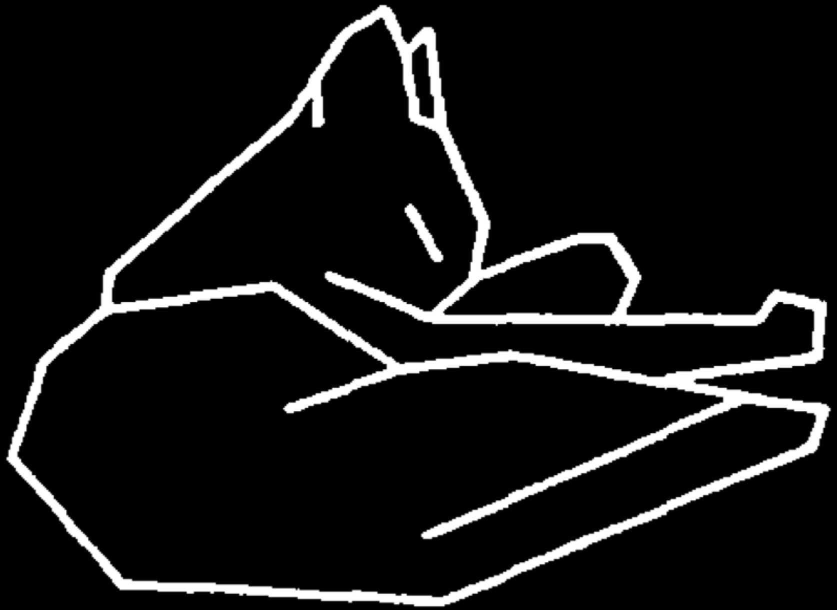
Computer vision: machines that see

Richard Harvey
IT Livery Company Professor of Information Technology, Gresham College

Professor of Computer Science, School of Computing Sciences,
University of East Anglia

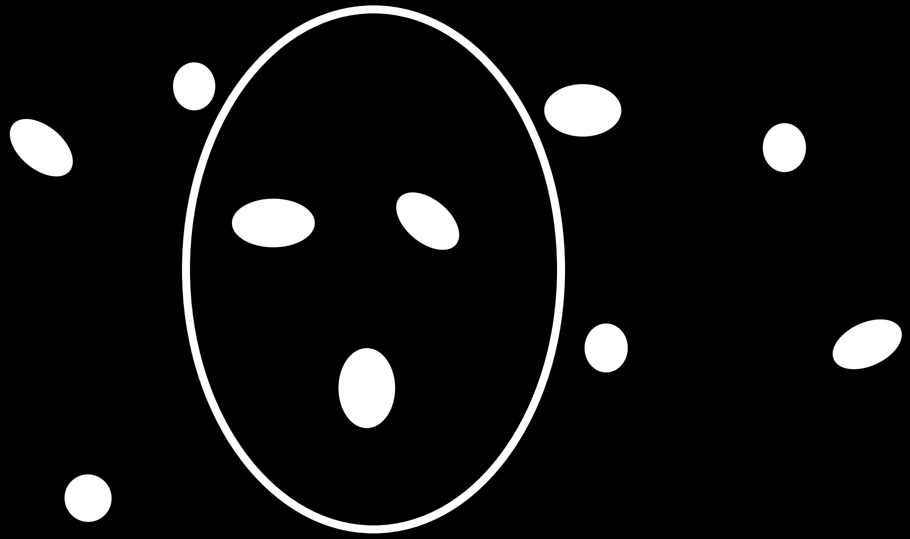
#richardwharvey

The mystery of vision in 64 lines

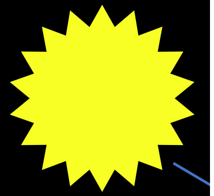


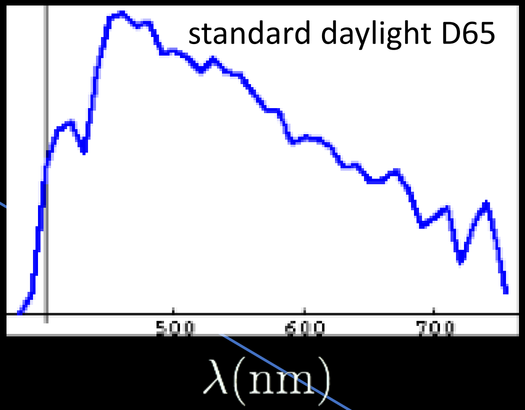
Some information aspects of visual attention, Fred Attneave, Psychological Review, 61, (3), 1954, pp 183--193

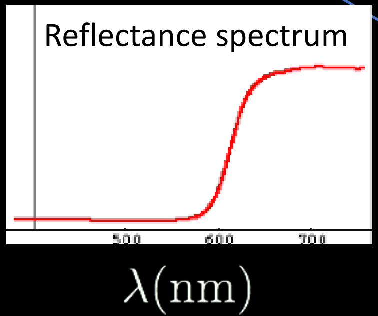
The mystery of vision

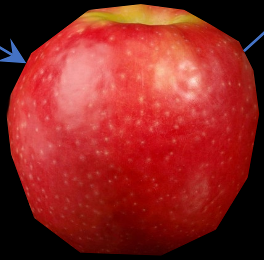


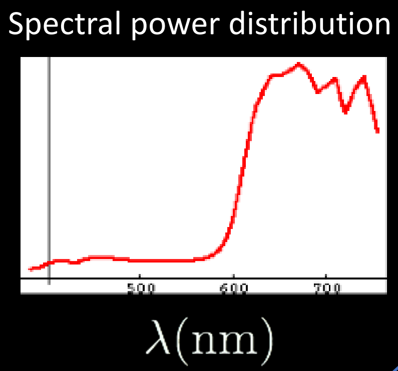
Colour imaging



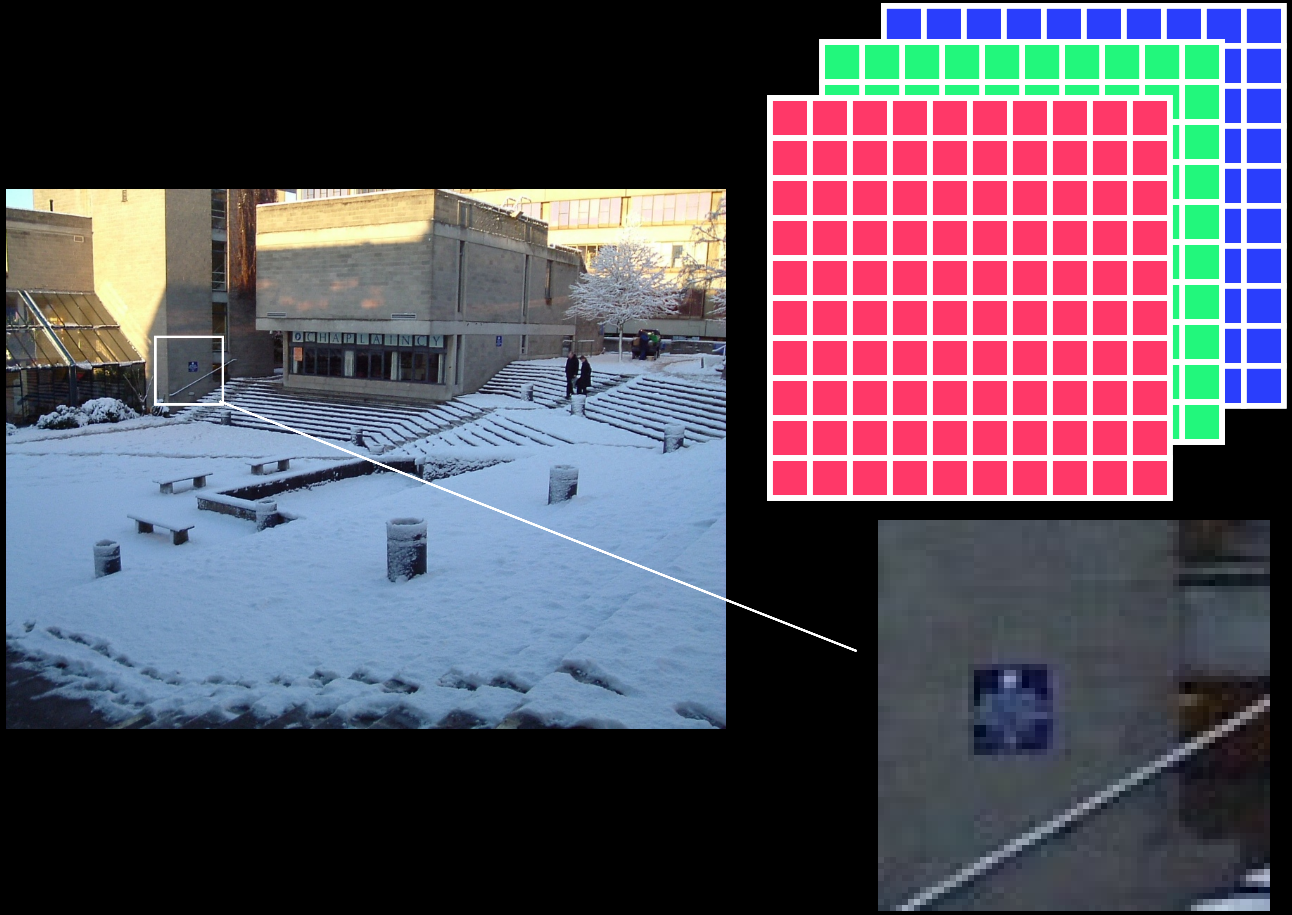




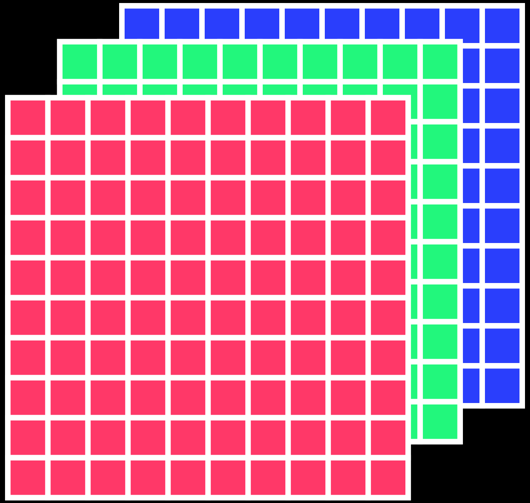




My colour sensors are giving a signal that I have previously learned to associate with the word “red”


CHAPLAINCY

Images

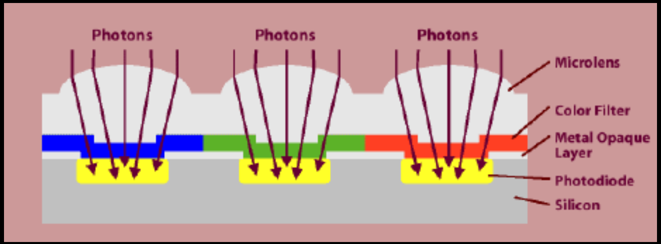


Typically:

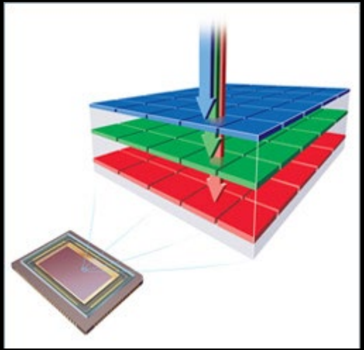
- three planes: R, G and B;
- 8-bit unsigned integers represent colours
- 0 = Black or no intensity
- 255 = White or full intensity

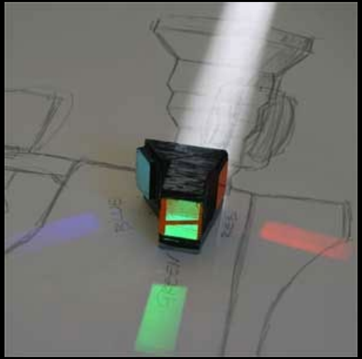
Nikon D70 CCD

Colour sensors



G	R	G	R
B	G	B	G
G	R	G	R
B	G	B	G

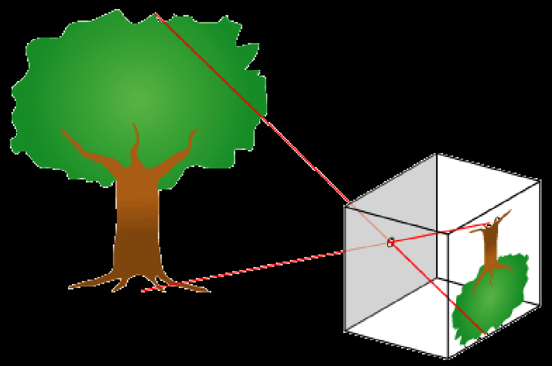




Colour images

Not all colour images are reliable

Pinhole Camera



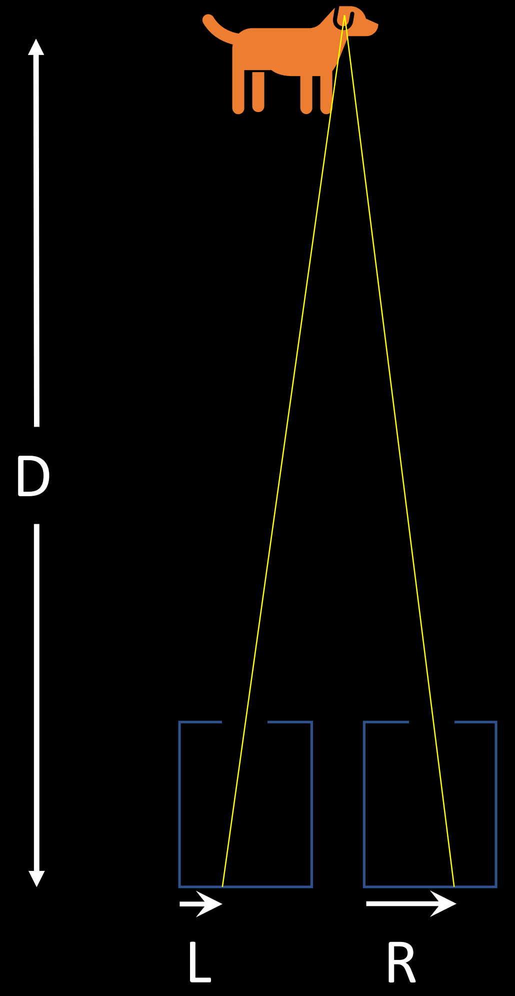
Camera obscura

Camera geometry

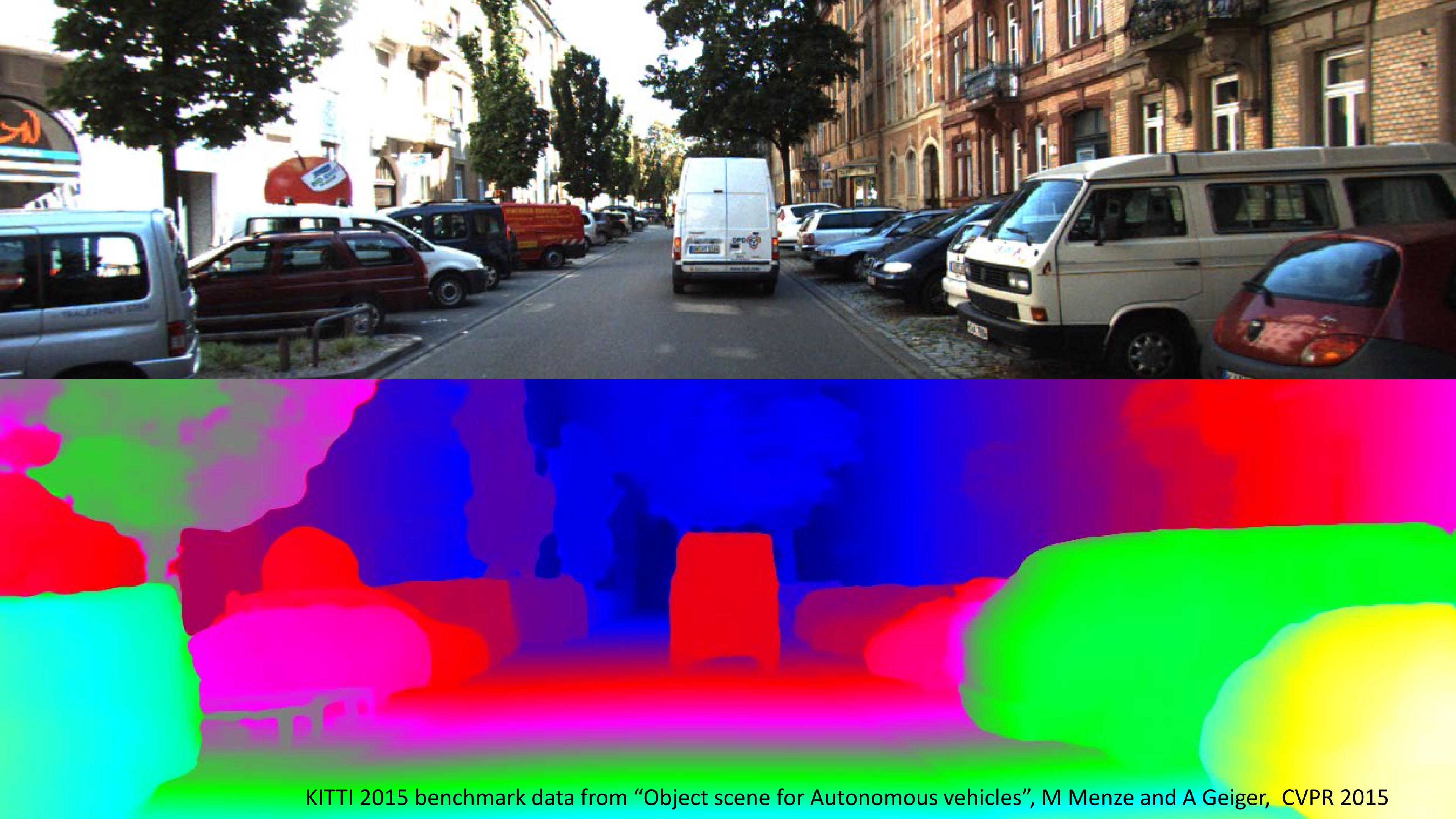
Intrinsic parameters
- focal length
- principal point
- pixel size
- pixel aspect ratio

Extrinsic parameters
- location of camera
- orientations of camera

Stereo



Disparity = $R - L \propto 1/D$



KITTI 2015 benchmark data from “Object scene for Autonomous vehicles”, M Menze and A Geiger, CVPR 2015

Video data

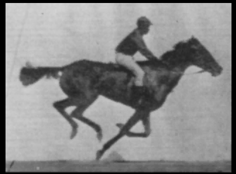
Fifteen cameras positioned along a single stride (6.85m) of "Annie" galloping at 54 kmph

- equivalent to a single moving camera

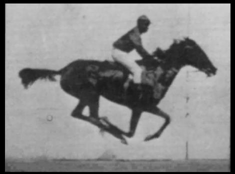
Muybridge wanted to know if a horse's feet were always in contact with the ground during a gallop

- a question that is easily answered…

From *The Horse in Motion by* Eadweard Muybridge (Edward Muggeridge), 1878. Republished in 1899 as part of *Animals in Motion: An Electro-Photographic Investigation of Consecutive Phases of Animal Progressive Movements* (Chapman and Hall, London).



Frame 1



Frame 2

Video

If a sequence of images is displayed fast enough, the motion looks smooth

This is a *sampling problem*:

- Given a signal varying at a known rate, how quickly must we sample it to be reconstruct the signal?

Sampling

But video data are enormous

4k Television, without compression

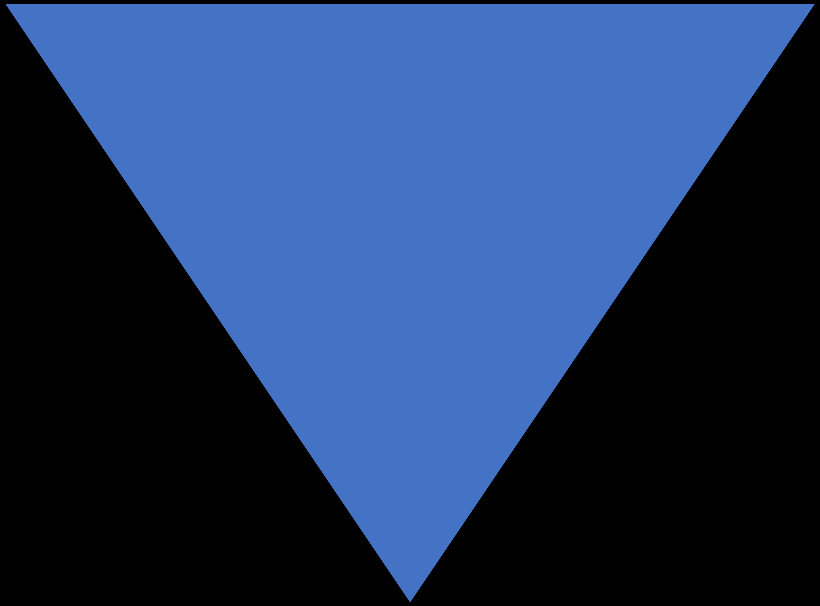
$$3840 \times 2160 \times 3 \text{ bytes} \times 8 \text{ bits} \times 30 \text{ fps} = 6 \text{ Gbits s}^{-1}$$

with compression

$$25 \text{ Mbits s}^{-1}$$

How to handle all that data?

The image pyramid



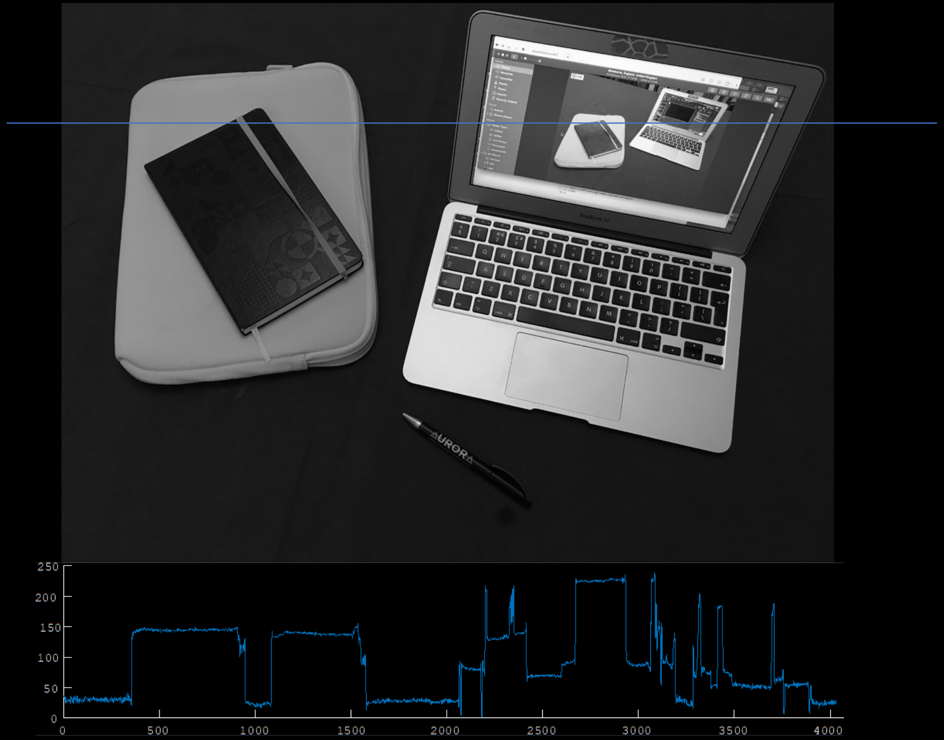
Images: millions of bits fine scale

Inference: a few bits at coarse scale

Scale-space theory

A grand theory of everything

Images have extrema



Objects are likely to be extreme

So it is imperative to simplify images
without introducing any new extrema

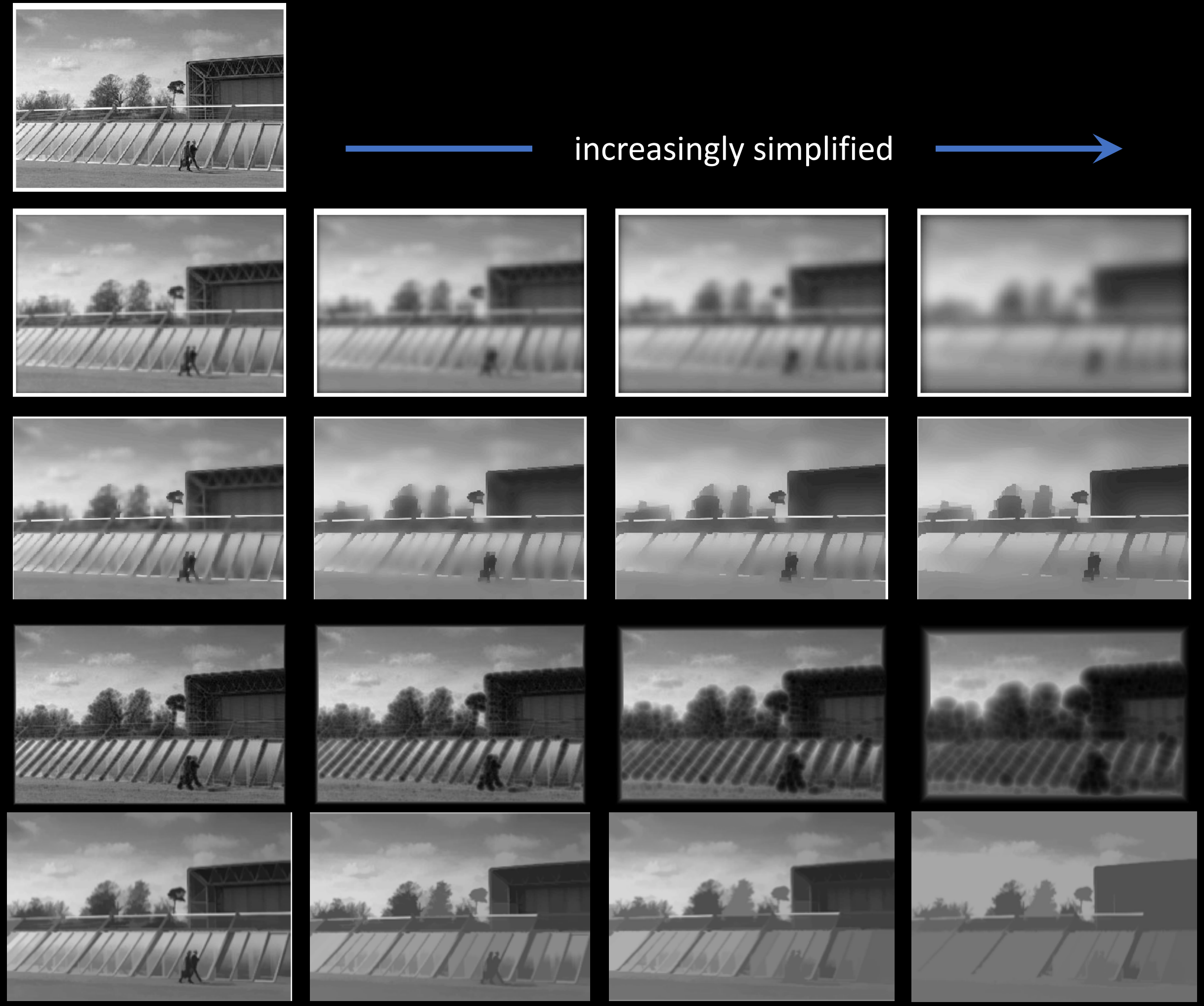
original image

linear diffusion

nonlinear diffusion

elliptic poweroid

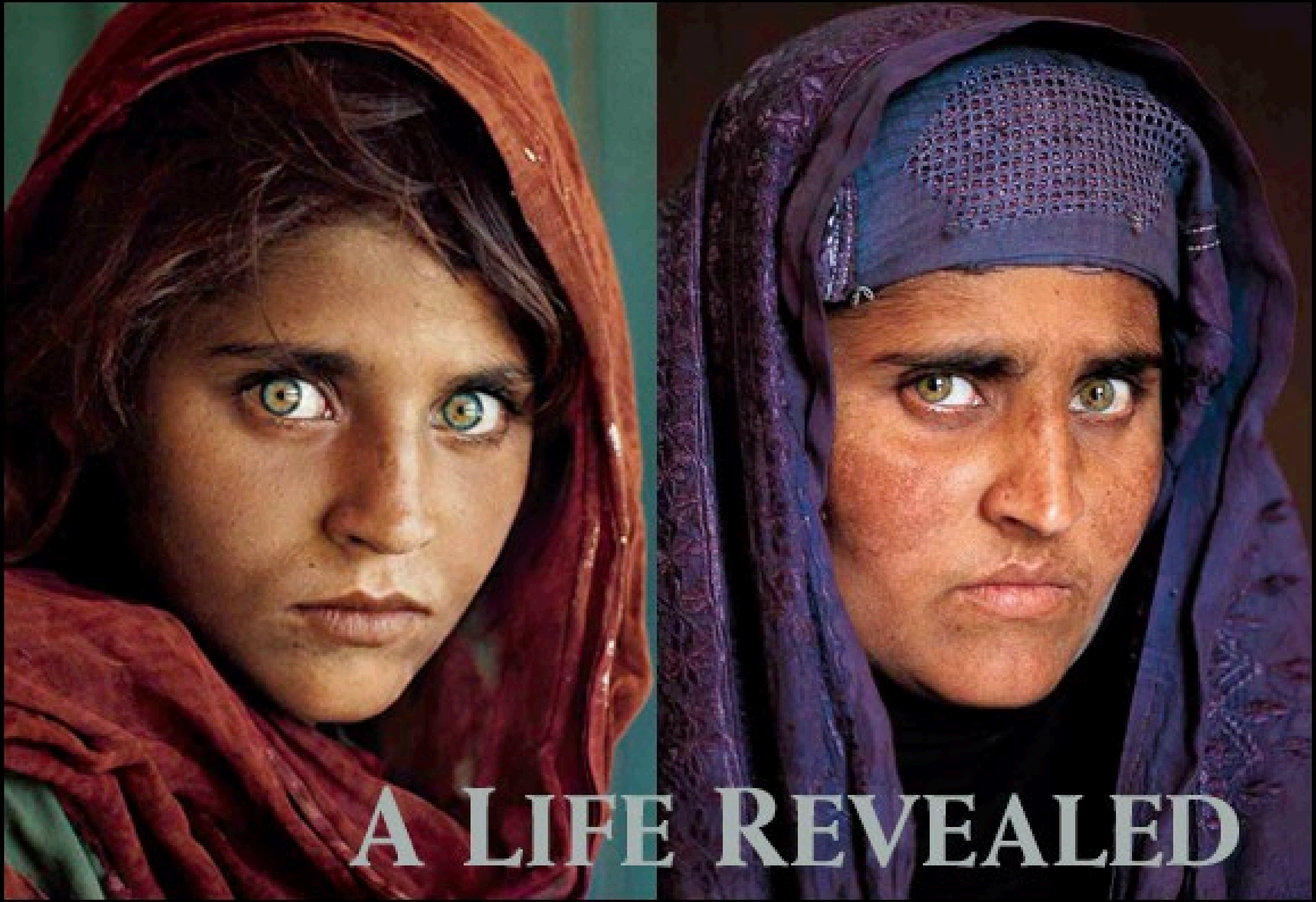
sieve



Coarse-to-fine strategies

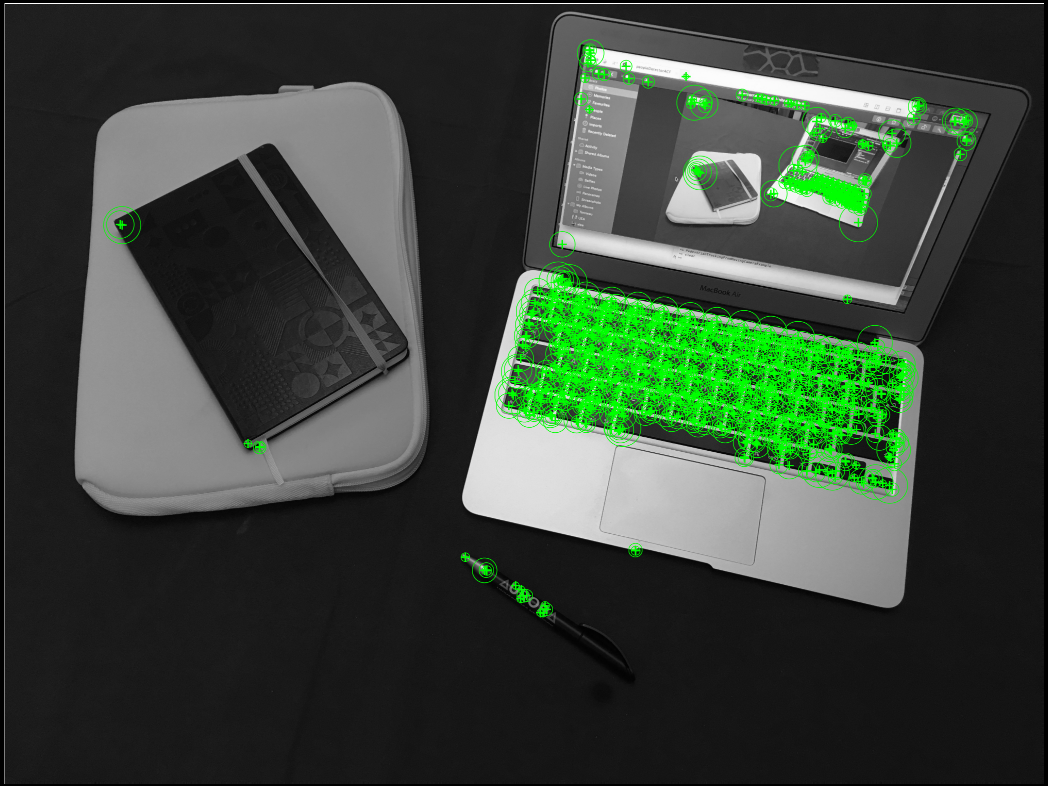
- Scale space
- Wavelets
- Ad hoc smoothing and down-sampling
- Lines and corners via the *primal sketch*

Iris sca



New method
B, Vol 37, No

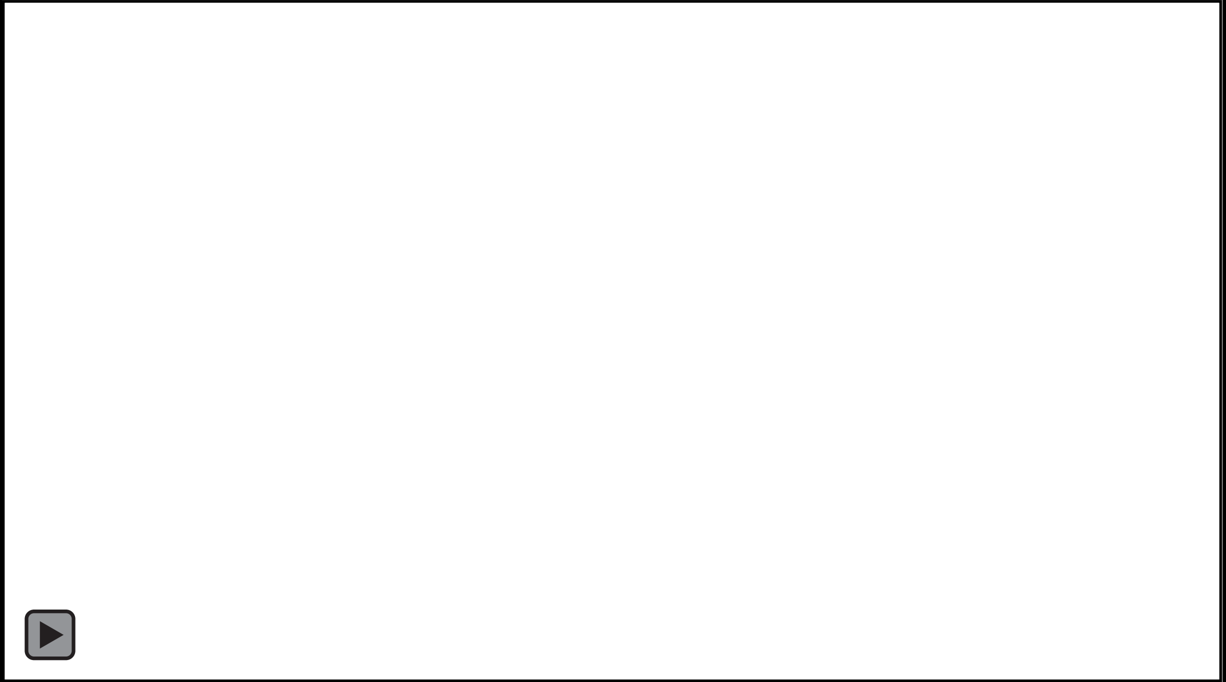
Interest-point operators



Speeded-up Robust Features (SURF), H Bay, A Ess, Ttuytelaars ,L van Gool, Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346--359, 2008

Face tracking

People tracking



Dollar, P., R. Appel, S. Belongie, and P. Perona. "Fast Feature Pyramids for Object Detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 36, Issue 8, 2014, pp. 1532–1545.

But what about machine learning?

Porn detection

Interesting example of *supervised* machine learning

11,005 images, Classified by eye: porn (1994); nudity (1973); people (1626); portraits (1803); graphics (1767); miscellaneous (1842)

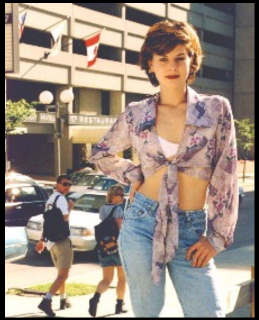
Alison Bosson, Gavin C. Cawley, Yi Chan and Richard Harvey, *Non-retrieval: blocking pornographic images* in Proceedings International Conference on the Challenge of Image and Video Retrieval, Lecture Notes in Computer Science, Vol 2383, pp 60 — 60, Springer-Verlag, London, 2002

Segmentation

Feature extraction

Classification

STEP 1: SEGMENTATION

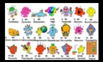


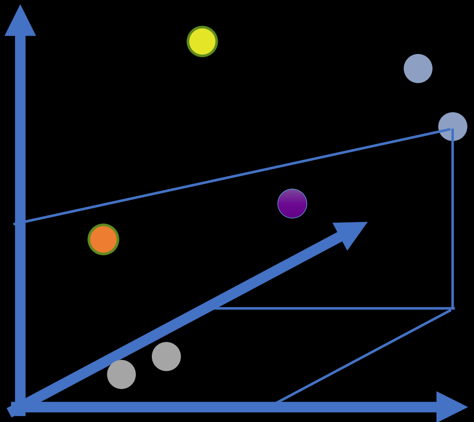
STEP 2: FEATURE EXTRACTION

Num of skin blobs, Skin area, Number of colours …

4, 0.15, 16436, …

STEP 3: CLASSIFICATION

	3	0.2	0.5
	3	0.1	0.6
	1	0.01	0.1
	1	0.03	0.01
	1	0.02	0.5
	3	0.02	0.5
	0	0.01	0.7



STEP 3: CLASSIFICATION

3 0.2 0.5

3 0.1 0.6

Feature vectors Classes

1 0.01 0.1

Training data

1 0.03 0.01

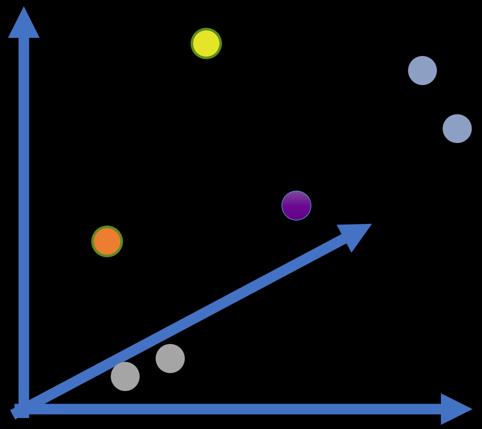
1 0.02 0.5

3 0.02 0.5

Test data

0 0.21 0.3 ?

0 0.01 0.7



CONFUSION MATRIX

	naughty	innocent
naughty	0.89	0.11
innocent	0.15	0.85

Ten years ago…

- Numerous segmentation algorithms
- Even more numerous machine learning algorithms
- Feature extraction
  - a dark art
  - one overarching theory – scale-space but widely misunderstood and ignored

And then along came…

DEEP LEARNING

er

omdeiningerart.com

Next lectures:

learning (19th March)
text (16th April)
creativity (28th May)