Bayes in the Big: 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 approximately and for the tasks that we care about – attention, change blindness.*

Richard Gregory

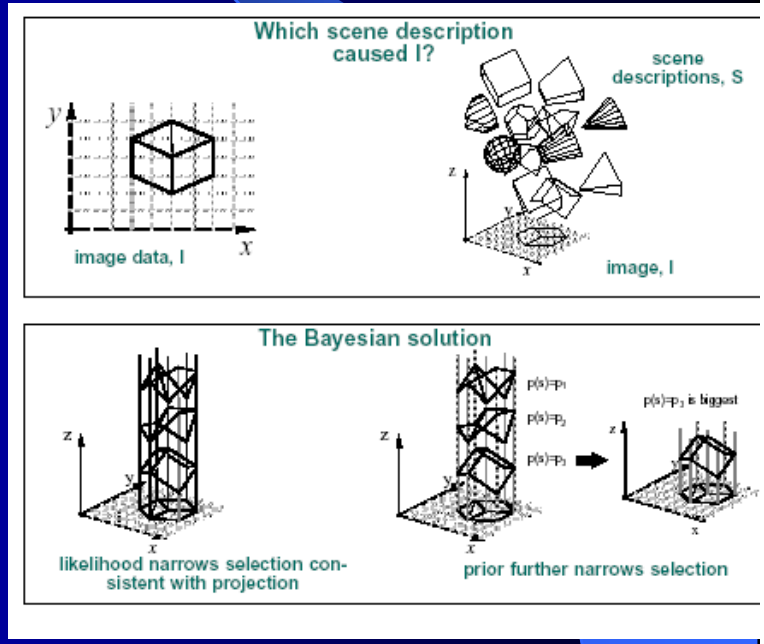
- "Perception (vision) as hypotheses“.
- Perception is not just a passive acceptance of stimuli, but an active process involving memory and other internal processes.
- Humans have internal representations – we see images when we dream, we can imagine what animals and people will do, we can hallucinate.

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 interpetations 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.***

Inverse Computer Graphics versus Deep Networks

Computer Graphic gives method to generate real images.

This yields P(I|S). This will typically be a deterministic function, if S is specified exactly, specified by radiosity equations (see later)

The prior P(S) will be a distribution over all possible configurations S of the 3D world.

The world is complex, so the posterior distribution P(S|I) is extremely complicated.

Inverse Computer Graphics

- The complexity of P(I|S) and P(S) means that there are very few generative models of realistic images.
- In recent years, this is changing due to the success of computer graphics models.
- But work that does inverse computer graphics, though important, is almost non-existant.

Deep Networks as Alternative

- Deep Networks seek to learn P(S|I) directly.
- They assumes that P(S|I) can be expressed as parameterized functional form:
- P(S|I) = f(S,I;w), where f(.) is a deep network which is a differentiable function of the weights w.
- It is, conceptually, easy to learn the weights w from training data (because f(.) is differentiable).
- After learning, it is easy to estimate the best S by calculating f(S,I;w) and taking the maximum

Deep Networks as Alternatives

- Deep Networks are attractive, since they only require learning probability distributions for S – which is enormously simpler than learning a distribution on I.
- But for inverse computer graphics, the space of possible 3D worlds S is combinatorially large, so impractical to learn a deep network to estimate P(S|I).

Deep Networks

- Even for simpler, more restricted, domains it is questionable whether the generative process $P(S|I) = P(I|S)P(S)/P(I)$ can be approximated by a function like $f(S,I:w)$.
- This means that Deep Networks can only address a limited number of constrained visual tasks, without major modifications.

Inverse Computer Graphics

- Inverse computer graphics is very hard.
- But arguably vision will not be solved until we can achieve it.
- And plenty of evidence suggests that humans can perform approximate and as-needed inverse computer graphics.