Computer Vision

Computer Science Tripos Part II

Dr Christopher Town

1. Overview. Goals of computer vision; why they are so difficult.

UNIVERSITY OF CAMBRIDGE

1. Overview. Goals of computer vision; why they are so difficult.
2. Image sensing, pixel arrays, cameras. Image coding. Biological vision.
3. Mathematical operations for extracting structure from images.
4. Edge detection operators; the Laplacian and its zero-crossings.
5. Multi-scale feature detection and matching.
6. Texture, colour, stereo, and motion descriptors. Disambiguation.
7. Lambertian and specular surface properties. Reflectance maps.
8. Shape description. Codons; superquadrics and surface geometry.
9. Perceptual psychology and visual cognition. Visual illusions.
10. Bayesian inference. Classifiers; probabilistic methods.
11. Learning and statistical methods in vision. Optical character recognition and Content based image retrieval.
12. Face detection, face recognition, and facial interpretation.

- **Goals of computer vision; why they are so difficult.** How images are formed, and the ill-posed problem of making 3D inferences from them about objects and their properties.
- **Image sensing, pixel arrays, cameras.** Elementary operations on image arrays; coding and information measures. Sampling and aliasing. Biological vision.
- **Mathematical operators for extracting image structure.** Finite differences and directional derivatives. Filters; convolution; correlation. Fourier and wavelet transforms.
- **Edge detection operators; the information revealed by edges.** The Laplacian operator and its zero-crossings. Logan's theorem.
- **Multi-scale feature detection and matching.** Gaussian pyramids and SIFT (scale-invariant feature transform). Active contours; energy-minimising snakes. 2D wavelets as visual primitives.
- **Texture, colour, stereo, and motion descriptors.** Disambiguation and the achievement of invariances. Image and motion segmentation.

- **Lambertian and specular surfaces.** Reflectance maps. Image formation geometry. Discounting the illuminant when inferring 3D structure and surface properties.
- **Shape representation.** Inferring 3D shape from shading; surface geometry. Boundary descriptors; codons; superquadrics and the "2.5-Dimensional" sketch.
- **Perceptual psychology and visual cognition.** Vision as model-building and graphics in the brain. Learning to see. Visual illusions, and what they may imply about how vision works.
- **Bayesian inference in vision; knowledge-driven interpretations.** Classifiers and pattern recognition. Probabilistic methods in vision.
- **Applications of machine learning in computer vision.** Discriminative and generative methods. Optical character recognition. Content based image retrieval.
- **Approaches to face detection, face recognition, and facial interpretation.** Appearance and model based representations. 2D and 3D approaches. Cascaded detectors.

Objectives

At the end of the course students should:

- understand visual processing from both "bottom-up" (data oriented) and "top-down" (goals oriented) perspectives;
- be able to decompose visual tasks into sequences of image analysis operations, representations, specific algorithms, and inference principles;
- understand the roles of image transformations and their invariances in pattern recognition and classification;
- be able to describe and contrast techniques for extracting and representing features, edges, shapes, and textures
- be able to analyse the robustness, brittleness, generalizability, and performance of different approaches in computer vision;
- understand some of the major practical application problems, such as face interpretation, character recognition, and image retrieval.

Recommended books

* Forsyth, D. A. & Ponce, J. (2003). *Computer Vision: A Modern Approach.* Prentice Hall.

Shapiro, L. & Stockman, G. (2001). *Computer vision.* Prentice Hall.

Further resources can be found on the course website:

http://www.cl.cam.ac.uk/teaching/1011/CompVision

- Exercises (to be discussed in supervisions or examples classes)
- Links to demos, sample code, research papers, libraries etc.

Online resources

The OpenCV Computer Vision Library: [an excellent C++ open source library with interfaces for some other languages]
http://opencv.willowgarage.com

"CVonline: The Evolving, Distributed, Non-Proprietary, On-Line Compendium of Computer Vision" (Edinburgh University):
http://homepages.inf.ed.ac.uk/rbf/CVonline/

Matlab Functions for Computer Vision and Image Processing:
http://www.csse.uwa.edu.au/~pk/Research/MatlabFns

Annotated Computer Vision Bibliography:
http://iris.usc.edu/Vision-Notes/bibliography/contents.html

"The Computer Vision Homepage" (Carnegie Mellon University): [somewhat outdated]
http://www-2.cs.cmu.edu/~cil/vision.html

Computer Vision is concerned with automated interpretations and representations of visual information

General goals are the detection, recognition and characterisation of objects, scenes, and events of interest

CV is increasingly an engineering discipline

Computer Vision systems are generally task-specific, brittle, and perform worse than biological vision systems

Paradox of "cognitive penetrance"

Which two pictures show the same person?

Intra-class variation

Fei-Fei Li.

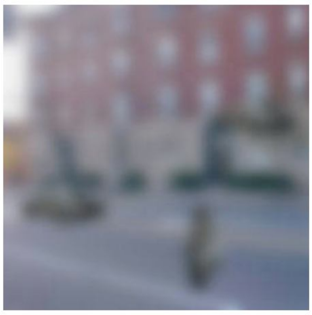
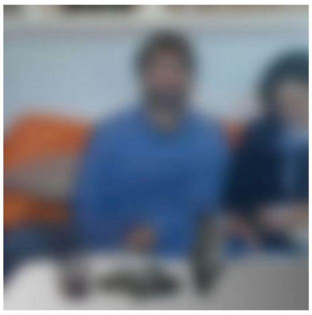
Can you recognise the objects in the following scene?

Antonio Torralba

Dr Chris Town

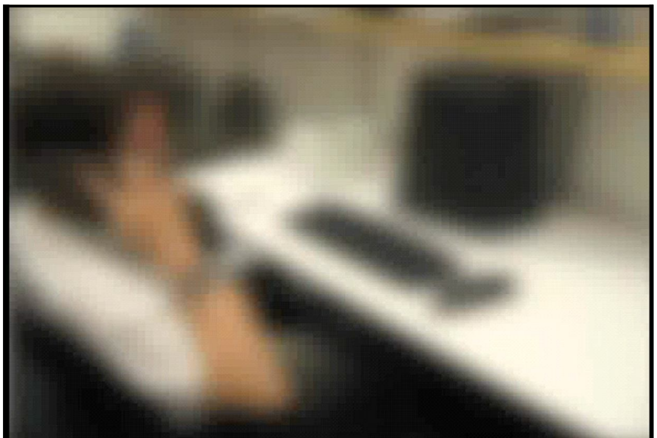
Antonio Torralba

Dr Chris Town

Antonio Torralba

Dr Chris Town



Dr Chris Town

Dr Chris Town

Human face recognition only takes about 100 to 300ms.
It is highly reliable for <u>familiar</u> faces.

Who is in the next image?

Dr Chris Town

P.Sinha

Dr Chris Town

Who was that?

Dr Chris Town

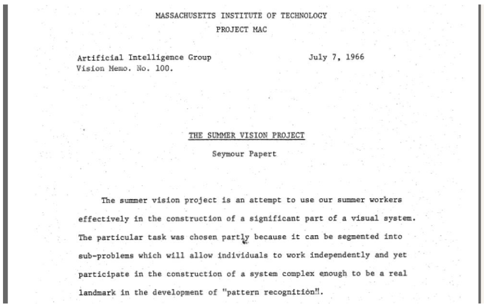
Thompson, P. (1980). "Margaret Thatcher: a new illusion." *Perception* **9**:483-484

Dr Chris Town

Thompson, P. (1980). "Margaret Thatcher: a new illusion." *Perception* **9**:483-484

Dr Chris Town

Computer Vision as an undergraduate summer project at MIT in 1966 – urban myth?.



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group — July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

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

Dr Chris Town

Biological vision is usually robust and appears effortless without cognitive insight

But:

Biological vision systems also have a number of weaknesses and “failure cases” (e.g. visual illusions)

Computer Vision is proving useful in an increasing number of application areas

Dr Chris Town

Some Applications of Computer Vision

- automatic face recognition, and interpretation of expression
- visual guidance of autonomous vehicles
- automated medical image analysis, interpretation, and diagnosis
- robotic manufacturing: manipulation, grading, and assembly of parts
- OCR (optical character recognition): recognition of printed or handwritten characters and words
- ANPR: automated number plate recognition
- agricultural robots: visual grading and harvesting of produce

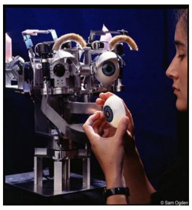
Dr Chris Town

Some Applications of Computer Vision

- smart offices: tracking of persons and objects; understanding gestures
- biometric-based visual identification of persons
- security monitoring and alerting; detection of anomaly
- intelligent interpretive prostheses for the blind
- tracking of moving objects; collision avoidance; stereoscopic depth
- object-based (model-based) compression of video streams
- general scene understanding
- image search and matching by content

Dr Chris Town

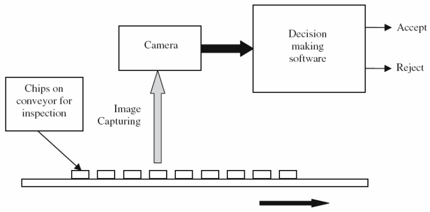
Robotics



Dr Chris Town

Industrial Inspection

Fig. 3 Schematic diagram of marking inspection setup at Texas Instruments



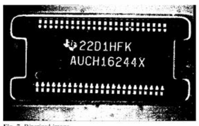
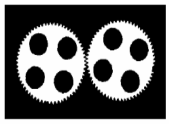


Fig. 7 Binarized image

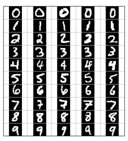
Fig. 9 Row sum for separating a row



R. Nagarajan et al. "A real time marking inspection scheme for semiconductor industries", 2006

Dr Chris Town

Optical Character Recognition



Handwritten digits

Text in images

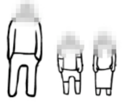
Scanned documents

Dr Chris Town

Face detection

[Face priority AE] When a bright part of the face is too bright

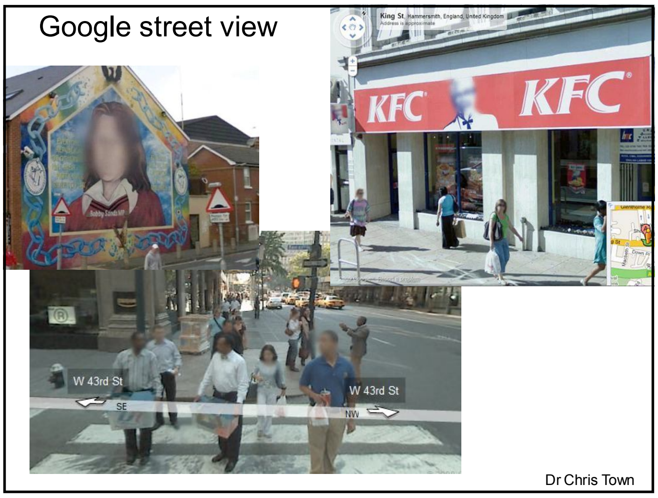
Dr Chris Town



Google street view

Dr Chris Town

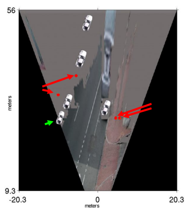
Google street view

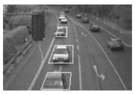


Dr Chris Town

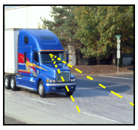
Assisted driving

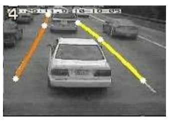
Pedestrian and car detection





Lane detection

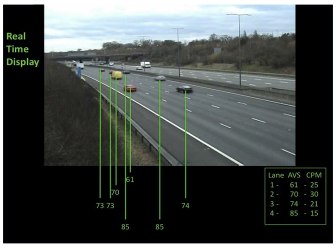


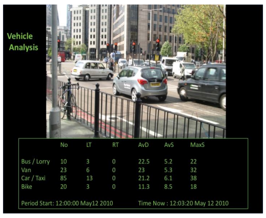


- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,

Dr Chris Town

Traffic monitoring

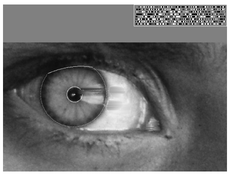


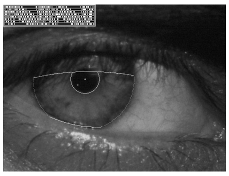


Dr Chris Town

Vision-based biometrics

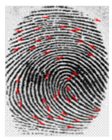
Fingerprints, face, iris, retina, gait, earlobes…





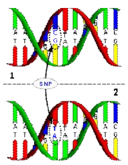
Dr Chris Town

Some examples of biometric methods and applications

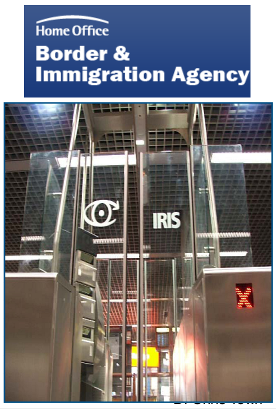


Forensics

"IrisKids" (US) missing children registration and identification



Face recognition ??



Iris recognition

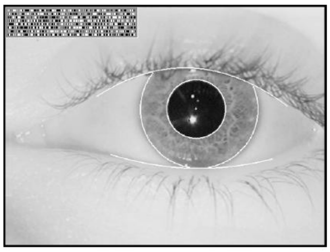
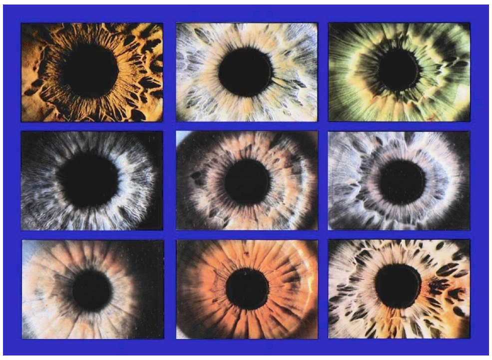


Fig. 1. Example of an iris pattern, imaged monochromatically at a distance of about 35 cm. The outline overlay shows results of the iris and pupil localization and eyelid detection steps. The bit stream in the top left is the result of demodulation with complex-valued two-dimensional (2-D) Gabor wavelets to encode the phase sequence of the iris pattern.

JOHN DAUGMAN

Dr Chris Town

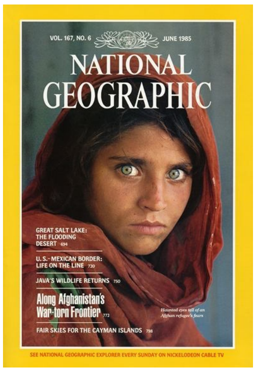
Iris recognition



http://www.cl.cam.ac.uk/~jgd1000/iriscollage.jpg

Dr Chris Town

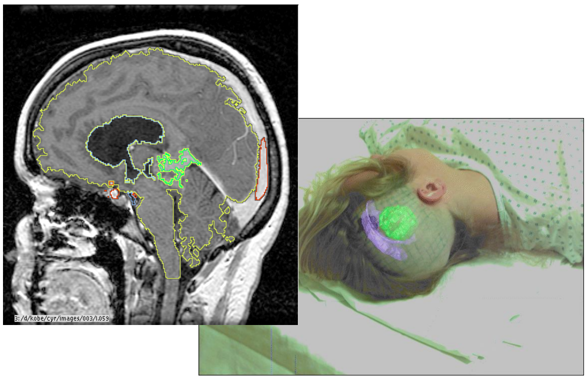
Iris recognition



Sharbat Gula (1984); identified (2002) using the Daugman iris recognition algorithms (based on photographs taken by *National Geographic*)

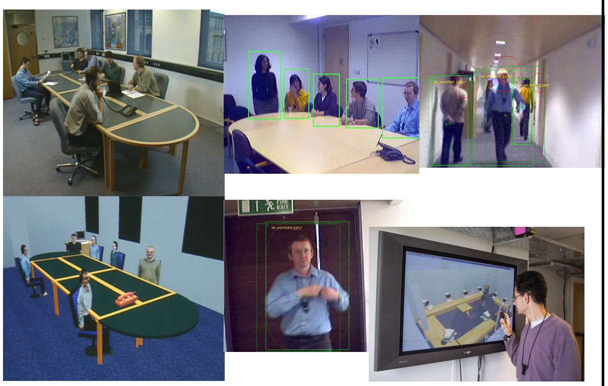
Dr Chris Town

Medical Image Analysis



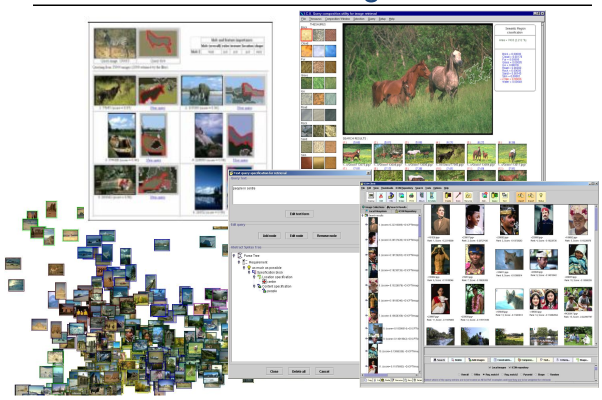
Dr Chris Town

Human-Computer Interfaces



Dr Chris Town

Content-Based Image Retrieval



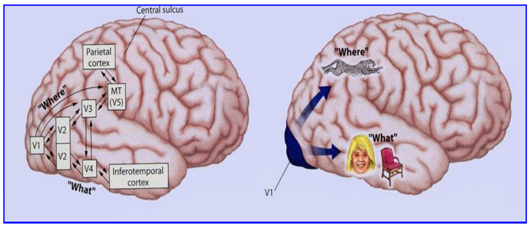
Dr Chris Town

CV can be regarded as being “**AI-complete**” (requires intelligent interpretation of signal data, symbolic reasoning, inference, attention, goal directed behaviour, robust recognition).

Computer Vision as a ***Signal-to-Symbol Converter***

Vision as *Knowing What is Where*

- Vision is hard: noise, ambiguity, complexity
- Prior knowledge is essential to constrain the problem
- Combining multiple cues: motion, contour, shading, texture, stereo



Vision as *Inverse Optics or Graphics*

- Graphics is concerned with rendering images given scene descriptions

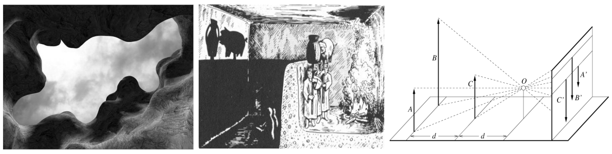
3D ➡2D

- Computer vision seeks to do the opposite

2D ➡3D

-> an ill-posed, heavily underdetermined problem

Plato's Cave



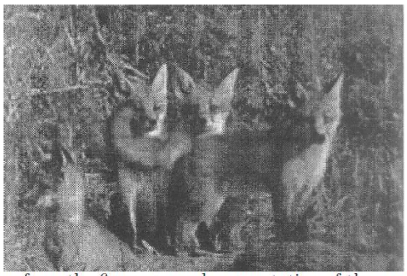
Vision as *Graphics*

pointlight.mov

Vision as *Graphics*

Human visual perceptions are in some sense illusory - elaborate constructs of the brain which are highly generative, context-dependent and knowledge-driven rather than accurate representations or objective descriptions of the visual world.

Vision as inference



- perform the figure-ground segmentation of the scene (into its objects and background)
- infer the 3D arrangements of objects from their mutual occlusions
- infer surface properties (texture, colour) from the 2D image statistics
- infer volumetric object properties from their 2D image projections
- and do all of this in “real time?” (This matters quite a lot in the natural world “red in tooth and claw,” since survival depends on it.)

Vision has to solve an ill-posed problem

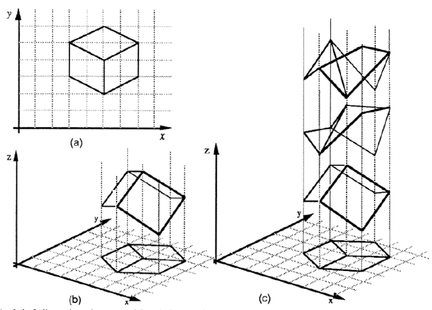


Figure 1. (a) A line drawing provides information only about the x, y coordinates of points lying along the object contours. (b) The human visual system is usually able to reconstruct an object in three dimensions given only a single 2D projection (c) Any planar line-drawing is geometrically consistent with infinitely many 3D structures.

Sinha & Adelson 93

Dr Chris Town

Hadamard's criteria for well-posed problems

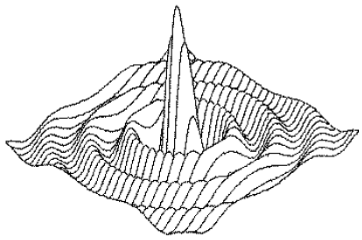
- Solution must exist
- Solution must be unique
- Solution must vary continuously with the data

Dr Chris Town

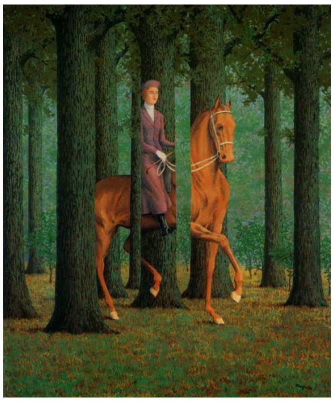
ill-posed problems

- infering depth properties from an image
- infering surface properties from image properties
- infering colours in an illuminant-invariant manner
- infering structure from motion, shading, texture, shadows, ...
- infering a 3D shape unambiguously from a 2D line drawing:
- interpreting the mutual occlusions of objects, and stereo disparity
- recognising a 3D object regardless of its rotations about its three axes in space (e.g. a chair seen from many different angles)
- understanding an object that has never been seen before:

Dr Chris Town

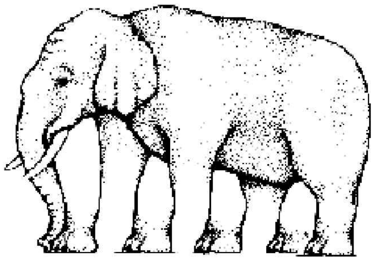


Dr Chris Town

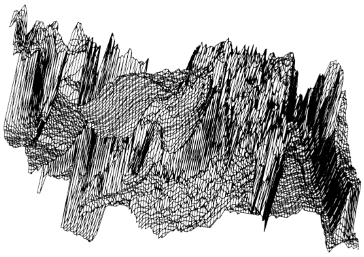


Magritte, 1957

Dr Chris Town



Dr Chris Town



Dr Chris Town

So how do we go from an array of numbers to "perception"?

				49	151	176	182	179
				45	148	175	183	181
				42	146	176	185	184
				35	140	172	184	184
66	64	64	84	129	134	168	181	182
59	63	62	88	130	128	166	185	180
60	62	60	85	127	125	163	183	178
62	62	58	81	122	120	160	181	176
63	64	58	78	118	117	159	180	176

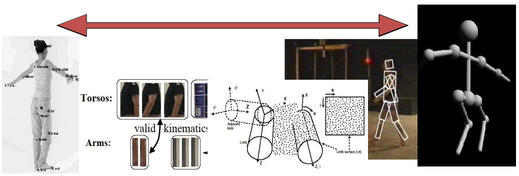
Irani and Basri

Dr Chris Town

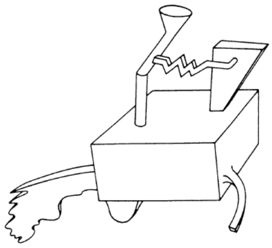
Levels of Representation

- Much of computer vision can be regarded as lying on a continuum between **appearance-based** representations at one end and **explicit models** at the other

- In specific domains, these categories are also known as
  - **view-based** vs **object-centred**
  - **non-parametric** vs **parametric**
  - **exemplars** vs **models**



Dr Chris Town



Dr Chris Town