Lecture 1: Introduction to Computer Vision

Note Title 7/3/2011

Goals: To introduce vision, theoretical concepts, and techniques.

A rapid tour of Computer Vision.

Vision is the task of interpreting, or understanding, the world from images.

Images are rays of light that reach a camera or our eyes.

Vision is very difficult. People think it is easy because we open our eyes and immediately understand the scene – the part of the world we are looking at.

But this case is misleading. Humans find it easy because our brain is specialized to do vision. About half of the cortex does vision. The cortex is the most advanced part of the brain.

In 1970's scientists working in Artificial Intelligence thought Vision was easy. A student at MIT was told to solve Vision as a summer project. Scientists thought that playing Chess was much harder than Vision.

But, by 1995, there were chess programs that could beat the world champion Kasparov. But Vision researchers could not find faces in images.

Researchers only realized how hard vision really is when they started to design computer vision algorithms to interpret images.

(Page 2)

Why is Vision difficult?

Images are complex and ambiguous.

Complexity (1): An image is a set of numbers defined on an image lattice $\{I_{ij} : i = 1 \text{ to } 1,024, j = 1 \text{ to } 1,024\}$, $0 \le I_{ij} \le 255$.

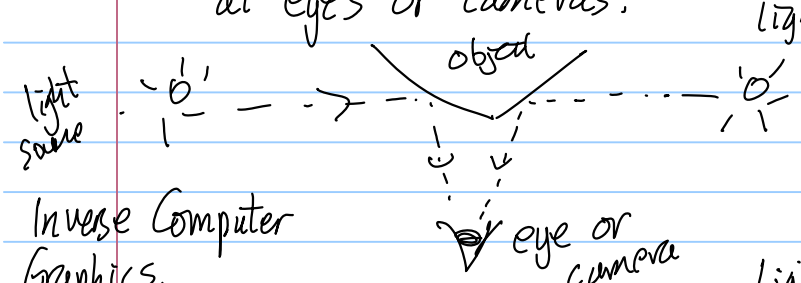
The total number of $1,024 \times 1,024$ images is astronomically big $256^{(1,024 \times 1,024)}$

The number of 10×10 images $256^{10 \times 10}$

>> Total number of images seen by all humans lived.
$50,000,000,000$ humans
$\times$ $60 \times 365 \times 24 \times 60$ total number of minutes
$\times$ 30 number of images seen each minute.

Complexity (2): Images are formed in extremely complex ways.

Light rays are emitted from many light sources, are reflected off objects, and images at eyes or cameras.



Inverse Computer Graphics.

Light sources can be localized - e.g., Sun or diffuse - e.g. Sky.
Light can be reflected from many objects

Ambiguity: Two images of the same object can be very different - due to changes of lighting, small changes of viewpoint, small changes in the position and orientation of the object.

Image Sequences and Active Vision:

- In many vision applications we have a sequence of images. Also the vision system can be active. The vision system can move the camera to take new images.

Vision as Bayesian Inference:

This provides a conceptual framework for vision.

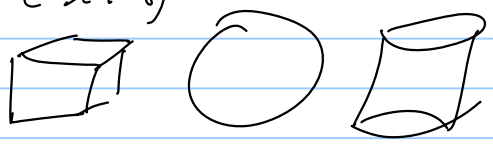
W - represents the state of the world.

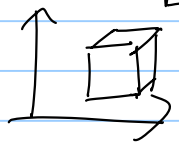
$P(I|W)$ - probabilistic model of image formation (e.g. Computer Graphics)

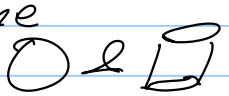
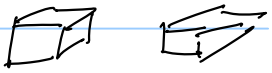
$P(W)$ - prior probability of states of the word.

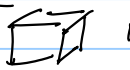
Bayes Thm $P(W|I) = \frac{P(I|W)P(W)}{P(I)}$

Simple Example (Sinha)

$W \rightarrow$  many types of shapes

I 

$P(I|W)$ eliminates some world states - e.g.,  but allows several solutions: 

$P(W)$ resolves the ambiguity by saying that a cube  is more common in the real world.

These probability distributions are defined over graph structures (see later in course).

Graph Structures → Knowledge Representation (Computer Vision)

Probabilities → Uncertainty / Lack of knowledge / Don't care about details.

Note: Probability on Structured Distributions is proposed as a common theoretical framework to model all aspects of Intelligence — vision, language, reasoning, motor control. (IPAM, Summer School 2011, Tenenbaum & Yuille)

Challenges → how to represent worlds and images?
how to put probability distributions for them?

This is much more complicated that language → because the number of words in a language, like English, is only 20,000.

Alternative → Model $P(W|I)$ directly → not $P(I|W)$ and $P(W)$

This is often easier for specific tasks.

E.G. To see if there is a face in an image region → $W \in \{\text{Face}, \text{Non-Face}\}$ has two states

But the intensity I in the region has an exponential number of states → e.g. 256^{100}, if region is 10×10,

It is easier to learn $P(W|I)$, which is a distribution on a variable W with two states, than a distribution $P(I|W)$ on 256^{100} states.

Relationship of Bayes/Probability approaches to Machine Learning (ML). Claim that Machine learning is an approximation to Bayes approaches. The approximations makes learning and inference possible.

There are a large of techniques — mathematical and computational — which are needed for vision. These come from Computer Science, Engineering, Mathematics, and Statistics. Very interdisciplinary field.

The 'full task' of computer vision is to understand an image, or image sequence. This means identifying all objects that are present, specifying their positions and other properties, and understanding the structure of the scene.

But probably most animals cannot do this. Instead they can use vision to detect prey, predators, and do navigation.

So they may ignore large parts of the image.

Many computer vision applications will also only perform a limited set of tasks.