Reconstructing force-dynamic models from video sequences

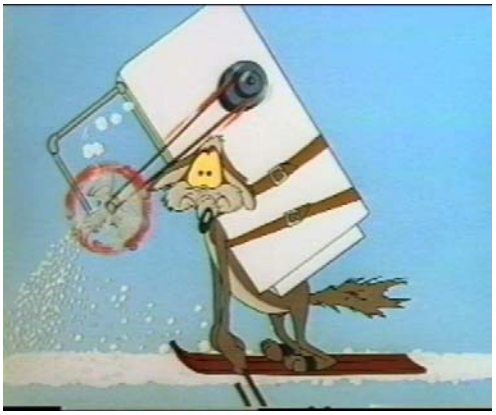
Jeffrey Mark Siskind
School of Electrical and Computer Engineering, Purdue
Presenter: Juan Ramos

Overview

- What is stability? Why pursue this problem?
- LEONARD at an (extended) glance
  - Bases for LEONARD
  - Decision process
  - Experiment and results
- Questions to consider

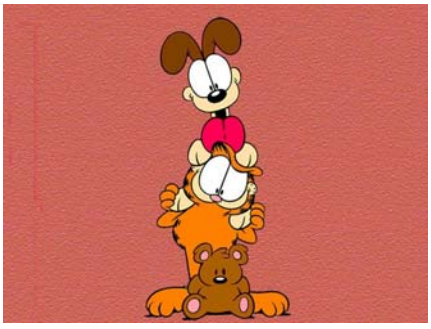
Stability – what is it?

- Which of the following are “stable”?



Stability – what is it?

- What about these?



Stability – why?

- How did you come to those conclusions?
- Why study stability?
  - Understanding the visual process
  - Make judgments on the world
- How would an agent benefit from having stability analysis?
- How do humans perform stability analysis?

Stability

- Everyday concept: when objects are prevented from falling.
- Siskind's definition: a scene is stable if all objects are supported globally.
- Support: attachment joints, grounded property, substantiality constraint.
- What is our notion of grounded? How do we identify a grounded object other than just presuming it?

Bases for LEONARD

- 3D image to 2D layer polygon interpretation (Pylyshyn?)
- Differentiation and Integration
- Predicate Logic
- Linear programming
- Approximations of nature, e.g. friction
- Accounting for noisy data

LEONARD

- Find admissible and stable minimal cost models that interpret data
- Prioritized circumscription: choose models based on assumptions
- Cardinality circumscription: choose models with minimal cost
- Temporal circumscription: choose models that fit better over sequential data

Stability Analysis

- Determine if a given scene is stable under a given interpretation
  - Original assumptions: {Groundness, Rigid, Revolute, Prismatic1, Prismatic2, Same-layer}
  - New predicate logic assumptions: {Groundness, Rigid1, Rigid2, Rigid-theta, Same-layer}
- Kinds of contacts: edge/edge, edge/corner, corner/edge, corner/corner
- A touch of Physics: potential energy, friction
- Reducing noise: tolerance

Admissibility

- A joint must be of only one kind
- Objects cannot “merge”
- Same-layer must be an equivalence relation: reflexive, symmetric, transitive

Prioritized Circumscription

- Set-theoretic four stage process to find admissible and stable interpretations
  - Groundness assertion
  - Rigid joints
  - Revolute joints
  - Same-layer relationships
- Final set forms lattice, search space top-down or bottom-up for preferred models

Temporal/Cardinality Circumscription

- Search through (sequence of) sets of preferred models
- Find most-preferred models that minimize given cost function
- Can eliminate need for cardinality circumscription in some cases

Implementation

- Capture simple actions and pass stability judgments on these
  - Pick Up (hand, block*2)
  - Put Down (hand, block*2)
  - Stack (hand, block*3)
  - Unstack (hand, block*3)
  - Assemble (hand, block*3)
  - Disassemble (hand, block*3)
  - Move (hand, block*3)

Results and Further Work

- 74.3% accurate classifications on 30 movies of each action
- Seven specific, discernible and reproducible error modes
- Some segmentation and tracking errors; a few unknown errors
- Some redundant but acceptable states

Connections to Other Disciplines

- LEONARD is not a complete system for visual analysis. It requires a separate system to transform a given 3D image into a 2D representation. However, LEONARD does pass judgment and thus has some manifestation and effect on the world.
- How applicable is this to evolutionary psychology/darwinian modules?

Connections to Other Disciplines

- In the implementation of LEONARD, verbs are used to represent what action is taking place. Syntactic theory tells us that verbs require objects to be the performers or recipients of the represented action (theta-roles). E.g. in “push”, there is a pushing object and a pushed object.
- Is it then surprising that the actions LEONARD categorizes contains such performers and recipients? What else can be said about verbs and their “meaning”?

Connections to Other Disciplines

- Ambiguity and noisy data are ever-present obstacles in agent interaction with the real world. One data set movie in LEONARD was not used because it was too segmented and noisy. LEONARD can also return multiple most-preferred models when only one is required.
- What does this tell us about our own ability to discern through noise and eliminate ambiguity? In what ways could noise be eliminated to enhance LEONARD's (or any other agent's) performance?