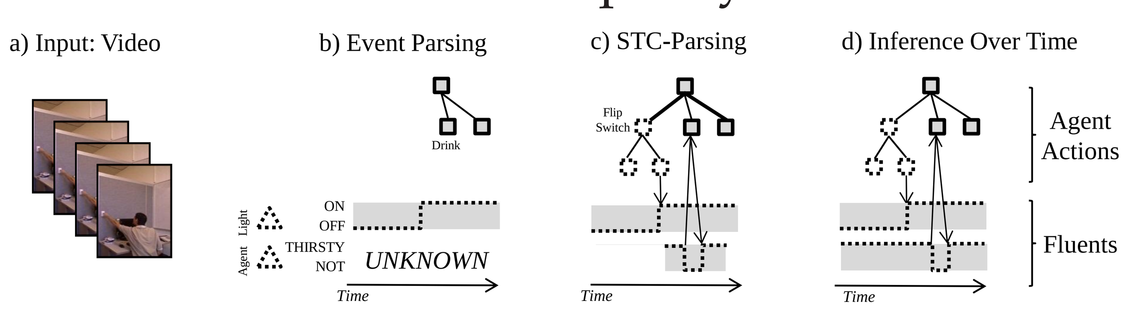
Learning Perceptual Causality from Video

Amy Fire and Song-Chun Zhu
Center for Vision, Cognition, Learning and Art (VCLA), University of California, Los Angeles

Introduction

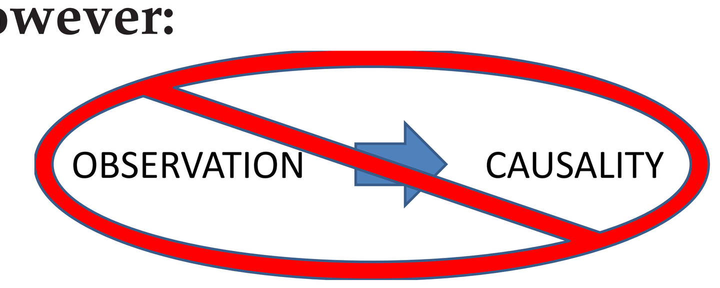
Goal: Learn causality from raw video

Motivation: Increased capacity for inference



1. Answer why events occur
2. Correct misdetections and infer hidden objects/actions
3. Infer triggers, goals, and intents

However:



(generally)

Causality in Vision

Works in vision that incorporate causality can be categorized as

1. Using pre-specified causal relationships for action detection
2. Using causal measures to aid action detection
3. Using a pre-specified grammar for learning causality

Learning causality from video is largely missing from the vision literature.

Causality and Video Data

The focus of causality research is often disjoint from the needs of a vision system:

1. Learning causal networks via constraint satisfaction or Bayesian methods: Intractable on vision sensors
2. First-order logic: Not probabilistic
3. Markov logic networks: Intractable, not learned

Perceptual Causality

Cognitive science research suggests infants use heuristics in judging causal relationships:

1. Agentive actions are causes
2. Measure co-occurrence

cr :

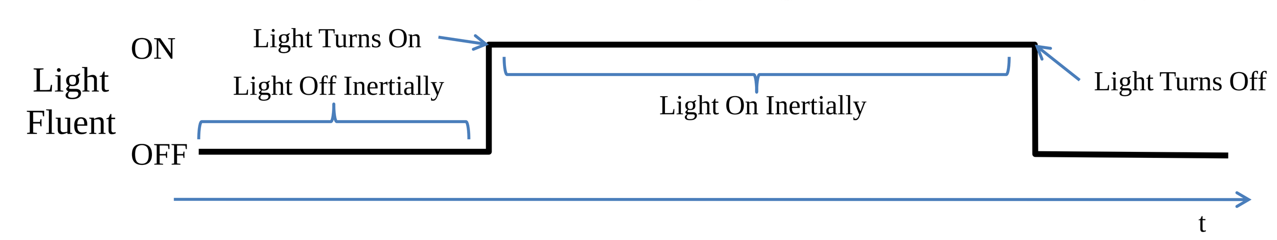
	Action	$\neg$Action
Effect	c_0	c_1
$\neg$Effect	c_2	c_3

3. Temporal lag between the two is short
   $\text{Time(Action)} - \text{Time(Effect)} < \epsilon$
4. Cause precedes effect
   $\text{Time(Action)} - \text{Time(Effect)} > 0$

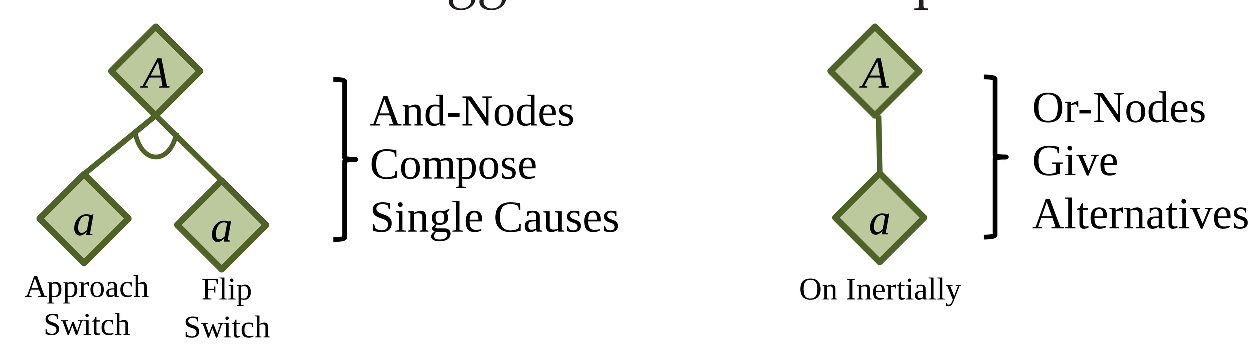
To learn perceptual causality in video, we restrict co-occurrence of detected events and effects to these heuristics.

Causes and Effects: The Causal And-Or Graph

Effects: Fluents are time-varying statuses of objects.

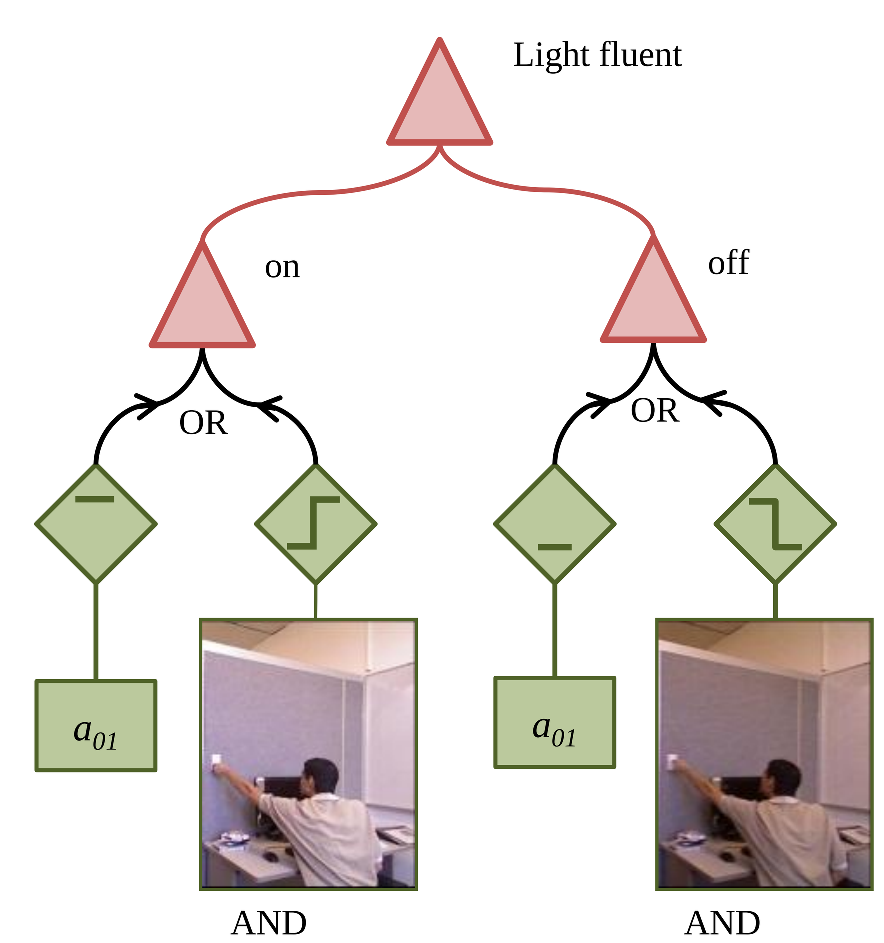


Causes: Actions suggest an And-Or representation.



Pairing cause and effect: Fluent changes are matched with corresponding causing actions. In the absence of change-inducing actions, fluent values are causally attributed to the *inertial action*.

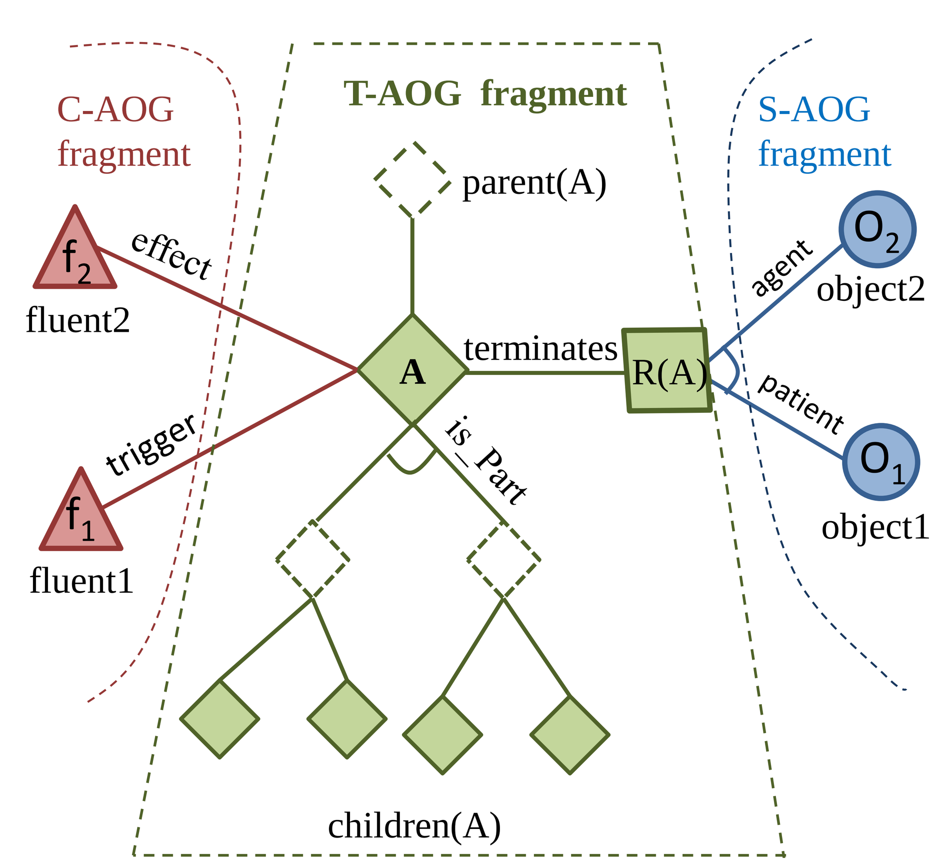
The Causal And-Or Graph



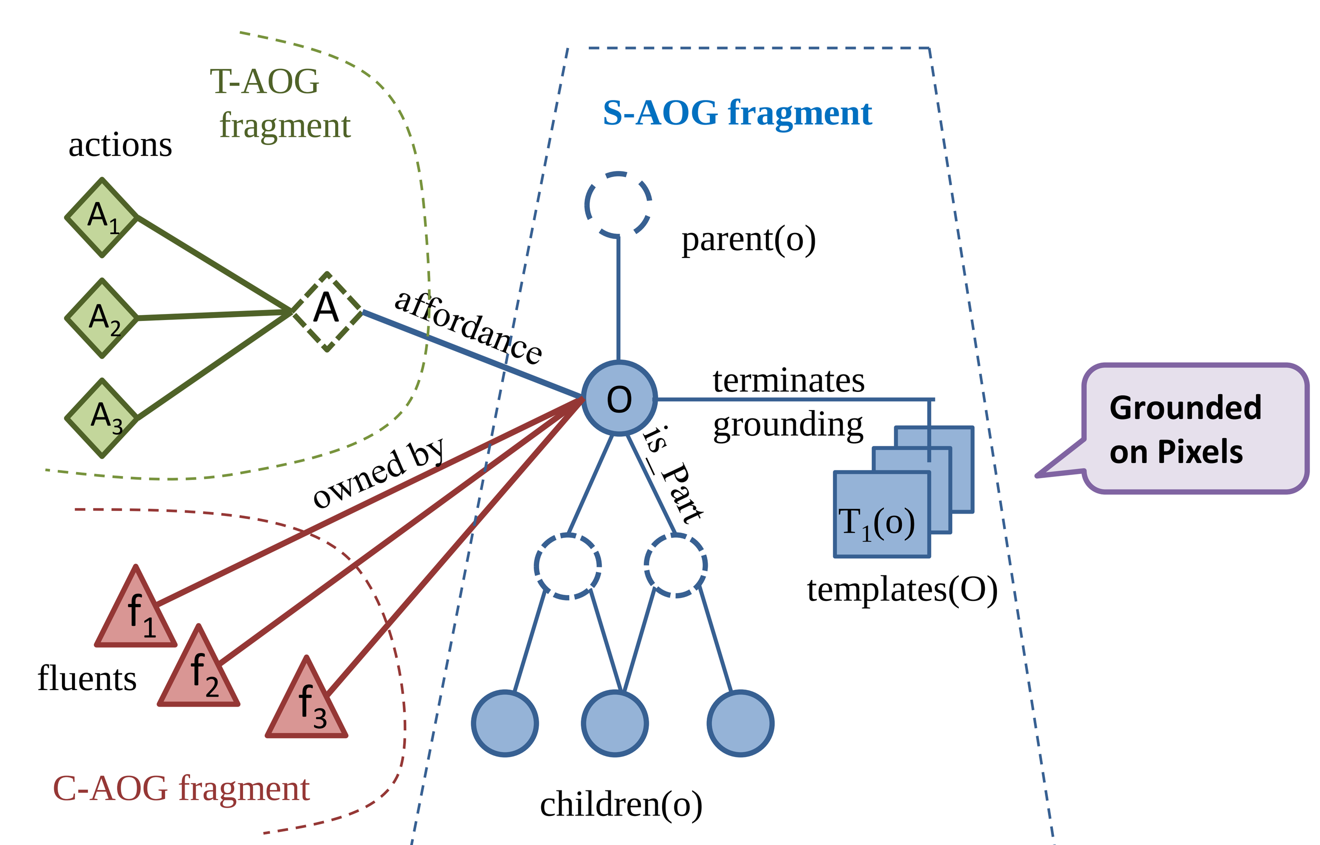
Grounding Causality on Video

Fluent changes in the Causal And-Or Graph are detected using classifiers. Actions are detected as instances from the Temporal And-Or Graph, which are grounded on relationships between objects. Objects are detected on the Spatial And-Or Graph using templates.

The Temporal And-Or Graph

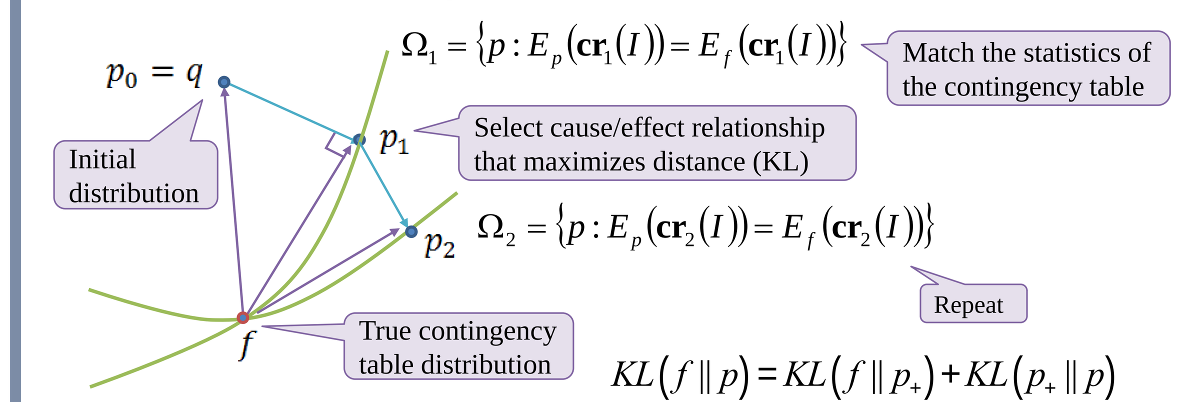


The Spatial And-Or Graph



Principled Learning

Information Projection:



Model Pursuit: Incrementally pursue a model, adding a contingency table at each iteration:

$$p_+(pg) = \frac{1}{z_+} p(pg) \exp\left(-\langle \lambda_+ \mathbf{cr}_+ \rangle\right)$$

Prop. 1: Matching statistics on the model to the observed data, $E_p(\mathbf{cr}_+) = E_f(\mathbf{cr}_+)$, gives

$$\lambda_{+,i} = \log\left(\frac{h_i}{h_0} \cdot \frac{f_0}{f_i}\right)$$

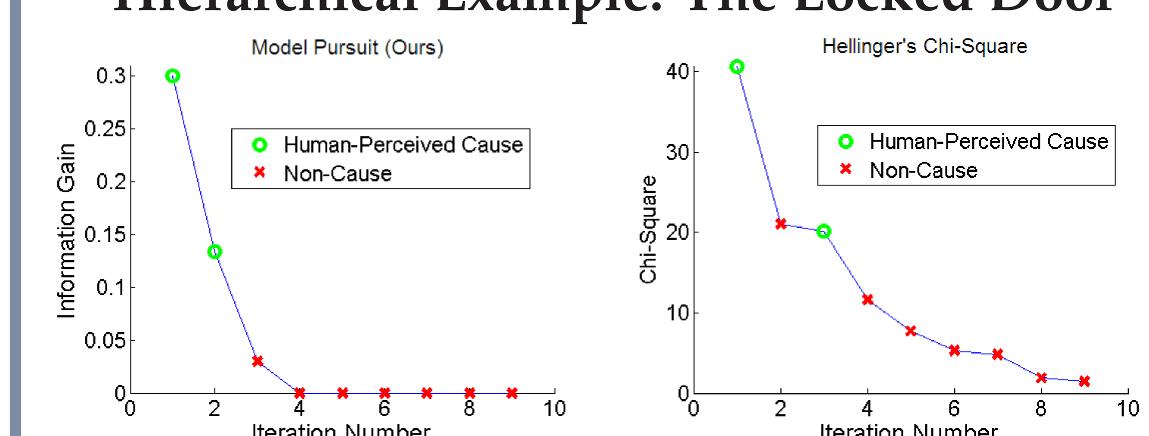
Prop. 2: Adding causal relations

$$\mathbf{cr}_+ = \underset{\mathbf{cr}}{\operatorname{argmax}} \, \mathrm{KL}(p_+ \| p) = \underset{\mathbf{cr}}{\operatorname{argmax}} \, \mathrm{KL}(\mathbf{f} \| \mathbf{h})$$

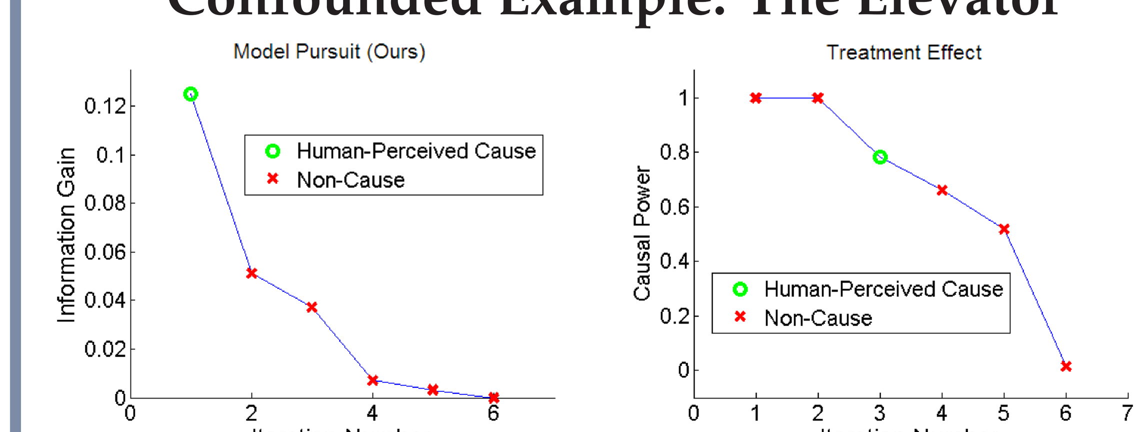
Preliminary Results

Pursuit orders of causal relations.

Hierarchical Example: The Locked Door



Confounded Example: The Elevator



Our method acquires true causes before non-causes, outperforming Hellinger's Chi-Square and Treatment Effect.

Contact Information

Web http://www.stat.ucla.edu/~amyfire
Email amy.fire@ucla.edu