Computer Vision (7CCSMCVI / 6CCS3COV)

Teaching

Course Web-Page

Available through KEATS.

Lectures

Lecture Slides – available through KEATS.

Recommended Reading – available through KEATS.

Tutorials

Start **next** week

Questions – available through KEATS.

Answers – available on KEATS after corresponding tutorials.

Assessment

Exam

65% of mark

- 3 hours
- Answer 4 out of 6 questions (first question compulsory)

Coursework

35% of mark

3 assessed lab reports

Labs

Start **next** week

Programming exercises in MATLAB – if you don't already know MATLAB you will need to learn!

Timetable - available through KEATS

Instructions - available through KEATS

Deadlines - available through KEATS

Labs

TAs for Lab sessions

Zhuoling Huang

Guangyu Jia

Prabhakar Ray

Mai Alzamel

Labs

Start **next** week

Programming exercises in MATLAB – if you don't already know MATLAB you will need to learn!

Timetable - available through KEATS

Instructions - available through KEATS

Deadlines - available through KEATS

Computer Vision (7CCSMCVI / 6CCS3COV)

Pre-requisites

Geometry

Matrix and Vector Mathematics

Willingness to learn to programme in MATLAB

Contact Details

Use Forums on KEATS

michael.spratling@kcl.ac.uk

Office Hours

Today

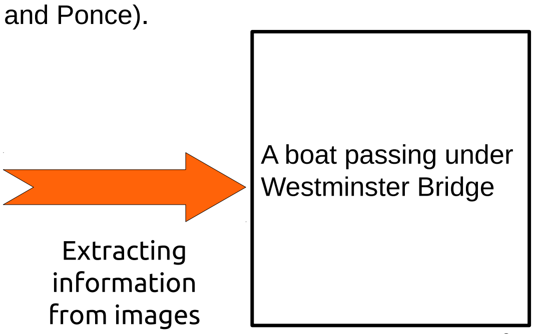
- **What is Computer Vision?**
  - Aims
  - Relation to other subjects
- **Why is it important?**
  - Need to understand biological vision
  - Applications of machine vision
- **Why is it difficult?**
  - One image → many interpretations
  - One object → many images
- **How do we tackle this problem?**
  - Computational approach
  - Biological approach

What is computer vision?

What is computer vision?

Possible Definitions

- "To know what is where, by looking." (Marr).
- "Computing properties of the 3D world from one or more digital images" (Trucco and Verri).
- "To make useful decisions about real physical objects and scenes based on sensed images" (Stockman and Shapiro).
- "Extracting descriptions of the world from pictures or sequences of pictures" (Forsyth and Ponce).



Related disciplines

Image processing – manipulation of an image

Computer graphics – digitally synthesizing images

Image Processing:	Image	→	Image
Computer Vision:	Image	→	Description
Computer Graphics:	Description	→	Image

Pattern recognition – recognising and classifying stimuli in images and other datasets

Photogrammetry – obtaining measurements from images

Biological Vision – understanding visual perception in humans and animals (studied in Neuroscience, Psychology, Psychophysics)

Related disciplines

Computer vision also has other (equivalent) names:

machine vision
image analysis
image understanding
computational vision

Why is Vision Important?

Why is it worth studying?

- Biological Motivation
  - Understanding how we see.
  - Vision is main way in which we experience the world.
  - Evolutionary important:
    - » ~ 50% of cerebral cortex is devoted to vision (→ important, difficult).
    - » Vision consumes ~10% of entire human energy consumption.

- Artificial (Computational) Motivation
  - Want machines to interact with world.
  - Digital images are everywhere.
  - Lots of applications...

Applications

- Industrial inspection, quality control
- Robot navigation
- Autonomous vehicles
- Guiding tools for blind
- Surveillance and security
- Object / face / character recognition
- Medical image analysis
- Digital libraries and video search

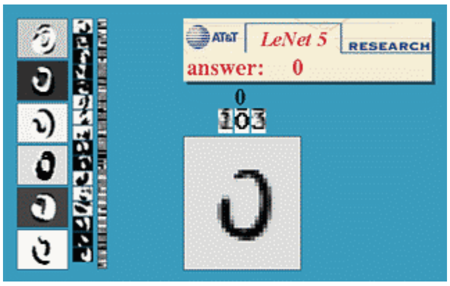
excellent overview of vision industry:
http://www.cs.ubc.ca/spider/lowe/vision.html

The next slides show some examples of what current state-of-the-art vision systems can do...

Appln: character recognition

• Optical character recognition (OCR) converts scanned documents to text.

• Automatic numberplate recognition (ANR) reads car licence plates.



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

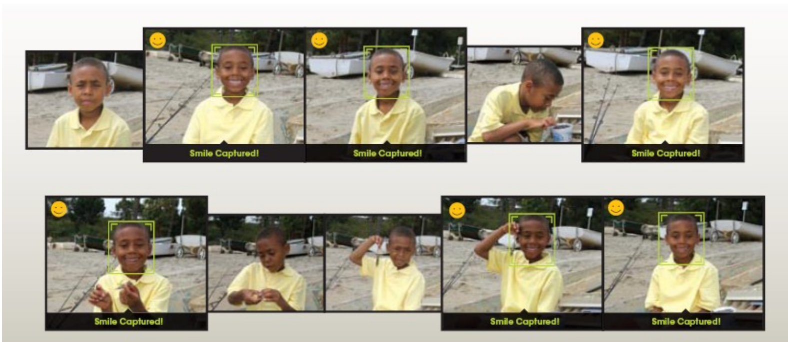
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Appln: face detection

Many digital cameras use face detection to focus on the likely subject.

Some webcams detect faces so that the camera can automatically pan/tilt and zoom to follow someone.

Appln: smile detection



Camera can be set to automatically take photos when a chosen subject laughs, smiles, and grins
e.g. Sony Cyber-shot® T70 Digital Still Camera

Appln: face detection / recognition

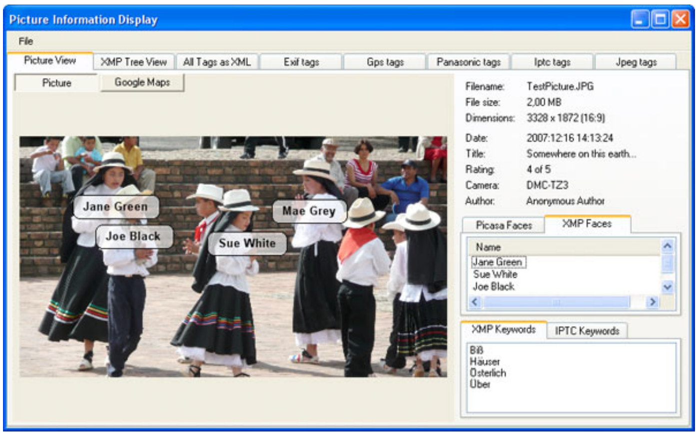


Photo managers detect faces to allow user to tag people.
Recognise tagged people in new photos.
e.g. Picasa

Appln: content-based image retrieval

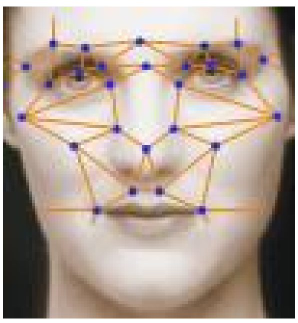
... also known as query by image content.

“beetle” →

results that look like insect

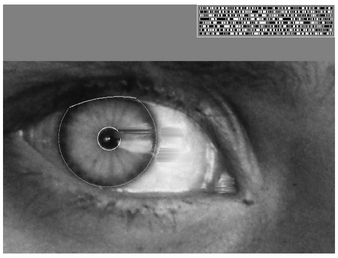
results that look like car

Appln: face recognition



Face recognition
http://www.face-rec.org/
e.g. for controlling access to buildings, identifying terrorists at airports, and for controlling access to computers (e.g. http://www.sensiblevision.com/)

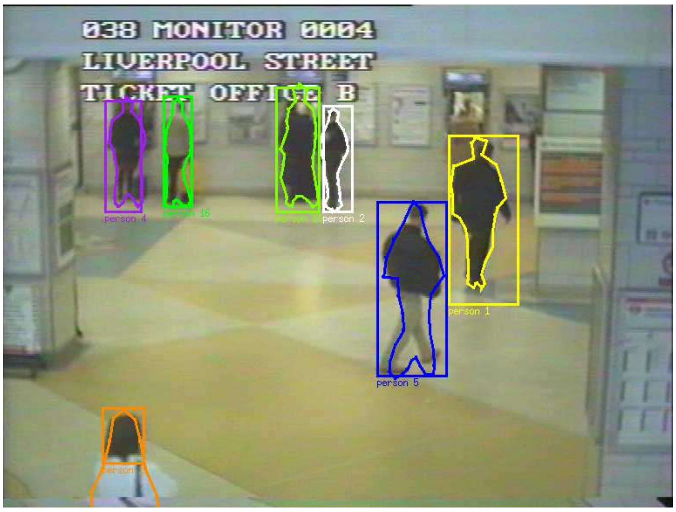
Appln: biometrics



Iris Recognition
http://www.iris-recognition.org/

Fingerprint Recognition.
e.g. Fingerprint scanners on many
laptops, phones, and other devices

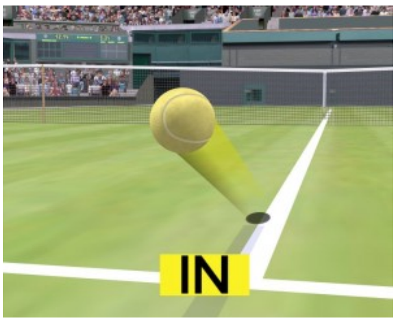
Appln: people tracking

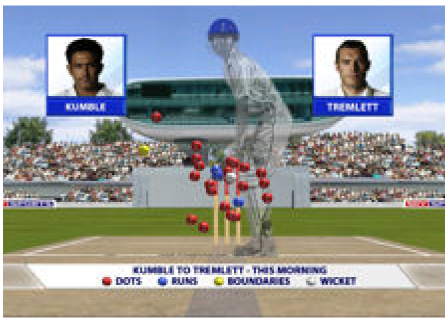


People tracking for visual surveillance and crime detection
(e.g. generate warning if someone is breaking into a car)

Appln: object tracking

e.g. in sport for instant replay and analysis
(http://www.hawkeyeinnovations.co.uk/)





Appln: advertising

Detect ground plane in video and introduce pictures on them.

Appln: place recognition

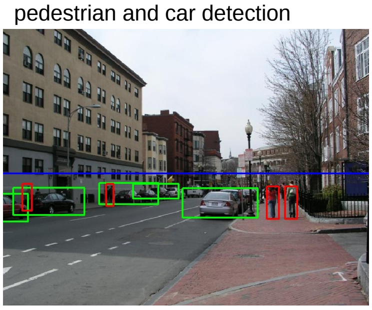
Point & Find by Nokia: lets people point a camera phone at an object or picture and find out more about it.

Appln: driver assistance

lane departure warning

collision warning / automatic braking

pedestrian and car detection



also driver impairment monitoring
e.g. http://www.mobileye.com/

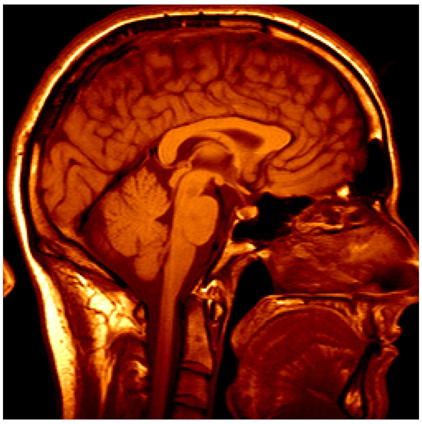
...soon driver replacement, with self-driving cars

Appln: space exploration

Vision systems used for several tasks
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking

(see "Computer Vision on Mars" by Matthies et al.)

Appln: medical imaging

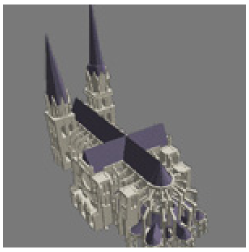


3D imaging in MRI, CT and ultrasound
automatic measurement and analysis

Appln: 3D models from images

 +  +  + ...

From a set of photos of an object or building, generate a 3D virtual (CAD) model that can be viewed from any angle.



e.g
http://photosynth.net/about.aspx
http://www.3dsom.com/

Why is vision difficult?

Note that all the previous examples of vision systems are limited to operating in a specific (small) domain:

- specific task
  - e.g. locate a tennis ball, identify a finger print
- specific environment
  - e.g. on a road, given a frontal view of a face

The challenge of developing general purpose vision systems that can match human performance still remains:

- any task
  - e.g. recognise many different objects
- any environment
  - e.g. under many viewing conditions

Why is vision difficult?

A boat passing under Westminster Bridge

Vision is easy for us, so it is difficult to appreciate how difficult it is to develop algorithms for computer vision.

Why is vision difficult?

210	209	204	202	197	247	143	71
206	196	203	197	195	210	207	56
207	210	211	199	217	194	183	177
201	207	192	201	198	213	156	69
216	206	211	193	202	207	208	57
221	206	211	194	196	197	220	56
209	214	224	199	194	193	204	173
204	212	213	208	191	190	191	214
214	215	215	207	208	180	172	188
209	205	214	205	204	196	187	196
208	209	205	203	202	186	174	185
208	205	209	209	197	194	183	187
149	71	63	55	55	45	56	98
209	90	62	64	52	93	52	76
187	239	58	68	61	51	56	24
86	62	66	87	57	60	48	31

A boat passing under Westminster Bridge

If we replace the image by its numerical representation, the transformation into a description is less obvious.

Major Challenges:

1. **One image → many interpretations**
   problem is ill-posed
2. **One object → many images**
   problem is exponentially large

Vision is an ill-posed problem

Mapping from world to image (3D to 2D) is unique (well-posed).

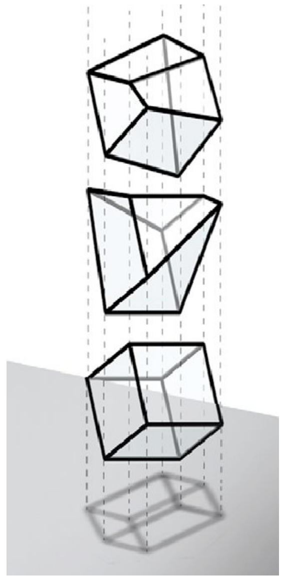
- This is a “forward problem” (i.e. imaging).

Mapping from image to world (2D to 3D) is NOT unique (ill-posed)

- This is an “inverse problem” (i.e. vision).

For any given image there are many objects that could have generated that image.

Solved using constraints or priors: which make some interpretations more likely than others (usually the brain produces one interpretation from the many possible ones).



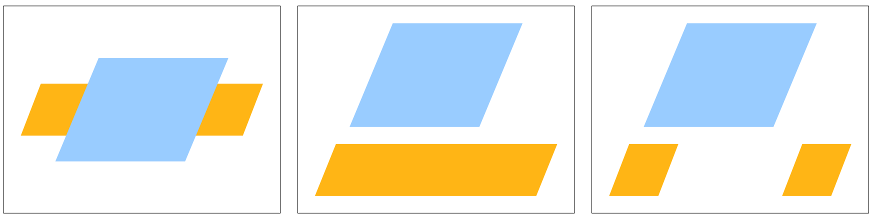
Multiple interpretations of an image

What does this image show?

Multiple interpretations of an image

What does this image show?

Three possible interpretations:

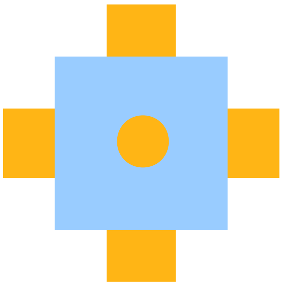


One object

Two objects
Most likely?

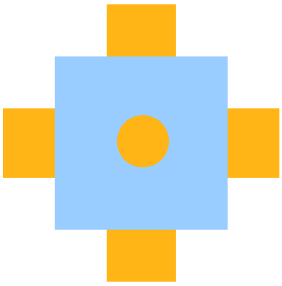
Three objects

Multiple interpretations of an image



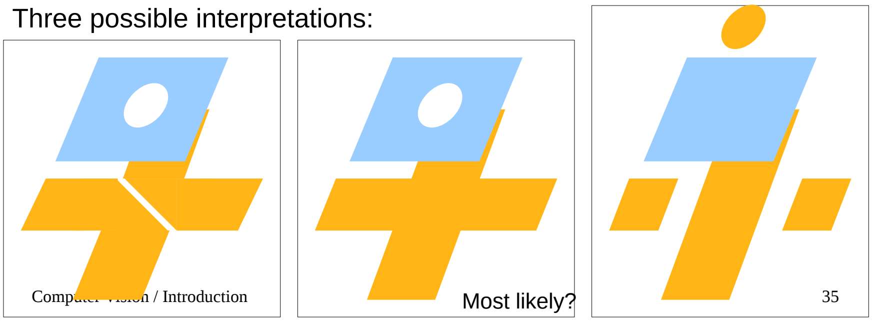
What does this image show?

Multiple interpretations of an image



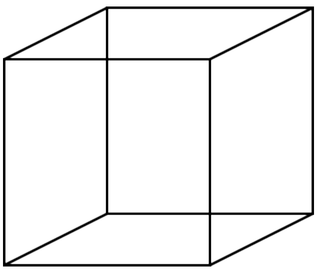
What does this image show?

Three possible interpretations:



Multiple interpretations of an image

Necker cube



Rubin's Face / Vase illusion

In both case there are two possible interpretations which are both equally likely, so either is perceived spontaneously.
Note, differing interpretations never perceived simultaneously.

Vision scales exponentially

Consider trying to recognize an object.

Suppose the object can:

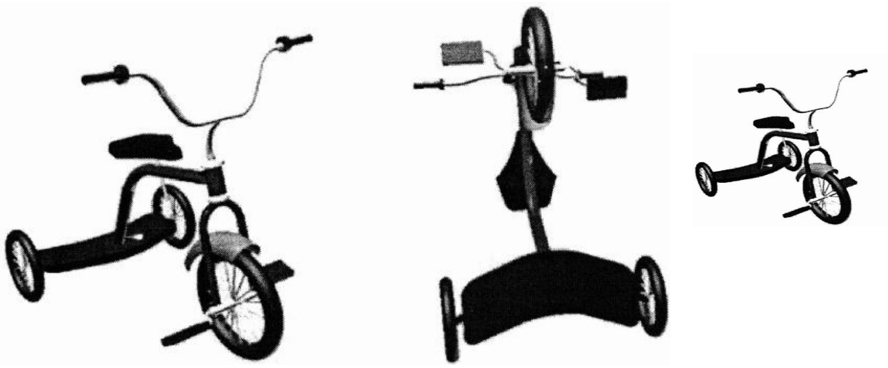
- appear at any one of *l* locations in the image
- appear at any one of *s* different scales (i.e. sizes)
- appear at any one of *o* orientations
- appear in any one of *c* colours
- ...

This one object can give rise to $l \times s \times o \times c$ different images.

The number of images increases exponentially with the number of parameters.

Solved by using invariant representations and priors.

Viewpoint affects appearance



A single object seen from different viewpoints can vary greatly in appearance (object orientation, retinal location, scale, etc. all affect appearance).

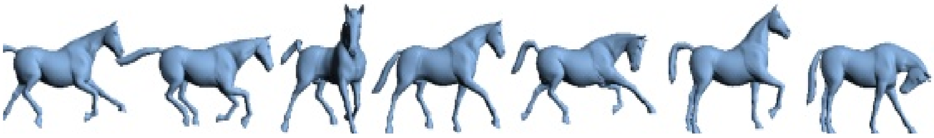
The resulting images have very little similarity.

Illumination affects appearance

A single object seen under different lighting conditions can vary greatly in appearance.

The resulting images have little similarity.

Non-rigid deformations affect appearance



A single object can undergo deformations which cause it to vary greatly in appearance.

The resulting images have little similarity.

Within-category variation in appearance

Objects forming a single category can vary greatly in appearance.

The resulting images have little similarity.

Discrimination despite variation

"Objects that look very similar can be represented and recognized as different objects, whereas objects that look very different can be recognized as the same basic-level objects" (Bar, 2004).

Despite the variation in appearance of a single object, or a single category, it is necessary to be able to distinguish one object/category from another.

Other objects affect appearance

Images usually contain multiple objects.
This leads background clutter and occlusion.
Resulting in images of a single object having little similarity.

Need for constraints (priors)

Previous slides illustrated the two major challenges for Computer Vision:

- One image → many interpretations
- One object → many images

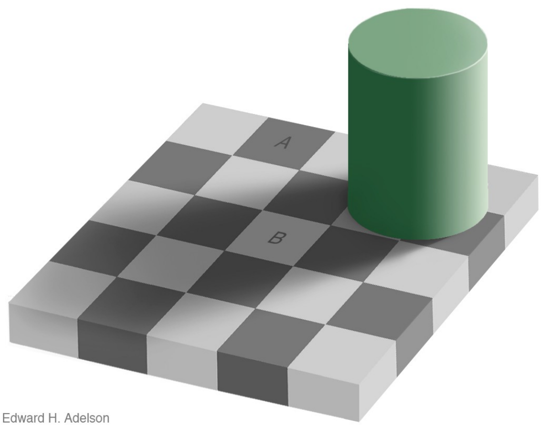
To solve these challenges we need to employ constraints / priors / expectations.

Perception involves inference:

> We must combine prior information about the world with evidence from our senses (e.g. vision) to infer what is in the world.

The next slides illustrate some effects that priors have on human visual inference...

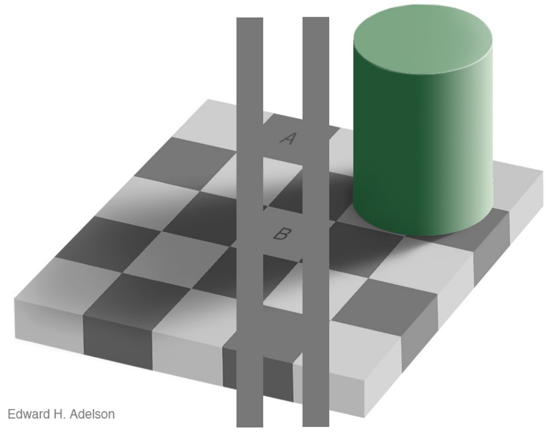
Effects of inference (illumination)



Edward H. Adelson

Which is darker, A or B?
Prior knowledge about shadows results in perceived intensity not reflecting image intensity.

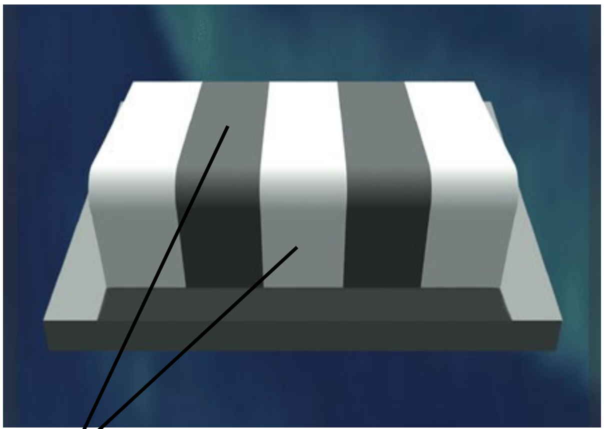
Effects of inference (illumination)



Which is darker, A or B?

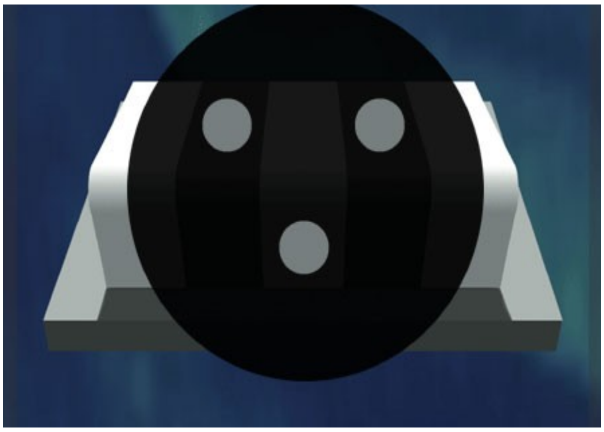
Prior knowledge about shadows results in perceived intensity not reflecting image intensity.

Effects of inference (illumination)



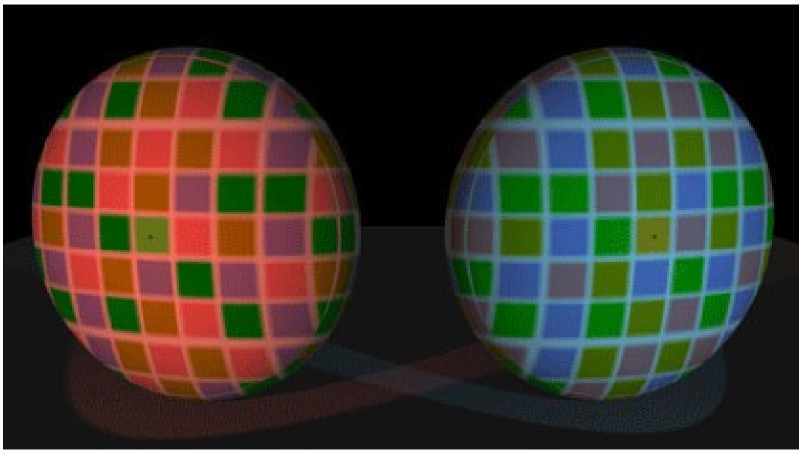
Which is darker?

Effects of inference (illumination)



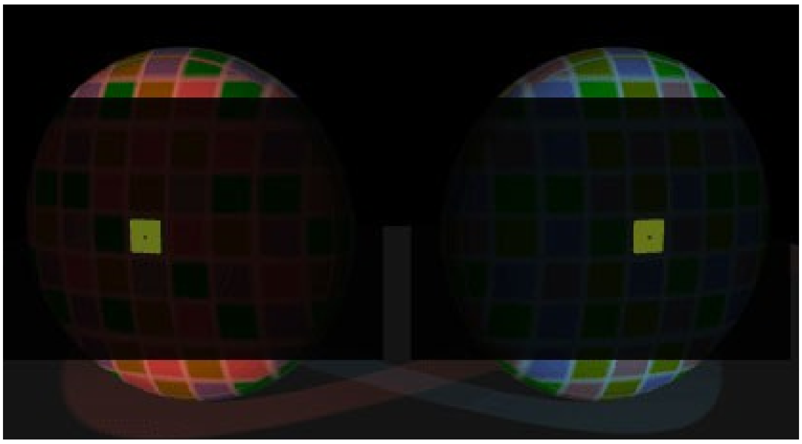
Which is darker?

Effects of inference (illumination)



Are the central patches the same colour?

Effects of inference (illumination)



Are the central patches the same colour?

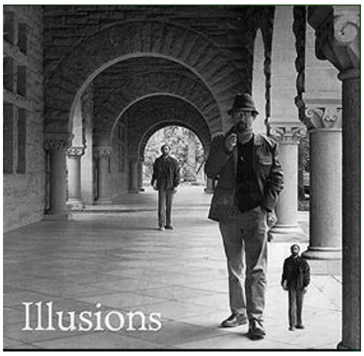
Visual system sees them as different due to inference about different lighting conditions

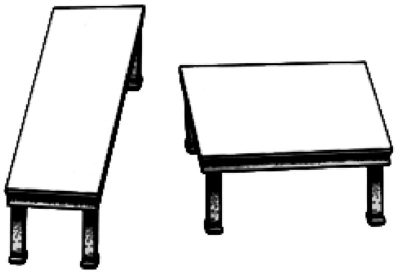
Effects of inference (illumination)

Are the craters convex or concave?

Prior expectation about the direction of illumination (from above) effects the interpretation of a single image.

Effects of inference (perspective)



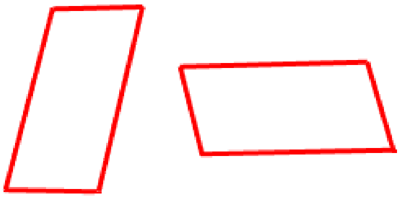


Which is larger?

Which is squarer?

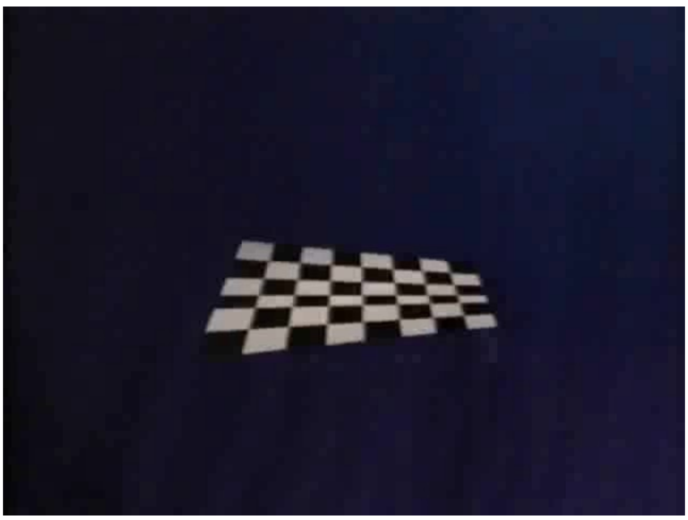
Prior expectation about image formation (perspective geometry) effects size/shape perception

Effects of inference (perspective)



Effects of inference (perspective)

The Ames Room illusion



Effects of inference (prior knowledge)

What does this image show?

Prior knowledge about the image content enables us to easily see something that was previously invisible.

Effects of inference (prior knowledge)

Our prior expectation to see a face is so strong that we see them everywhere.

“Virgin Mary” toast fetches $28,000 on eBay!

Effects of inference (prior knowledge)

Who are these people?

What does this say?

Prior expectation about the likely content of an image prevents us from seeing what is actually there.

Effects of inference (prior exposure)

What is changing in these images?

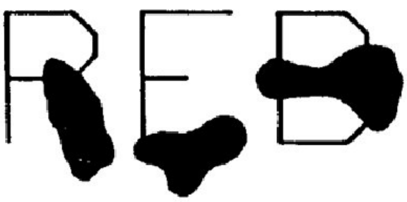
For better demo see:
http://csclab.ucsd.edu/~alan/vision/change_blindness/

Effects of inference (prior exposure)

Vision is sensitive to temporal discontinuities, so a sudden change is easy to spot. Disrupting the temporal continuity (with a flicker, or by flashing up some other stimulus) makes us insensitive to significant changes to the scene. This is called “change blindness”.

Effects of inference (context)

What is the middle letter in each word?



What does this word say?

Contextual information from the whole image enables us to disambiguate parts of the image.

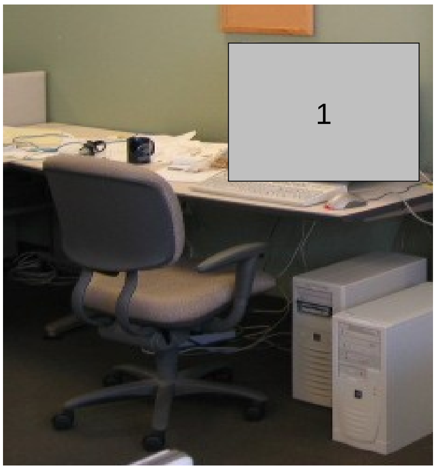
Effects of inference (context)

What are these objects?

Effects of inference (context)

Contextual information from the whole image enables us to disambiguate parts of the image.

Effects of inference (context)



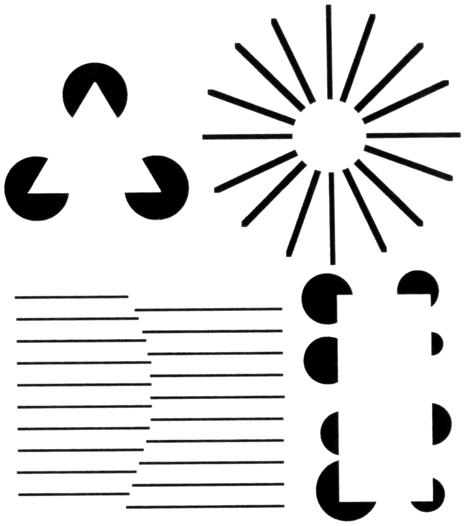
What are the hidden objects?

Effects of inference (context)

Chance ≈ 1/20000!

Contextual information from the whole image enables us to predict contents of parts of the image.

Effects of inference (context)



Prior expectation about the image formation (occlusion) and context causes perception of illusory contours.

Illusions as effects of inference

Several of the preceding examples are illusions.

• Illusions are traditionally considered to reveal "mistakes" made by the visual system.

• However, illusions actually reveal the assumptions that the visual system is making in order to solve the under-constrained problem of vision.

The assumption does not reflect a "flaw" in the visual system but represents an adaptation to the the way things *usually* are.

Our visual system excels because it has learned rules about our world, that work well in typical situations.

Influence of priors on human vision

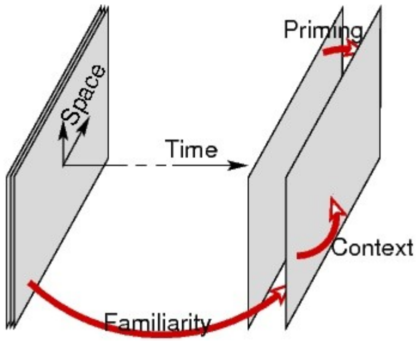
Preceding examples demonstrate that Human perception is influenced by prior expectations coming from many sources.

We can categorise these sources as priors from:

prior knowledge / *familiarity*
- » learned familiarity with certain objects
- » knowledge of image formation process in general

– prior exposure / motion / *priming*
- » recent / preceding sensory input

– current *context*
- » surrounding visual scene (and concurrent input in other sensory modalities)



How do we tackle the problem of vision?

(Forward) Engineering Approach.
- determine what the system needs to do (requirements).
- design a system to perform this task.
- implement the system, test and refine it.
  - "top-down": start with computational theory and fill out details.

Reverse Engineering Approach.
- find a system that performs the task (e.g. the brain).
- analyse the system to determine how it does it.
- implement a new system using the same mechanisms.
  - "bottom-up": start with mechanisms and build a model.

This course will consider both approaches...

This course

...Hence, this course is interdisciplinary.

It considers both:

Artificial (machine) vision:
- How can we get computers to see?
- Implementing algorithms for perception

Biological (human) vision:
- How do people see?
- Modelling biological perception

Course Outline

Introductory course on computer vision – aiming to provide a comprehensive introduction to the main issues and methods.

• **Image formation**: the physics of image formation, cameras, the geometry of image formation, image coding and representation, the eye.

• **Low-level vision**: image processing (filtering, convolution), feature detection (edges), neural representations in V1.

• **Mid-level vision**: grouping and segmentation, the correspondence problem, stereo and depth, video and motion, Gestalt principles, border ownership.

• **High-level vision**: object recognition and categorisation.

Summary

- Vision is concerned with determining properties of the world from images.
- Vision is difficult due to the problem being ill-posed (one image can have many interpretations) and being exponentially large (one object can generate many images).
- Overcoming these problems requires a combination of prior information and from evidence within the image in order to make inferences about image content.
- Prior knowledge about a restricted domain has enabled the development of many impressive vision applications.