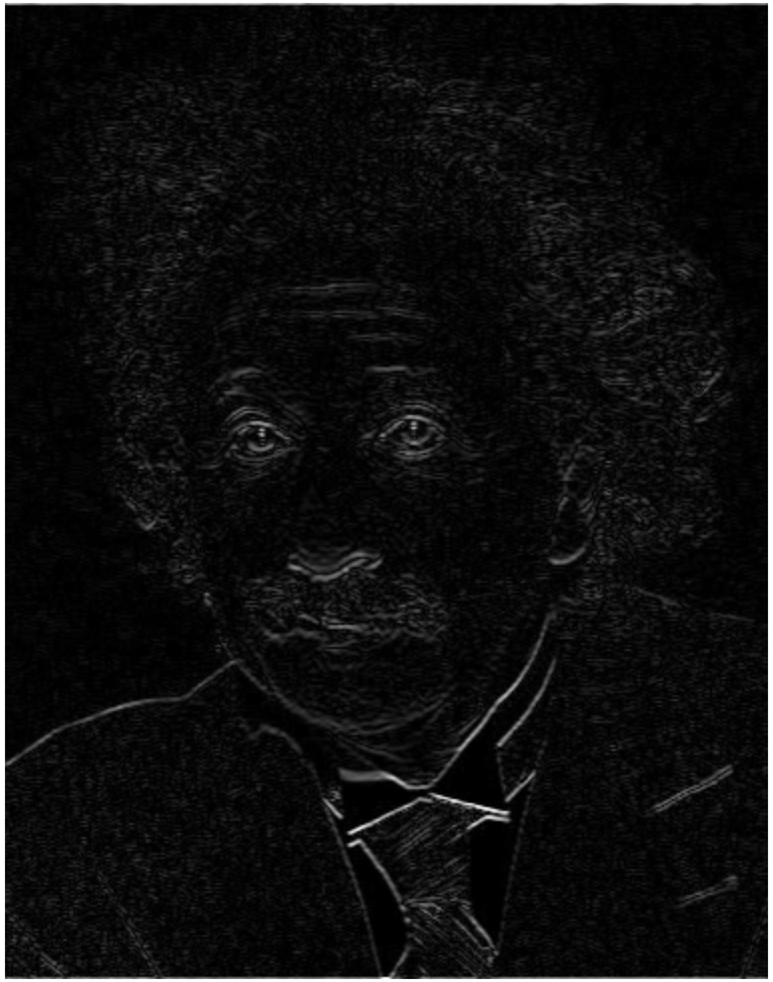
Vertical Edge
(absolute value)

Other filters



1	2	1
0	0	0
-1	-2	-1

Sobel



Horizontal Edge
(absolute value)

Basic gradient filters

Horizontal Gradient

0	0	0
-1	0	1
0	0	0

or

-1	0	1

Vertical Gradient

0	-1	0
0	0	0
0	1	0

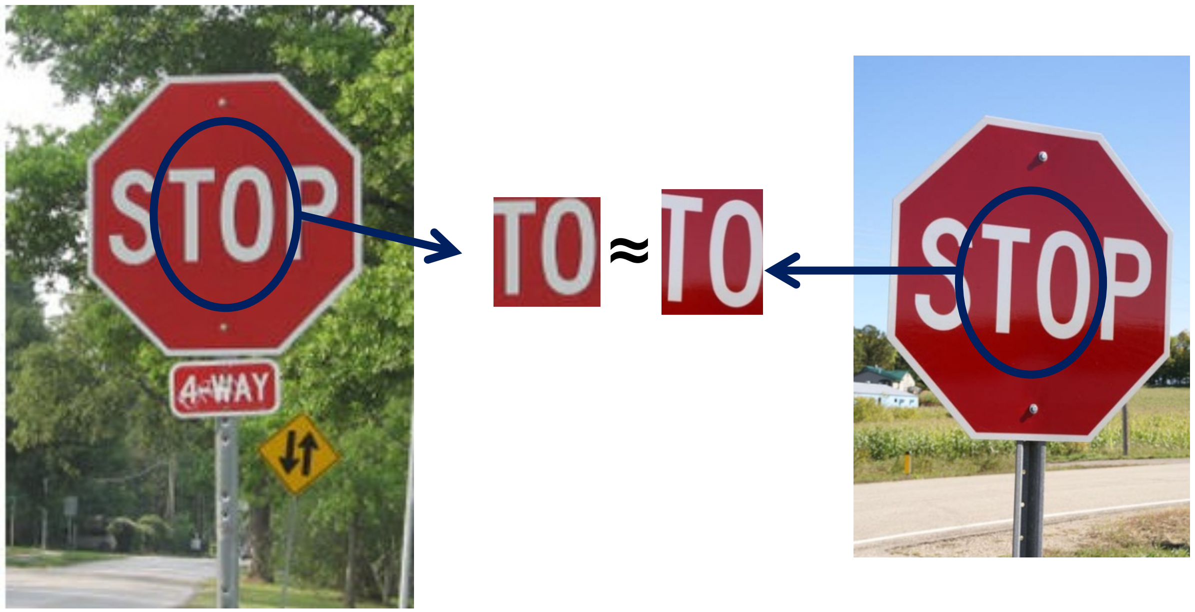
or

-1
0
1

Demo - image filtering

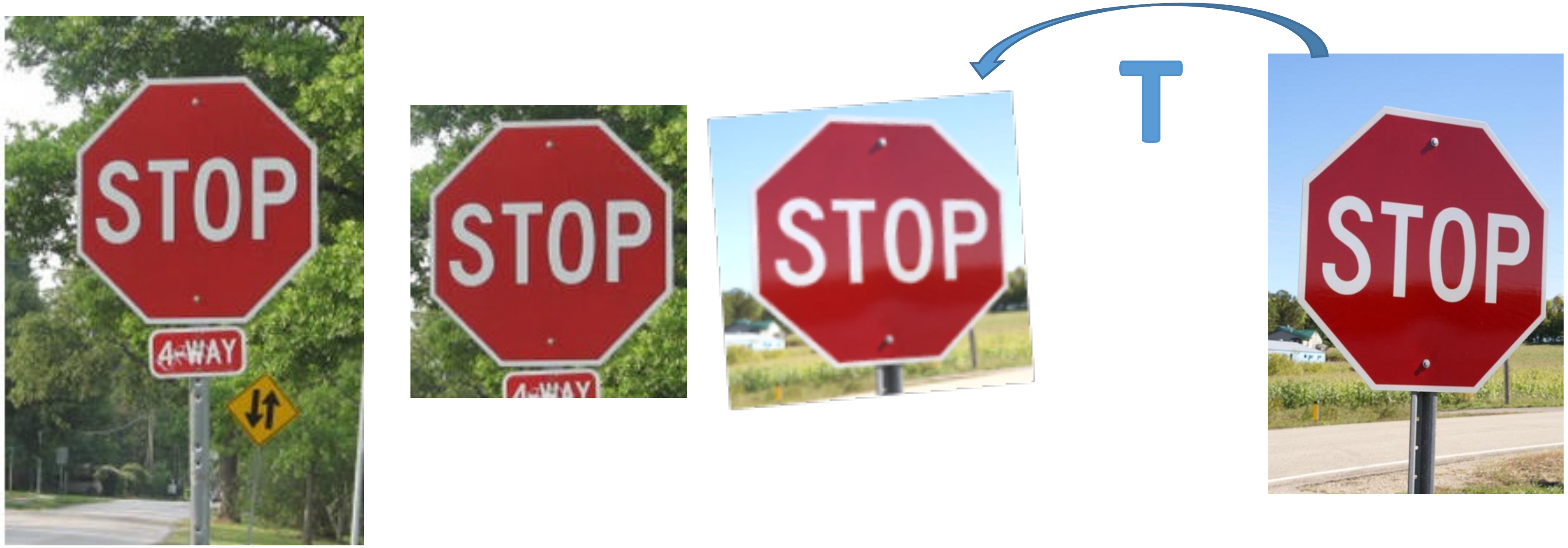
Correspondence and alignment

- Correspondence: matching points, patches, edges, or regions across images

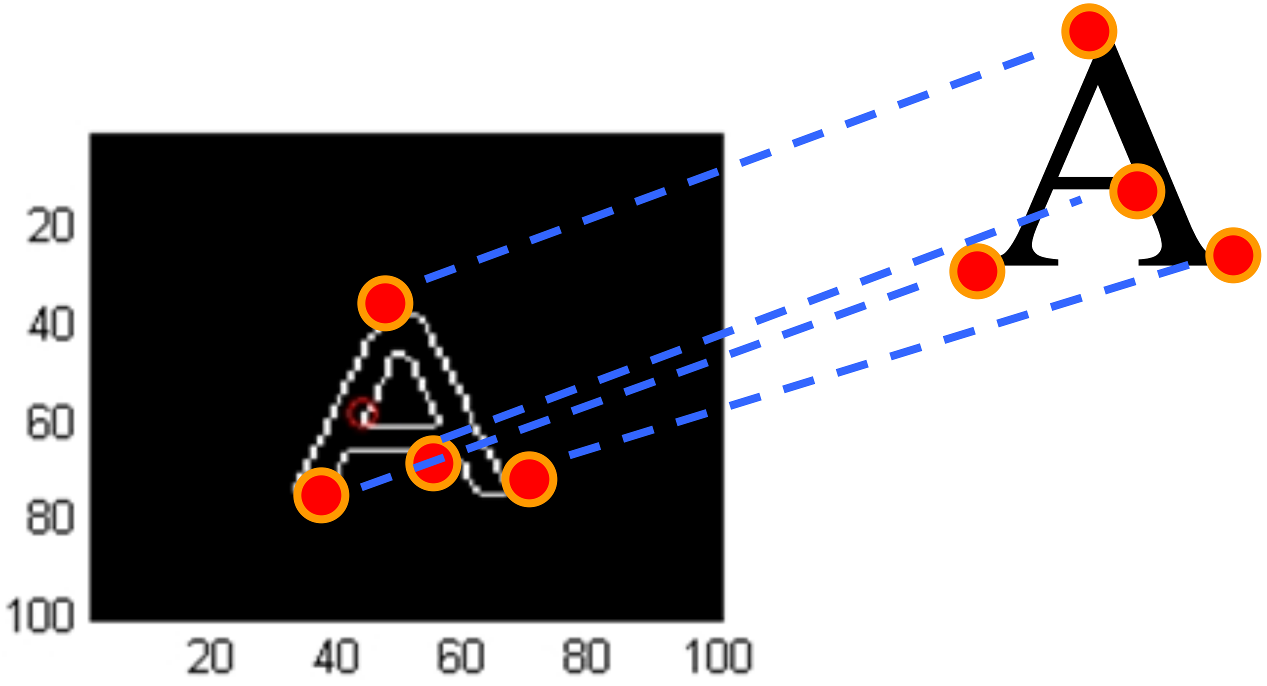


Correspondence and alignment

- Alignment: solving the transformation that makes two things match better

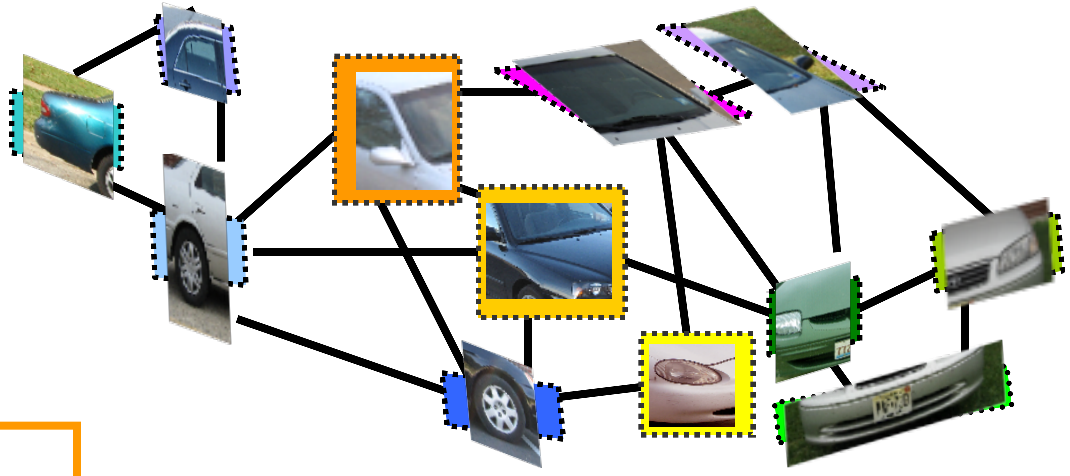


Example: fitting an 2D shape template



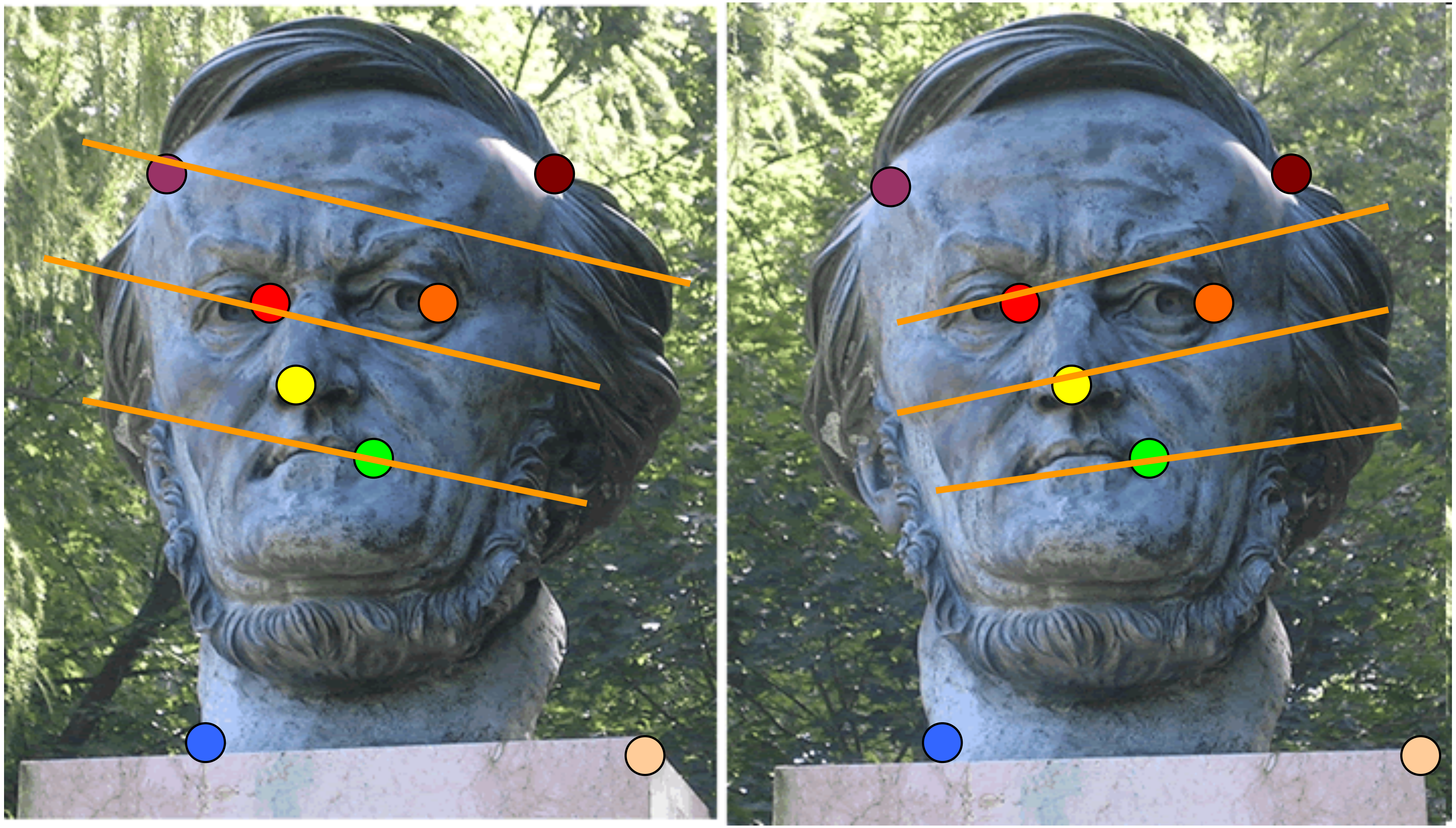
Slide from Silvio Savarese

Example: fitting a 3D object model



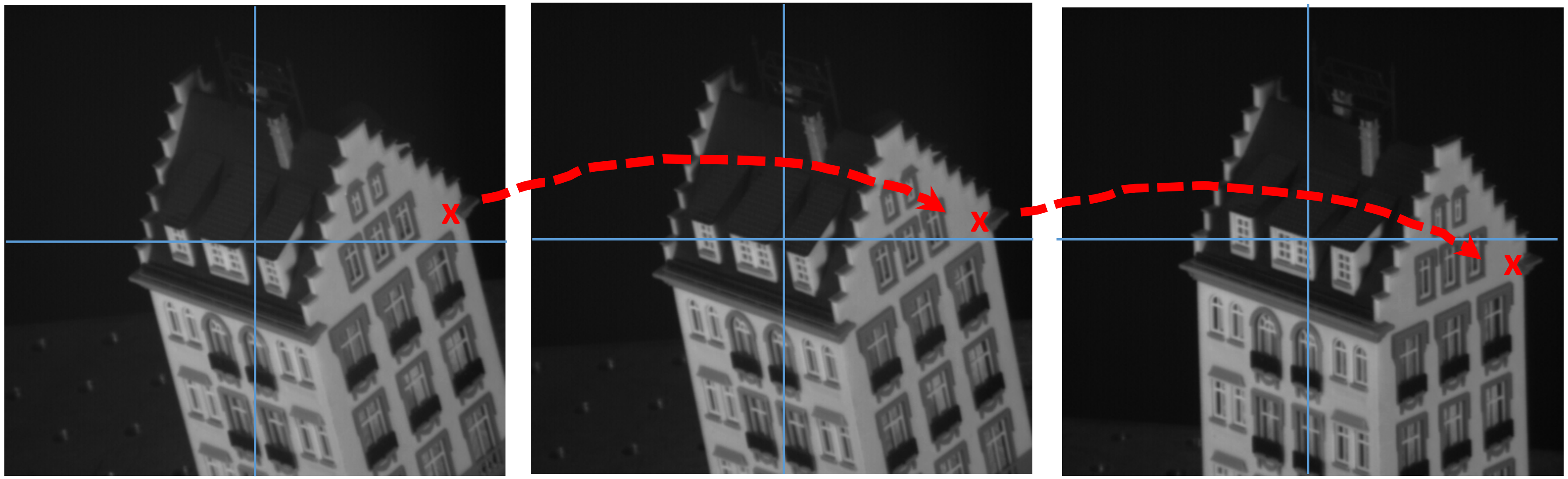
Slide from Silvio Savarese

Example: estimating “fundamental matrix” that corresponds two views



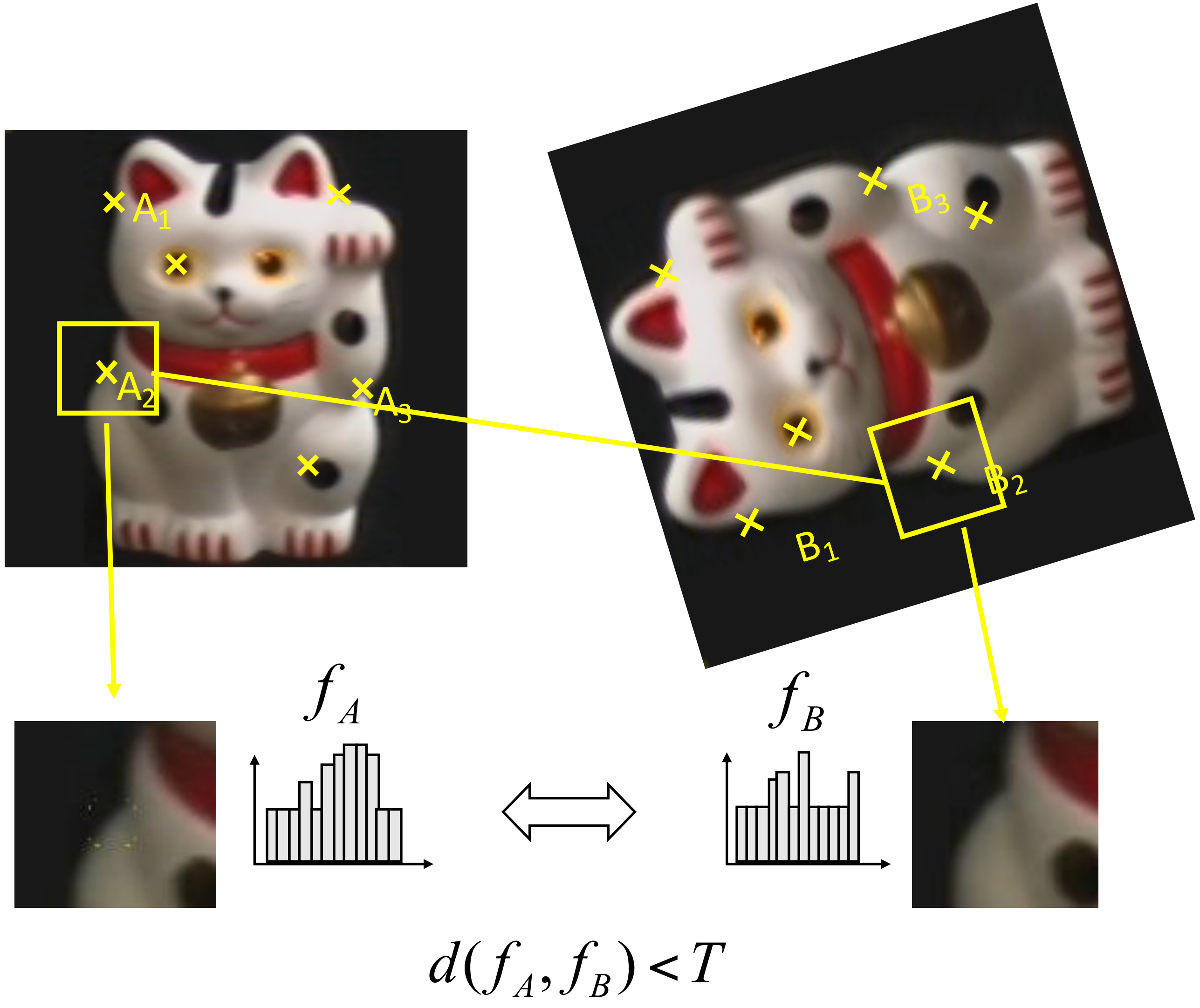
Slide from Silvio Savarese

Example: tracking points



frame 0 frame 22 frame 49

Overview of Keypoint Matching



1. **Find a set of distinctive key-points**

2. **Define a region around each keypoint**

3. **Extract and normalize the region content**

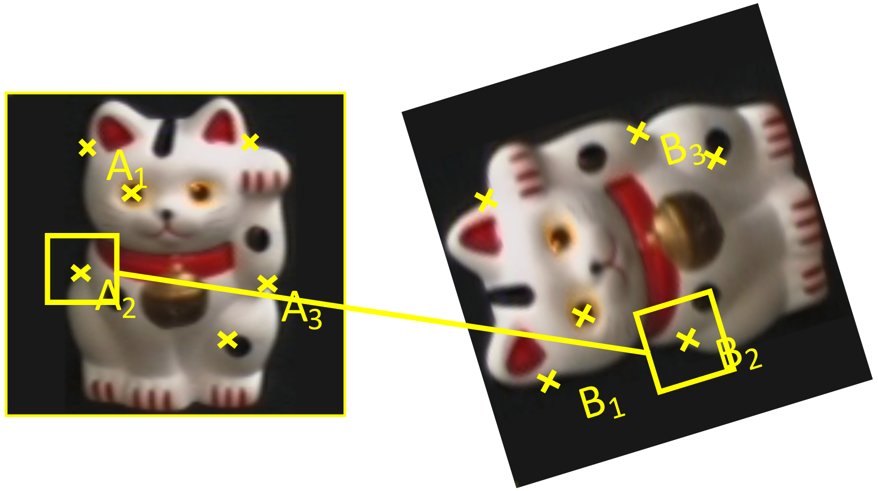
4. **Compute a local descriptor from the normalized region**

5. **Match local descriptors**

Goals for Keypoints

Detect points that are *repeatable* and *distinctive*

Key trade-offs



Detection

More Repeatable

Robust detection
Precise localization

More Points

Robust to occlusion
Works with less texture

Description

More Distinctive

Minimize wrong matches

More Flexible

Robust to expected variations
Maximize correct matches

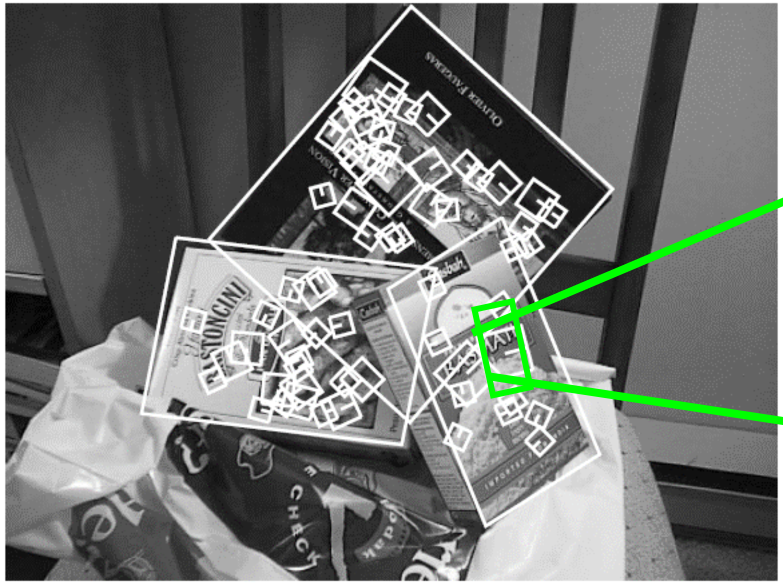
Choosing interest points

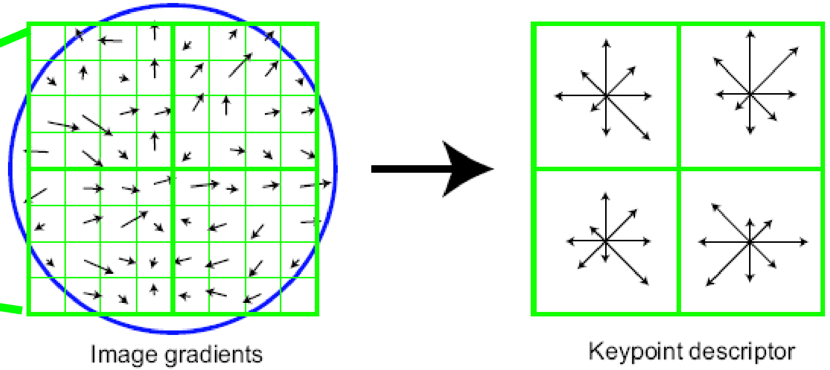
Where would you tell your friend to meet you?

Choosing interest points

Where would you tell your friend to meet you?

Local Descriptors: SIFT Descriptor



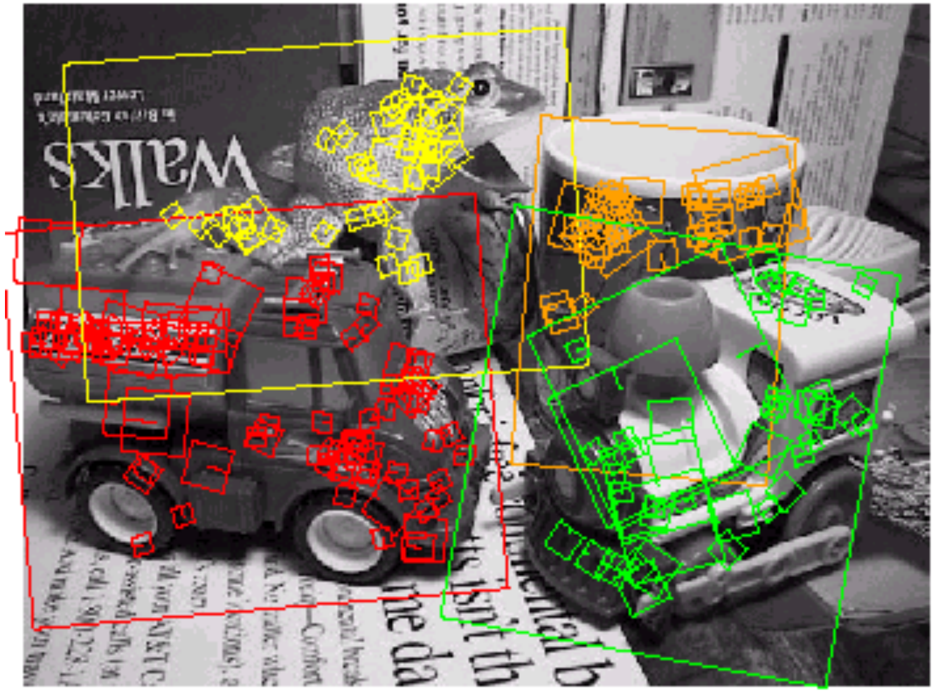


Histogram of oriented gradients

- **Captures important texture information**
- **Robust to small translations / affine deformations**

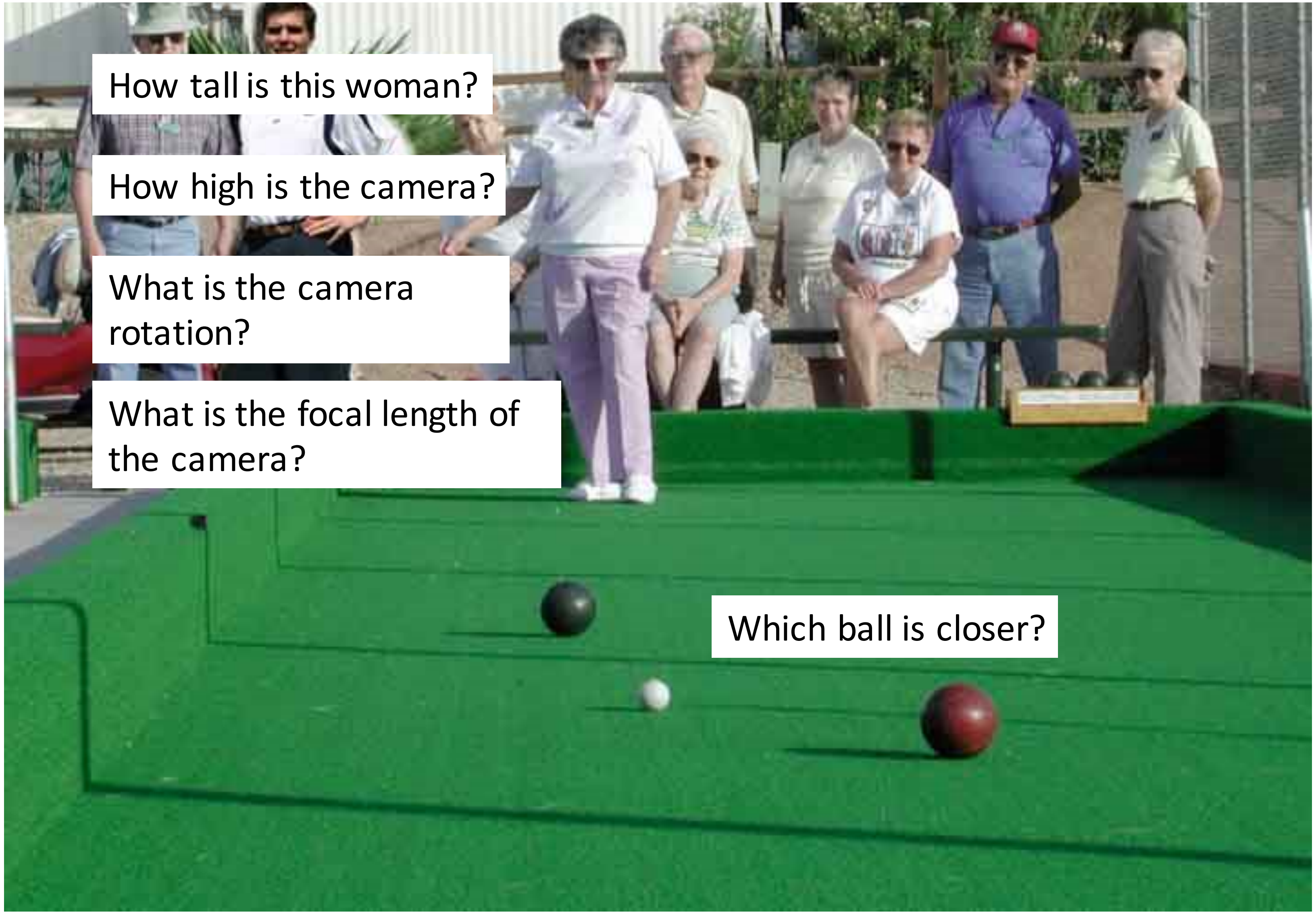
[Lowe, ICCV 1999]

Examples of recognized objects



Demo – interest point detection

Single-view Geometry

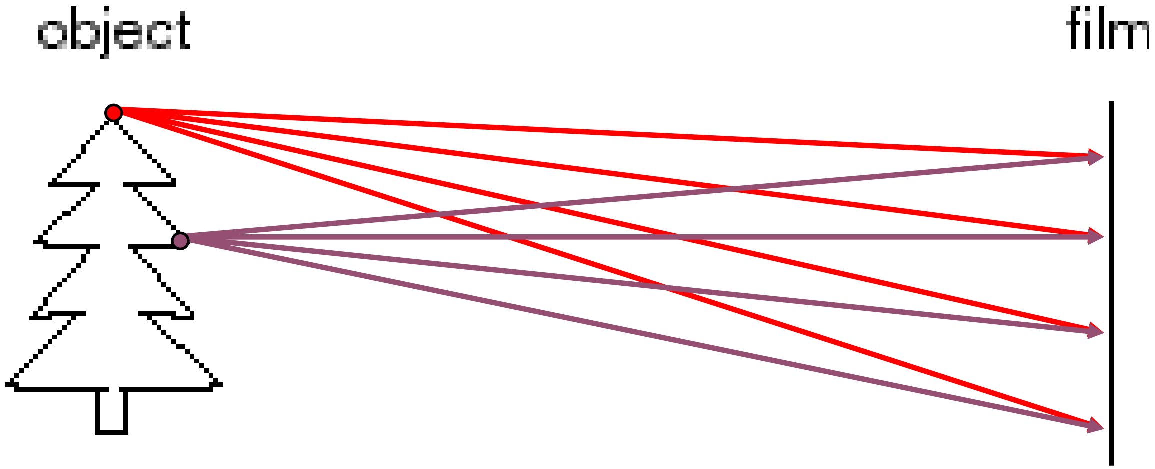


Single-view Geometry

Mapping between **image** and **world** coordinates

- Pinhole camera model
- Projective geometry
  - Vanishing points and lines
- Projection matrix

Image formation

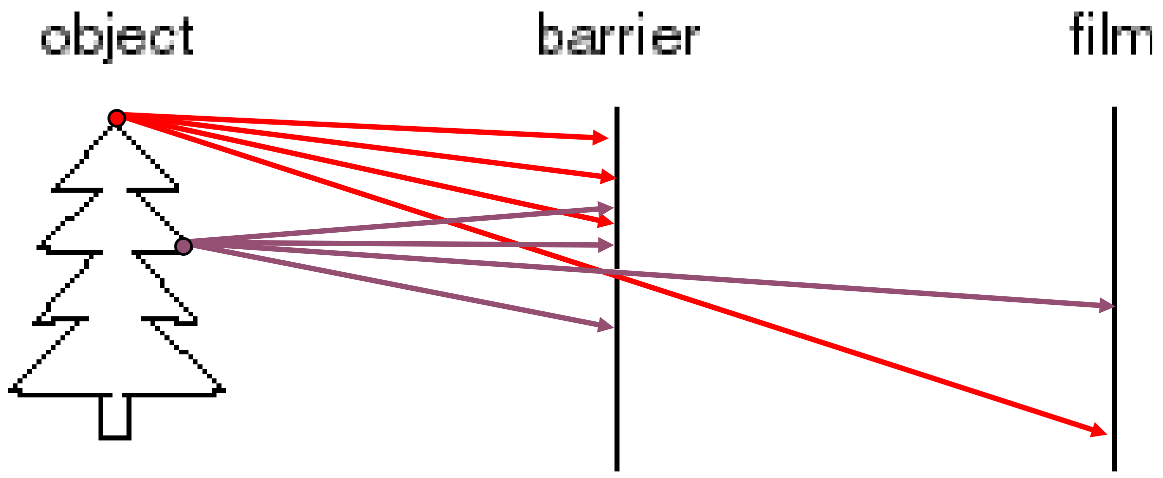


Let’s design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

Slide credit: Steve Seitz

Pinhole camera

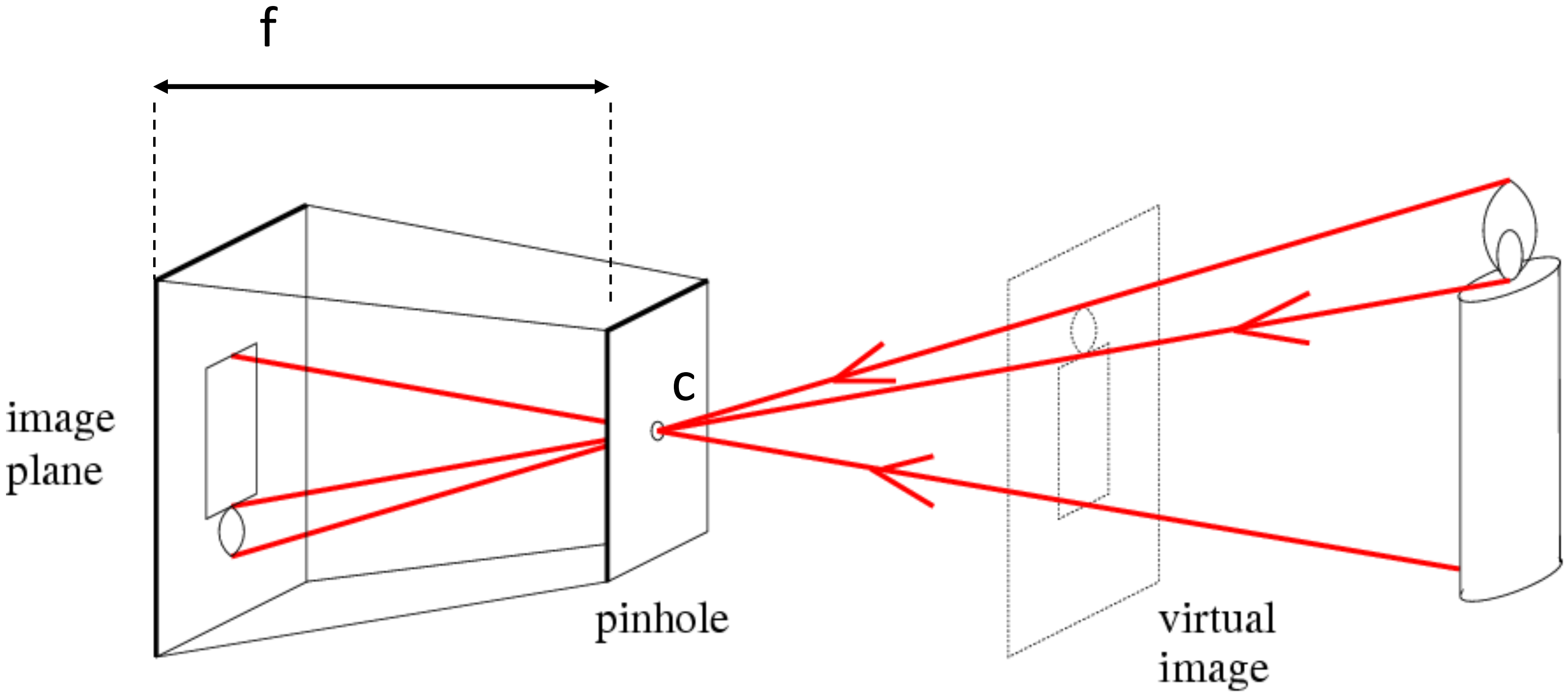


Idea 2: add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**

Slide credit: Steve Seitz

Pinhole camera



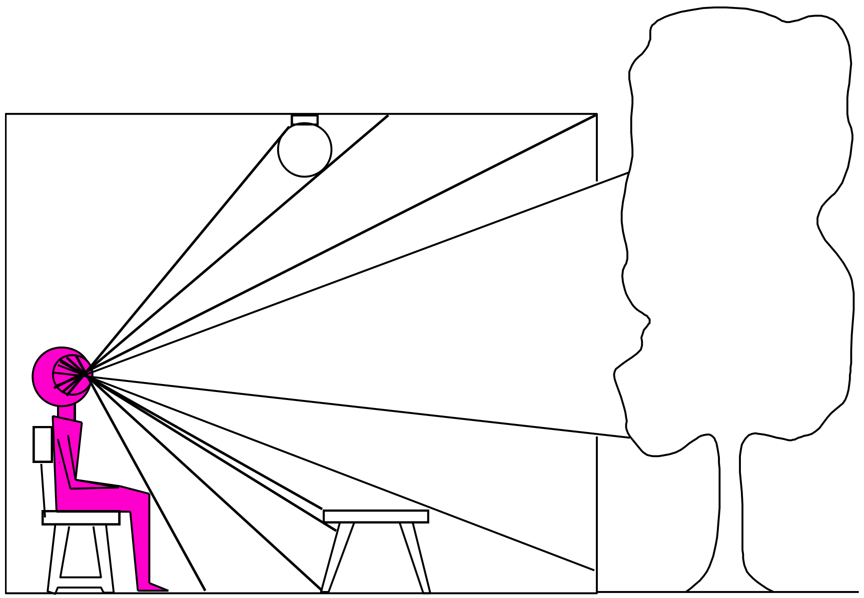
f = focal length

c = center of the camera

Figure from David Forsyth

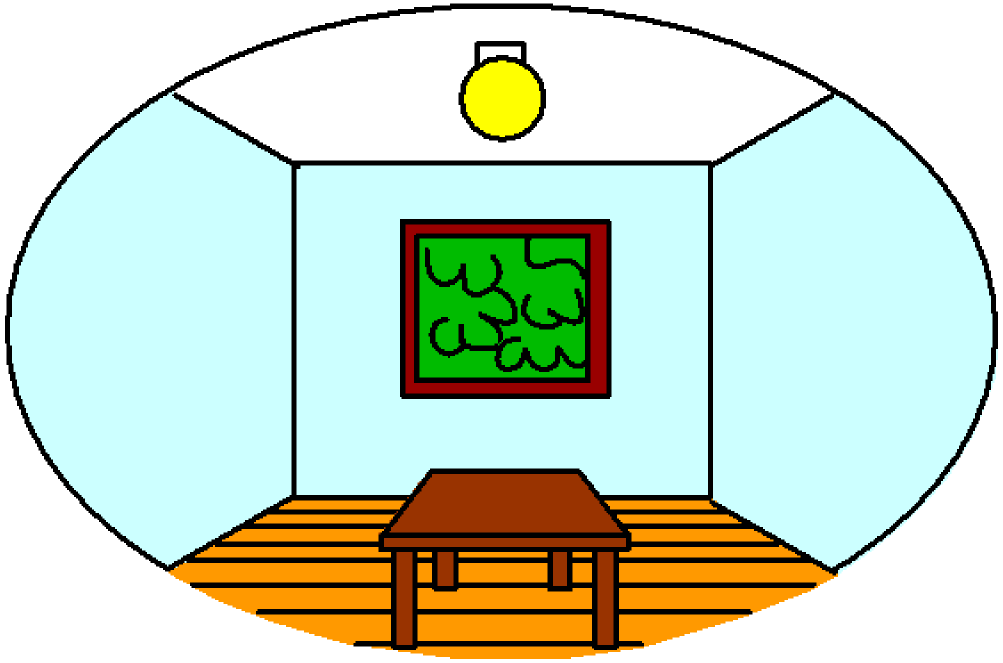
Dimensionality Reduction Machine (3D to 2D)

3D world



Point of observation

2D image



Slide source: Seitz

Projection can be tricky…

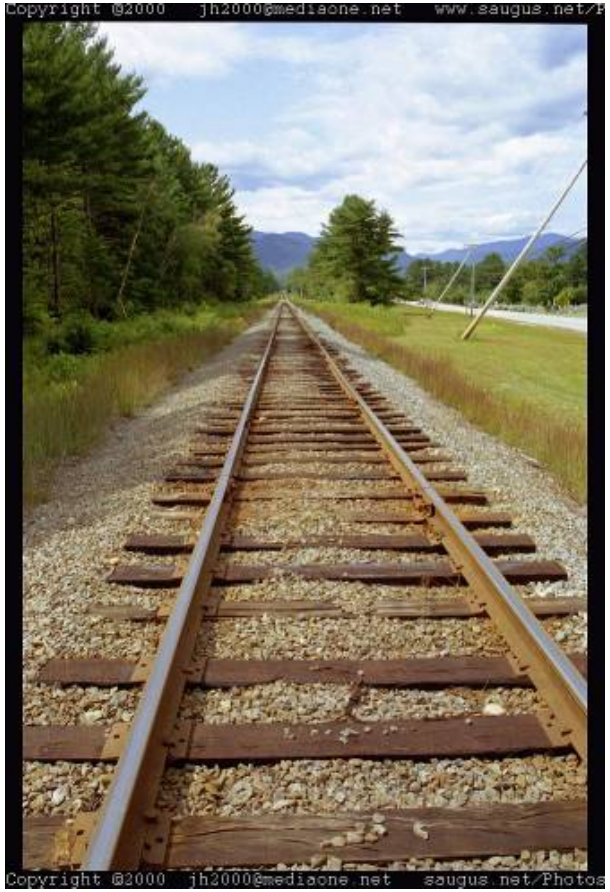
Slide source: Seitz

Projection can be tricky...

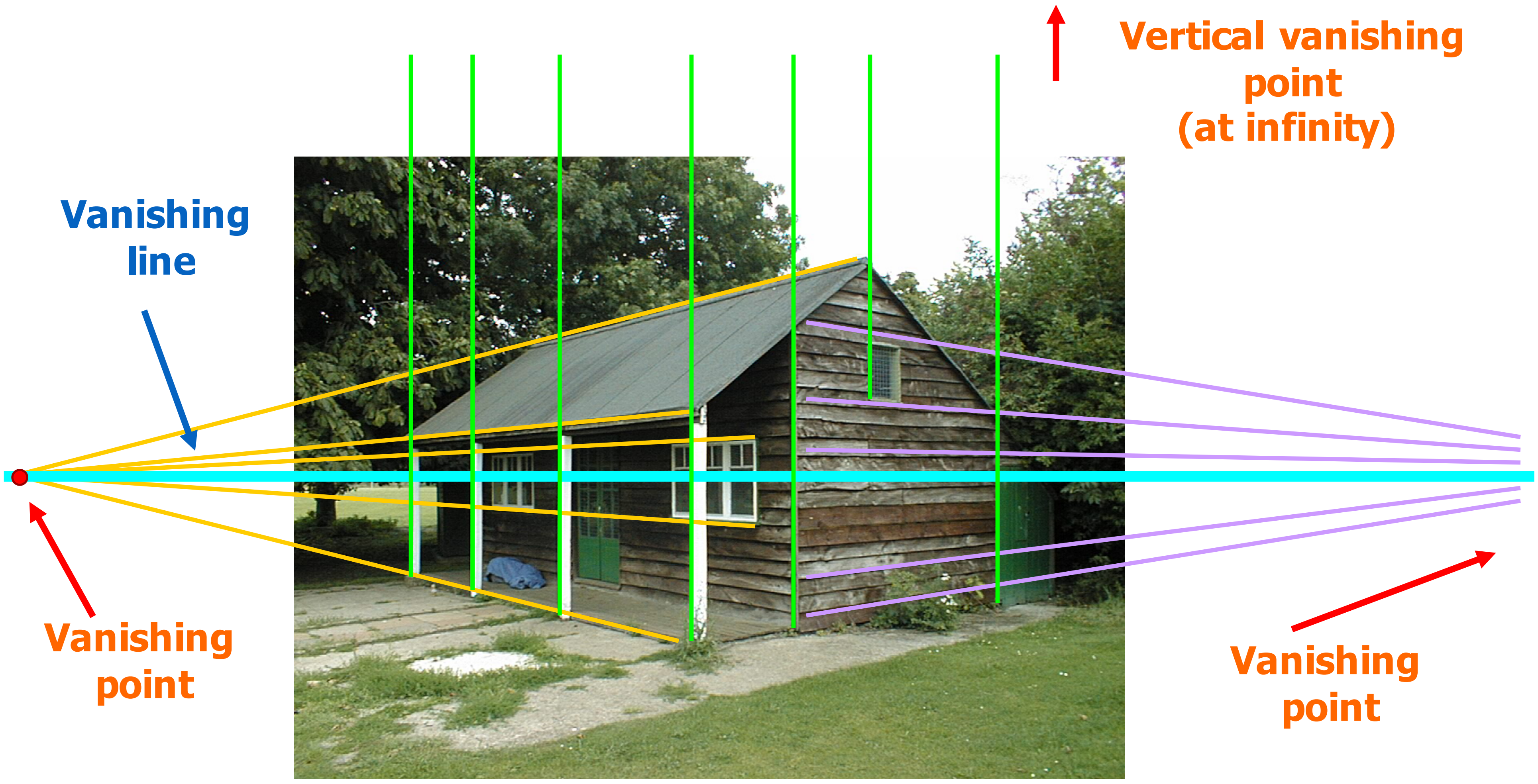
Making of 3D sidewalk art: http://www.youtube.com/watch?v=3SNYtd0Ayt0

Vanishing points and lines

Parallel lines in the world intersect in the image at a “vanishing point”

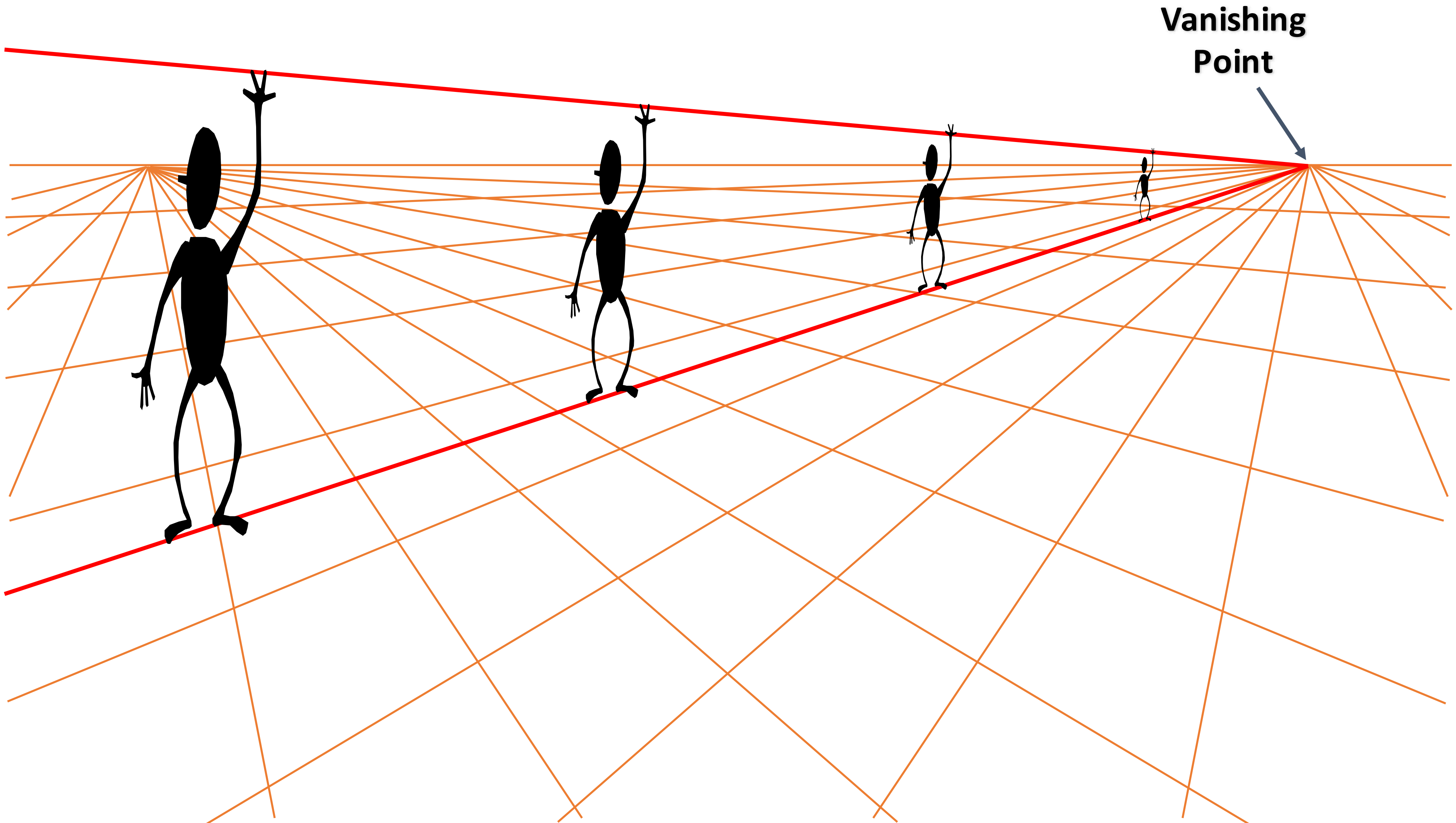


Vanishing points and lines

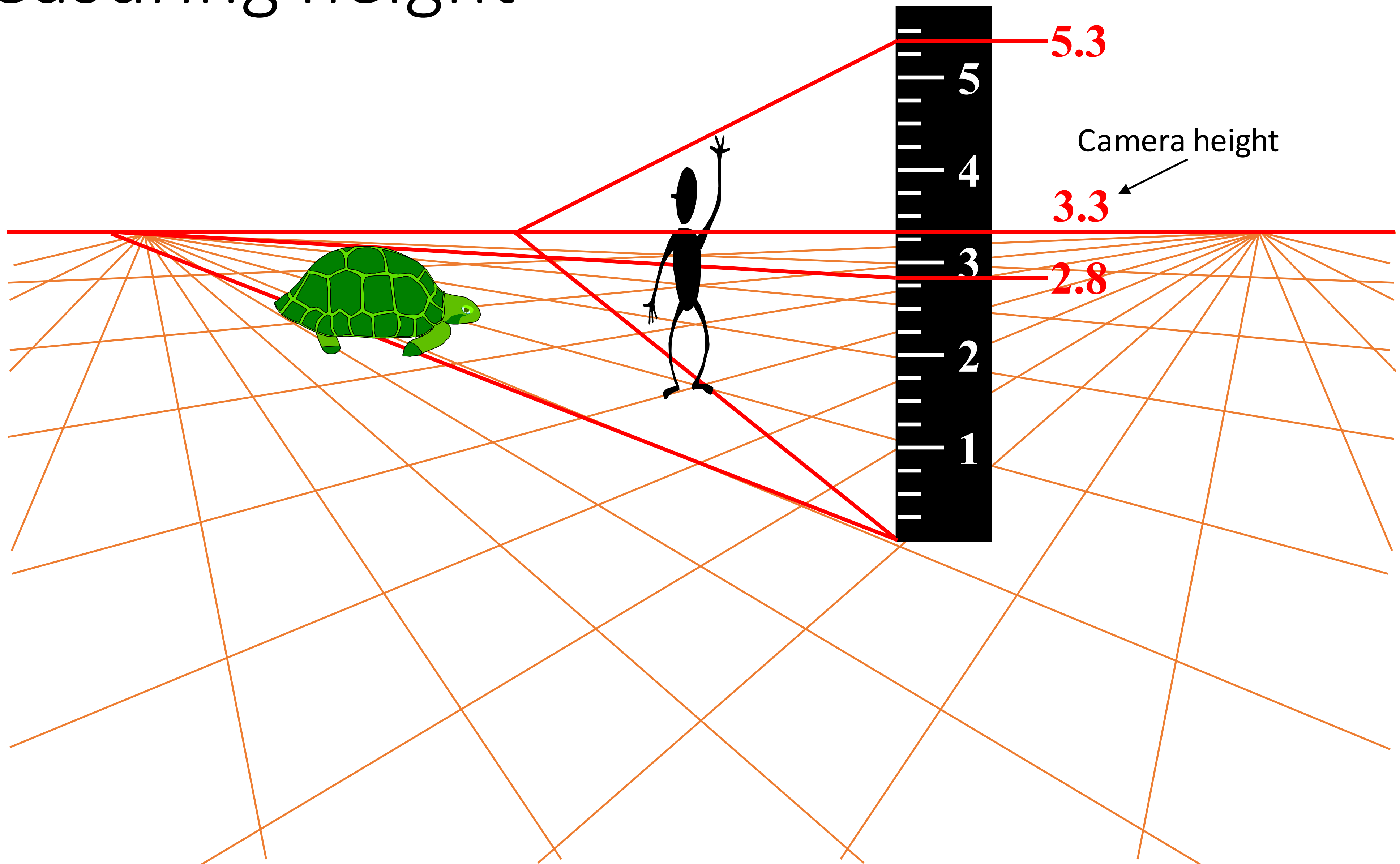


Slide from Efros, Photo from Criminisi

Comparing heights



Measuring height



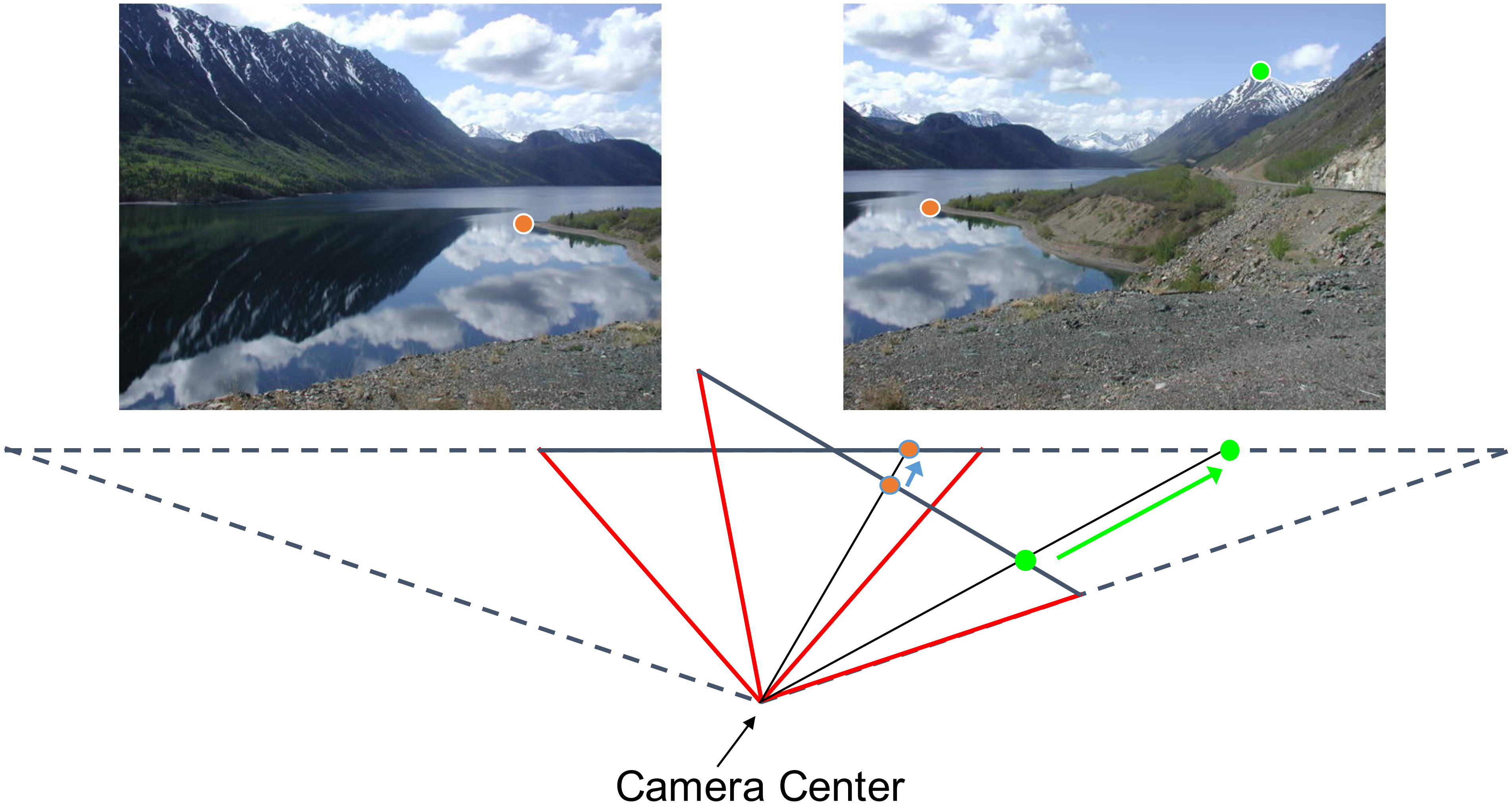

ETTV
東森新聞
范佐憲
CH10
6/27/2013 10:5

Image Stitching

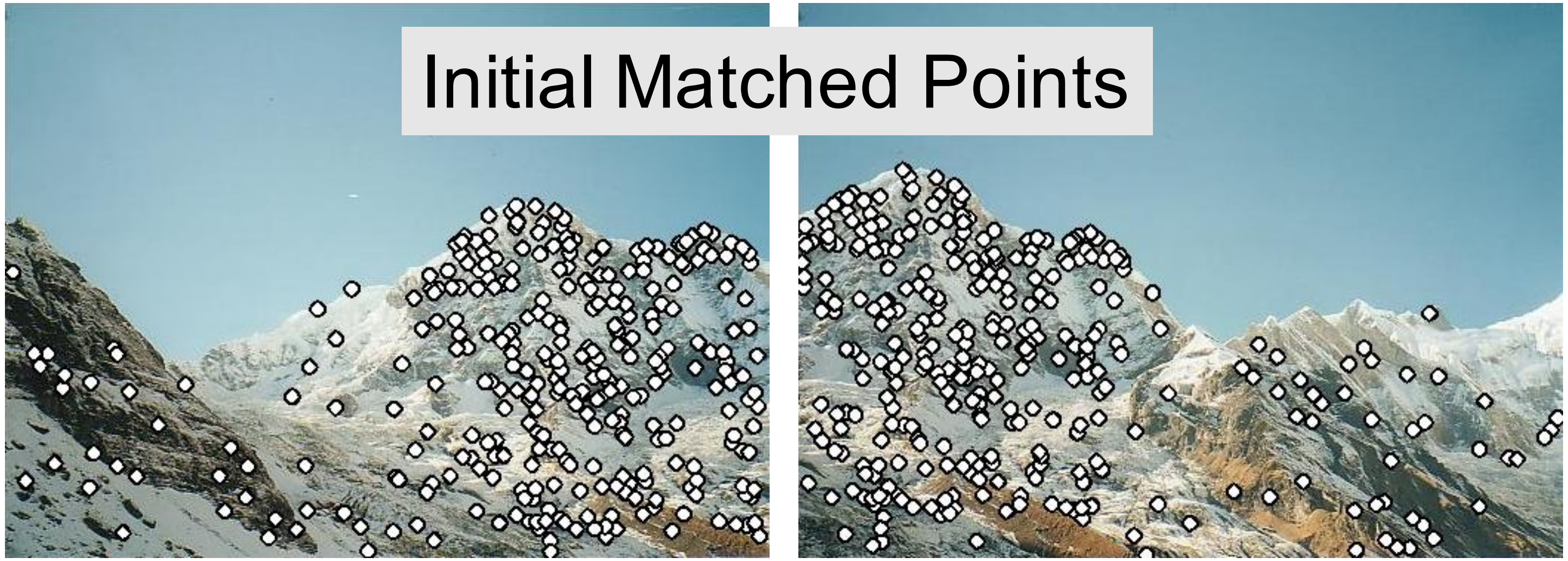
- Combine two or more overlapping images to make one larger image

Slide credit: [Vaibhav Vaish]

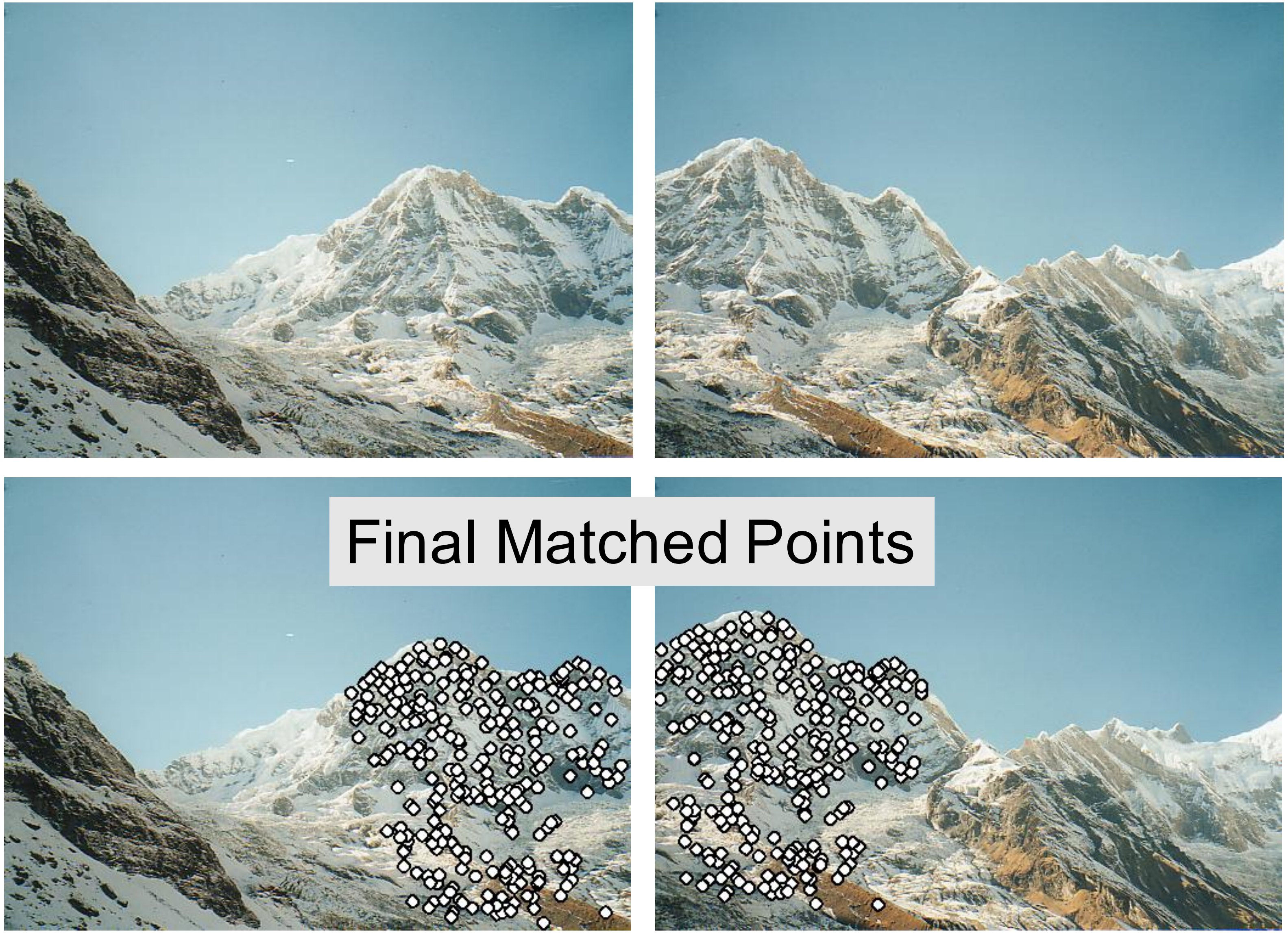
Illustration



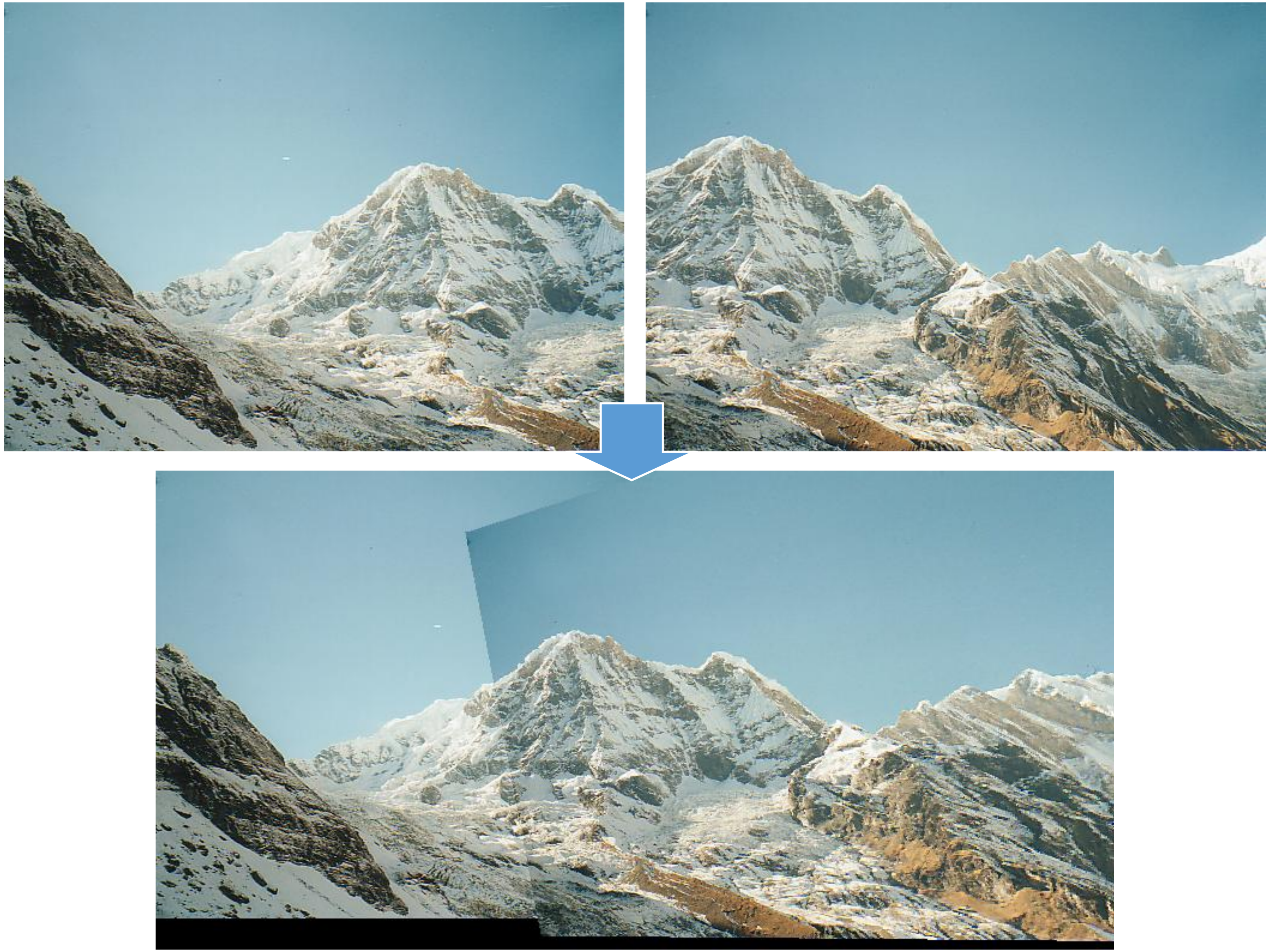
RANSAC for Homography



RANSAC for Homography



RANSAC for Homography



Bundle Adjustment

- New images initialised with rotation, focal length of best matching image

Bundle Adjustment

- New images initialised with rotation, focal length of best matching image

Details to make it look good

- Choosing seams
- Blending

Choosing seams

- Better method: dynamic program to find seam along well-matched regions

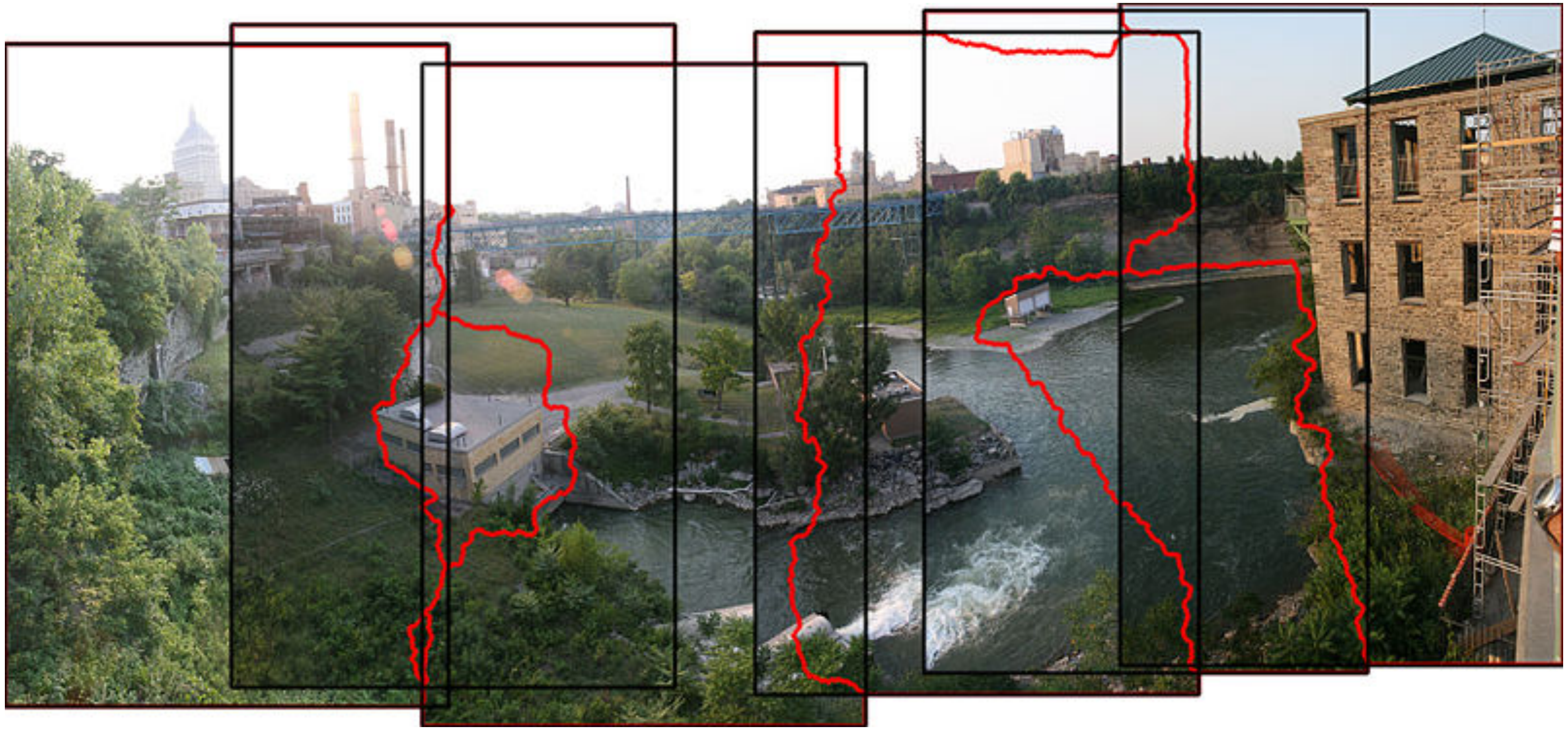


Illustration: http://en.wikipedia.org/wiki/File:Rochester_NY.jpg

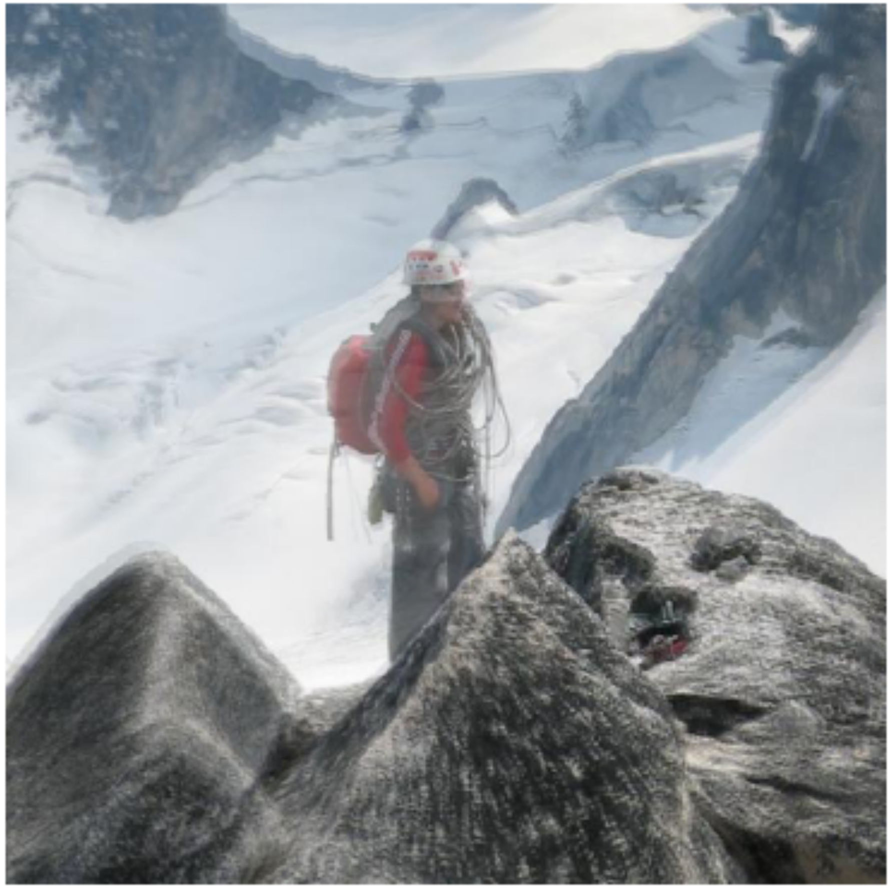
Gain compensation

- Simple gain adjustment
  - Compute average RGB intensity of each image in overlapping region
  - Normalize intensities by ratio of averages

Multi-band Blending

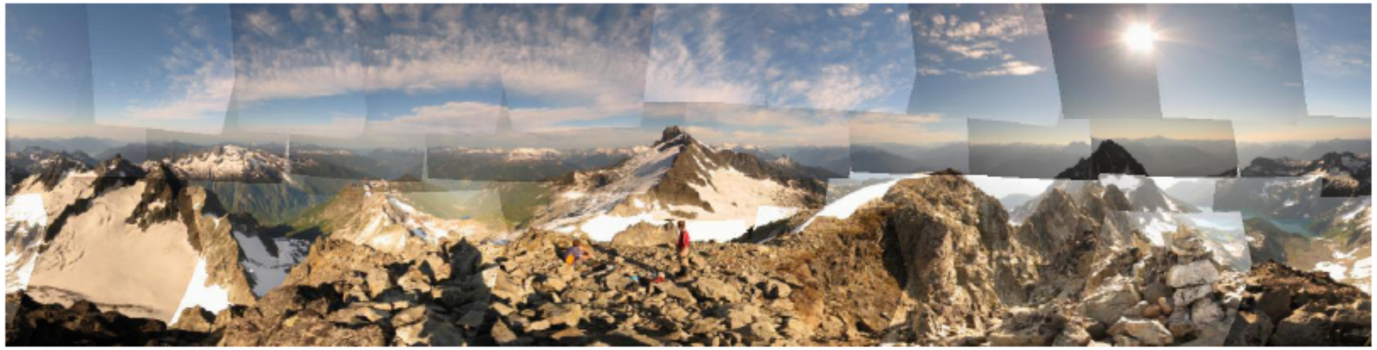
- Burt & Adelson 1983
  - Blend frequency bands over range $\propto \lambda$

Blending comparison (IJCV 2007)

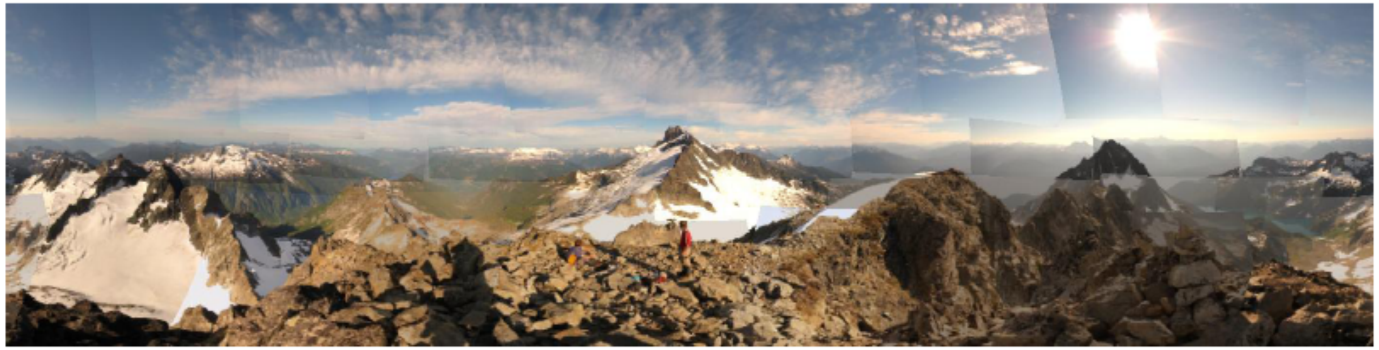


(a) Linear blending

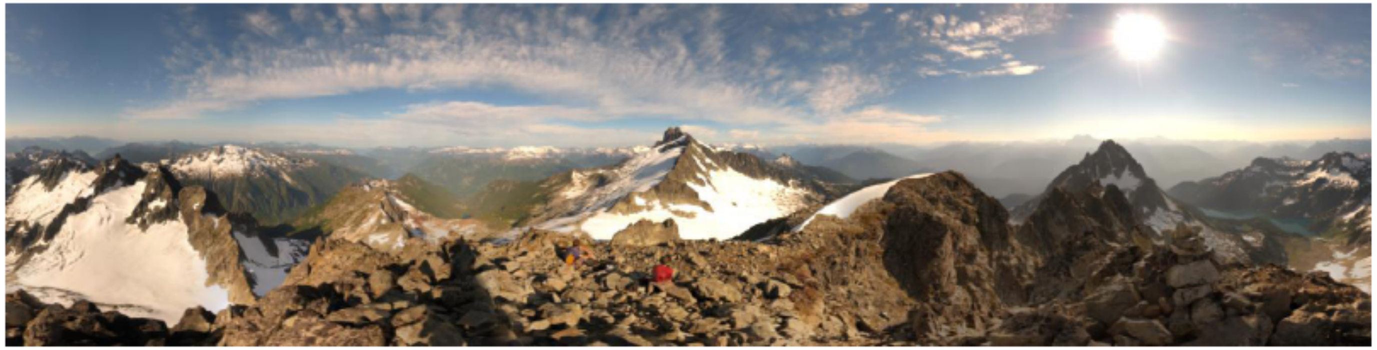
(b) Multi-band blending



(b) Without gain compensation



(c) With gain compensation

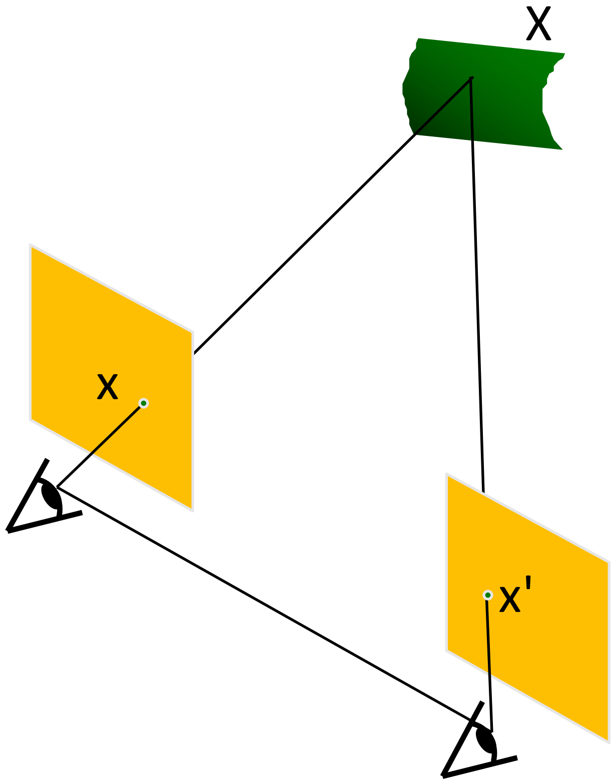


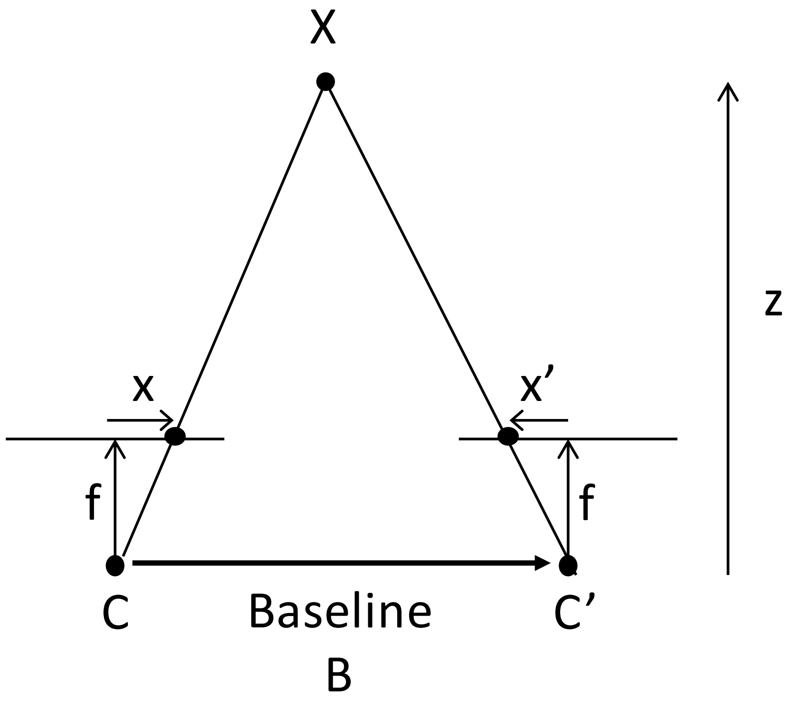
(d) With gain compensation and multi-band blending

Demo – feature matching and image stitching

Depth from Stereo

- Goal: recover depth by finding image coordinate x’ that corresponds to x



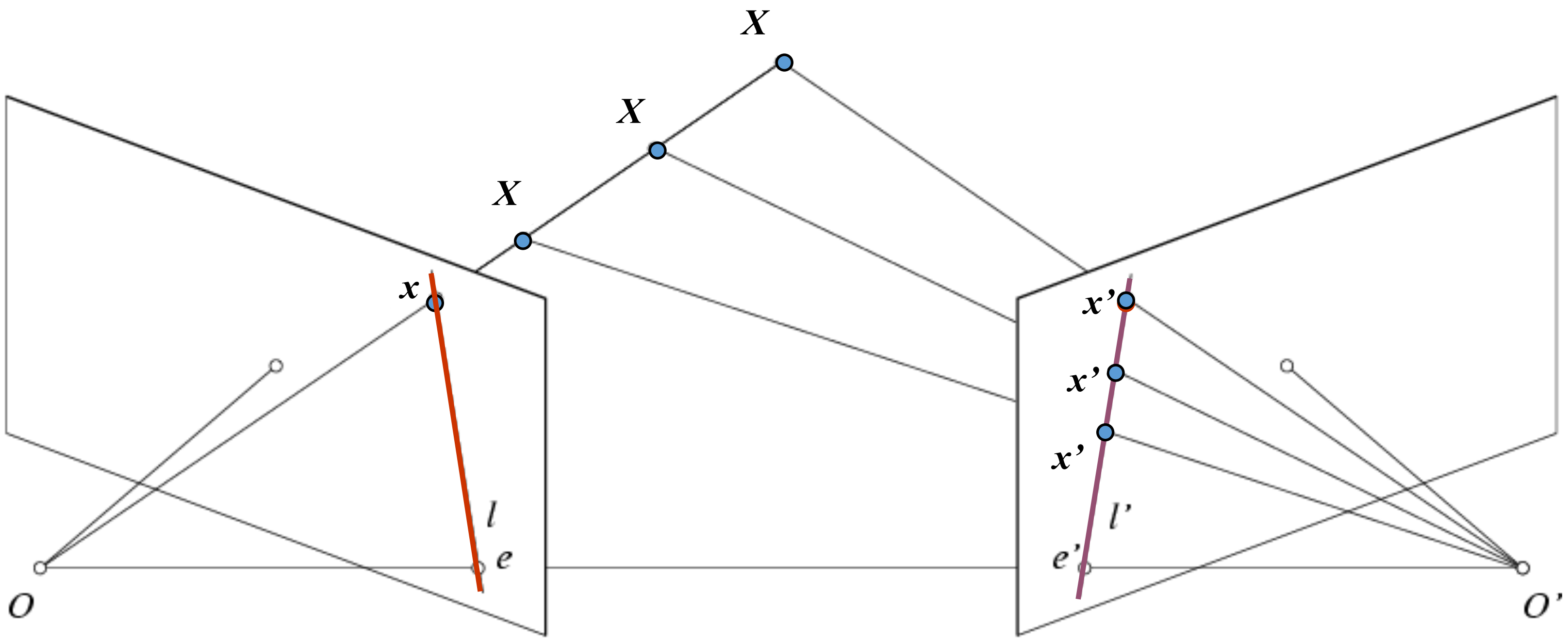


Correspondence Problem

- We have two images taken from cameras with different intrinsic and extrinsic parameters

- How do we match a point in the first image to a point in the second? How can we constrain our search?

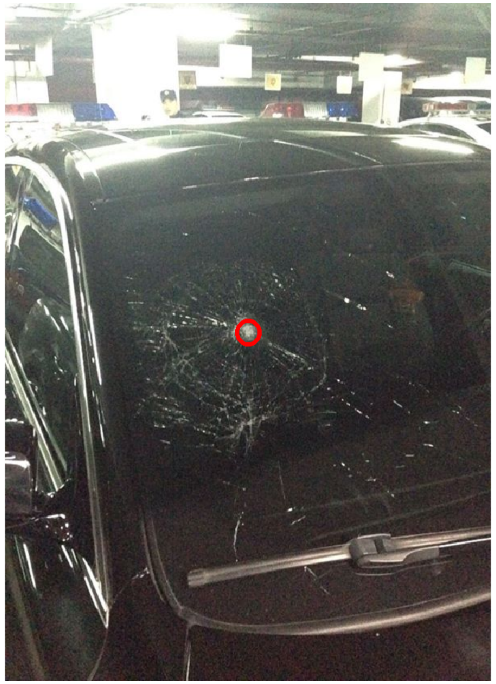
Key idea: Epipolar constraint



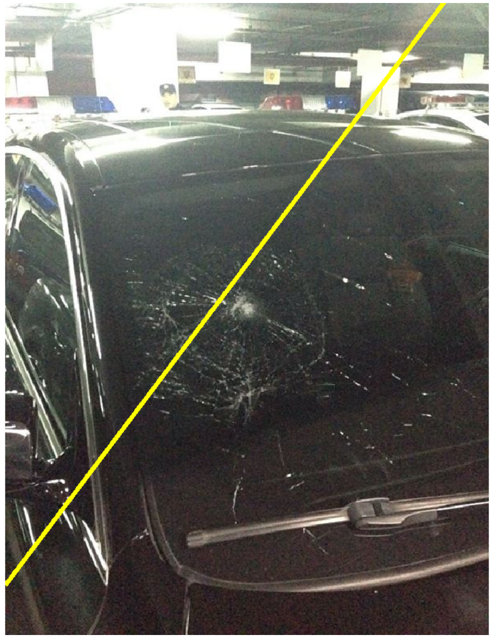
Potential matches for *x* have to lie on the corresponding line *l'*.

Potential matches for *x'* have to lie on the corresponding line *l*.


中時電子報
chinatimes.com

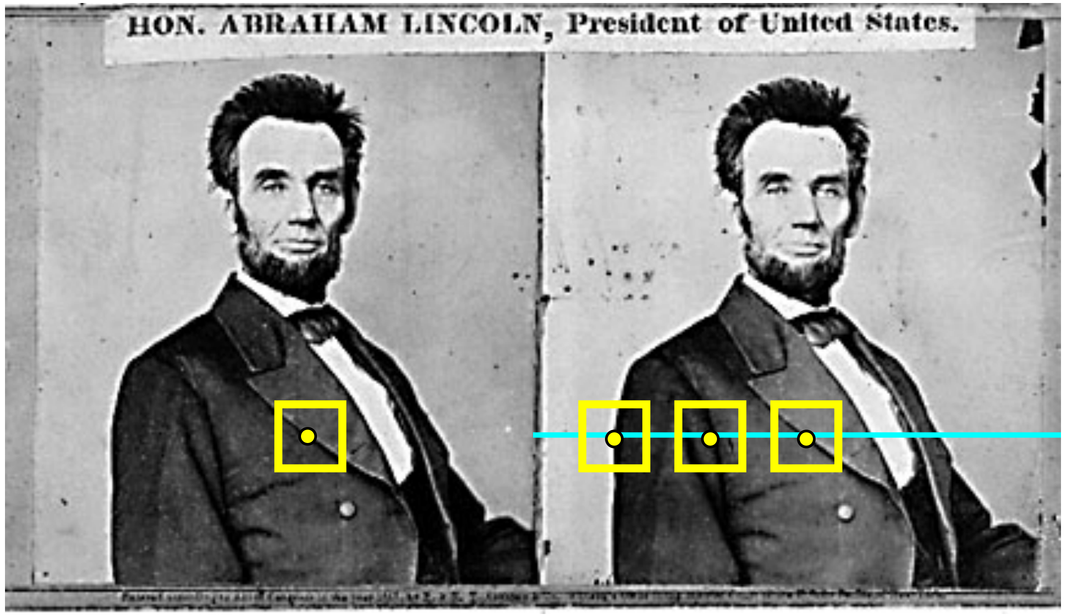



中時電子報
chinatimes.com



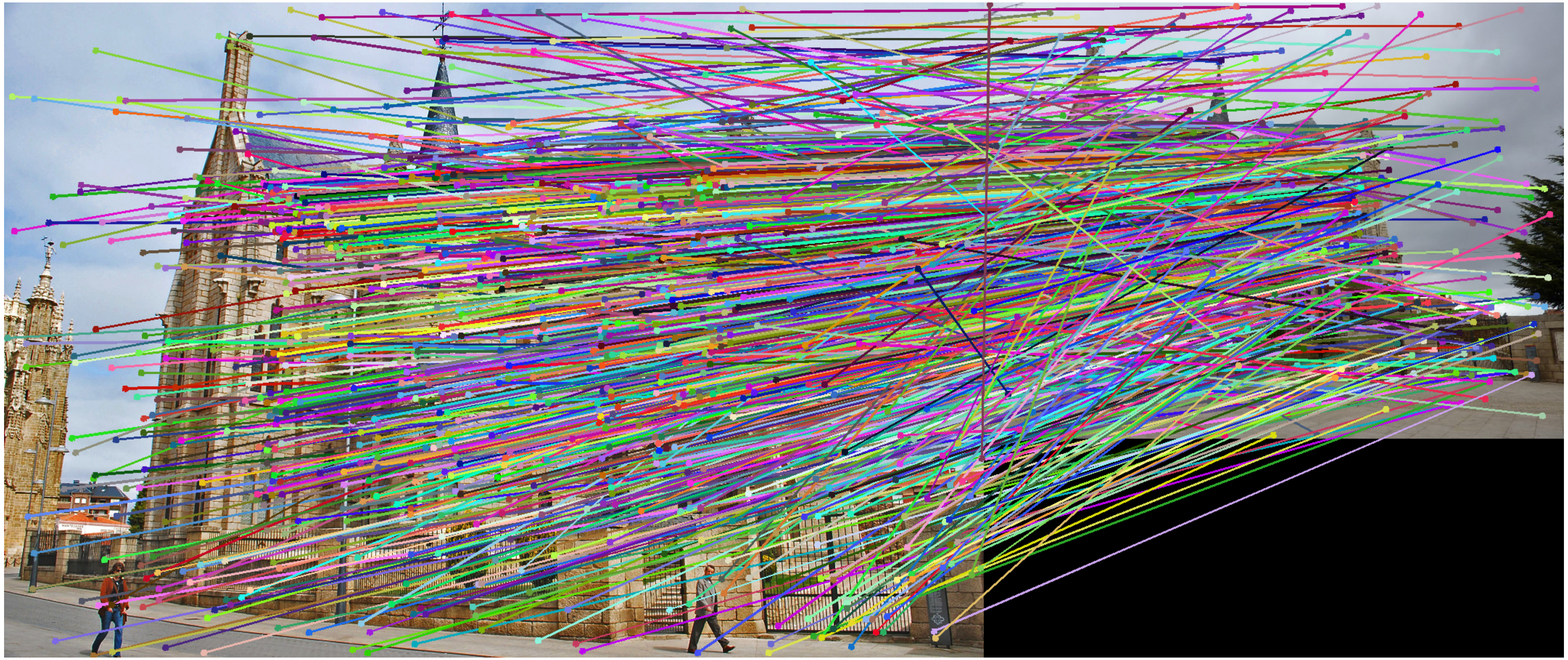

中時電子報
chinatimes.com

Basic stereo matching algorithm

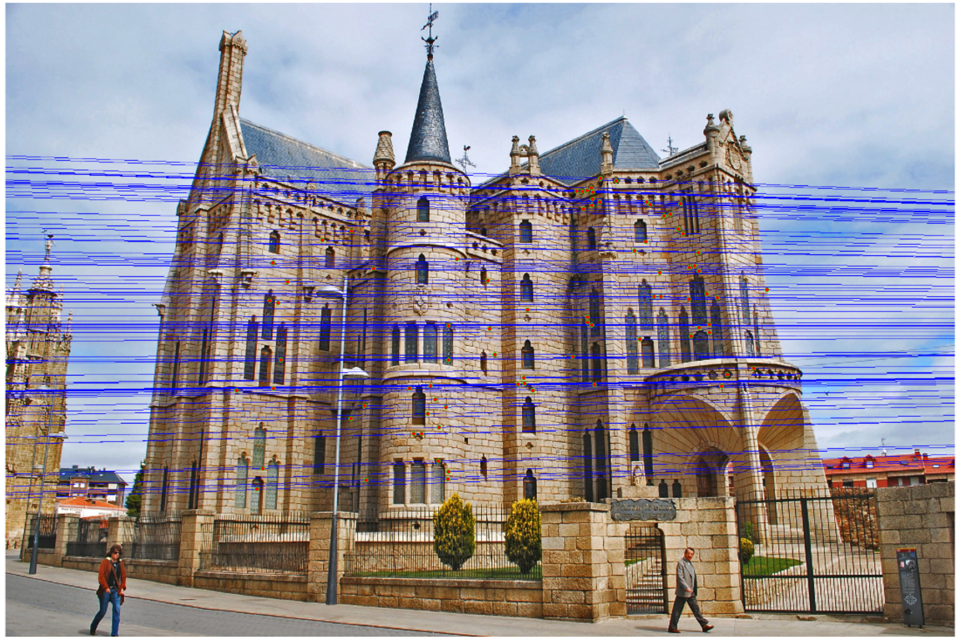


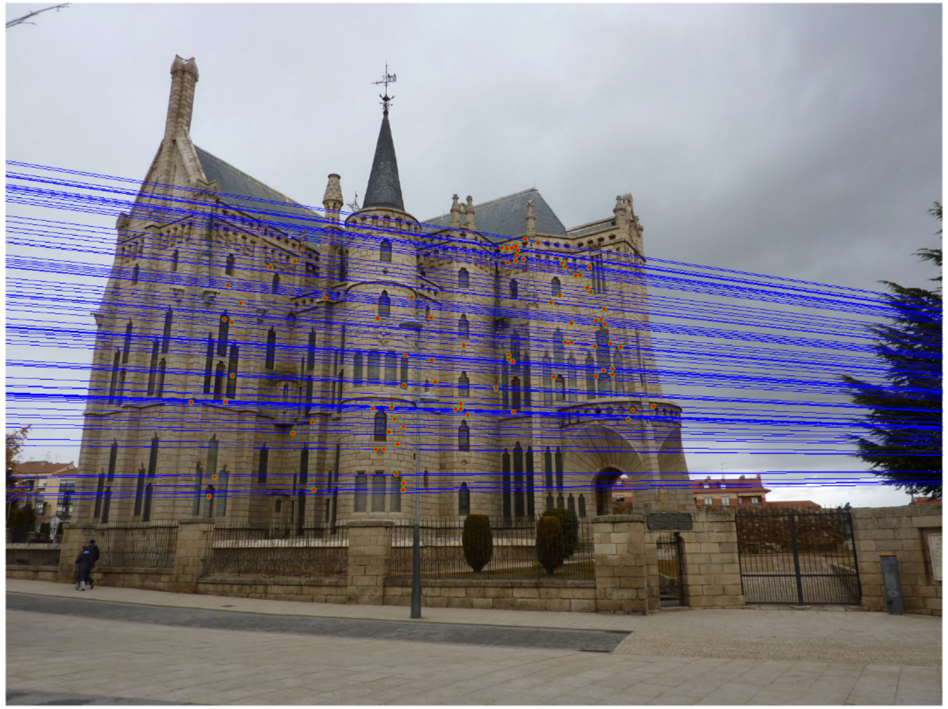
- If necessary, rectify the two stereo images to transform epipolar lines into scanlines
- For each pixel x in the first image
  - Find corresponding epipolar scanline in the right image
  - Search the scanline and pick the best match x'
  - Compute disparity x-x' and set depth(x) = fB/(x-x')

800 most confident matches among 10,000+ local features.

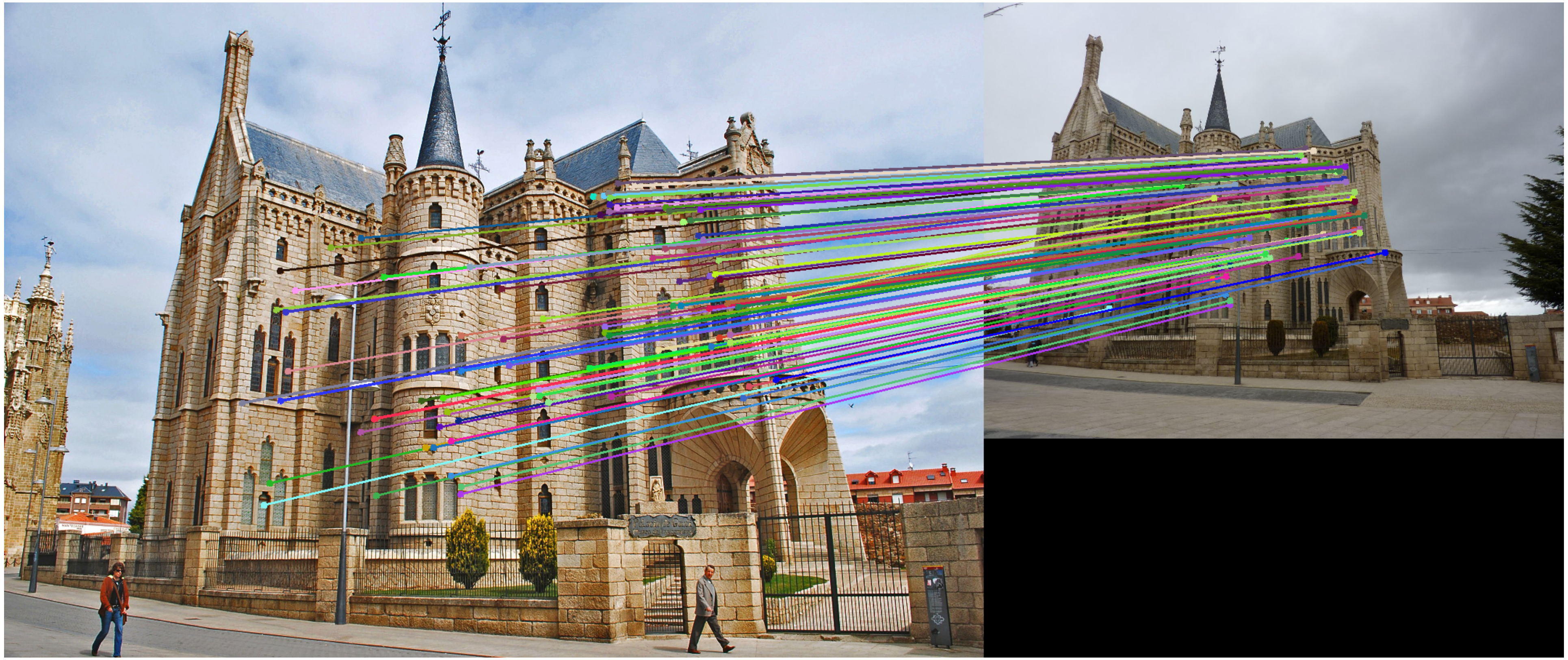


Epipolar lines

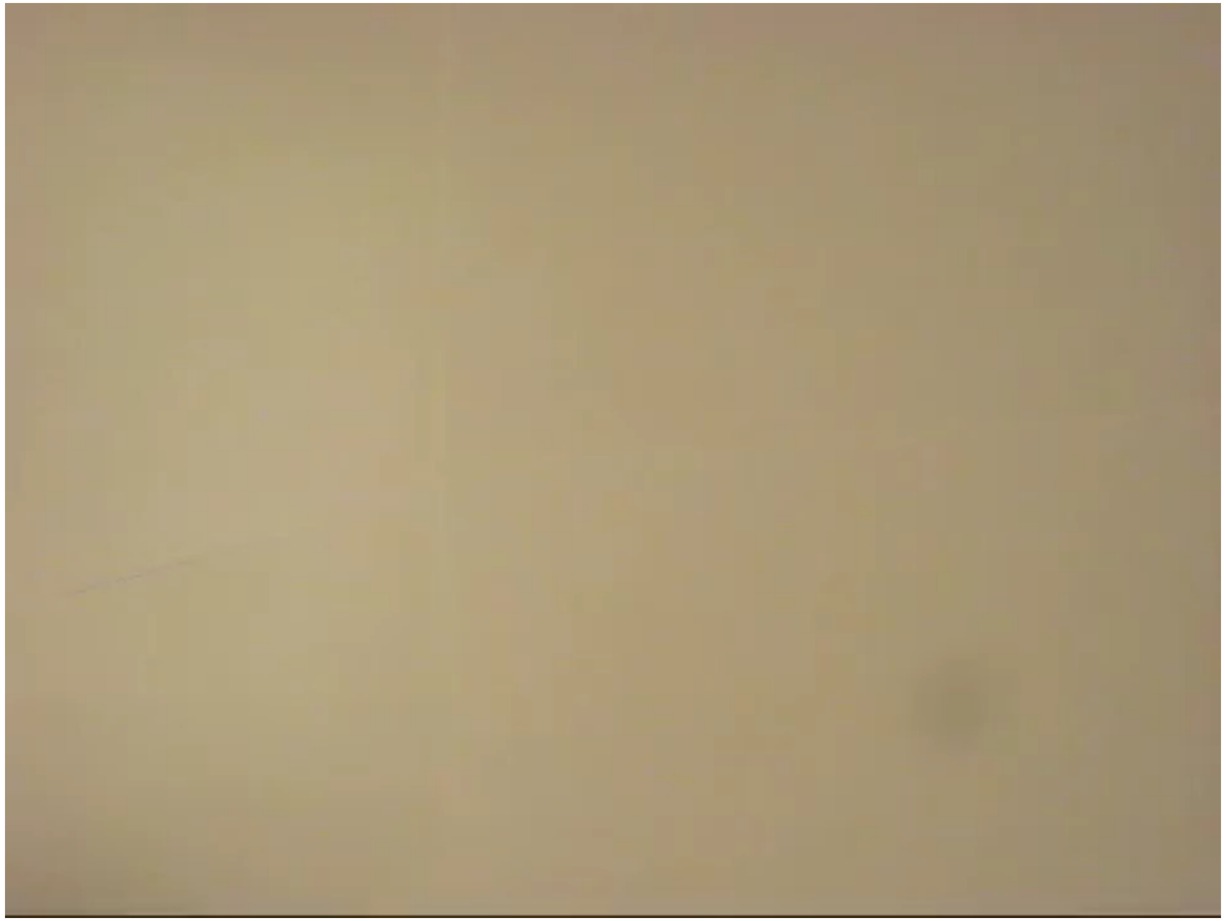




Keep only the matches at are “inliers” with respect to the “best” fundamental matrix

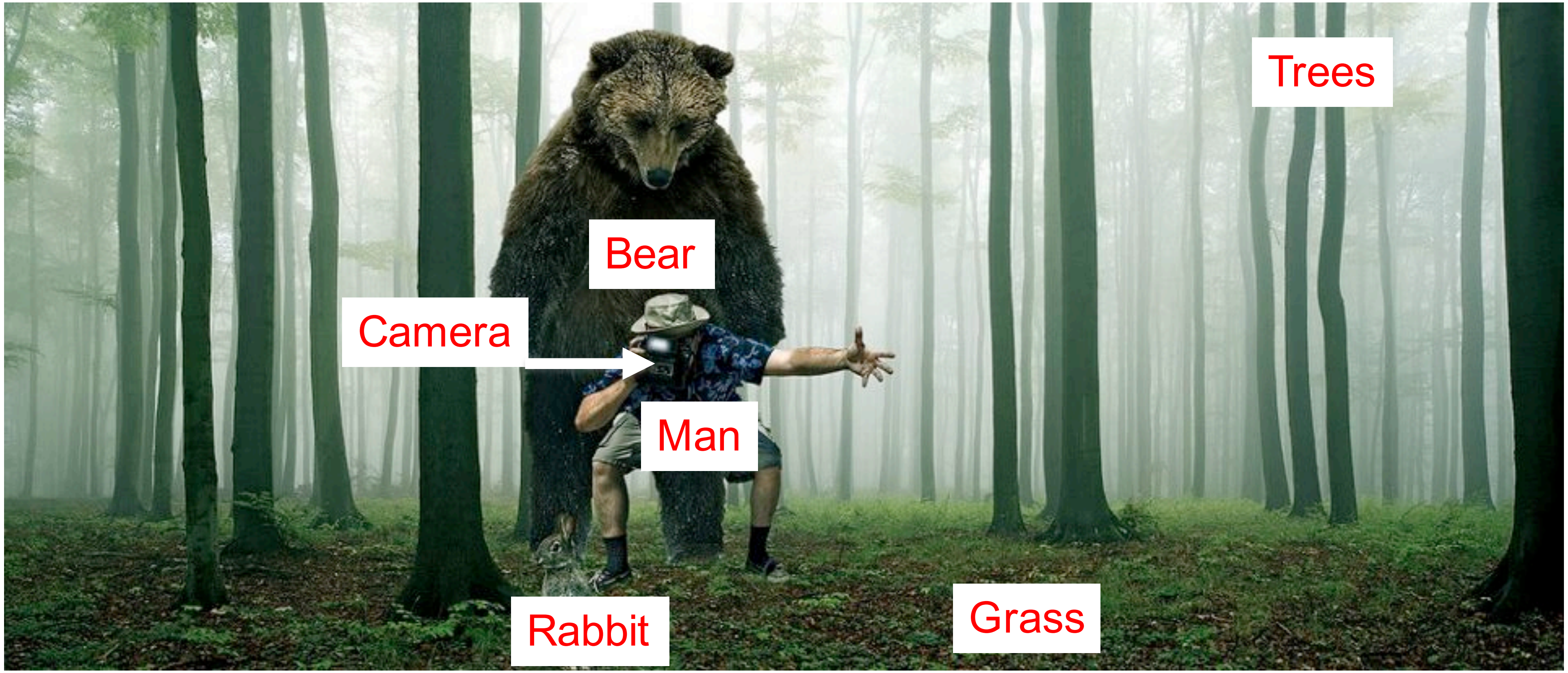


The Fundamental Matrix Song



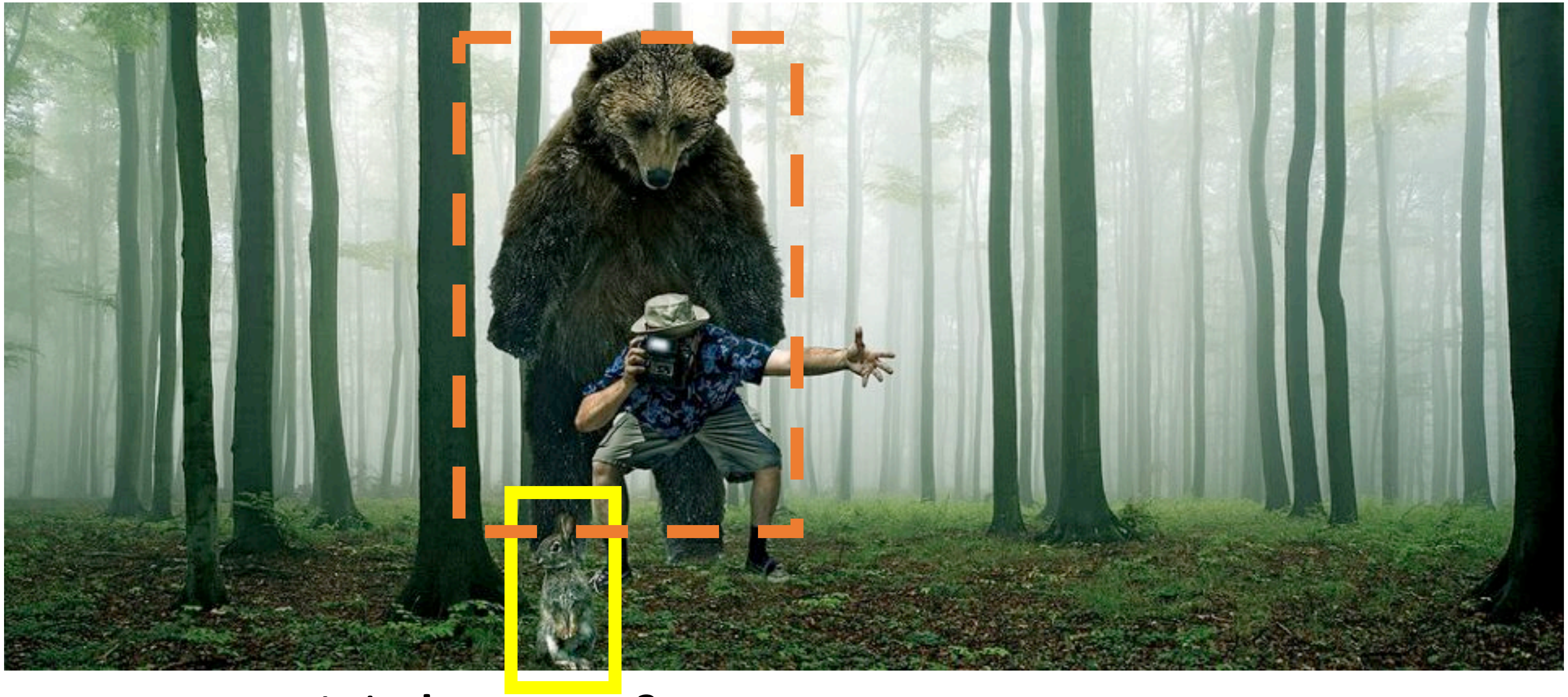
5-mins break

What do you see in this image?



Forest

Describe, predict, or interact with the object based on visual cues



Is it **dangerous**?

Is it **alive**?

How **fast** does it run?

Is it **soft**?

Does it have a **tail**?

Can I **poke with it**?

Why do we care about categories?

- From an object's category, we can make predictions about its behavior in the future, beyond of what is immediately perceived.
- Pointers to knowledge
  - Help to understand individual cases not previously encountered
- Communication

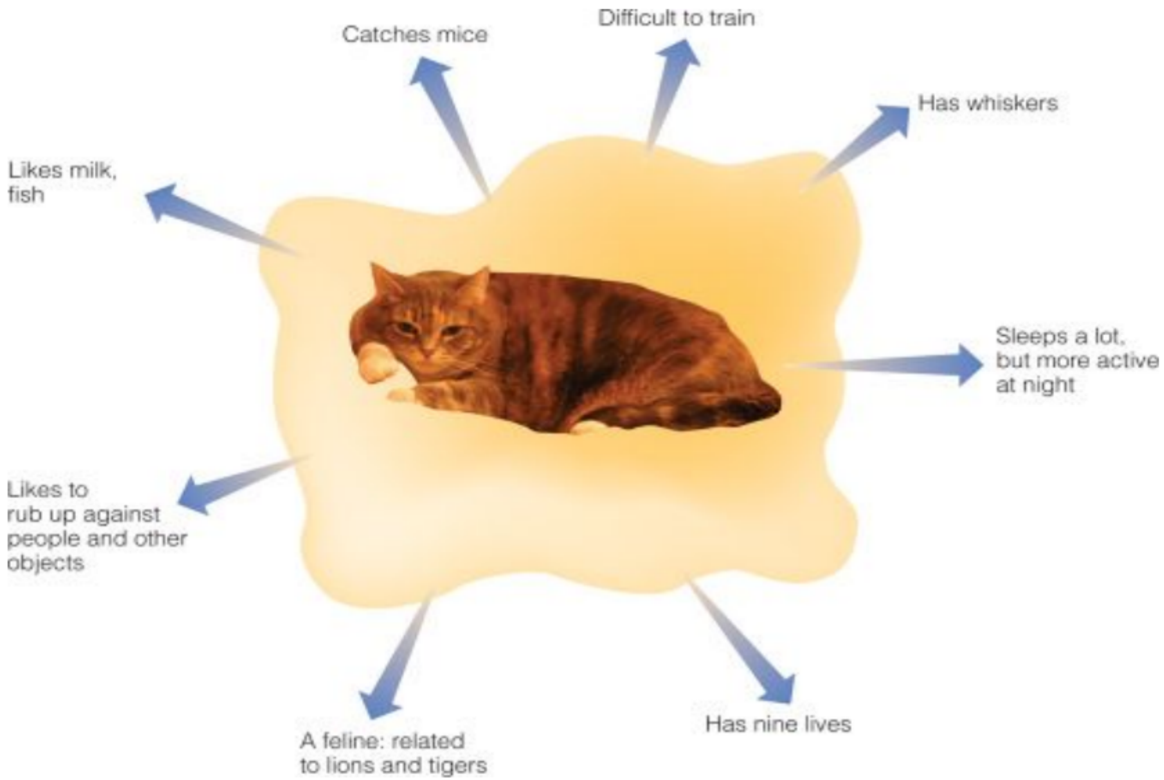
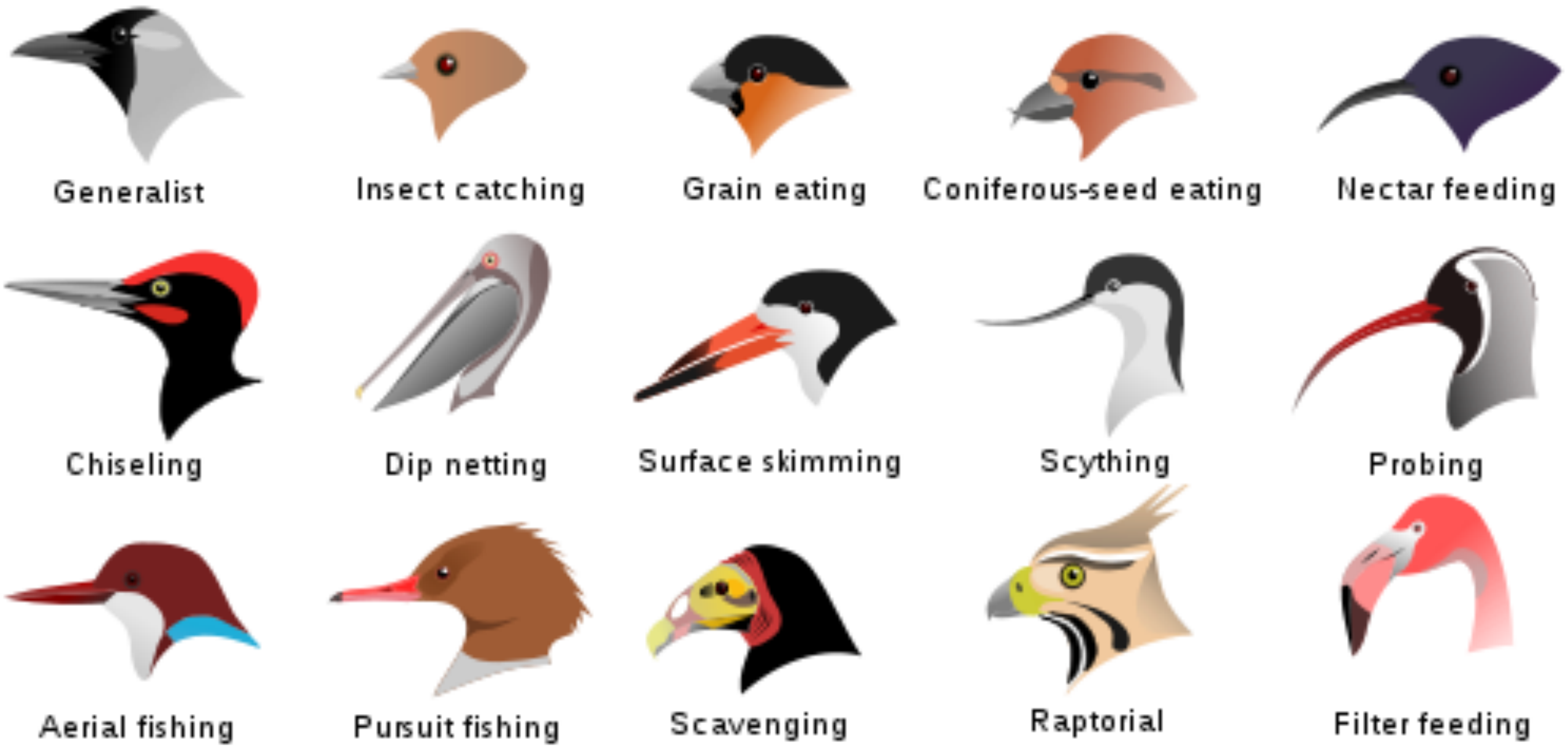


Image categorization: Fine-grained recognition

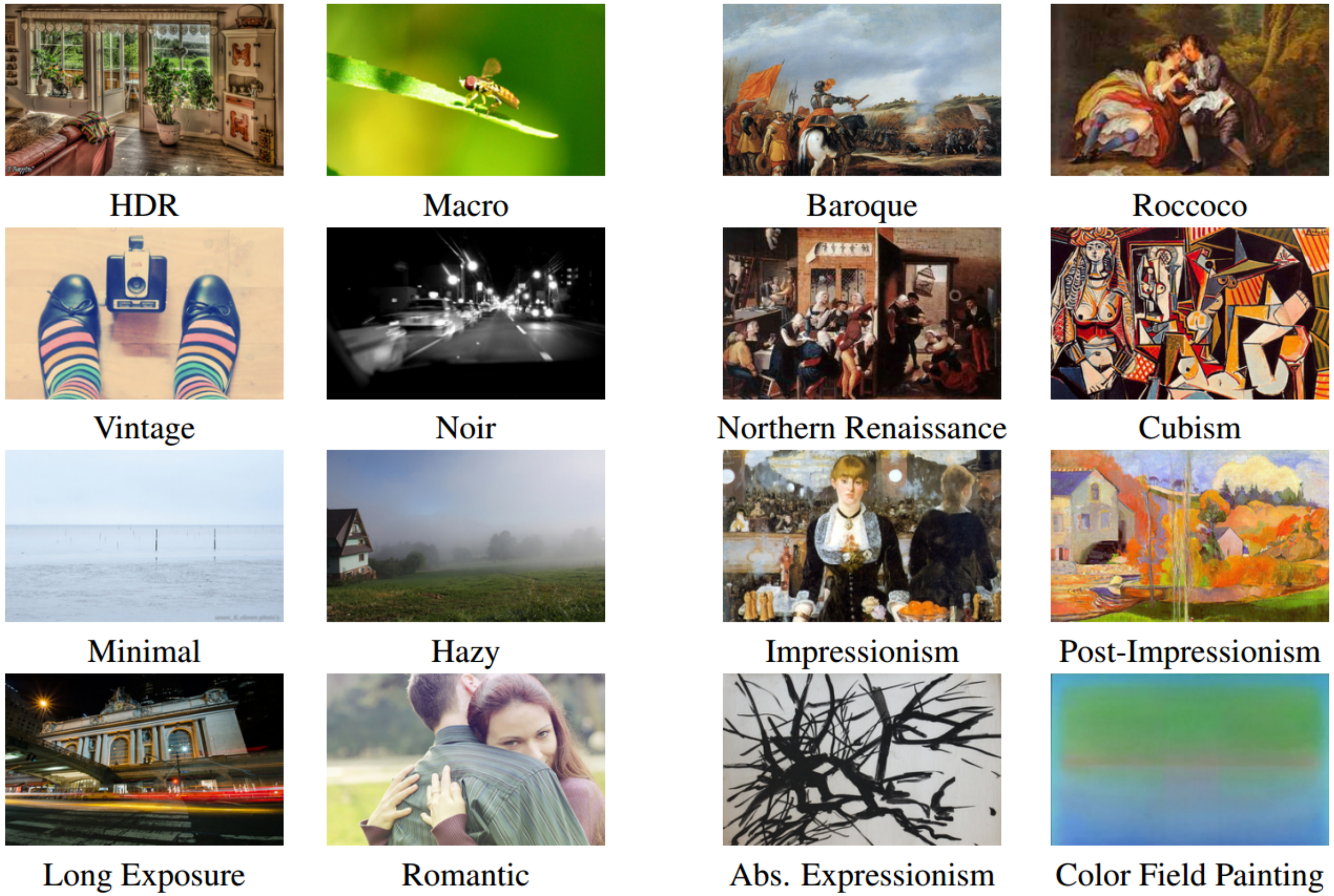


[Visipedia Project]

Place recognition

Places Database [Zhou et al. NIPS 2014]

Image style recognition

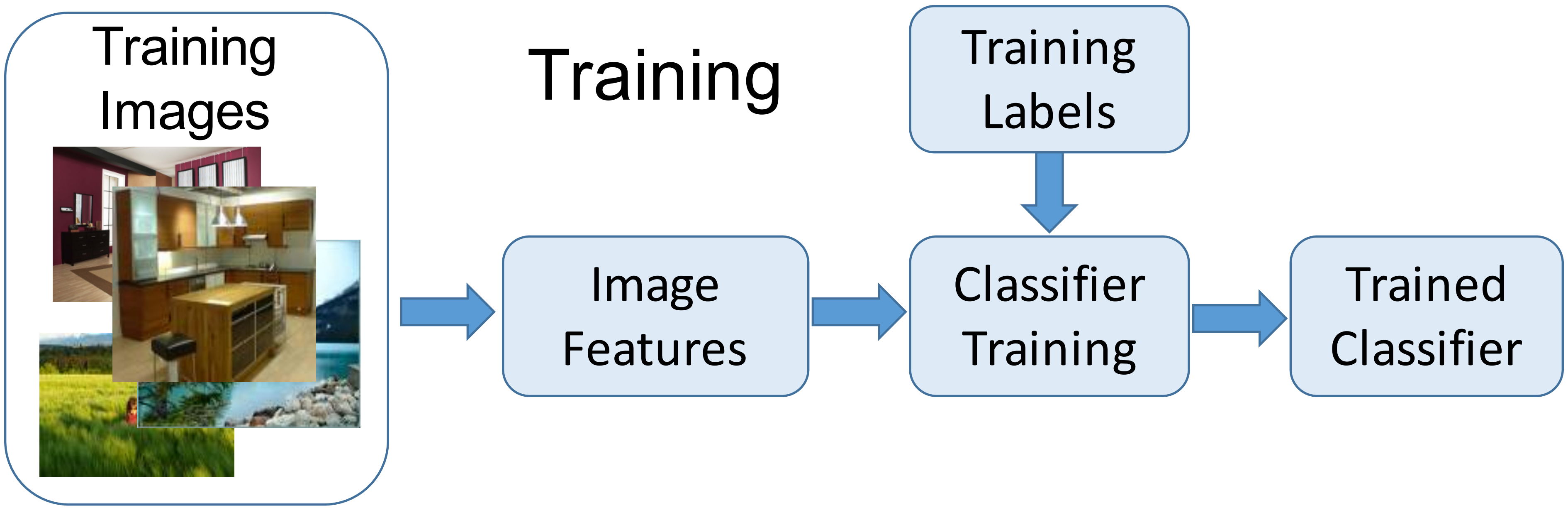


Flickr Style: 80K images covering 20 styles.

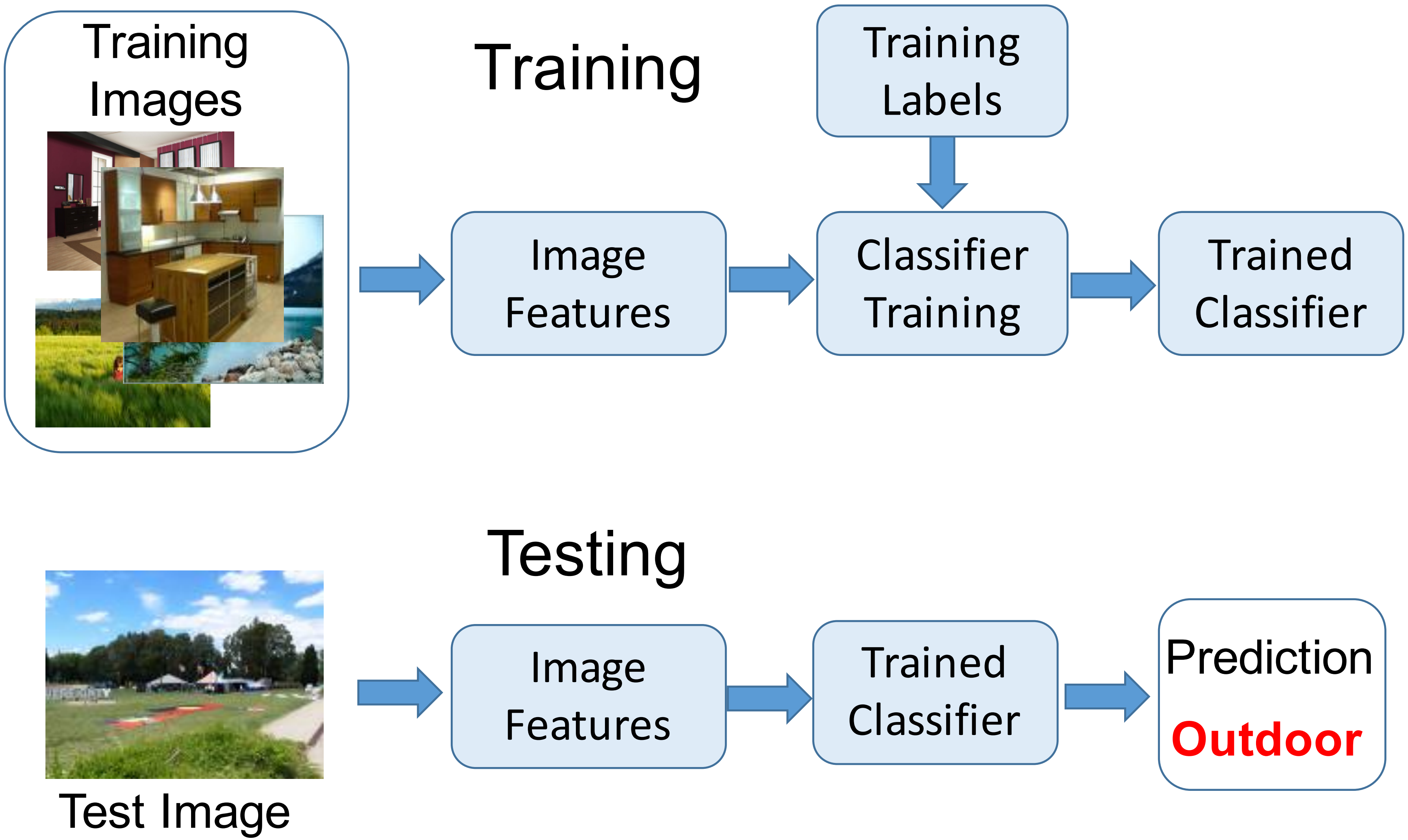
Wikipaintings: 85K images for 25 art genres.

[Karayev et al. BMVC 2014]

Training phase



Testing phase



Q: What are good features for…

- recognizing a beach?

Q: What are good features for…

- recognizing cloth fabric?






Q: What are good features for…

- recognizing a mug?





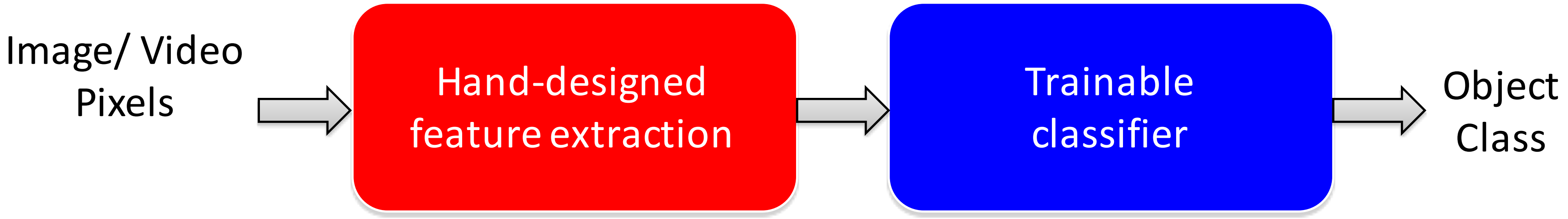

What are the right features?

Depend on what you want to know!

- Object: shape
  - Local shape info, shading, shadows, texture
- Scene : geometric layout
  - linear perspective, gradients, line segments
- Material properties: albedo, feel, hardness
  - Color, texture
- Action: motion
  - Optical flow, tracked points

“Shallow” vs. “deep” architectures

Traditional recognition: “Shallow” architecture



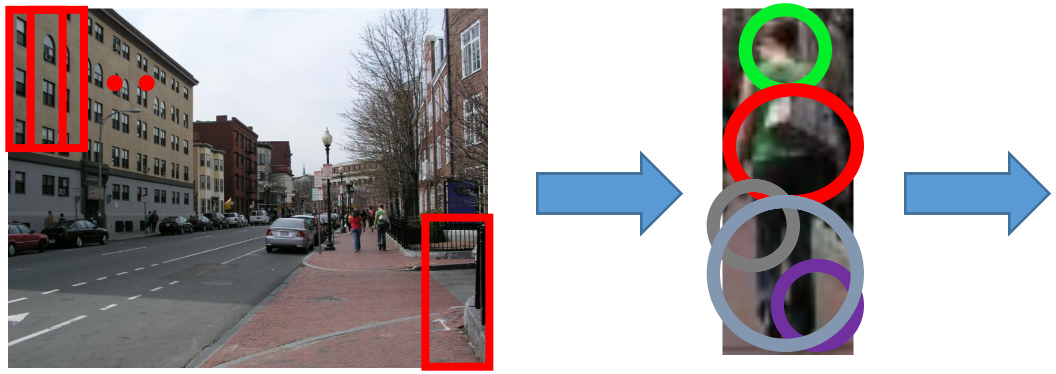
Deep learning: “Deep” architecture

Object Detection

Search by Sliding Window Detector

- May work well for rigid objects

- Key idea: simple alignment for simple deformations



Object or Background?

Object Detection

Search by Parts-based model

- Key idea: more flexible alignment for articulated objects
- Defined by models of **part appearance**, **geometry** or spatial layout, and **search** algorithm

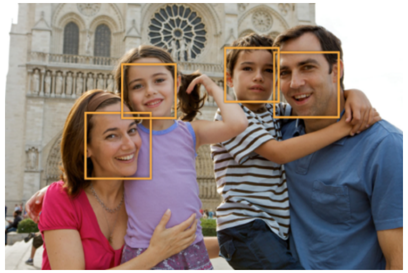
Demo – face detection and alignment

Vision as part of an intelligent system

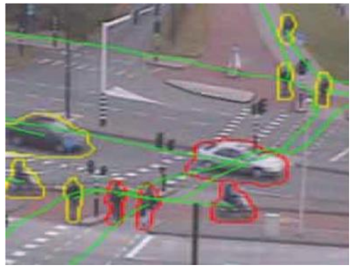
3D Scene

Feature Extraction	Texture	Color	Optical Flow	Stereo Disparity	
Grouping	Surfaces	Bits of objects	Sense of depth	Motion patterns	
Interpretation	Objects	Agents and goals	Shapes and properties	Open paths	Words
Action	Walk, touch, contemplate, smile, evade, read on, pick up, …				

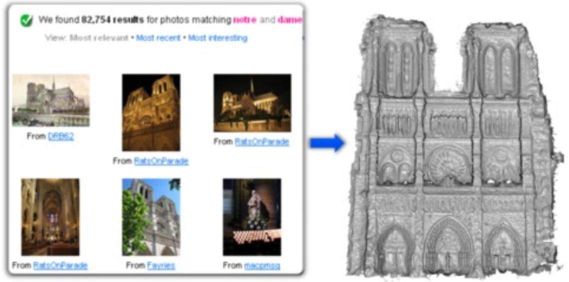
Well-Established (patch matching)



Face Detection/Recognition



Object Tracking / Flow

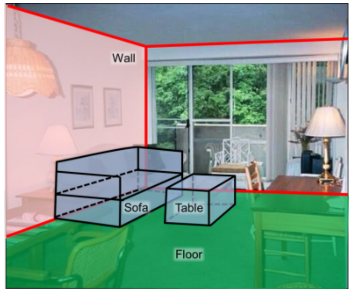


Multi-view Geometry

Major Progress (pattern matching++)

Category Detection

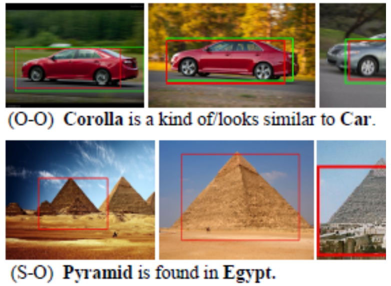
Human Pose



3D Scene Layout

New Opportunities (interpretation/tasks)

Entailment/Prediction



Life-long Learning

Vision for Robots

Slide credit: Derek Hoiem

This course has provided fundamentals

- What is computer vision?
- Examples of computer vision applications
- Basic principles of computer vision:
  - Image formation
  - Filtering
  - Correspondence and alignment
  - Geometry
  - Recognition.

How do you learn more?

Explore!

Resources – where to learn more

- Awesome Computer Vision
  - A curated list of awesome computer vision resources
- Computer Vision Taiwanese Group
  - > 2,200 members in facebook
- Videolectures and ML&CV Talks
  - Lectures, Keynotes, Panel Discussions
- OpenCV – C++/Python/Java

Thank You!

Image: www.clubtuzki.com