Object recognition and Computer Vision 2017

http://www.di.ens.fr/willow/teaching/recvis17/

Introduction and basic camera geometry

Josef Sivic

http://www.di.ens.fr/~josef

INRIA, WILLOW, ENS/INRIA/CNRS UMR 8548

Departement d'Informatique, Ecole Normale Supérieure, Paris

With slides from: **S. Lazebnik**, **J. Ponce**, S. Seitz, R. Szeliski

Object recognition and Computer Vision 2017

http://www.di.ens.fr/willow/teaching/recvis17/

Lectures:

Jean Ponce

Cordelia Schmid

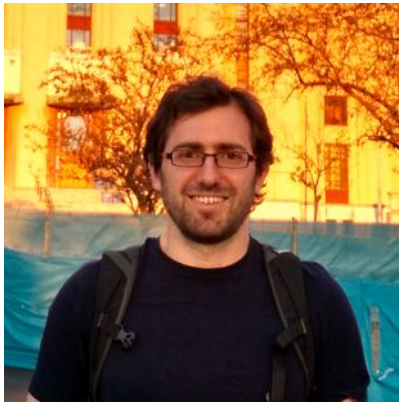
Ivan Laptev

Armand Joulin

Mathieu Aubry

Teaching assistants:

Gul Varol



Ignacio Rocco

Lecture	**Date**	**Topic and reading materials.**	**Slides**
1	Oct 3	**Introduction; Camera geometry (3hrs, J. Sivic)**	
2	Oct 10	**Instance-level recognition I. - Local invariant features, correspondence, image matching (3hrs, J. Sivic);**	
3	Oct 17	**Instance-level recognition II. - Efficient visual search (1.5hrs, C. Schmid)** **Bag-of-feature models for category-level recognition (1.5hrs, C. Schmid)**	
4	Oct 24	**ICCV 2017. No lecture.** Assignments: Assignment 1 due.	
5	Oct 31	**Sparse coding and dictionary learning for image analysis (3hrs, J. Ponce)**	
6	Nov 7	**Neural networks; Optimization methods (3hrs, A. Joulin)** Assignments: Assignment 2 due.	
7	Nov 14	**Convolutional neural networks for visual recognition I. (I. Laptev)** Final project topics are out. Due date for project proposals: Nov 28.	
8	Nov 21	**Convolutional neural networks for visual recognition II. (J. Sivic)** Assignments: Assignment 3 due.	
9	Nov 28	**Motion and human actions I. (C. Schmid)** Assignments: Final project proposal due.	
10	Dec 5	**Human pose estimation; Weakly-supervised learning I (I. Laptev)**	
11	Dec 12	**3D object recognition and Convolutional neural networks (M. Aubry)** **Weakly-supervised learning II (I. Laptev)**	
12	**Jan 15** **Jan 16** **Jan 17**	**Final project presentations and evaluation (I. Laptev, J. Sivic)** **Jan 15: 13:00-17:00** **Jan 16: 13:00-17:00** **Jan 17: 13:00-17:00** The presentations will take place at **Salle Alan Turing - 1st floor at Inria Paris research center, 2 Rue Simone Iff, 75012, Paris. Directions are here.** When you enter the building tell the receptionist you are going for the presentation and go directly to the first floor (no special access card is needed)	

Object recognition and computer vision 2017

Class webpage:

http://www.di.ens.fr/willow/teaching/recvis17

Grading:

- 3 programming assignments (50%)
  - Instance-level recognition
  - Image classification
  - Convolutional Neural Networks
- Final project (50%)

  More independent work, resulting in a report and a class presentation.

Policy

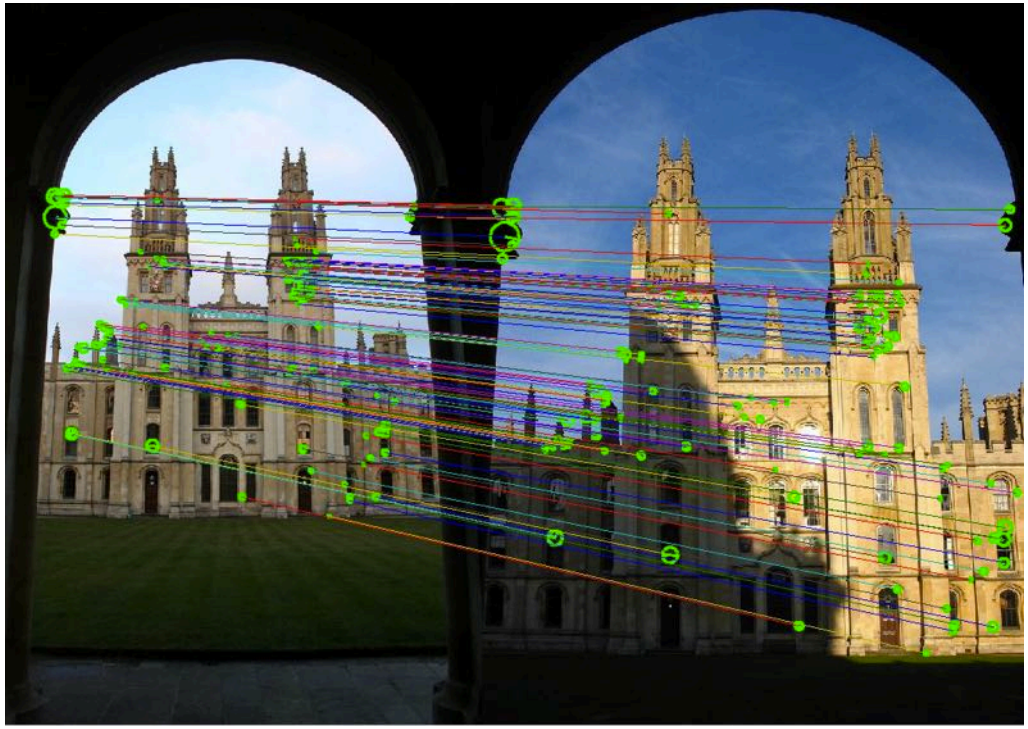
Assignments are strictly individual

Copy-paste of the code, results, parts of the report ⇨ 0p.

FPs can be done in groups of max 2 people

Assignment I: Instance level recognition

- Part I: Sparse features for matching specific objects in images
- Part II: Affine co-variant detectors
- Part III: Towards large scale retrieval
- Part IV: Large scale retrieval

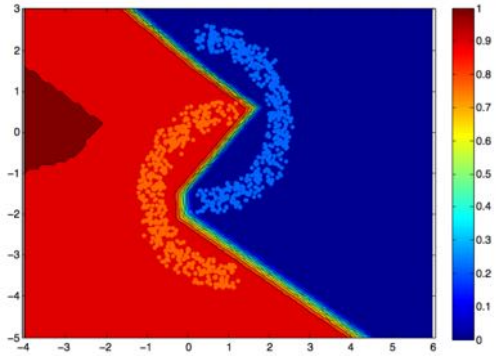


Assignment II: Image Classification

- Part 1: Training and testing an Image Classifier
- Part 2: Training an Image Classifier for Retrieval using Google images

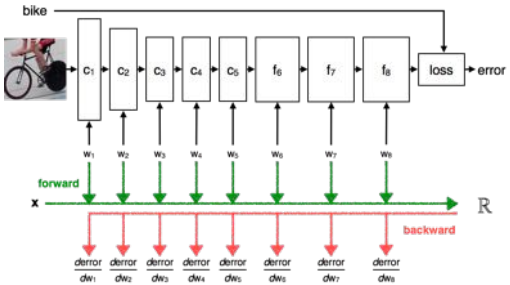
Assignment III: Image Classification with Convolutional Neural Networks

- Part 1: Neural Network's theory:
  - Forward pass, Backward pass
  - Parameter update



- Part 2: PASCAL VOC Image classification with ConvNet features

- Part 3: Convolutional neural networks



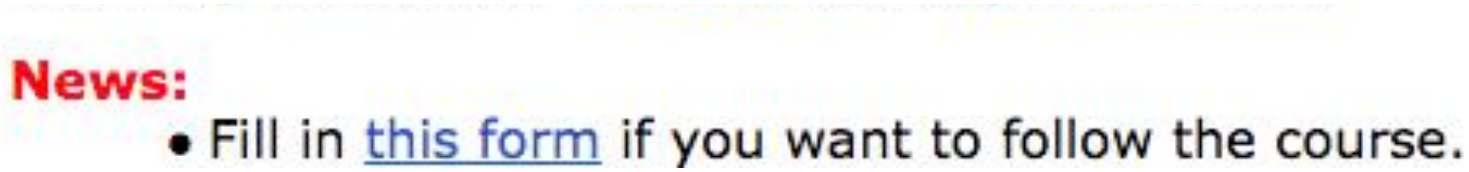
Final project

- Select the topic + write project proposal
- Present the work in the class
- Write project report

- Can be done individually or as a **group of max 2 people**
- The proposed project topics are from the recent top-conference publications in computer vision, see example topics from 2016 here: http://www.di.ens.fr/willow/teaching/recvis16/
- Student-defined projects are welcome
- Final project can be joint with another MVA course

Register on the class webpage

On the class website, fill-in the registration form (name, email, school)



We will create an account for you on the class Moodle.

Assignments / reports will be collected using the Moodle.

Assignment 1 is out. Due on Oct 24th.

Recap: TODO

1. Register by filling-in form on the class webpage (asap)

2. Fill-in Doodle for Matlab tutorial (by Wed Oct 10 6pm)

3. Assignment 1 – Instance-level recognition

http://www.di.ens.fr/willow/teaching/recvis17/assignment1/

Due in 3 weeks: Oct 24 2017

Matlab tutorial

Please fill-in a Doodle linked from the class webpage
by **Wed Oct 10 6pm.**

The tutorial will be at:

INRIA/Willow, 2 Rue Simone Iff, Paris

Who should participate?

- Students with no or limited experience with Matlab.

Introduction to computer vision

http://imagine.enpc.fr/~aubrym/lectures/introvis17/

Thursdays 9:00 - 12:00 at Salle R, ENS ULM.

Taught by Mathieu Aubry.

M1 course

Covers the basics of computer vision in detail.

Mathieu Aubry

Research

Both WILLOW (J. Ponce, I. Laptev, J. Sivic) and LEAR (C. Schmid) groups are active in computer vision and visual recognition research.

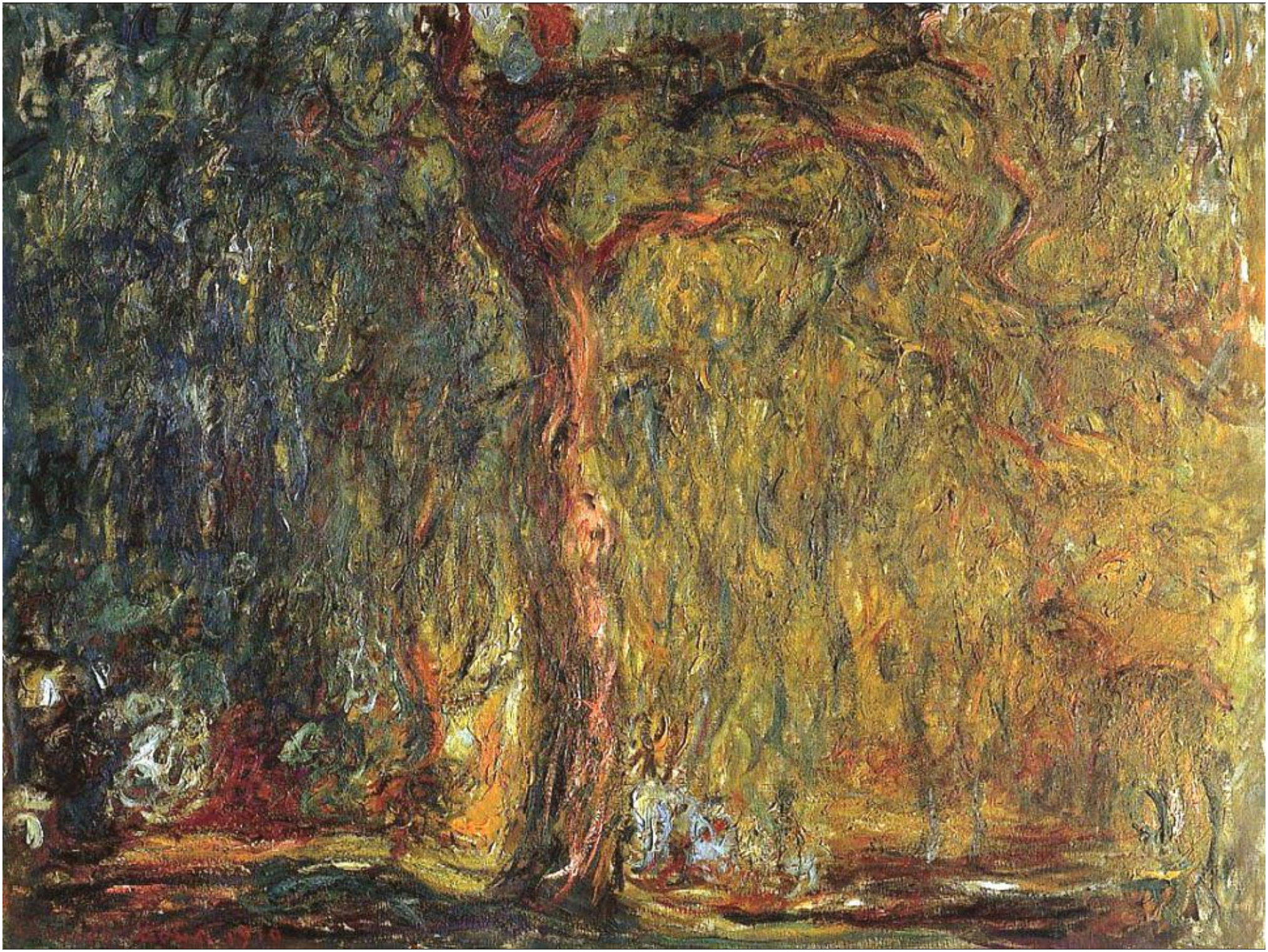
http://www.di.ens.fr/willow/

http://lear.inrialpes.fr/

with close links to SIERRA – machine learning (F. Bach)

http://www.di.ens.fr/sierra/

There will be master internships available.
Talk to us if you are interested!

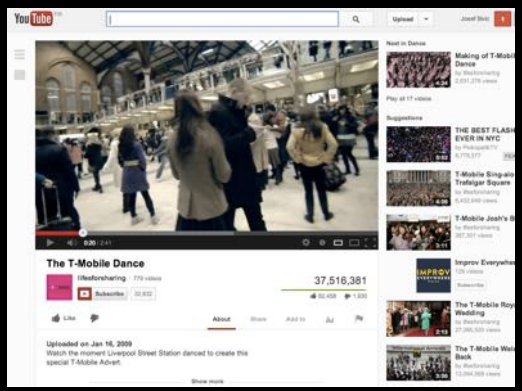


Outline

- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry

Why is visual recognition important?
Images are all around us

Archives of visual information



Internet videos

10,000+ **TV channels**

Historical imagery

Cameras around us

2M+ **surveillance cameras**

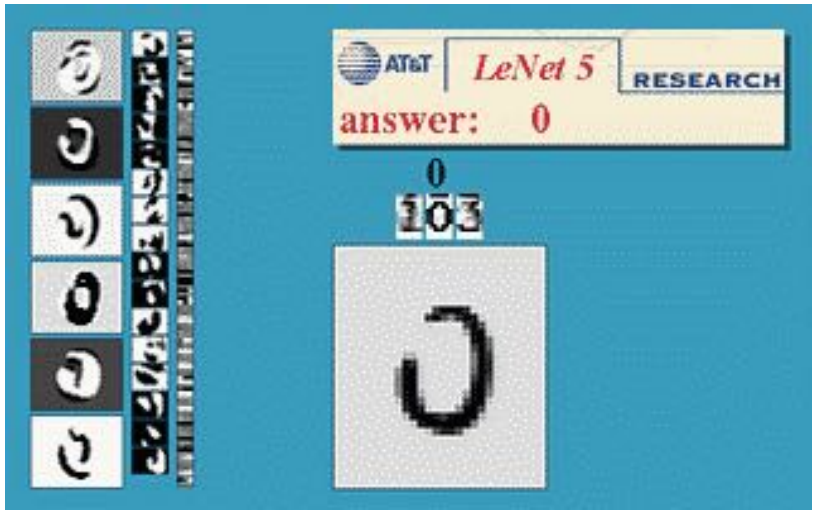
Car cameras

Personal cameras

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



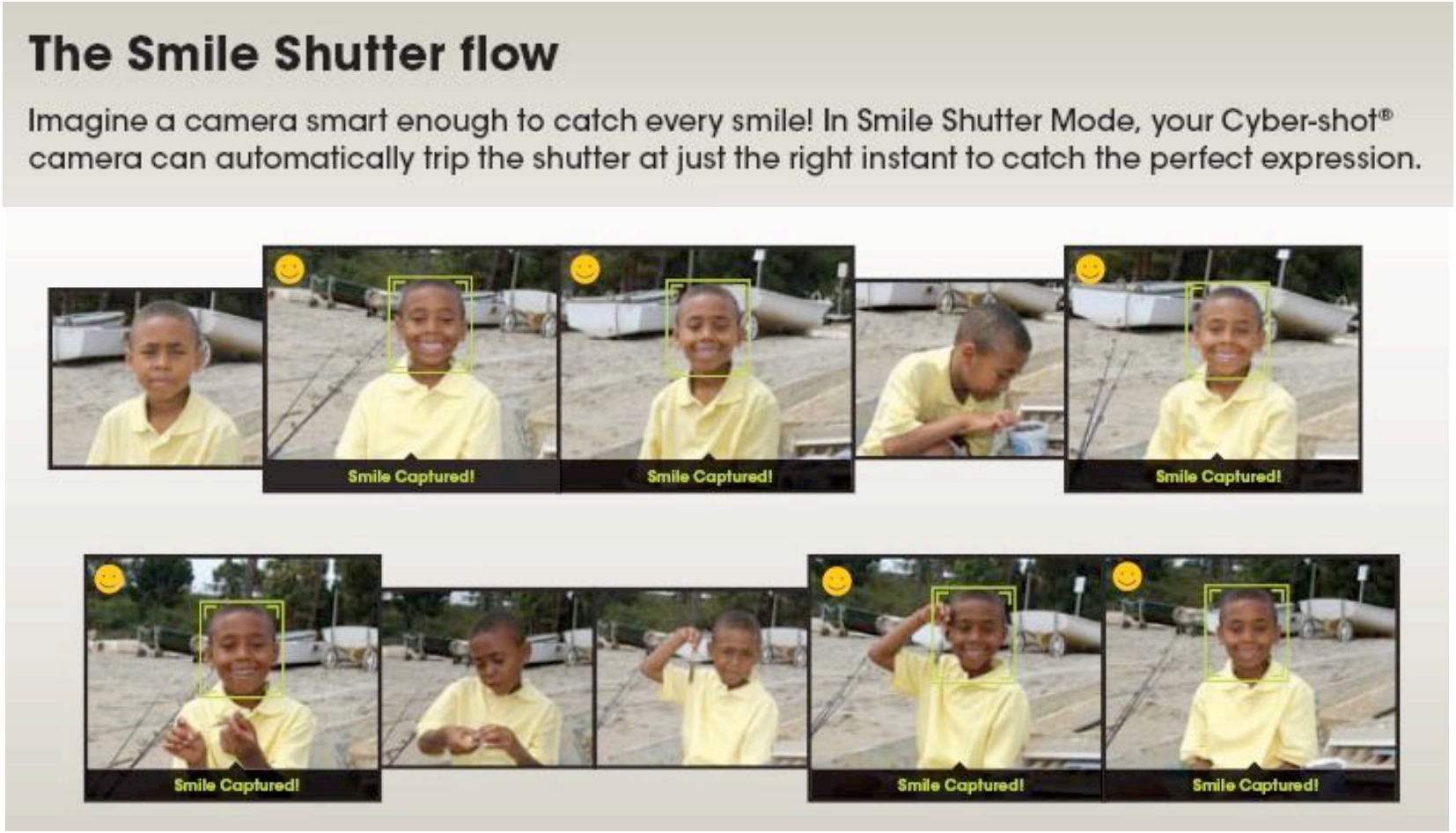
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

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

Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Smile detection



Sony Cyber-shot® T70 Digital Still Camera

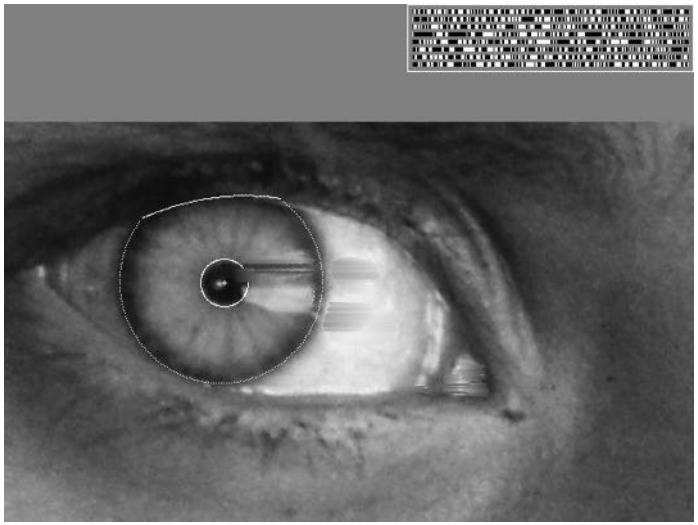
Object recognition (in supermarkets)

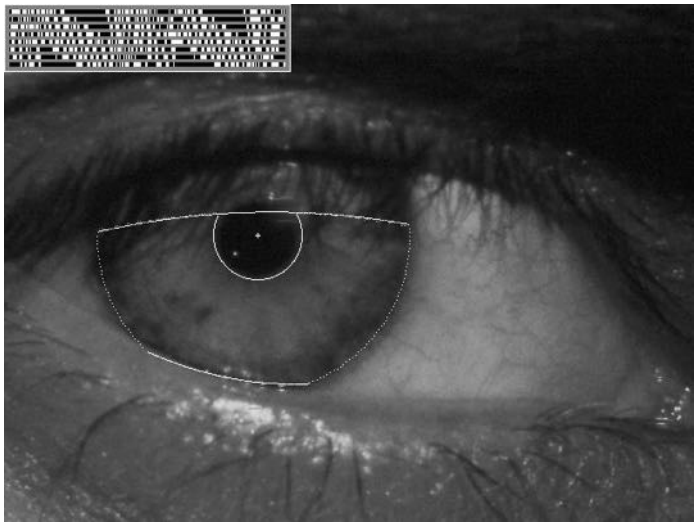
LaneHawk by EvolutionRobotics

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… “

Vision-based biometrics

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





Login without a password…

Fingerprint scanners on many new laptops, other devices

Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

FaceID on the latest iPhone

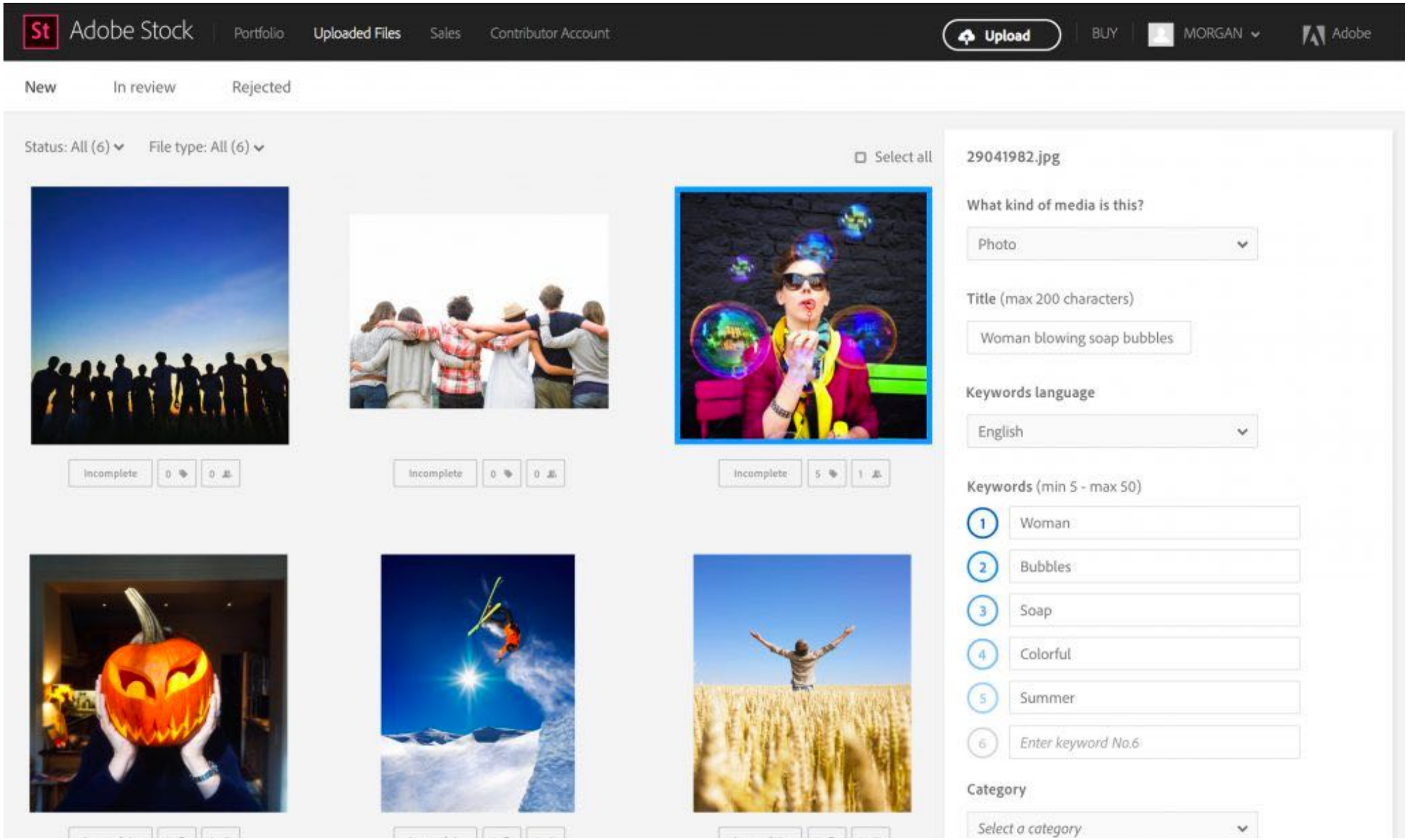
Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

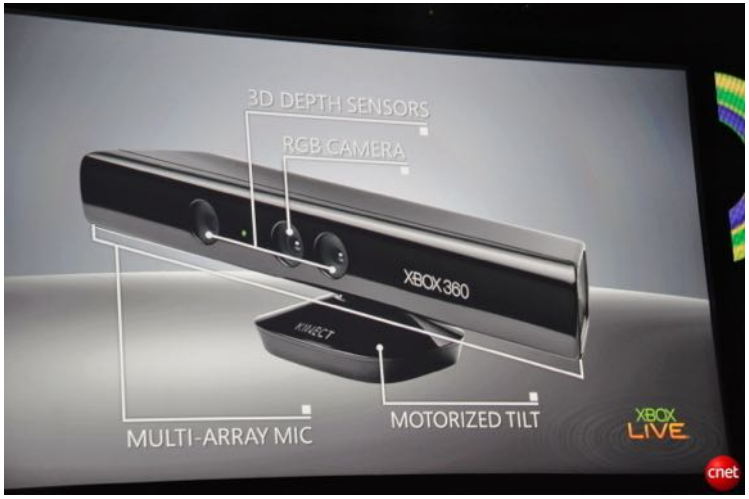
Helping to find content

- Adobe Stock auto-keywording



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



Smart cars

Slide content courtesy of Amnon Shashua

- Mobileye
  - Bought by Intel: 15.3 Billion USD
  - See also CVPR 2016 keynote

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)

Ford acquires SAIPS for self-driving machine learning and computer vision tech

Posted Aug 16, 2016 by Darrell Etherington (@etherington)

Ford outlined a few of the ways it's aiming to ship driverless cars by 2021, and part of the plan involves acquisitions. CEO Mark Fields revealed at a press event in Palo Alto today that the automaker acquired SAIPS, an Israeli company focusing on machine learning and computer vision. It's also partnering exclusively with Nirenberg Neuroscience, to bring more "humanlike intelligence" to machine learning components of driverless car systems.

SAIPS' technology brings image and video processing algorithms, as well as deep learning tech focused on processing and classifying input signals, all key ingredients in the special sauce that makes up autonomous vehicle tech. This company's expertise should help with on-board interpretation of data captured by sensors on Ford's self-driving cars, and turning that data into usable info for the car's virtual driver system. SAIPS' offerings include detection of anomalies, persistent tracking of objects detected by sensors, and much more. The company's past clients include HP and Trax, but its partner group doesn't appear to have included much in the way of driving-specific applications.

Paris, 14 June 2017 – Valeo launches the first global research center in artificial intelligence and deep learning dedicated to automotive applications.

Industrial robots

Vision-guided robots position nut runners on wheels

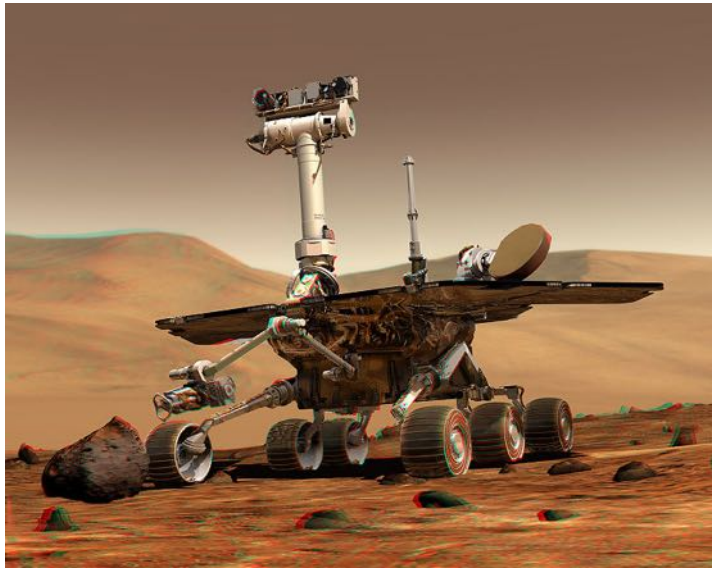
Vision in space

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

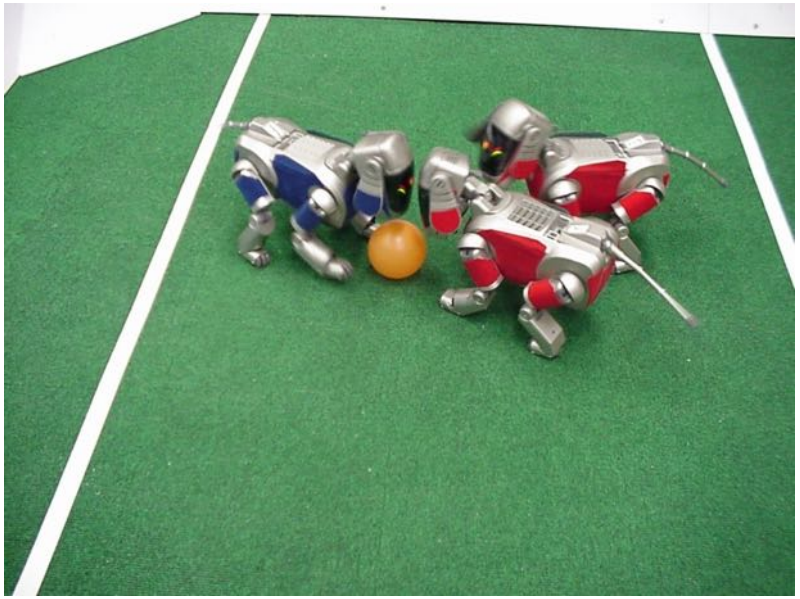
Vision systems (JPL) used for several tasks

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

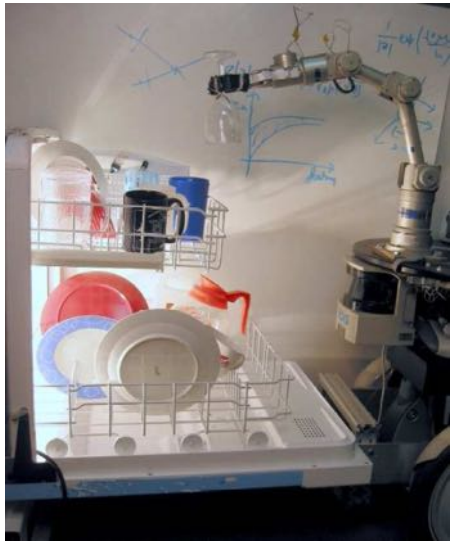
Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



http://www.robocup.org/

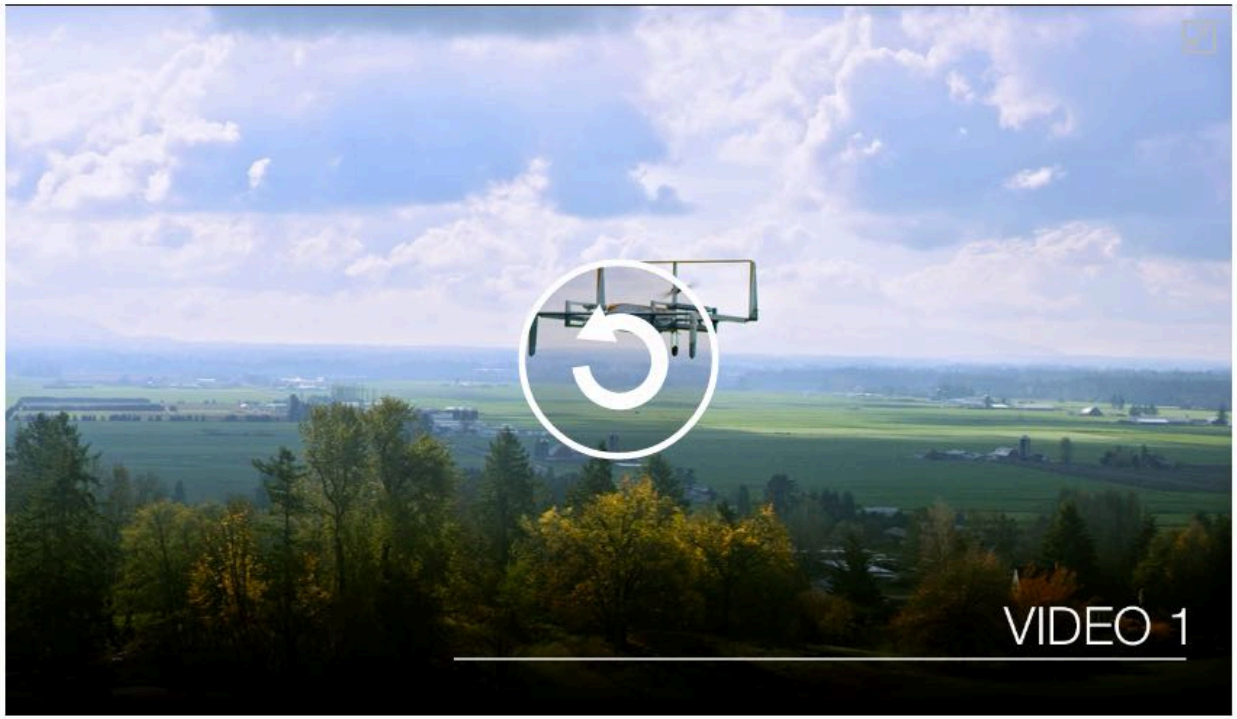


Saxena et al. 2008
STAIR at Stanford

Amazon Prime Air

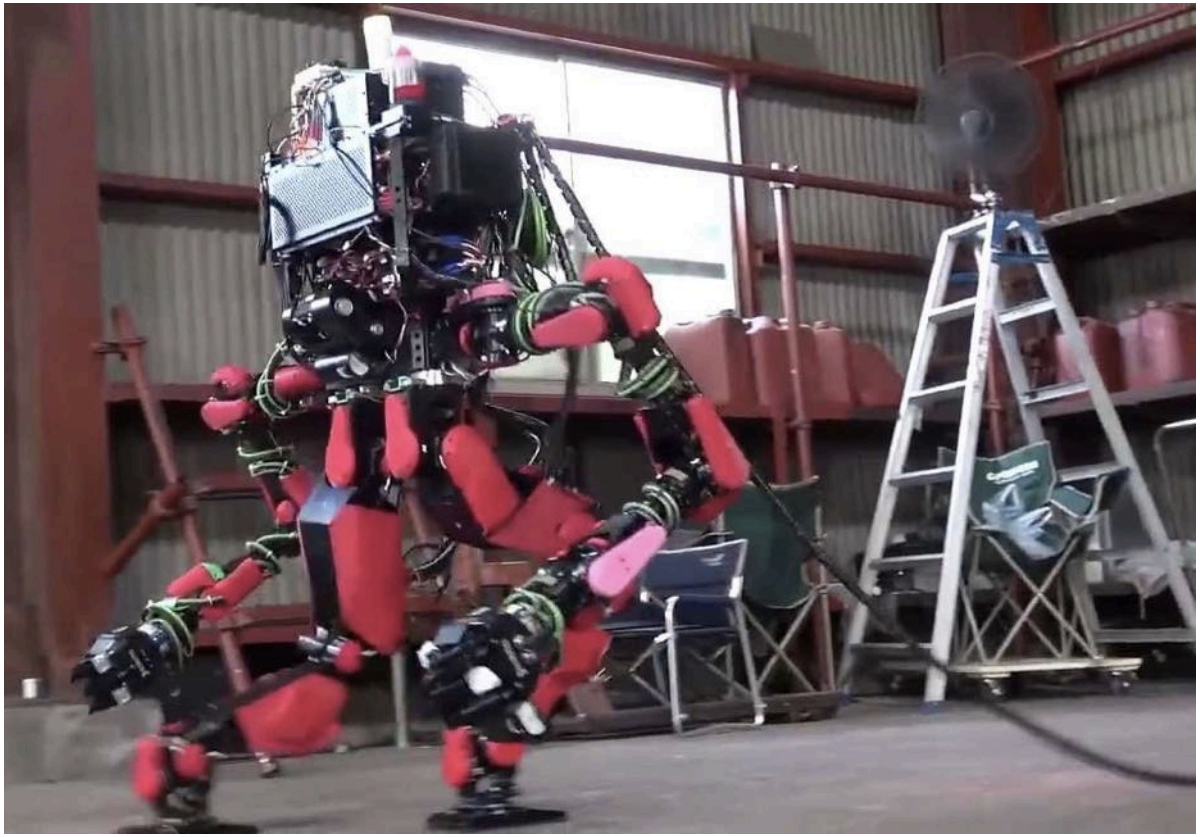
amazon Prime Air

We're excited about Prime Air — a future delivery system from Amazon designed to safely get packages to customers in 30 minutes or less using small unmanned aerial vehicles, also called drones. Prime Air has great potential to enhance the services we already provide to millions of customers by providing rapid parcel delivery that will also increase the overall safety and efficiency of the transportation system. Putting Prime Air into service will take some time, but we will deploy when we have the regulatory support needed to realize our vision.



https://www.amazon.com/b?node=8037720011

Assistive robotics for industry and home



[Robot by Schaft, DARPA robot challenge 2015]

Augmented Reality and Virtual Reality

Magic Leap, Oculus, Hololens, etc.

Virtual assistants at home and industry

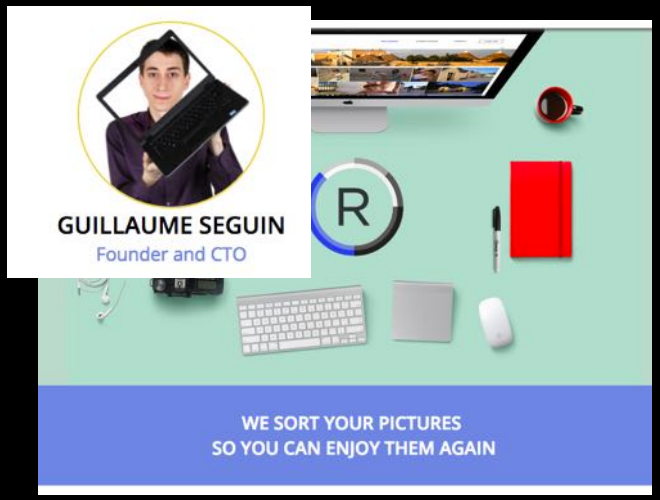
To assist people
[Microsoft HoloLens 2015]

Computer vision as a job

Vincent Delaitre
Phd, 2015
Start-up Deepomatic.com

https://www.inria.fr/centre/paris/actualites/deepomatic-un-shazam-de-l-image-pour-le-e-commerce-et-les-medias

Guillaume Seguin
Phd, 2016
Start-up regaind.io



Piotr Bojanowski
Phd, 2016
Facebook AI Research

Oliver Whyte
Phd, 2012
Engineer at Microsoft

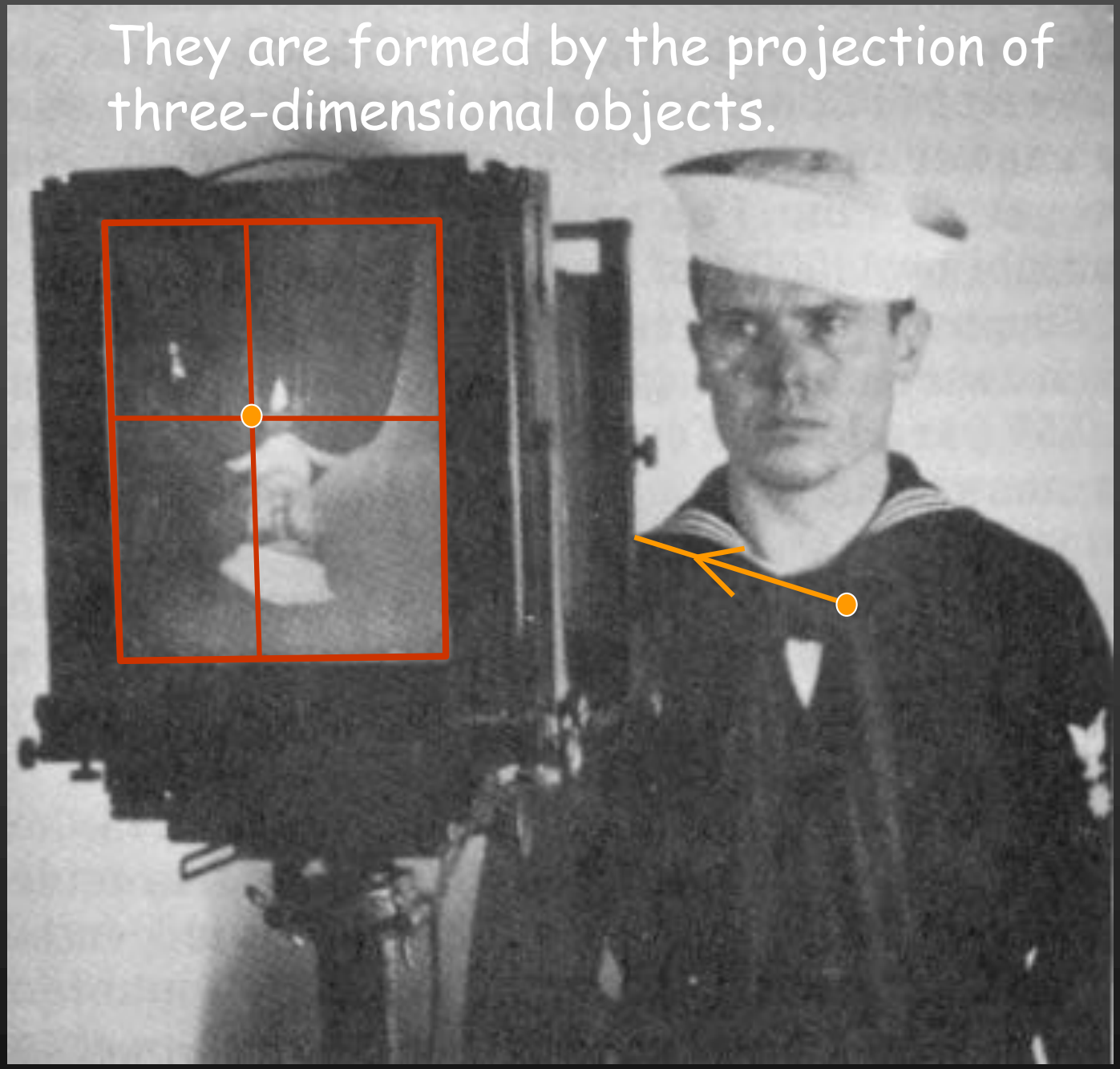
Mathieu Aubry
Phd, 2015
Faculty at ENPC

Relja Arandjelovic
Post-doc, 2016
Google DeepMind

Why is computer vision difficult?

Image source: J. Koenderink

They are formed by the projection of three-dimensional objects.



Images are brightness/color patterns drawn in a plane.

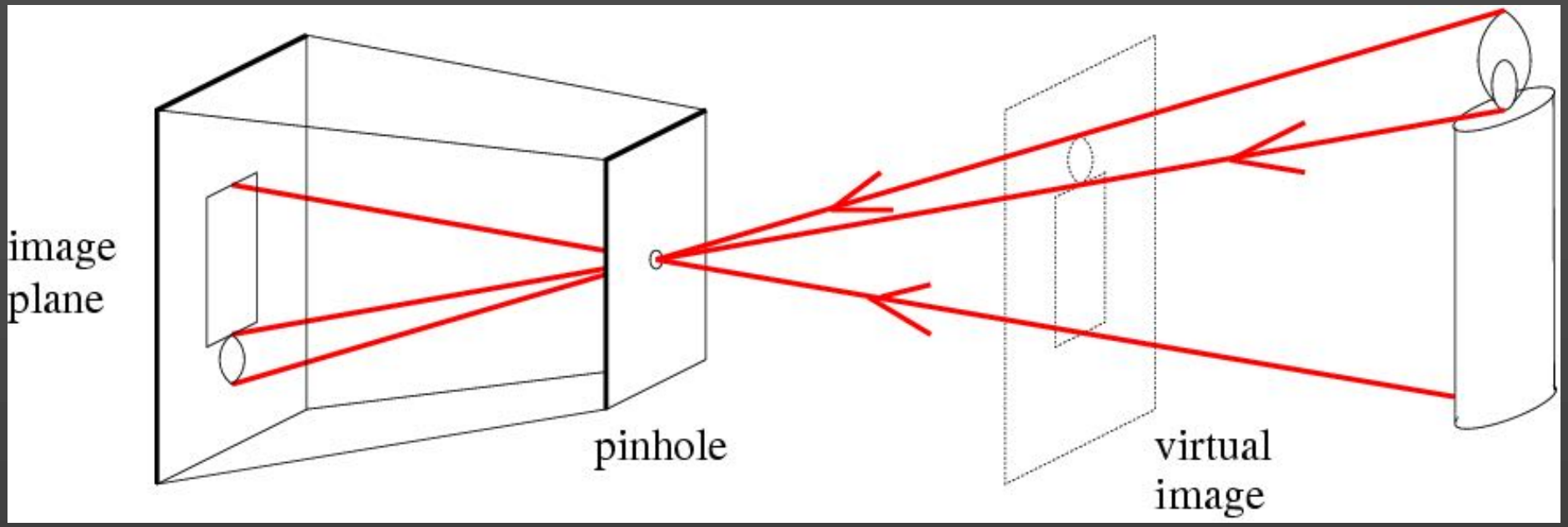
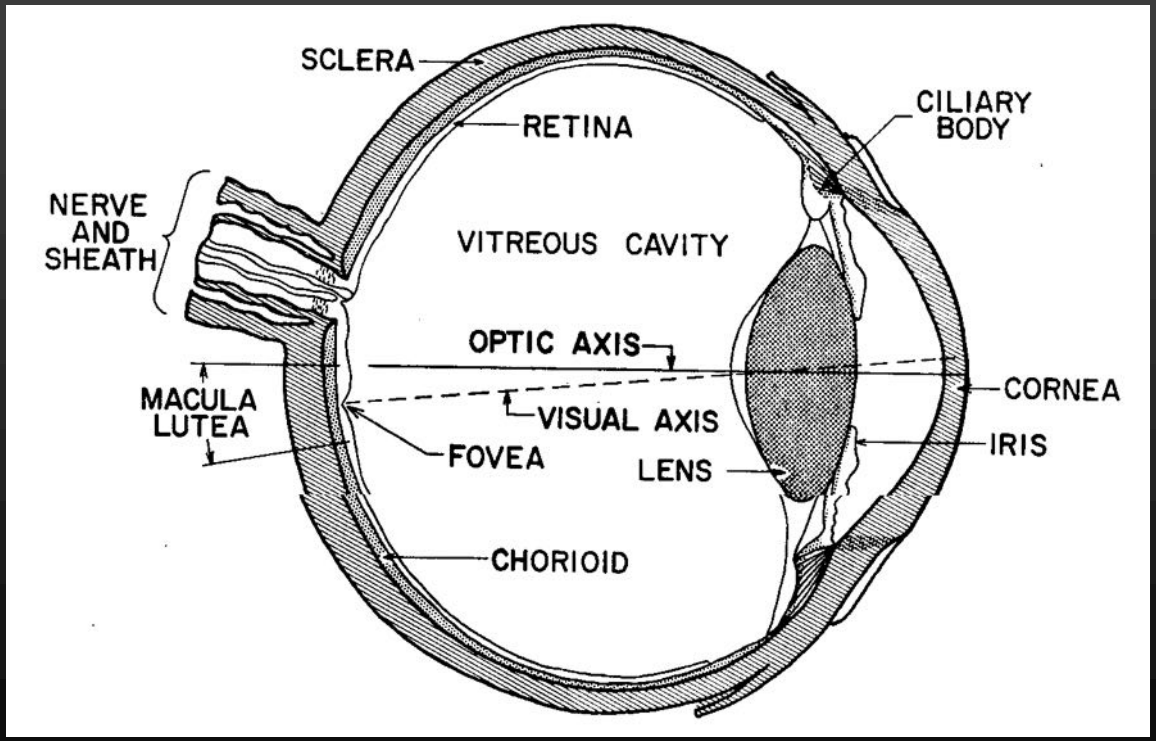
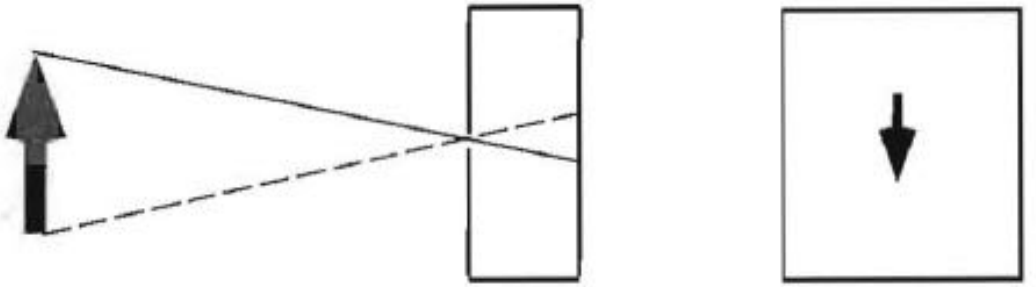


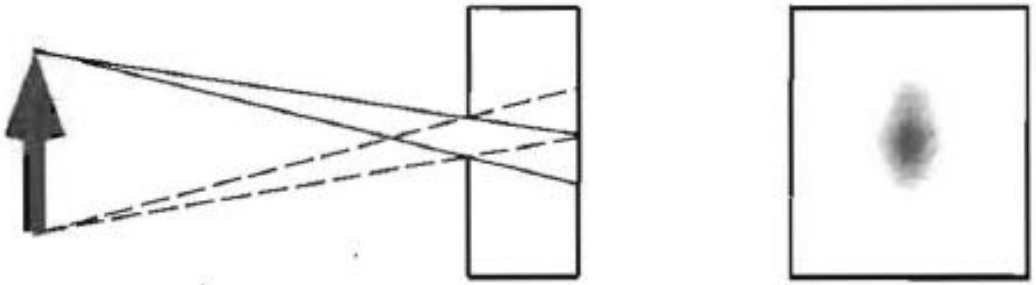
image
plane
pinhole
virtual
image


SCLERA
RETINA
CILIARY
BODY
NERVE
AND
SHEATH
VITREOUS CAVITY
OPTIC AXIS
CORNEA
MACULA
LUTEA
VISUAL AXIS
FOVEA
IRIS
LENS
CHORIOID

Pinhole camera: trade-off between sharpness and light transmission



A. Pinhole Aperture without Lens --> Sharp Image

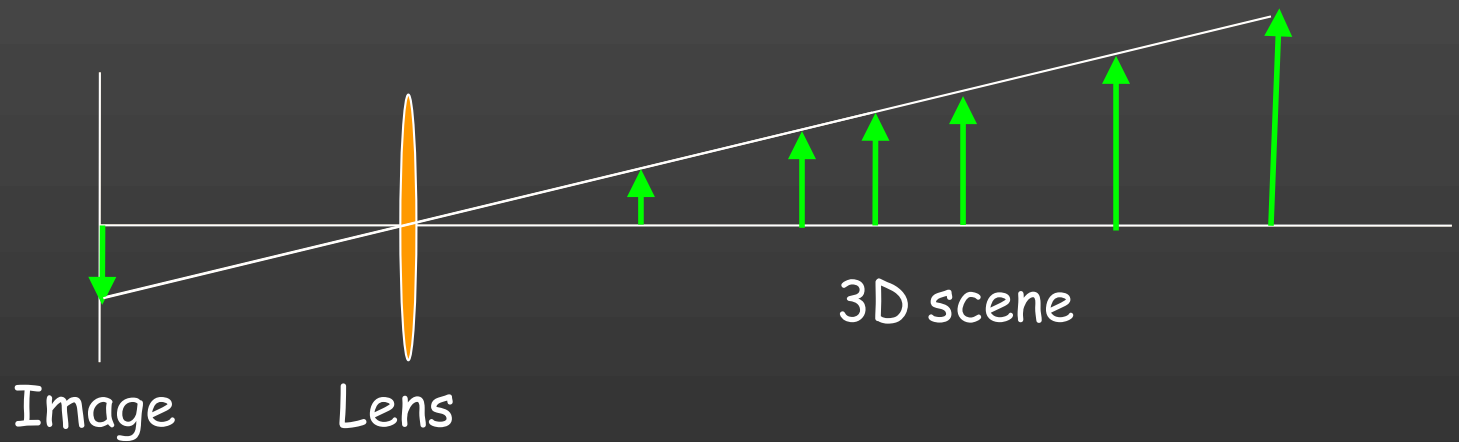


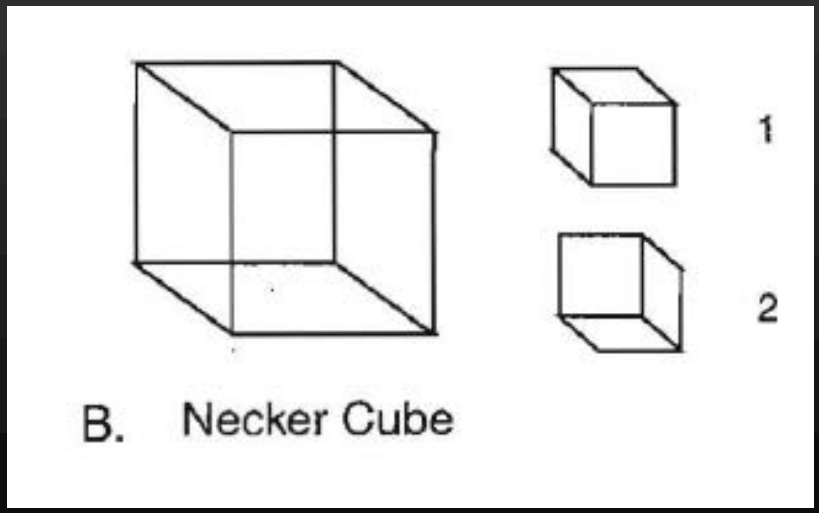
B. Large Aperture without Lens --> Fuzzy Image

Camera Obscura in Edinburgh

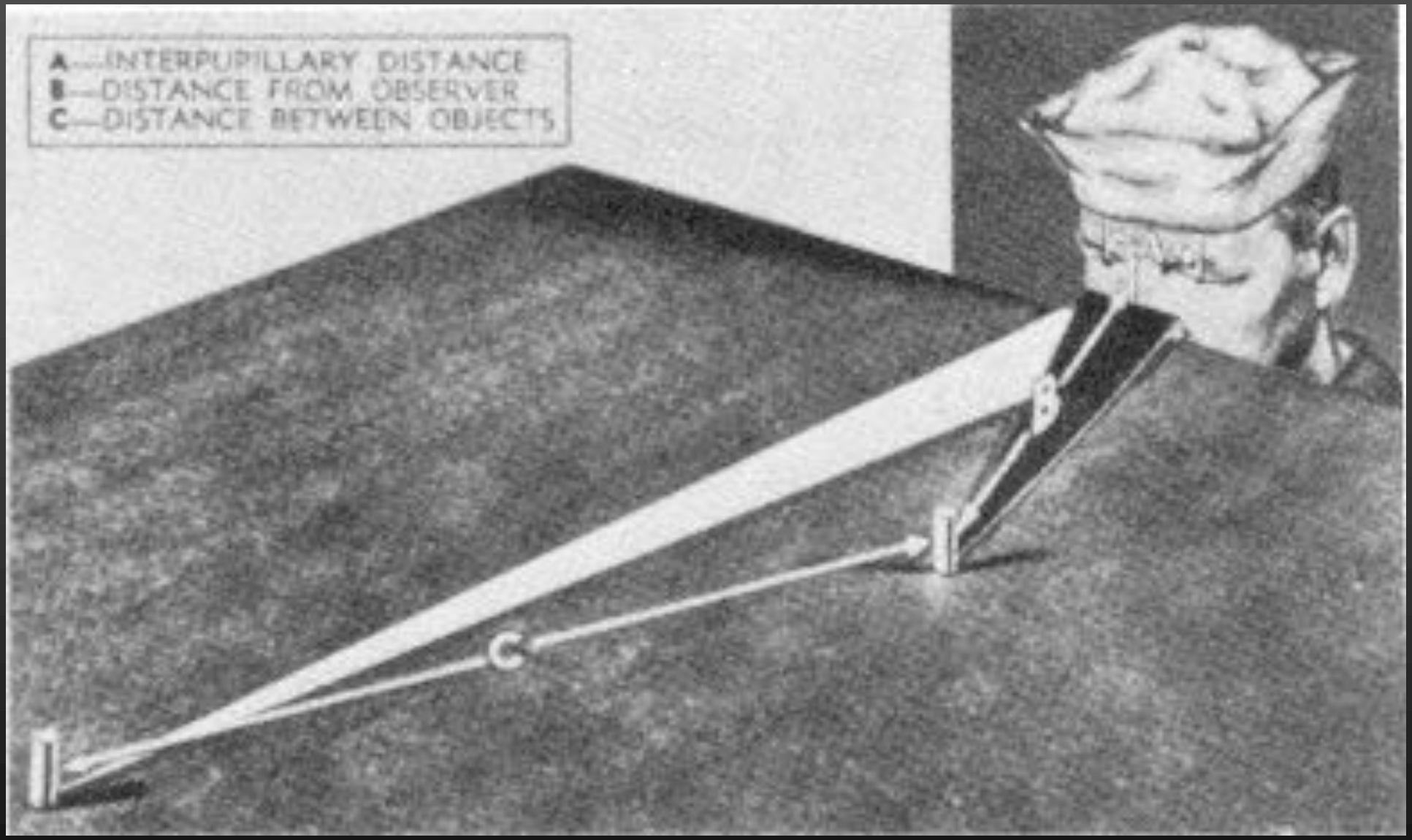
Fundamental problem I:
3D world is "flattened" to 2D images

➡ Loss of information





Question : how do we see "in 3D" ?

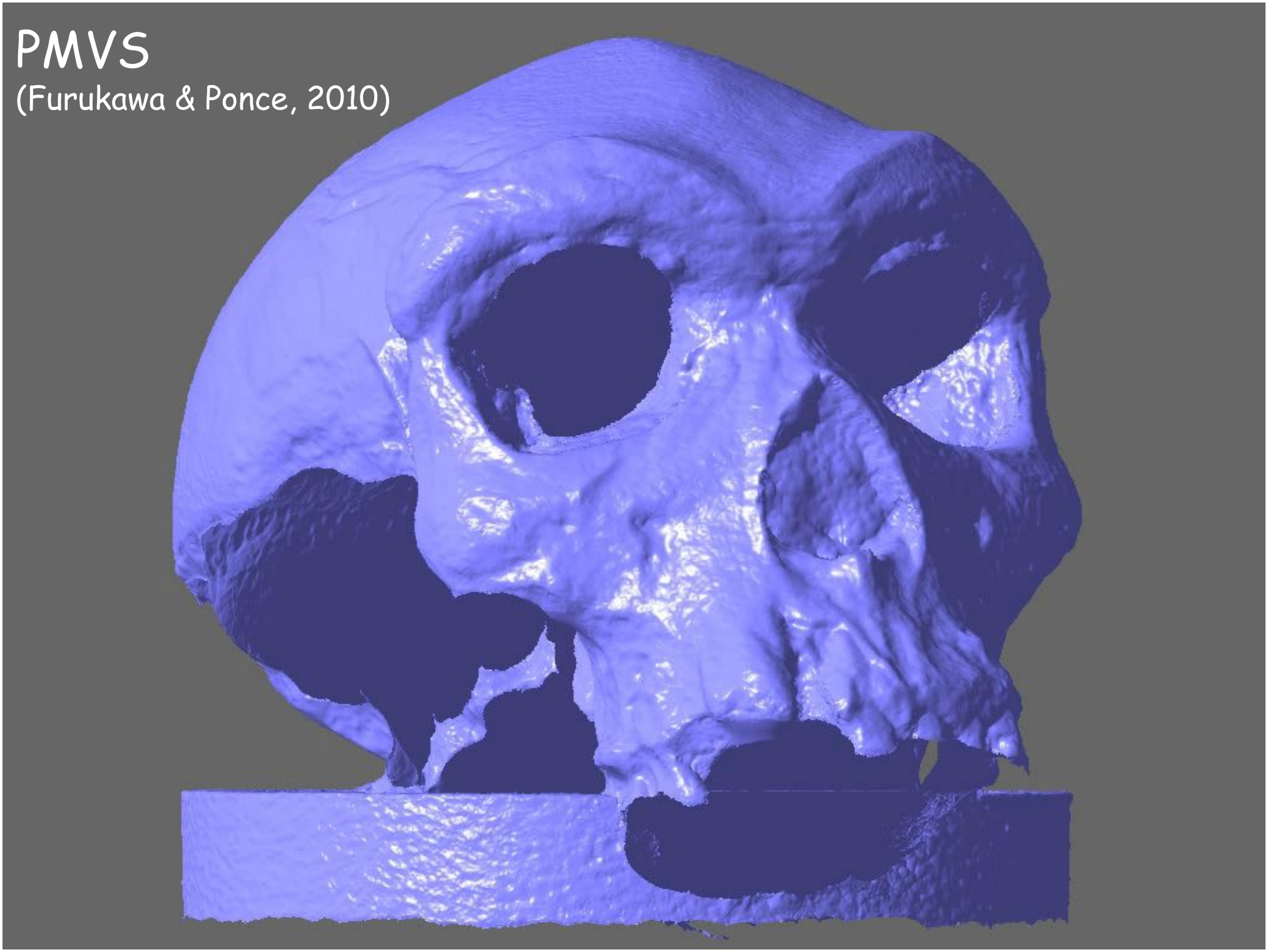


(First-order) answer: with our two eyes.

Simulated 3D perception

PMVS

(Furukawa & Ponce, 2010)



But there are other cues..

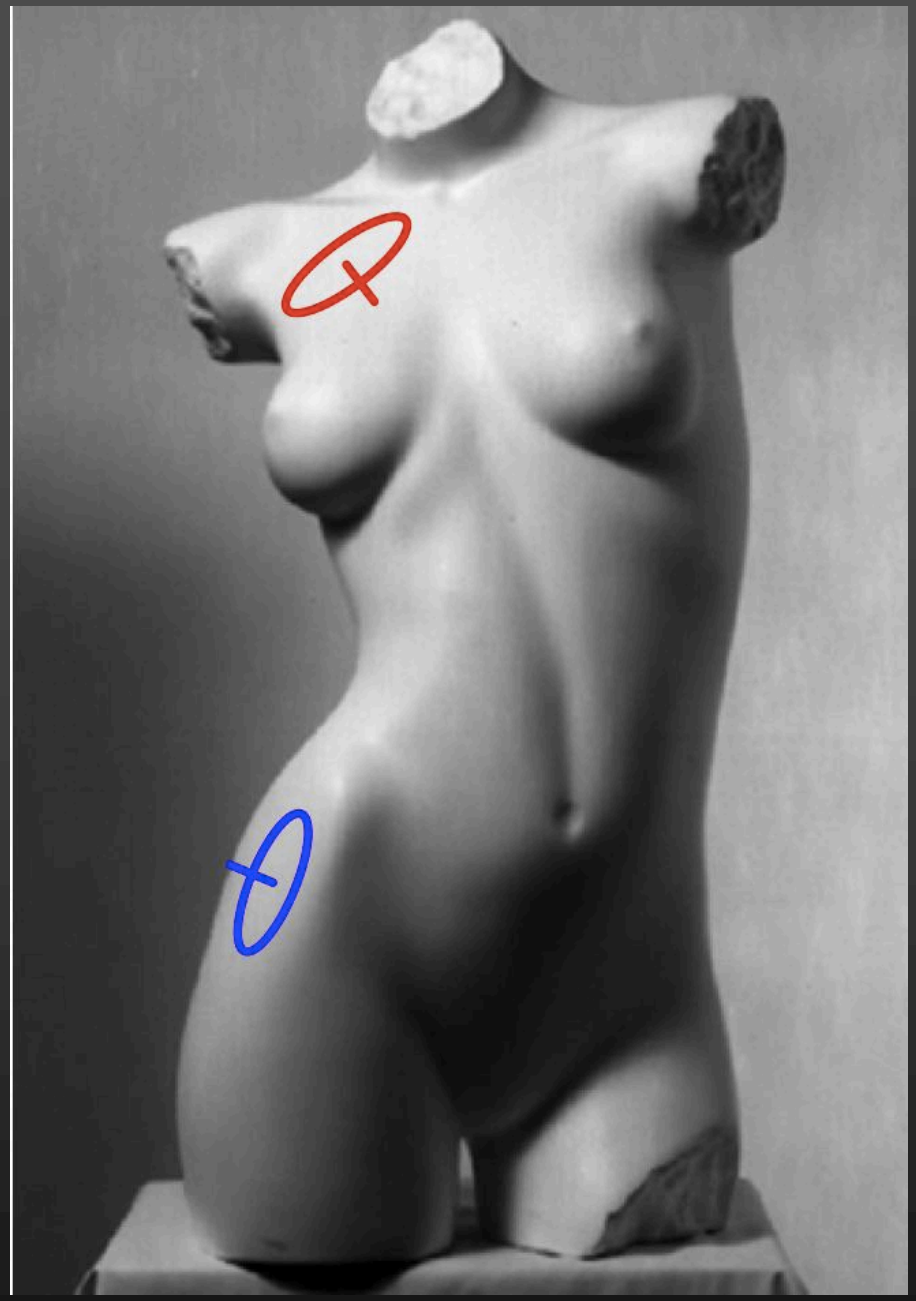
Depth from haze

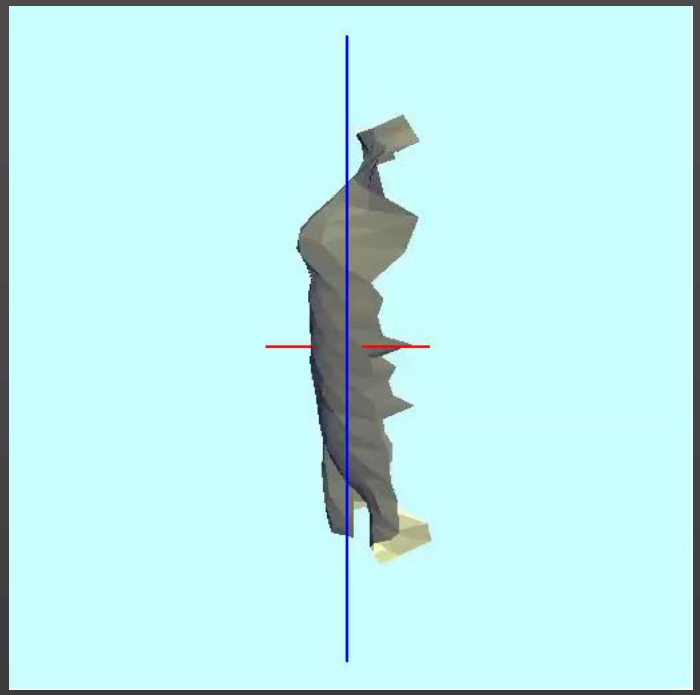
Input haze image

Reconstructed images

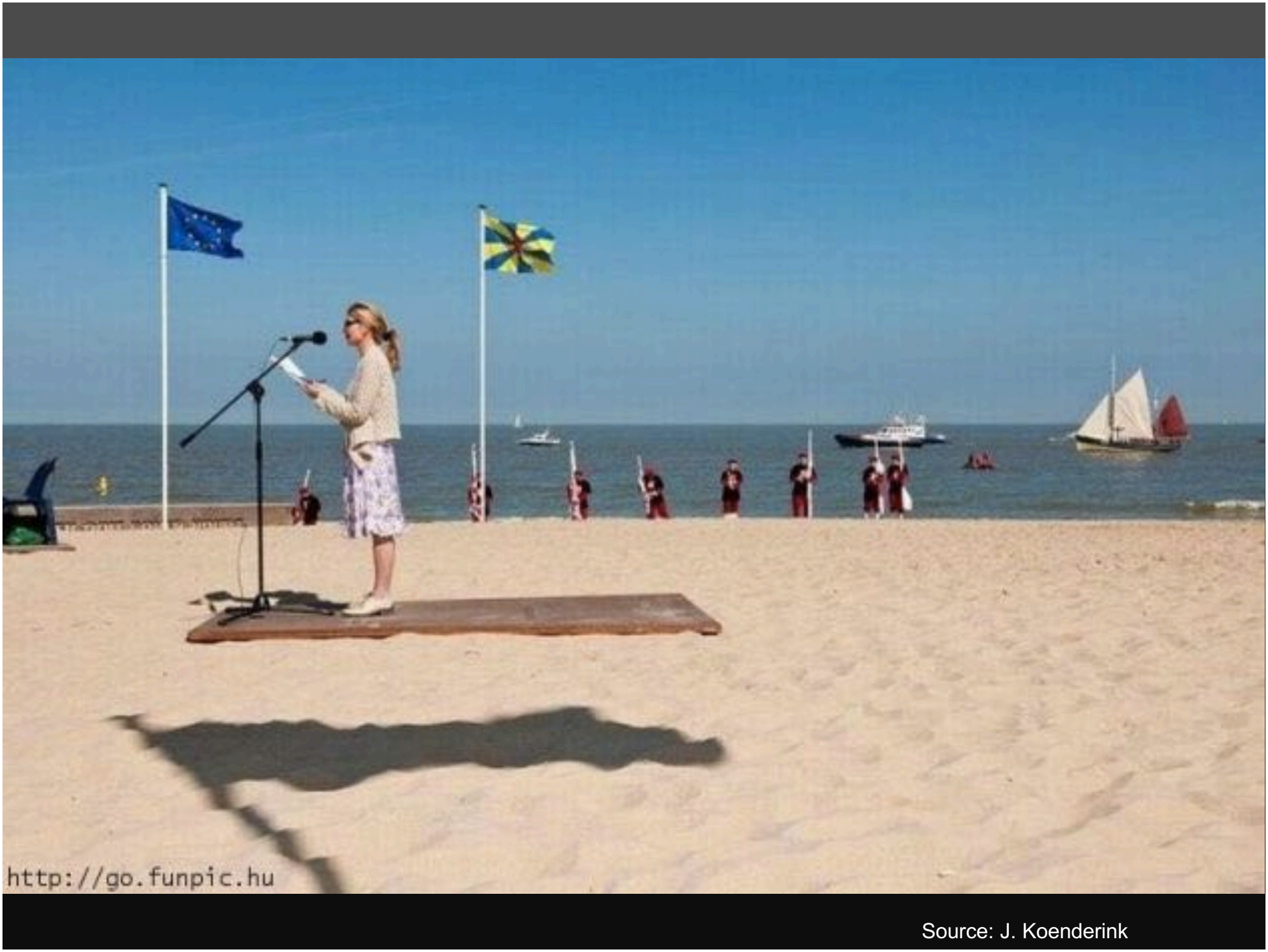
Recovered depth map

[K. HE, J. Sun and X. Tang, CVPR 2009]





Depth perception without disparity (Koenderink & Van Doorn, 1995)



Source: J. Koenderink

Challenges or opportunities?

Image source: J. Koenderink

- Images are confusing, but they also reveal the structure of the world through numerous cues.
- Our job is to interpret the cues!

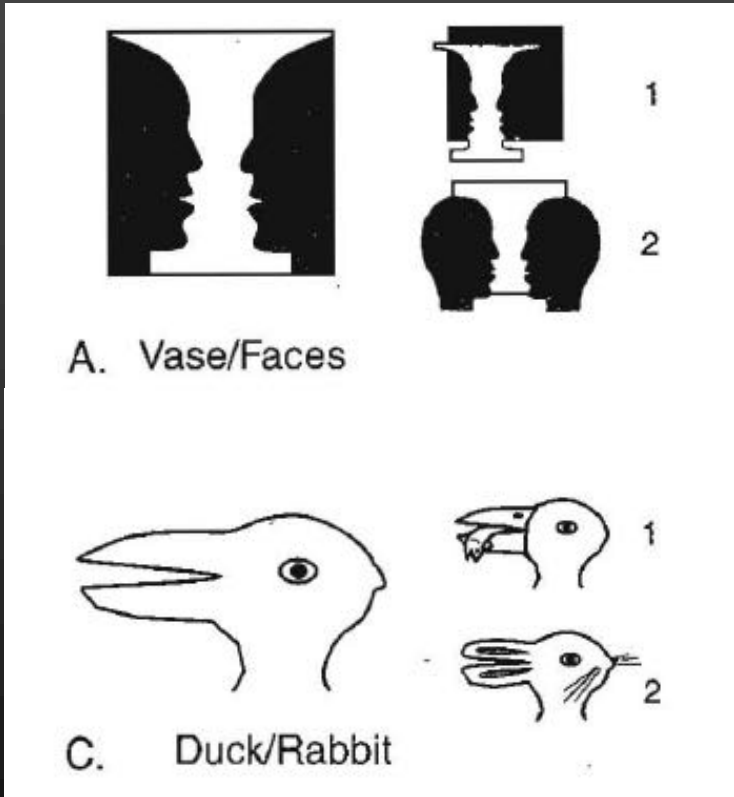
Vision is more than 3D, e.g. scene analysis

The true output of a digital camera

(Courtesy of Ivan Laptev)

Fundamental problem II: Cameras do not measure semantics

- We need lots of prior knowledge to make meaningful interpretations of an image



Outline

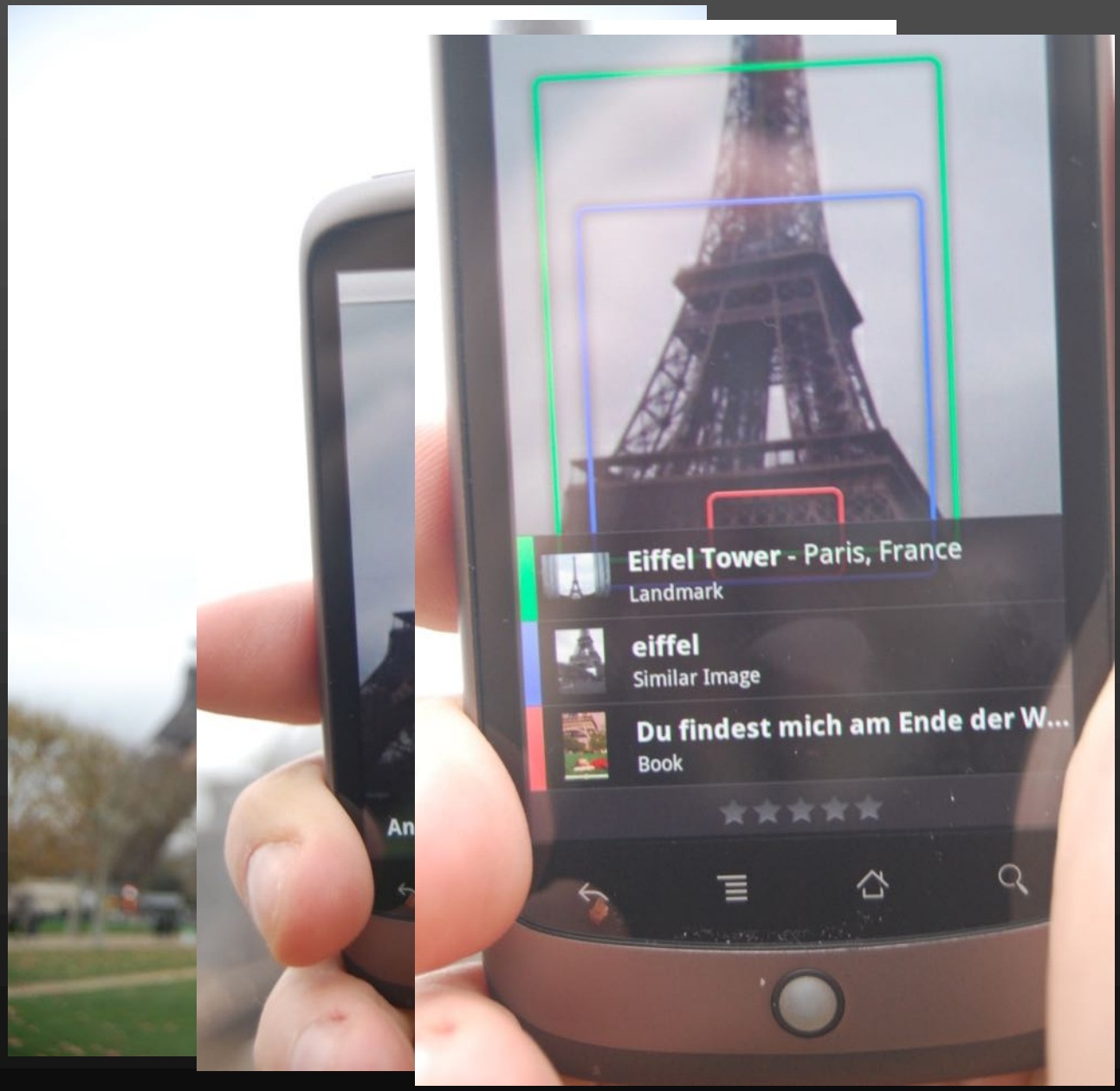
- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry

Course outline

1. **Instance-level recognition**
   - Camera geometry and image alignment
   - Local invariant features
   - Large-scale visual search

2. **Category-level recognition**
   - Bags-of-features
   - Sparse coding and dictionary learning
   - Convolutional neural networks

3. **Advanced topics**
   - Motion and human actions
   - 3D object recognition
   - Weakly-supervised learning


Eiffel Tower - Paris, France
Landmark
eiffel
Similar Image
Du findest mich am Ende der W...
Book

Specific object detection

(Lowe, 2004)

Lecture	**Date**	**Topic and reading materials.**	**Slides**
1	Oct 3	**Introduction; Camera geometry (3hrs, J. Sivic)**	
2	Oct 10	**Instance-level recognition I. - Local invariant features, correspondence, image matching (3hrs, J. Sivic);**	
3	Oct 17	**Instance-level recognition II. - Efficient visual search (1.5hrs, C. Schmid)** **Bag-of-feature models for category-level recognition (1.5hrs, C. Schmid)**	
4	Oct 24	**ICCV 2017. No lecture.** Assignments: Assignment 1 due.	
5	Oct 31	**Sparse coding and dictionary learning for image analysis (3hrs, J. Ponce)**	
6	Nov 7	**Neural networks; Optimization methods (3hrs, A. Joulin)** Assignments: Assignment 2 due.	
7	Nov 14	**Convolutional neural networks for visual recognition I. (I. Laptev)** Final project topics are out. Due date for project proposals: Nov 28.	
8	Nov 21	**Convolutional neural networks for visual recognition II. (J. Sivic)** Assignments: Assignment 3 due.	
9	Nov 28	**Motion and human actions I. (C. Schmid)** Assignments: Final project proposal due.	
10	Dec 5	**Human pose estimation; Weakly-supervised learning I (I. Laptev)**	
11	Dec 12	**3D object recognition and Convolutional neural networks (M. Aubry)** **Weakly-supervised learning II (I. Laptev)**	
12	**Jan 15** **Jan 16** **Jan 17**	**Final project presentations and evaluation (I. Laptev, J. Sivic)** **Jan 15: 13:00-17:00** **Jan 16: 13:00-17:00** **Jan 17: 13:00-17:00** The presentations will take place at **Salle Alan Turing - 1st floor at Inria Paris research center, 2 Rue Simone Iff, 75012, Paris. Directions are here.** When you enter the building tell the receptionist you are going for the presentation and go directly to the first floor (no special access card is needed)	

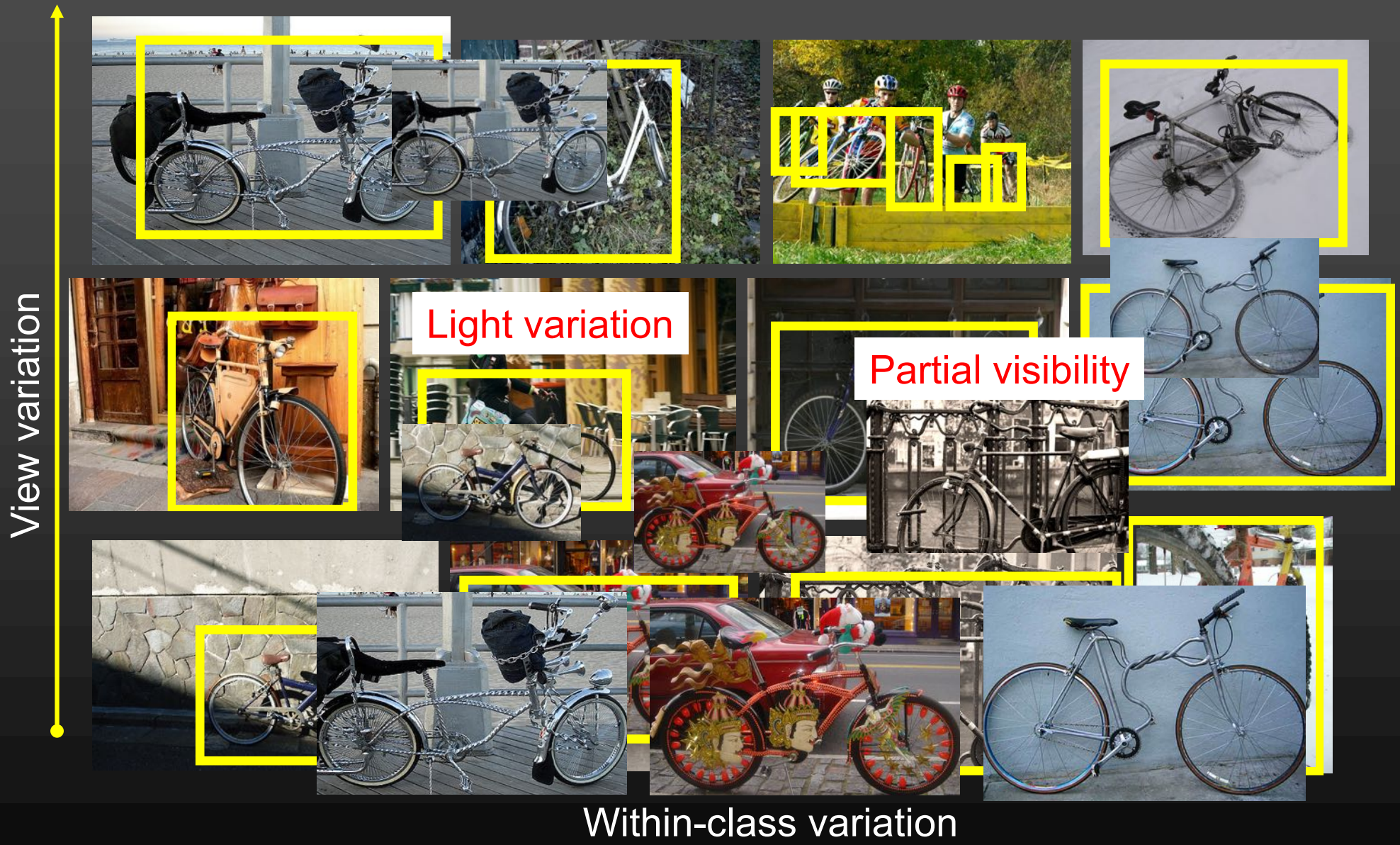
Course outline

1. **Instance-level recognition**
   - Camera geometry and image alignment
   - Local invariant features
   - Large-scale visual search

2. **Category-level recognition**
   - Bags-of-features
   - Sparse coding and dictionary learning
   - Convolutional neural networks

3. **Advanced topics**
   - Motion and human actions
   - 3D object recognition
   - Weakly-supervised learning

Category-level recognition



Convolutional neural networks (CNNs)

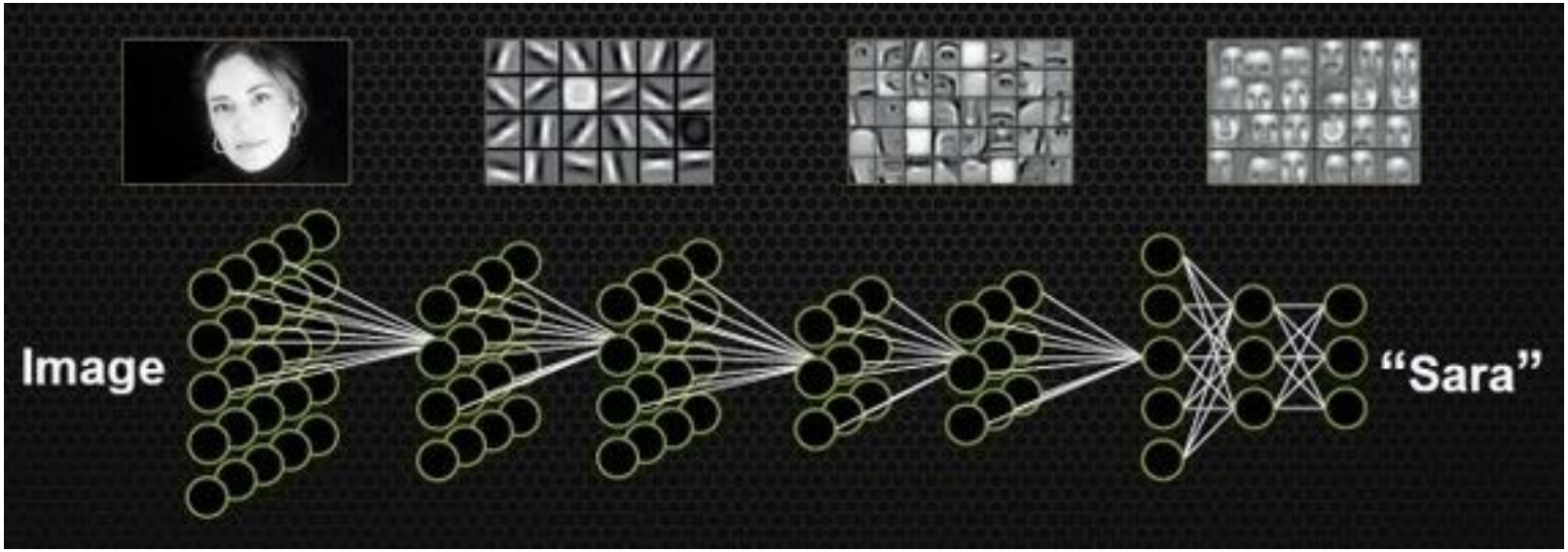
Multi-layer nested image representation

Output representation

Input image

$$\hat{z} = f^n(\ldots f^2(f^1(x))\ldots)$$

"Layer" n "Layer" 2 "Layer" 1



Source:
A. Shivkumar

Guest lecture by A. Joulin on neural networks and optimization.

[Rosenblatt'57], [Hubel&Wiesel'59], [Fukushima'80], [Rumelhart'86], [LeCun et al.'89], [LeCun et al.'98], [Hinton&Salakhutdinov'06], [Krizhevsky'12], …

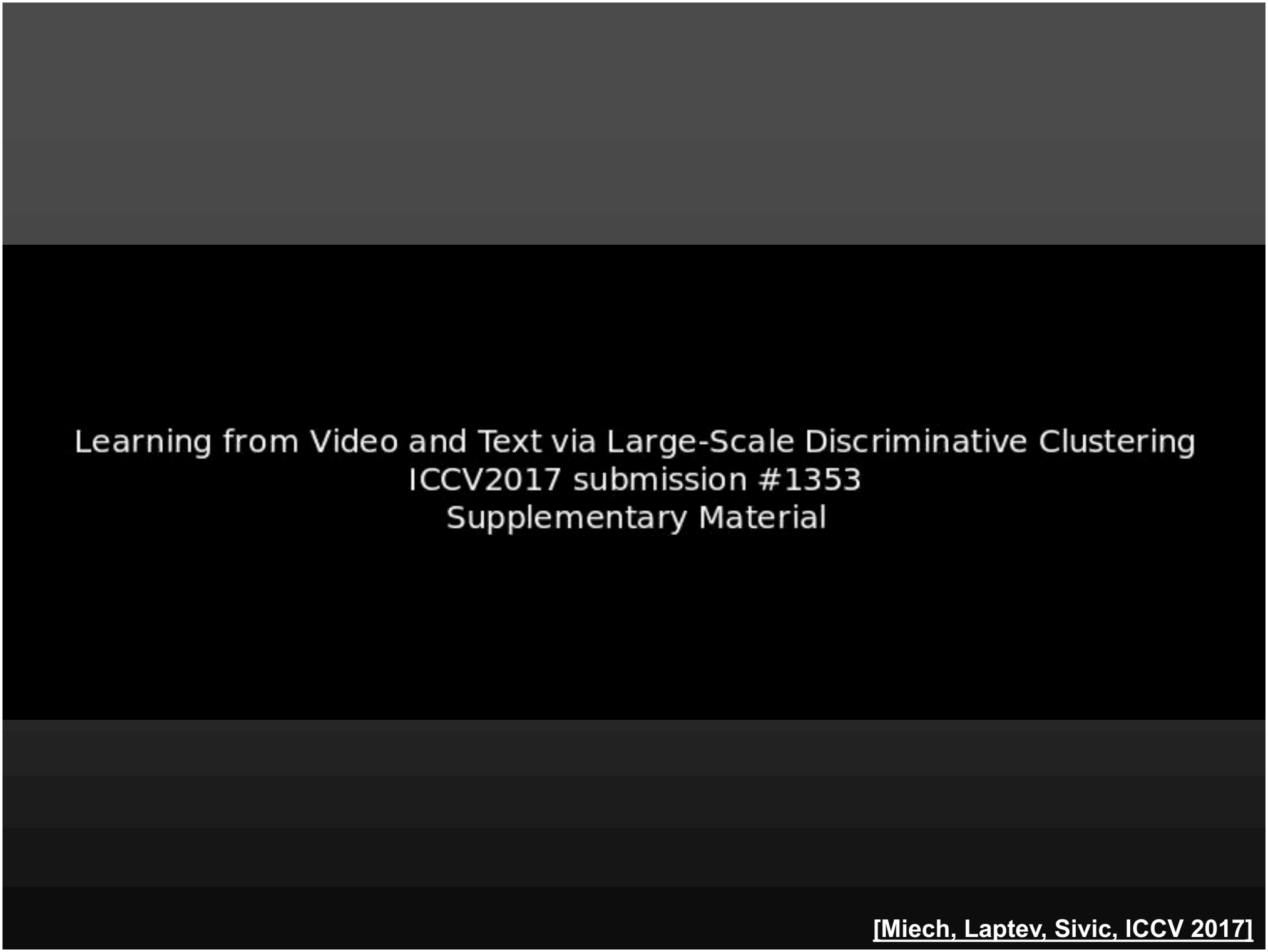
Lecture	**Date**	**Topic and reading materials.**	**Slides**
1	Oct 3	**Introduction; Camera geometry (3hrs, J. Sivic)**	
2	Oct 10	**Instance-level recognition I. - Local invariant features, correspondence, image matching (3hrs, J. Sivic);**	
3	Oct 17	**Instance-level recognition II. - Efficient visual search (1.5hrs, C. Schmid)** **Bag-of-feature models for category-level recognition (1.5hrs, C. Schmid)**	
4	Oct 24	**ICCV 2017. No lecture.** Assignments: Assignment 1 due.	
5	Oct 31	**Sparse coding and dictionary learning for image analysis (3hrs, J. Ponce)**	
6	Nov 7	**Neural networks; Optimization methods (3hrs, A. Joulin)** Assignments: Assignment 2 due.	
7	Nov 14	**Convolutional neural networks for visual recognition I. (I. Laptev)** Final project topics are out. Due date for project proposals: Nov 28.	
8	Nov 21	**Convolutional neural networks for visual recognition II. (J. Sivic)** Assignments: Assignment 3 due.	
9	Nov 28	**Motion and human actions I. (C. Schmid)** Assignments: Final project proposal due.	
10	Dec 5	**Human pose estimation; Weakly-supervised learning I (I. Laptev)**	
11	Dec 12	**3D object recognition and Convolutional neural networks (M. Aubry)** **Weakly-supervised learning II (I. Laptev)**	
12	**Jan 15** **Jan 16** **Jan 17**	**Final project presentations and evaluation (I. Laptev, J. Sivic)** **Jan 15: 13:00-17:00** **Jan 16: 13:00-17:00** **Jan 17: 13:00-17:00** The presentations will take place at **Salle Alan Turing - 1st floor at Inria Paris research center, 2 Rue Simone Iff, 75012, Paris. Directions are here.** When you enter the building tell the receptionist you are going for the presentation and go	

Course outline

1. **Instance-level recognition**
   - Camera geometry and image alignment
   - Local invariant features
   - Large-scale visual search

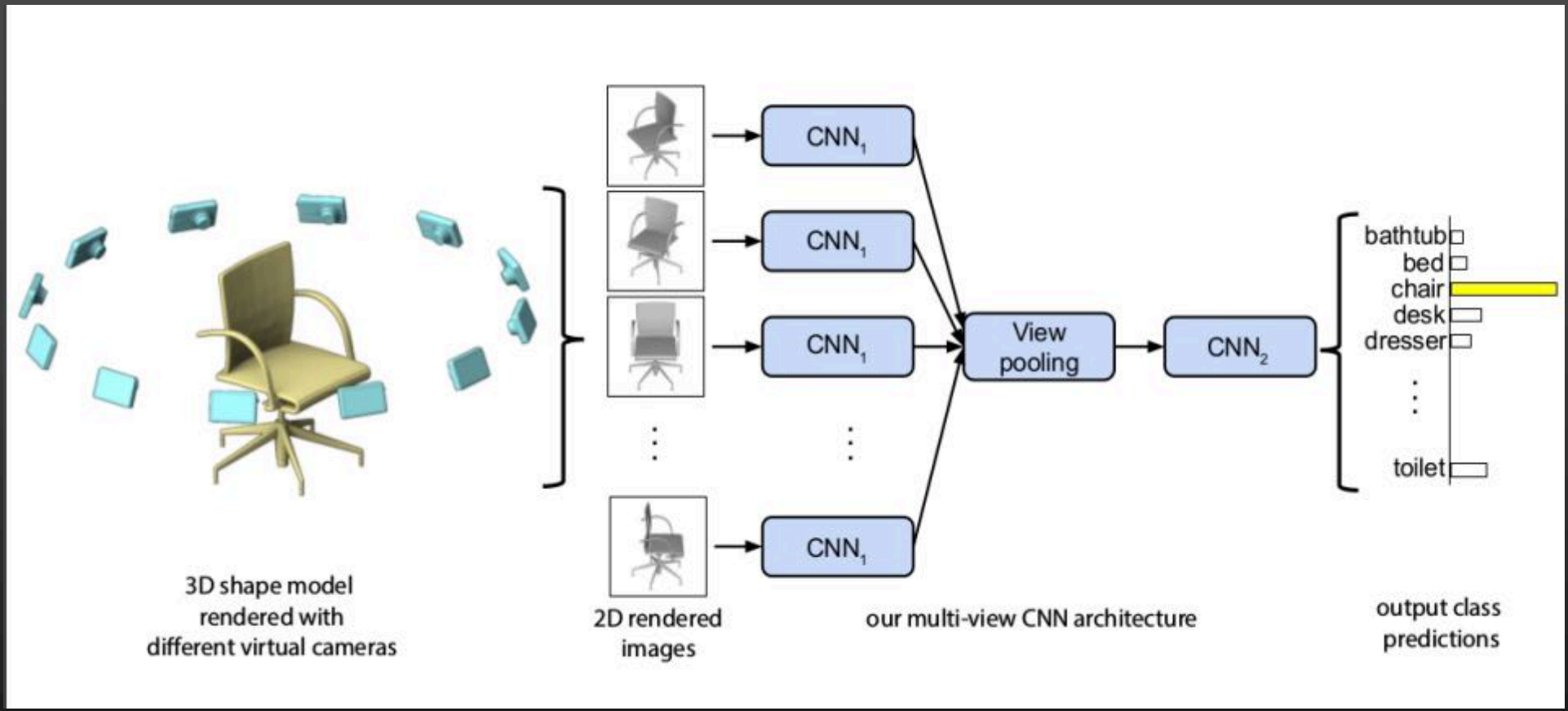
2. **Category-level recognition**
   - Bags-of-features
   - Sparse coding and dictionary learning
   - Convolutional neural networks

3. **Advanced topics**
   - Motion and human actions
   - 3D object recognition
   - Weakly-supervised learning



[Miech, Laptev, Sivic, ICCV 2017]

3D object recognition



[Su et al., ICCV'15]

Weakly supervised learning:
learning from **incomplete and noisy** meta-data

"Public bikes in Warsaw during night"

[Barnard et al.'03], [Berg et al.'04], [Gupta and Davis'08], [Ordonez et al.'11], [Kulkarni et al.'11], [Karpathy et al.'14], …

Towards Scene understanding

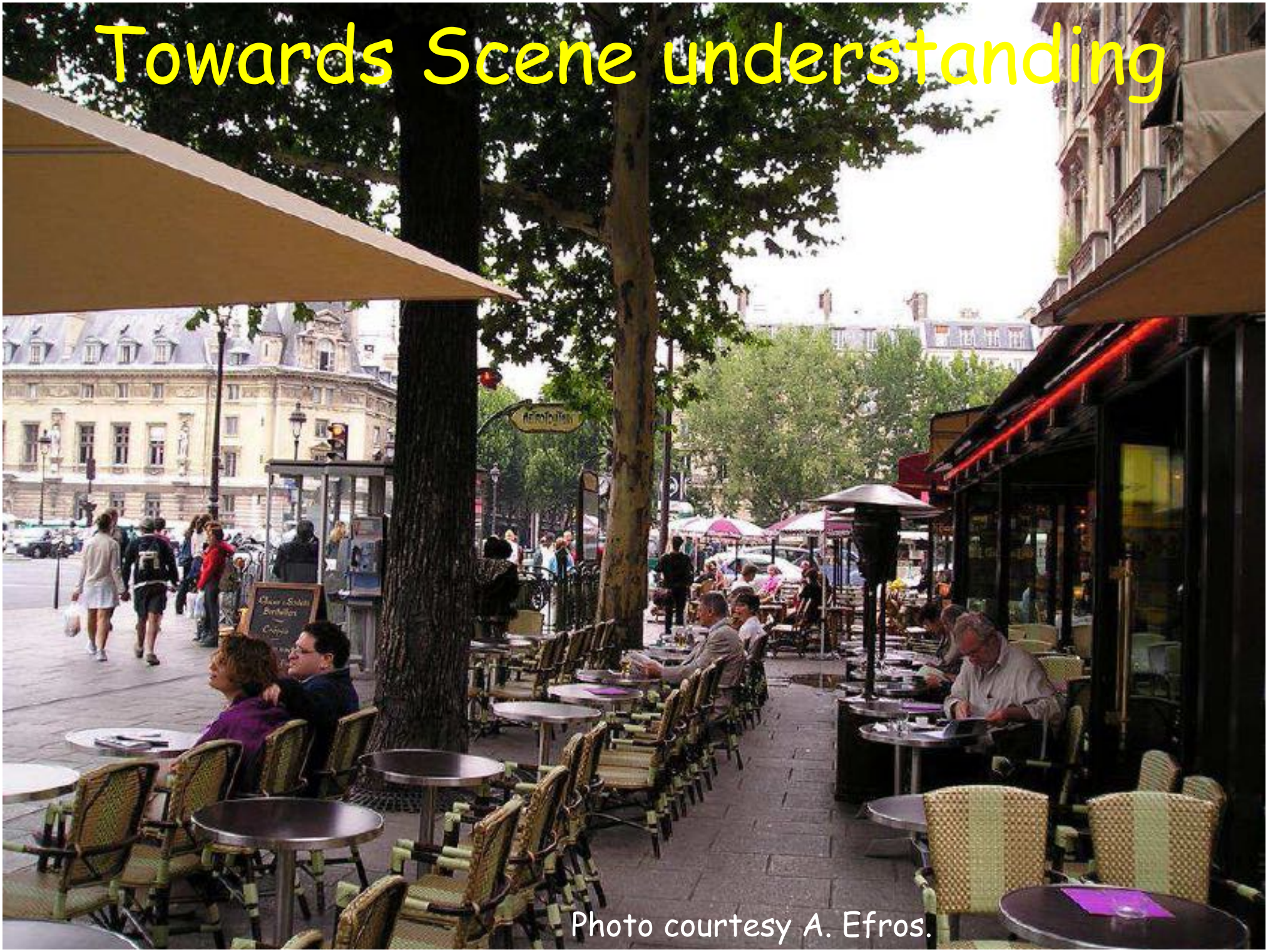
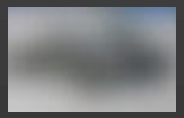
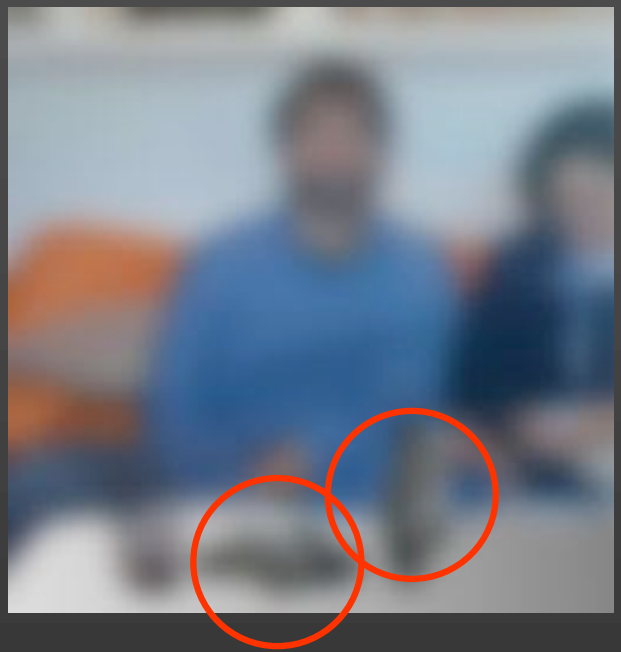
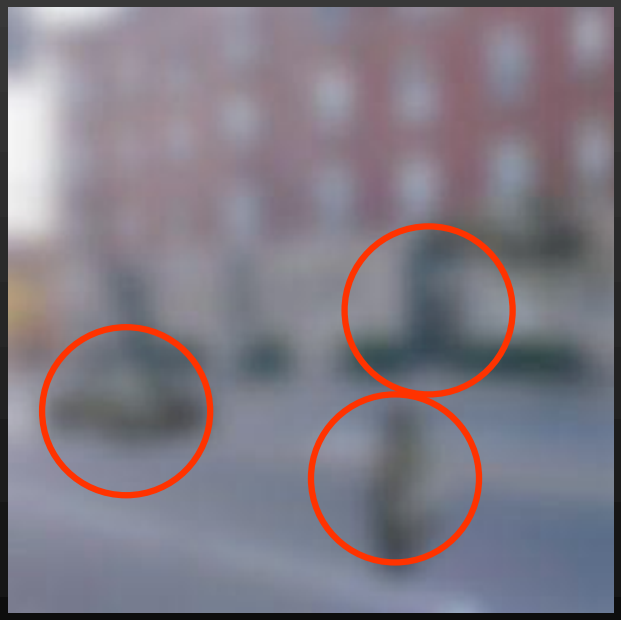


Photo courtesy A. Efros.

Local ambiguity and global scene interpretation







slide credit: Fei-Fei, Fergus & Torralba

Lecture	**Date**	**Topic and reading materials.**	**Slides**
1	Oct 3	**Introduction; Camera geometry (3hrs, J. Sivic)**	
2	Oct 10	**Instance-level recognition I. - Local invariant features, correspondence, image matching (3hrs, J. Sivic);**	
3	Oct 17	**Instance-level recognition II. - Efficient visual search (1.5hrs, C. Schmid)** **Bag-of-feature models for category-level recognition (1.5hrs, C. Schmid)**	
4	Oct 24	**ICCV 2017. No lecture.** Assignments: Assignment 1 due.	
5	Oct 31	**Sparse coding and dictionary learning for image analysis (3hrs, J. Ponce)**	
6	Nov 7	**Neural networks; Optimization methods (3hrs, A. Joulin)** Assignments: Assignment 2 due.	
7	Nov 14	**Convolutional neural networks for visual recognition I. (I. Laptev)** Final project topics are out. Due date for project proposals: Nov 28.	
8	Nov 21	**Convolutional neural networks for visual recognition II. (J. Sivic)** Assignments: Assignment 3 due.	
9	Nov 28	**Motion and human actions I. (C. Schmid)** Assignments: Final project proposal due.	
10	Dec 5	**Human pose estimation; Weakly-supervised learning I (I. Laptev)**	
11	Dec 12	**3D object recognition and Convolutional neural networks (M. Aubry)** **Weakly-supervised learning II (I. Laptev)**	
12	**Jan 15** **Jan 16** **Jan 17**	**Final project presentations and evaluation (I. Laptev, J. Sivic)** **Jan 15: 13:00-17:00** **Jan 16: 13:00-17:00** **Jan 17: 13:00-17:00** The presentations will take place at **Salle Alan Turing - 1st floor at Inria Paris research center, 2 Rue Simone Iff, 75012, Paris. Directions are here.** When you enter the building tell the receptionist you are going for the presentation and go	

Books for this class

- D.A. Forsyth and J. Ponce, "Computer Vision: A Modern Approach", Prentice-Hall/Pearson, 2nd edition, 2011.
- J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, "Toward category-level object recognition", Springer LNCS, 2007.
- R. Szeliski, "Computer Vision: Algorithms and Applications", Springer, 2010.
- C. Bishop: Pattern Recognition and Machine Learning, 2006

R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2004.

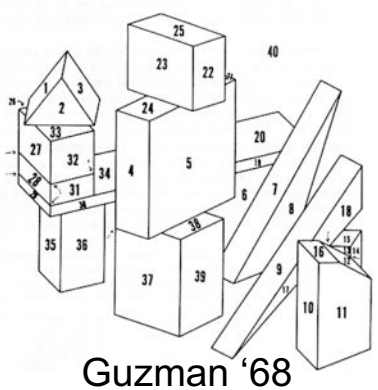
- J.J. Koenderink, "Solid Shape", MIT Press, 1990.
- J.J. Koenderink, http://www.gestaltrevision.be/en/resources/clootcrans-press

Outline

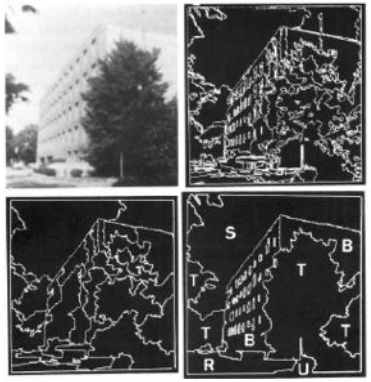
- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry

A very brief history of computer vision

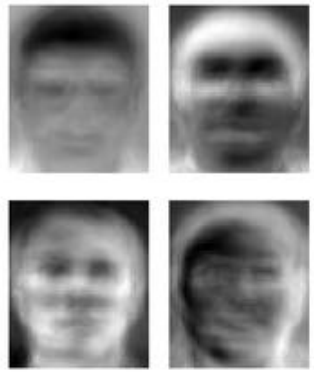
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: Deep learning with ConvNets
- 2030's: …



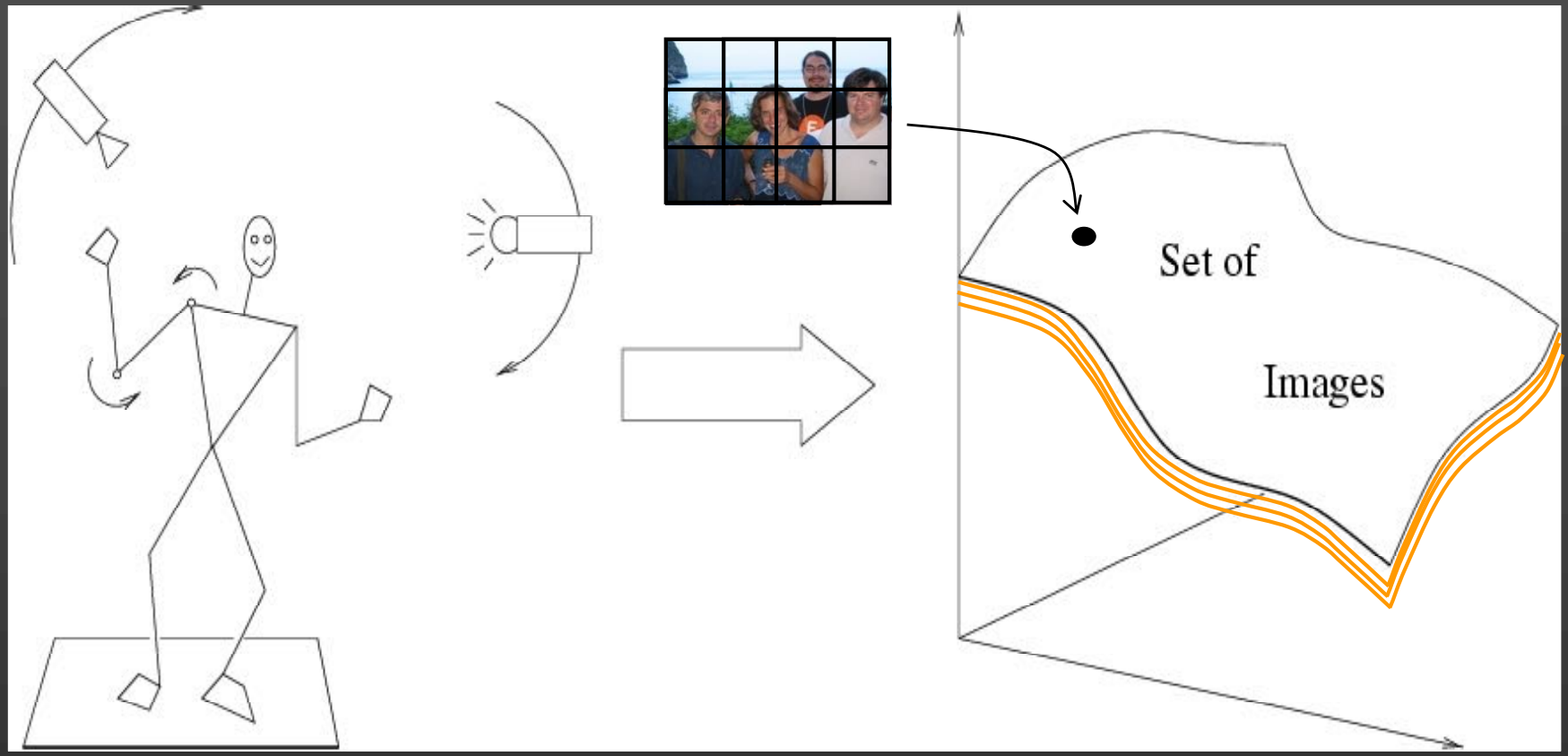
Guzman '68



Ohta Kanade '78

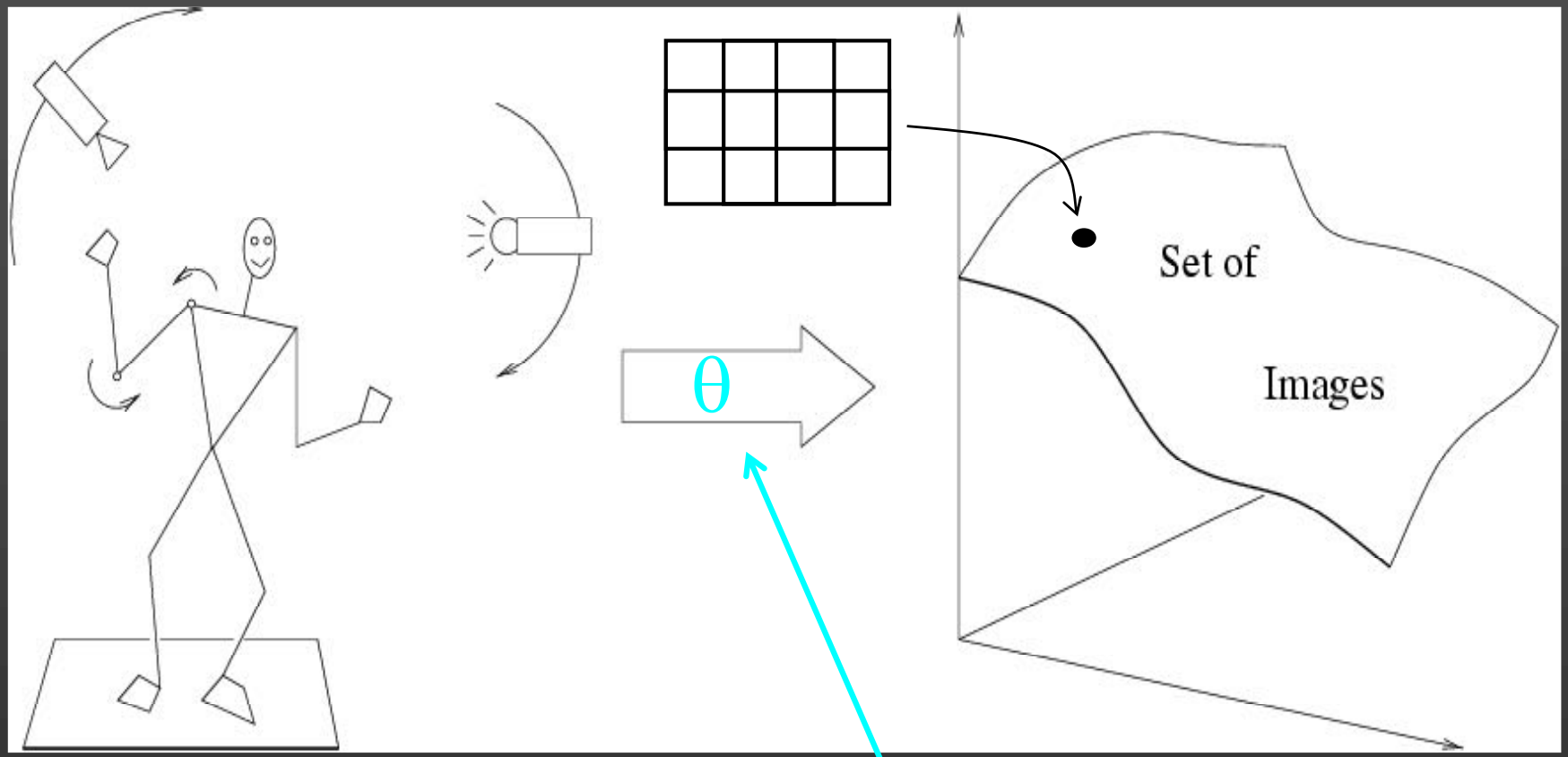


Turk and Pentland '91



Variability:

Camera position
Illumination
Internal parameters
Within-class variations

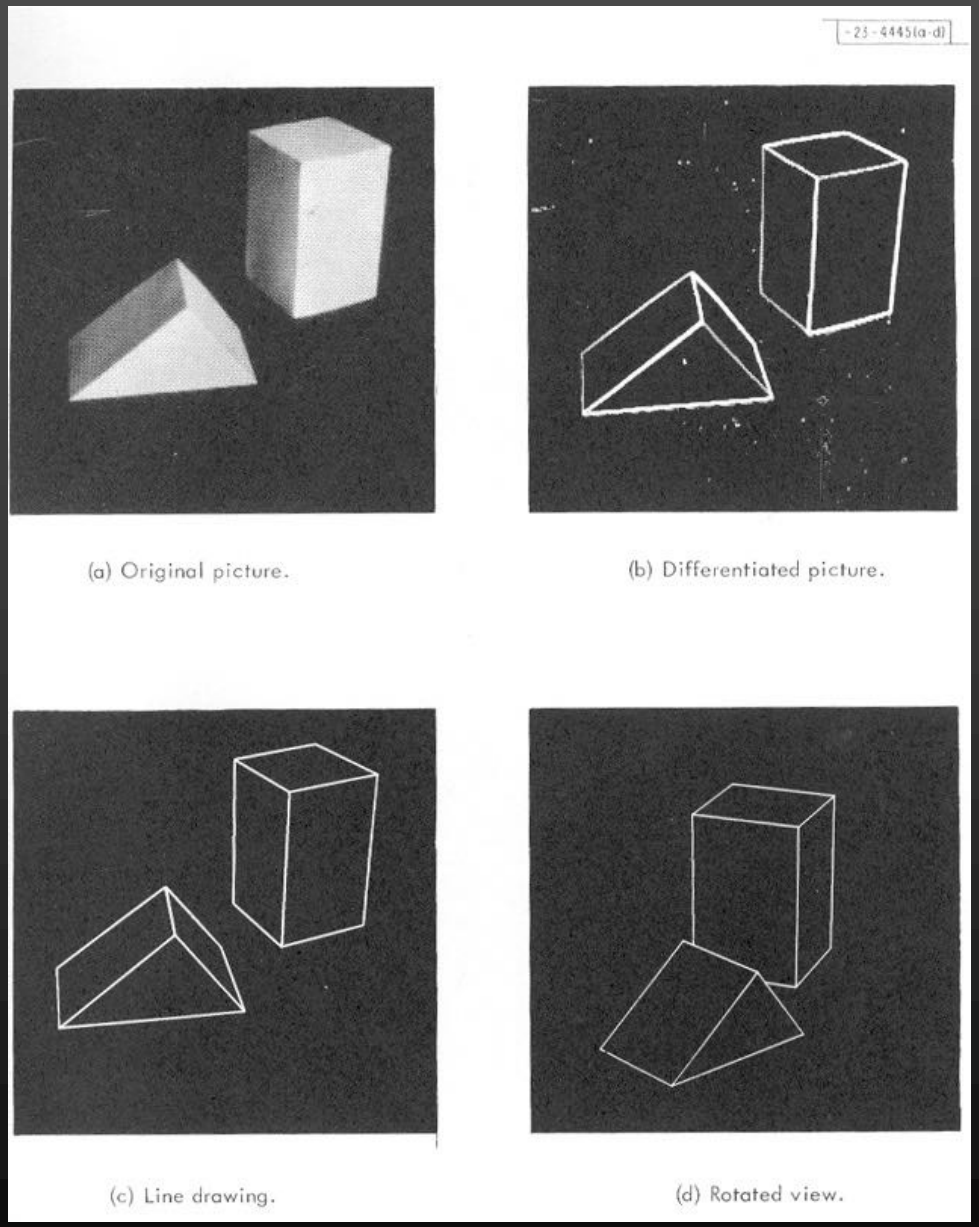


Variability:

Camera position
Illumination
Internal parameters

Roberts (1963); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)

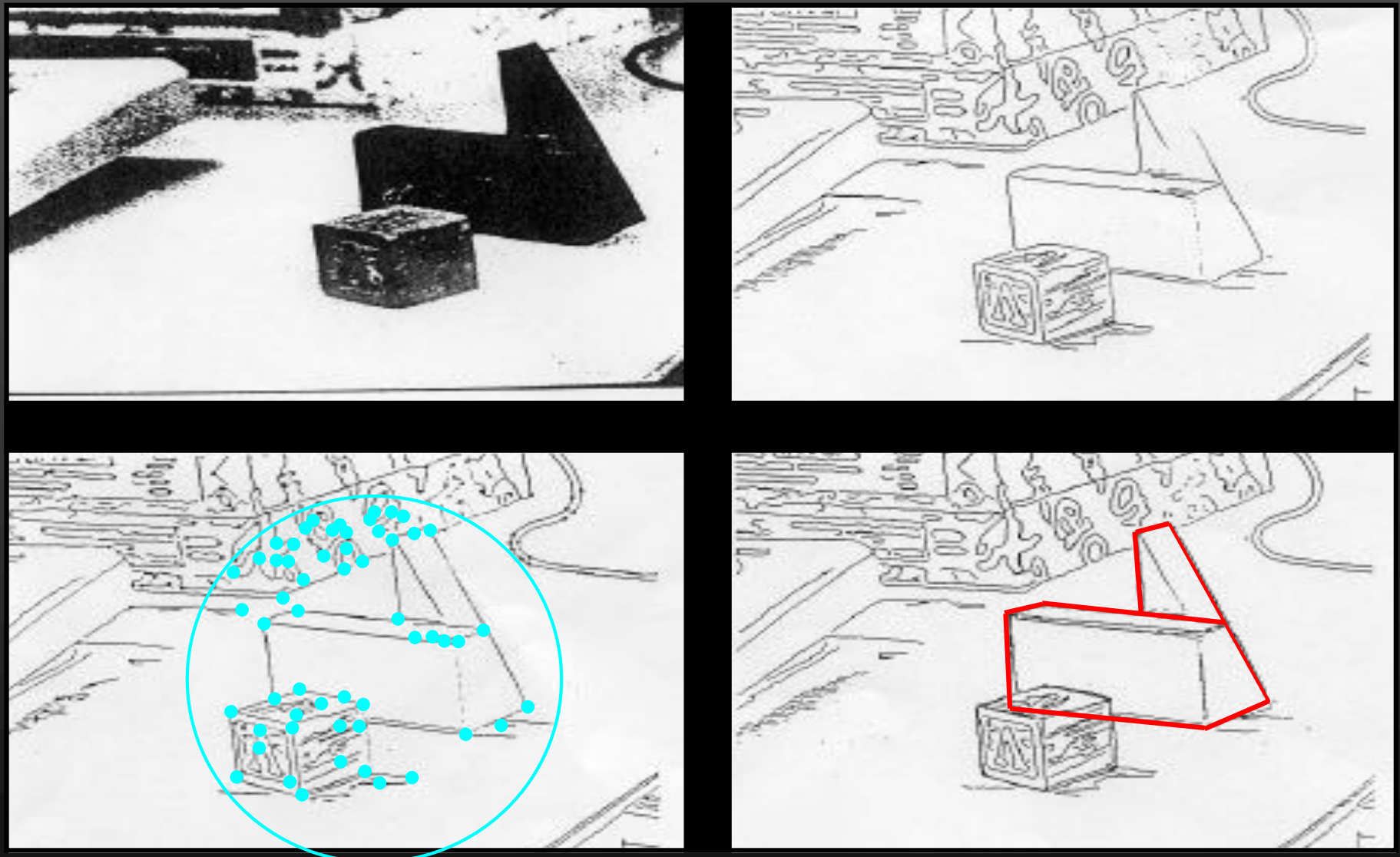
Origins of computer vision

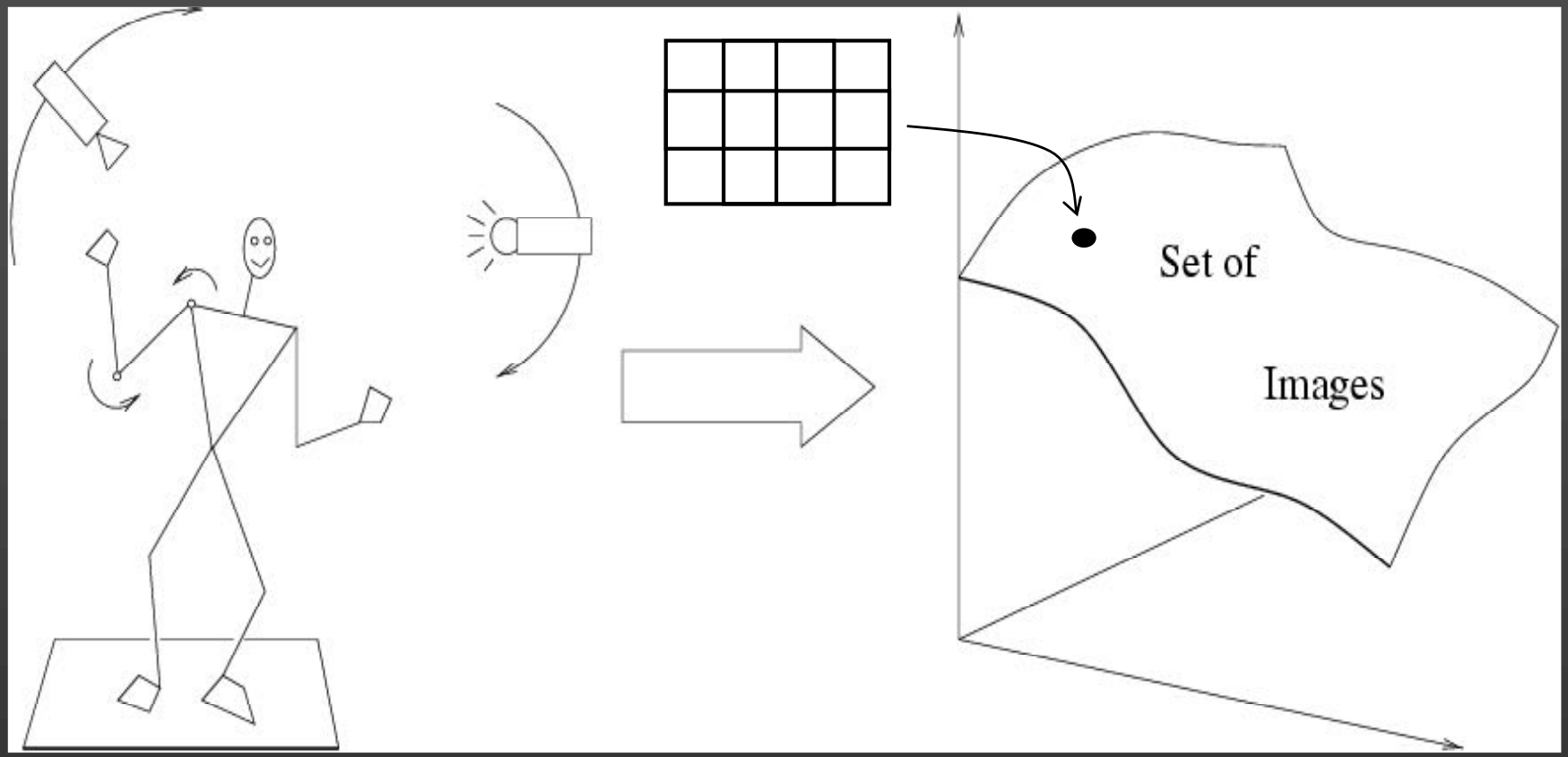


L. G. Roberts, *Machine Perception of Three Dimensional Solids,* Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

photo credit: Joe Mundy

Huttenlocher & Ullman (1987)





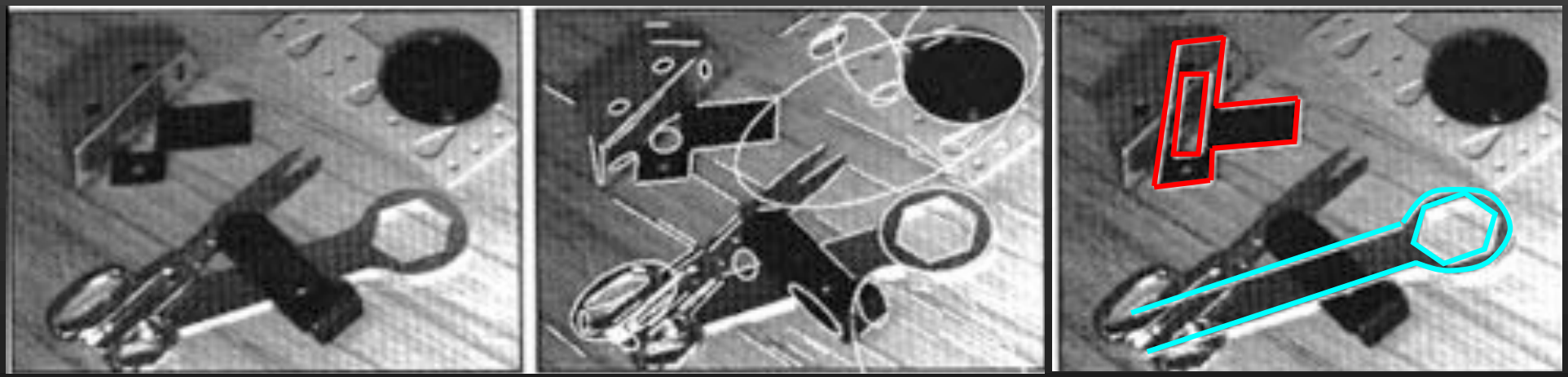
~~Variability~~ Invariance to:

- Camera position
- Illumination
- Internal parameters

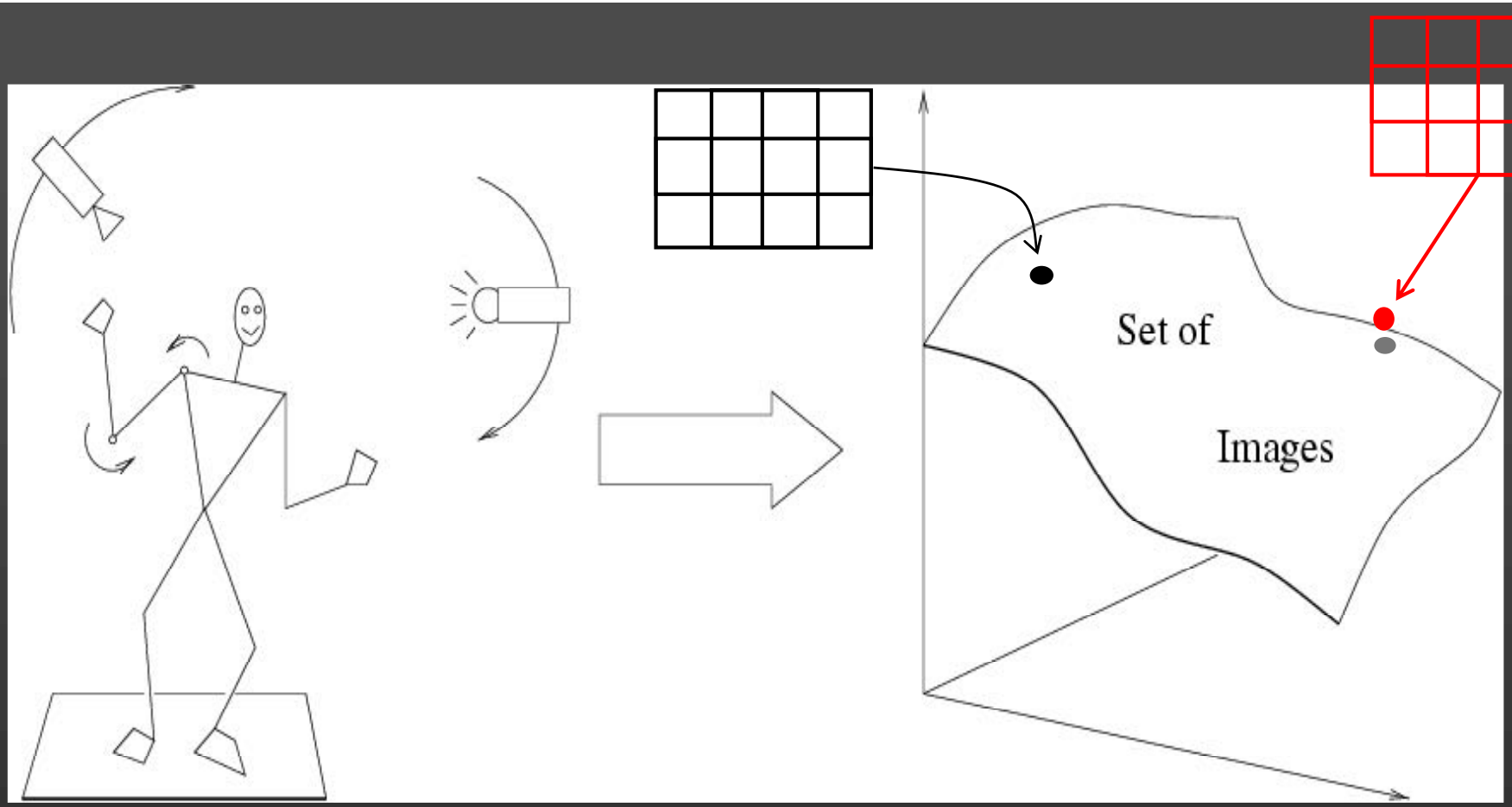
Duda & Hart (1972); Weiss (1987); Mundy et al. (1992-94); Rothwell et al. (1992); Burns et al. (1993)

Example: rotation invariants

Projective invariants (Rothwell et al., 1992):



BUT: True 3D objects do not admit monocular viewpoint invariants (Burns et al., 1993) !!



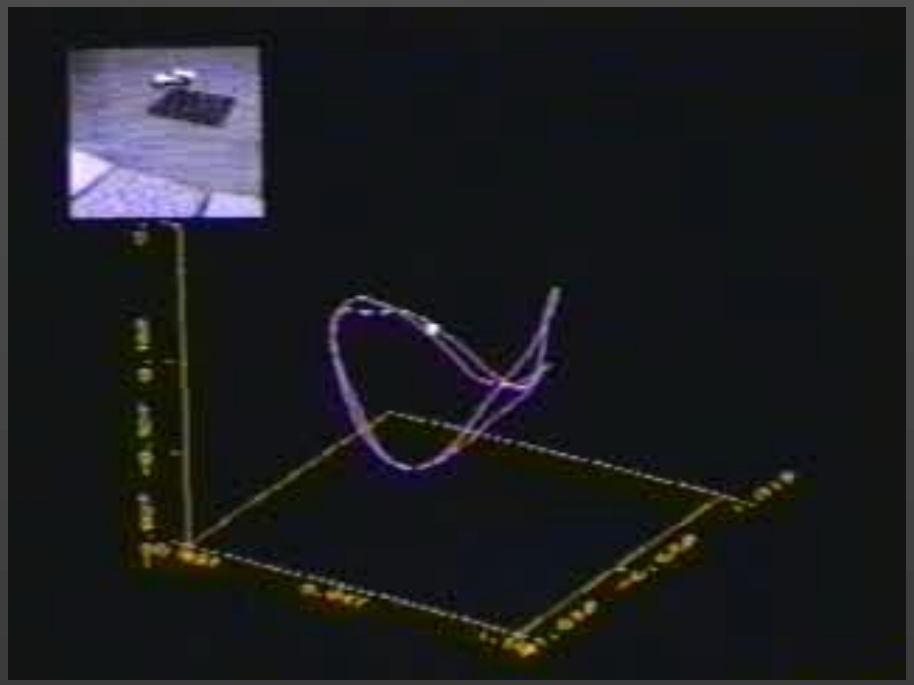
Empirical models of image variability:

Appearance-based techniques

Turk & Pentland (1991); Murase & Nayar (1995); etc.

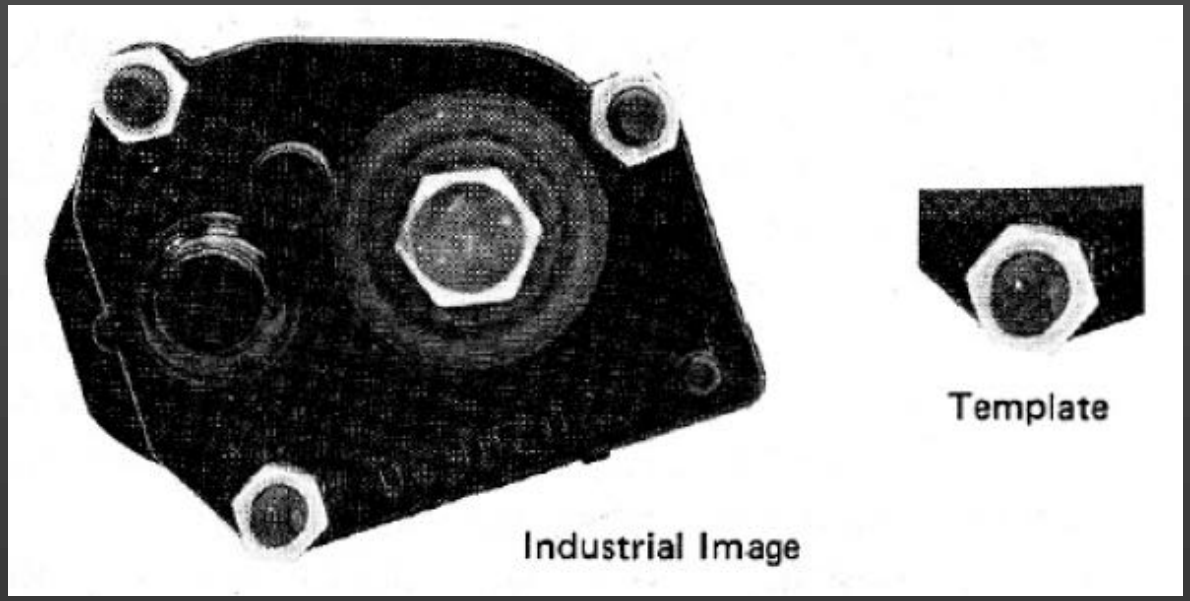
Eigenfaces (Turk & Pentland, 1991)

Experimental	Correct/Unknown Recognition Percentage		
Condition	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20



Appearance manifolds
(Murase & Nayar, 1995)

Correlation-based template matching (60s)



Ballard & Brown (1980, Fig. 3.3). Courtesy Bob Fisher and Ballard & Brown on-line.

- Automated target recognition
- Industrial inspection
- Optical character recognition
- Stereo matching
- Pattern recognition

In the late 1990s, a new approach emerges:
Combining local appearance, spatial constraints, invariants, and classification techniques from machine learning.

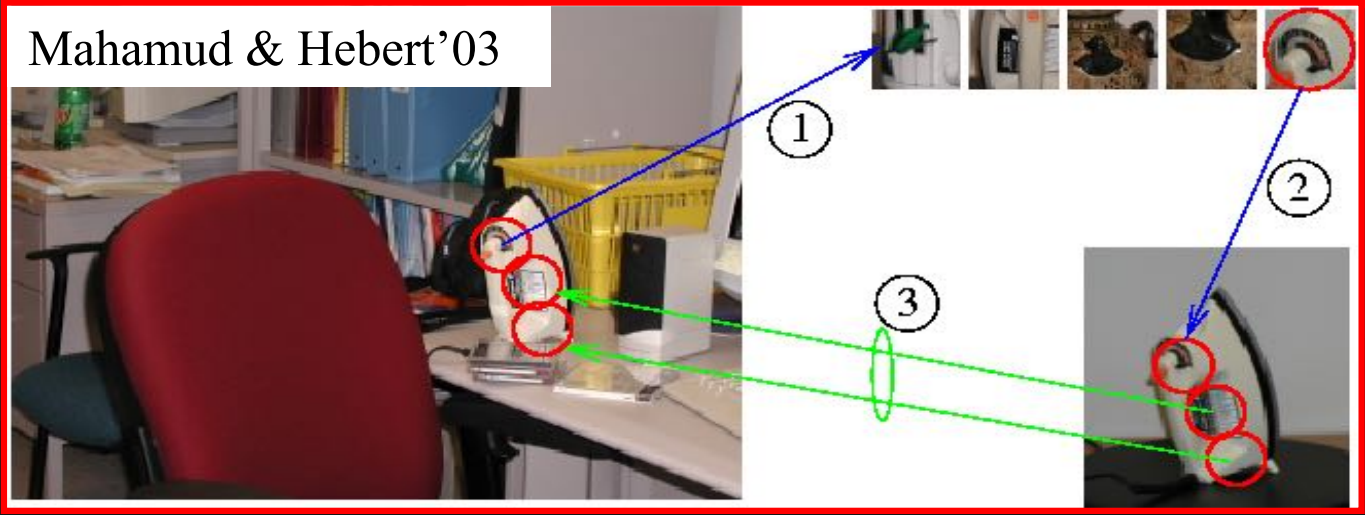
Query

Retrieved (10° off)

Schmid & Mohr'97

Lowe'02

Mahamud & Hebert'03



Late 1990s: Local appearance models

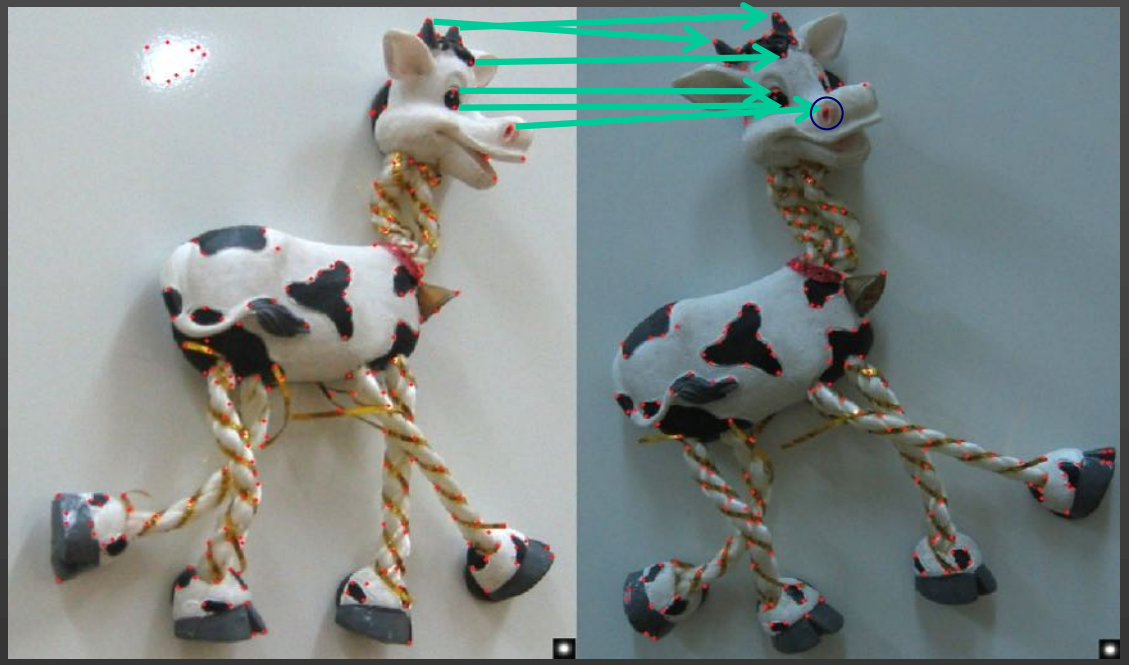
(Image courtesy of C. Schmid)

Late 1990s: Local appearance models

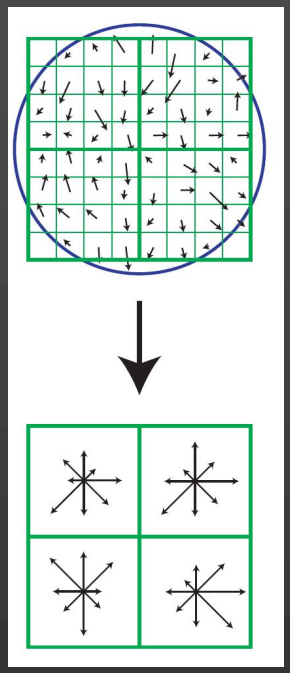
(Image courtesy of C. Schmid)

- Find features (interest points).

Late 1990s: Local appearance models



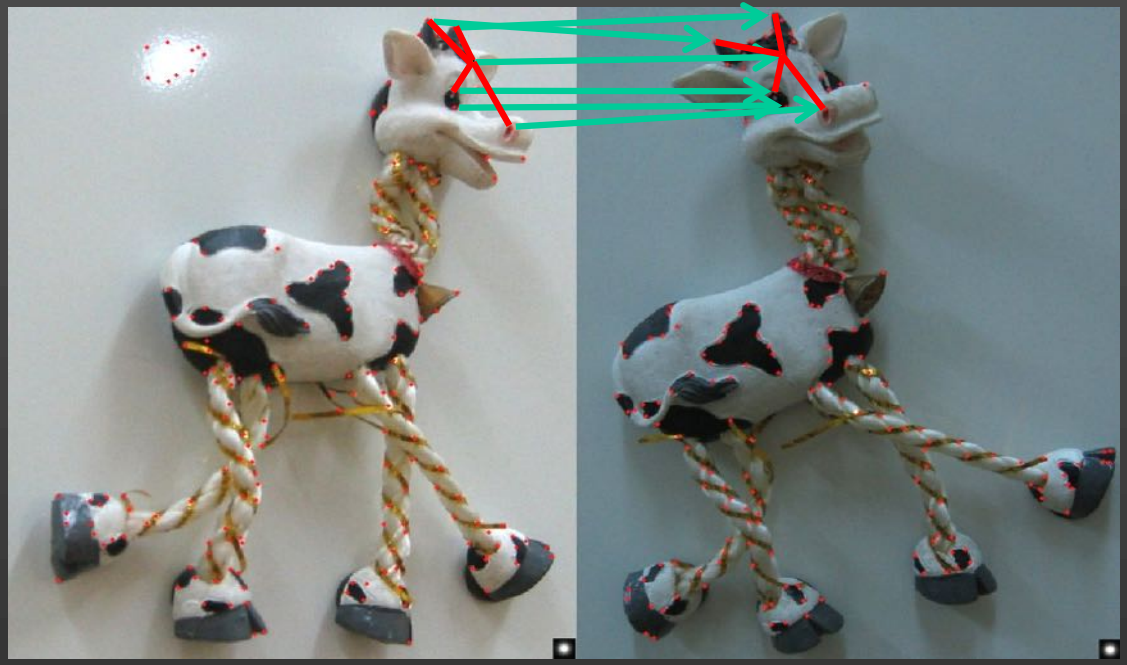
(Image courtesy of C. Schmid)



(Lowe 2004)

- Find features (interest points).
- Match them using local invariant descriptors (jets, SIFT).

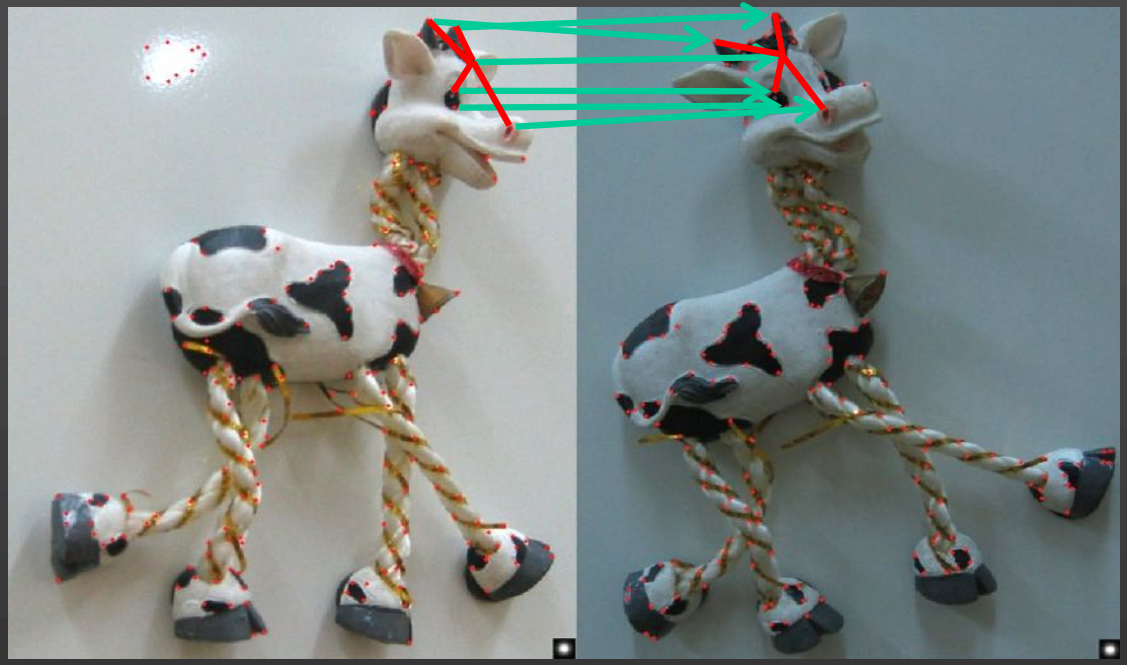
Late 1990s: Local appearance models



(Image courtesy of C. Schmid)

- Find features (interest points).
- Match them using local invariant descriptors (jets, SIFT).
- Optional: Filter out outliers using geometric consistency.

Late 1990s: Local appearance models



(Image courtesy of C. Schmid)

- Find features (interest points).
- Match them using local invariant descriptors (jets, SIFT).
- Optional: Filter out outliers using geometric consistency.
- Vote.

See, for example, Schmid & Mohr (1996); Lowe (1999);Tuytelaars & Van Gool, (2002); Rothganger et al. (2003); Ferrari et al., (2004).

Bags of words: Visual "Google"

(Sivic & Zisserman, ICCV' 03)

"Visual word" clusters

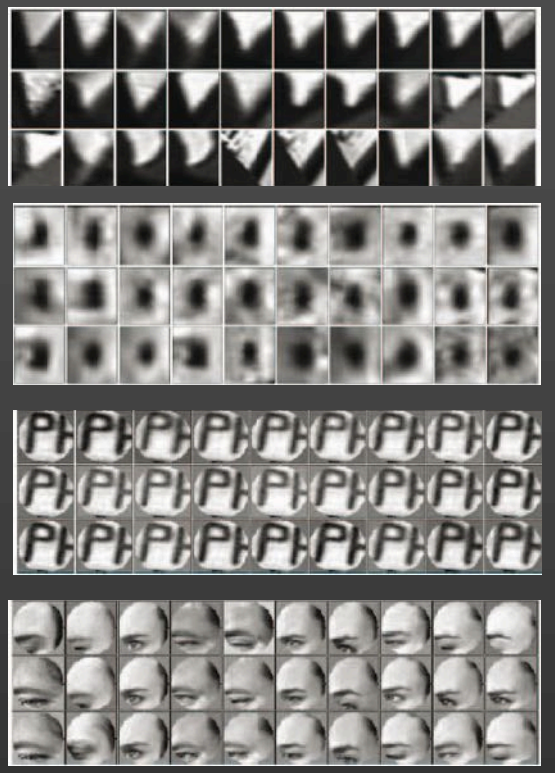
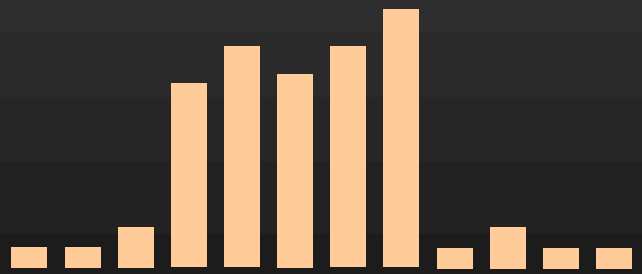


Image retrieval in videos



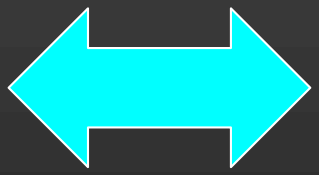
Vector quantization into histogram (the "bag of words")

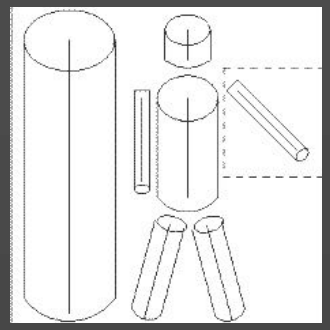
Bags of words: Visual "Google"
(Sivic & Zisserman, ICCV' 03)

Retrieved shots

Select a region

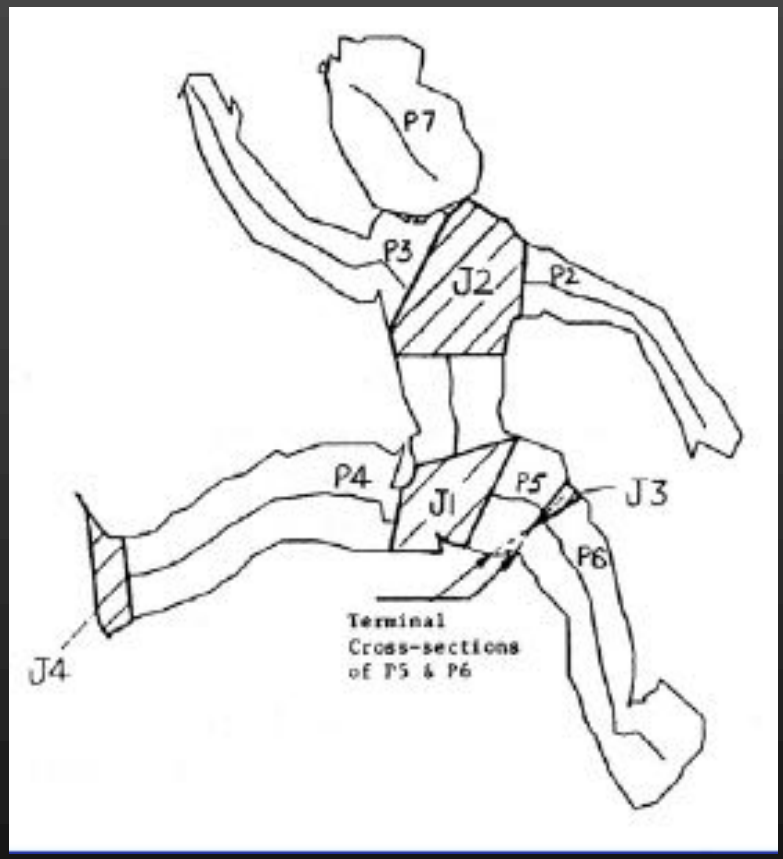
Image categorization is harder





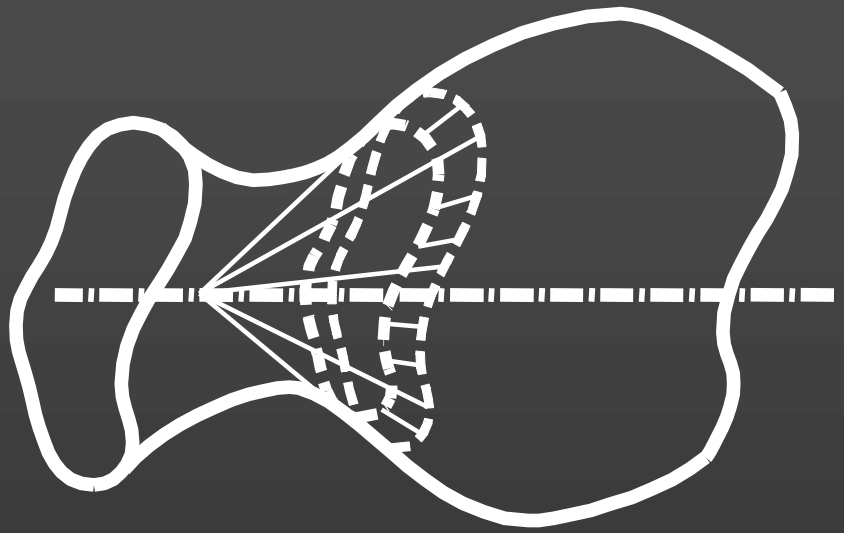
Structural part-based models

(Binford, 1971; Marr & Nishihara, 1978)

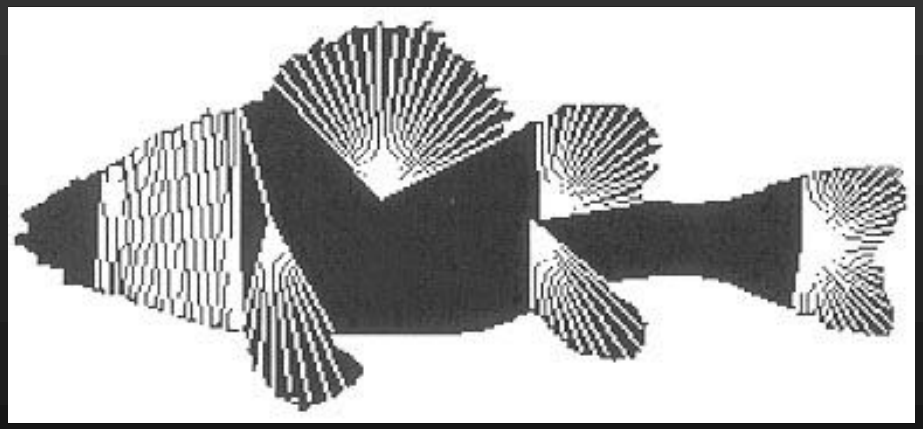


(Nevatia & Binford, 1972)

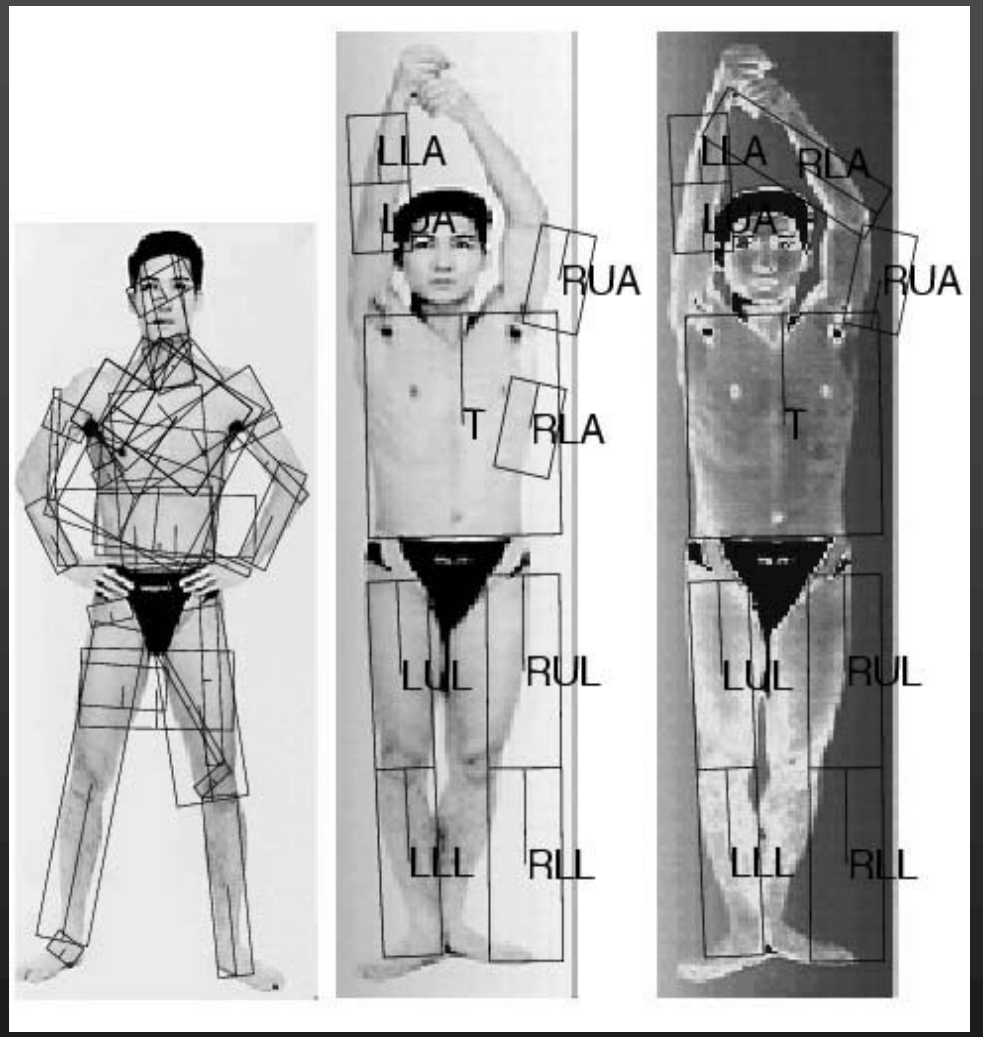
This is hard to operationalize



Ponce et al. (1989)

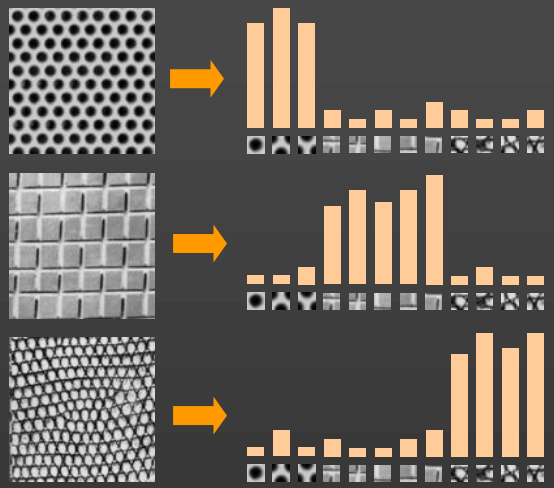


Zhu and Yuille (1996)



Ioffe and Forsyth (2000)

Bags of words and their variants have become the dominant model for image categorization



(Swain & Ballard'91; Lazebnik, Schmid, Ponce'03; Sivic & Zisserman,'03; Csurka et al.'04; Zhang et al.'06)

Locally orderless image models

(Koenderink & Van Doorn'99; Dalal & Triggs'05; Lazebnik, Schmid, Ponce'06; Chum & Zisserman'07)

Image categorization as supervised classification

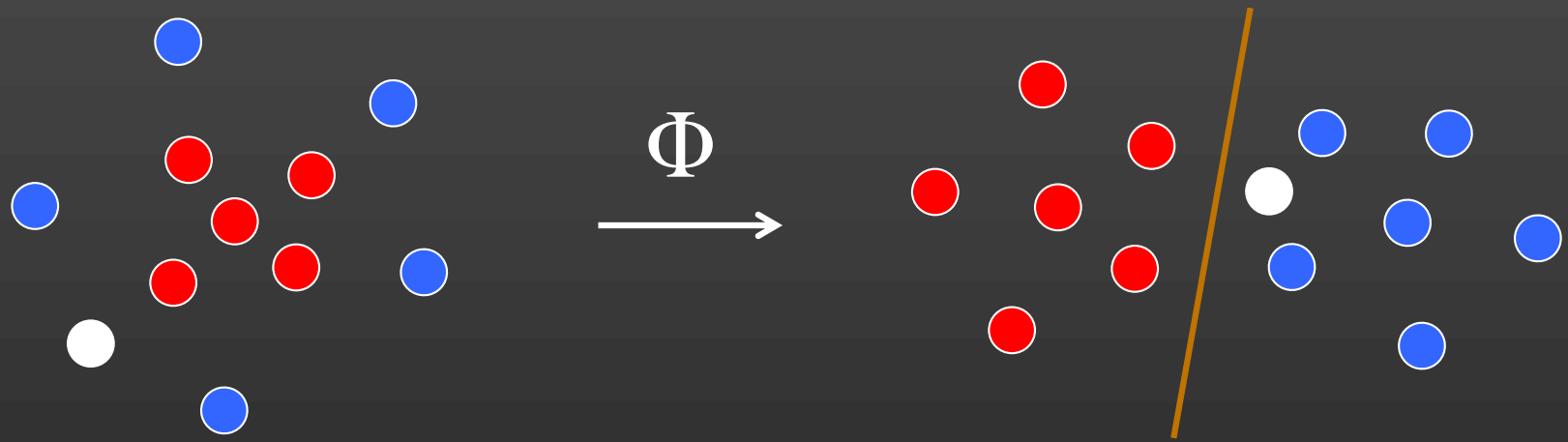


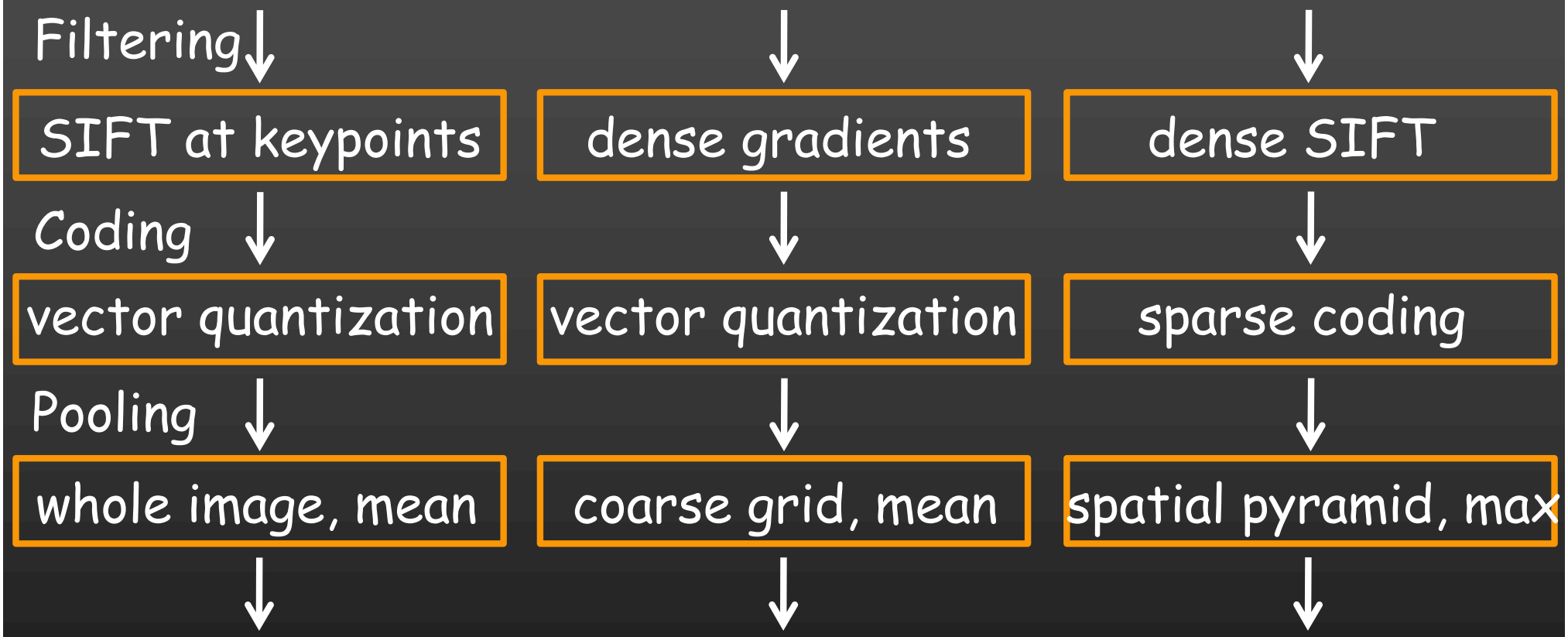
Image "space"

Feature (Hilbert) space

$$\min_{f \in \mathcal{F}} \frac{1}{N} \sum_{\mathrm{n}} \ell(z_n, f(\phi(x_n))) + \Omega(f)$$

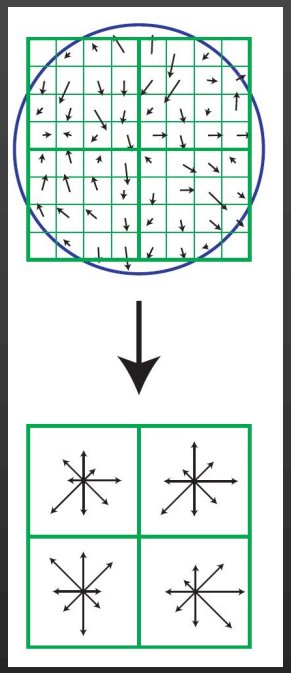
affine

A common architecture for image classification



(Lowe'04, Csurka et al.'04, Dalal & Triggs'05)
(Yang et al.'09-10, Boureau et al.'10, Mallat'11)

A common architecture for image classification



SIFT

dense gradients

vector quantization

coarse grid, mean

HOG

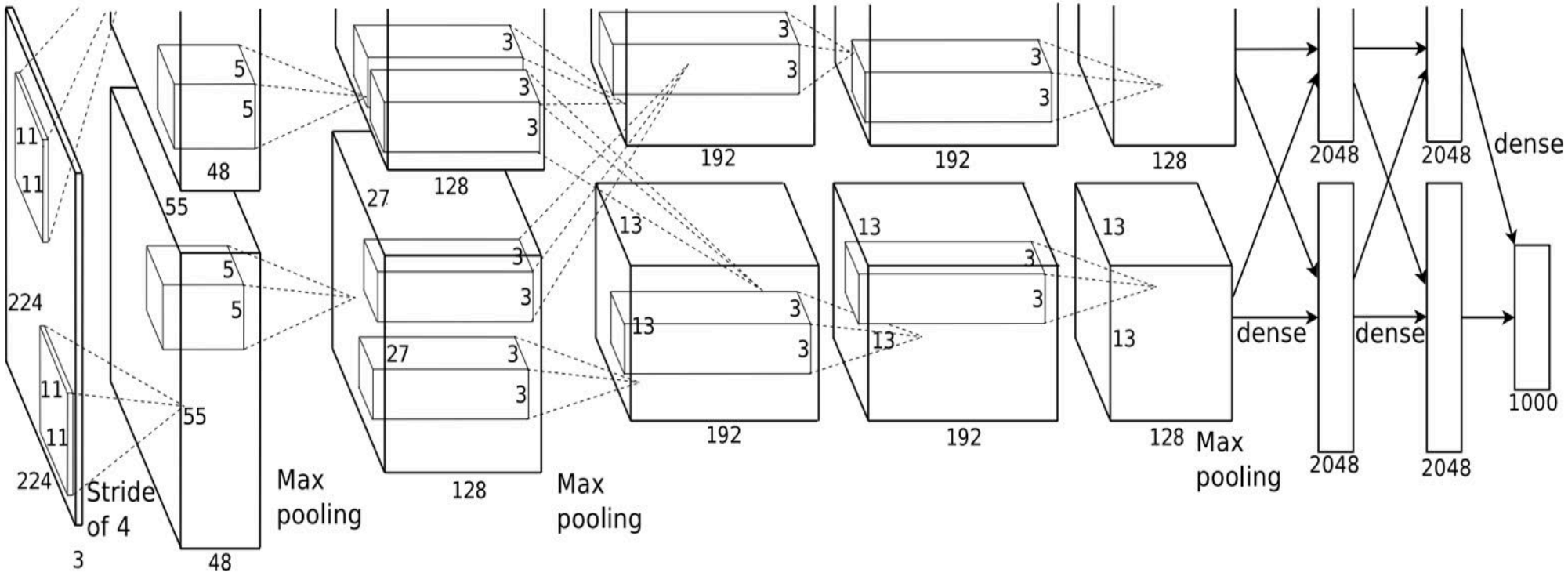
dense gradients

vector quantization

coarse grid, mean

(Lowe'04, Csurka et al.'04, Dalal & Triggs'05)
(Yang et al.'09-10, Boureau et al.'10, Mallat'11)

A common architecture for image classification

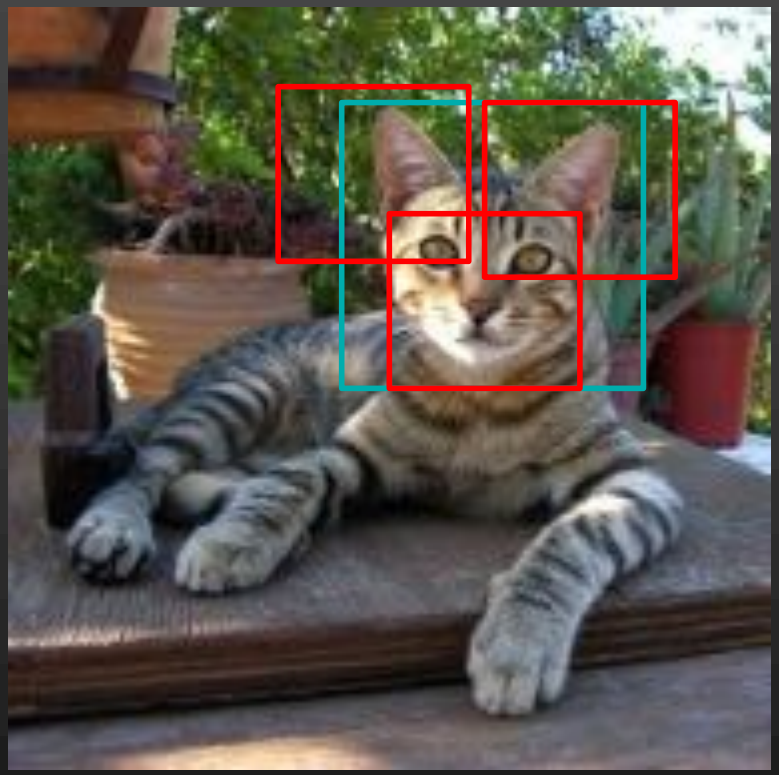


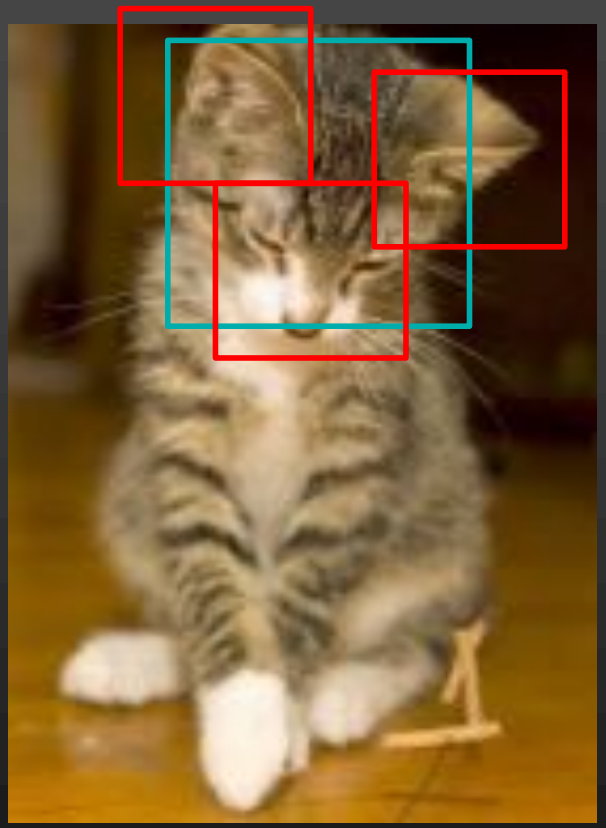
(Deep learning: Krizhevsky, Sutskever, Hinton, 2012)

A common architecture for image classification

(Convolutional neural networks, LeCun, 1998)

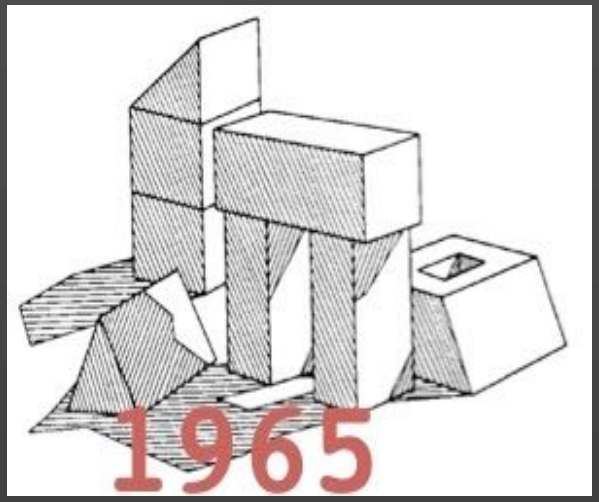
Object detection
(vue "d'artiste")



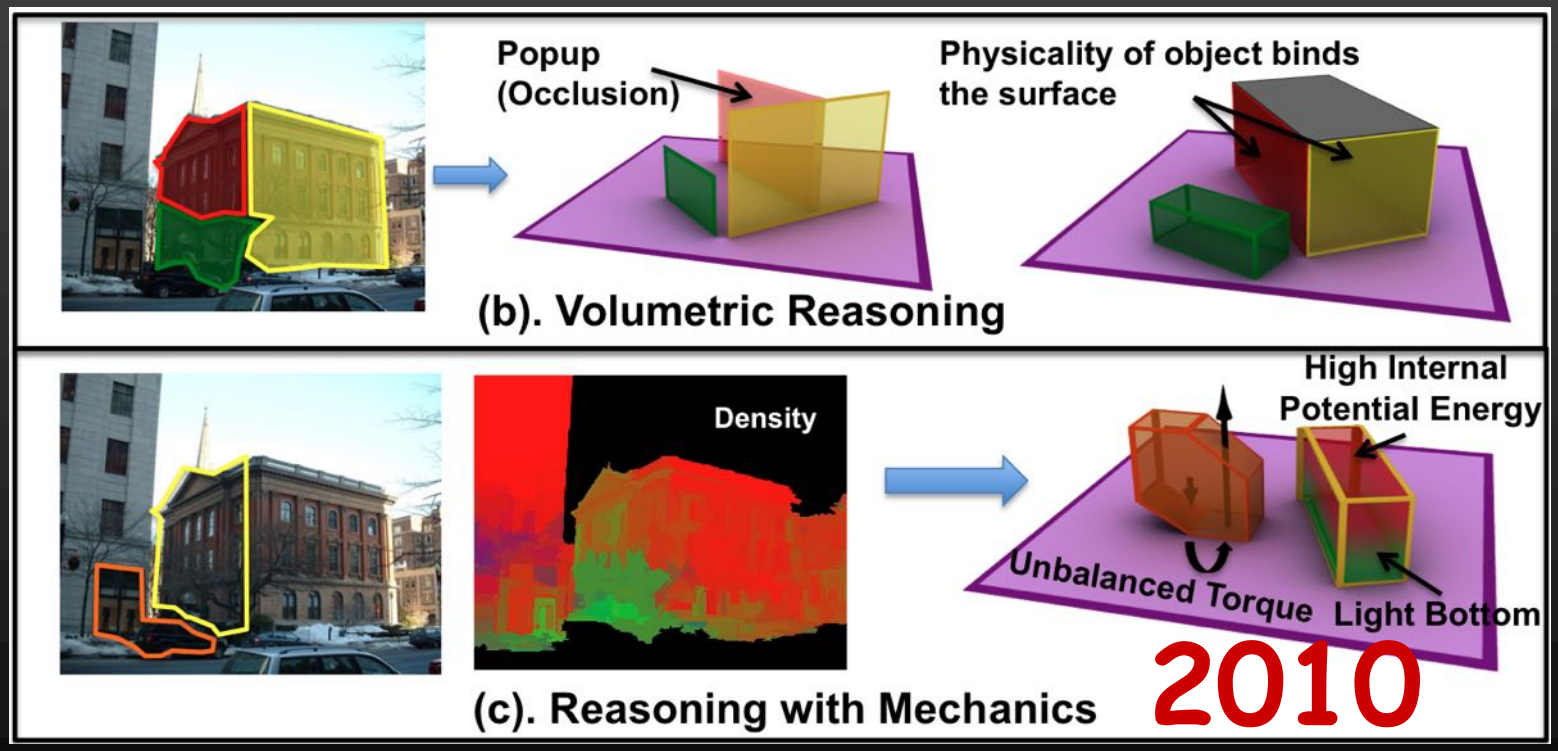


(A la Felzenszwalb, McAllester, Ramanan, 2008)

What about scene understanding?



The blocks world revisited

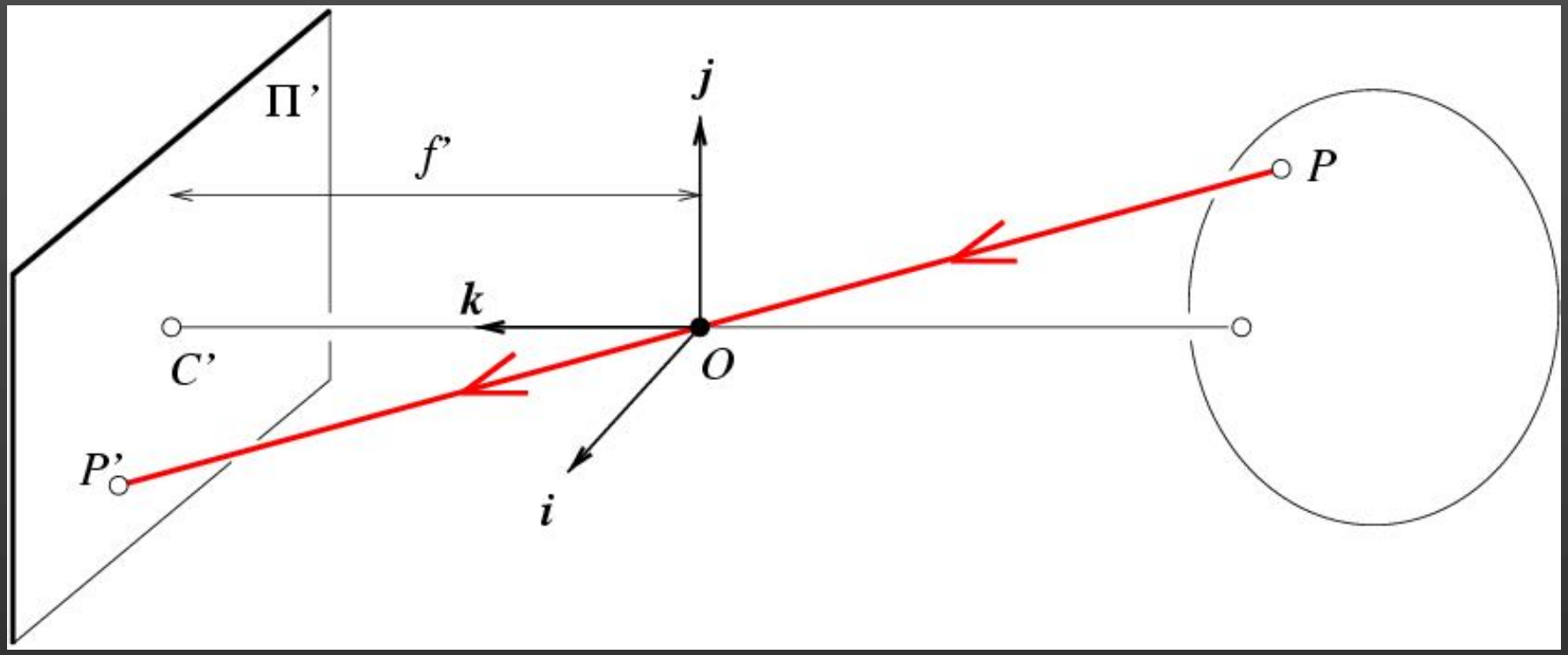


(Gupta, Efros, Hebert, ECCV'10)

Outline

- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry

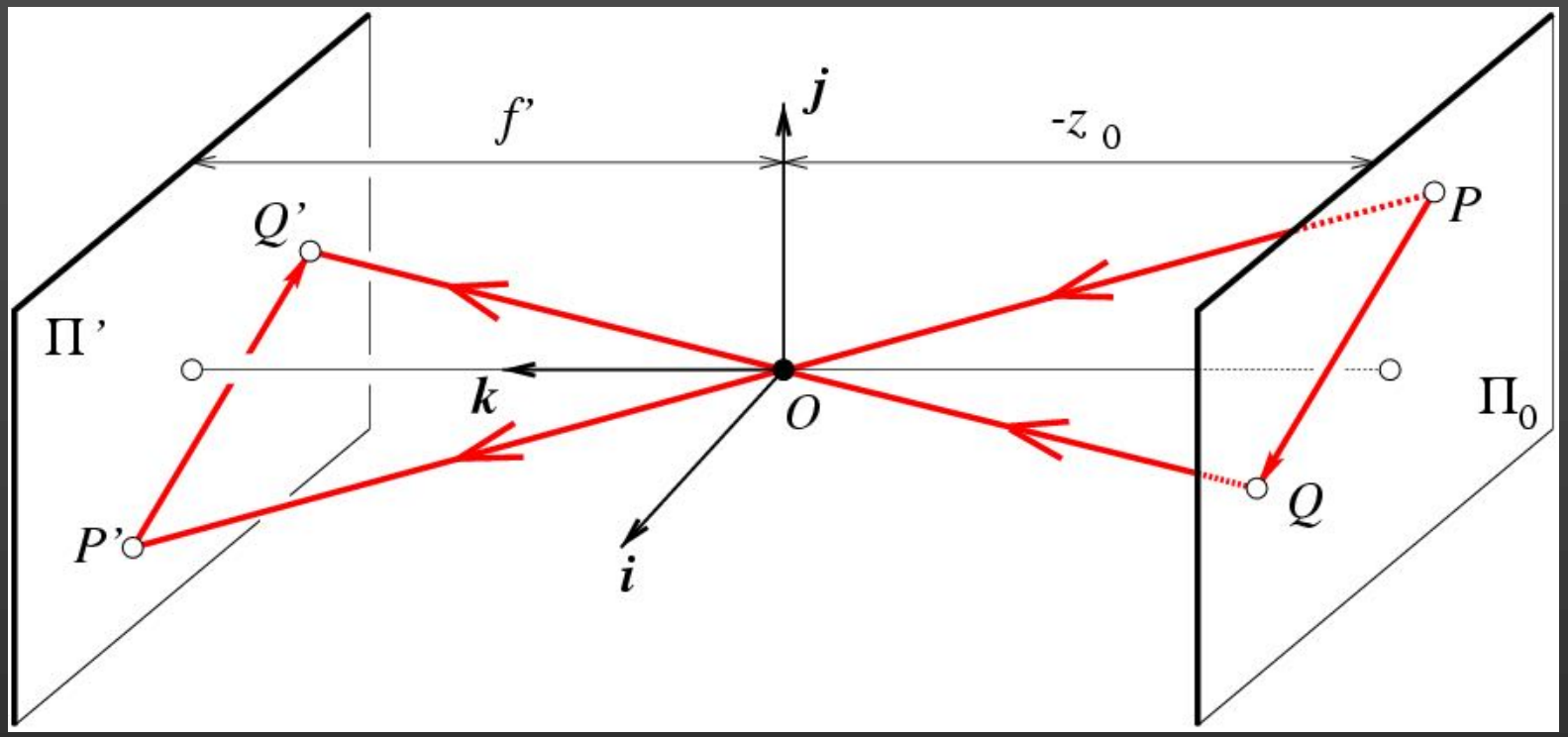
Pinhole perspective equation



$$\begin{cases} x' = f' \dfrac{x}{z} \\ y' = f' \dfrac{y}{z} \end{cases}$$

NOTE: z is always negative..

Affine models: Weak perspective projection

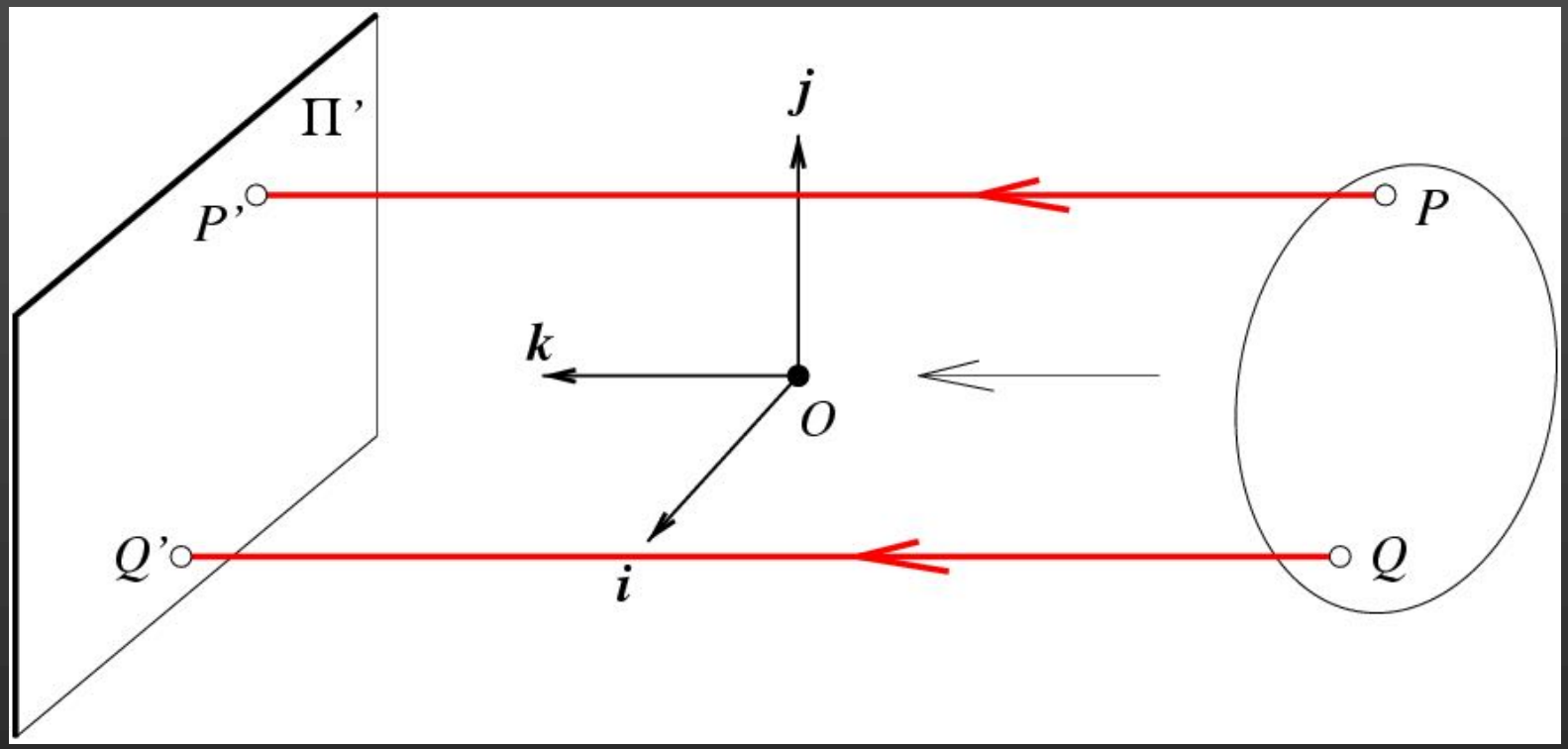


$$\begin{cases} x' = -mx \\ y' = -my \end{cases} \quad \text{where} \quad m = -\frac{f'}{z_0}$$

is the magnification.

When the scene relief is small compared its distance from the camera, *m* can be taken constant: weak perspective projection.

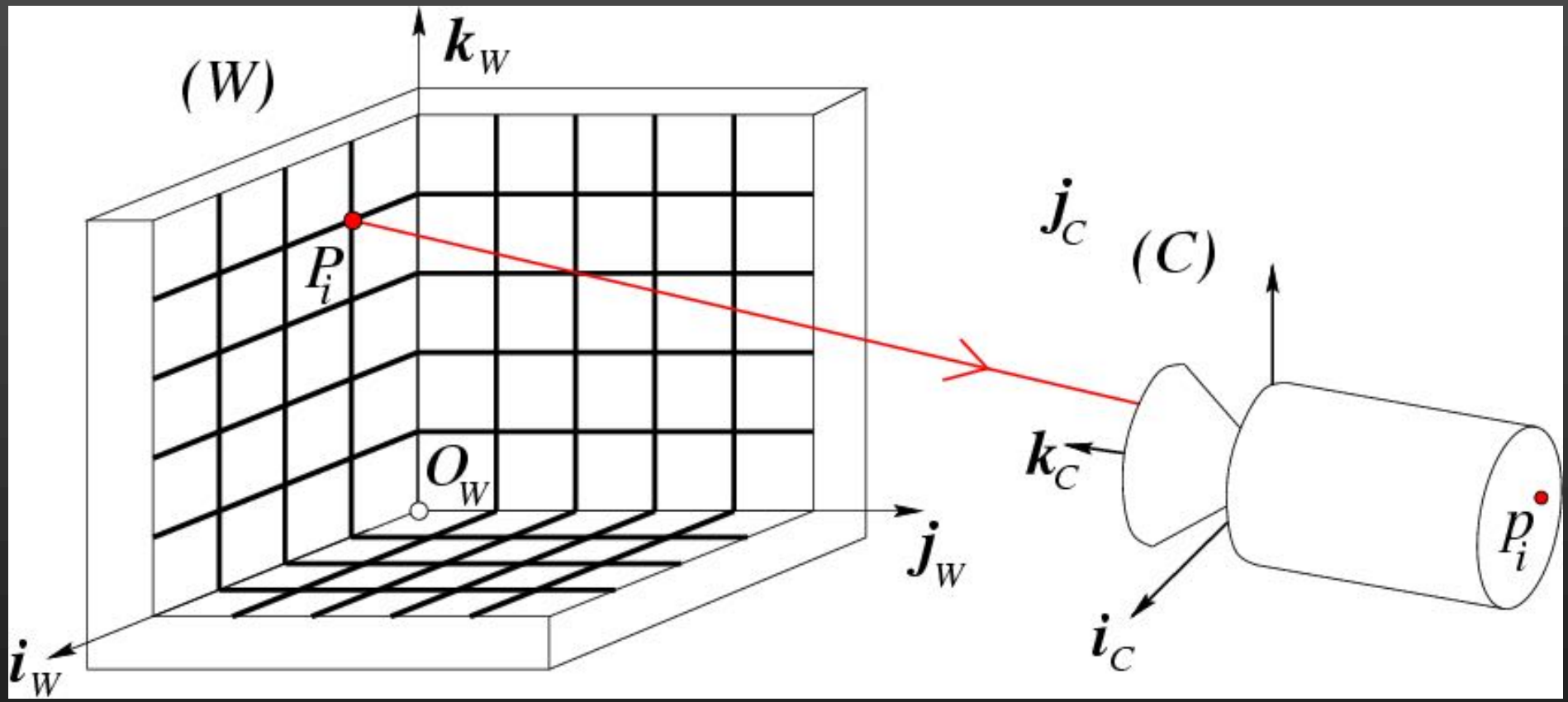
Affine models: Orthographic projection



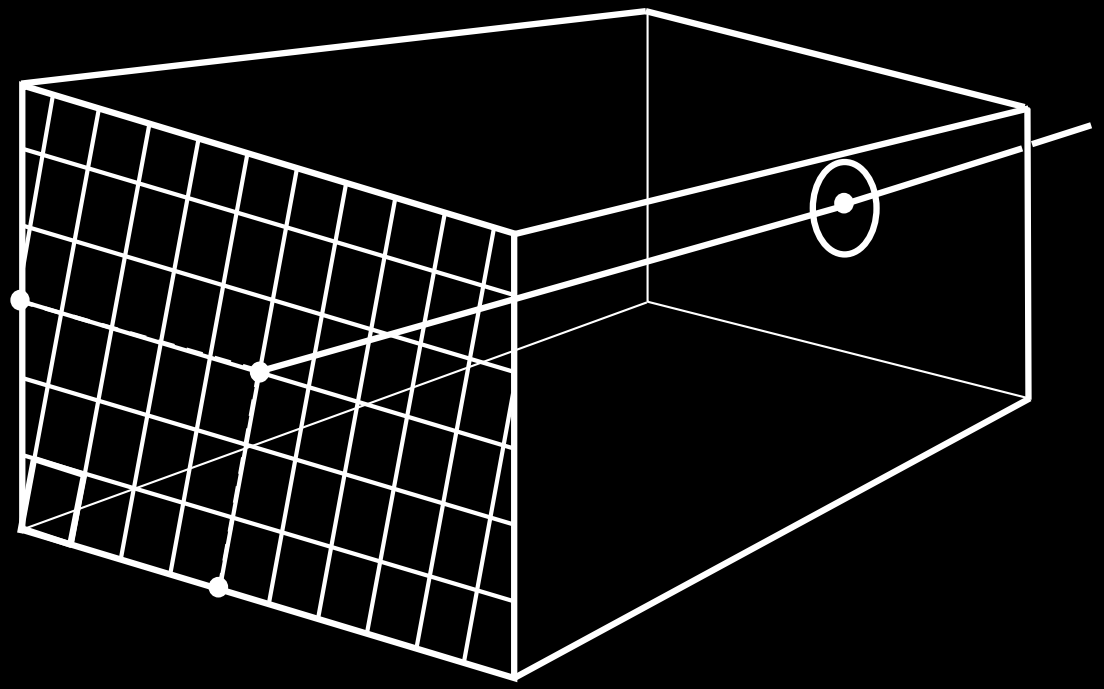
$$\begin{cases} x' = x \\ y' = y \end{cases}$$

When the camera is at a (roughly constant) distance from the scene, take *m*=1.

Analytical camera geometry



Cameras and their parameters



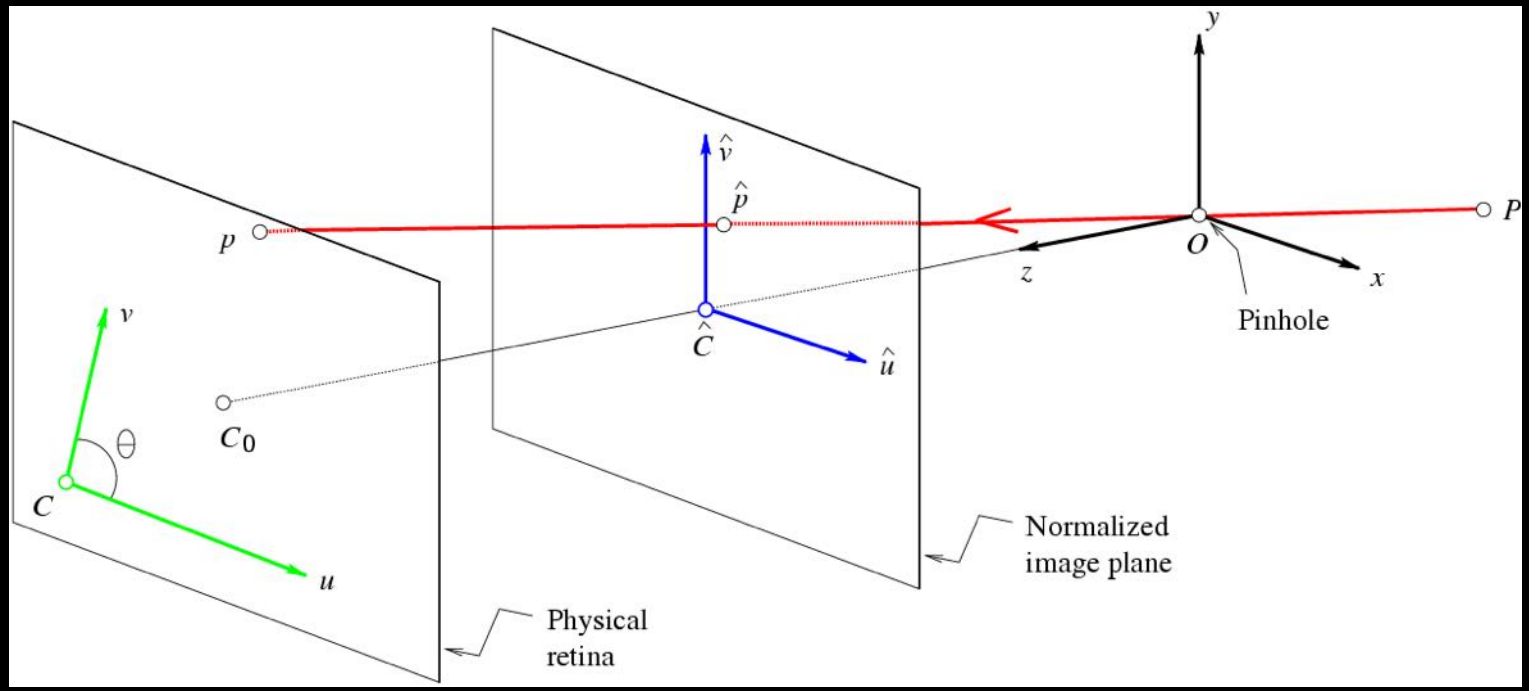
The intrinsic parameters of a camera

Units:

k,l : pixel/m

f : m

α,β : pixel



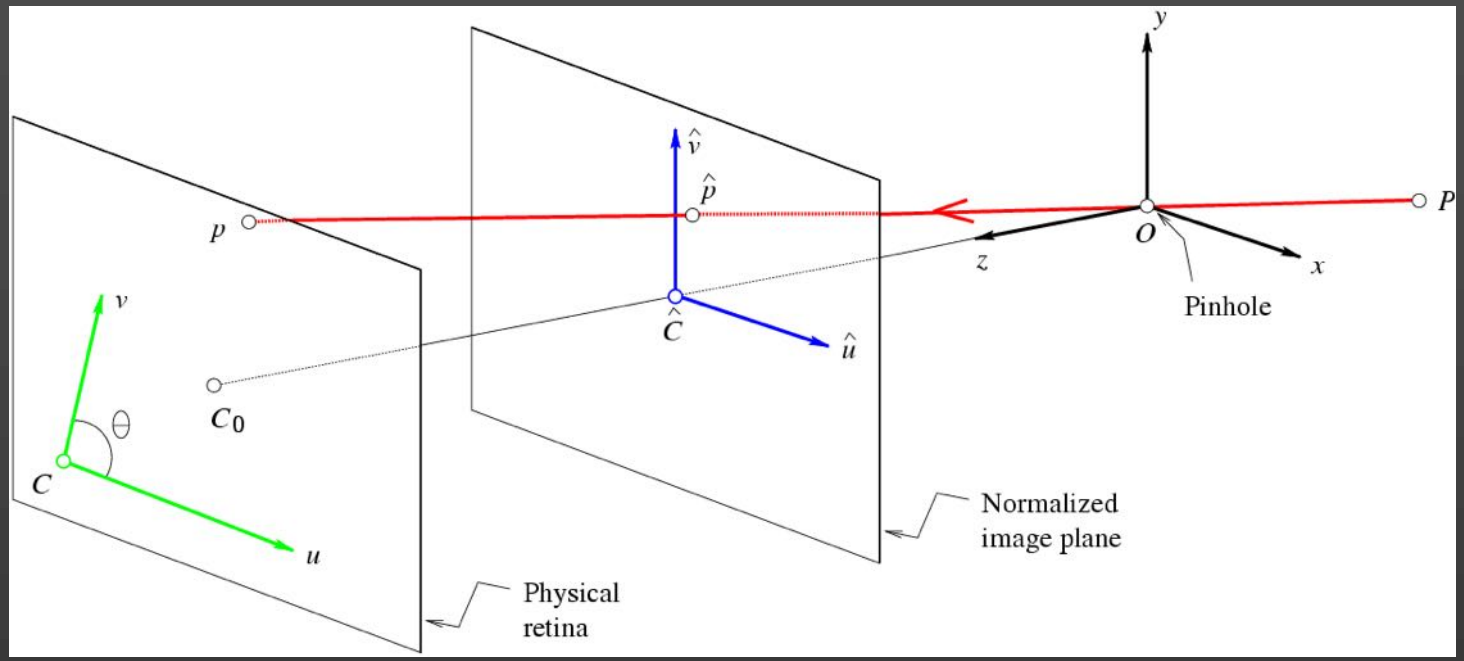
$$\begin{cases} \hat{u} = \dfrac{x}{z} \\ \hat{v} = \dfrac{y}{z} \end{cases} \iff \hat{\boldsymbol{p}} = \frac{1}{z}\begin{pmatrix} \mathrm{Id} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \boldsymbol{P} \\ 1 \end{pmatrix}$$

Normalized image coordinates

Physical image coordinates

$$\begin{cases} u = kf\dfrac{x}{z} \\ v = lf\dfrac{y}{z} \end{cases} \rightarrow \begin{cases} u = \alpha\dfrac{x}{z} + u_0 \\ v = \beta\dfrac{y}{z} + v_0 \end{cases} \rightarrow \begin{cases} u = \alpha\dfrac{x}{z} - \alpha\cot\theta\dfrac{y}{z} + u_0 \\ v = \dfrac{\beta}{\sin\theta}\dfrac{y}{z} + v_0 \end{cases}$$

The intrinsic parameters of a camera



Calibration matrix

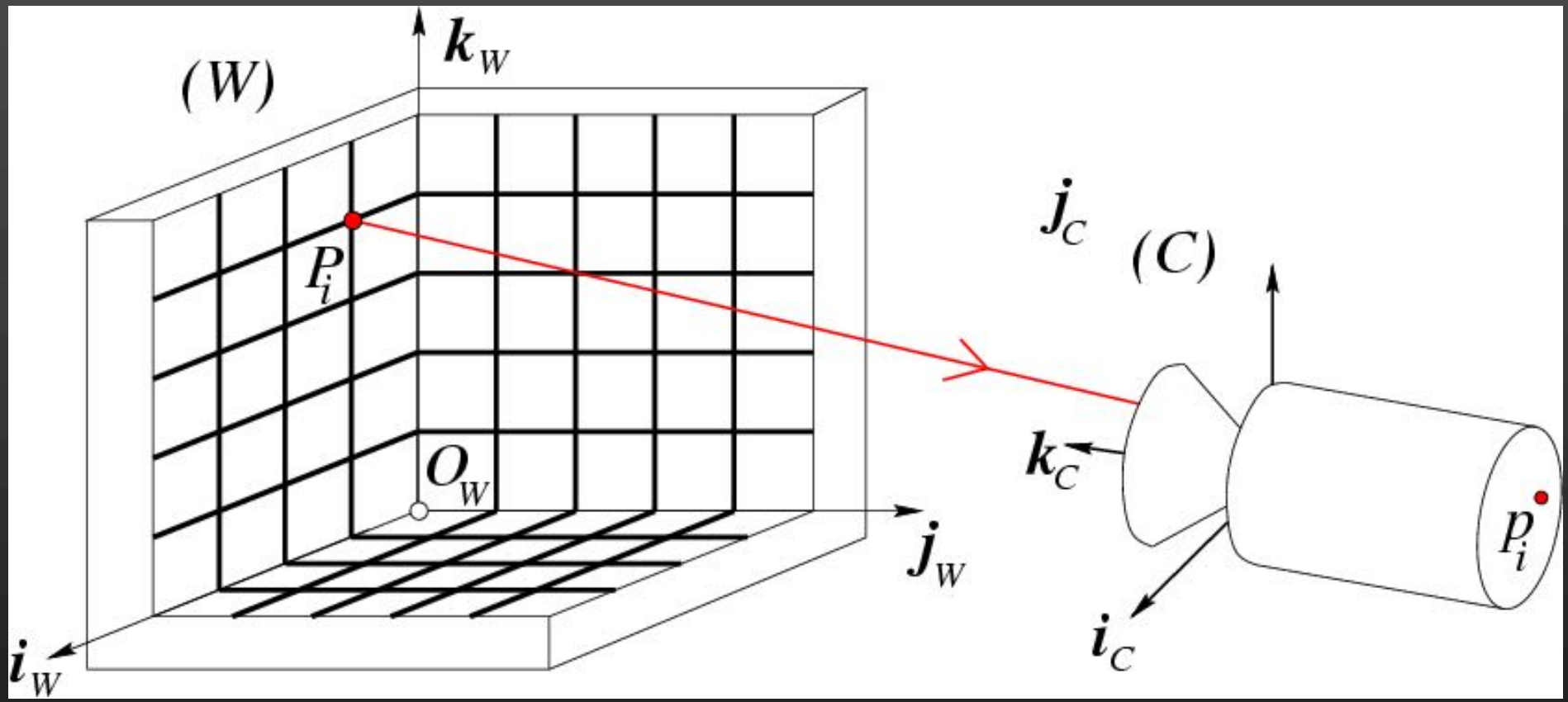
$$\boldsymbol{p} = \mathcal{K}\hat{\boldsymbol{p}}, \quad \text{where} \quad \boldsymbol{p} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad \text{and} \quad \mathcal{K} \stackrel{\text{def}}{=} \begin{pmatrix} \alpha & -\alpha \cot\theta & u_0 \\ 0 & \dfrac{\beta}{\sin\theta} & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Homogeneous coordinates

The perspective projection equation

$$\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P}, \quad \text{where} \quad \mathcal{M} \stackrel{\text{def}}{=} (\mathcal{K} \quad \mathbf{0})$$

Analytical camera geometry



The extrinsic parameters of a camera

- When the camera frame (C) is different from the world frame (W),

$$\begin{pmatrix} {}^{C}P \\ 1 \end{pmatrix} = \begin{pmatrix} {}^{C}_{W}\mathcal{R} & {}^{C}O_{W} \\ \mathbf{0}^{T} & 1 \end{pmatrix} \begin{pmatrix} {}^{W}P \\ 1 \end{pmatrix}.$$

- Thus,

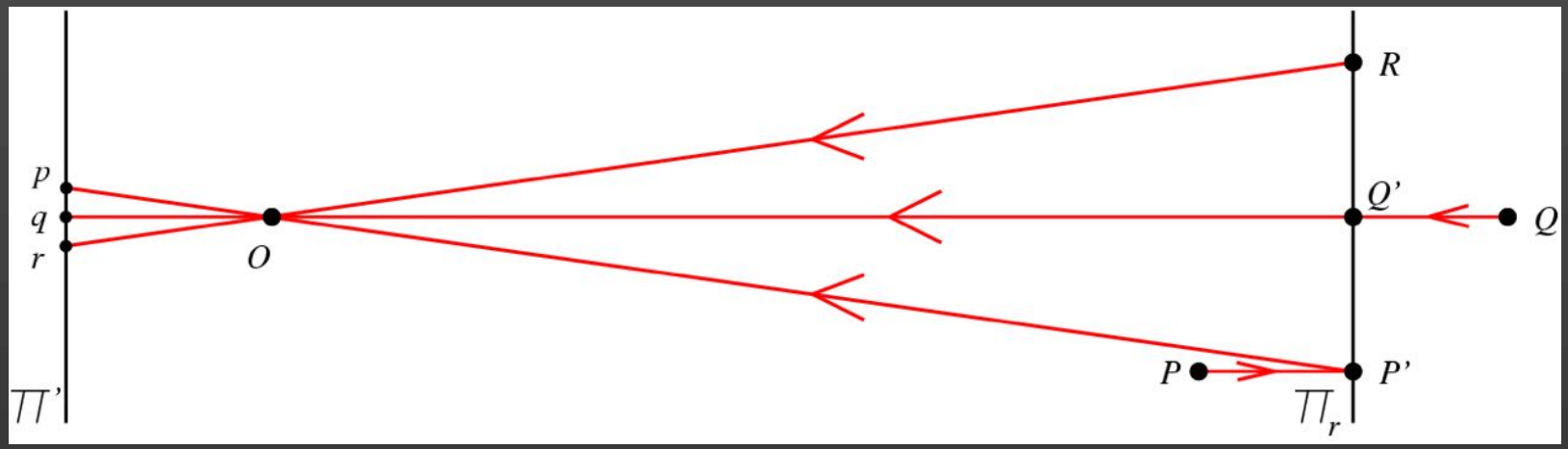
$$\boxed{\boldsymbol{p} = \frac{1}{z}\mathcal{M}\boldsymbol{P},} \quad \text{where} \quad \begin{cases} \mathcal{M} = \mathcal{K}(\mathcal{R} \quad \boldsymbol{t}), \\ \mathcal{R} = {}^{C}_{W}\mathcal{R}, \\ \boldsymbol{t} = {}^{C}O_{W}, \\ \boldsymbol{P} = \begin{pmatrix} {}^{W}P \\ 1 \end{pmatrix}. \end{cases}$$

- Note: z is *not* independent of $\mathcal{M}$ and $\boldsymbol{P}$:

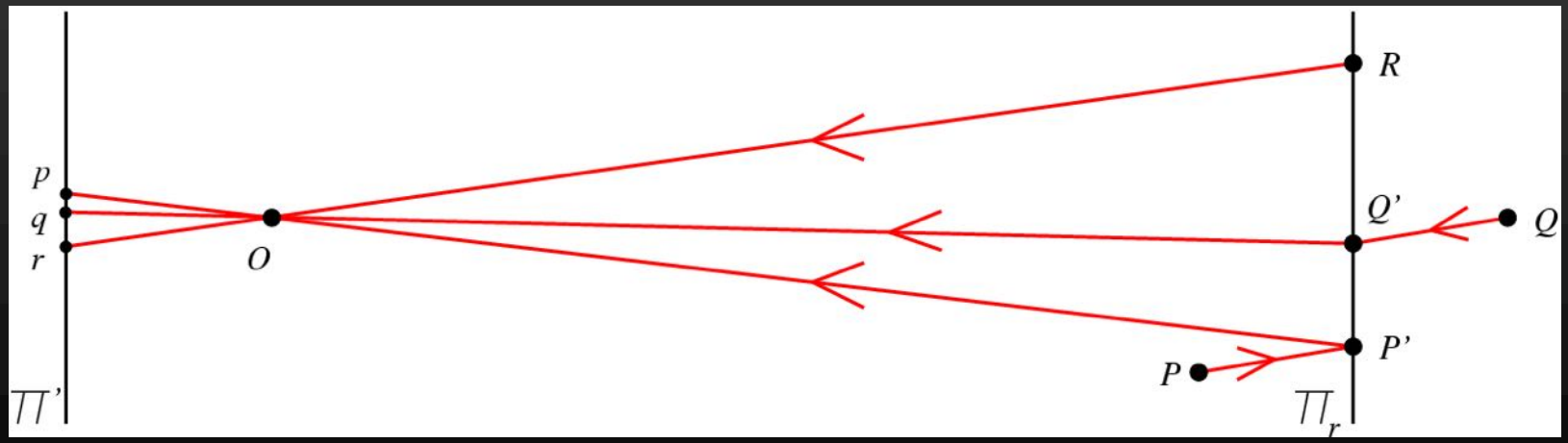
$$\mathcal{M} = \begin{pmatrix} \boldsymbol{m}_1^T \\ \boldsymbol{m}_2^T \\ \boldsymbol{m}_3^T \end{pmatrix} \Longrightarrow z = \boldsymbol{m}_3 \cdot \boldsymbol{P}, \quad \text{or} \quad \begin{cases} u = \dfrac{\boldsymbol{m}_1 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}}, \\ v = \dfrac{\boldsymbol{m}_2 \cdot \boldsymbol{P}}{\boldsymbol{m}_3 \cdot \boldsymbol{P}}. \end{cases}$$

Affine cameras

Weak-perspective projection

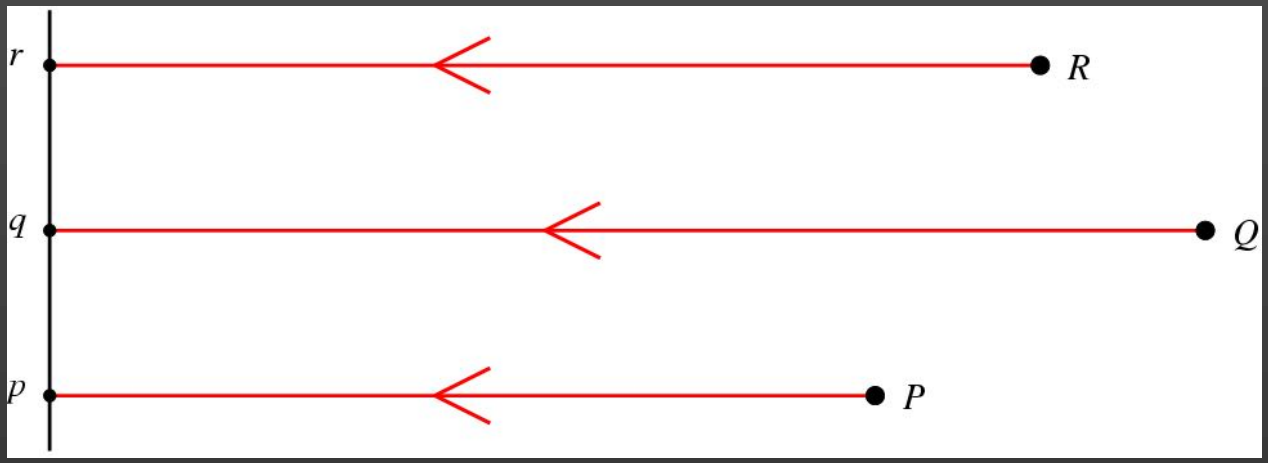


Paraperspective projection

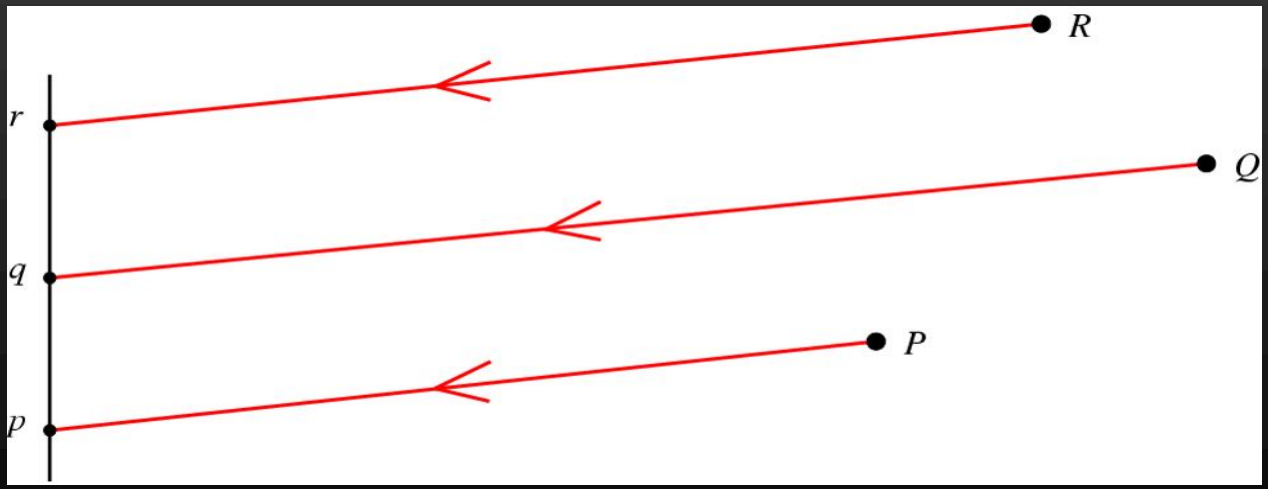


More affine cameras

Orthographic projection



Parallel projection



Weak-perspective projection model

$$\boldsymbol{p} = \frac{1}{z_r}\mathcal{M}\boldsymbol{P}$$

(p and P are in homogeneous coordinates)

$$\boldsymbol{p} = M\,\boldsymbol{P}$$

(P is in homogeneous coordinates)

$$\boldsymbol{p} = A\,\boldsymbol{P} + \boldsymbol{b}$$

(neither p nor P is in hom. coordinates)