The Confluence of Computational Vision and Computational Linguistics CS828U Spring '12

Angjoo Kanazawa

February 6, 2012

1 January 30th Class 1

Perception vs Language Perception is to understand the world through signals. Understanding is recursive, because you only understand in respect to something else. When does it stop?

Perception is a set of algorithms that are operating on a system. What do they do? The experts can't agree.

David Marr said the goal of perception is to assign labels to objects in the data. Lead to the field of Computer Vision. This hasn't gone anywhere because there's no interaction with the world.

Difference bewteen perception and understanding: human can perceive and relate the situation with a set of symbols in his head (the goat, monkey, ed, tom and the mars bar).

We may never know how animals perceive the world, the contexted needed to relate with the world.

Use language to relate things in the world, as well as the nature of those representations (of signals)

The Cognitive Dialogue Query using a lexicon to a vision system. The vision system answers the query i.e. "is there a car?", then the lexicon searches for more hypotheses. Can apply this dialog to any problem, so this is a (general model for intelligence Perception cannot start bottomw up.. log-polar space normalizes the space so that it's scale invariant

2 February 6th 2012 - Lecture 2

Concepts: An apple is a concept, we don't have a memory of a specific apple. It's like an idea.

Three approaches:

1. Bayesian: Build a memory of concepts by looking at many.
2. Attribute: Every object have a vector of properties.
3. Theory: A concept is such a complicated thing, it's not just one function or datastructure, it's more like a scientific theory. A formal system defined on a set of objects and the relationship between such objects.

2.1 Douglas Stay: Visual Filter

in PASCAL, people start with superpixel segmentation. Problems with this is that where do you divide the pieces? What happens with occlusion? How to figure out texture vs object?

A different approach, is by using a filter:

1. Take a feature (multiscale patch compressed with PCA in $\mathbf{R}^k$)
2. Train a NN classifier
3. Run the classifier on the images, return to step 1 but with results from step 2 (feature now in $\mathbf{R}^{2k}$), with a new NN

This iteration is the key because it actively improves/motivates the classification.

Finding objects people are holding in image streams: Using background substraction (motion cue) + human filter = figure out objects.

Seems like we recognize objects instantly without reasoning about why it maybe a “cup”.

Can we combine filteres to make new ones? By using intersets and unions. Do we have a limited number of concepts and combine them to make sentences. The ability to generalize from categories.

A draw back: extremely supervised.

2.2 Intelligence

Action is fundamental. A very important part of intelligence is the ability to understand others, understanding human action.