

Reconnaissance d'objets et vision artificielle

Jean Ponce (ponce@di.ens.fr)

<http://www.di.ens.fr/~ponce>

Equipe-projet WILLOW

ENS/INRIA/CNRS UMR 8548

Laboratoire d'Informatique

Ecole Normale Supérieure, Paris

Jean Ponce



<http://www.di.ens.fr/~ponce/>

Cordelia Schmid



<http://lear.inrialpes.fr/~schmid/>

Josef Sivic



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

Ivan Laptev

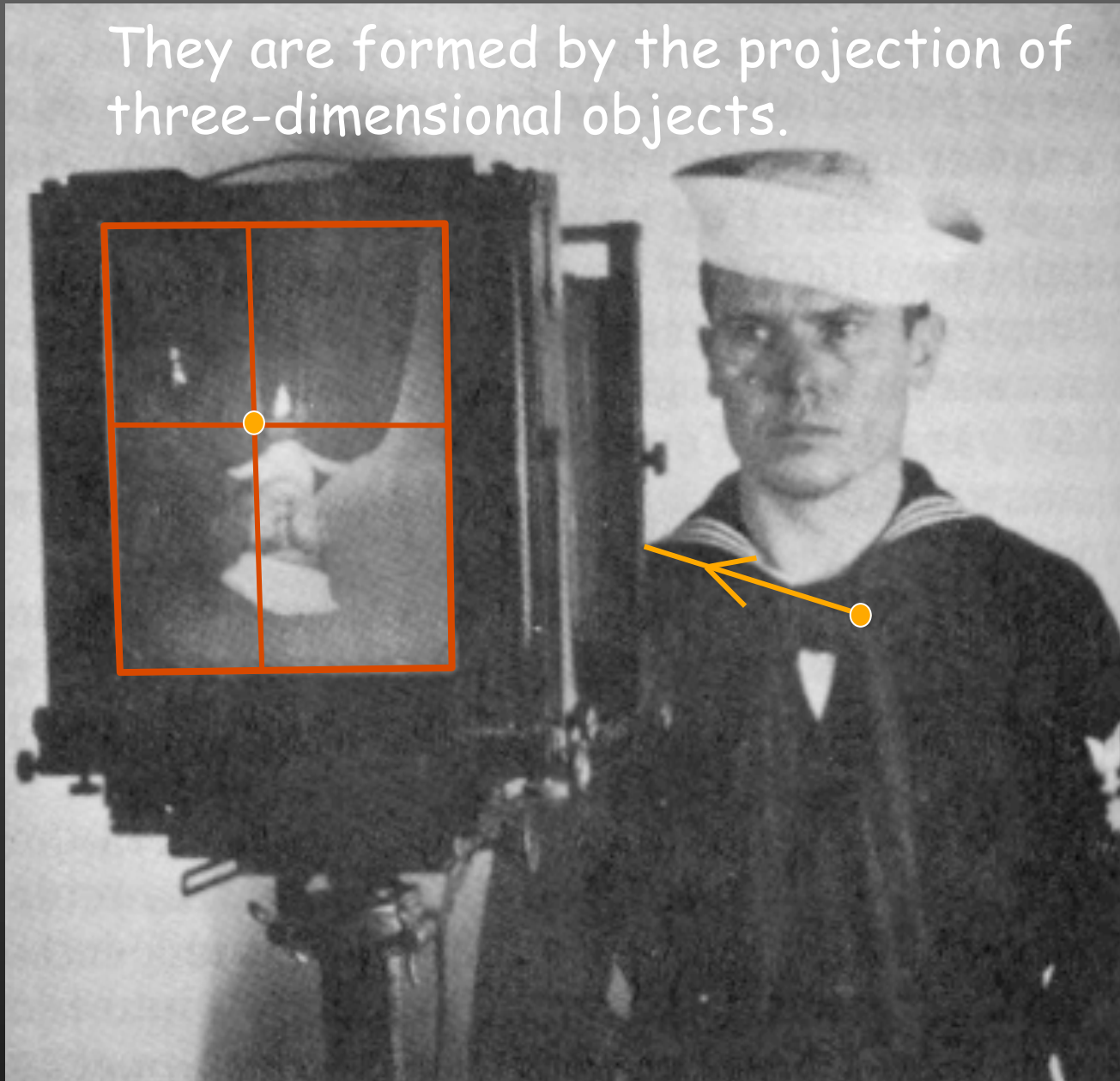


<http://www.irisa.fr/vista/Equipe/People/Ivan.Laptev.html>

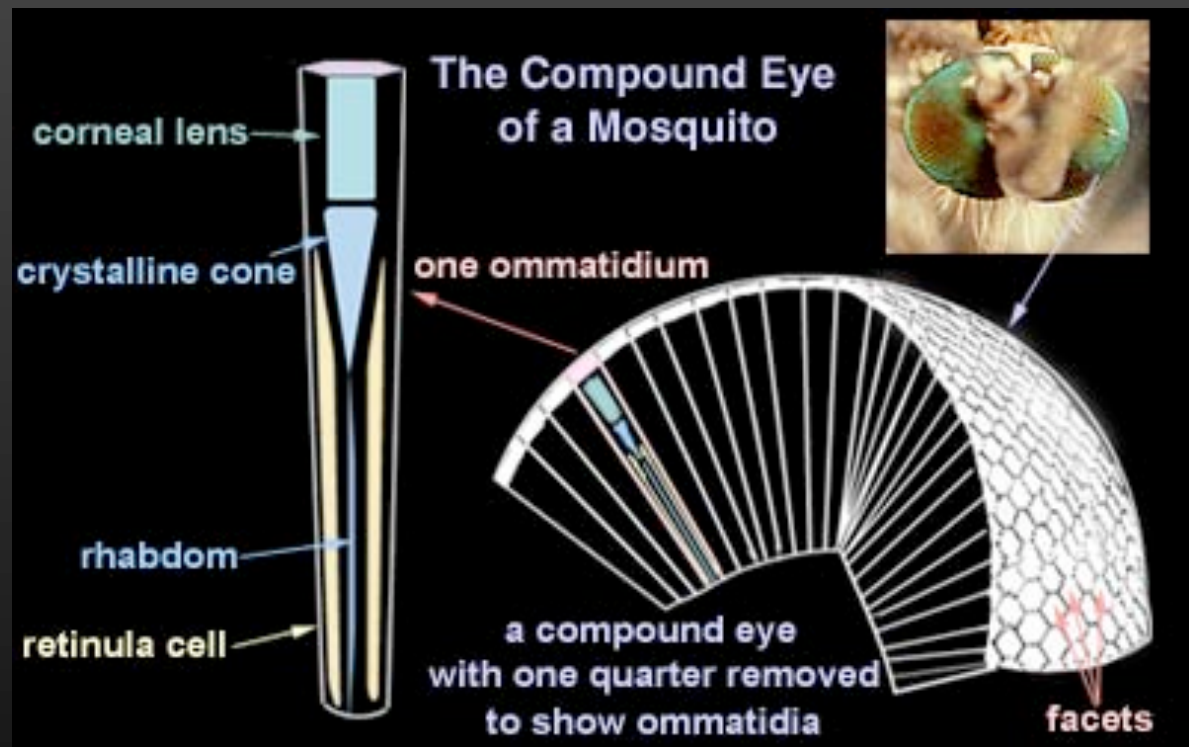
Outline

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

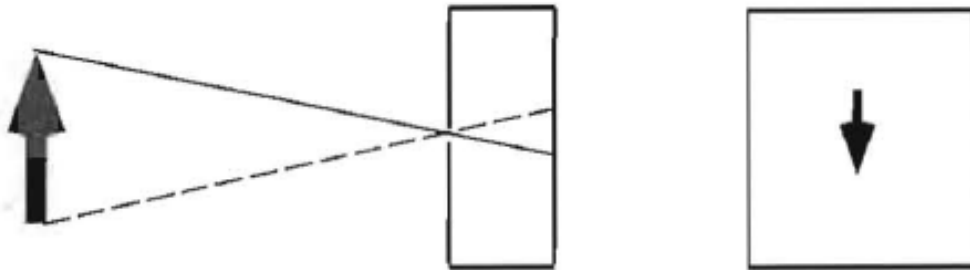
They are formed by the projection of three-dimensional objects.



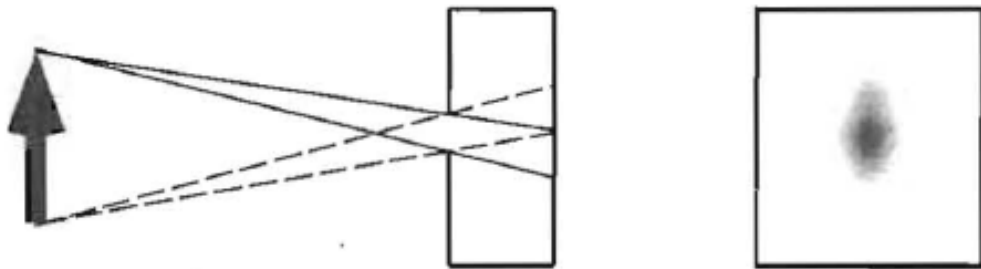
Images are brightness/color patterns drawn in a plane.



Pinhole camera: trade-off between sharpness and light transmission



A. Pinhole Aperture without Lens --> Sharp Image

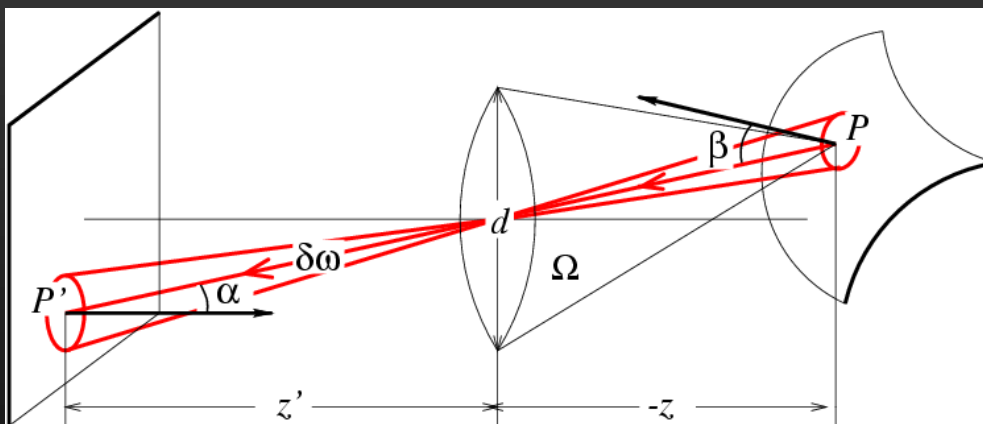
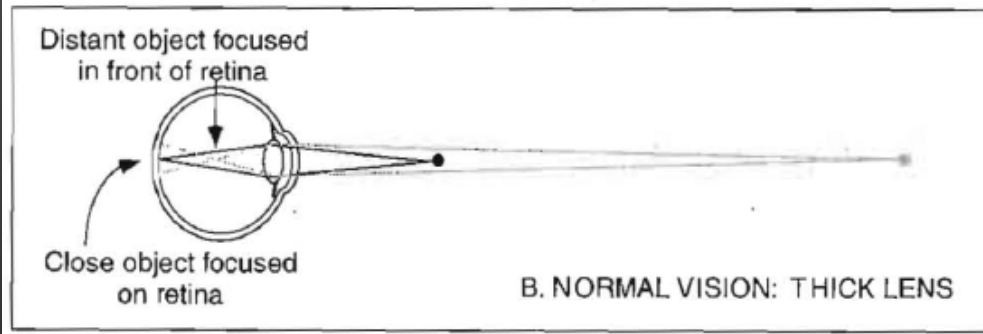
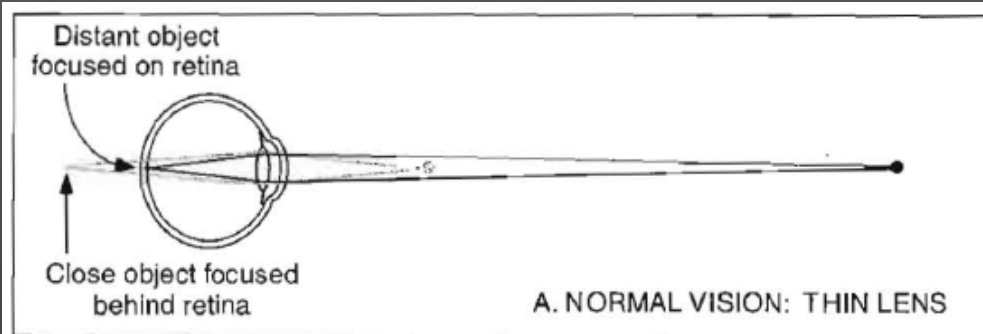


B. Large Aperture without Lens --> Fuzzy Image



Camera Obscura in
Edinburgh

Advantages of lens systems

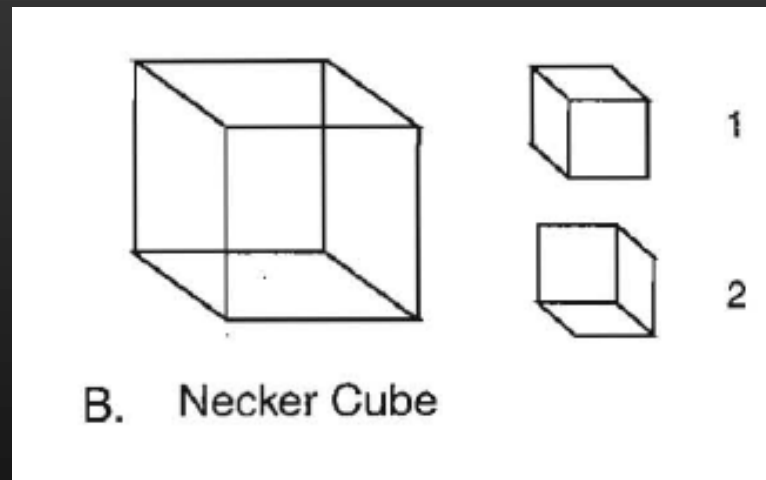
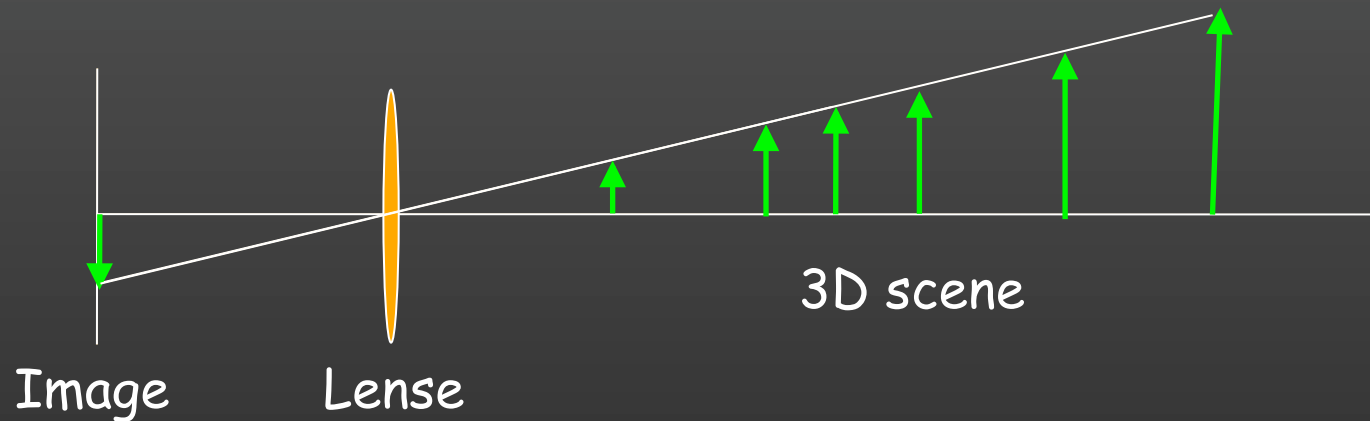


- Can focus sharply on close and distanced objects
- Transmits more light than a pinhole camera

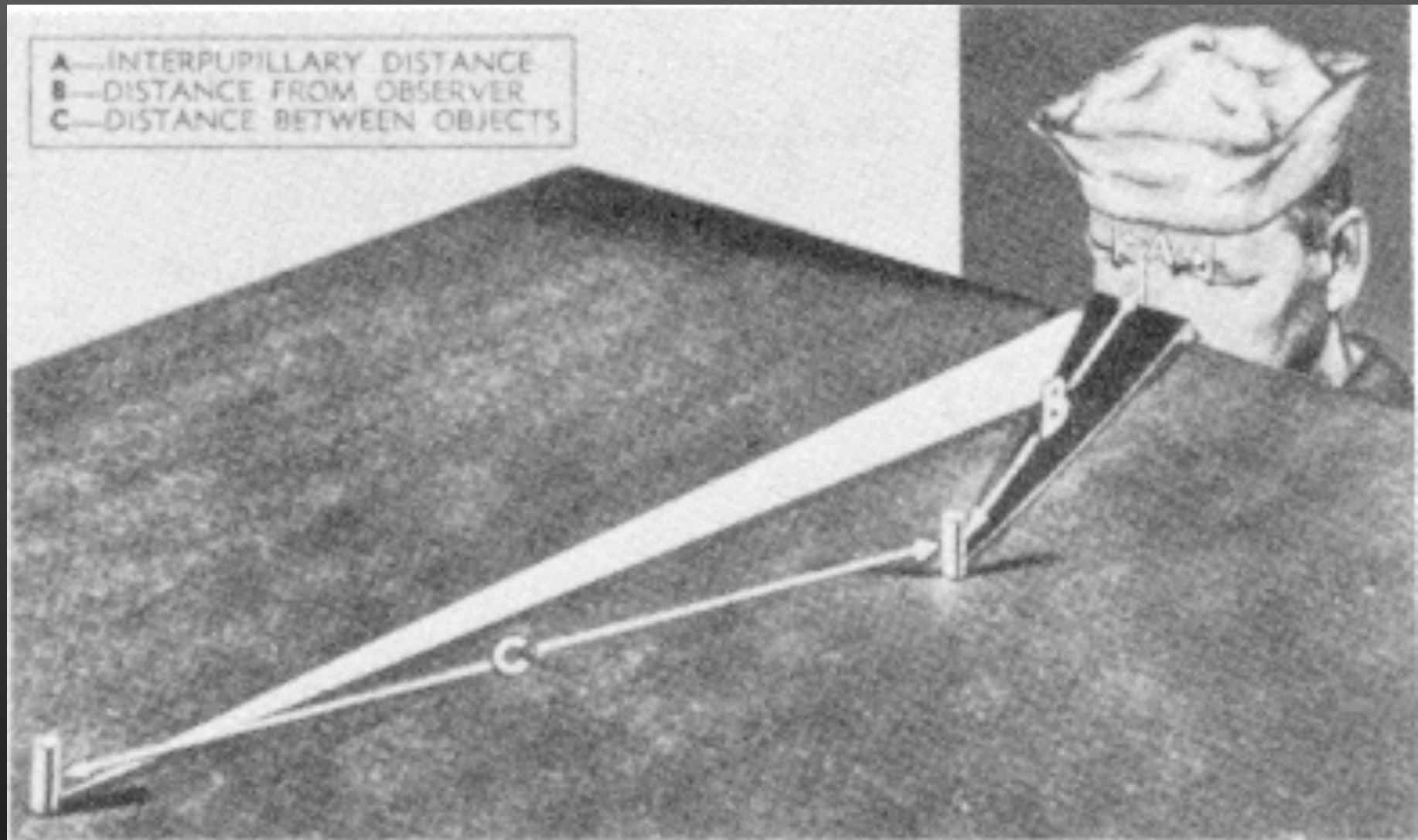
$$E = (\pi/4) \left[(d/z')^2 \cos^4 \alpha \right] L$$

Fundamental problem I:

3D world is "flattened" to 2D images
Loss of information

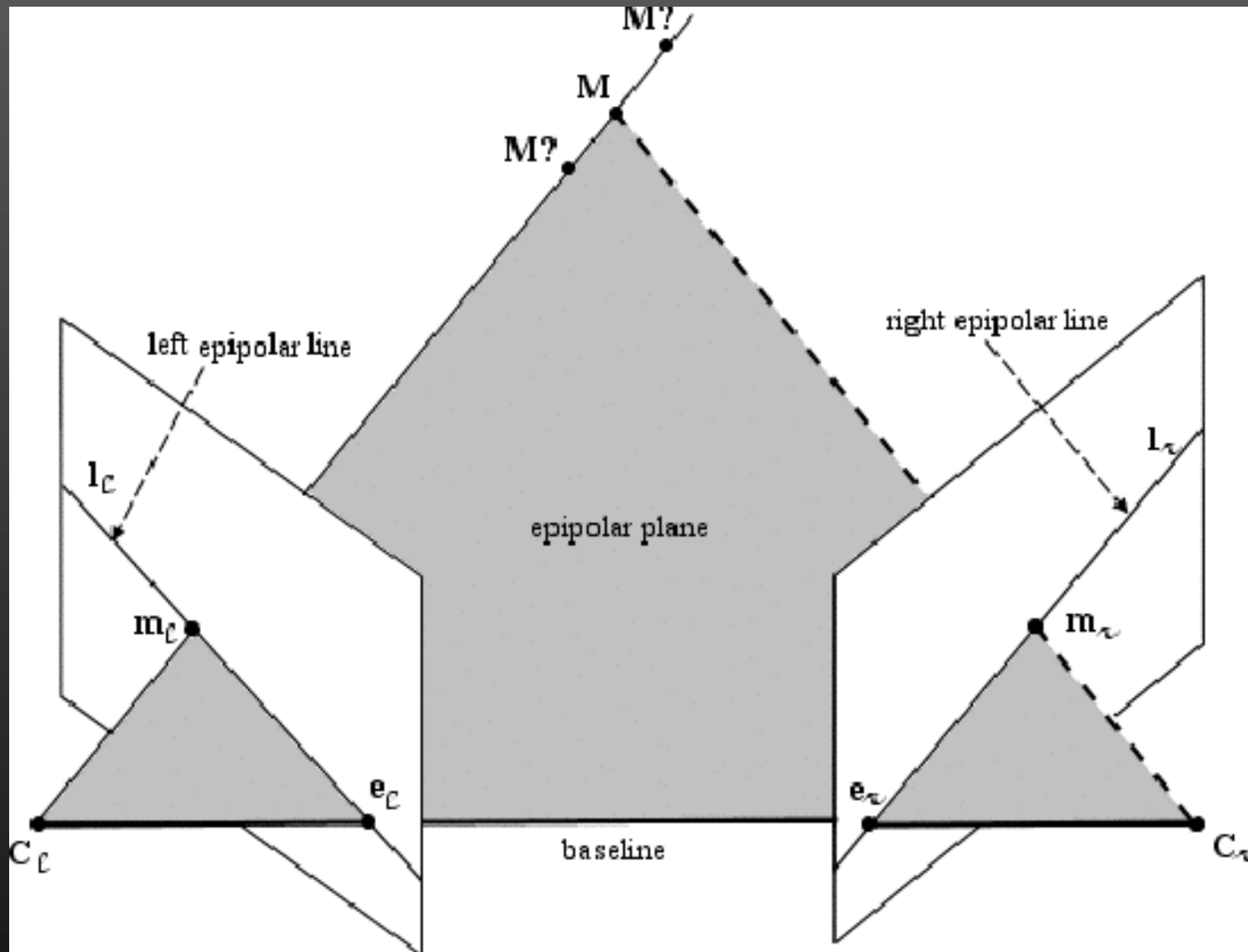


Question : how do we see "in 3D" ?



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

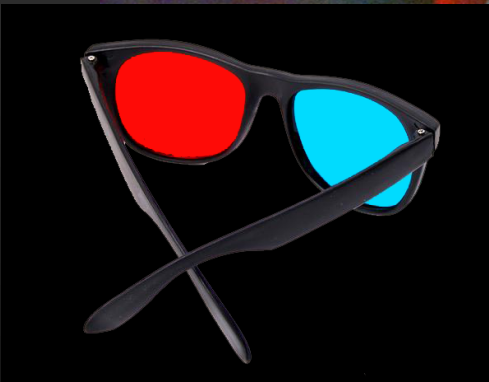
Epipolar Geometry



Simulated 3D perception



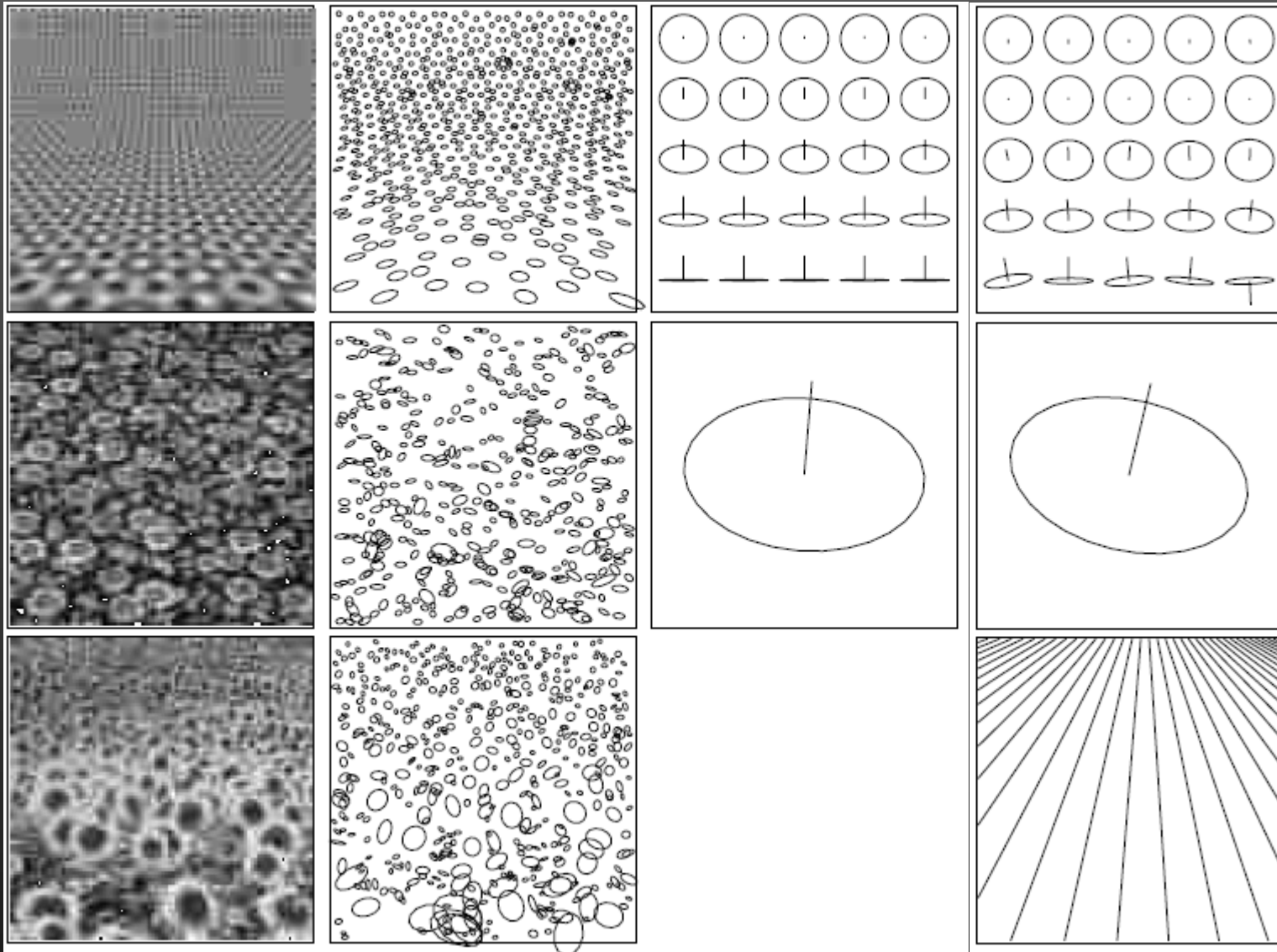
Disparity



But there are other cues..



Shape from texture





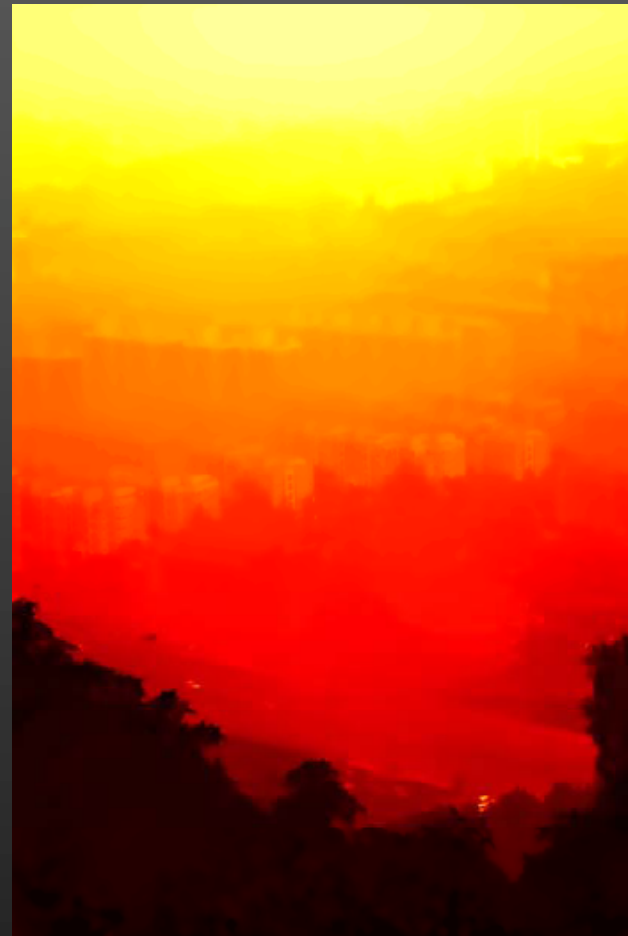
Depth from haze



Input haze image

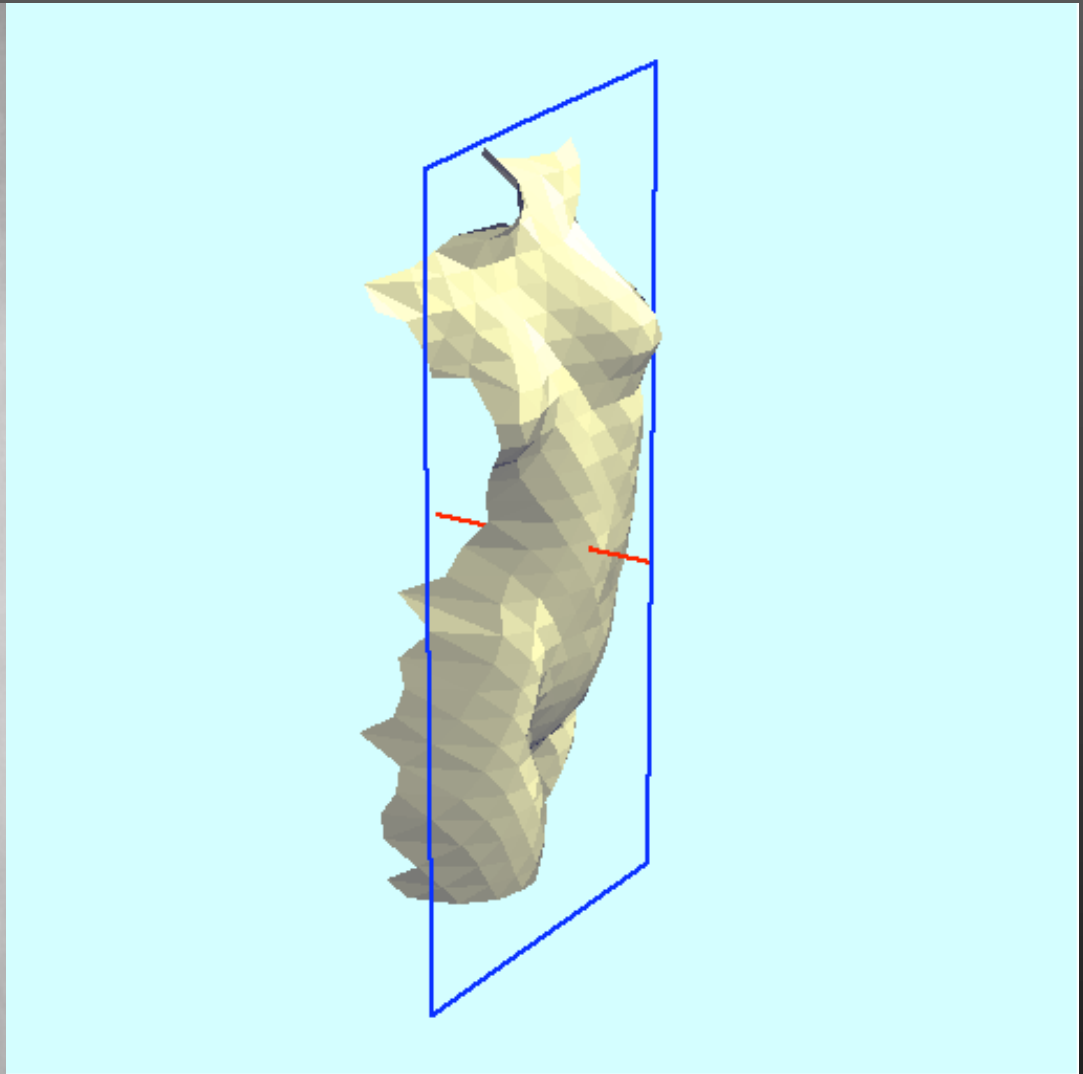


Reconstructed images



Recovered depth map

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



Source: J. Koenderink



Source: J. Koenderink

What is happening with the shadows?





Image source: F. Durand

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!

The goal of computer vision



To perceive the "world behind the picture", e.g.,

- as a metric measurement device
- as a device for measuring "semantic" information

The goal of computer vision



A 20x20 grid of 400 small, illegible images, representing a dataset for computer vision. The images are arranged in a regular grid and appear to be a collection of small, low-resolution pictures, possibly of objects or scenes, used for training or testing a computer vision algorithm.

To perceive the "world behind the picture", e.g.,

- as a metric measurement device
- as a device for "measuring" semantic information

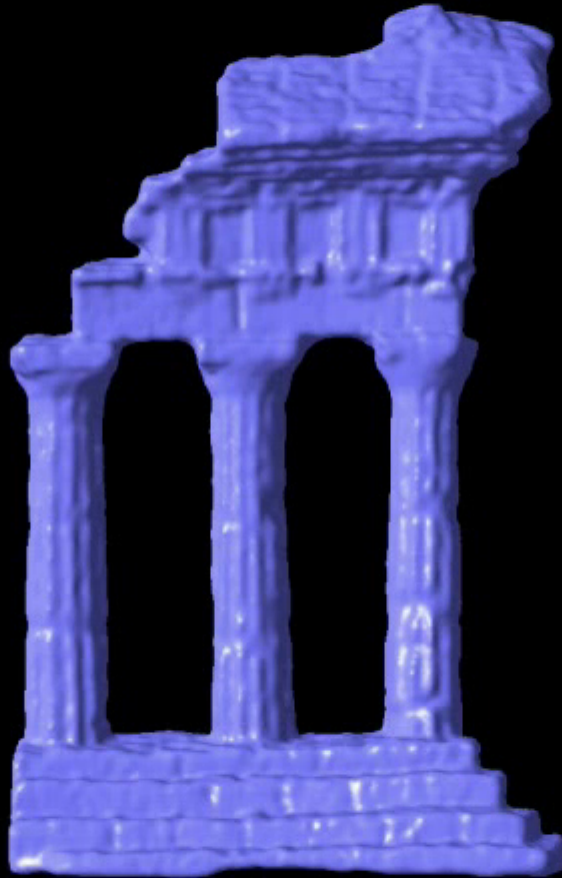
Vision as metric measurement device: Furukawa & Ponce (CVPR'07)
(cf also Keriven's class "Vision et reconstruction 3D")

Full (312)



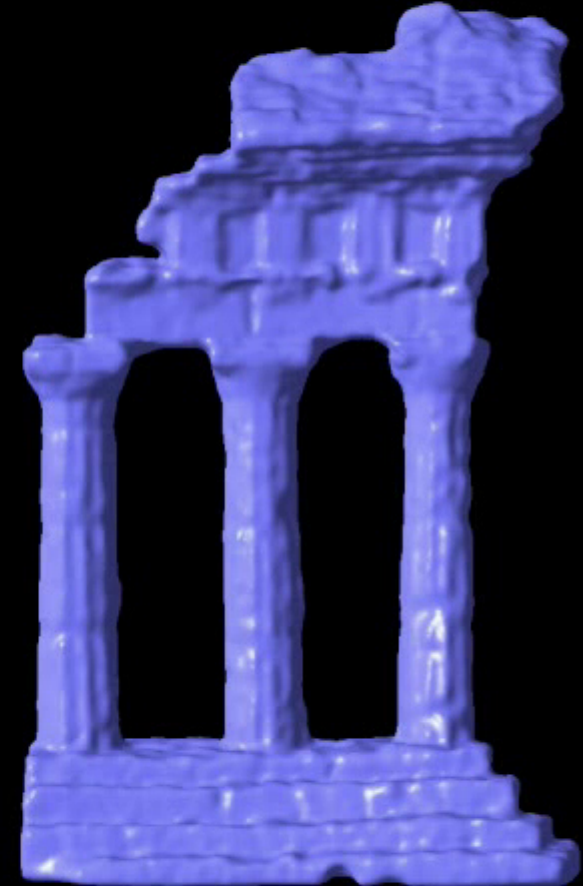
0.49mm (5th)
99.6% (4th)

Ring (47)



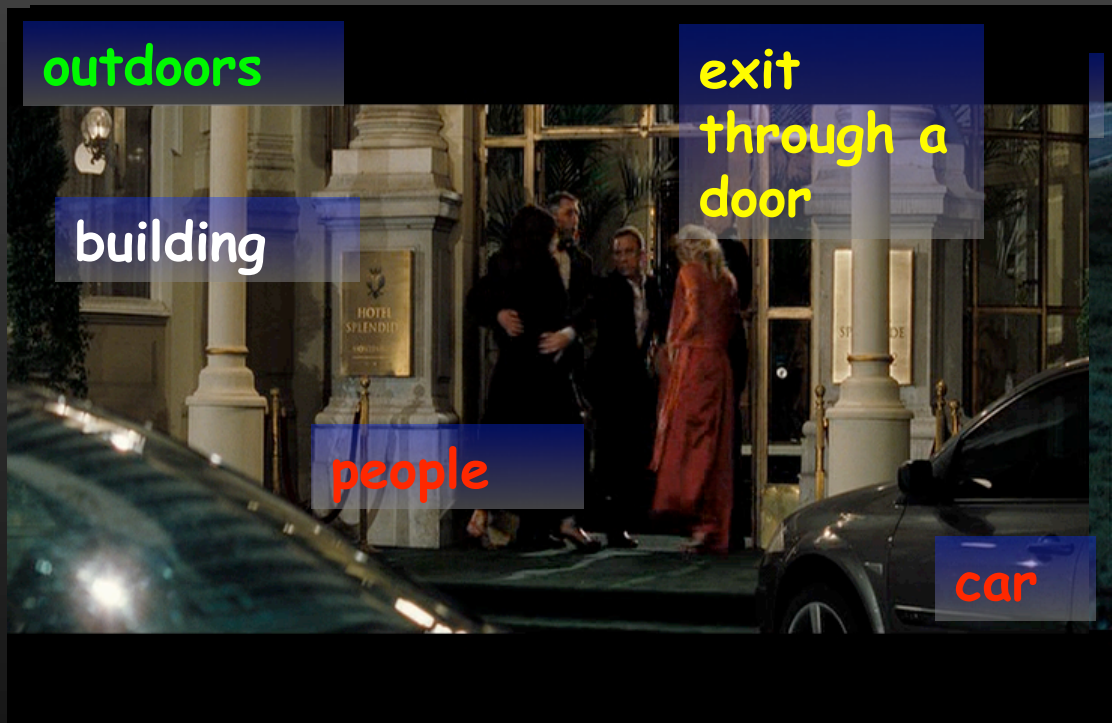
0.47mm (1st)
99.6% (1st)

SparseRing (15)



0.63mm (3rd)
99.3% (1st)
(excluding our previous results)

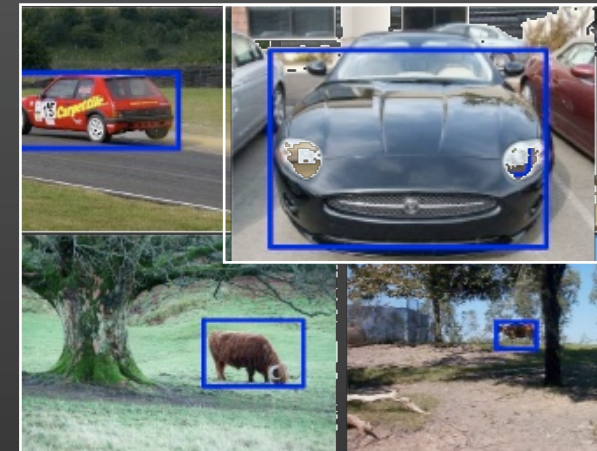
But we want much more than 3D:
ex: Visual scene analysis



How to make sense of "pixel-chaos"?

```
0812263625212131413321050710202322223633364245312220252623161414132094352413141313132823171409181921252721274442332636332191924333640424044352
08133242231010244837260708101827272228343748543631272928231817161440615352181515151417261108071121242930253250473431383218192631263139394940322
091129422300071647392811181820262626303948635442342930272118181615339954817141515141419140007081525252625375447353643024222531243135324640372
09092941394439495039291292327272728442655730162062722191716151727292014151515151415201308091222623234514654940321720253223333334938311
0910234384575246483030141212162923235374752592916135527273292102018181818181718171718140910132326232396442725453822232623233333339341404
07001839451831223840321440081328282831330466236161825202633302625233232323262829813128241313120822727223648402735302616222432340434344748441
05081639421427223037311705001228303139343845487117142225243033029272526272628303236329221217151726304203647382642453120252631243234454582820
0812143338171411263632210509102730312939404240311012192425171319262425293026262728329231613161623312923183340332237402716232530253331465448340
222516283141701112035312005009262828244340363016081017191712041727242236241008131128351406121718233037393033726184342819282729243331394142270
23261923312628091531292205060922191318272929181413172221191603182425162517050912153538160506192321215771272366342938402217232828213131535042180
1819181928180307112320150605102111031823231916131416161819160411182309311400061512283110040423312922262823263293546412120252927172927424747190
051824141810080912231814111720221603151315131212091212121612070611221429150402100723331004042119352524222122324334340261717273342839478393100
051325131616141719181819713161918048131807091615161721201711071031132817020410092734140506111024272927252122202480397412817186414264937100
0910281616121114151616150611141314071018181818231810172818192111133208141004070909490706030011223190725182628148403630302261817363630070
110824171413091213131414121617191611091010101215080413222019171016351100507090508331304072233804131506211830303339303352232631340443821050
572221170914172216222116252320201920141320191010000003071400020513130403040060515361409111417190402020518121604282421263231262423363530402050
482413200913193517191915242724252629221822120302010303060905081306020303040405121412141004050802090804071111081711142330282929263739363219301
080410220913123523282109071014100605091211090407080909070608050607070908060505040707030606070506004101610040708510307213432722213237383817363
0504041306070834221908303100020205058131107131914110810071209020408131210111110080407080826000508041510060211896703032546443311162631343116362
02050803030407100911040020100006050817223816110402010102030405060911111101108020102071010010709051713030403201205043129233312122425272118372
080503040506094904800001040000203071118030930182020303906070839091110111030101021316030913041410204030720503322181712142725241618262
050201030416140308081312080611140401010010426094020203034060606070708091310111714600021807011220041410020300819044291508161120262224161632
040403041618132019131317251214150302030306050504040404050606060707101918121219261101010901016230708020302030713036220705050616230618161021242
05060810101517040602010308181319190703040606060605040505060710162630201412153826060302000151702020305141712110305101412070602204101124262
14171714301405021401020305060503121201020406060606060505060507122336382311219373316090503060905020406051417120150702020304000002130512133523
101412240918030410040405050320104030203030606060606060507181020204472918171233535201006007070905050712170603000101010000621003012133523
13090708100804030302020308040010303030305060606050812050606070202233730181712547502509060606121507080609110700030201020201002060308180870996
151409070403020101010101010804002030405040405060705050403060602031291020191725443190906007121412080716170801010202020100000008131162976
2109030201010101000000000010102030405060606060609111030407090304203071716227231813110909081012116223516060000202010000000005061523311
0802120700000000000102030304050506070809070917322304823080004060811152535418425171109060709929429110400102010102020101000061005010050
010205030101010102030203040405060607080809090704070818010413110002163424214644437272015110907090728442415980203010001010202010010101071
0200201040403081207050304060606070809091009090906040302020105160105173091044494536312419131108050605183423226140704010000000010101000101010
00020910707040711130904061006060707080809101010100905030302020206161513021210044757494135282216151306050513362128241009040302000000000000101000
00000306050604020203020516070707080808091010100806020201010318042535020400195463514437292318121609040509293262612090702020101000000010100000
0000010405060300000202161405060606060707080910090606030202010127090304010101275349484739312318121015070504153336220813070402010100000010201000
00000001020401000010014309203030304040405060704040202010113060100000010936373535292116130906080703040916301604130604030101000000000501000
```

Object class recognition



Face recognition



Action recognition

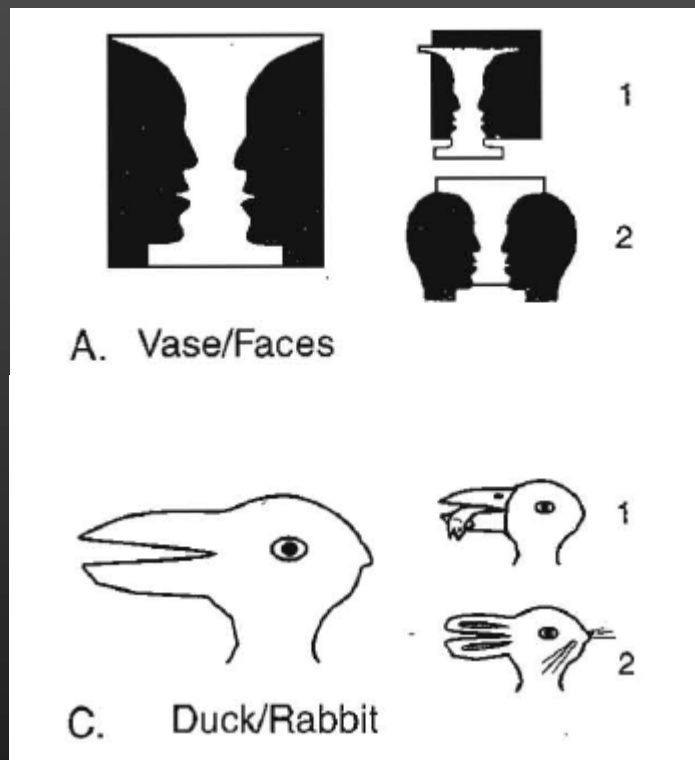


3D Scene reconstruction



Fundamental problem II: Images do not measure the meaning

→ 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

Specific object detection



(Lowe, 2004)

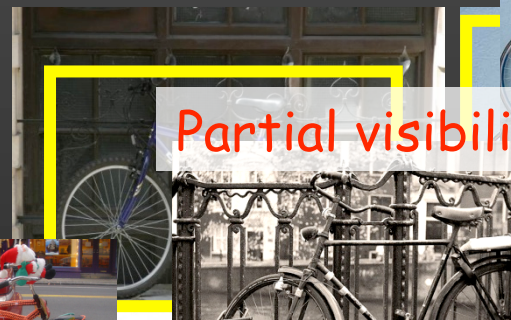
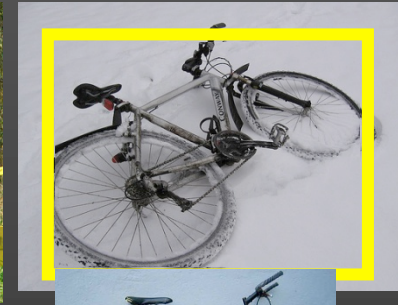
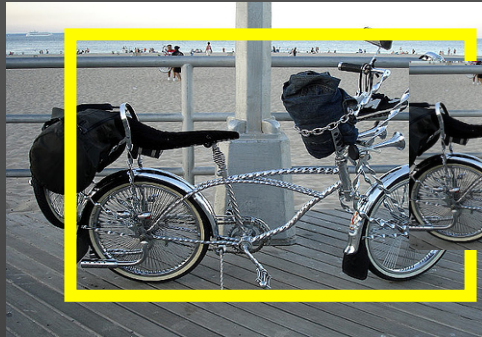
Image classification



Caltech 101 : http://www.vision.caltech.edu/Image_Datasets/Caltech101/

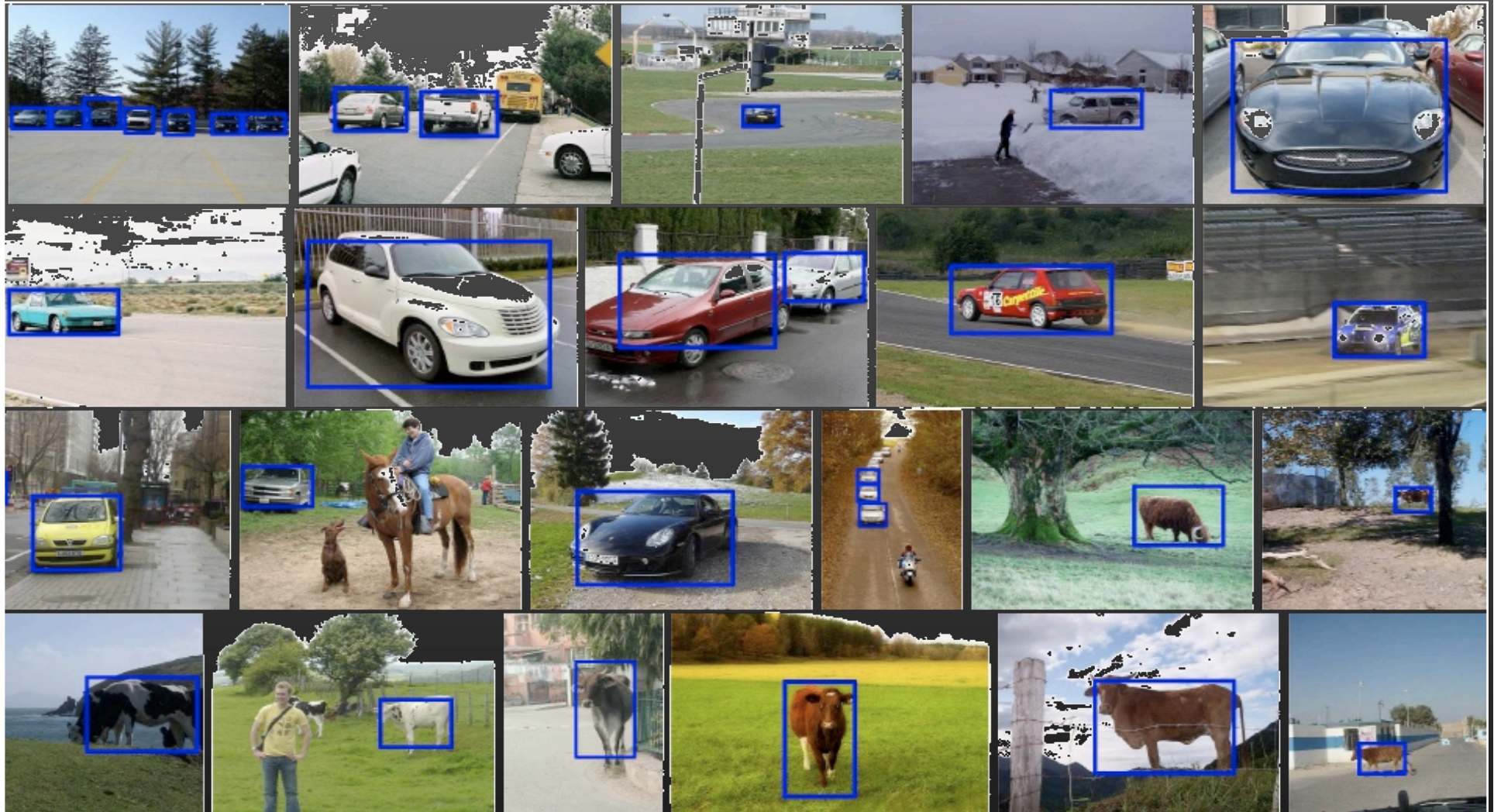
Object category detection

View variation



Within-class variation

Model \equiv locally rigid assembly of parts
Part \equiv locally rigid assembly of features



Qualitative experiments on Pascal VOC'07 (Kushal, Schmid, Ponce, 2008)

Scene understanding

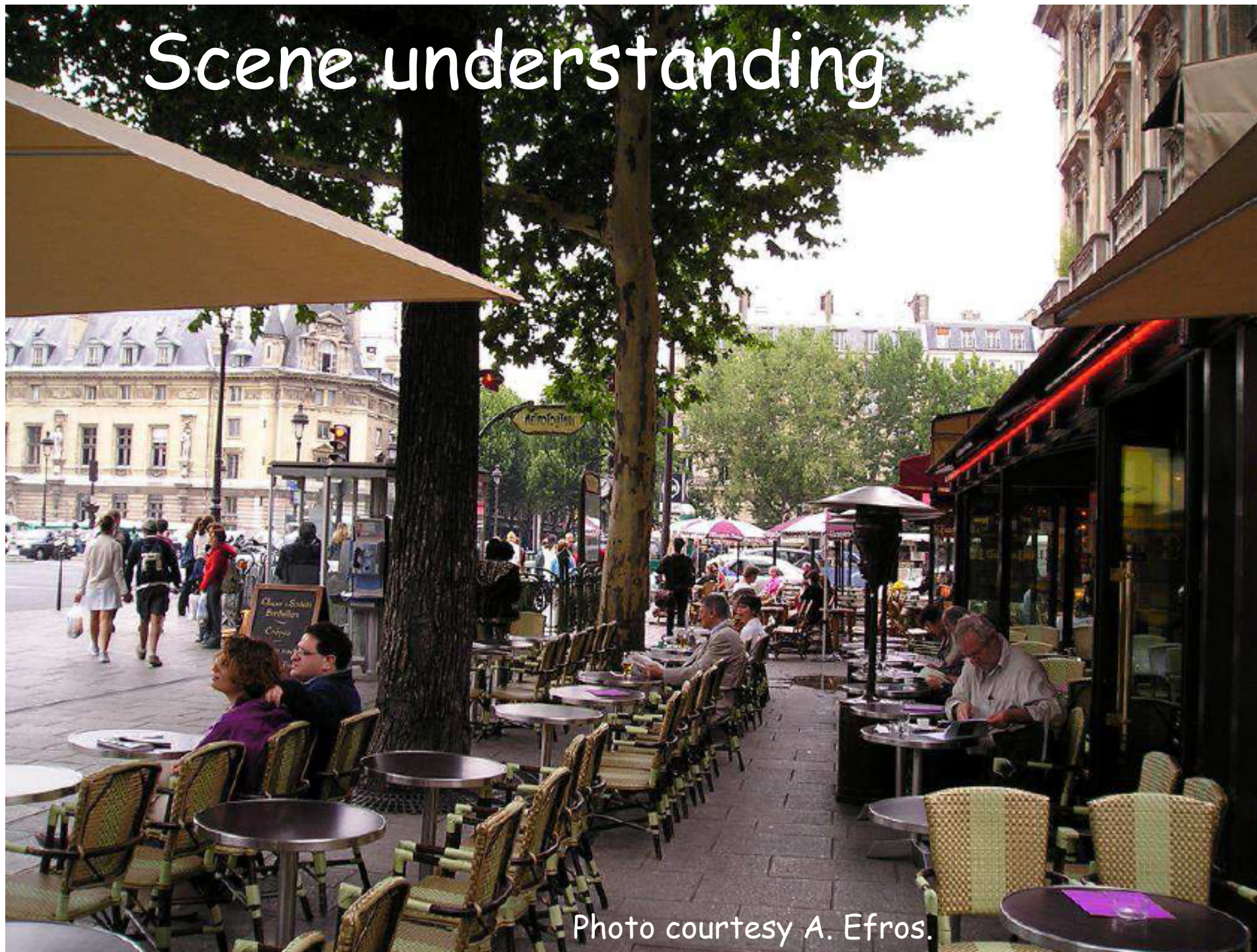
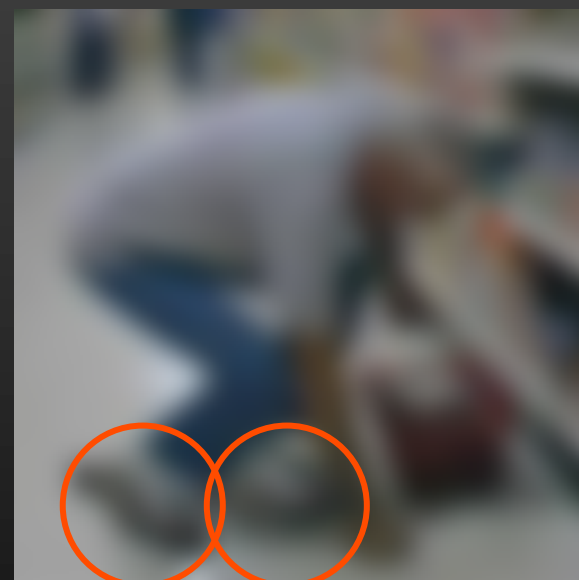
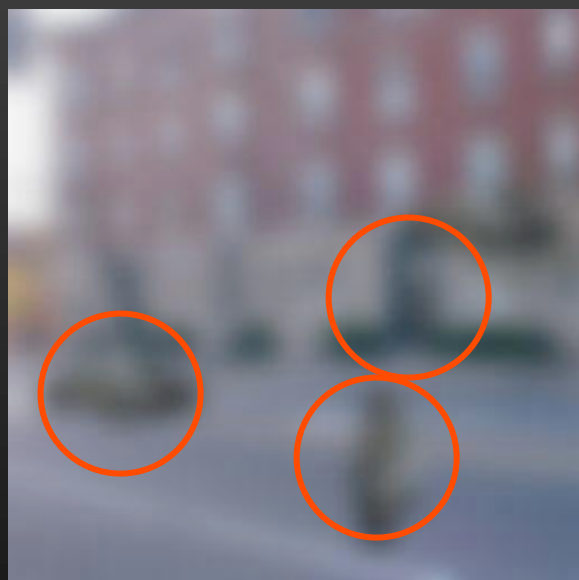
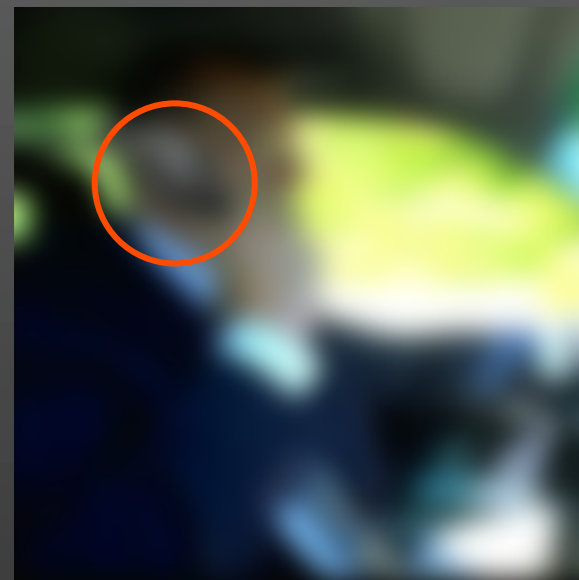
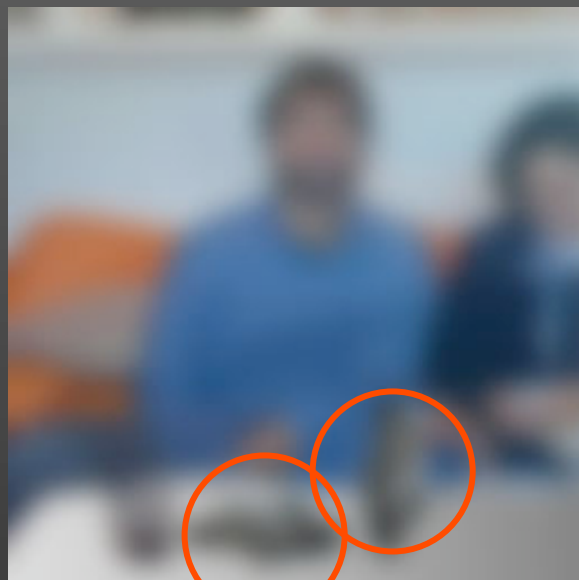
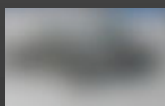


Photo courtesy A. Efros.

Local ambiguity and global scene interpretation



This class

1. Introduction plus recap on geometry (J. Ponce)
2. Instance-level recognition I. - Local invariant features (C. Schmid)
3. Instance-level recognition II. - Correspondence, efficient visual search (J. Sivic)
4. Very large scale image indexing. Bag-of-feature models for category-level recognition (C. Schmid)
5. Sparse coding and dictionary learning for image analysis (J. Ponce)
6. Part-based models and pictorial structures for object recognition (J. Sivic)
7. Motion and human actions I. (I. Laptev)
8. Motion and human actions II. (I. Laptev)
9. Neural networks; Optimization methods (J. Ponce)
10. Category level localization; Face detection and recognition (C. Schmid)
11. Multiple object categories; Context; Recognizing large number of object classes; Segmentation (I. Laptev, J. Sivic)
12. Final project presentations (J. Sivic, I. Laptev)

Computer vision books

- D.A. Forsyth and J. Ponce, "Computer Vision: A Modern Approach, Prentice-Hall, 2003.
- 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.
- O. Faugeras, Q.T. Luong, and T. Papadopoulos, "Geometry of Multiple Images," MIT Press, 2001.
- R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", Cambridge University Press, 2004.
- J. Koenderink, "Solid Shape", MIT Press, 1990.

Class web-page

<http://www.di.ens.fr/willow/teaching/recvis10>

Slides available after classes:

<http://www.di.ens.fr/willow/teaching/recvis10/lecture1.pptx>

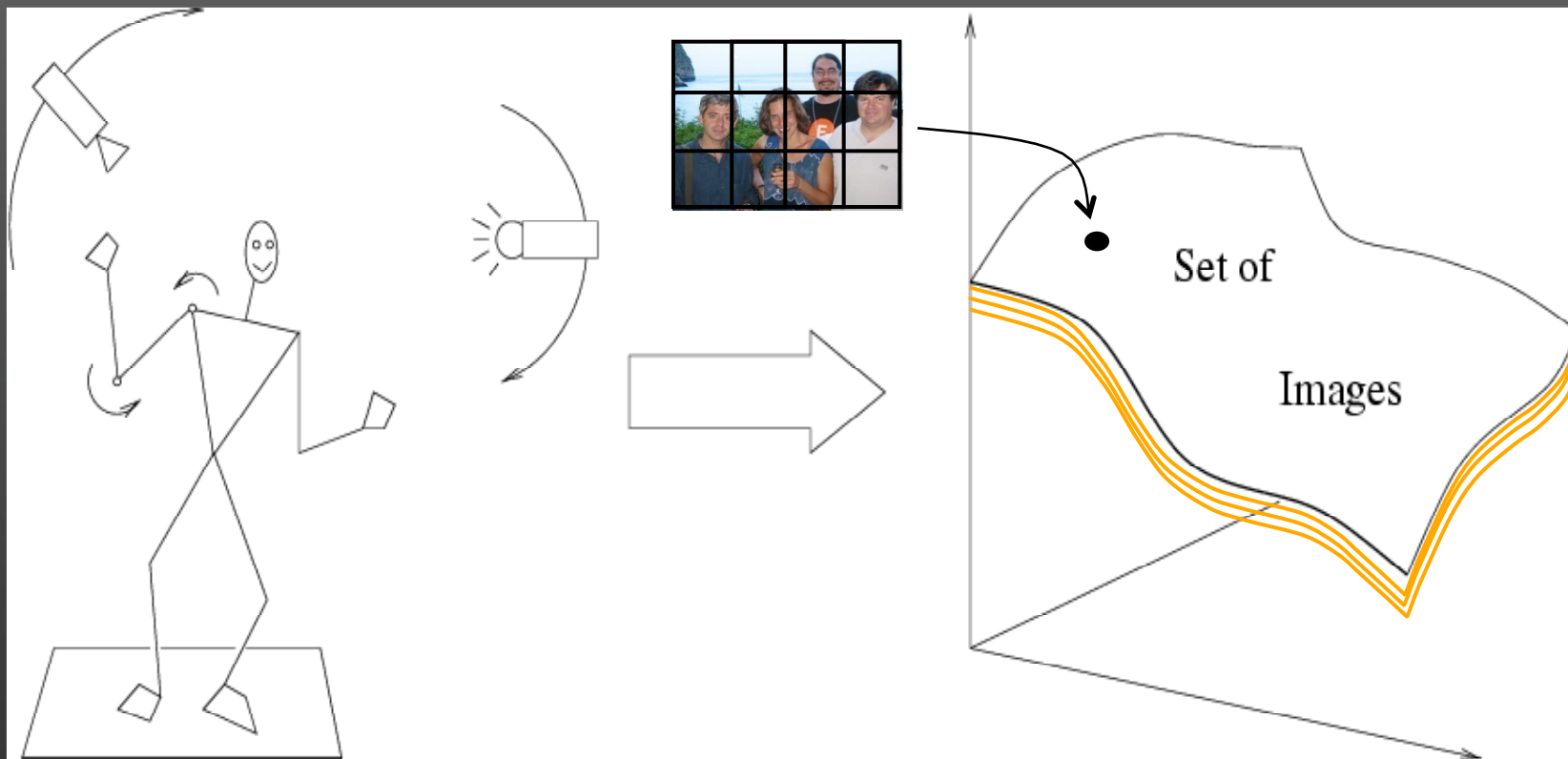
<http://www.di.ens.fr/willow/teaching/recvis10/lecture1.pdf>

Note: Much of the material used in this lecture is courtesy of Svetlana Lazebnik,

<http://www.cs.unc.edu/~lazebnik/>

Outline

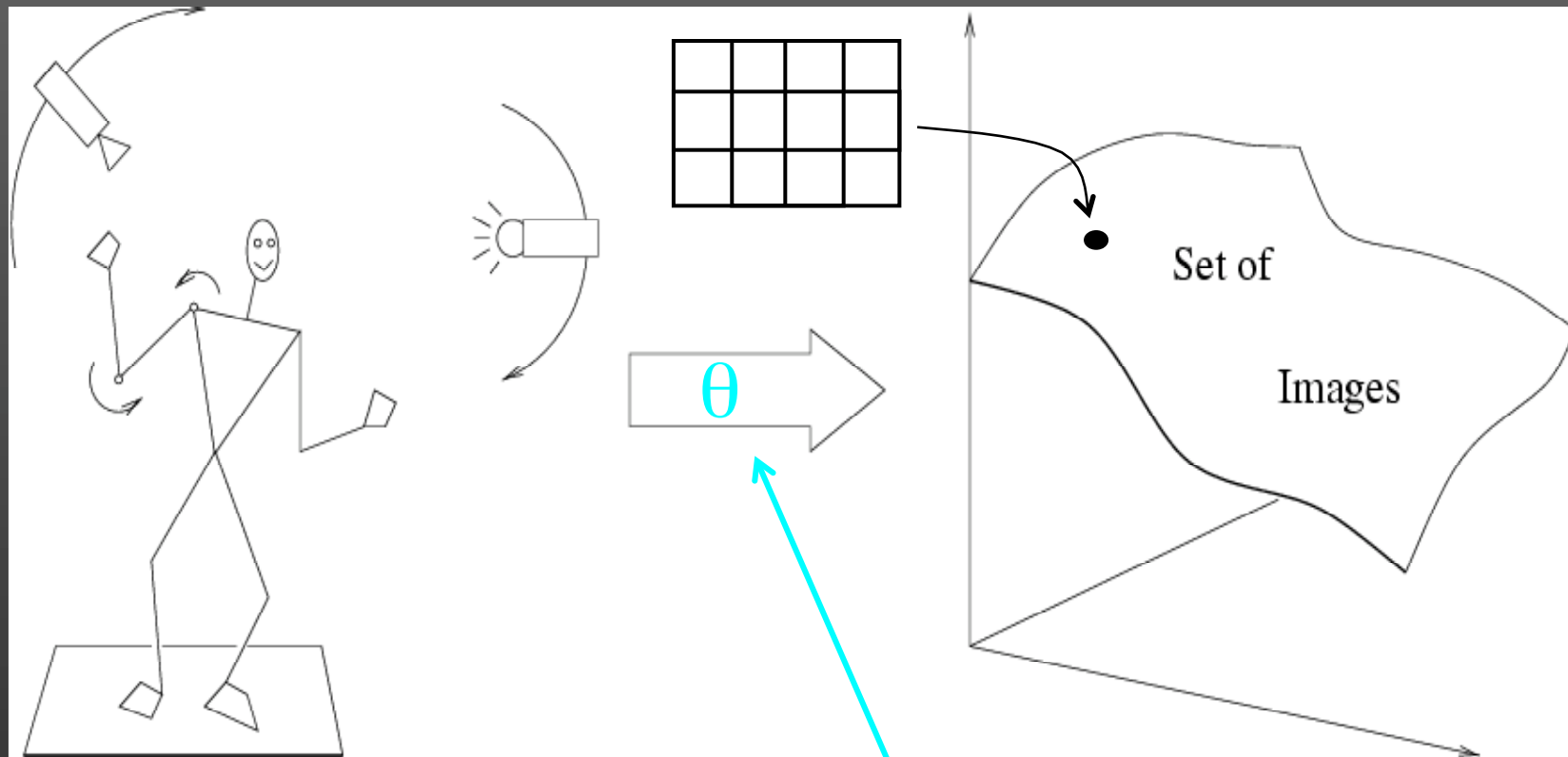
- What computer vision is about
- What this class is about
- A brief history of visual recognition
- A brief recap on geometry



Variability:

Camera position
Illumination
Internal parameters
Within-class variations



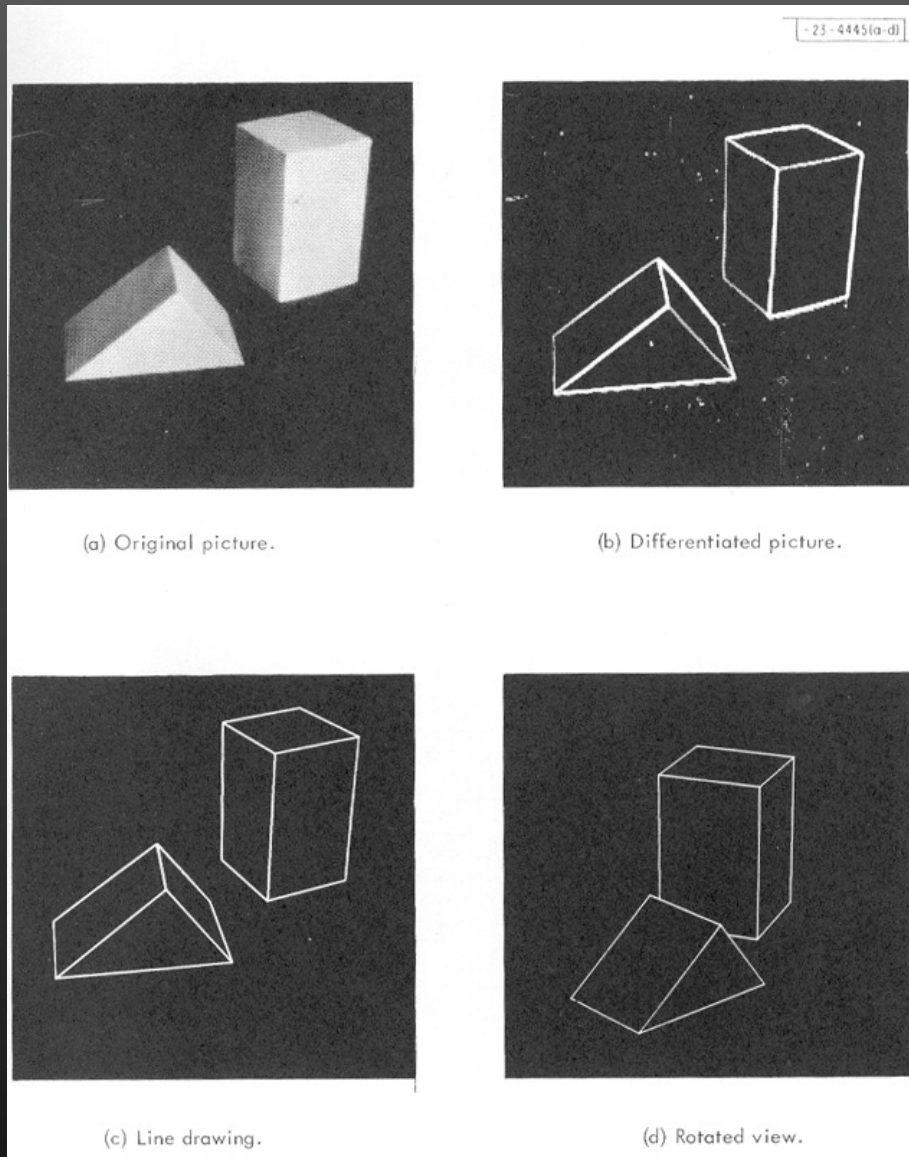


Variability:

Camera position
Illumination
Internal parameters

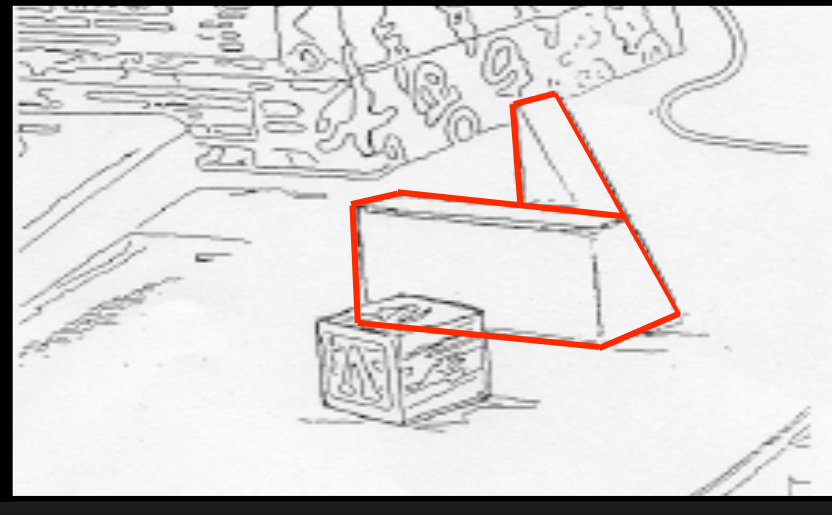
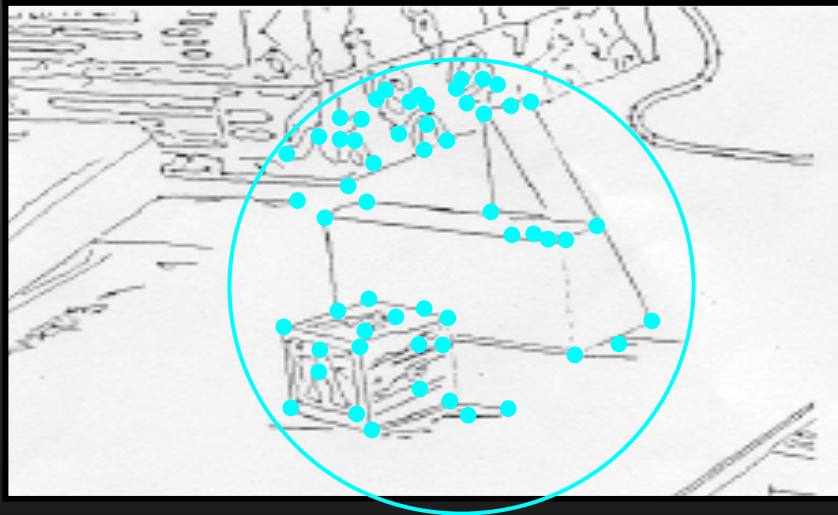
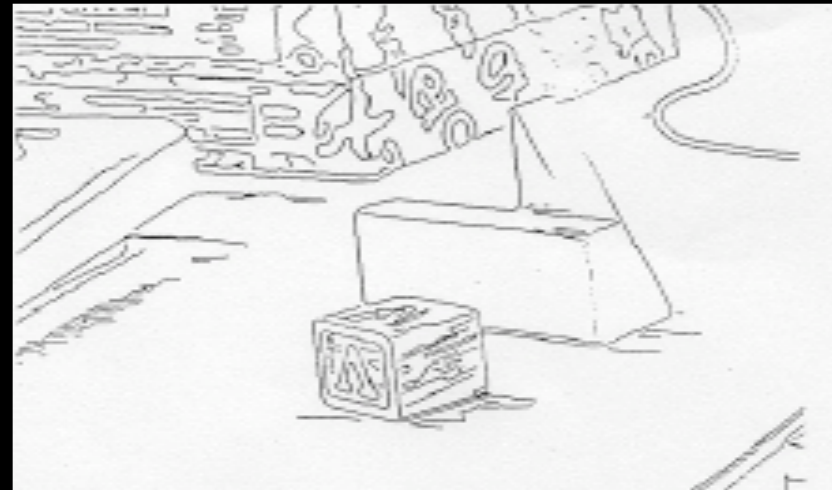
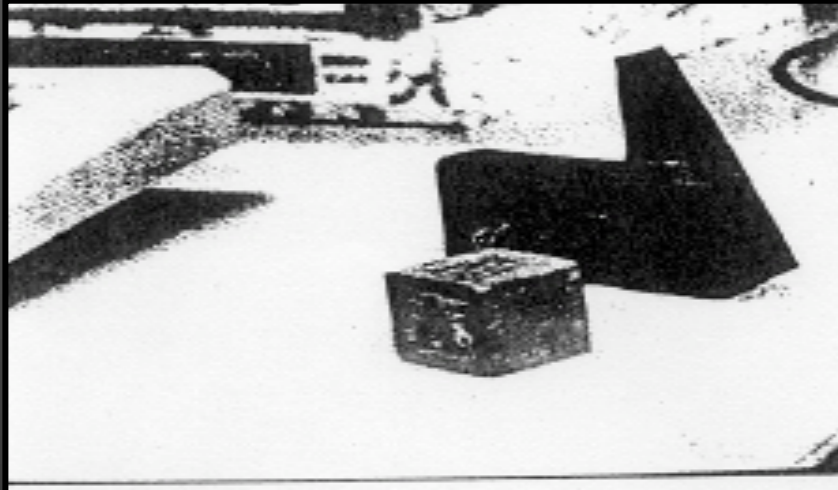
Roberts (1963); Lowe (1987); Faugeras & Hebert (1986); Grimson & Lozano-Perez (1986); Huttenlocher & Ullman (1987)

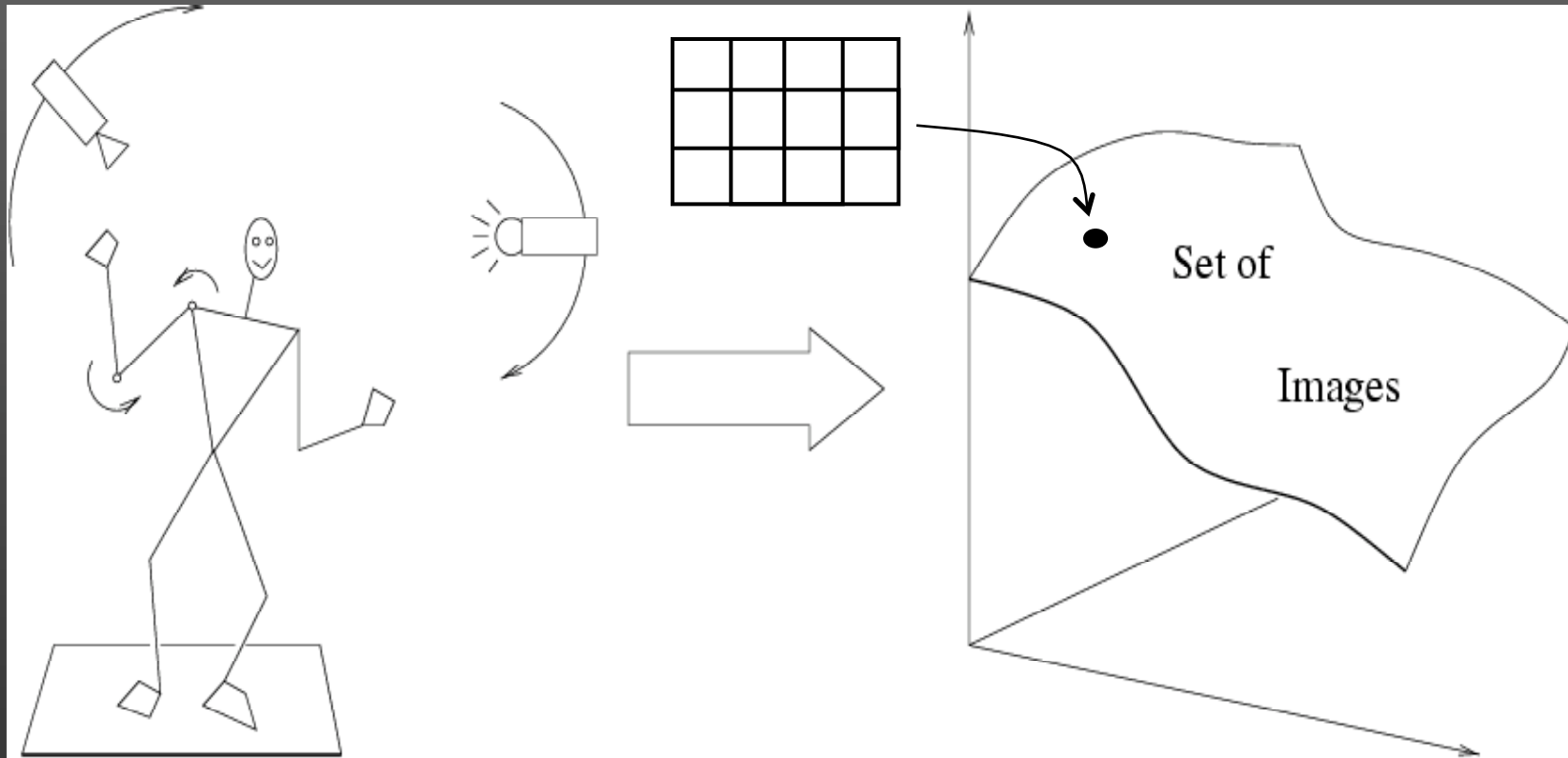
Origins of computer vision



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

Huttenlocher & Ullman (1987)





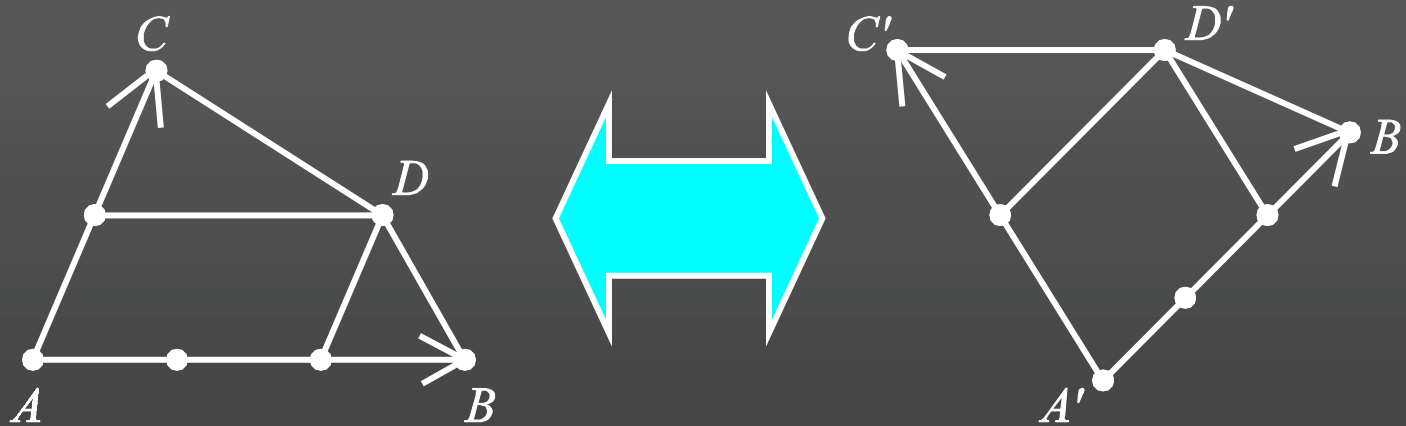
~~Variability~~

Invariance to:

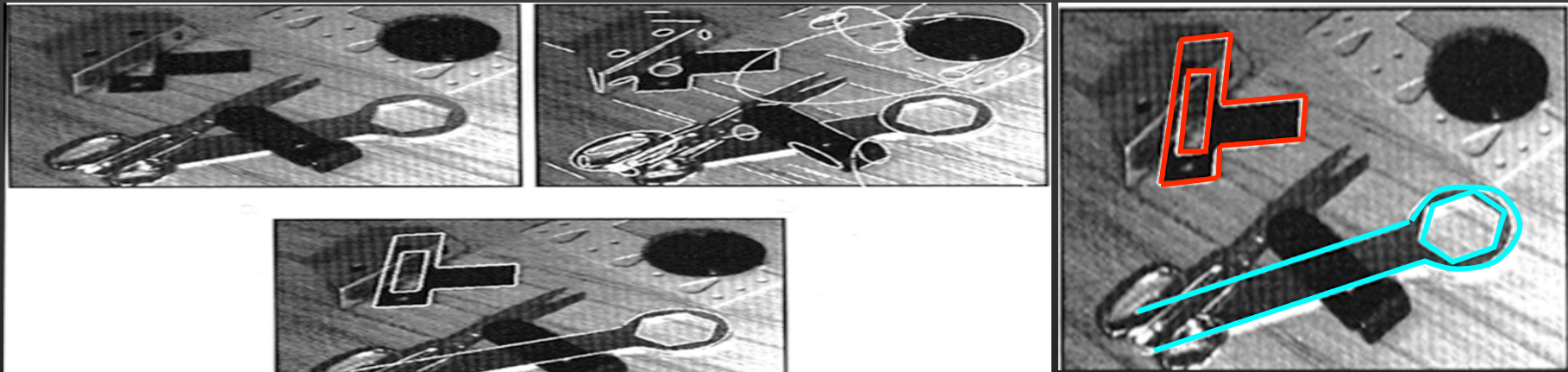
Camera position
Illumination
Internal parameters

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

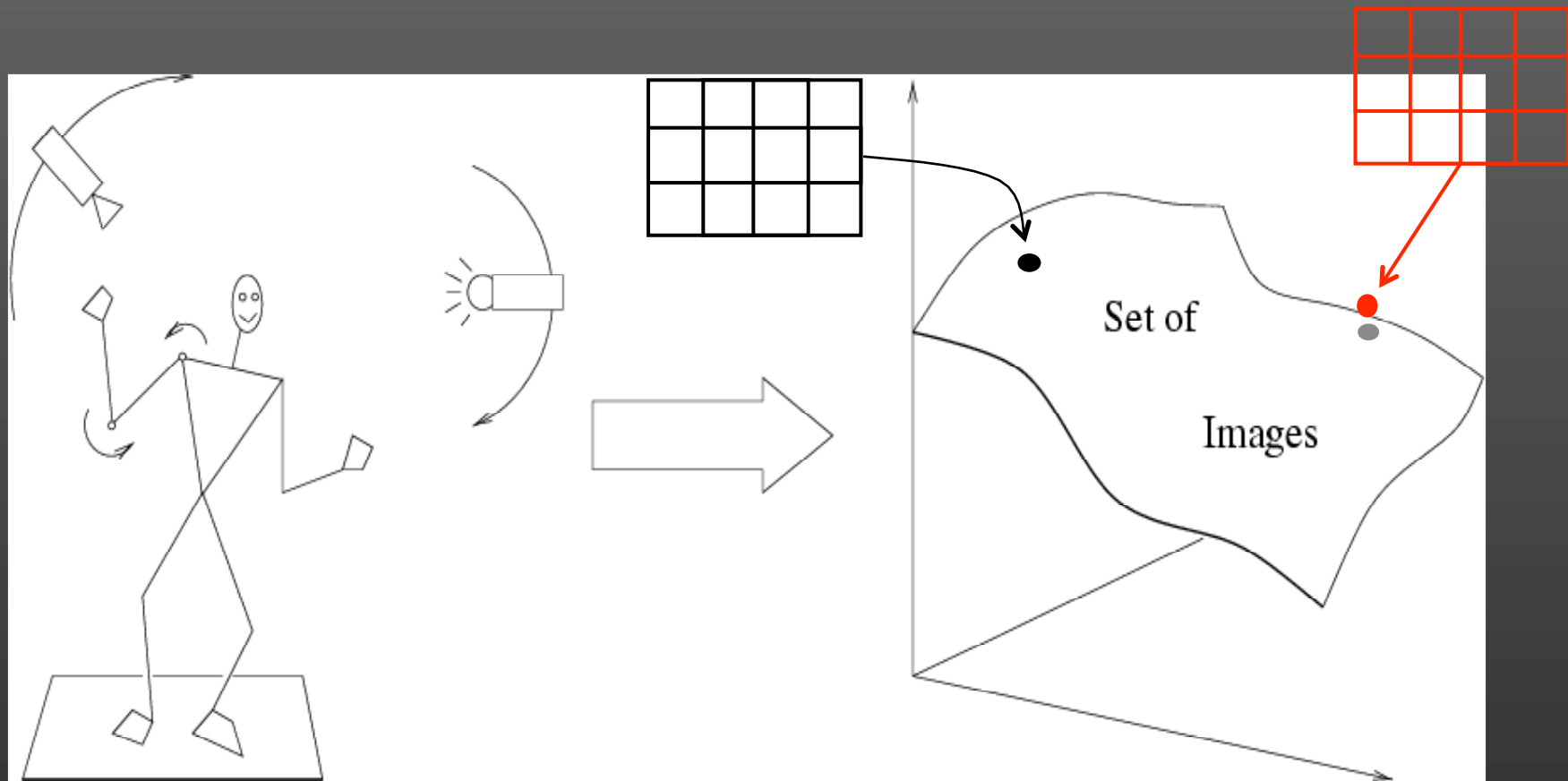
Example: affine invariants of coplanar points



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:

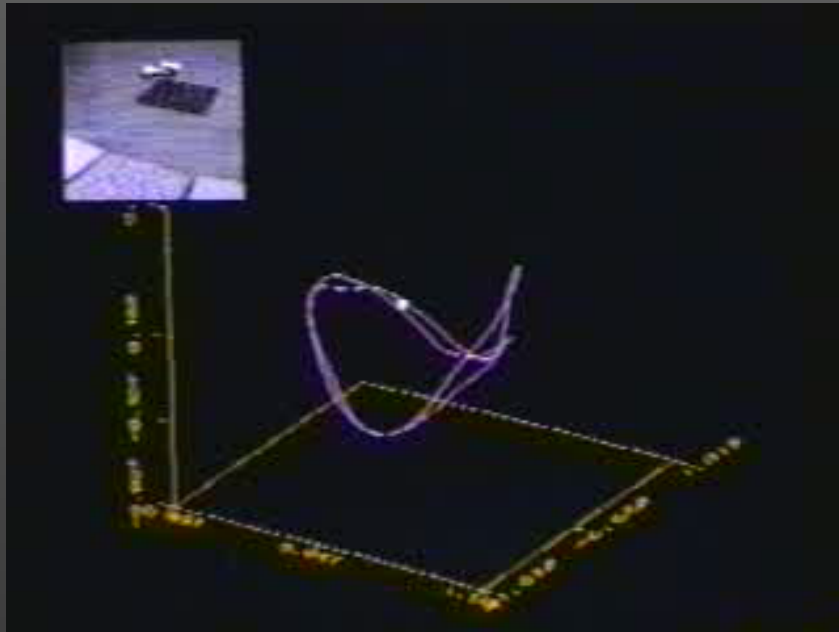
Appearance-based techniques

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

Eigenfaces (Turk & Pentland, 1991)



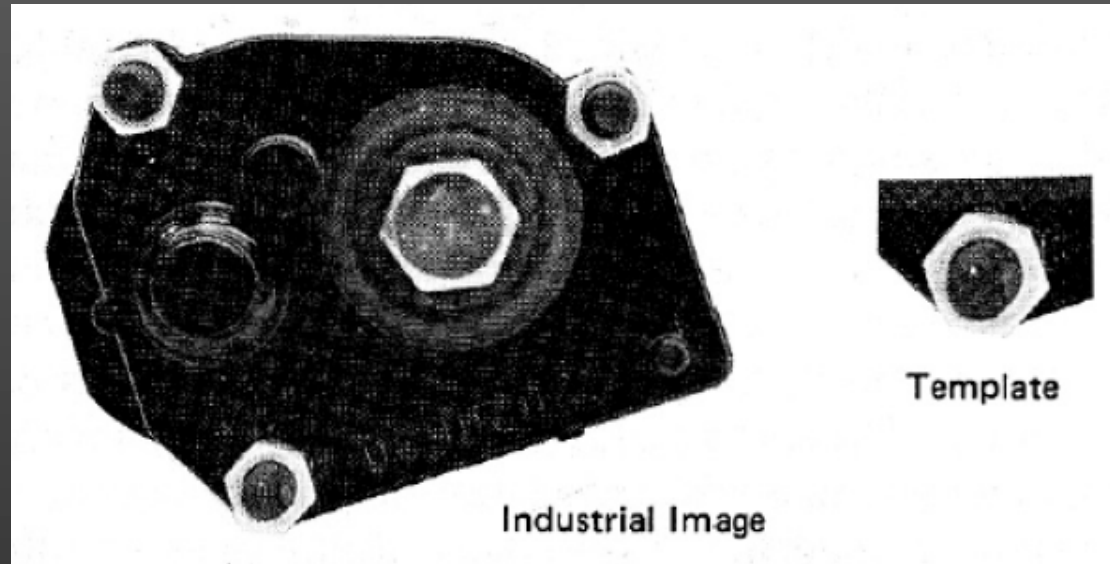
Experimental Condition	Correct/Unknown Recognition Percentage		
	Lighting	Orientation	Scale
Forced classification	96/0	85/0	64/0
Forced 100% accuracy	100/19	100/39	100/60
Forced 20% unknown rate	100/20	94/20	74/20



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)

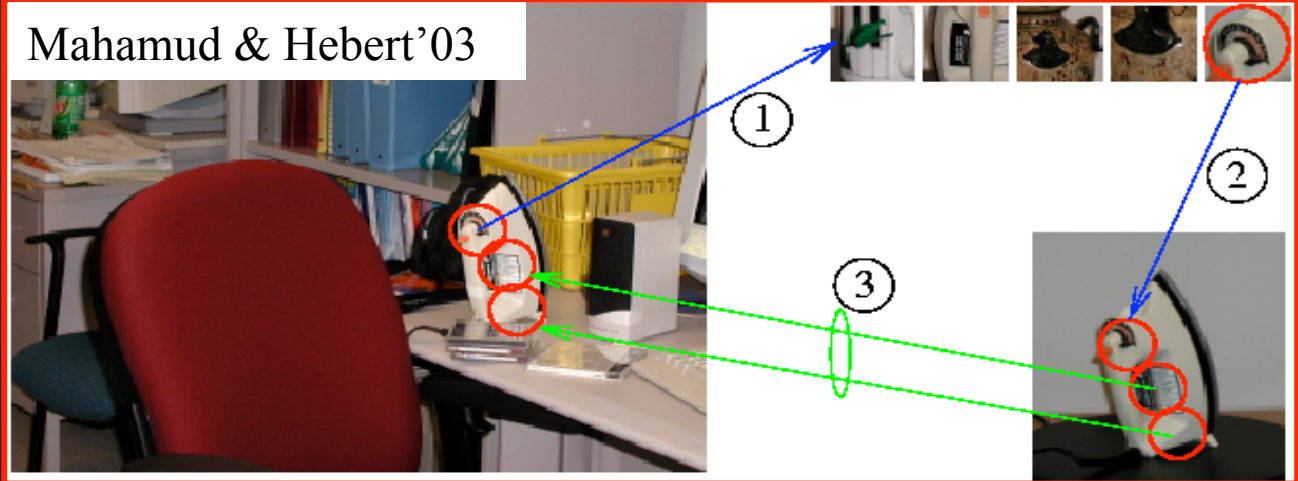


Schmid & Mohr'97

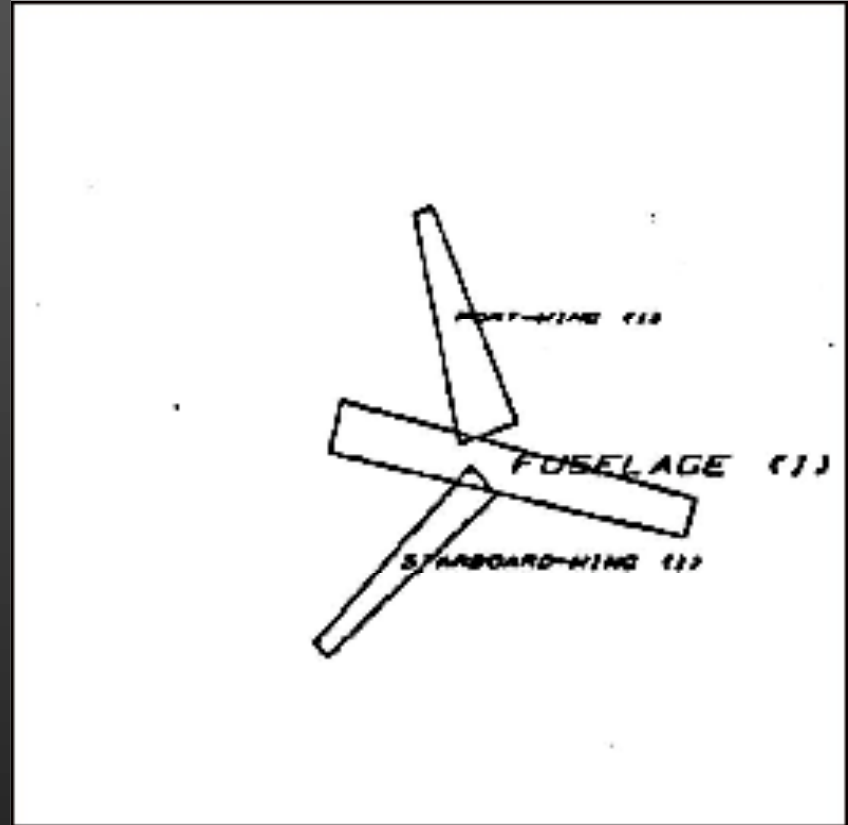
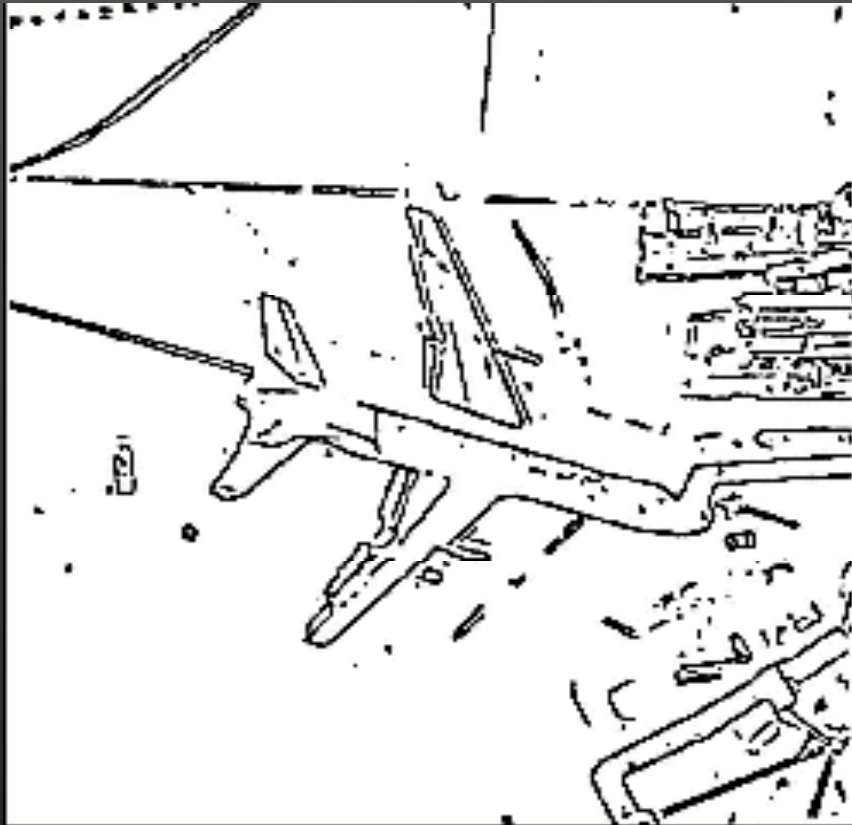
Lowe'02



Mahamud & Hebert'03



Representing and recognizing object categories is harder

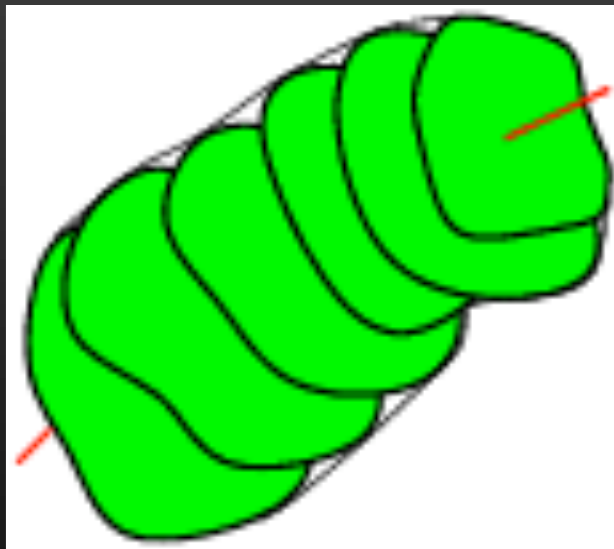
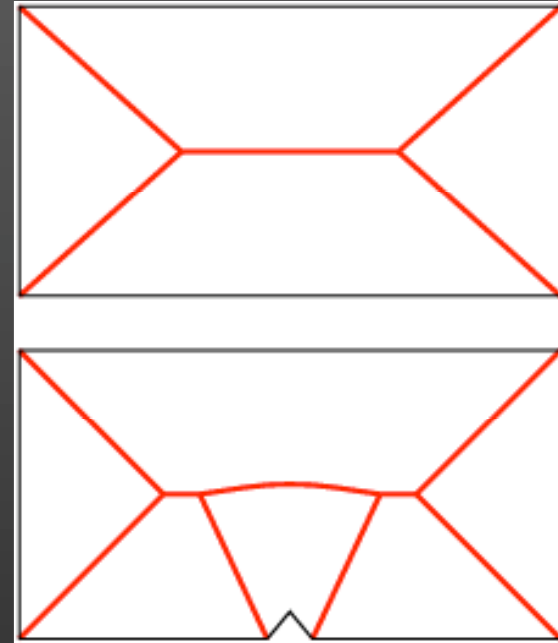
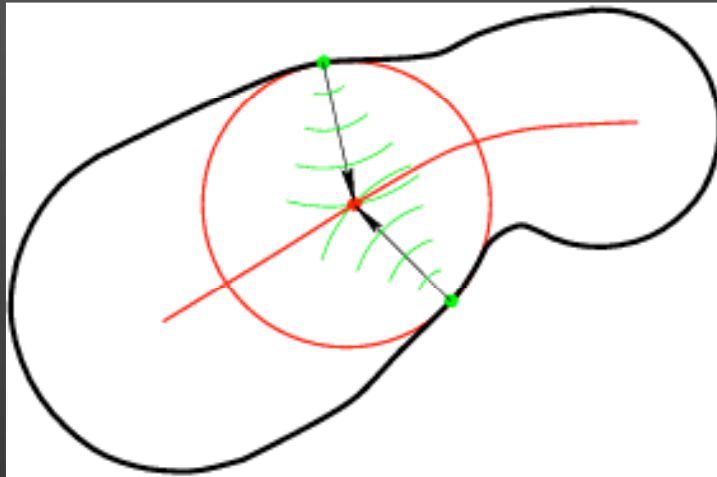


ACRONYM (Brooks and Binford, 1981)

Binford (1971), Nevatia & Binford (1972), Marr & Nishihara (1978)

Parts and invariants

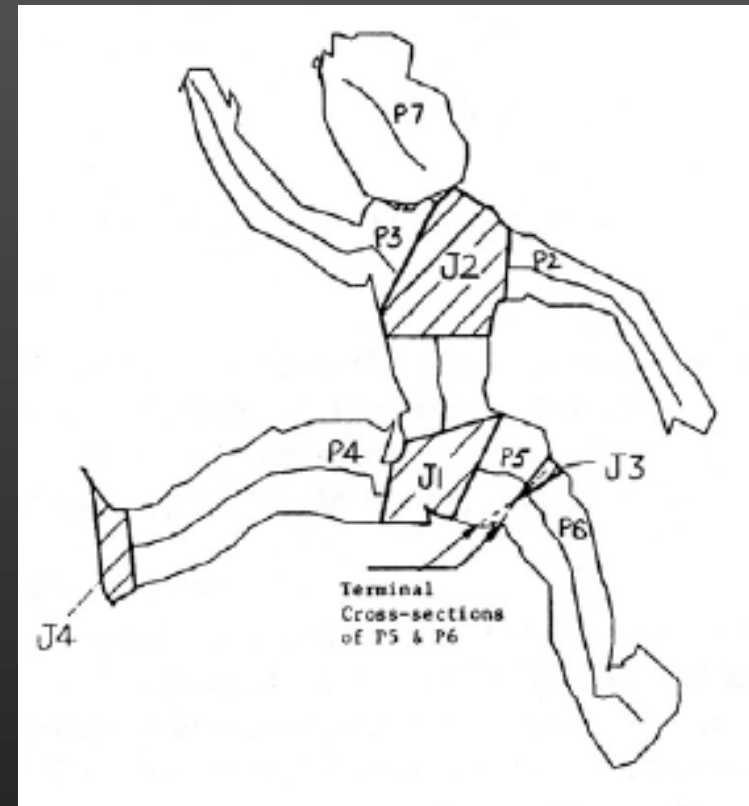
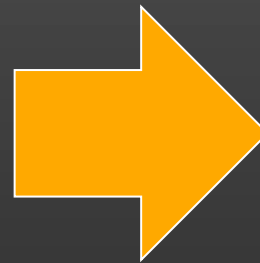
The Blum transform, 1967



Generalized cylinders
(Binford, 1971)

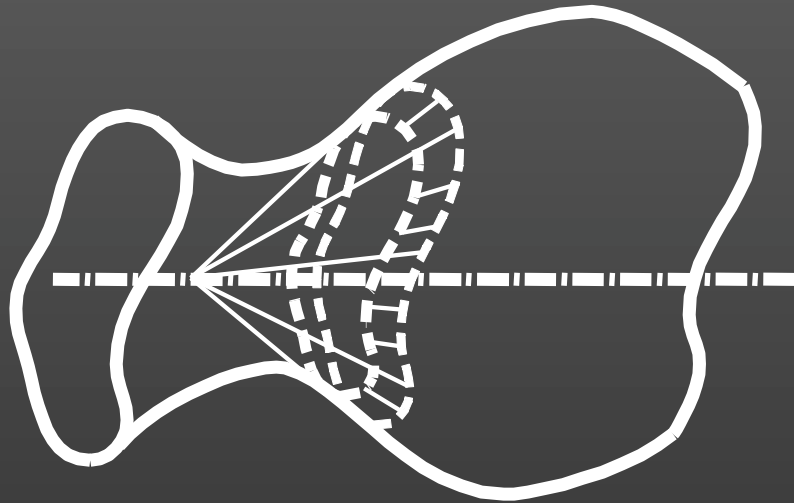
Generalized cylinders

(Binford, 1971; Marr & Nishihara, 1978)

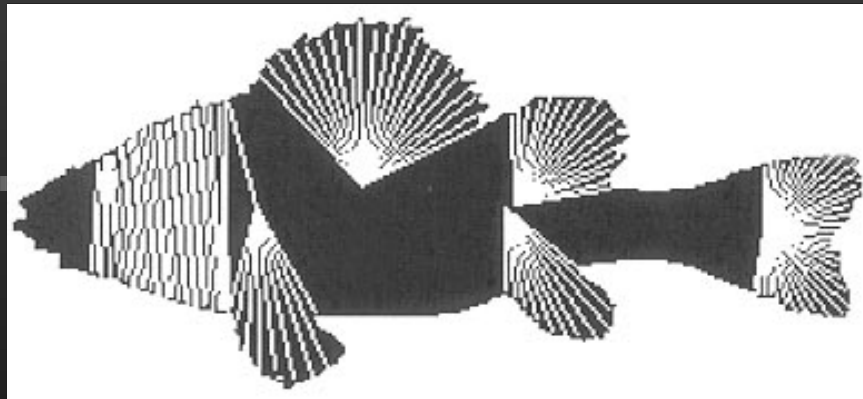


(Nevatia & Binford, 1972)

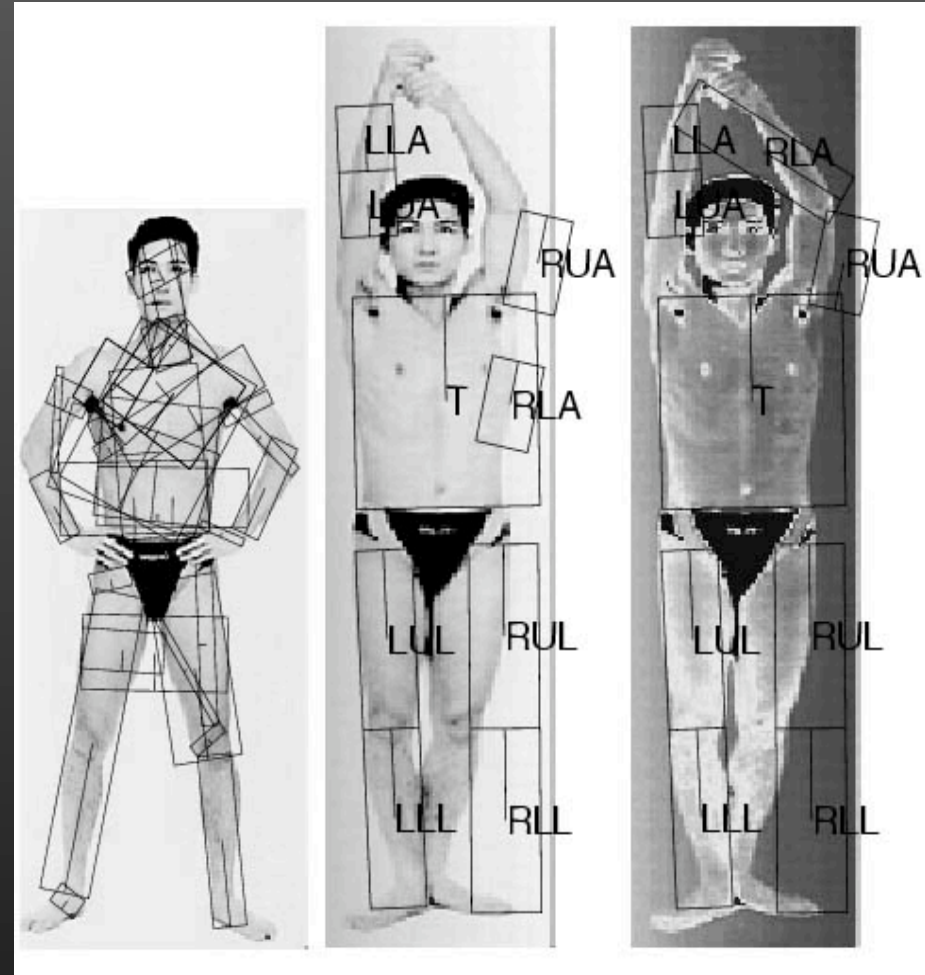
Parts and invariants II



Ponce et al. (1989)

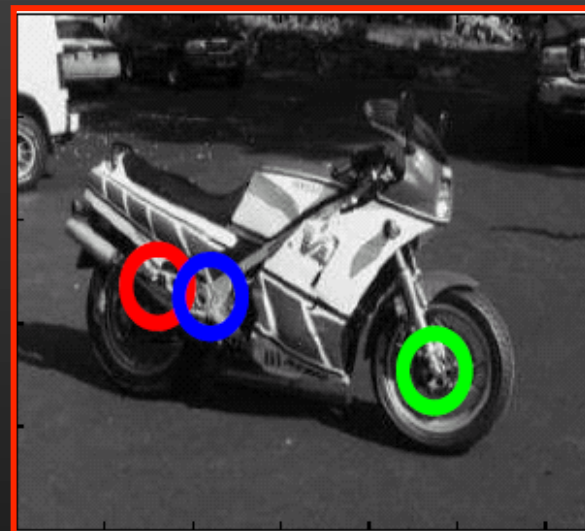
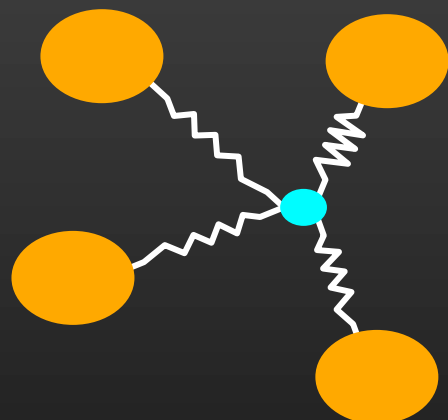
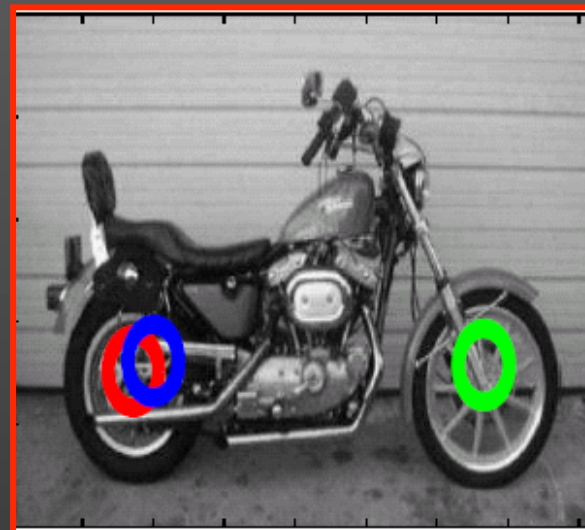


Zhu and Yuille (1996)



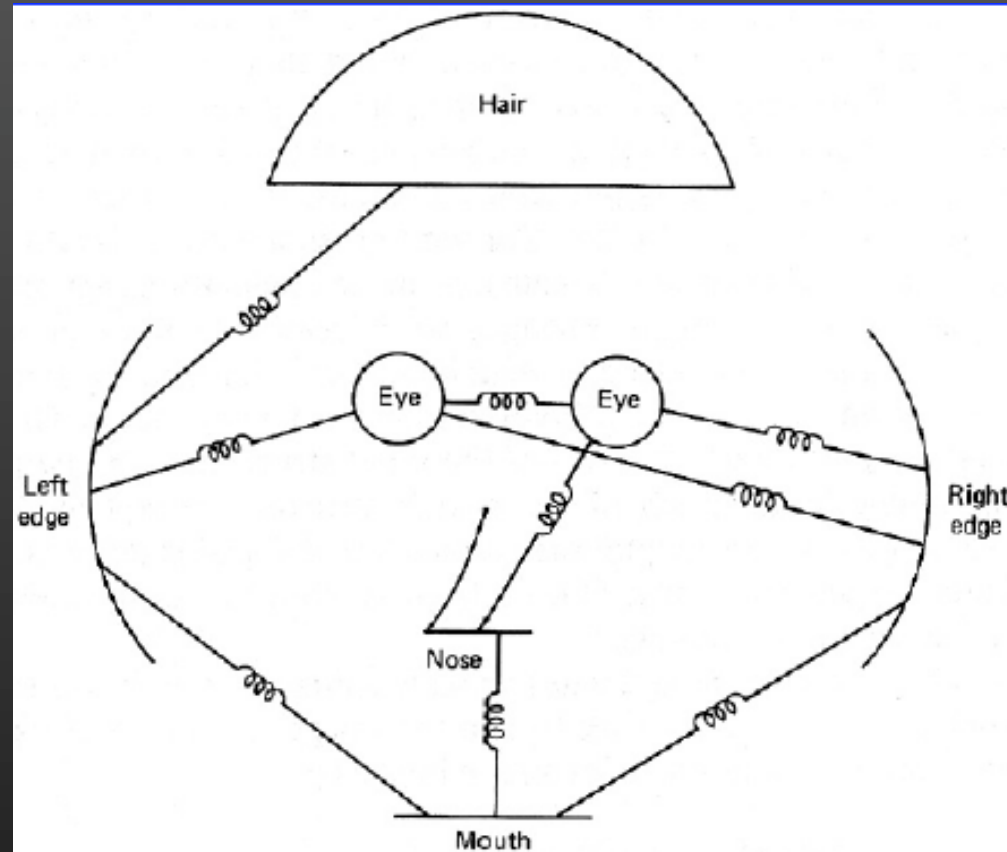
Ioffe and Forsyth (2000)

In the early 2000's, a new approach ?



Fergus, Perona & Zisserman (2003)

The "templates and springs" model (Fischler & Elschlager, 1973)



Ballard & Brown (1980, Fig. 11.5). Courtesy
Bob Fisher and Ballard & Brown on-line.

Object → Bag of 'words'



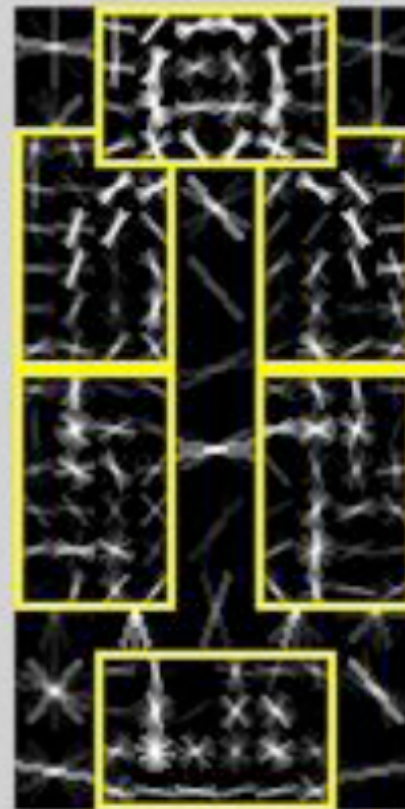
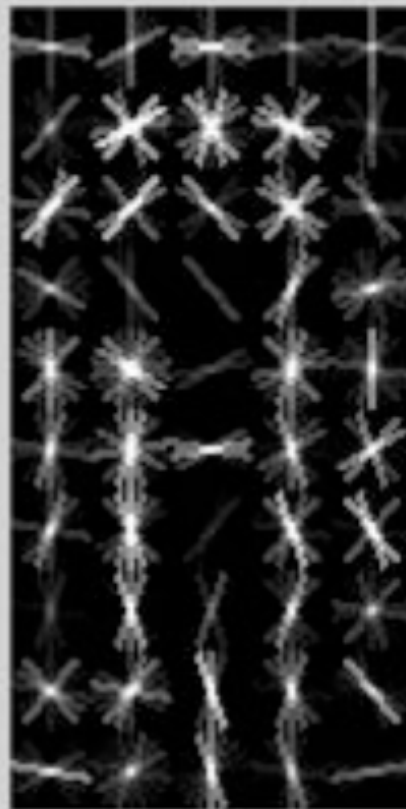


Color histograms (S&B'91)
Local jets (Florack'93)
Spin images (J&H'99)
Sift (Lowe'99)
Shape contexts (B&M'95)

Texton histograms (L&M'97)
Gist (O&T'05)
Spatial pyramids (LSP'06)
Hog (D&T'06)
Phog (B&Z'07)
Convolutional nets (LC'90)

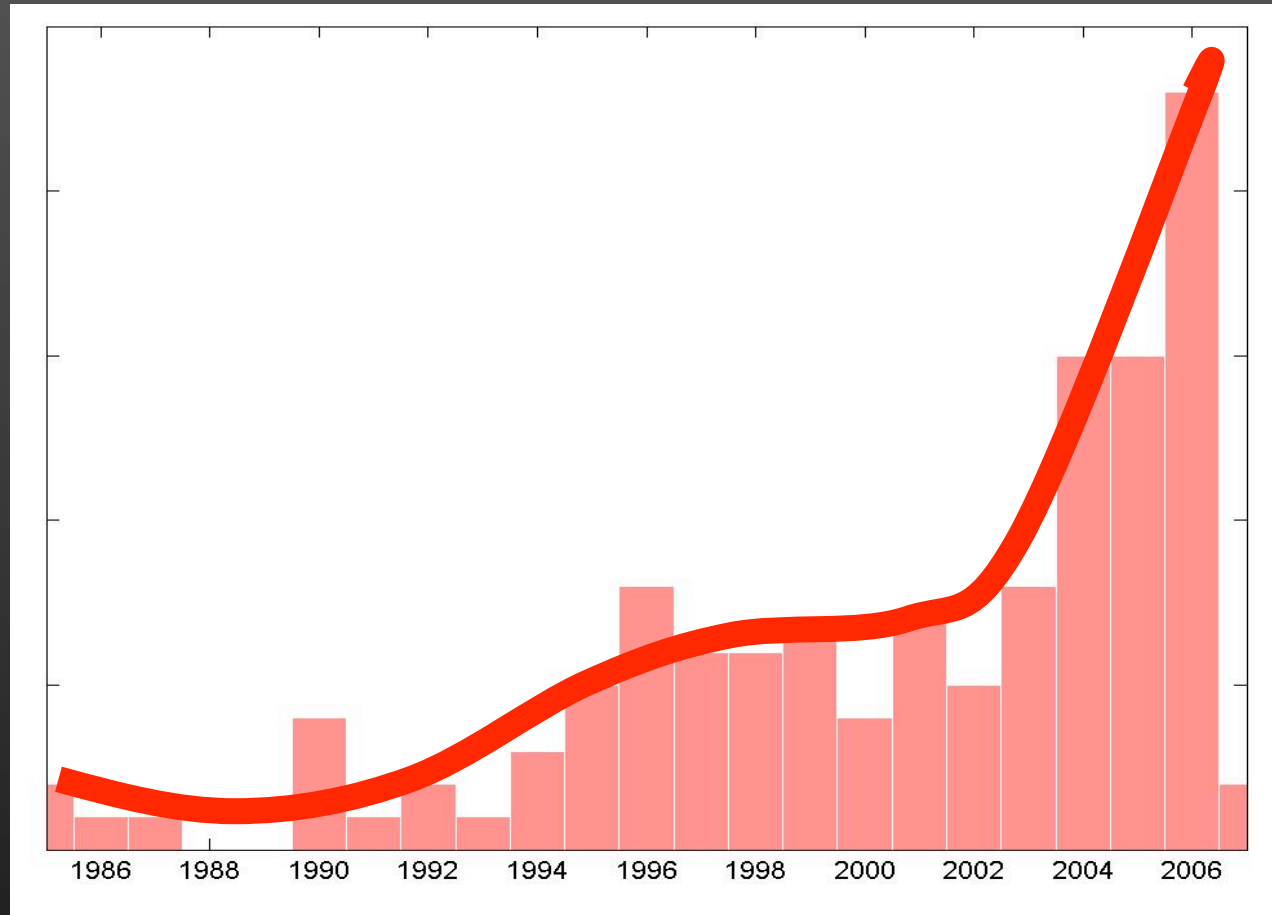


Locally orderless structure of images (K&vD'99)

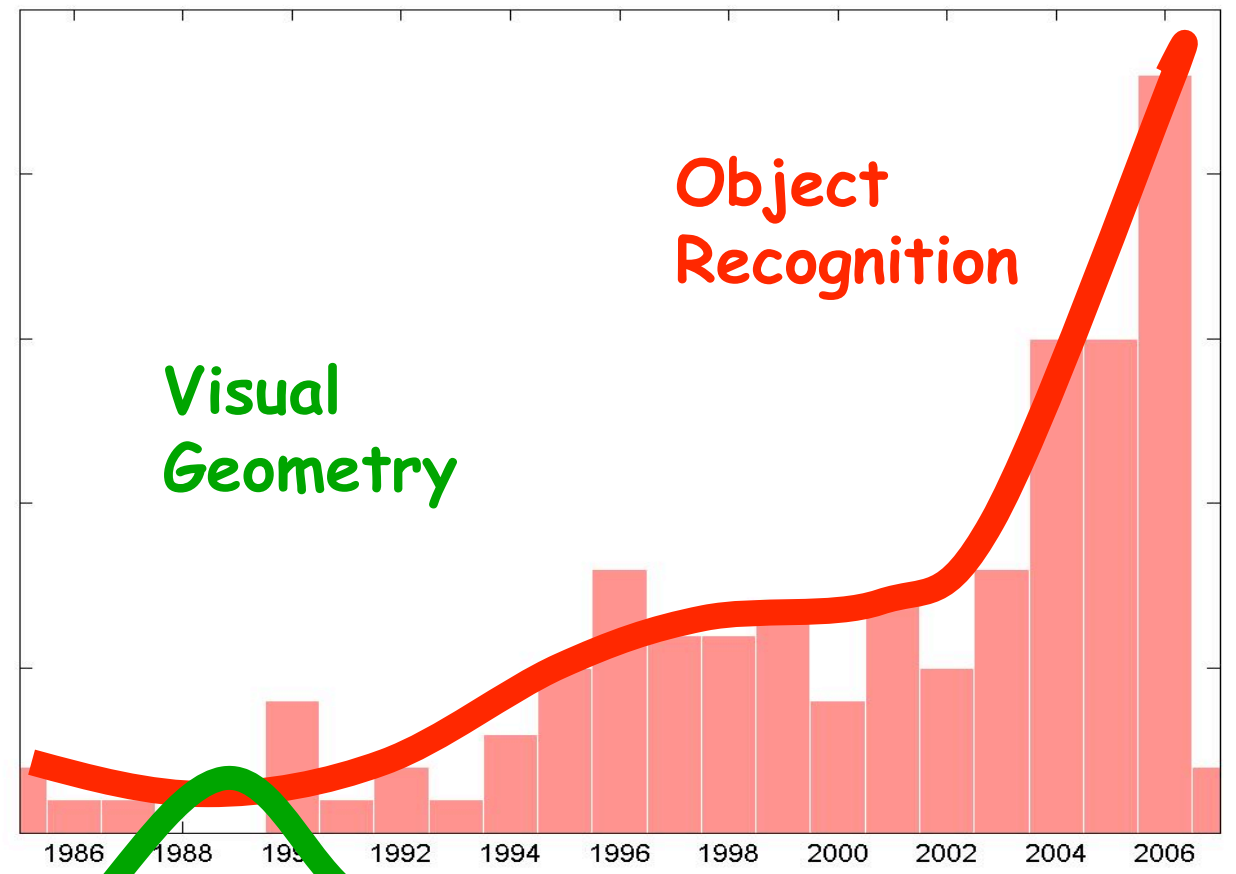
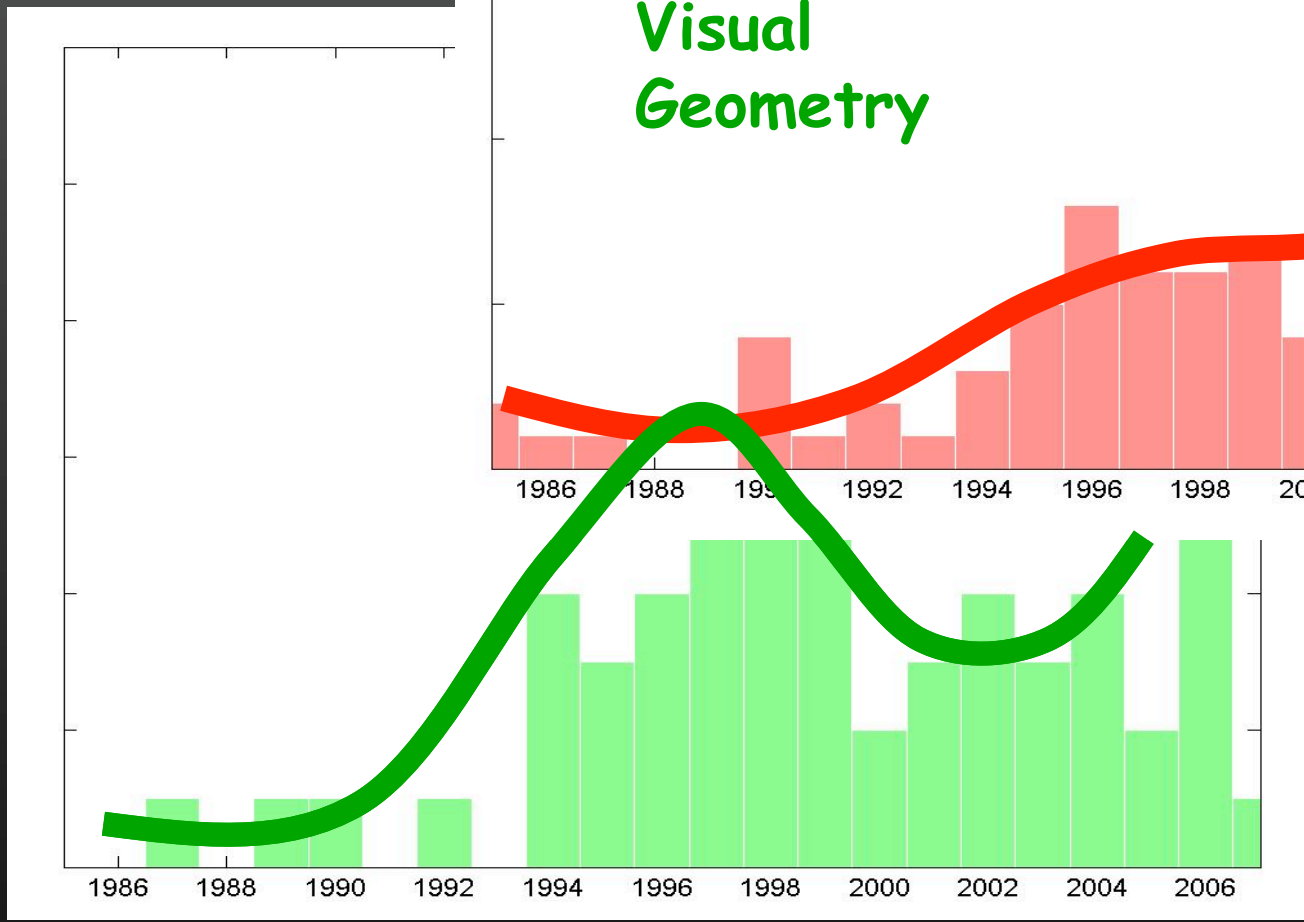


Felzenszalb, McAllester, Ramanan (2007)
[Wins on 6 of the Pascal'07 classes, see Chum
& Zisserman (2007) for the other big winner.]

Number of research papers with
key-words "object recognition",
source: Springer.com



Numbers of papers with key-words
"epipolar geometry"
source:
Springer.com



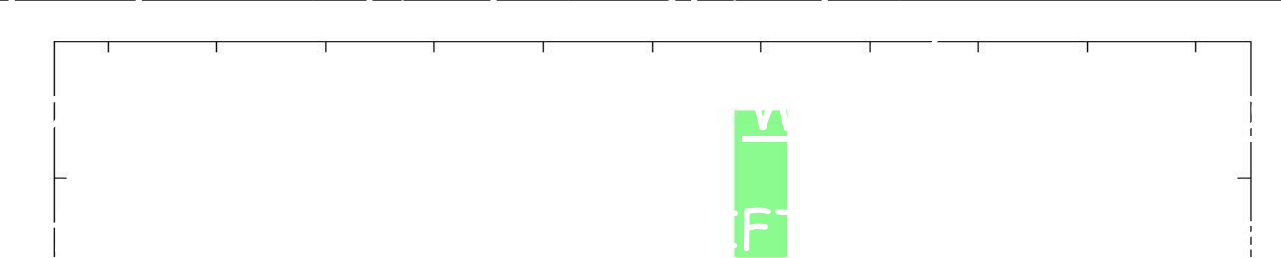
Visual Geometry:

Problems: Camera calibration, 3D reconstruction,

S

T

Scale



atching, ...

c.



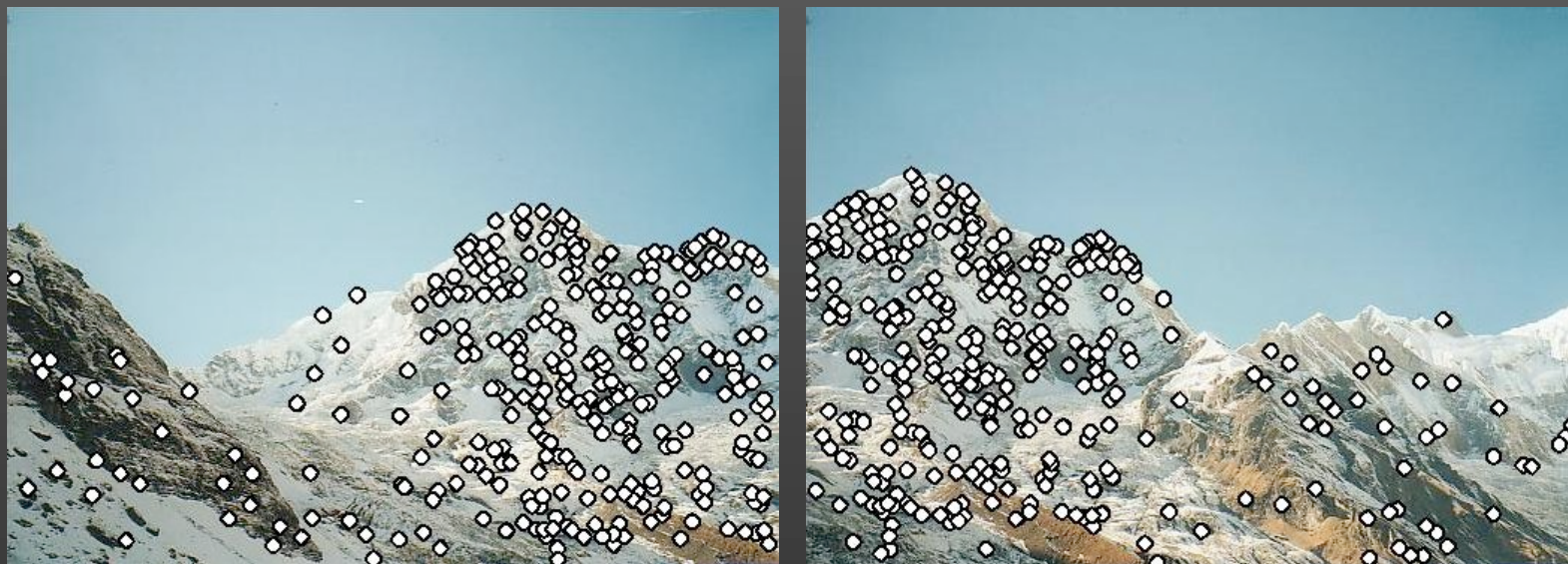
Outline

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

Feature-based alignment outline



Feature-based alignment outline



Extract features

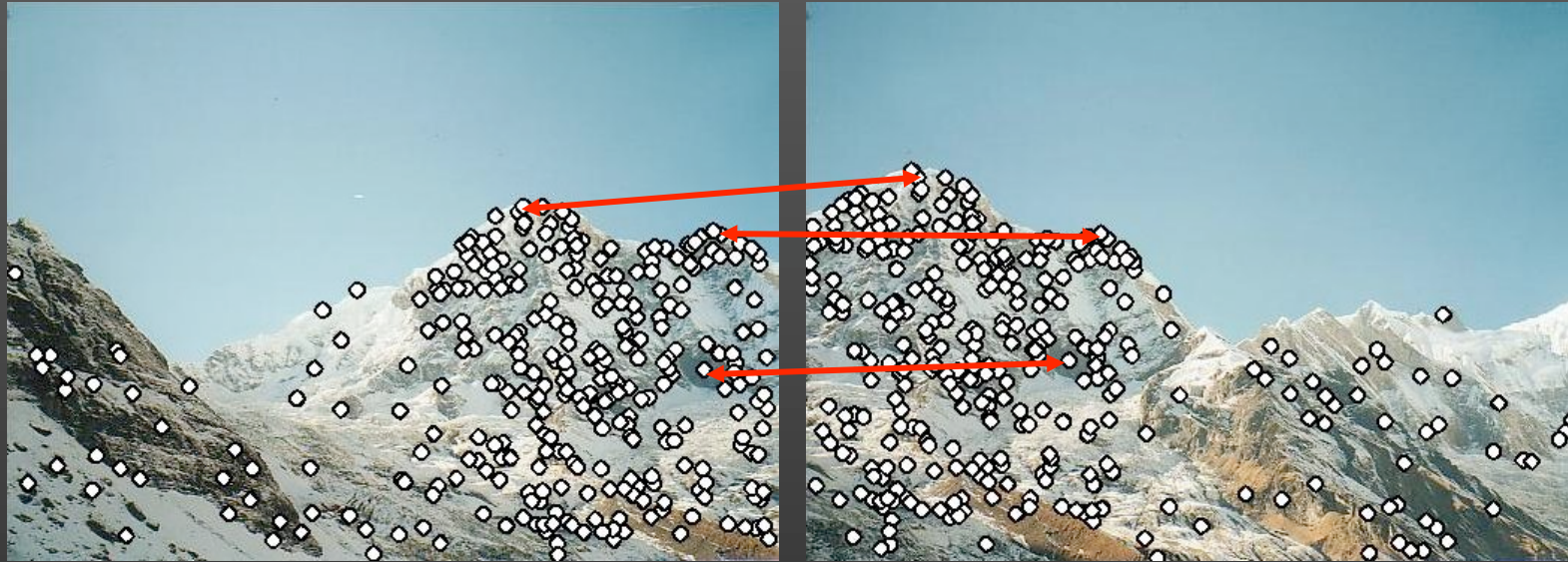
Feature-based alignment outline



Extract features

Compute *putative matches*

Feature-based alignment outline



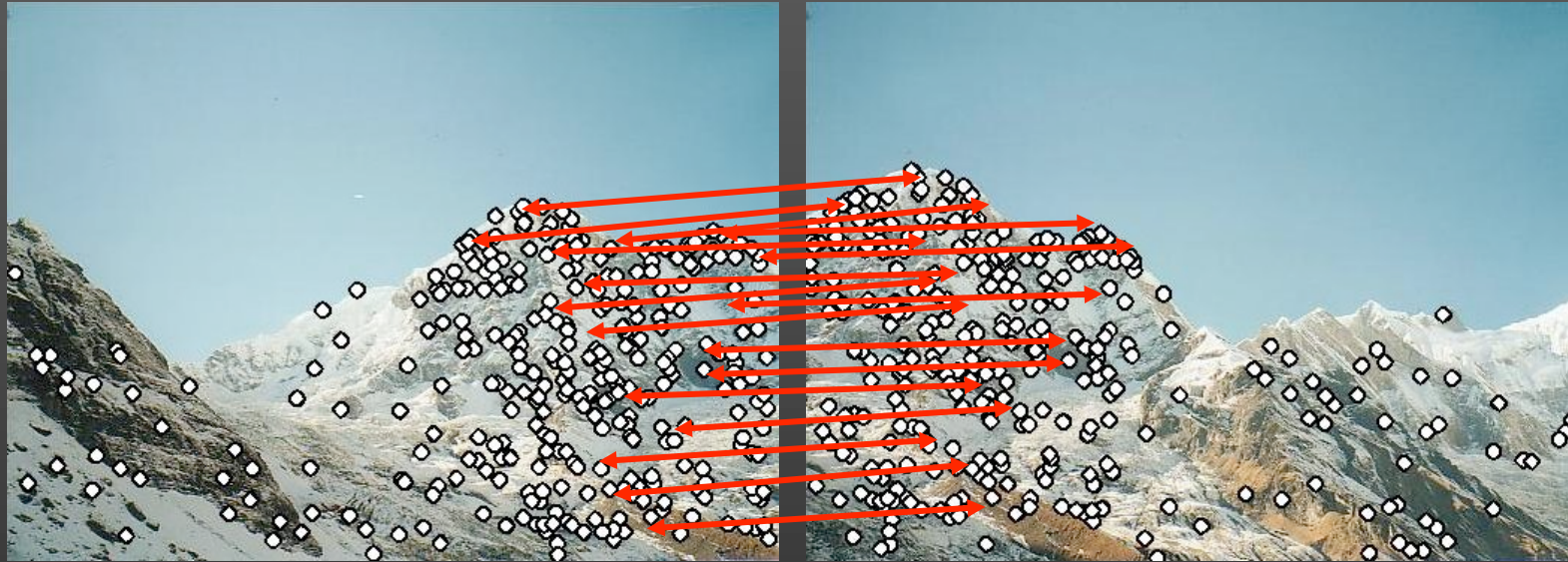
Extract features

Compute *putative matches*

Loop:

- *Hypothesize* transformation T (small group of putative matches that are related by T)

Feature-based alignment outline



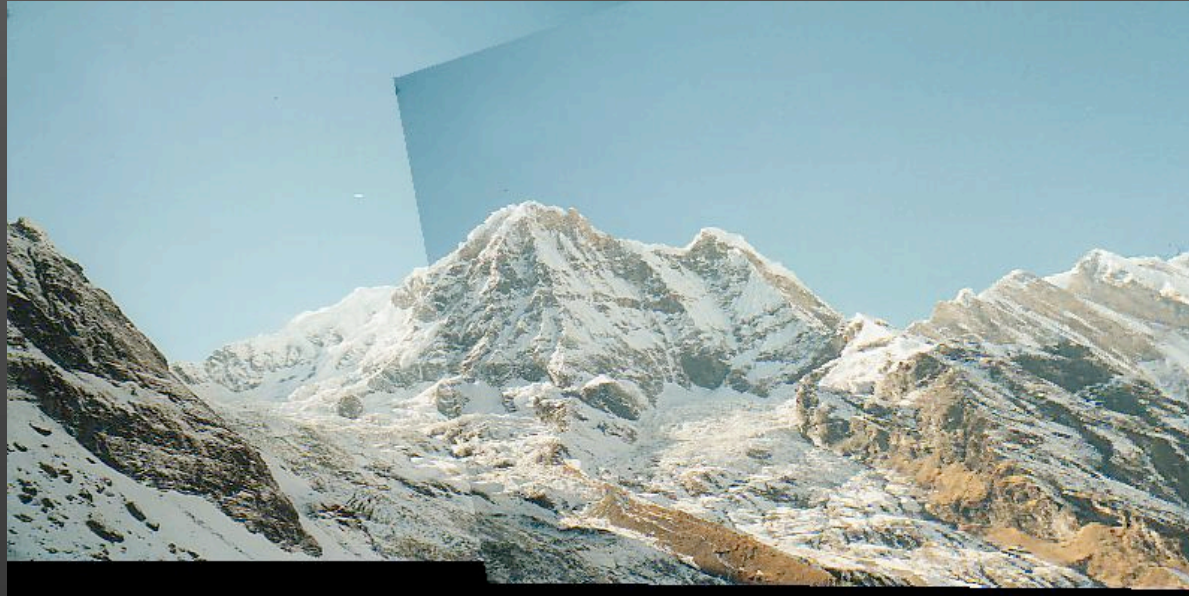
Extract features

Compute *putative matches*

Loop:

- *Hypothesize* transformation T (small group of putative matches that are related by T)
- *Verify* transformation (search for other matches consistent with T)

Feature-based alignment outline



Extract features

Compute *putative matches*

Loop:

- *Hypothesize* transformation T (small group of putative matches that are related by T)
- *Verify* transformation (search for other matches consistent with T)

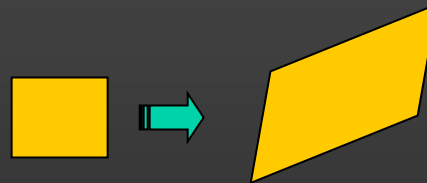
2D transformation models

Similarity

(translation,
scale, rotation)

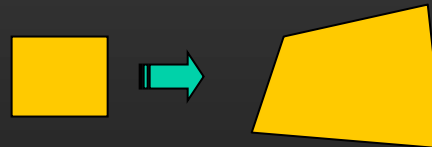


Affine



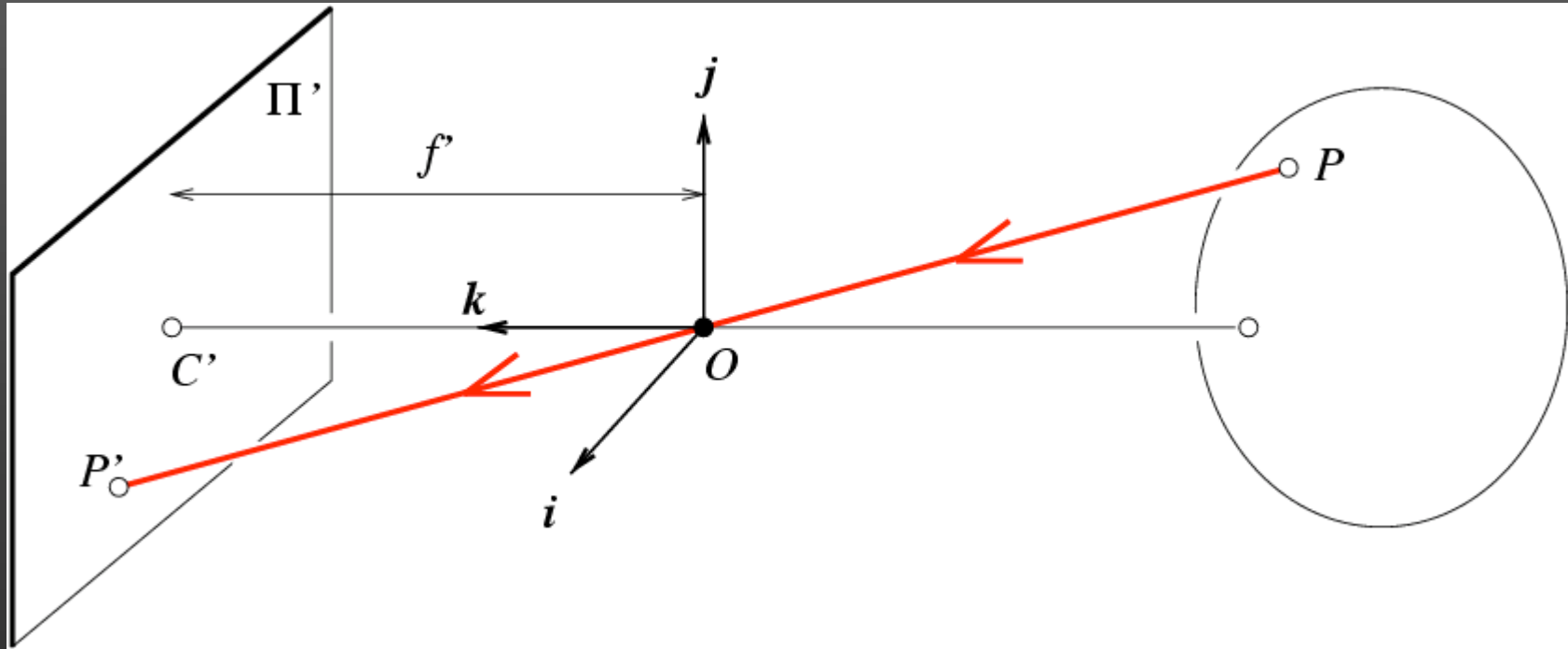
Projective

(homography)



Why these transformations ???

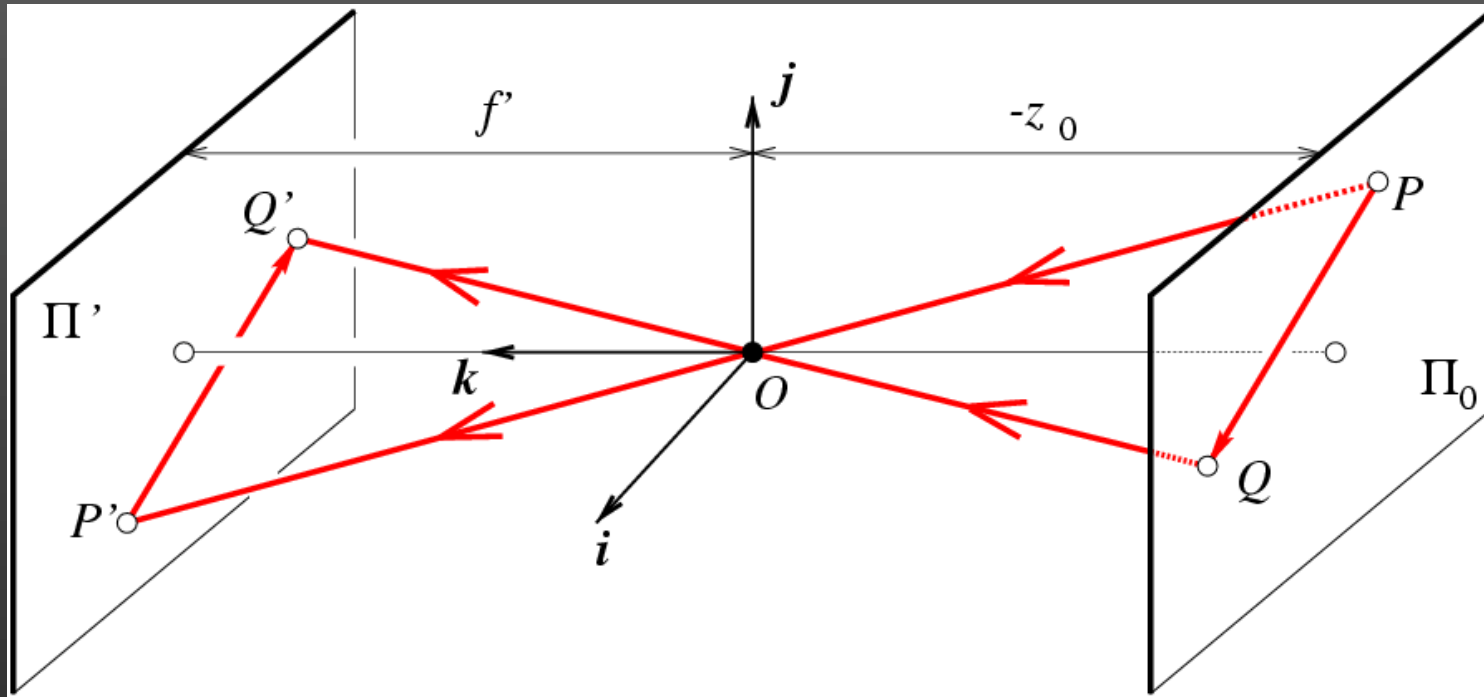
Pinhole perspective equation



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

NOTE: z is always negative..

Affine models: Weak perspective projection



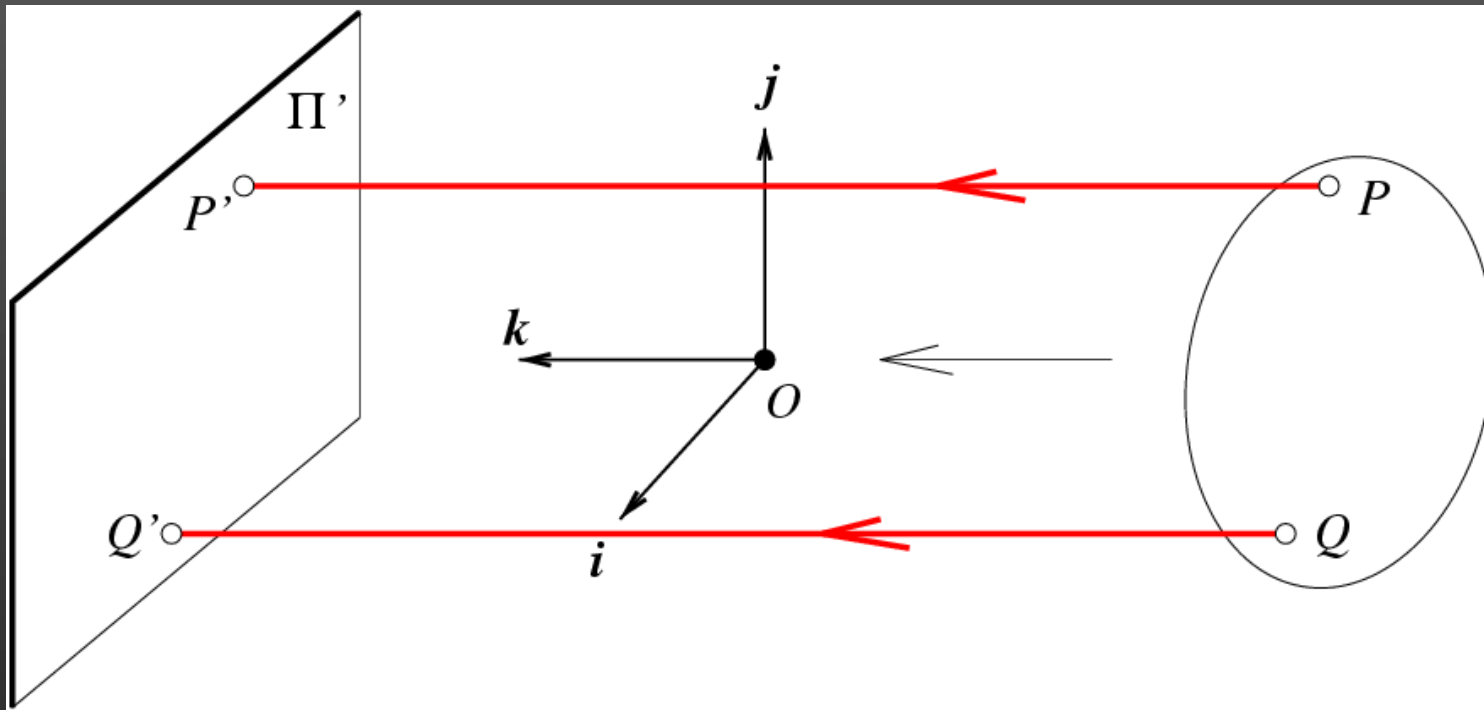
$$\begin{cases} x' = -mx \\ y' = -my \end{cases}$$

where $m = -\frac{f'}{z_0}$

is the magnification.

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

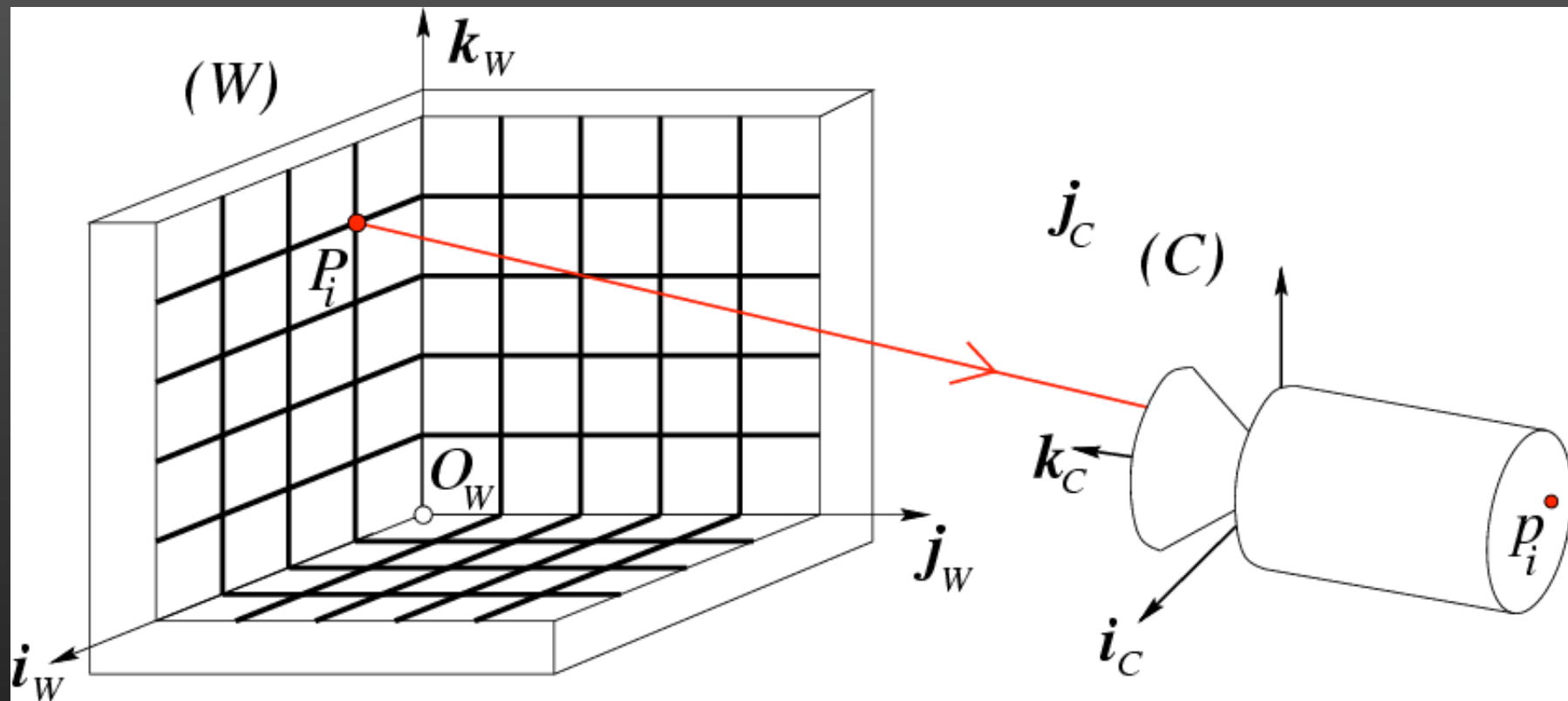
Affine models: Orthographic projection



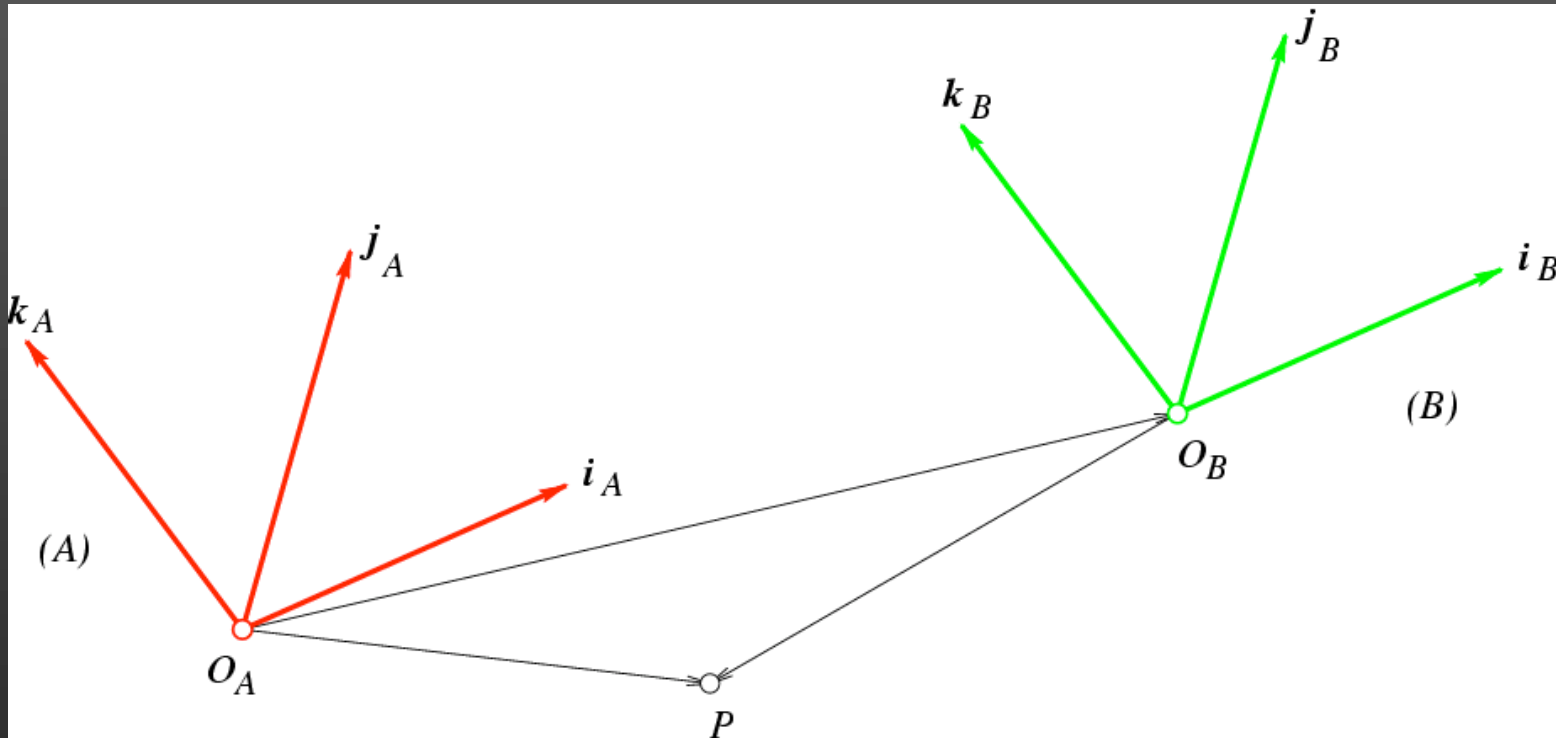
$$\begin{cases} x' = x \\ y' = y \end{cases}$$

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

Analytical camera geometry

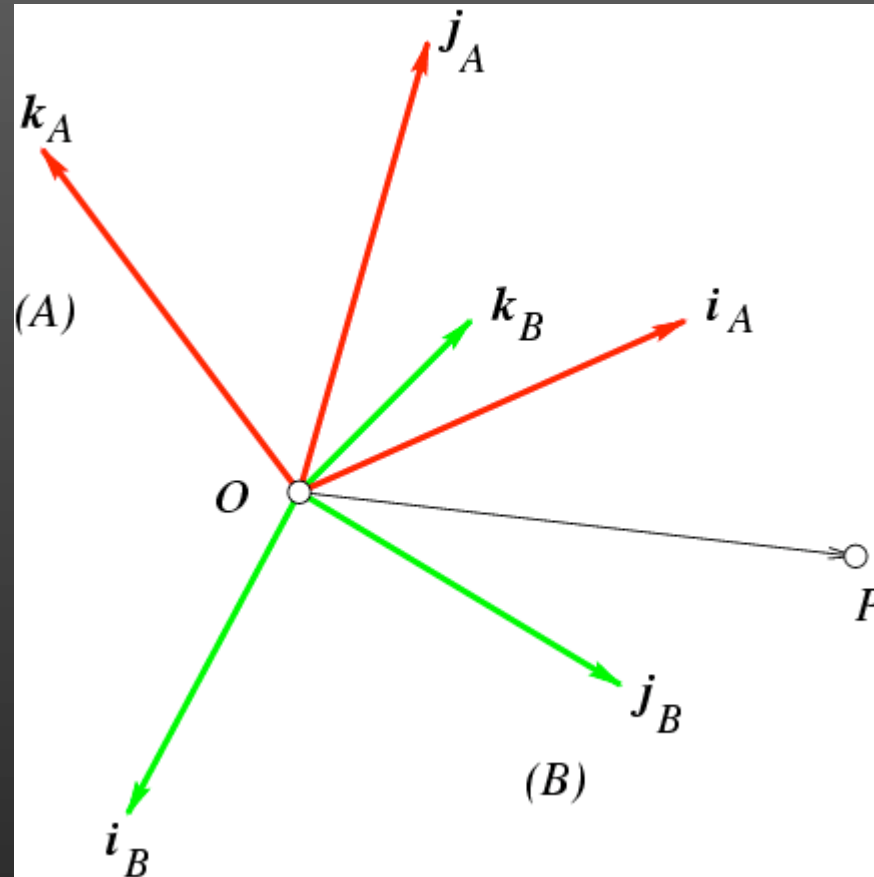


Coordinate Changes: Pure Translations



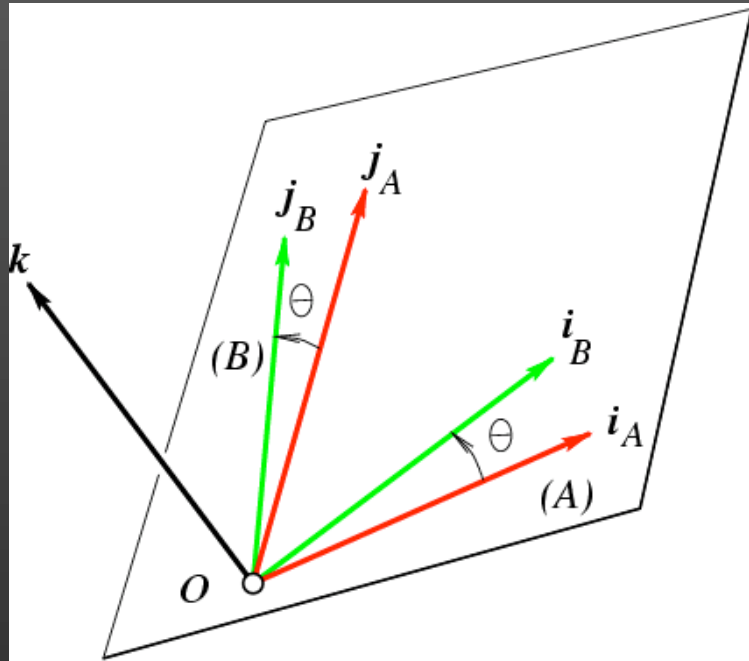
$$\vec{O_B P} = \vec{O_B O_A} + \vec{O_A P}, \quad BP = AP + {}^B O_A$$

Coordinate Changes: Pure Rotations

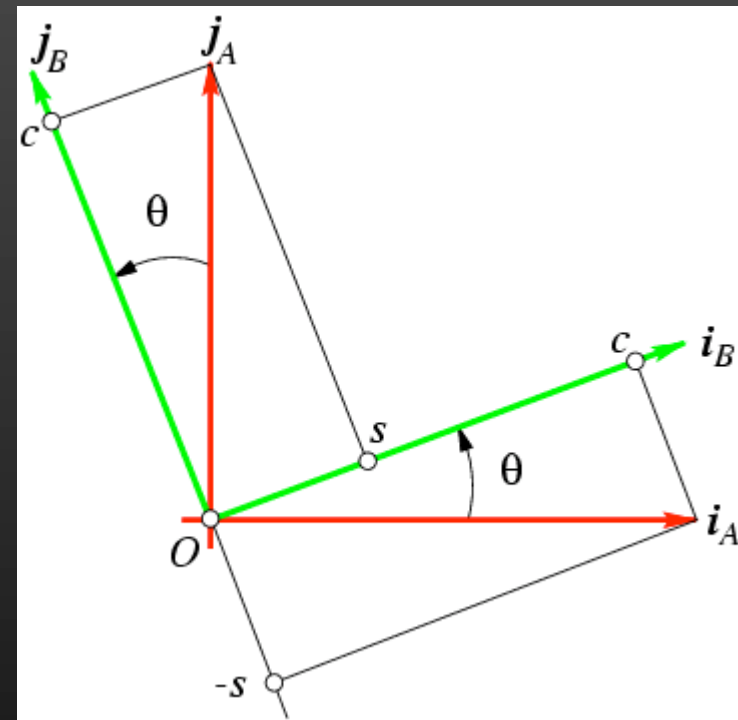


$${}^B_A R = \begin{bmatrix} \mathbf{i}_A \cdot \mathbf{i}_B & \mathbf{j}_A \cdot \mathbf{i}_B & \mathbf{k}_A \cdot \mathbf{i}_B \\ \mathbf{i}_A \cdot \mathbf{j}_B & \mathbf{j}_A \cdot \mathbf{j}_B & \mathbf{k}_A \cdot \mathbf{j}_B \\ \mathbf{i}_A \cdot \mathbf{k}_B & \mathbf{j}_A \cdot \mathbf{k}_B & \mathbf{k}_A \cdot \mathbf{k}_B \end{bmatrix} = \begin{bmatrix} {}^A \mathbf{i}_B^T \\ {}^B \mathbf{i}_A^T \\ {}^A \mathbf{k}_B^T \end{bmatrix} \begin{bmatrix} \mathbf{j}_A & \mathbf{k}_A \end{bmatrix}$$

Coordinate Changes: Rotations about the z Axis



$${}^B_A R = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



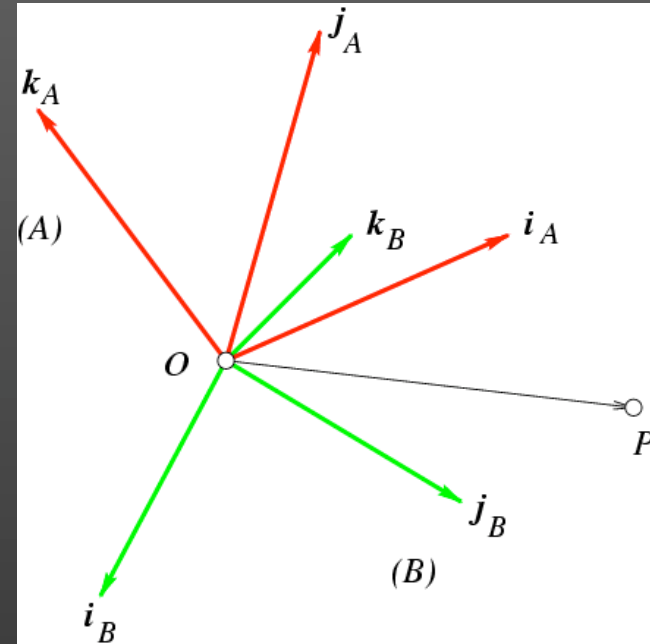
A rotation matrix is characterized by the following properties:

- Its inverse is equal to its transpose, and
- its determinant is equal to 1.

Or equivalently:

- Its rows (or columns) form a right-handed orthonormal coordinate system.

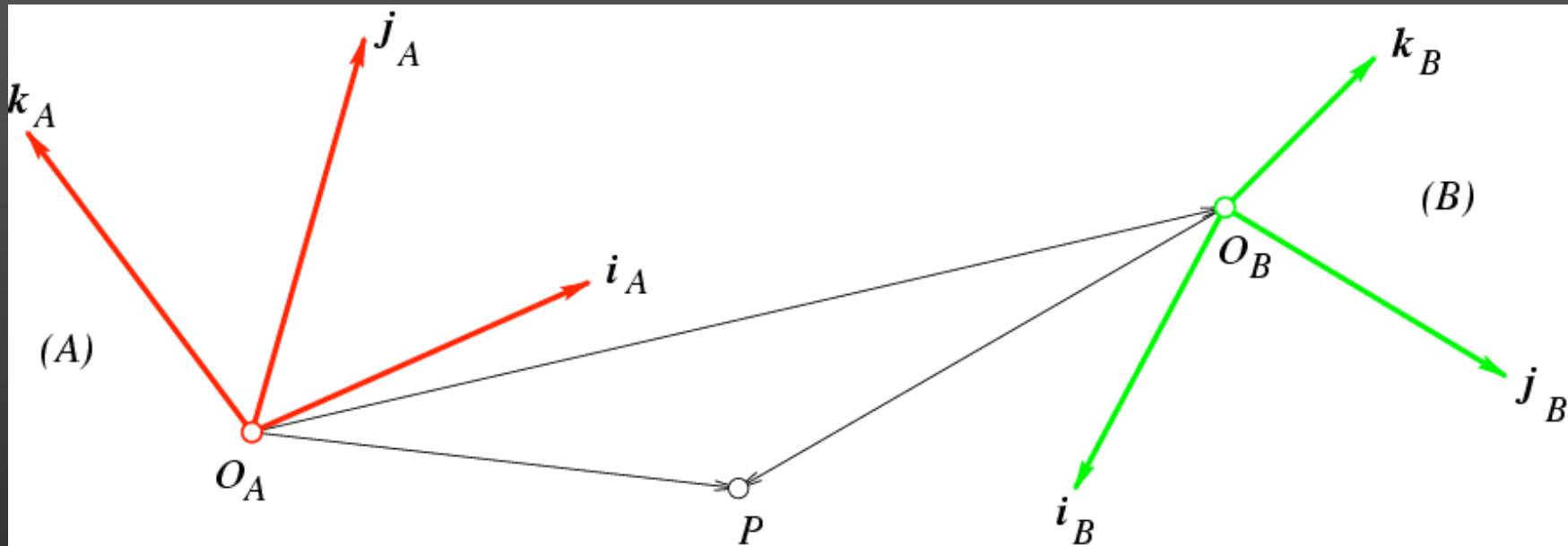
Coordinate changes: pure rotations



$$\overrightarrow{OP} = \begin{bmatrix} \mathbf{i}_A & \mathbf{j}_A & \mathbf{k}_A \end{bmatrix} \begin{bmatrix} {}^A x \\ {}^A y \\ {}^A z \end{bmatrix} = \begin{bmatrix} \mathbf{i}_B & \mathbf{j}_B & \mathbf{k}_B \end{bmatrix} \begin{bmatrix} {}^B x \\ {}^B y \\ {}^B z \end{bmatrix}$$

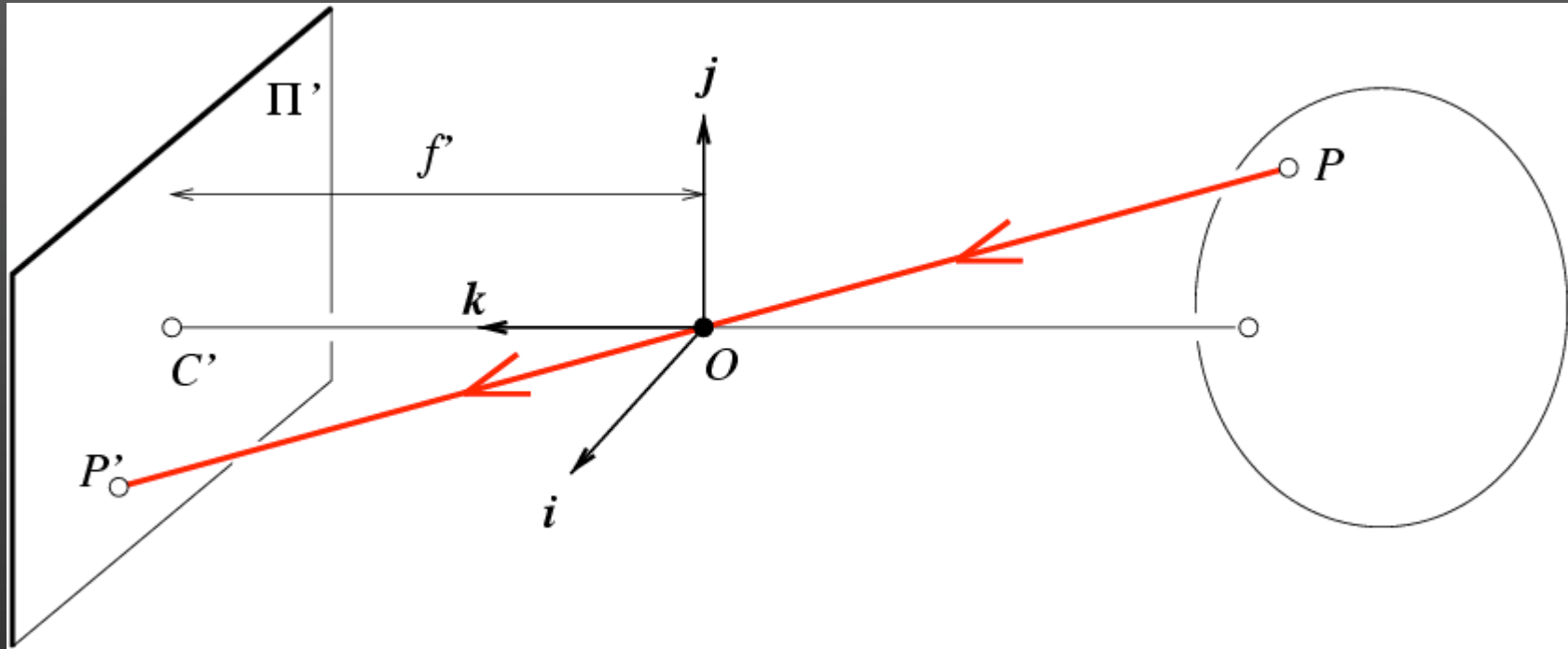
$$\Rightarrow {}^B P = {}^B R^A P$$

Coordinate Changes: Rigid Transformations



$$\begin{bmatrix} {}^B P \\ 1 \end{bmatrix} = \begin{bmatrix} {}^B A \mathbf{B} & {}^B O_A \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} {}^A P \\ 1 \end{bmatrix} = \begin{bmatrix} {}^A A \mathbf{B} & {}^A O_B \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} {}^B P \\ 1 \end{bmatrix} + \begin{bmatrix} {}^B O_B \\ 0 \end{bmatrix}$$

Pinhole perspective equation



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

NOTE: z is always negative..

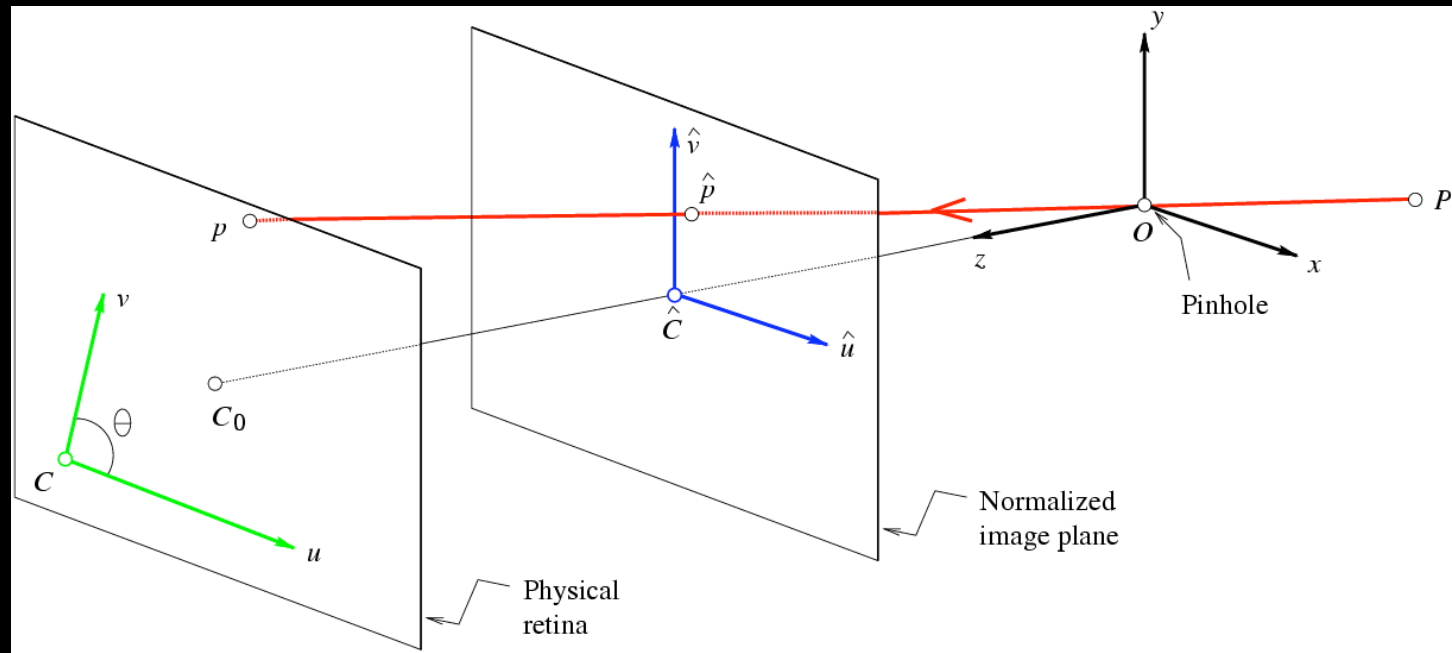
The intrinsic parameters of a camera

Units:

k, l : pixel/m

f : m

α, β : pixel



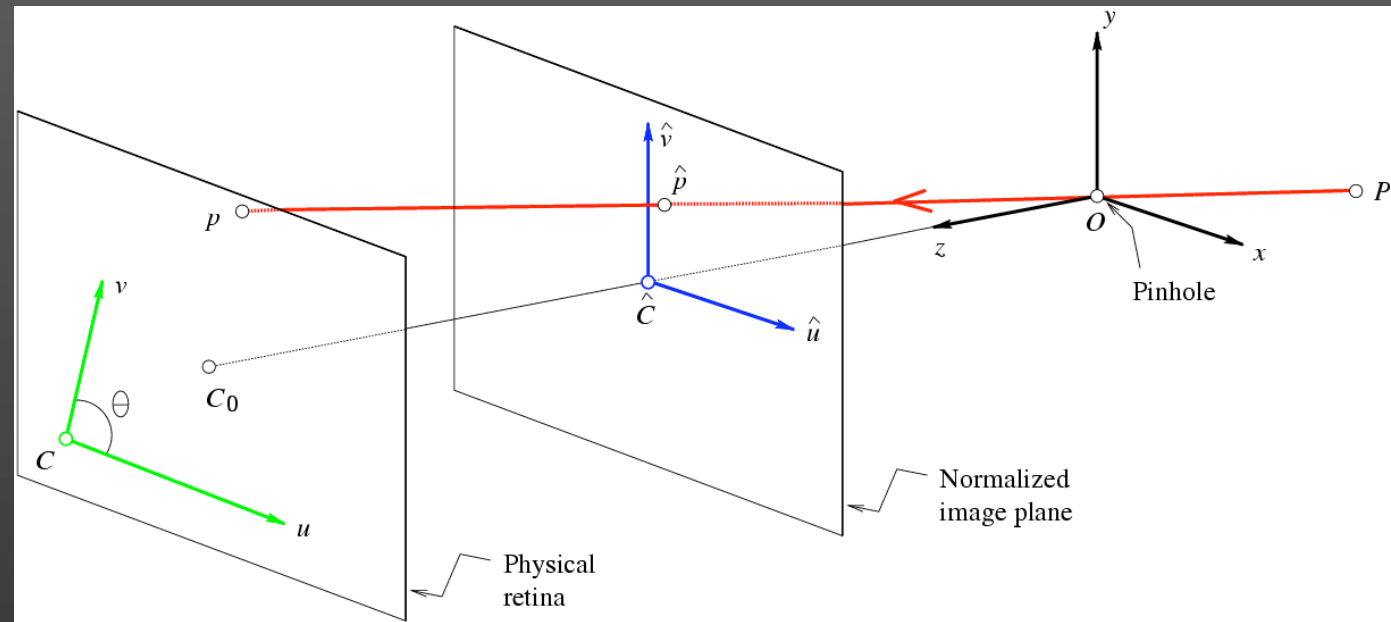
$$\begin{cases} \hat{u} = \frac{x}{z} \\ \hat{v} = \frac{y}{z} \end{cases} \iff \hat{\mathbf{p}} = \frac{1}{z} (\text{Id} \quad \mathbf{0}) \begin{pmatrix} \mathbf{P} \\ 1 \end{pmatrix}$$

Physical image coordinates

Normalized image coordinates

$$\begin{cases} u = kf \frac{x}{z} \\ v = lf \frac{y}{z} \end{cases}$$

The intrinsic parameters of a camera



Calibration matrix

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

The perspective projection equation

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

The extrinsic parameters of a camera

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

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

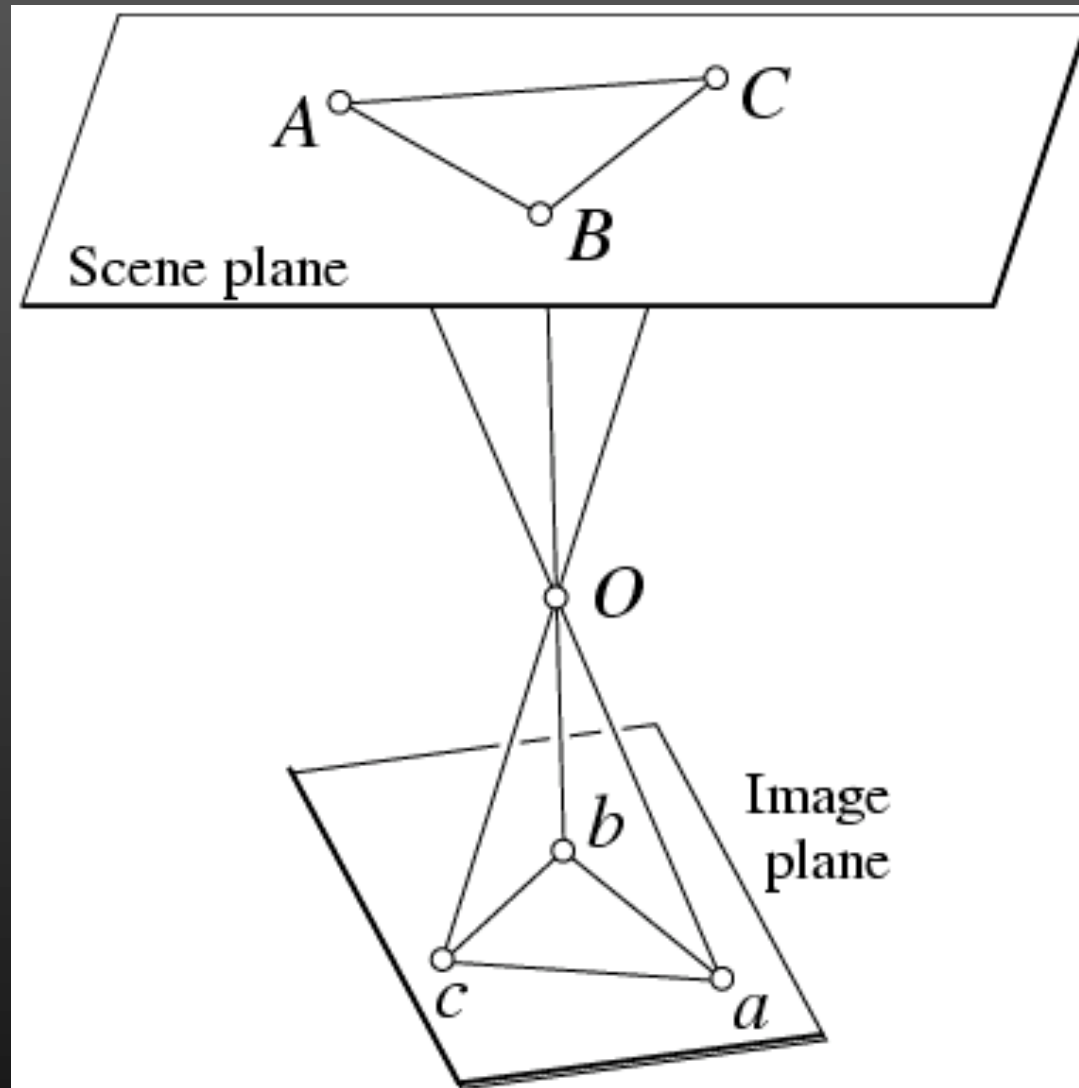
- Thus,

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

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

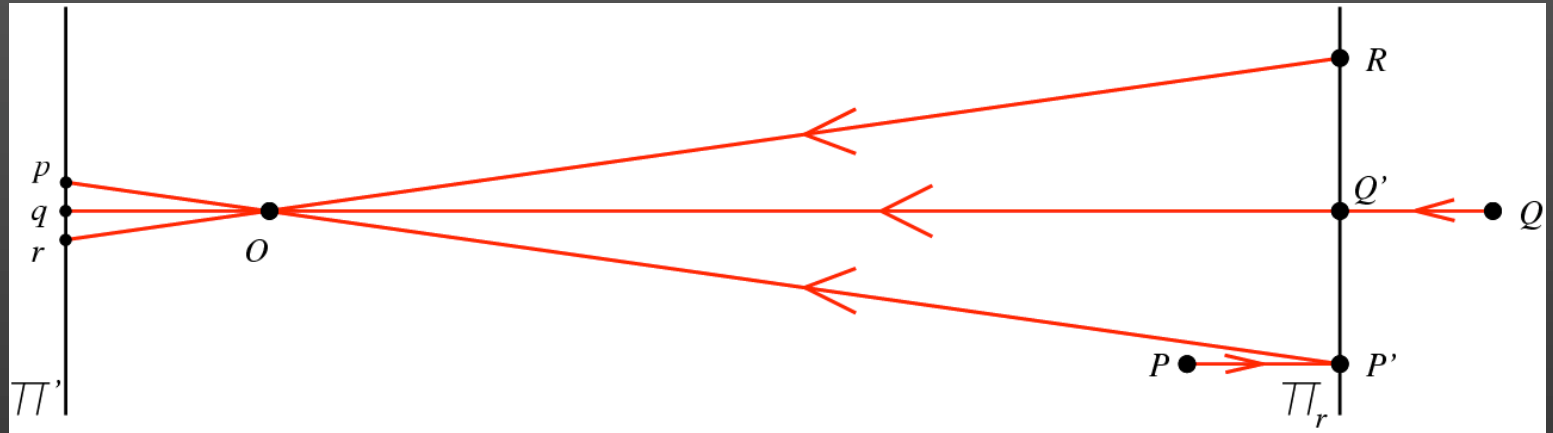
$$\mathcal{M} = \begin{pmatrix} \mathbf{m}_1^T \\ \mathbf{m}_2^T \\ \mathbf{m}_3^T \end{pmatrix} \implies z = \mathbf{m}_3 \cdot \mathbf{P}, \quad \text{or} \quad \begin{cases} u = \frac{\mathbf{m}_1 \cdot \mathbf{P}}{\mathbf{m}_3 \cdot \mathbf{P}}, \\ v = \frac{\mathbf{m}_2 \cdot \mathbf{P}}{\mathbf{m}_3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

Perspective projections induce projective transformations between planes

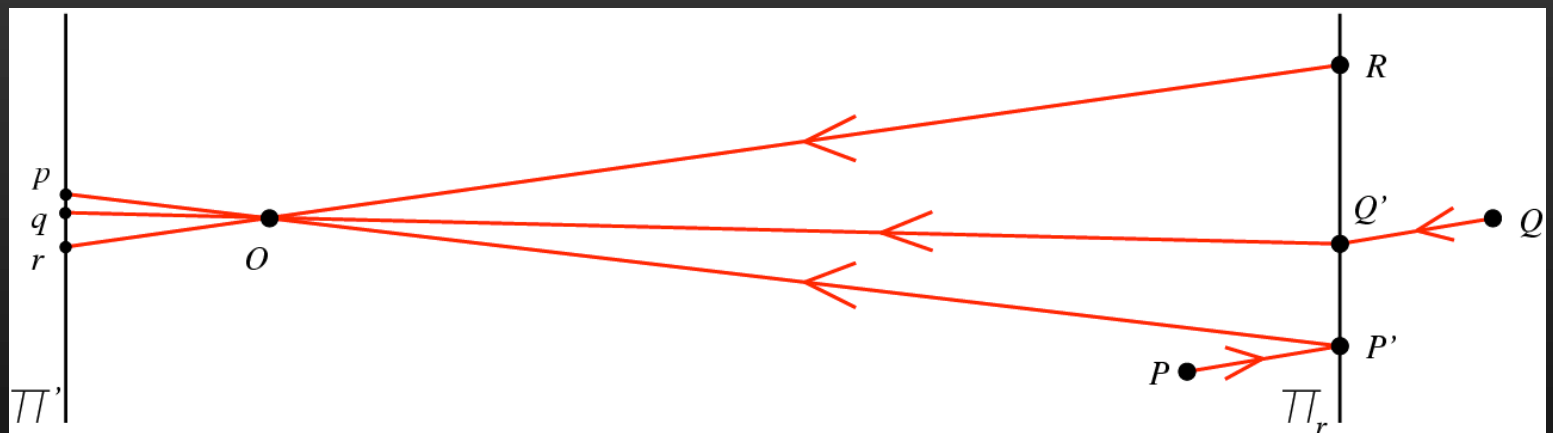


Affine cameras

Weak-perspective projection

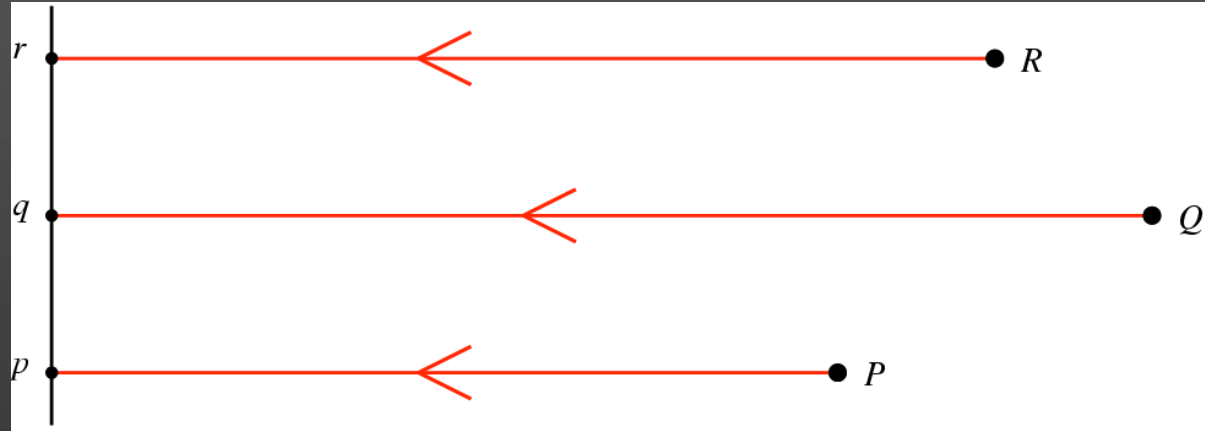


Paraperspective projection

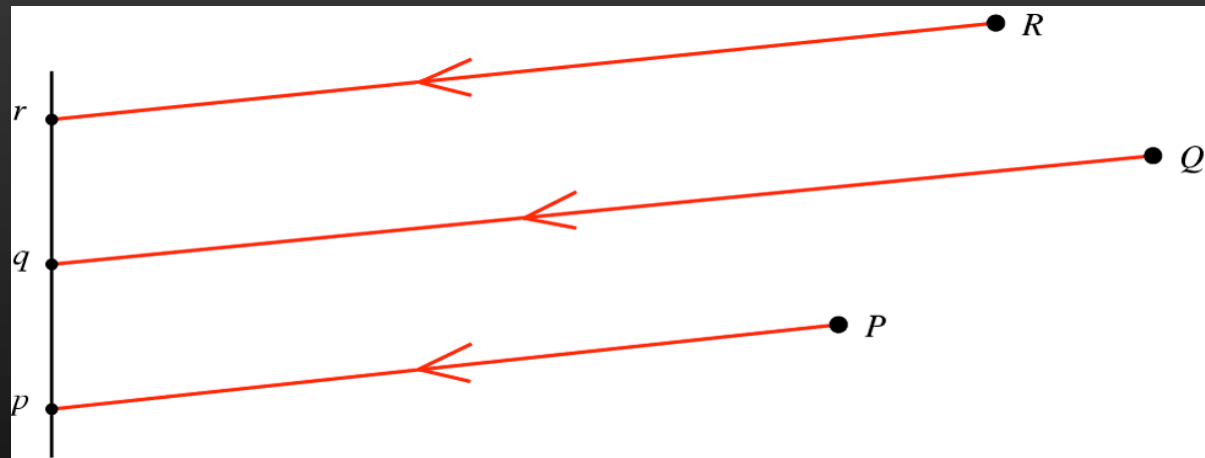


More affine cameras

Orthographic projection



Parallel projection



Weak-perspective projection model

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

(\mathbf{p} and \mathbf{P} are in homogeneous coordinates)

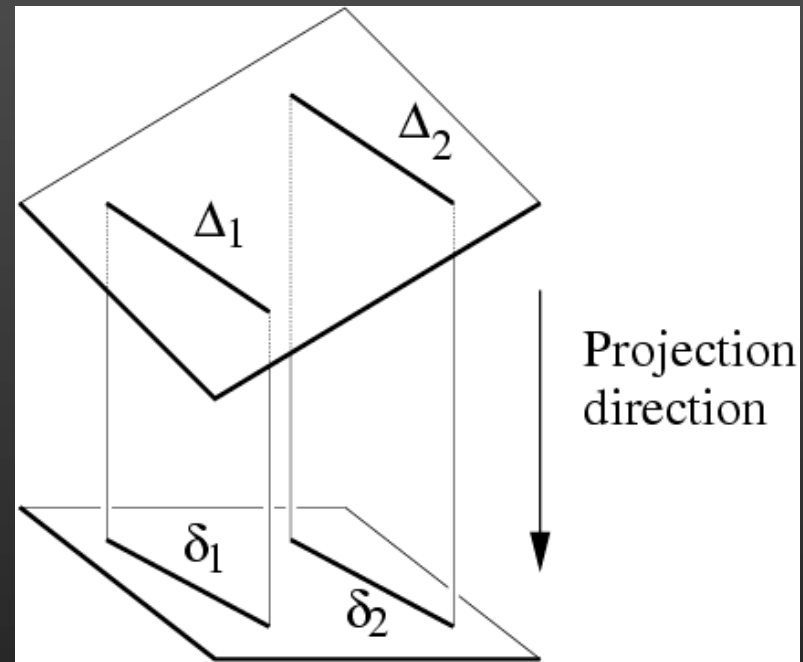
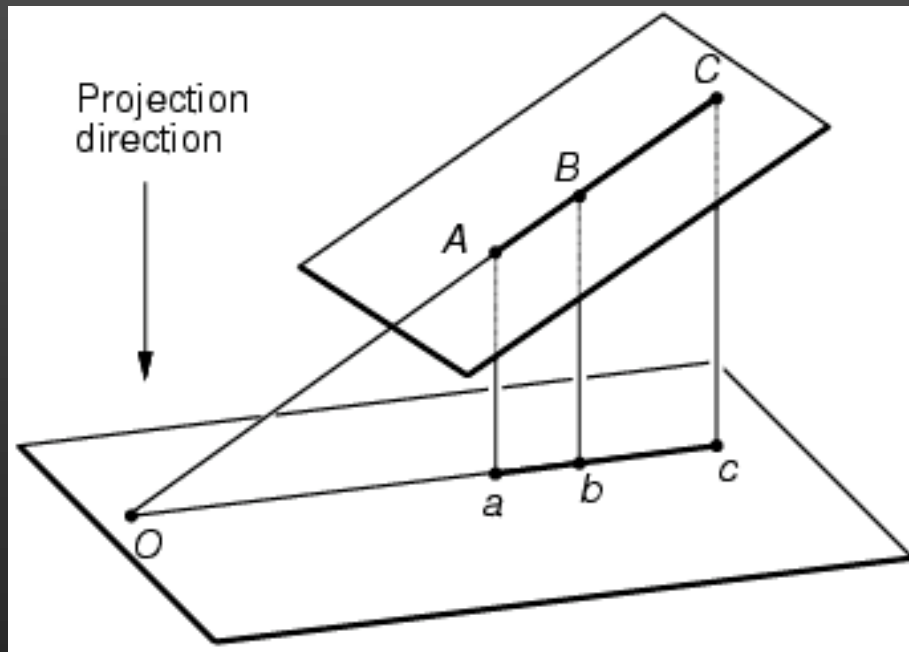

$$\mathbf{p} = \mathcal{M} \mathbf{P}$$

(\mathbf{P} is in homogeneous coordinates)


$$\mathbf{p} = \mathbf{A} \mathbf{P} + \mathbf{b}$$

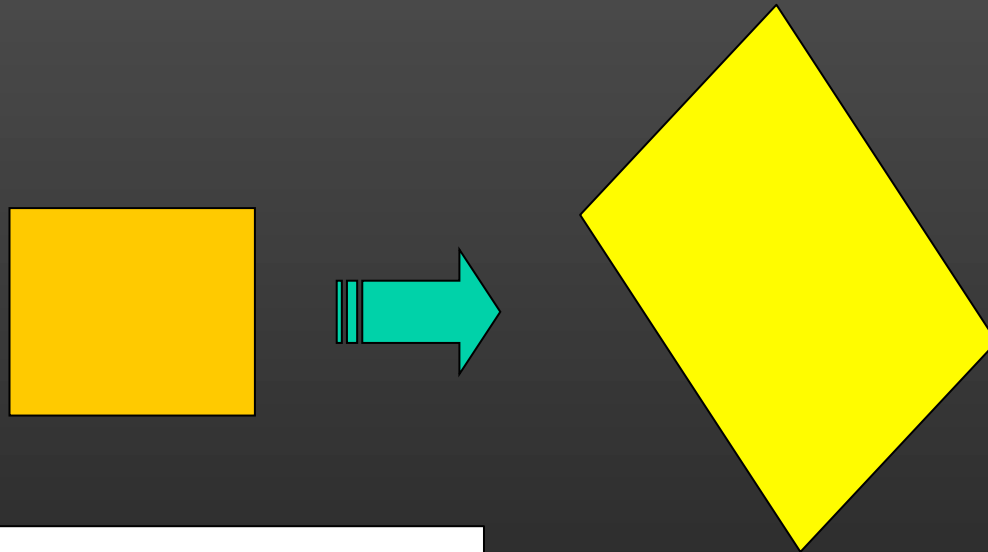
(neither \mathbf{p} nor \mathbf{P} is in hom. coordinates)

Affine projections induce affine transformations from planes onto their images.



Affine transformations

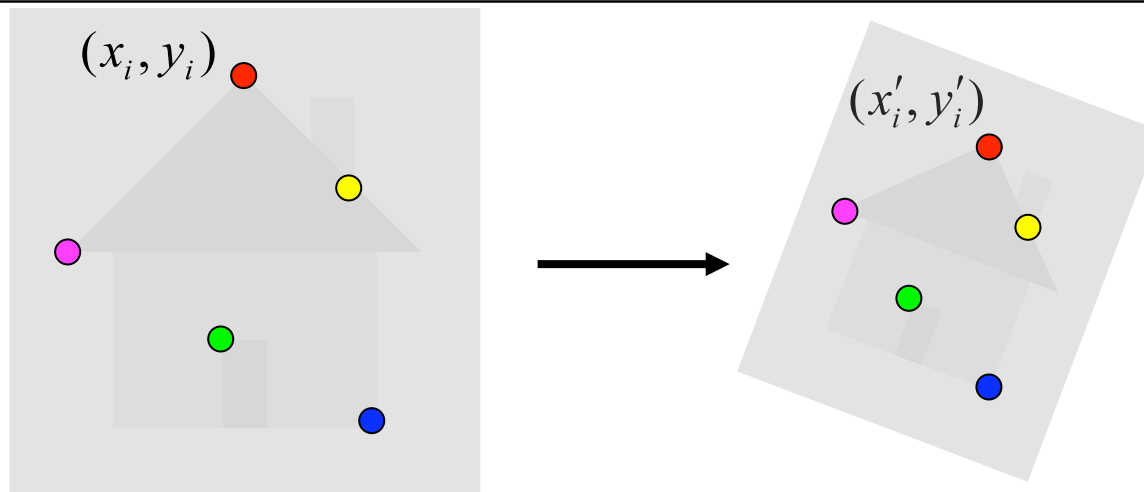
An affine transformation maps a parallelogram onto another parallelogram



$$\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

Fitting an affine transformation

Assume we know the correspondences, how do we get the transformation?



$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}$$

$$\begin{bmatrix} \dots & \dots & \dots & \dots & \dots & \dots \\ x_i & y_i & 0 & 0 & 1 & 0 \\ 0 & 0 & x_i & y_i & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \dots \\ x'_i \\ y'_i \\ \dots \end{bmatrix}$$

Fitting an affine transformation

$$\begin{bmatrix} \dots & & & & & & \\ x_i & y_i & 0 & 0 & 1 & 0 & \\ 0 & 0 & x_i & y_i & 0 & 1 & \\ \dots & & & & & & \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_1 \\ t_2 \end{bmatrix} = \begin{bmatrix} \dots \\ x'_i \\ y'_i \\ \dots \end{bmatrix}$$

Linear system with six unknowns

Each match gives us two linearly independent equations: need at least three to solve for the transformation parameters

Beyond affine transformations

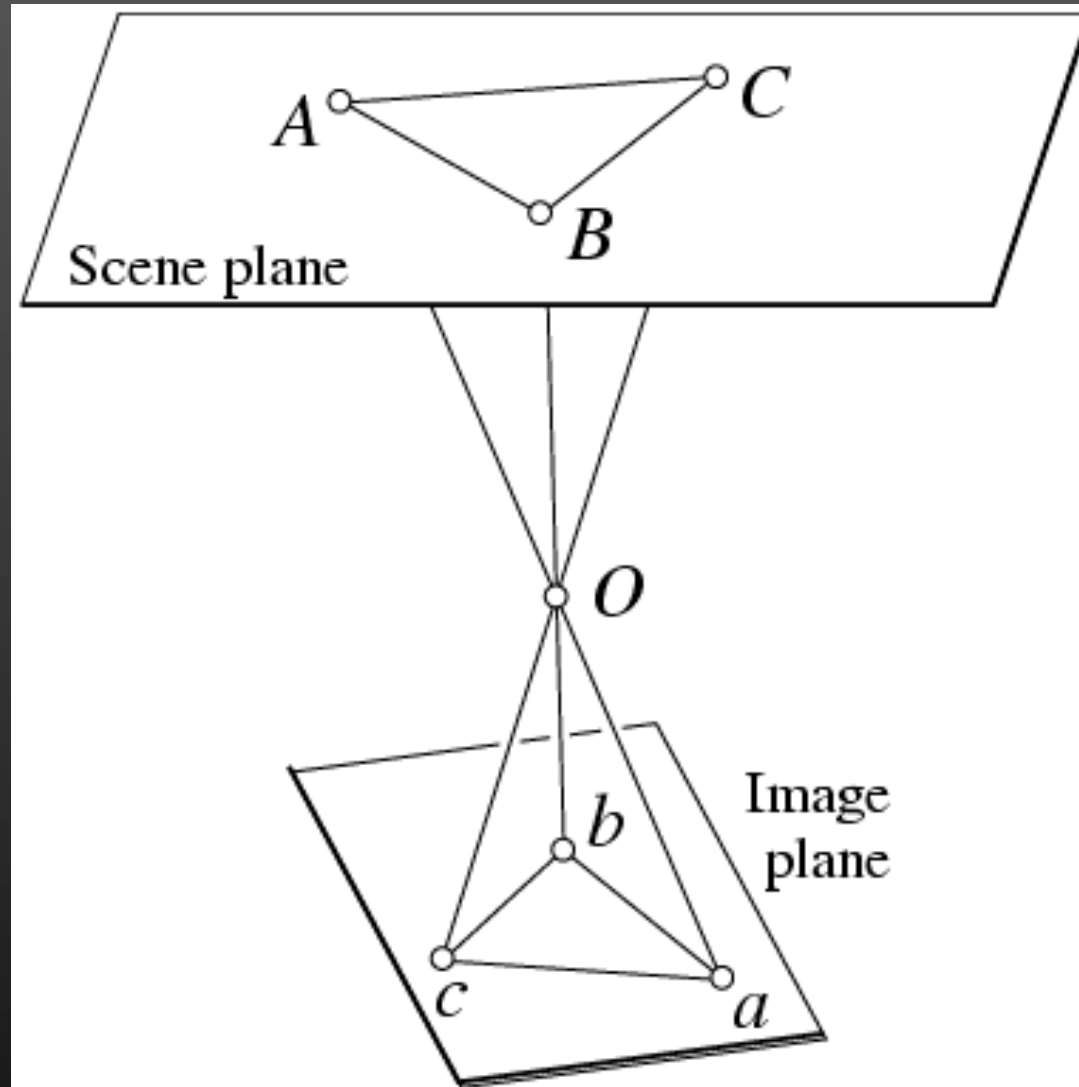
What is the transformation between two views of a planar surface?



What is the transformation between images from two cameras that share the same center?

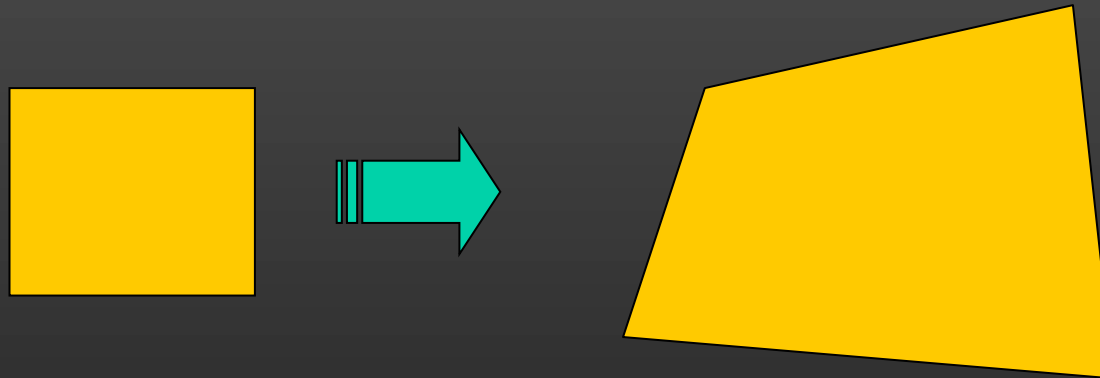


Perspective projections induce projective transformations between planes



Beyond affine transformations

Homography: plane projective transformation
(transformation taking a quad to another arbitrary quad)



Fitting a homography

Recall: homogenous coordinates

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Converting *to* homogenous
image coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

Converting *from* homogenous
image coordinates

Fitting a homography

Recall: homogenous coordinates

$$(x, y) \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Converting *to* homogenous
image coordinates

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$$

Converting *from* homogenous
image coordinates

Equation for homography:

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Fitting a homography

Equation for homography:

$$\lambda \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}$$

$$\lambda \mathbf{x}'_i = \mathbf{H} \mathbf{x}_i = \begin{bmatrix} \mathbf{h}_1^T \\ \mathbf{h}_2^T \\ \mathbf{h}_3^T \end{bmatrix} \mathbf{x}_i$$

9 entries, 8 degrees of freedom
(scale is arbitrary)

$$\mathbf{x}'_i \times \mathbf{H} \mathbf{x}_i = 0$$

$$\mathbf{x}'_i \times \mathbf{H} \mathbf{x}_i = \begin{bmatrix} y'_i \mathbf{h}_3^T \mathbf{x}_i - \mathbf{h}_2^T \mathbf{x}_i \\ \mathbf{h}_1^T \mathbf{x}_i - x'_i \mathbf{h}_3^T \mathbf{x}_i \\ x'_i \mathbf{h}_2^T \mathbf{x}_i - y'_i \mathbf{h}_1^T \mathbf{x}_i \end{bmatrix}$$

$$\begin{bmatrix} 0^T & -\mathbf{x}_i^T & y'_i \mathbf{x}_i^T \\ \mathbf{x}_i^T & 0^T & -x'_i \mathbf{x}_i^T \\ -y'_i \mathbf{x}_i^T & x'_i \mathbf{x}_i^T & 0^T \end{bmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{pmatrix} = 0$$

3 equations, only 2 linearly independent

Direct linear transform

$$\begin{bmatrix} \mathbf{0}^T & \mathbf{x}_1^T & -y'_1 \mathbf{x}_1^T \\ \mathbf{x}_1^T & \mathbf{0}^T & -x'_1 \mathbf{x}_1^T \\ \dots & \dots & \dots \\ \mathbf{0}^T & \mathbf{x}_n^T & -y'_n \mathbf{x}_n^T \\ \mathbf{x}_n^T & \mathbf{0}^T & -x'_n \mathbf{x}_n^T \end{bmatrix} \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \\ \mathbf{h}_3 \end{pmatrix} = 0$$

$$\mathbf{A} \mathbf{h} = 0$$

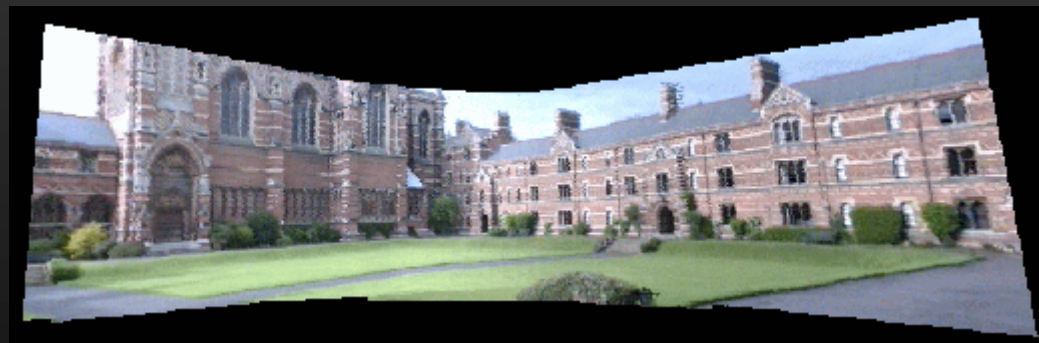
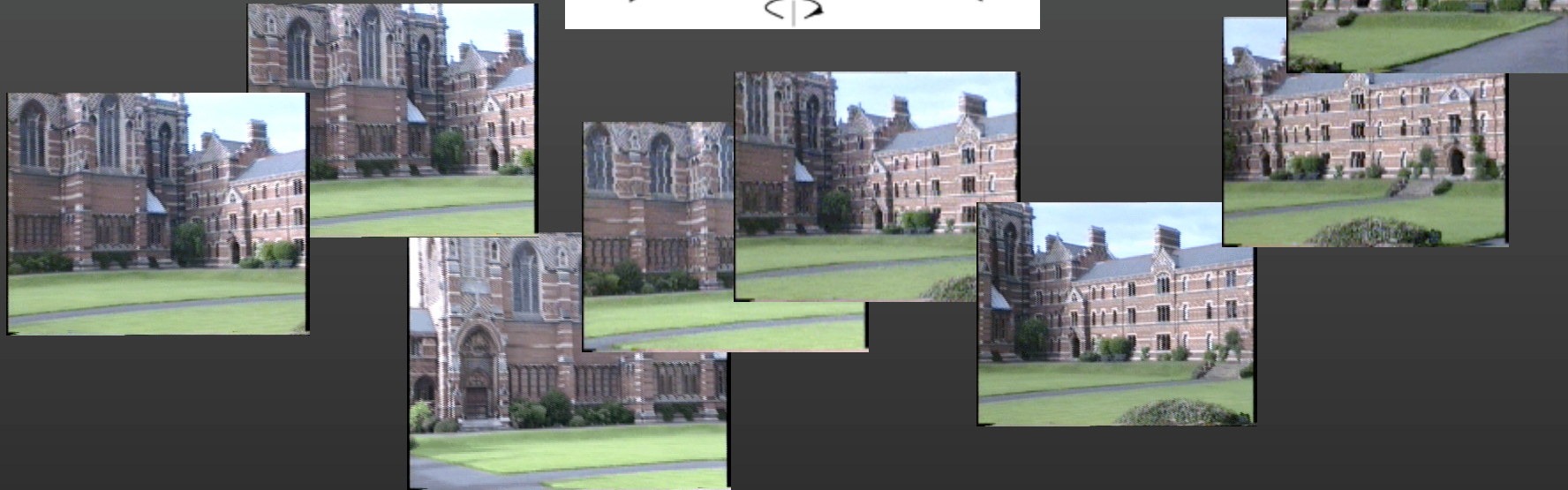
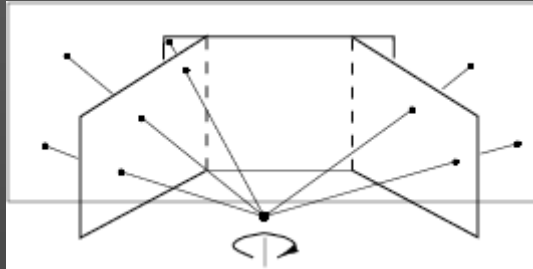
H has 8 degrees of freedom (9 parameters, but scale is arbitrary)

One match gives us two linearly independent equations

Four matches needed for a minimal solution (null space of 8x9 matrix)

More than four: homogeneous least squares

Application: Panorama stitching



Images courtesy of A. Zisserman.

