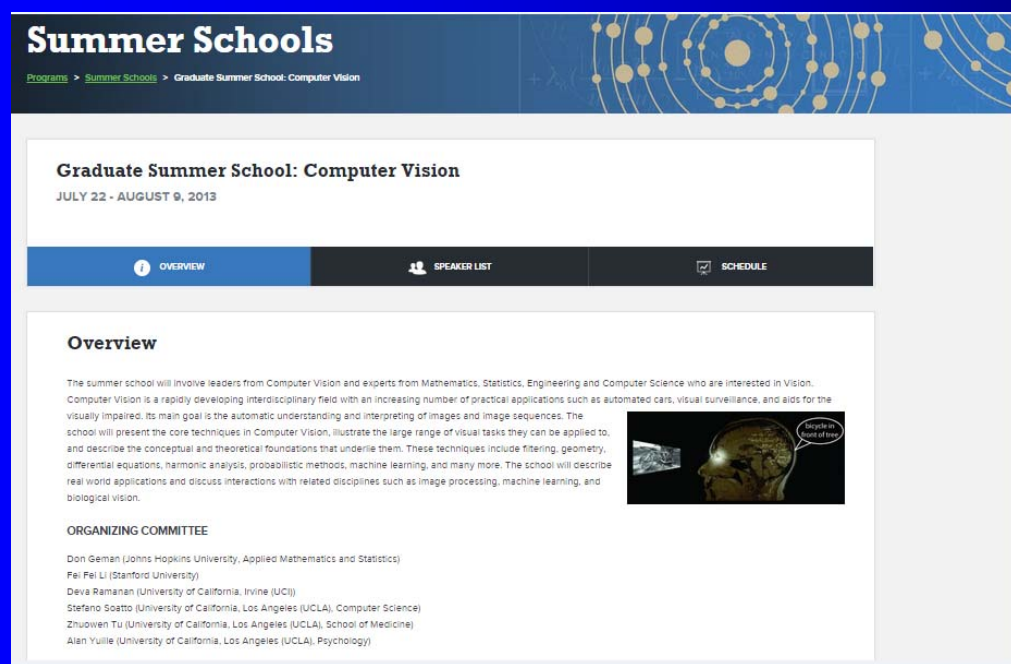


Introduction to Vision

Alan L. Yuille.
UCLA.

IPAM Summer School 2013

- 3 weeks of online lectures on Vision.
- “What papers do I read in computer vision?
There are so many and they are so different.”



The screenshot shows the website for the IPAM Summer Schools. The header features the text "Summer Schools" in a bold, sans-serif font, with a navigation path below it: "Programs > Summer Schools > Graduate Summer School: Computer Vision". The main content area is titled "Graduate Summer School: Computer Vision" and includes the dates "JULY 22 - AUGUST 9, 2013". Below this, there are three tabs: "OVERVIEW" (selected), "SPEAKER LIST", and "SCHEDULE". The "Overview" tab is active, displaying a paragraph about the summer school's focus on computer vision and interdisciplinary research. To the right of the text is a small image of a human head with a brain, labeled "Structure in the brain". Below the text, there is a section titled "ORGANIZING COMMITTEE" listing several names and their affiliations.

Summer Schools
Programs > Summer Schools > Graduate Summer School: Computer Vision

Graduate Summer School: Computer Vision
JULY 22 - AUGUST 9, 2013

OVERVIEW | SPEAKER LIST | SCHEDULE

Overview

The summer school will involve leaders from Computer Vision and experts from Mathematics, Statistics, Engineering and Computer Science who are interested in Vision. Computer Vision is a rapidly developing interdisciplinary field with an increasing number of practical applications such as automated cars, visual surveillance, and aids for the visually impaired. Its main goal is the automatic understanding and interpreting of images and image sequences. The school will present the core techniques in Computer Vision, illustrate the large range of visual tasks they can be applied to, and describe the conceptual and theoretical foundations that underlie them. These techniques include filtering, geometry, differential equations, harmonic analysis, probabilistic methods, machine learning, and many more. The school will describe real world applications and discuss interactions with related disciplines such as image processing, machine learning, and biological vision.

ORGANIZING COMMITTEE

Don German (Johns Hopkins University, Applied Mathematics and Statistics)
Fei Fei Li (Stanford University)
Deva Ramanan (University of California, Irvine (UCI))
Stefano Soatto (University of California, Los Angeles (UCLA), Computer Science)
Zhuowen Tu (University of California, Los Angeles (UCLA), School of Medicine)
Alan Yuille (University of California, Los Angeles (UCLA), Psychology)

Main Points of this Talk

- (I) Vision is extremely hard. Humans are vision experts, Vision is very complex and only partially understood.
- (III) Why is vision hard? *Complexity of images, ambiguity of images, complexity of visual tasks.*
- (IV) Taxonomy: Low-, middle-, high-level vision. Course structure.

The Purpose of Vision

- “To Know What is Where by Looking”. Aristotle. (384-322 BC).
- Information Theory: receive a signal by light rays and decode its information to understand the three-dimensional scene.
- *Vision appears deceptively simple to humans, but this is highly misleading.*

Humans can understand very complex images



- We can get the “gist” of an image in about 150 msec, the time to blink an eye.

But we make mistakes:
Accidental alignments fool us.



We are subject to

- Attention blindness.
- Change blindness.
- https://www.youtube.com/watch?v=IGQmdoK_ZfY
- Policeman was sent to jail because he was chasing a criminal and didn't stop to break up a fight. Was he guilty?

We don't agree on what we see

- The dress illusion:
- <http://www.michaelbach.de/ot/col-dress/index.html>
- This is an example of colour constancy. The colour that reaches our eye is a product of the colour of the object and of the lighting. We don't know the colour of the lighting, so our eyes guess it and sometimes are wrong.

But we can get information from very little

- From dots to people: biological motion
- http://www.michaelbach.de/ot/mot_biomot/index.html
- But computer vision can do better (3d from single image).
- *Also juggler.*

But we can perform many visual tasks

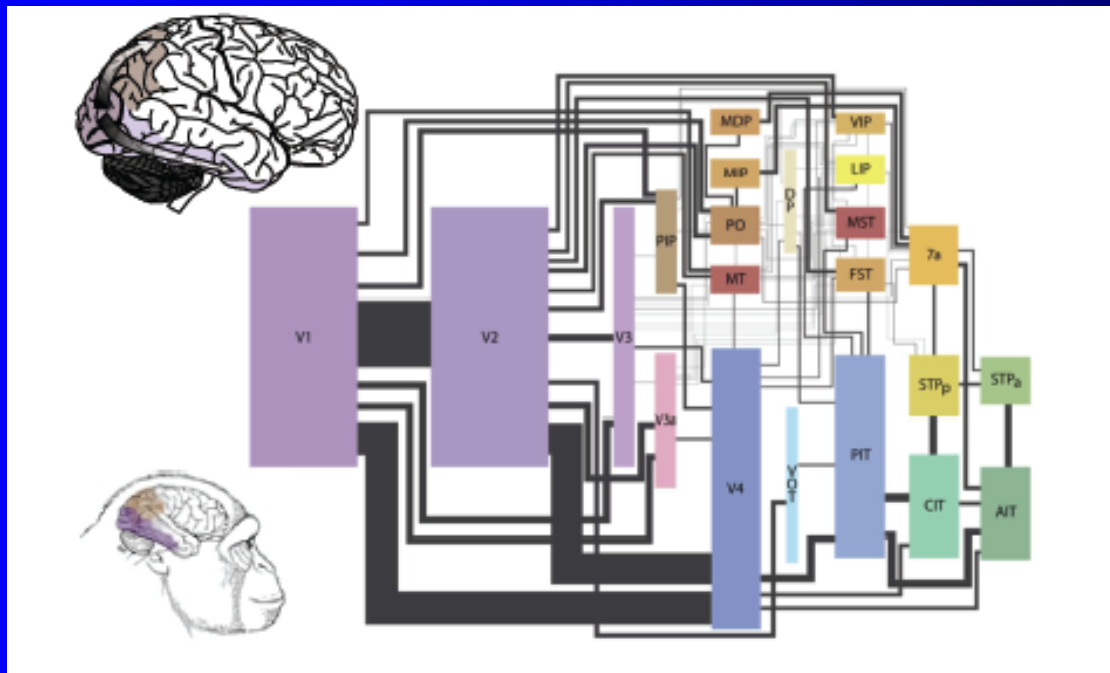
- And so can our relatives:
- https://www.youtube.com/watch?v=KY8c_a5YPDQ
- Why waste our brain power (takes 25% of our total energy) to do vision?
- Maybe our ancestors evolved because they could do vision? – For hunting, for foraging, for social interactions.

Vision and the Brain

- Humans appear to understand images effortlessly. But this is only because of the enormous amount of our brains that we devote to visual tasks.
- *It is estimated that 40-50% of neurons in the cortex are involved in doing vision.*
- Humans are very visual. We get much more information from our eyes than other animals.

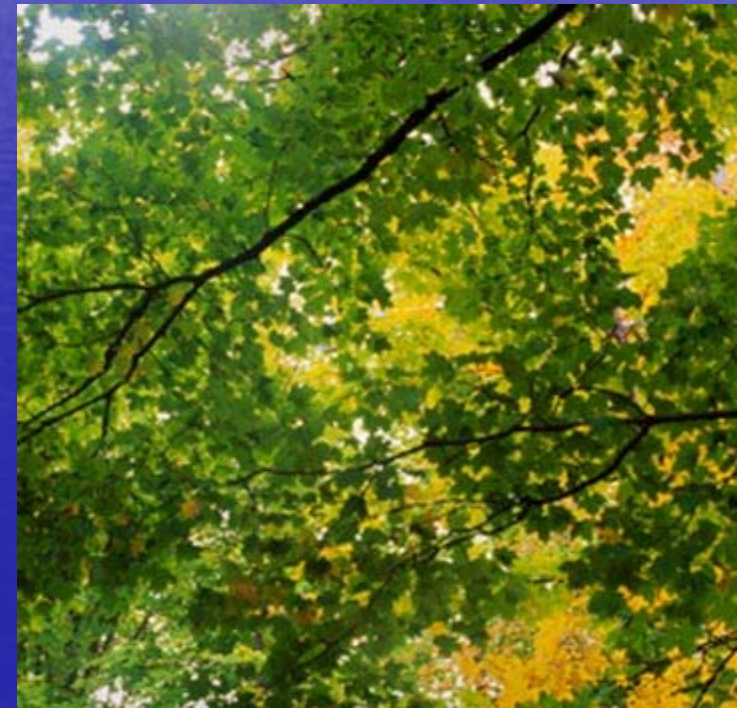
The Human and Monkey Visual System

- The visual cortex is roughly percent of the size of our cortex (the intelligent part of the brain)



Vision and the Brain: complexity

- The number of nerve cells in brain 100,000,000,000 is about the number of trees in the Amazon rainforest.
- The number of nerve cell connections 1,000,000,000,000,000 is about the number of leaves in rainforest.
- The number of connections in the world's telephone system (biggest machine on the planet).
- The brain is the most complicated (known) system in the Universe.
- The neurons in your brain would reach to the moon and back.



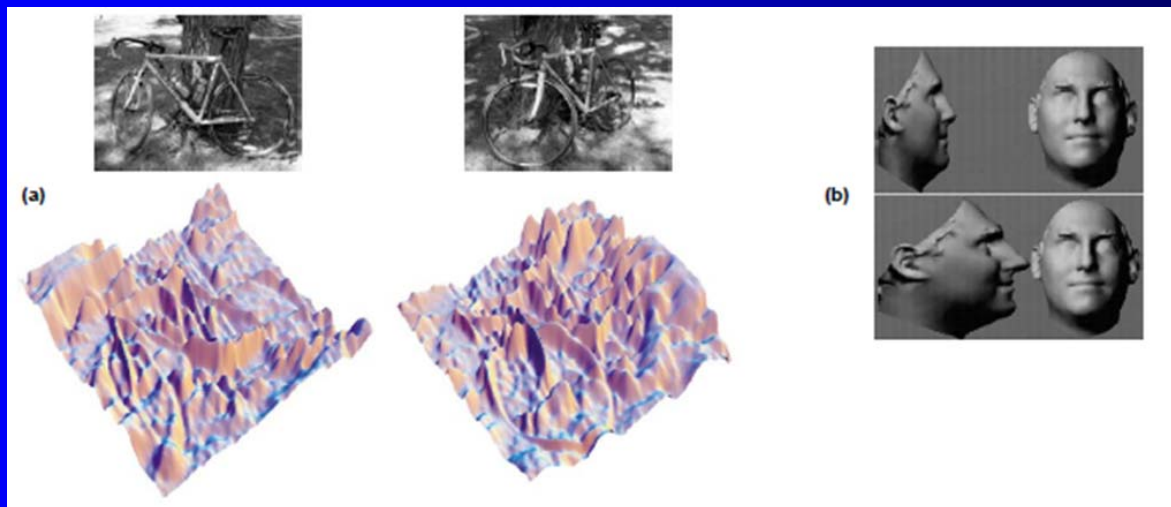
Why is Vision hard?



- The main difficulty of vision is due to the *complexity of images*, the *local ambiguity of images*, and the *complexity of visual tasks*.
- Artificial Intelligence researchers were too ambitious in the 1960's. "Solve vision in the summer".
- Now we have *computer power*, and are *developing the theories*, to deal with complexity.
- "The twenty-first century will be the century of complexity" S.W. Hawking (2000).

The Raw Input – Image Patterns

- The raw input is a set of numbers (bottom left).
- The image patterns are very complex.
- The patterns are very different even for similar objects (same bike and tree – changed viewpoint).



Vision can be very hard

- Hidden object:
- http://www.michaelbach.de/ot/cog_dalmatian/index.html
- Motion can help.

The Complexity of Images

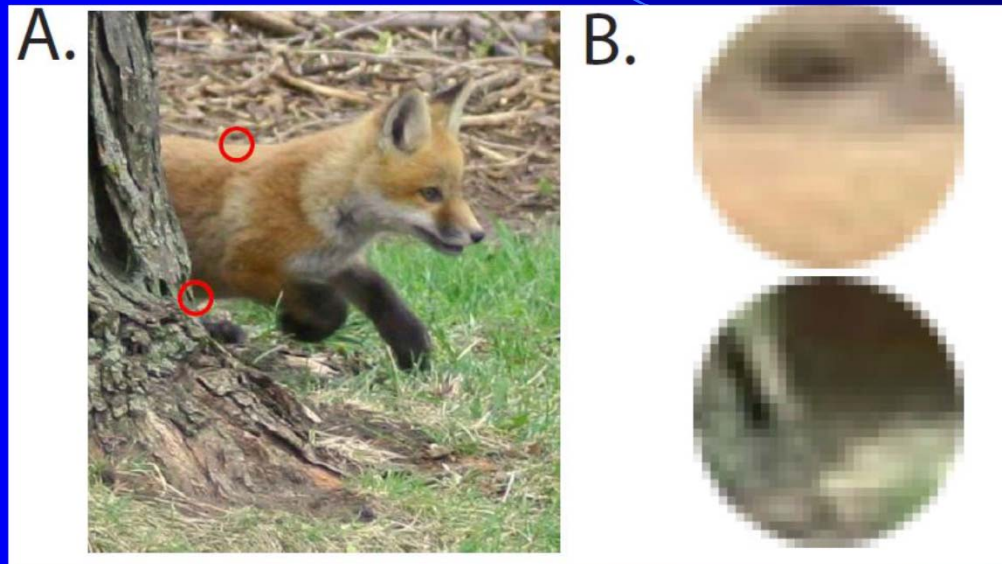
- The set of all images is practically infinite.
- *Only a tiny fraction of all possible images have been seen by humans.*
- The number of visual scenes is also enormous. There are 30,000 different types of objects. They can be arranged into 1,000 scene categories. This can be done in an exponential number of ways.

The Local Ambiguity of Images

Airplane
Car
Boat
Sign
Building



The Complexity of Visual Tasks



- Humans can easily detect the fox, the tree trunk, the grass and the background twigs.
- And can also estimate the shape of the fox's legs and head, its type fur, what it is doing, is it old or young, is this winter or summer?
- But the local regions are highly ambiguous.

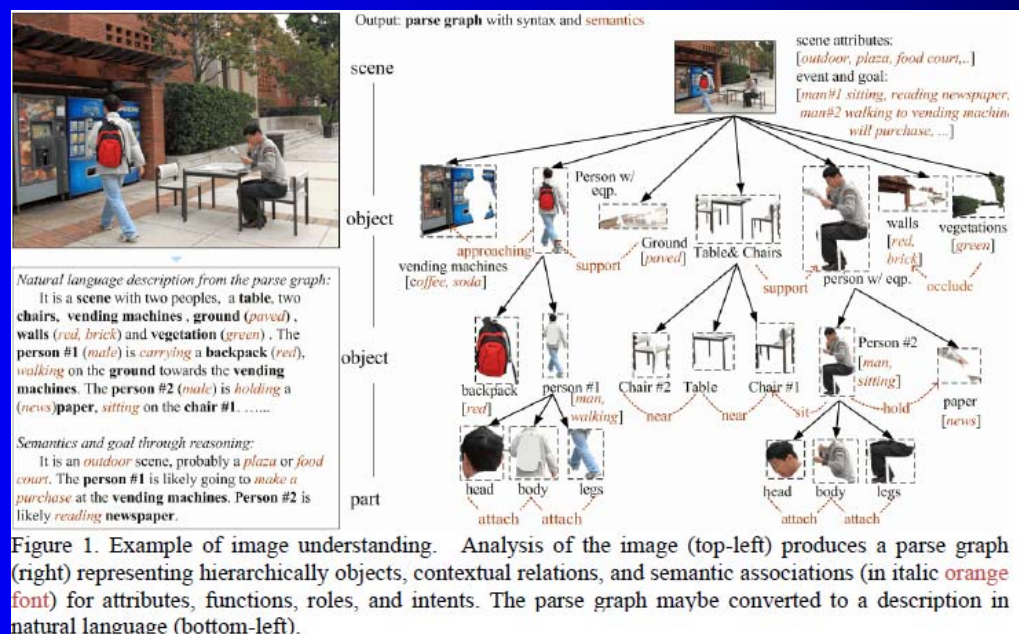
Visual Tasks and Image Detail

- Interpret images, reason about them, predict the future, estimate danger.
- Small details (feet, hands) can be very important.



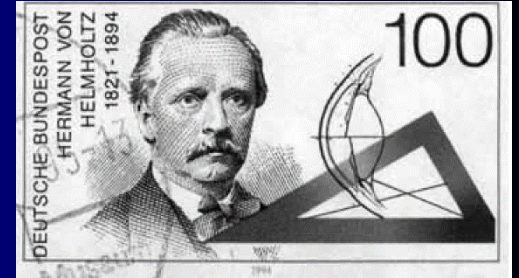
Vision is the full AI Problem

- Understanding of objects, scenes, and events.
- Reasoning about functions and roles of objects, goals and intentions of agents, predicting the outcomes of events.
- The full Artificial Intelligence problem.
- Visual Turing Tests: can computers answer all questions that a human can about images (or do it better)? (D. Geman and S. Geman).



How to Address Vision?

- One strategy – divide vision up into many subtasks that can be studied separately.
- *In this course, we will study subtasks in a unified way that will enable them to be integrated into a complete visual system and which takes complexity into account (research program).*



Vision as Inverse Inference

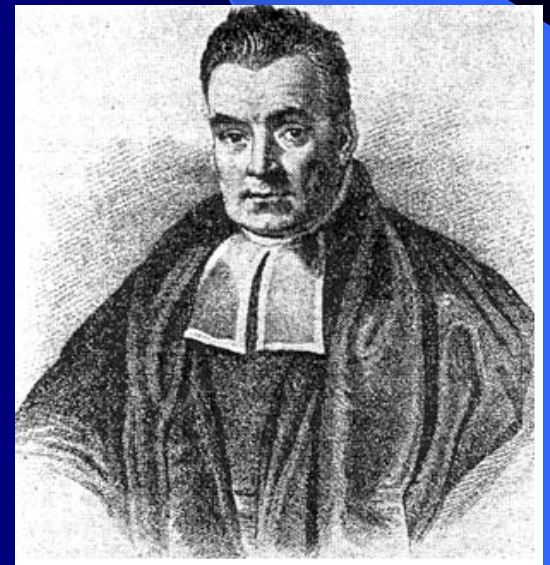
- Helmholtz. 1821-1894.
- More recently: inverse computer graphics.
- Vision must invert the forward process (CG) to discover the causal factors that have generated the images.
- *But only for the tasks that we care about – attention, change blindness.*

Richard Gregory

- "Perception (vision) as hypotheses".
- Perception is not just a passive acceptance of stimuli, but an active process involving memory and other internal processes.
- Humans have internal representations – we see images when we dream, we can imagine what animals and people will do, we can hallucinate.

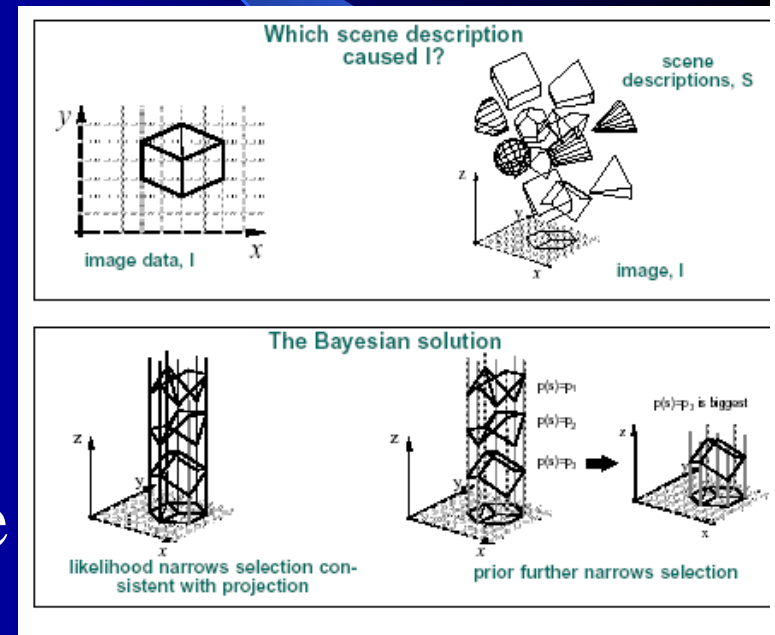
Bayesian Decision Theory

- Bayes' Theorem is a conceptual framework for inverse inference problems.
- We can infer the state S of the
- world from the image I using prior knowledge.
- $P(S/I) = P(I/S)P(S)/P(I)$.
- Rev. T. Bayes. 1702-1761
- $P(I/S)$ likelihood, $P(S)$ prior



Inverse Problems are hard

- There are an infinite number of ways that images can be formed.
- Why do we see a cube?
- The likelihood $P(I|S)$ rules out some interpretations S
- Prior $P(S)$ —cubes are more likely than other shapes consistent with the image.



Inverse inference requires priors

- Humans use prior knowledge about the world (obtained through experience). Often correct – but can fail occasionally.
- Flying carpet? Levitate?
- *Play ball-in-box.*



Taxonomy of Vision

- Hierarchy – Low-, Middle-, High-level vision.
- This gives a way to roughly taxonomize visual problems.
- Relates to theories of the human visual system. E.g., Poggio, Marr, Ullman.

Marr's Taxonomy

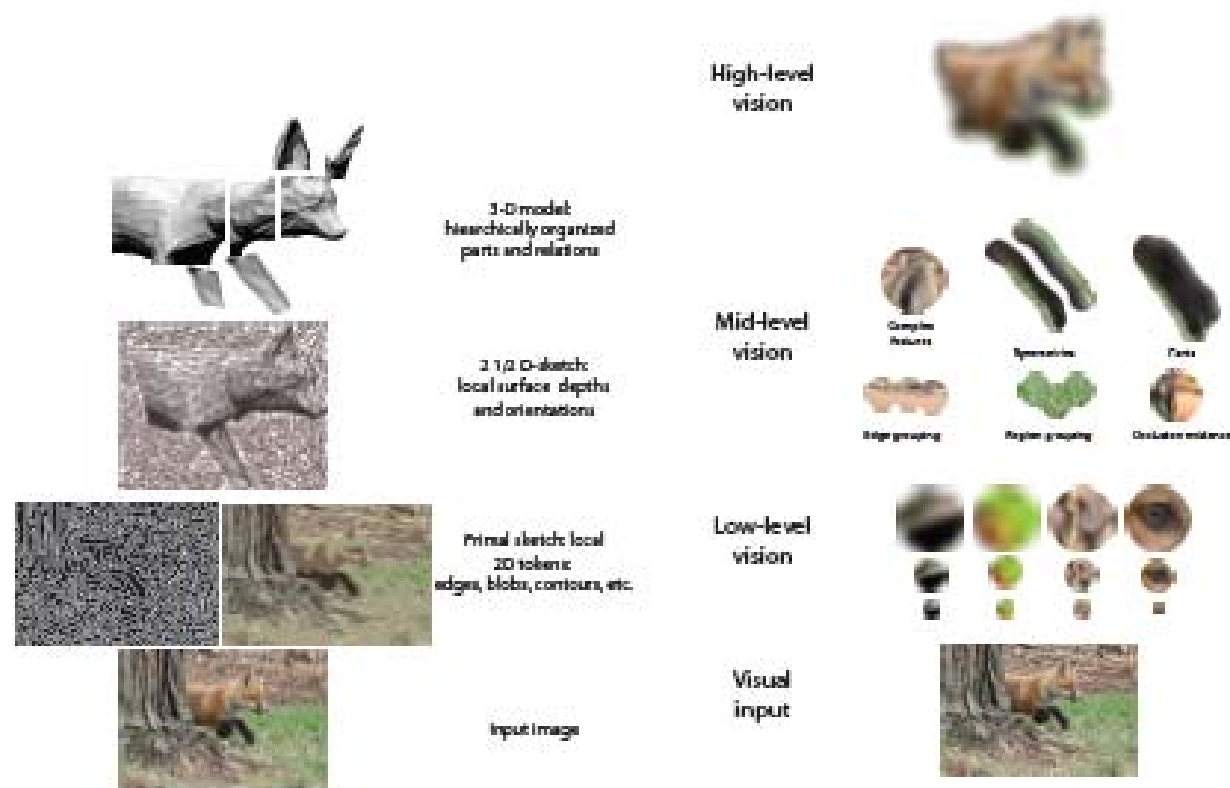


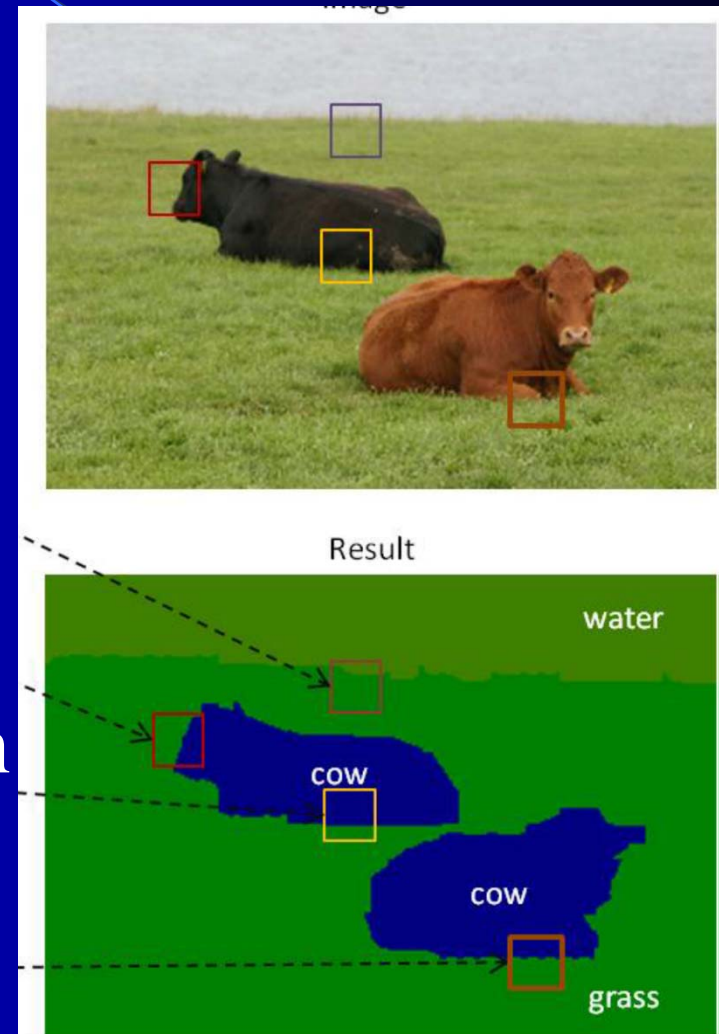
Fig. 4. Marr's framework for vision (left panel) consists a series of representations. Visual tasks can be classified into low-, mid-, and high-level tasks (right panel). This classification roughly relates to Marr's framework.

Low-, Mid-, and High-Level Vision

- Vision can be broken down into low-, mid-, and high-level vision (very roughly).
- Low-level vision – *local image* operations which have limited knowledge of the world.
- Mid-level vision – *semi-local* operations which know about surfaces and geometry.
- High-level vision – *non-local* operations knowing about objects and scene structures.
- *From generic to specify. From local to global.*

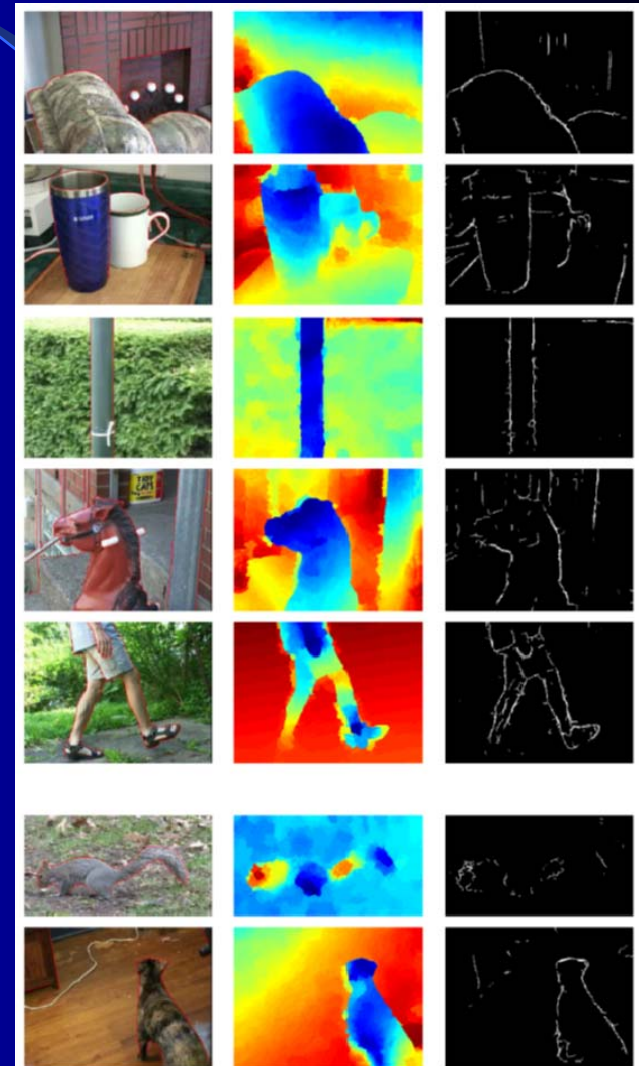
Low-level vision

- Image processing.
- Filtering, denoising, enhancement.
- Edge detection.
- Image segmentation.
- Right: ideal segmentation (followed by labeling).



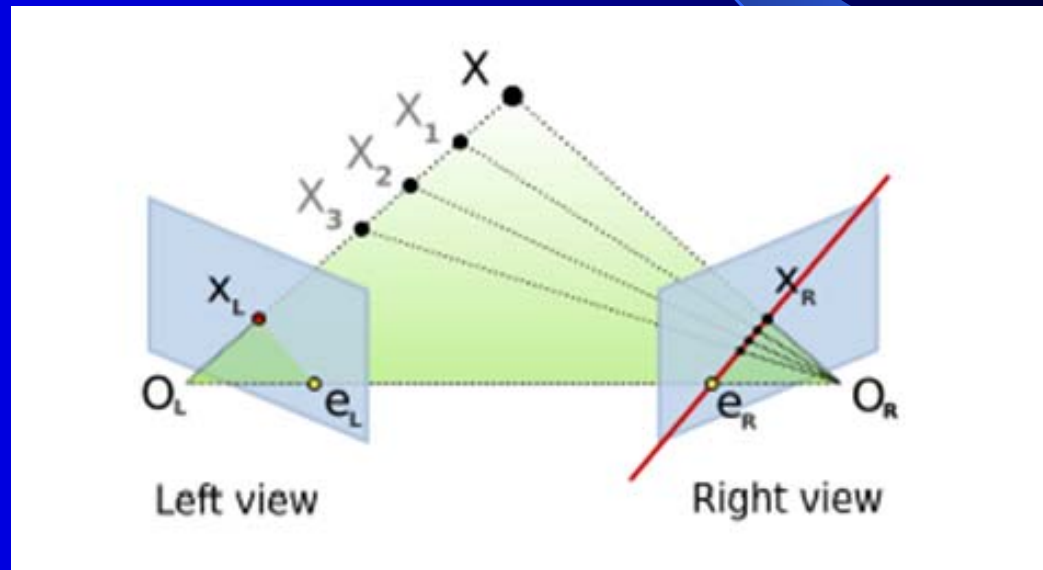
Mid-Level Vision: Depth

- Estimation of 3D surfaces:
- E.g., binocular stereo,
- structure from motion,
- Figure: Images, Depth,
- Segmentation.
- (Blue close, red far).



Stereo: Correspondence and Trigonometry

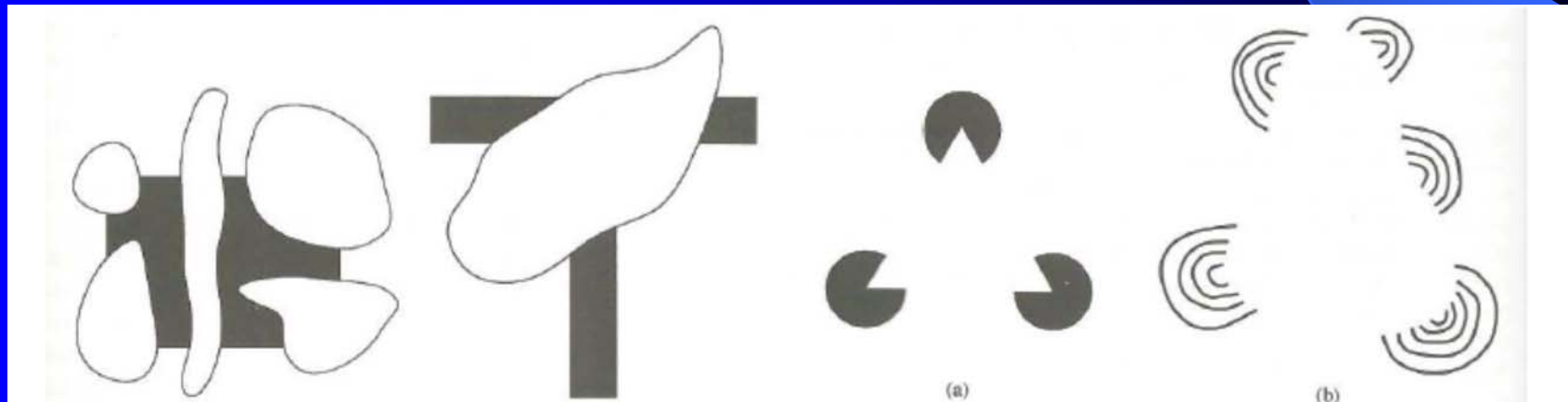
- The correspondence problem and depth estimation:



- There are no good one-eyed sportsplayers.

Mid-Level Vision: Grouping

- Kanisza. Humans have a strong tendency to group image structures as surfaces which can partially occlude each other.



More Kanisza

- From Michael Bach:
- <http://www.michaelbach.de/ot/cog-kanizsa/index.html>
- Kansiza with other cues.

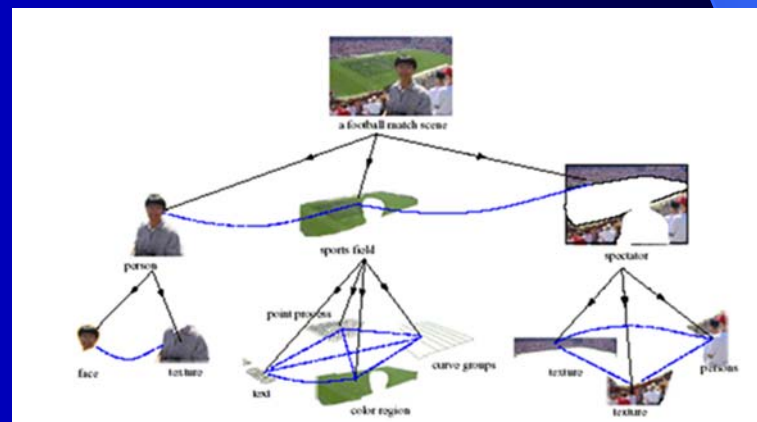
High-Level Vision

- Object detection and Scene Understanding.
- Example: detect objects and object parts.



Feedforward and Feedback

- How do these levels interact?
- Feedforward – from low-, to mid-, to high.
- Feedback – high to low -- analysis by synthesis.
- Low-level vision makes proposals which are validated by high-level vision.

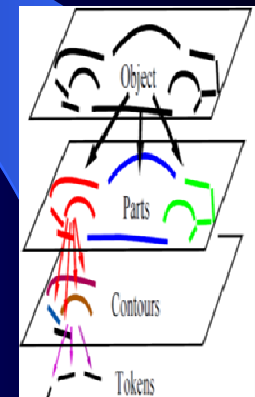
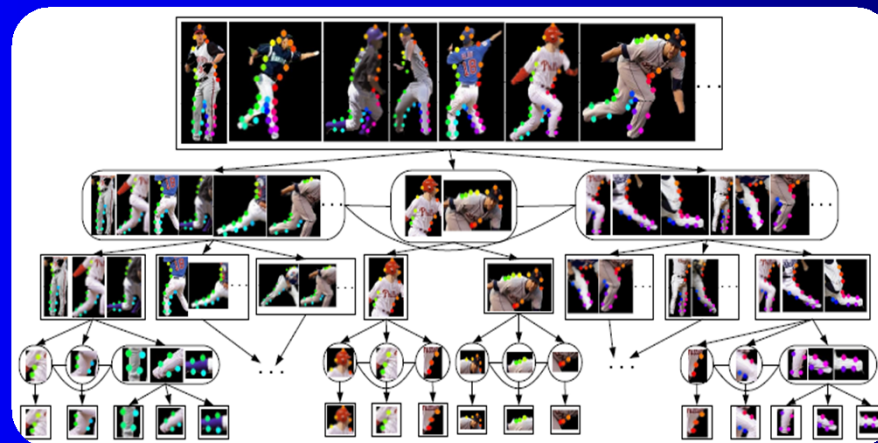


Does high-level always beat mid-level?

- Tanz Illusion:
- <https://www.youtube.com/watch?v=44mw37d8LQw>
- Inverted Face Mask:
- http://www.michaelbach.de/ot/fcs_hollow-face/index.html
- We see it as a real face, but it isn't.

Hierarchical Compositional Models

- Compositional models represent objects and scenes in terms of compositions of elementary shared parts.
- This offers a possible solution to the complexity problem of vision.



Summary

- The key challenges of vision are complexity of images, local ambiguity of images, and complexity of visual tasks.
- Vision is organized into a hierarchical structure. This leads to a rough taxonomy into low-, mid-, and high-level vision.
- Inverse inference, pattern theory, representation, inference and learning.
- Visual Turing tests – drive computer vision forward.

More Illusions

- A few more:
- <https://www.youtube.com/watch?v=MYJkM4wfyZI>