Seeing, Saying, Doing, and Learning

Integrating Computer Vision, Natural Language Processing, Robotics, and Machine Learning Through Multidirectional Inference

Jeffrey Mark Siskind, qobi@purdue.edu

PURDUE UNIVERSITY

Princeton, Thursday 15 October 2015

The effect of language and context on vision

The effect of language and context on vision

The effect of language and context on vision

Outline

1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

Outline

1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

Joint work

Andrei Barbu Daniel Paul Barrett N. Siddharth Haonan Yu

Object detection: Felzenszwalb et al. (2008)

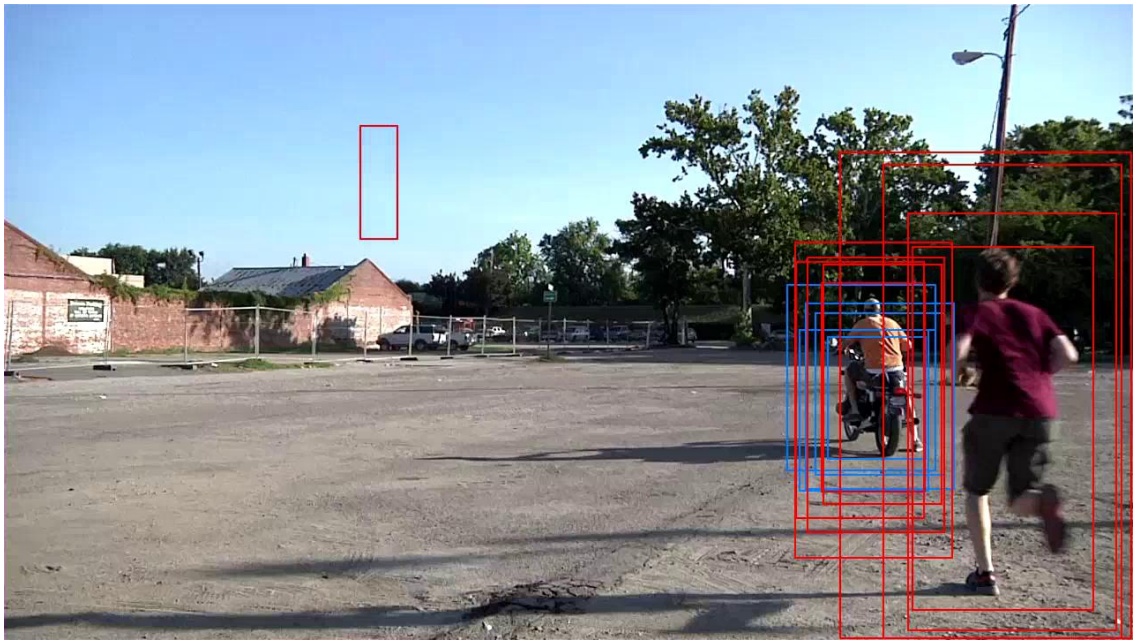
Object detection: Felzenszwalb et al. (2008)

False positives

Object detection: Felzenszwalb et al. (2008)

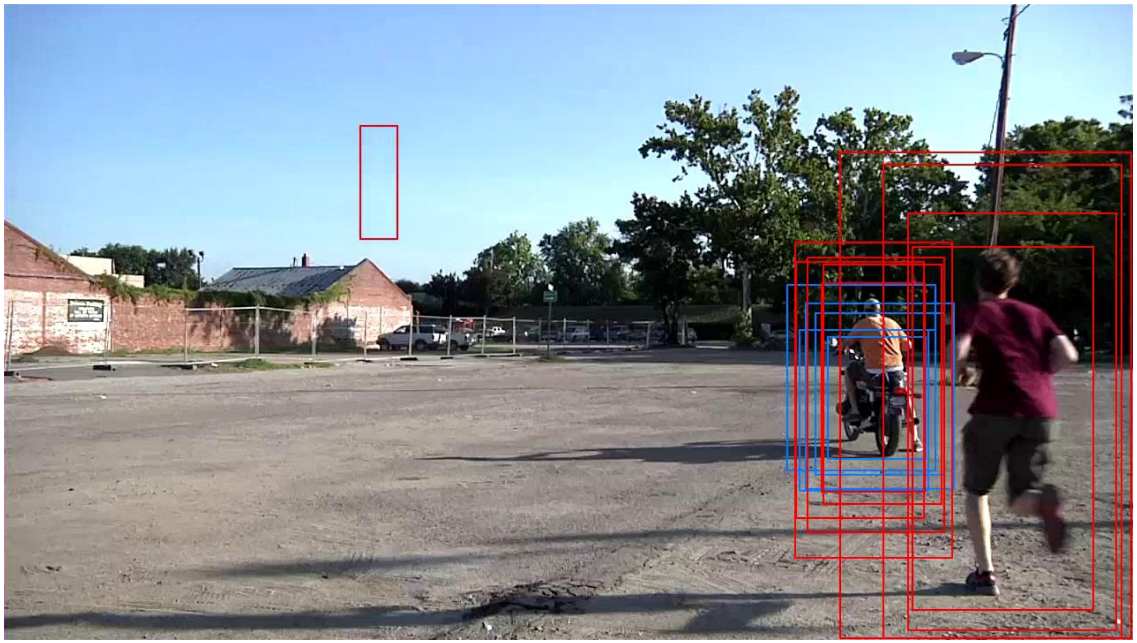
False negatives

From uncertain object detections to robust object tracks



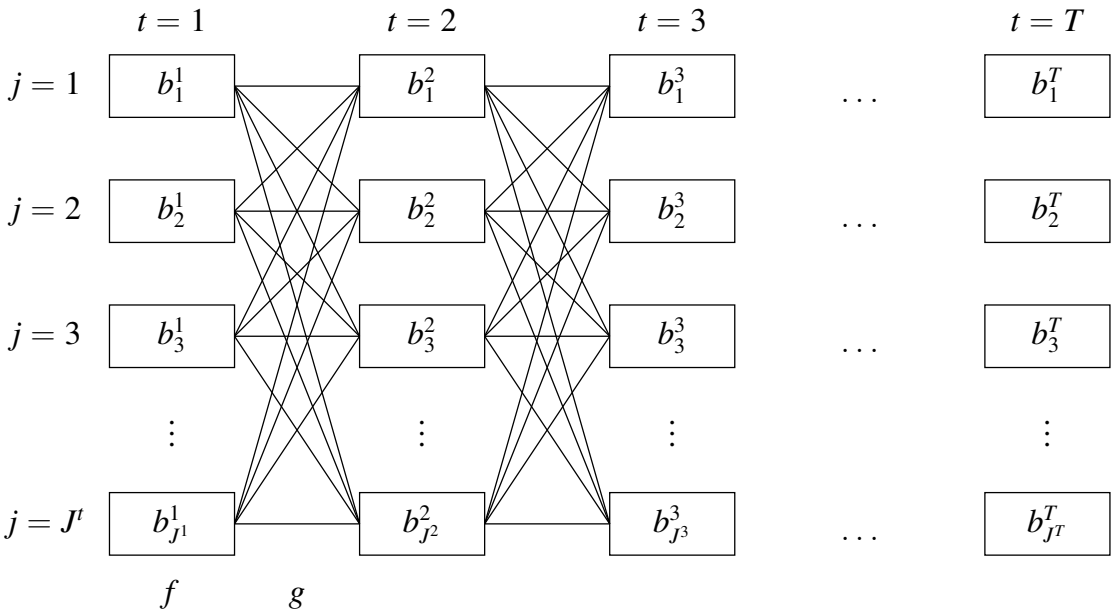
- for each object model in each frame
- overgenerate
  - many false positives
  - compensate for false negatives

From uncertain object detections to robust object tracks



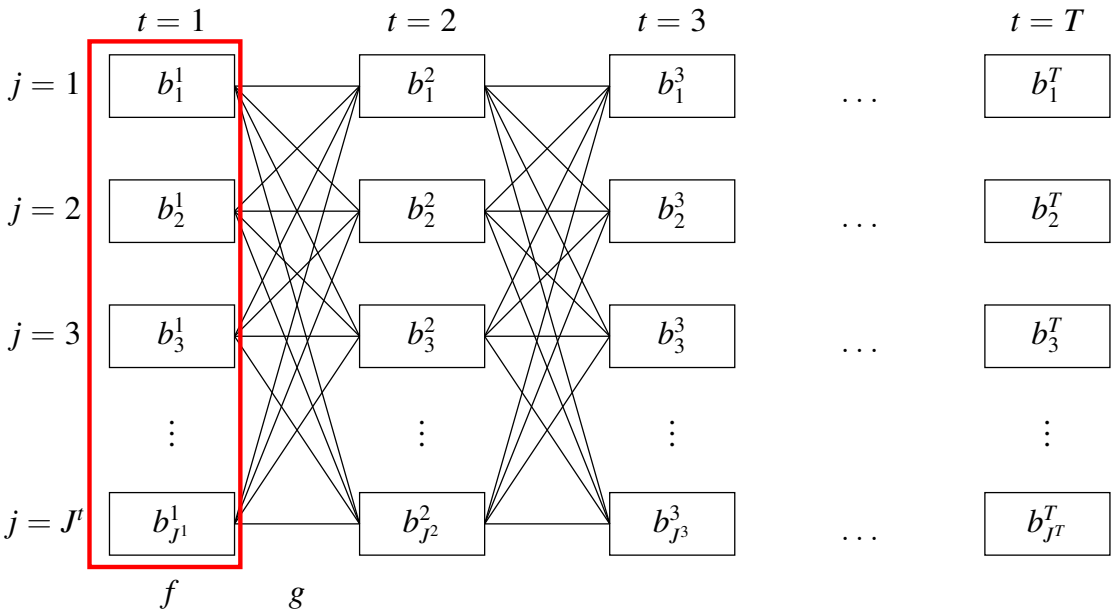
- for each object model in each frame
- overgenerate
  - many false positives
  - compensate for false negatives

From uncertain object detections to robust object tracks



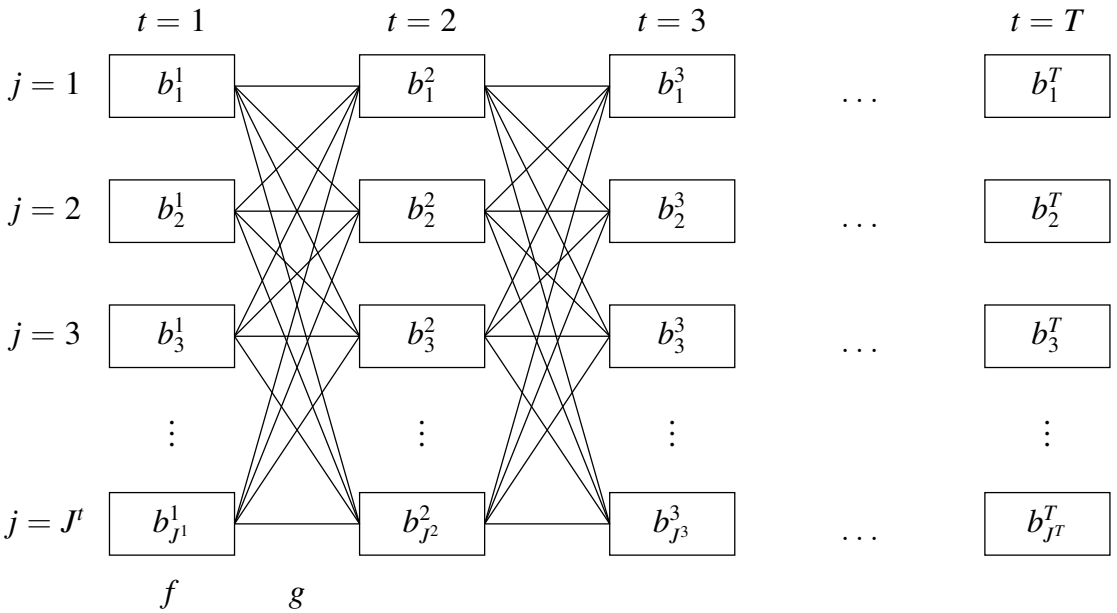
- find single detection for each object in each frame

From uncertain object detections to robust object tracks



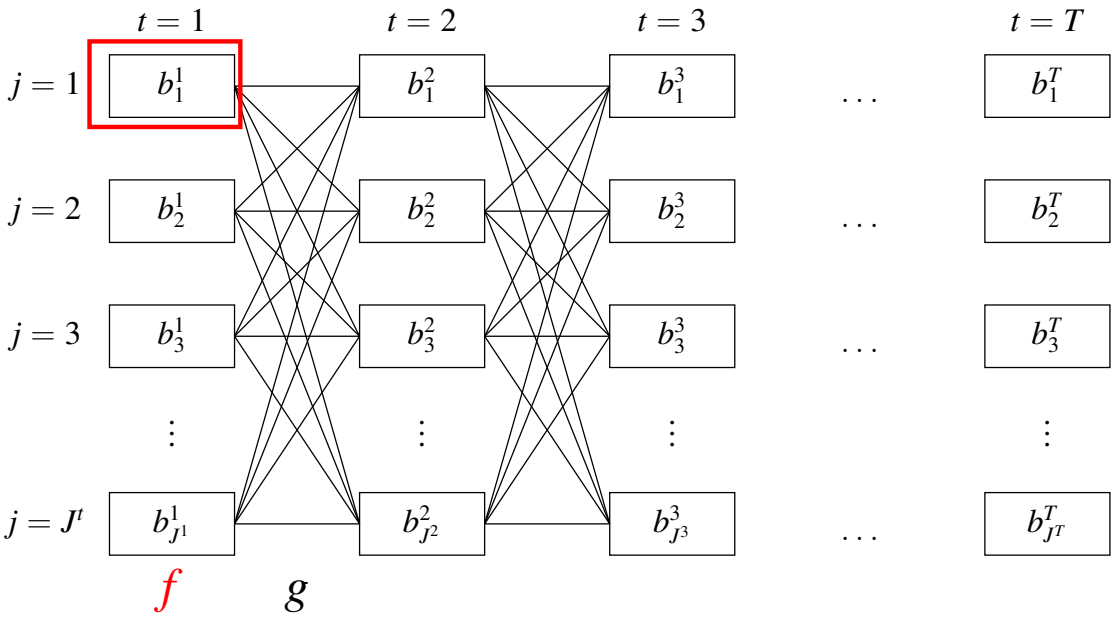
- find single detection for each object in each frame

From uncertain object detections to robust object tracks



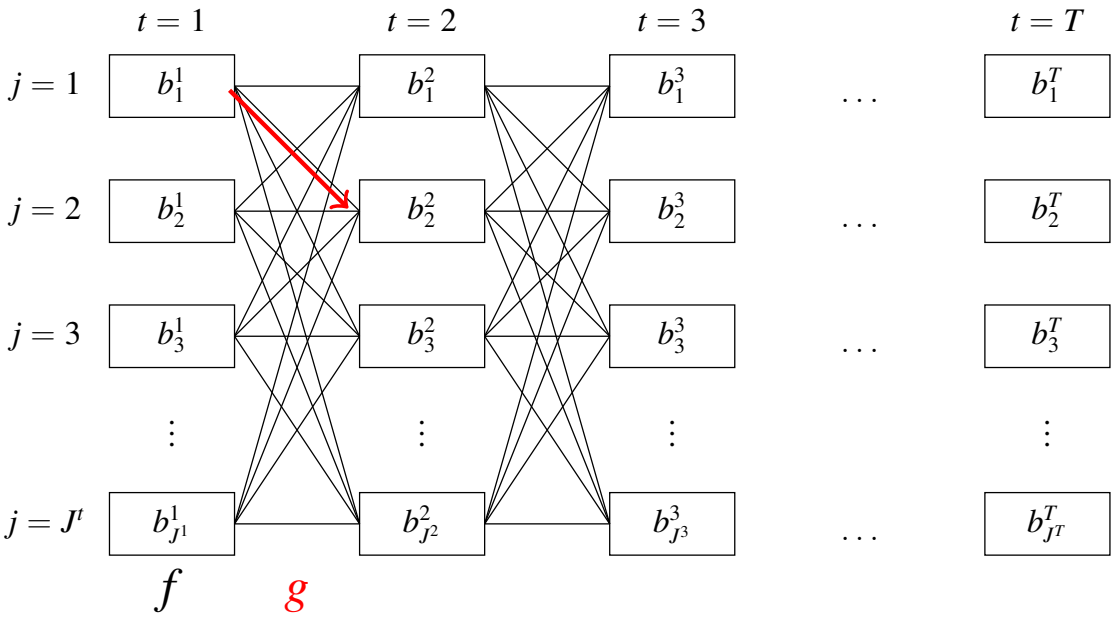
- find single detection for each object in each frame
- temporally coherent track

From uncertain object detections to robust object tracks



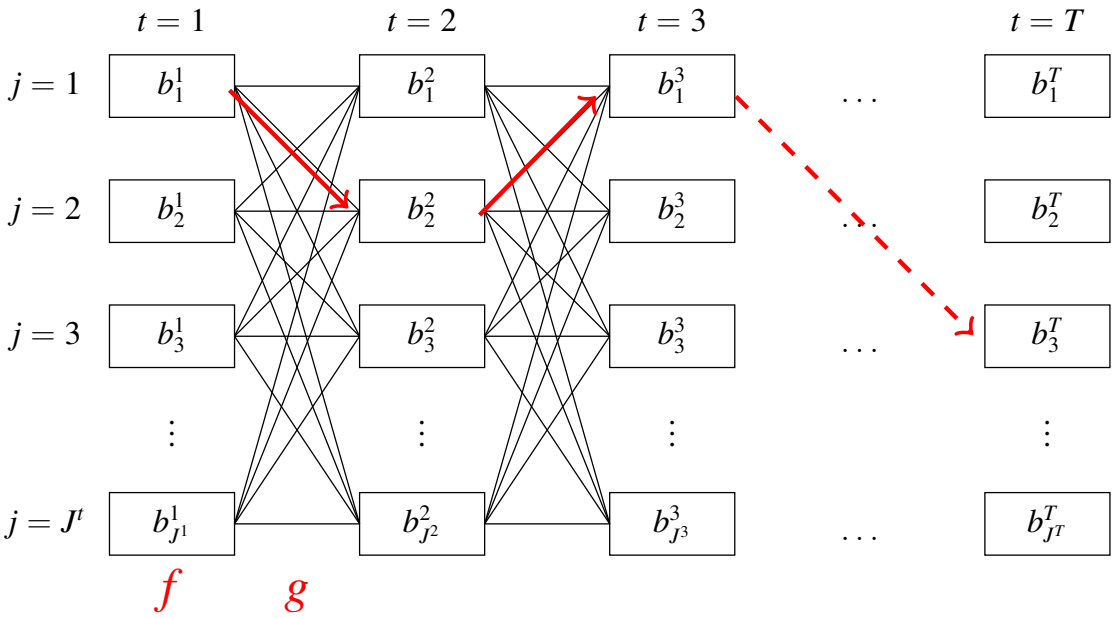
- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)

From uncertain object detections to robust object tracks



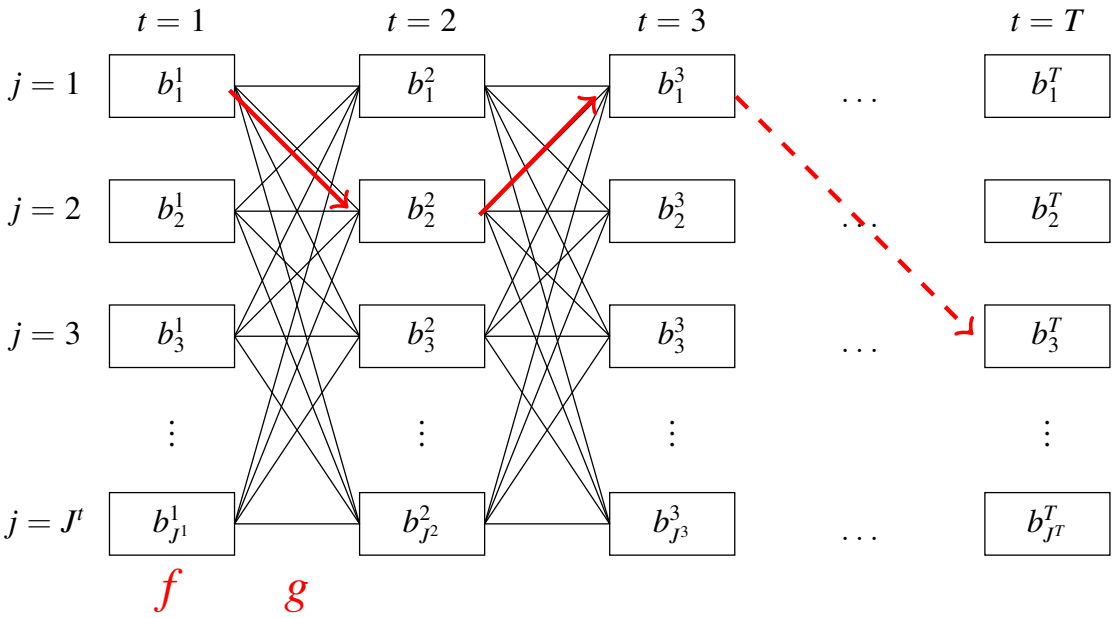
- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)
  - motion coherence (g)

From uncertain object detections to robust object tracks



- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)
  - motion coherence (g)
- optimal path through lattice of detections

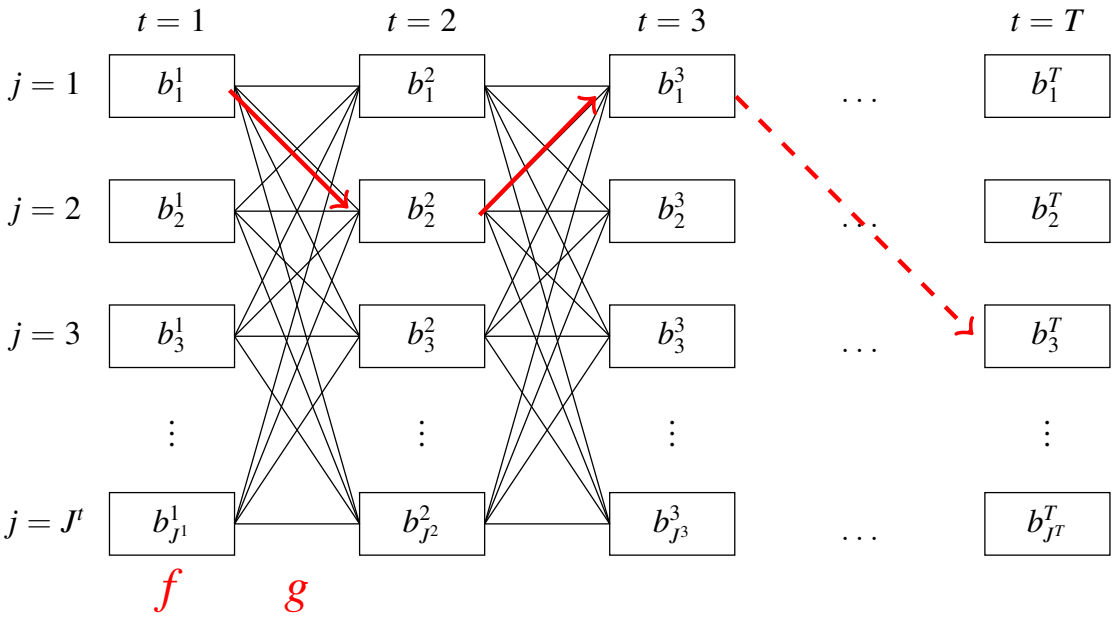
From uncertain object detections to robust object tracks



- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)
  - motion coherence (g)
- optimal path through lattice of detections

$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t)$$

From uncertain object detections to robust object tracks



- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)
  - motion coherence (g)
- optimal path through lattice of detections
- dynamic programming

  Bellman (1957), Viterbi (1967)

$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t)$$

From uncertain object detections to robust object tracks

- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)
  - motion coherence (g)
- optimal path through lattice of detections
- dynamic programming

  Bellman (1957), Viterbi (1967)

$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t)$$

From uncertain object detections to robust object tracks

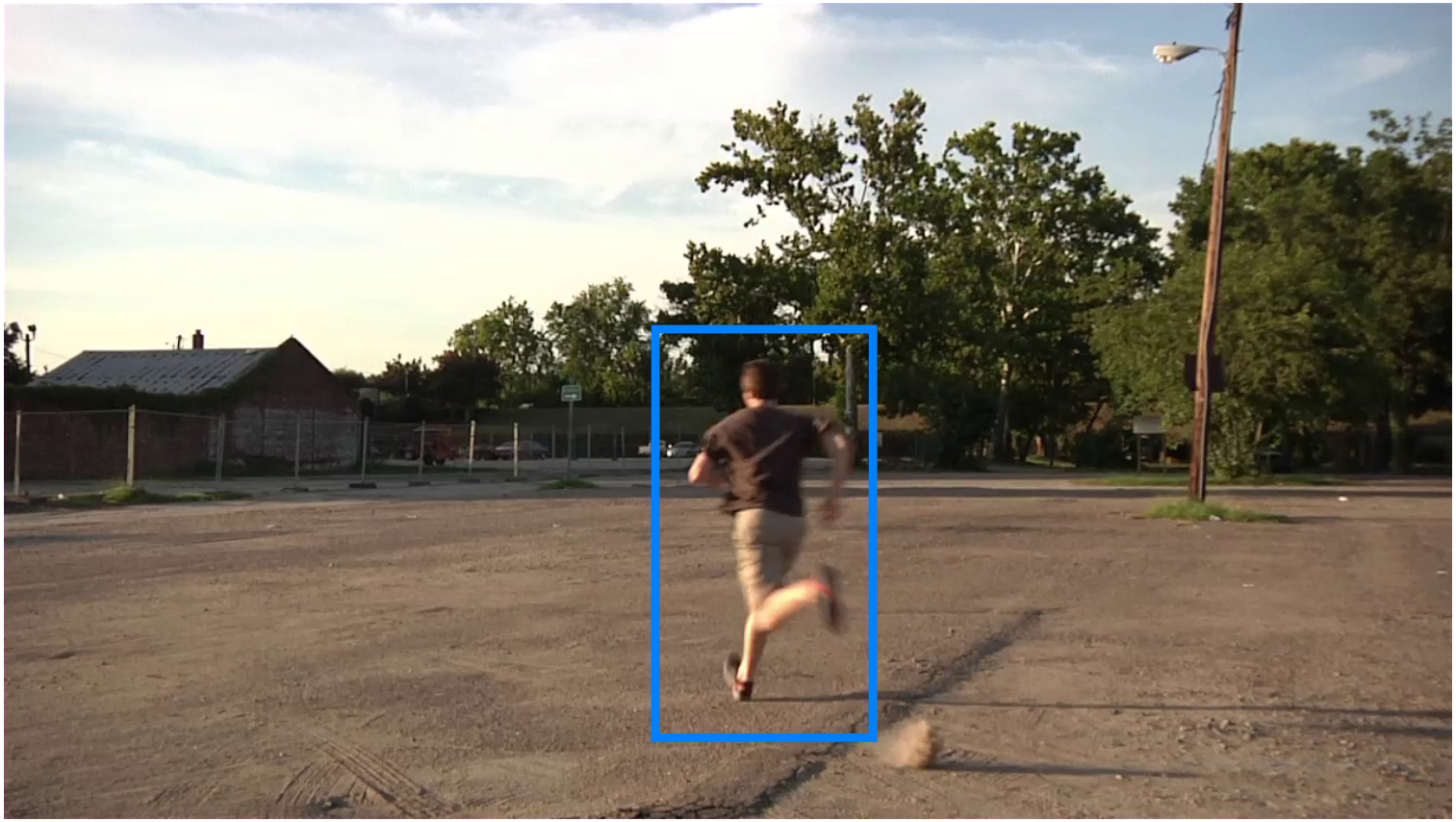
- find single detection for each object in each frame
- temporally coherent track
  - object detector confidence (f)
  - motion coherence (g)
- optimal path through lattice of detections
- dynamic programming

  Bellman (1957), Viterbi (1967)

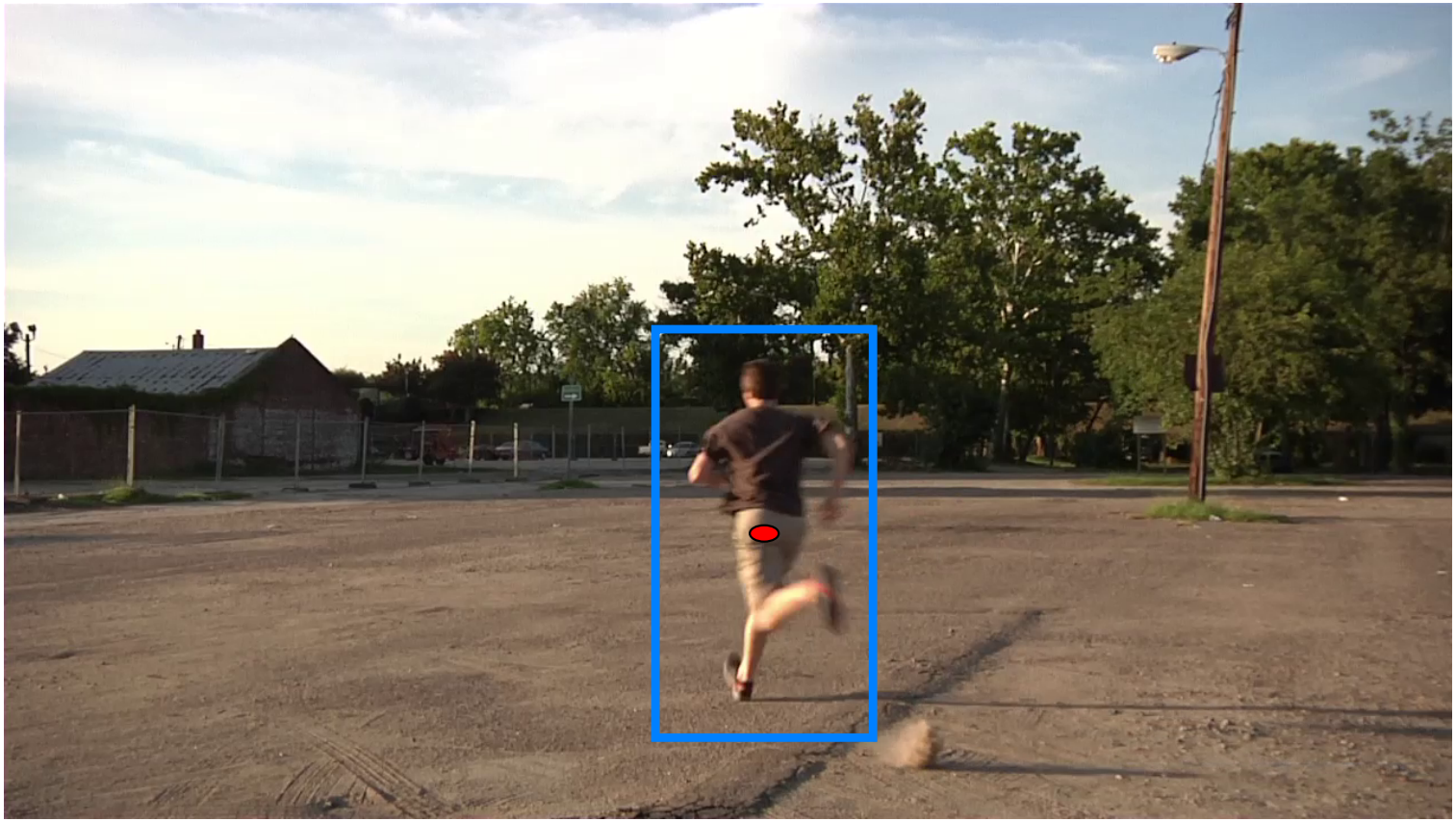
$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t)$$

Feature vector—single participant

Feature vector—single participant

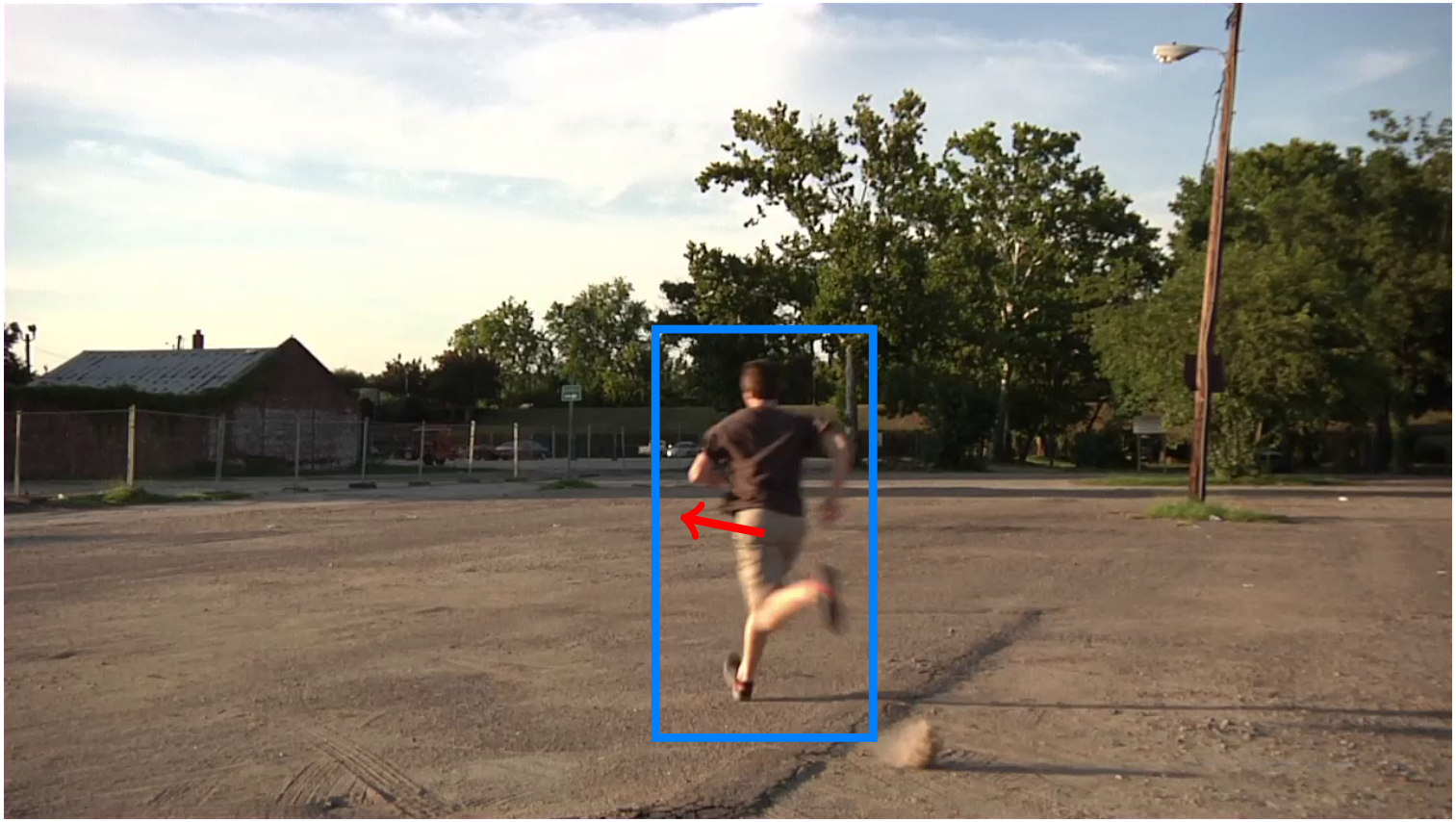


Feature vector—single participant



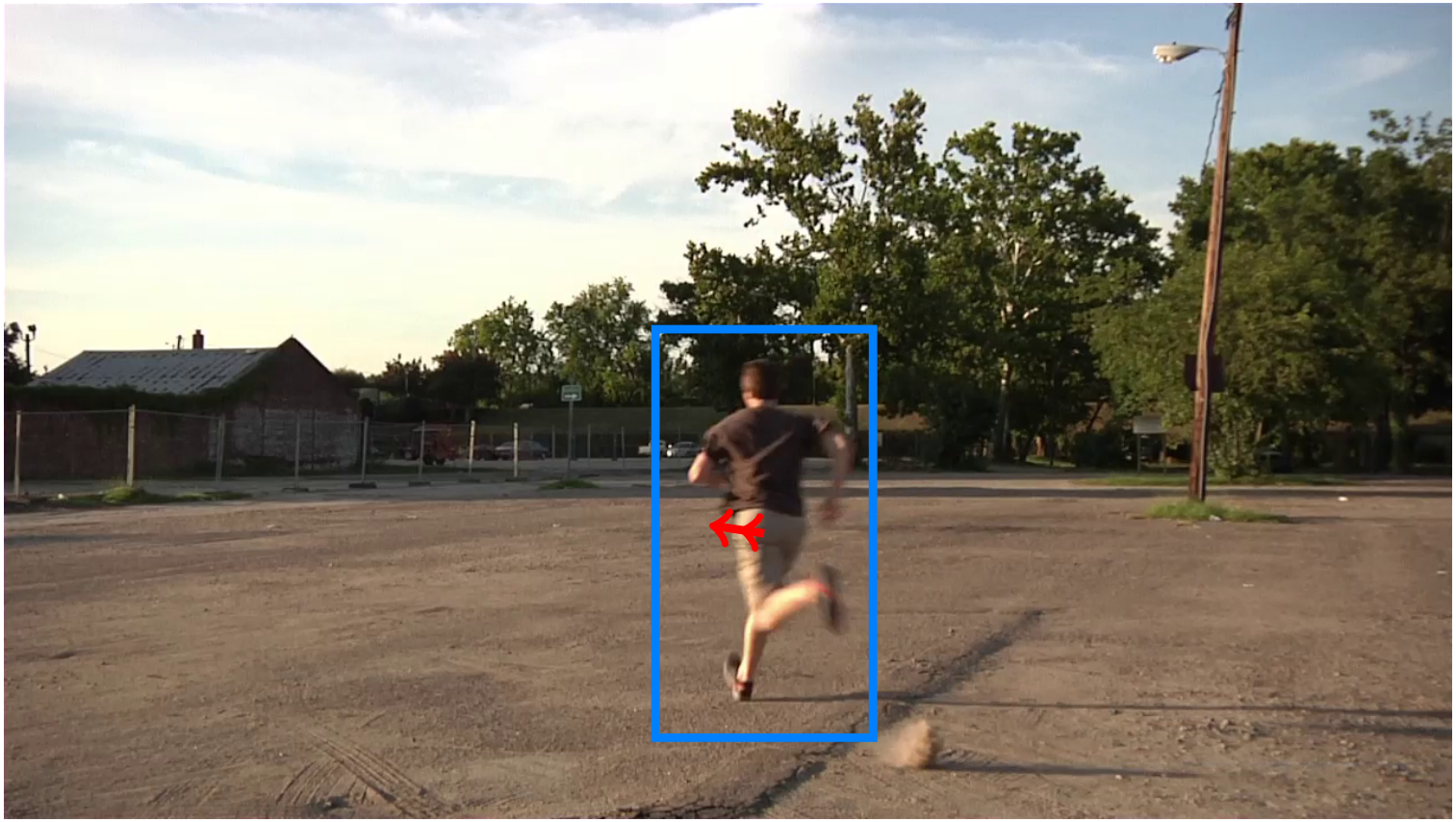
position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—single participant



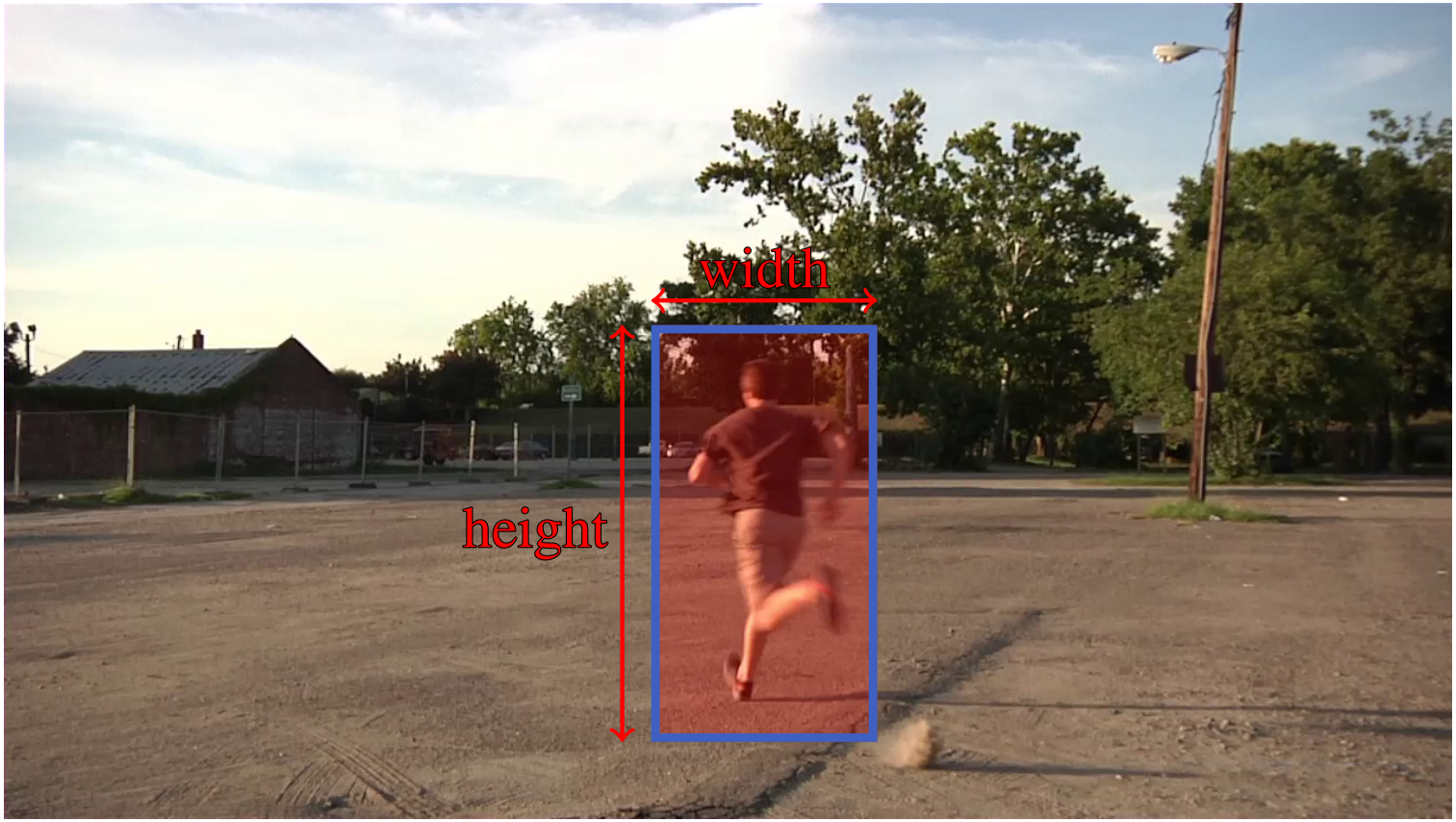
position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—single participant



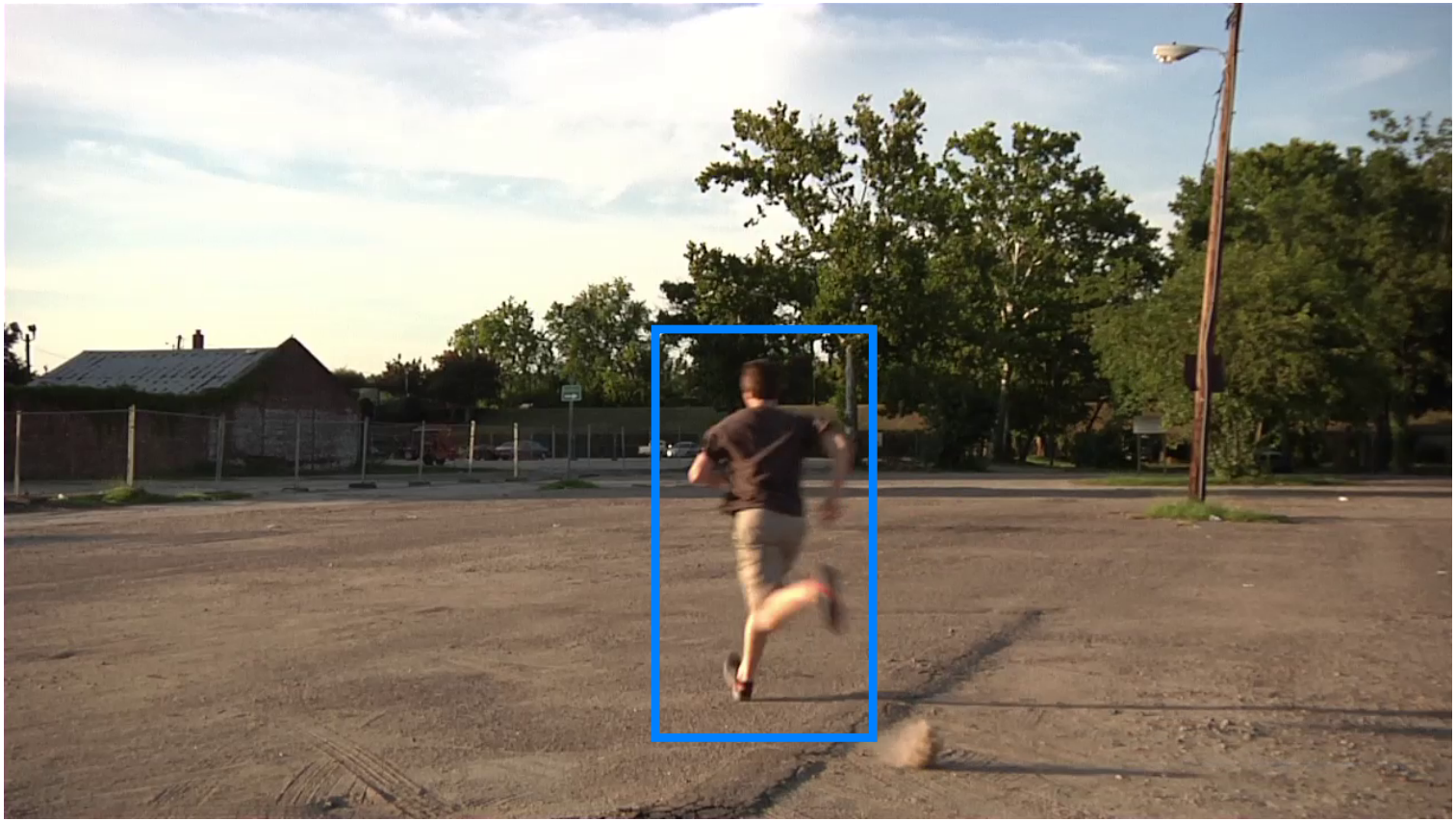
position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—single participant



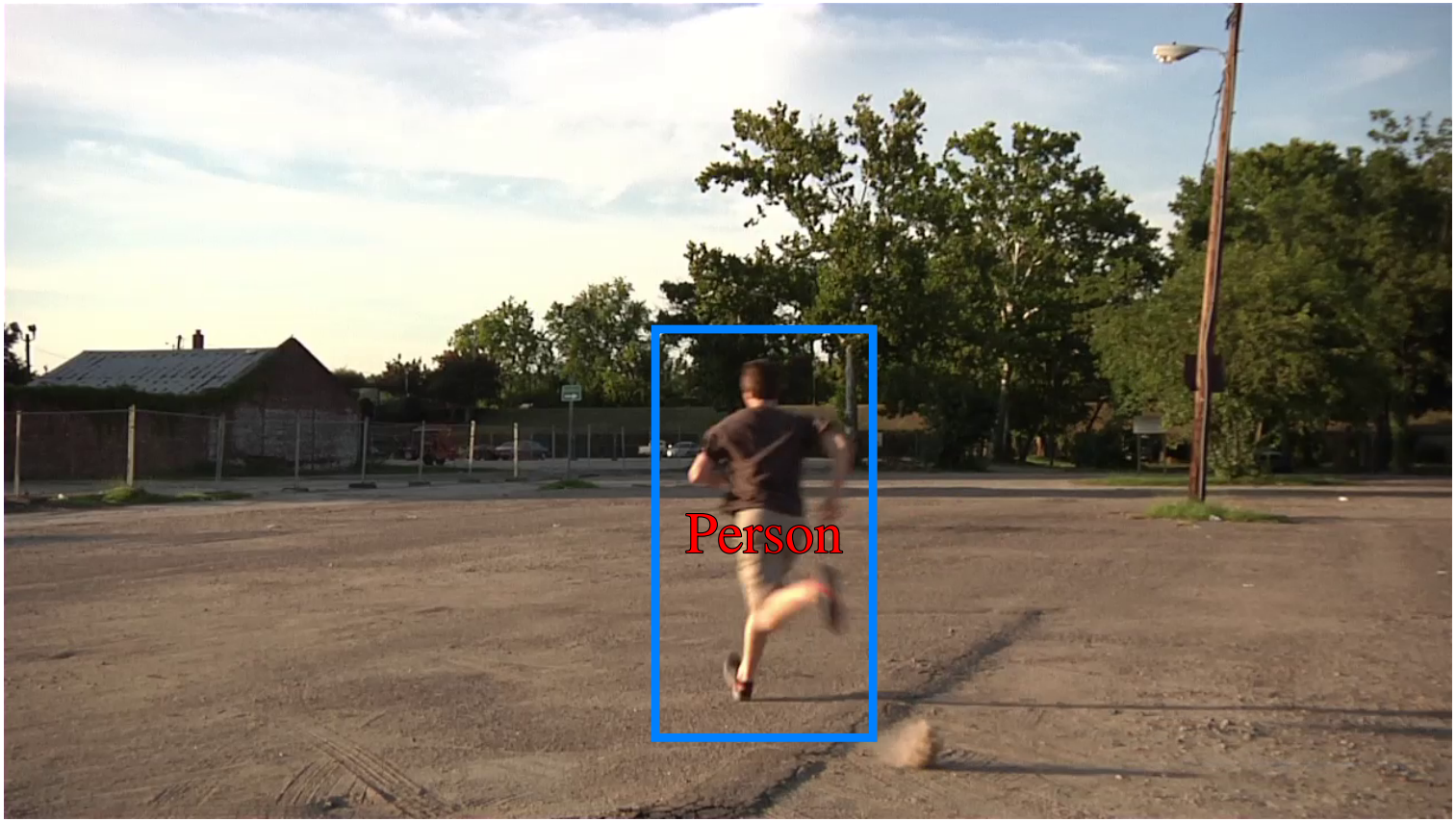
position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—single participant



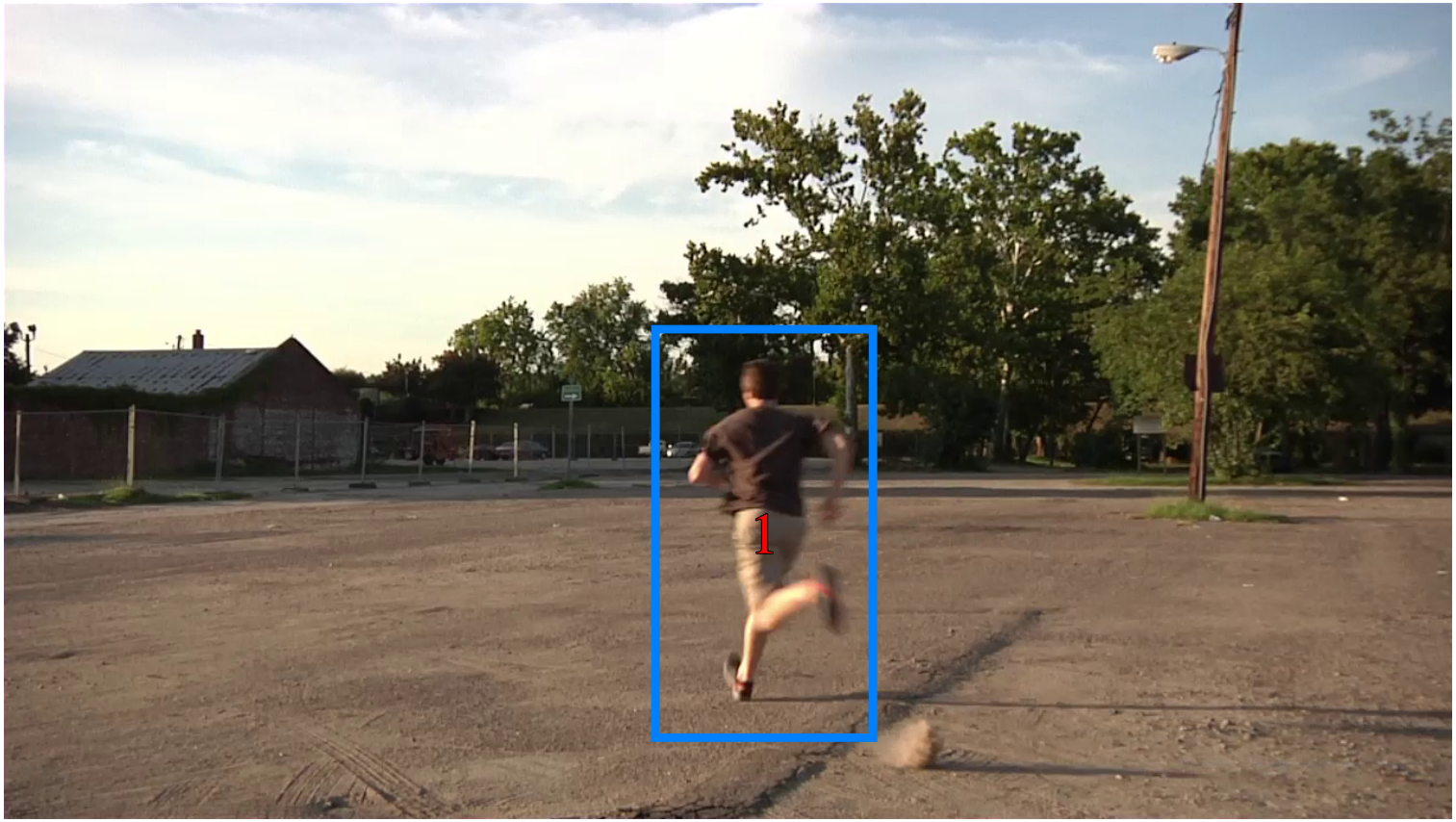
position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—single participant



position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—single participant



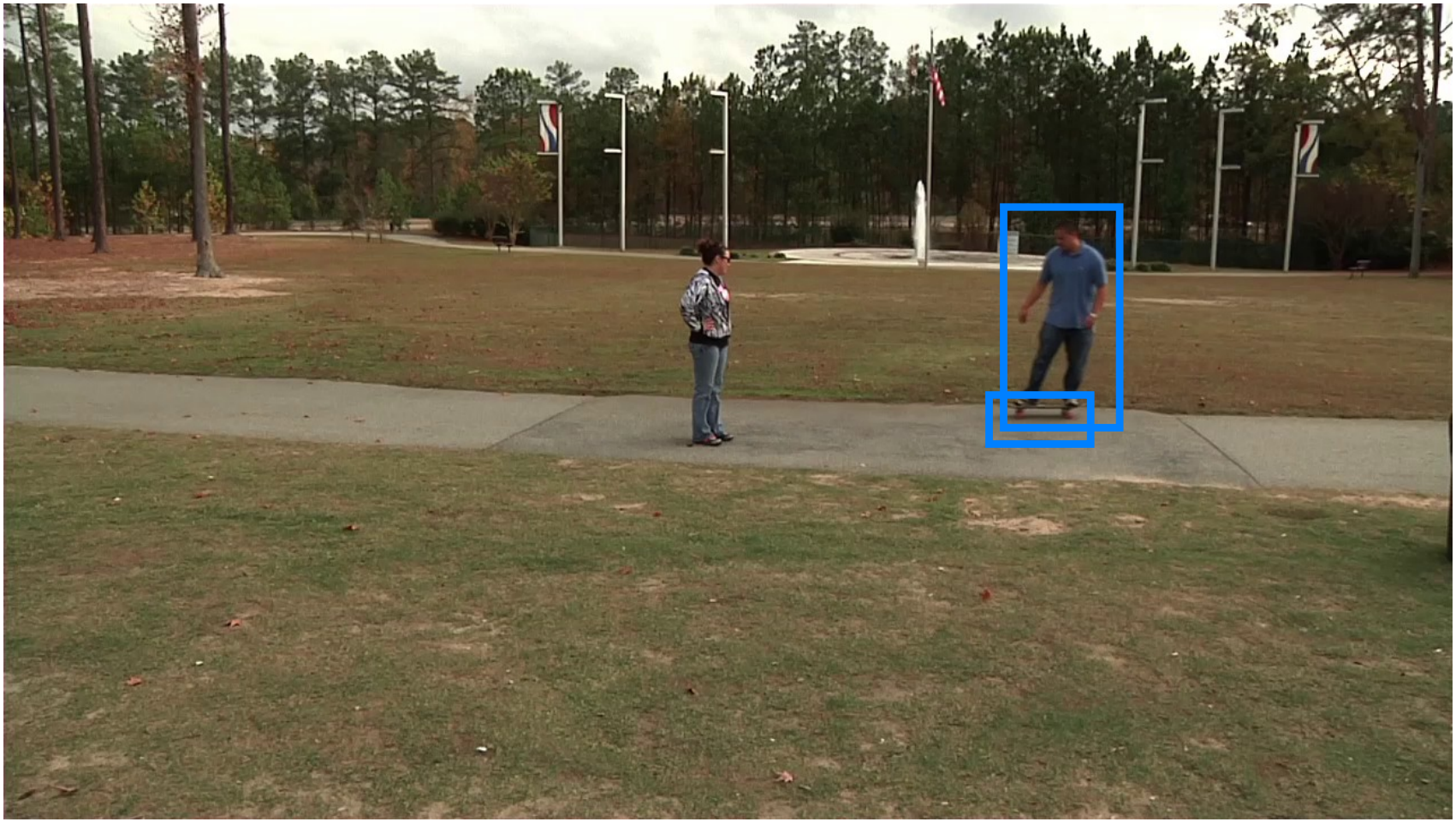
position	$\frac{\mathrm{d}}{\mathrm{d}t}$position	$\frac{\mathrm{d}^2}{\mathrm{d}t^2}$position	
aspect ratio	$\frac{\mathrm{d}}{\mathrm{d}t}$aspect ratio	area	$\frac{\mathrm{d}}{\mathrm{d}t}$area
object class	root-filter index		

Feature vector—dual participant

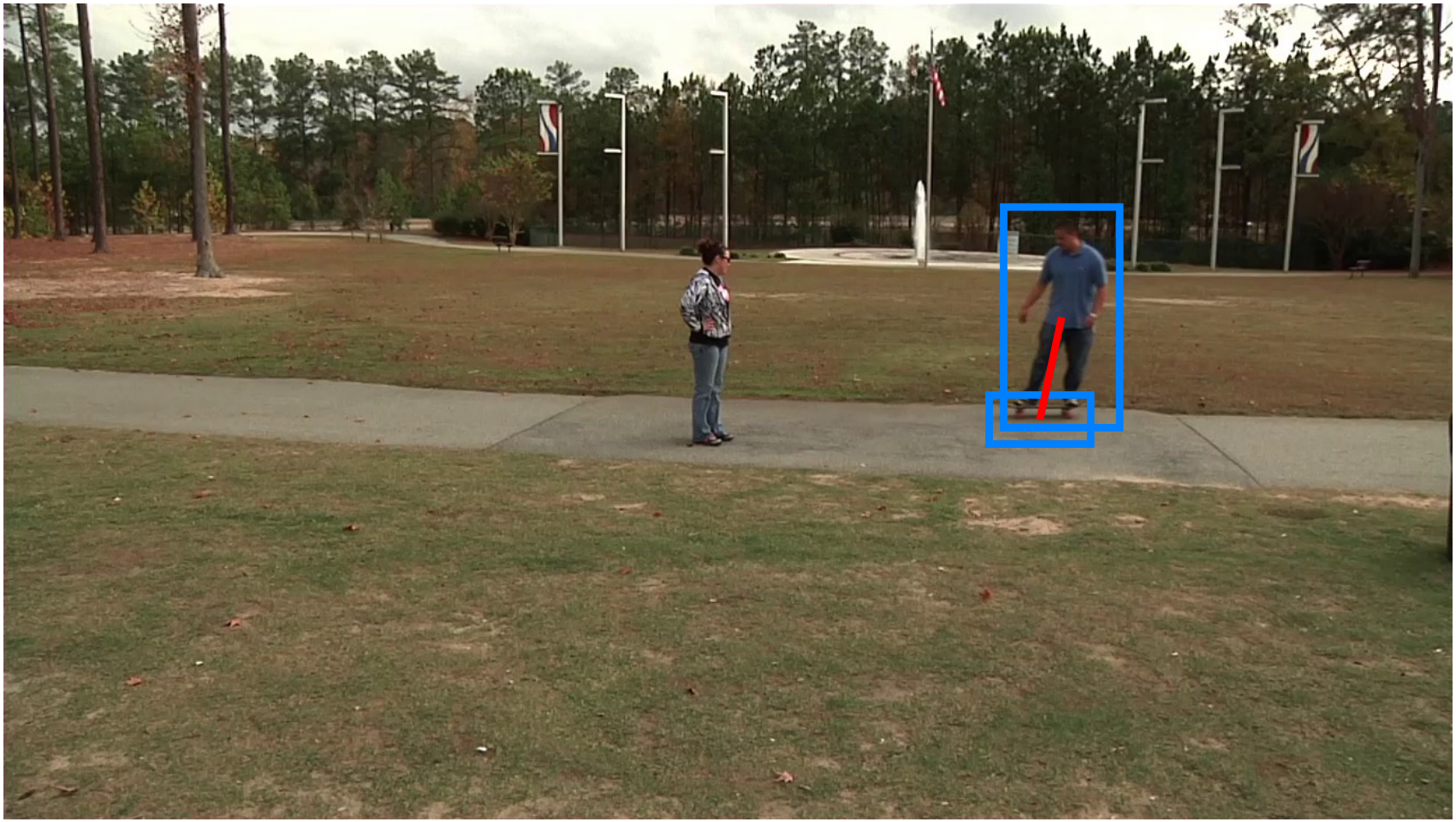
Feature vector—dual participant

Feature vector—dual participant

Feature vector—dual participant

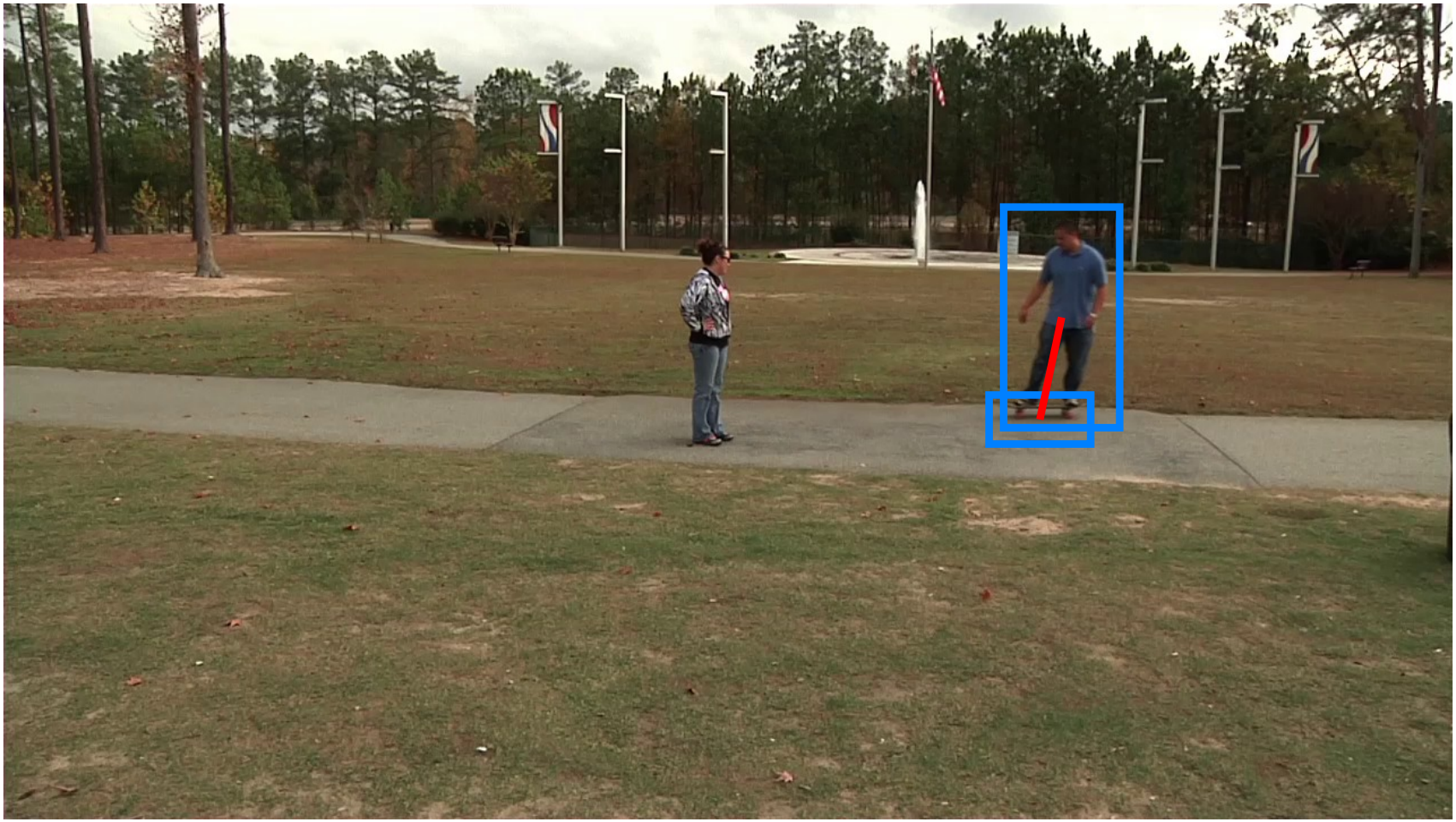


Feature vector—dual participant



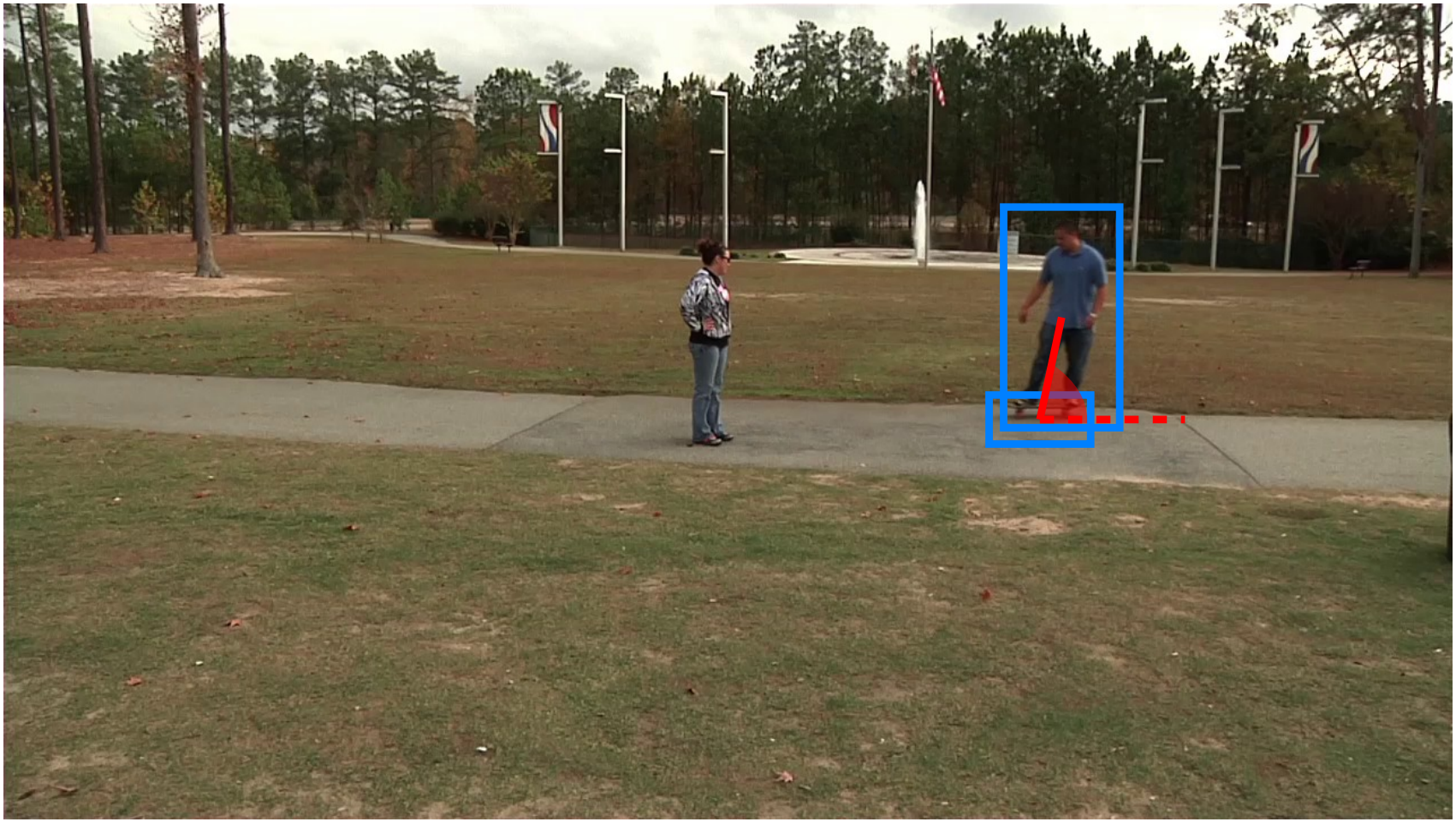
distance $\frac{\mathrm{d}}{\mathrm{d}t}$distance orientation $\frac{\mathrm{d}}{\mathrm{d}t}$orientation

Feature vector—dual participant



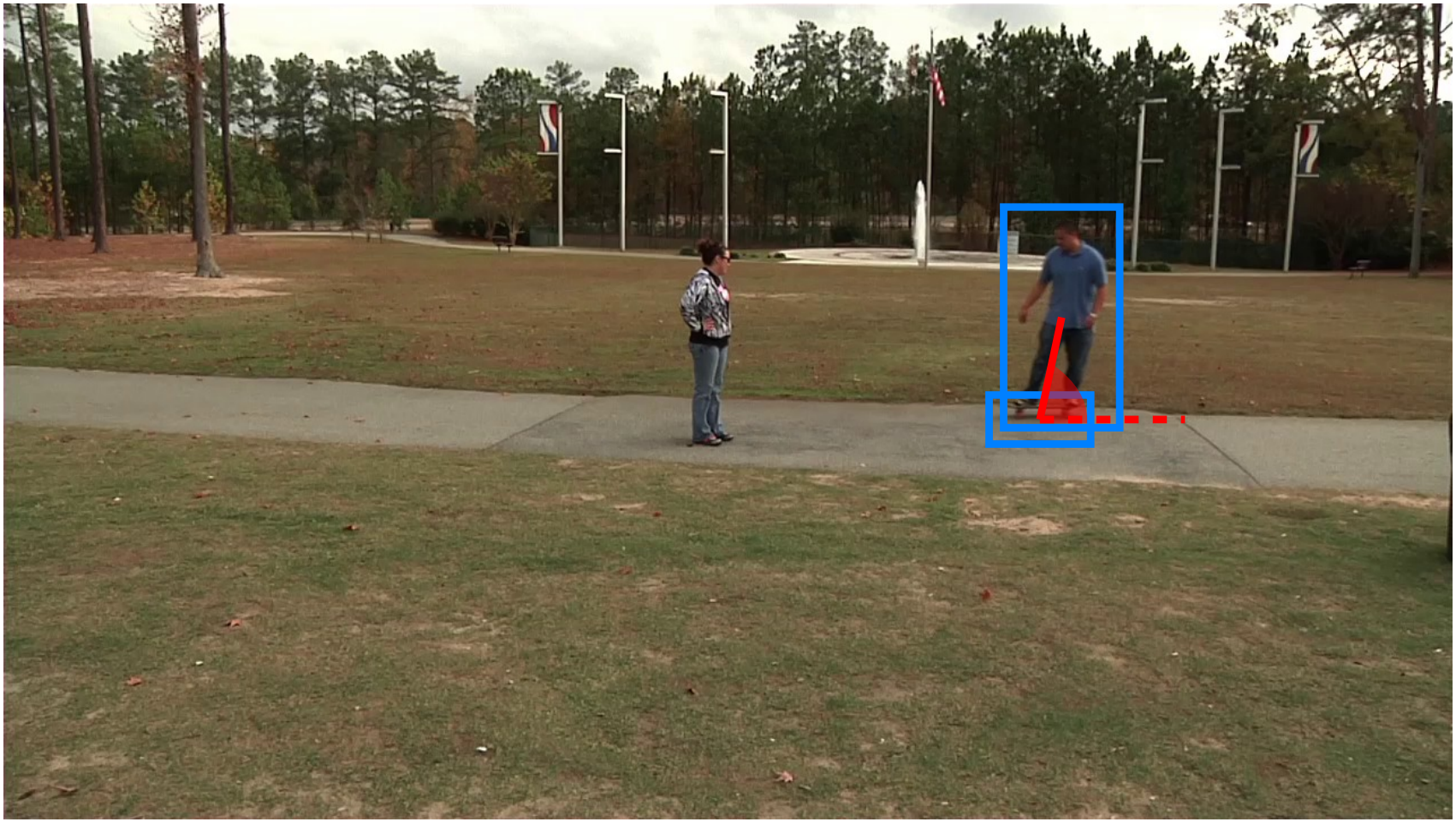
$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

Feature vector—dual participant



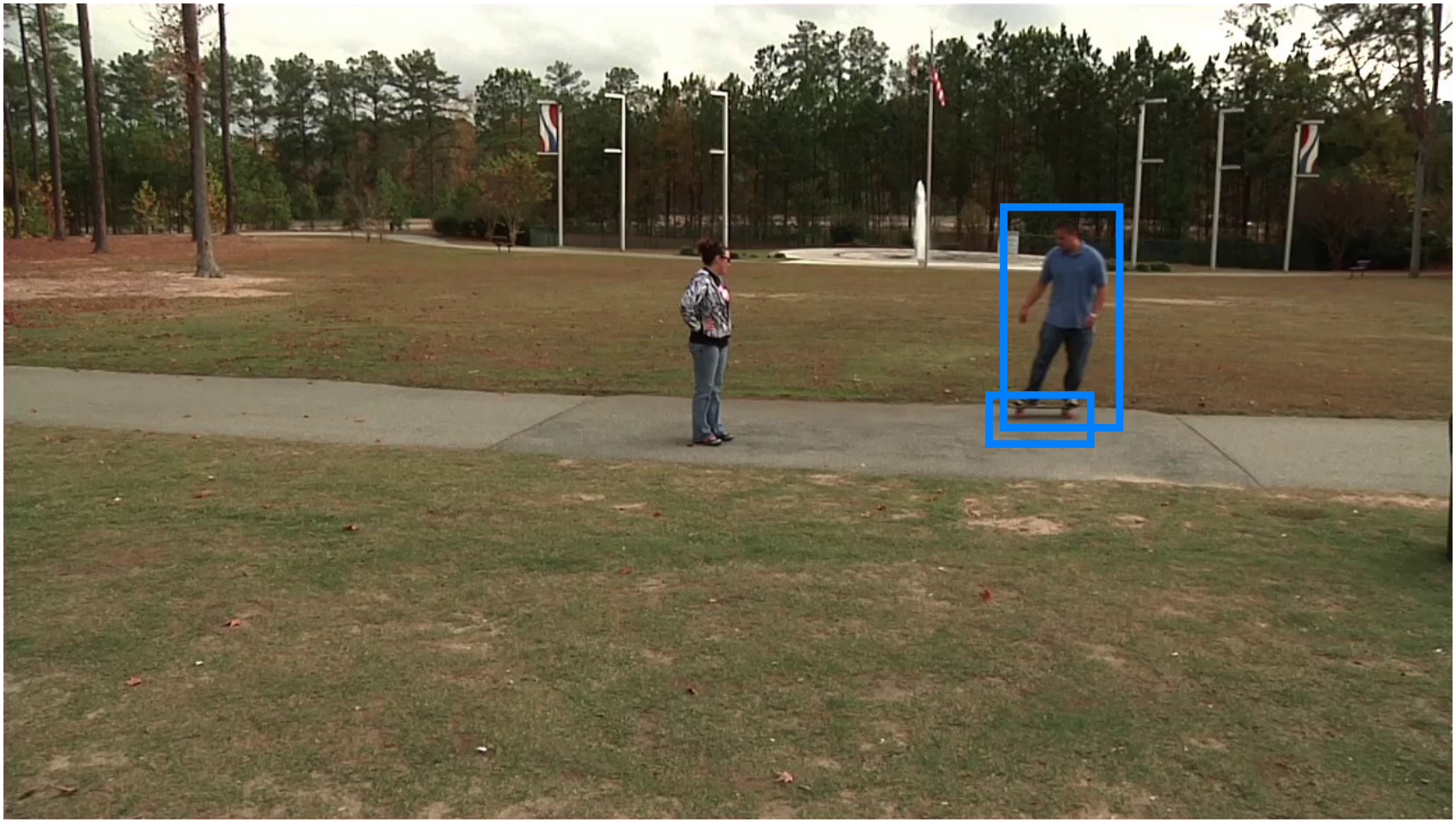
$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard

Feature vector—dual participant

$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

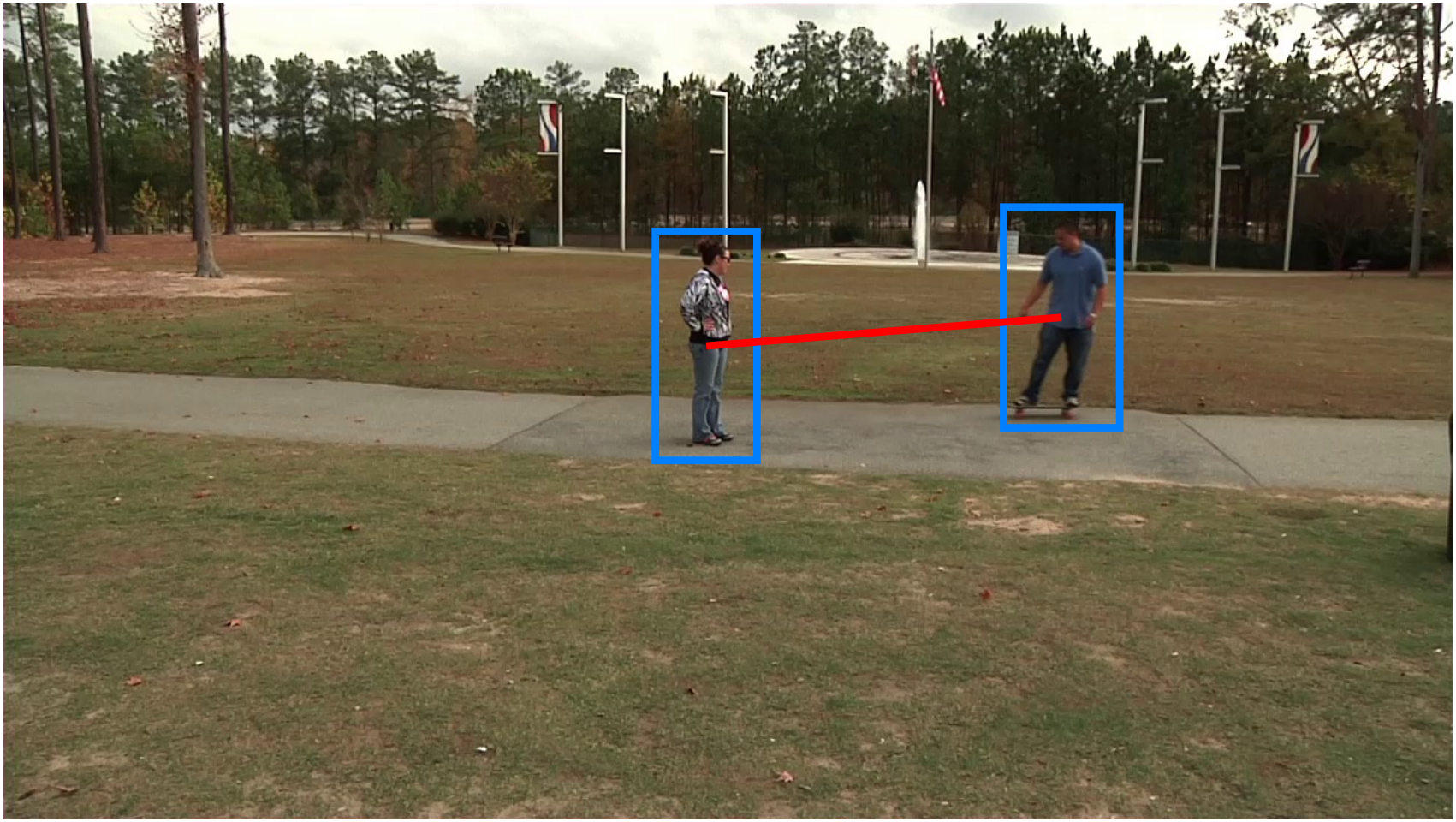
person riding skateboard

Feature vector—dual participant

$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard

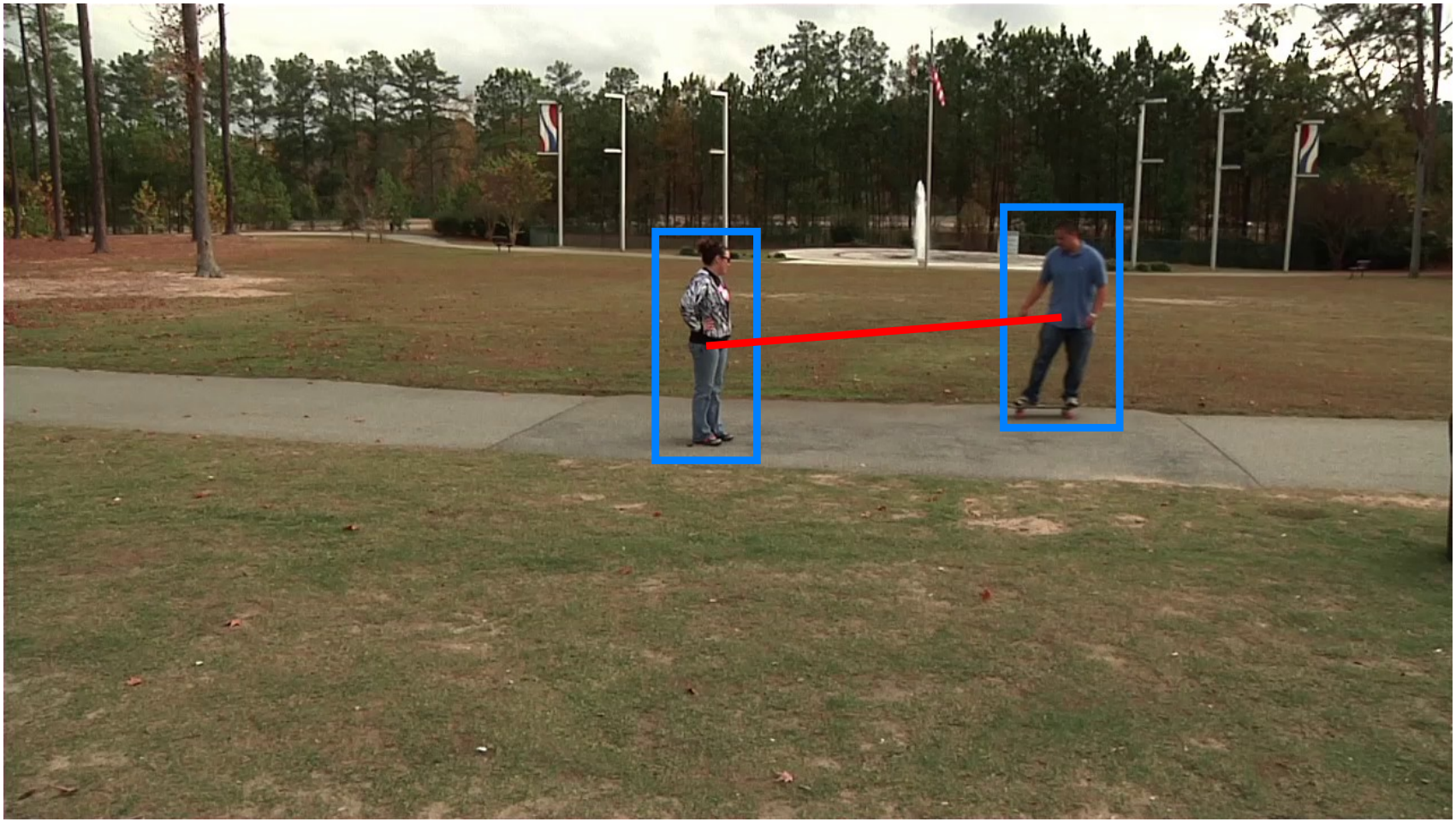
Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard

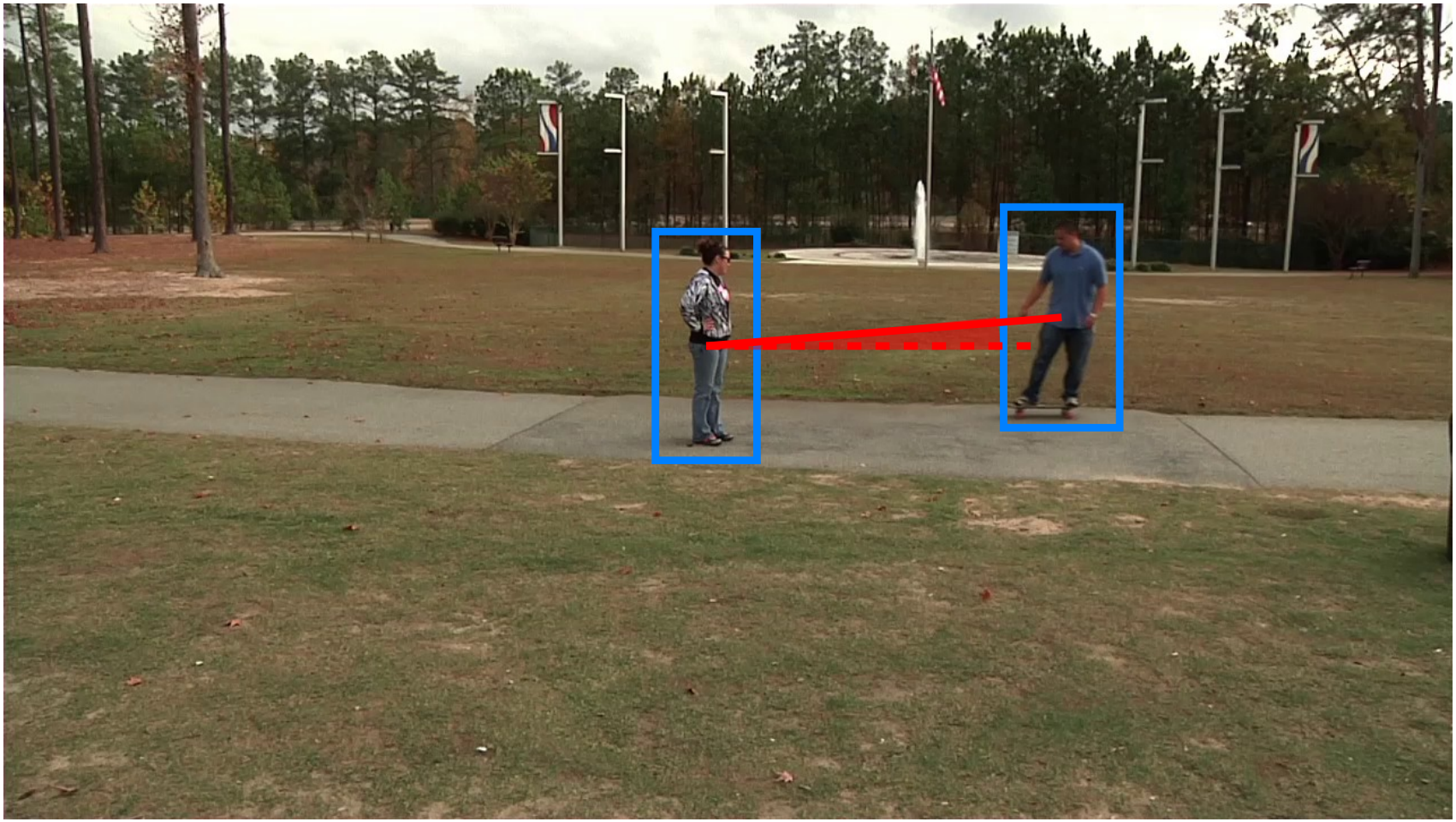
Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard

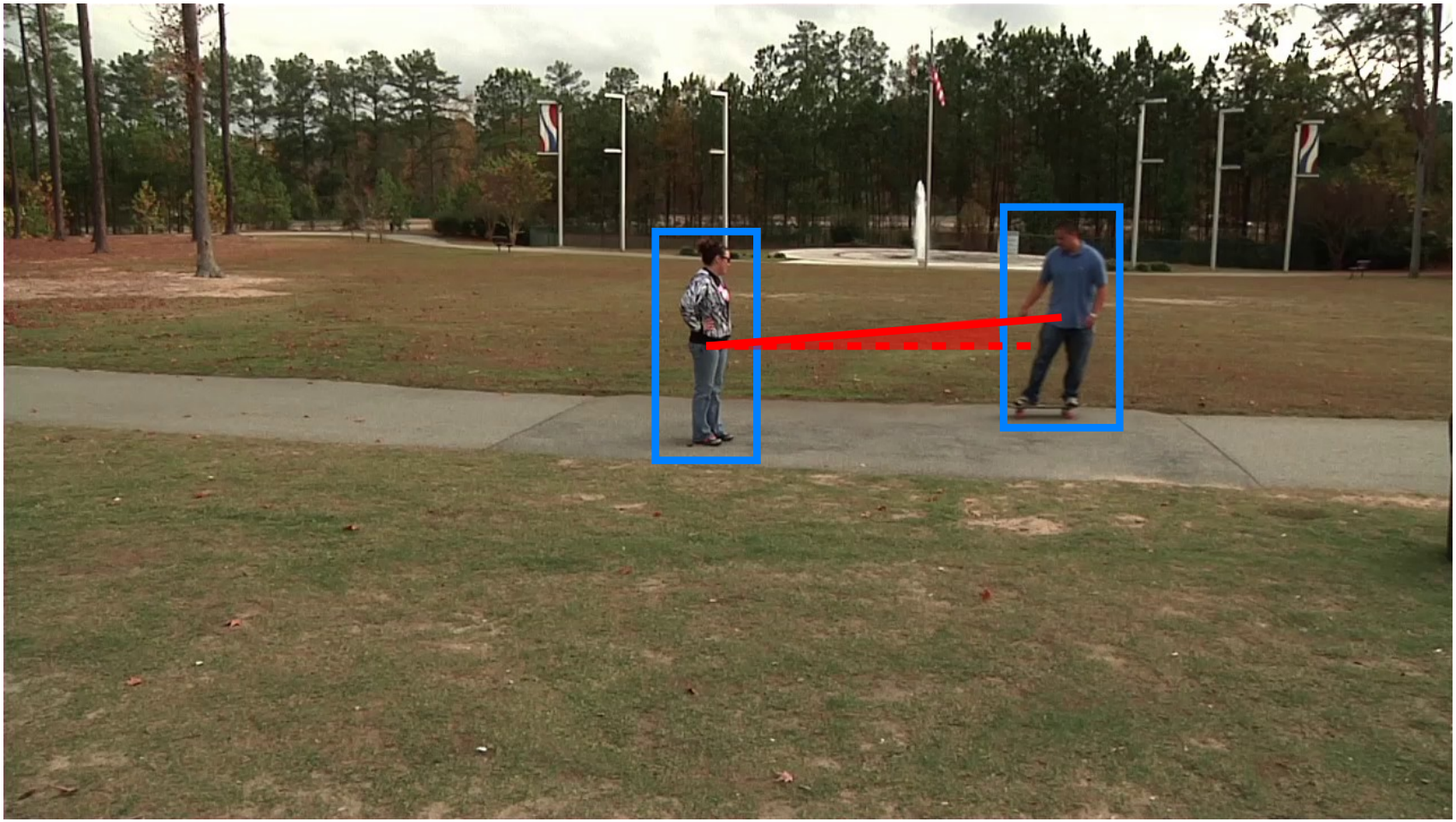
Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard

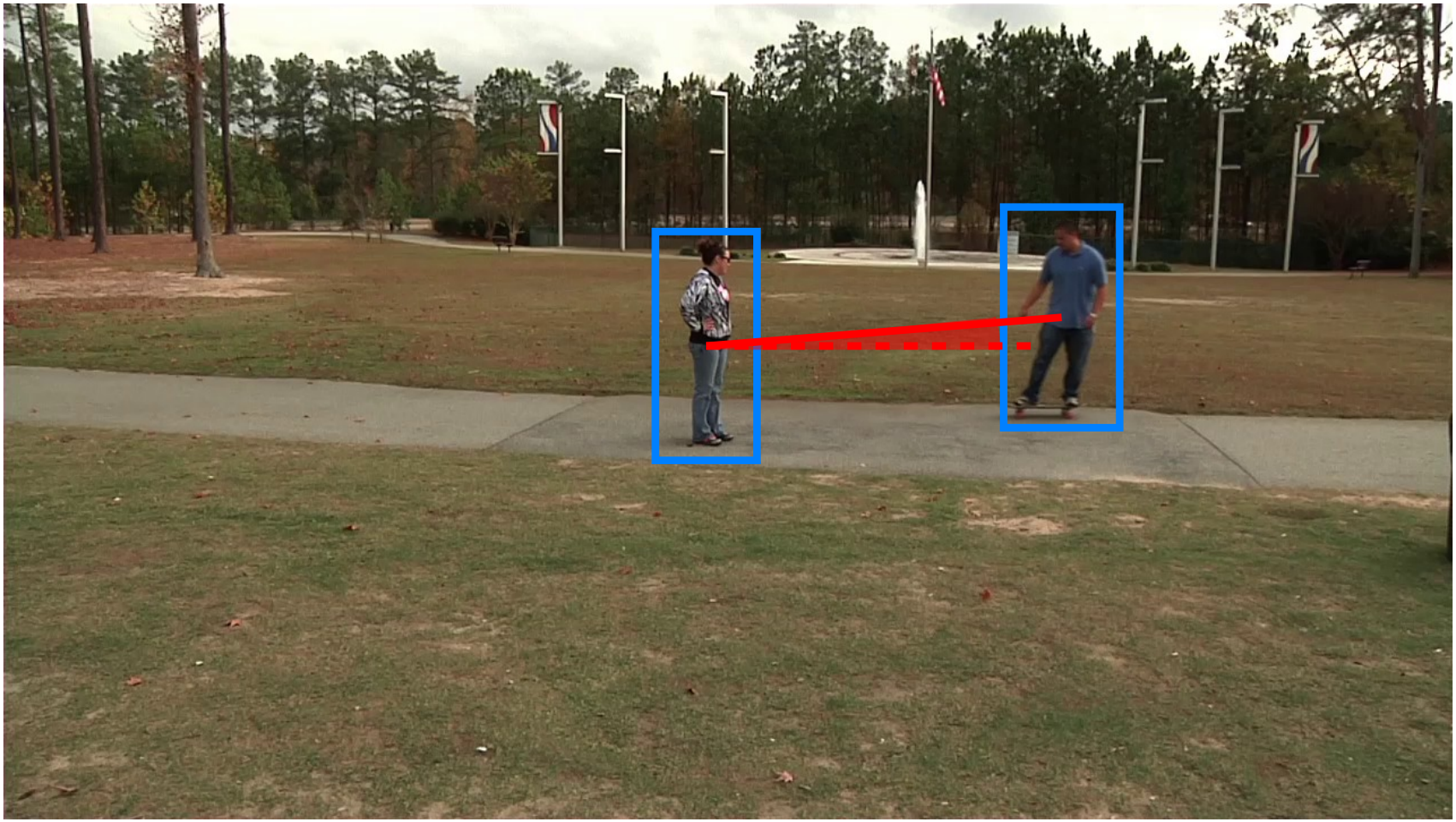
Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard

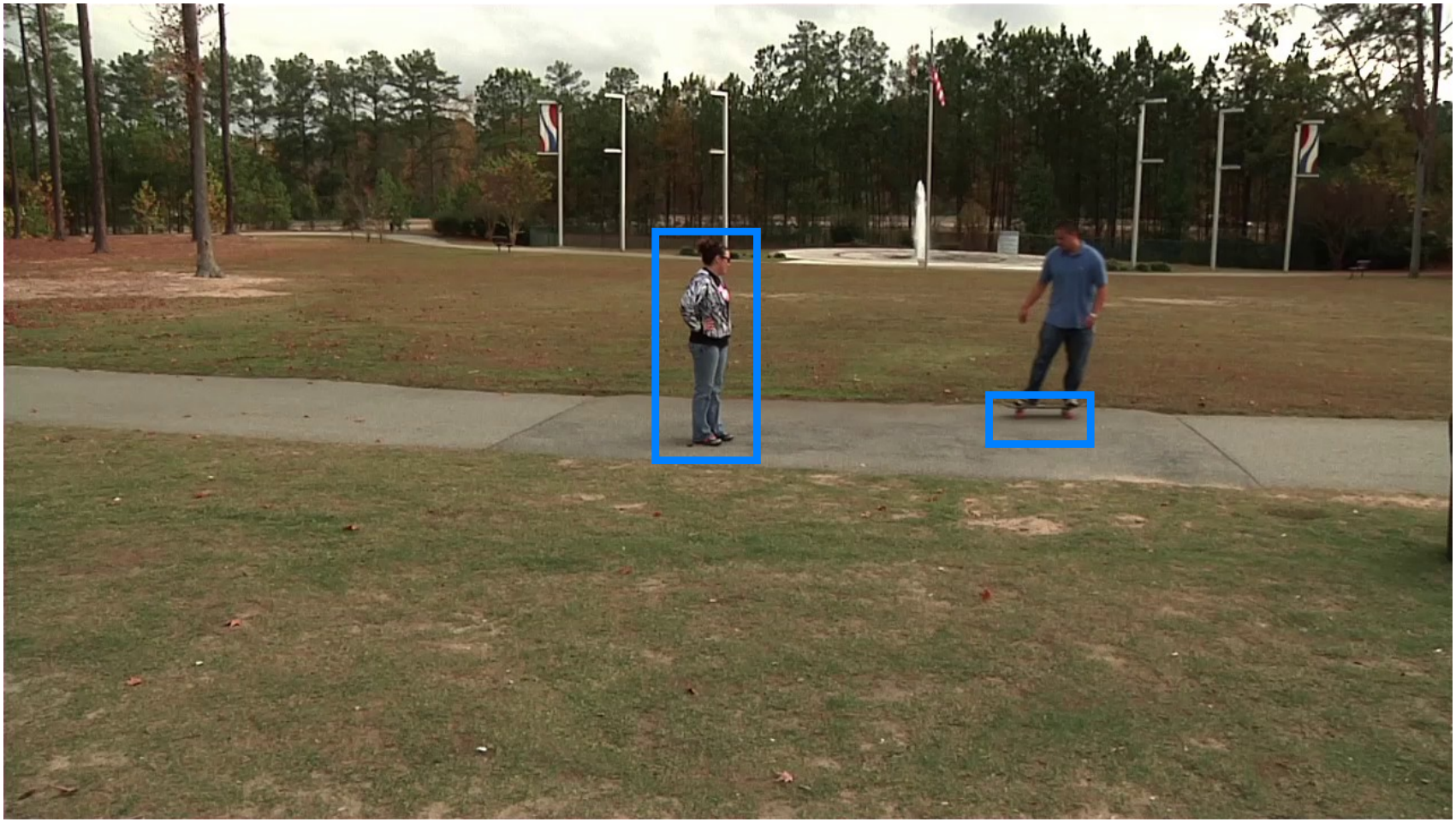
Feature vector—dual participant



$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard person approaching person

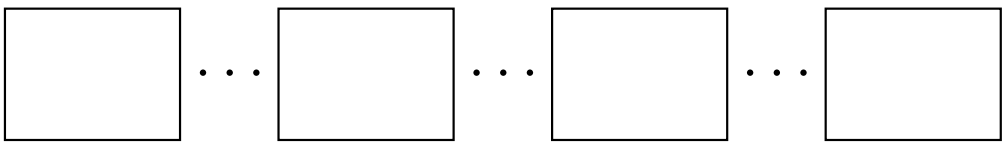
Feature vector—dual participant



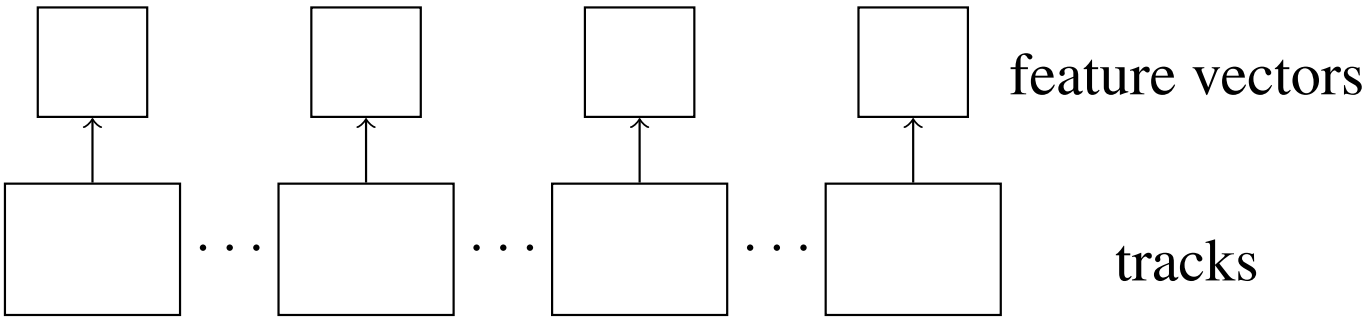
$$\text{distance} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{distance} \quad \text{orientation} \quad \frac{\mathrm{d}}{\mathrm{d}t}\text{orientation}$$

person riding skateboard | person approaching person | skateboard approaching person

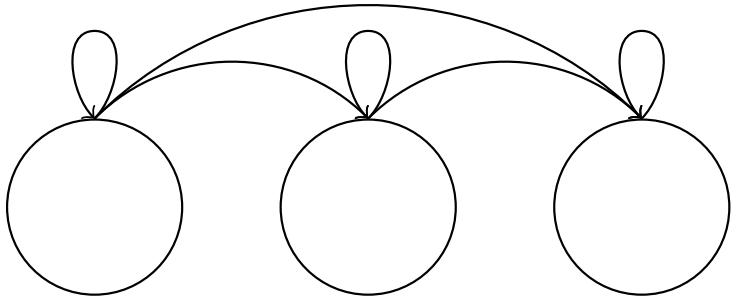
Event recognition



Event recognition



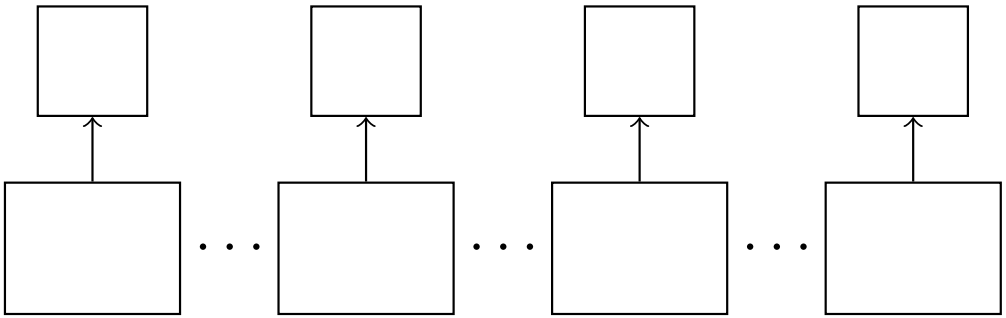
Event recognition



a

- HMMs

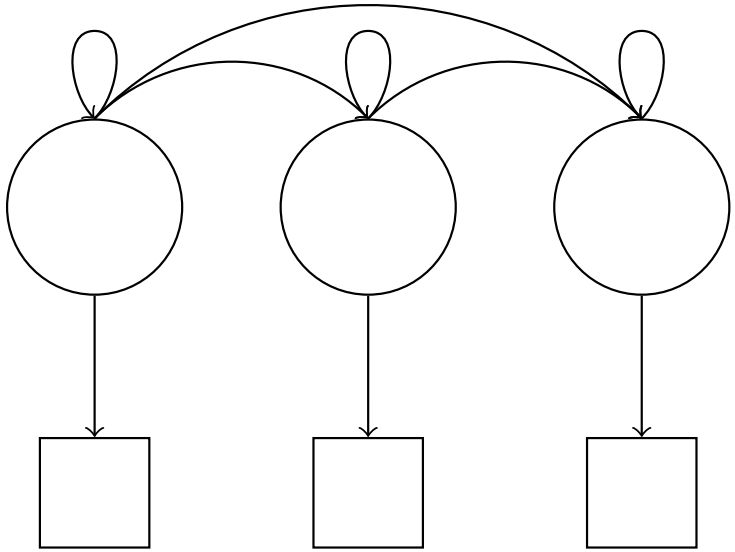
  Baum and Petrie (1966)



feature vectors

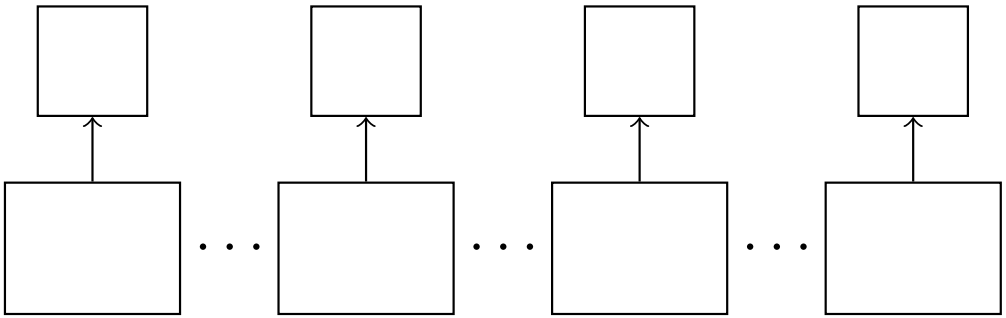
tracks

Event recognition



- HMMs

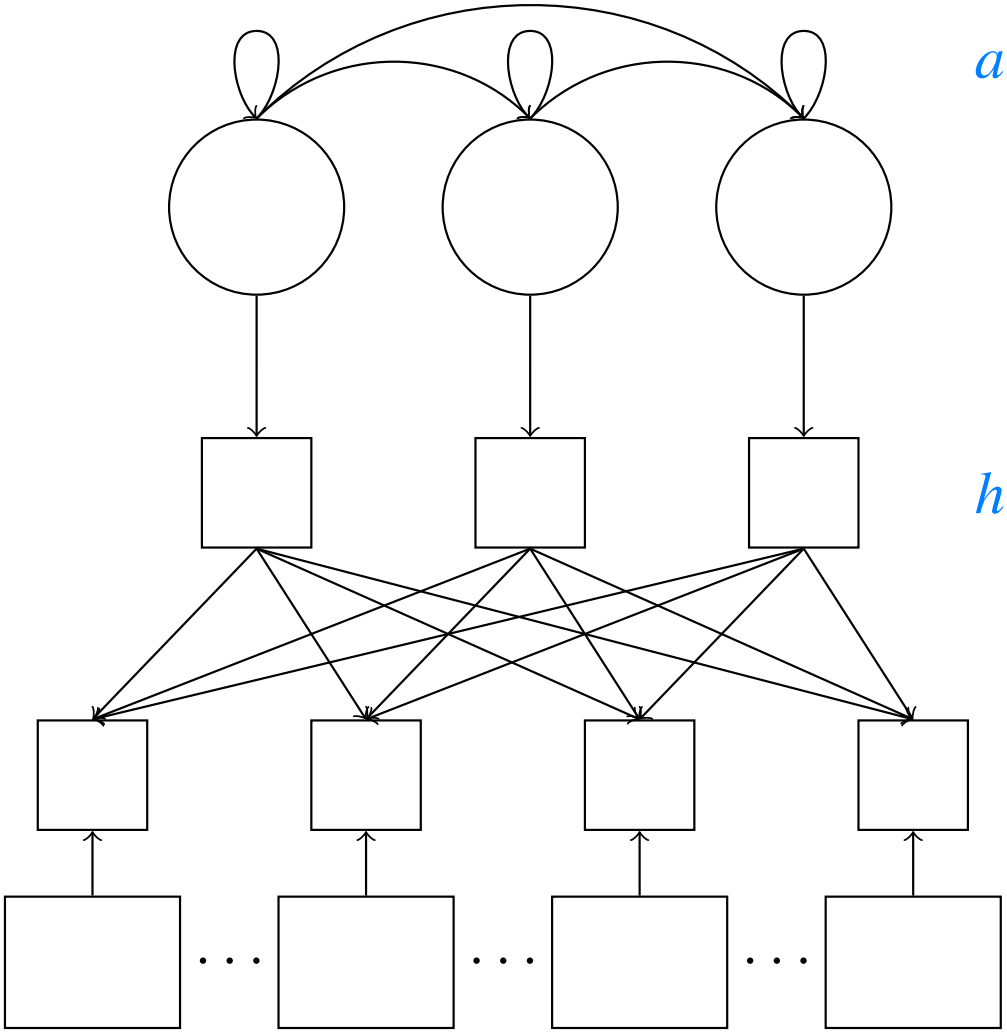
  Baum and Petrie (1966)



feature vectors

tracks

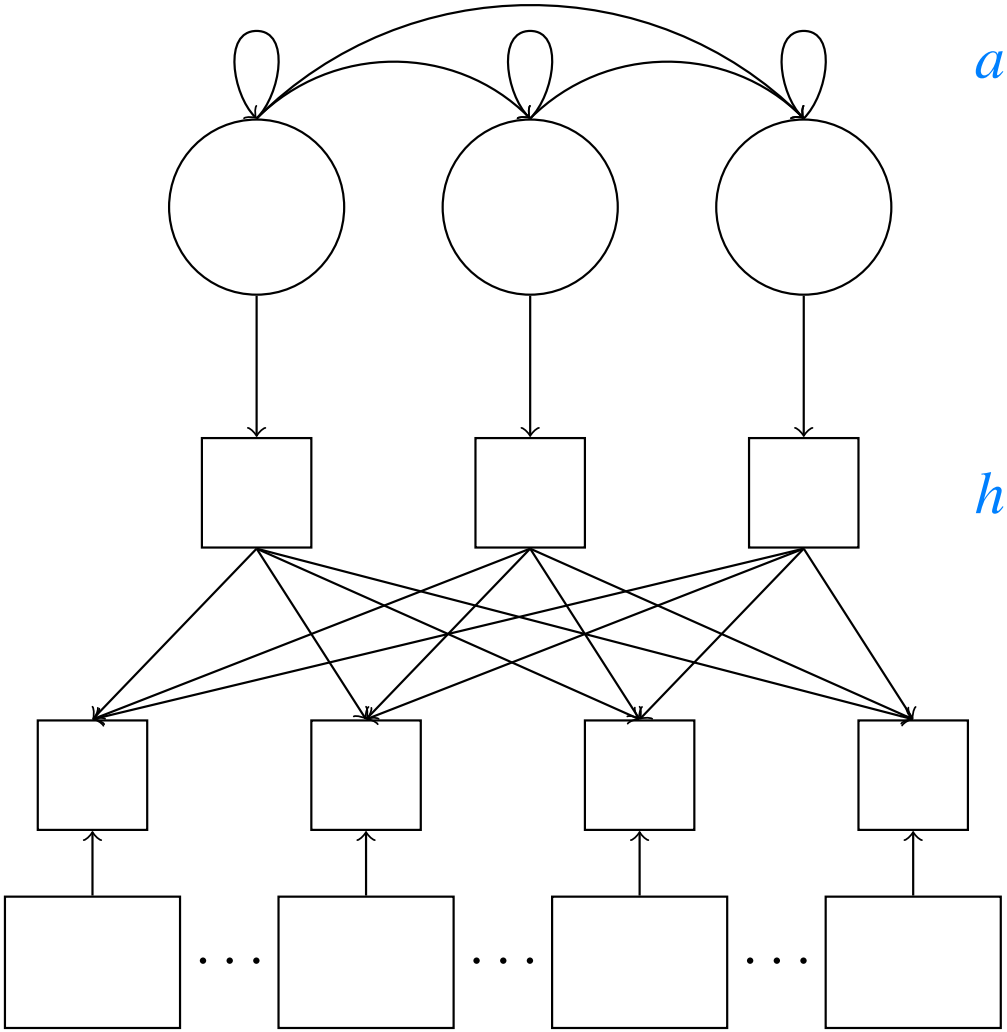
Event recognition



- HMMs

  Baum and Petrie (1966)

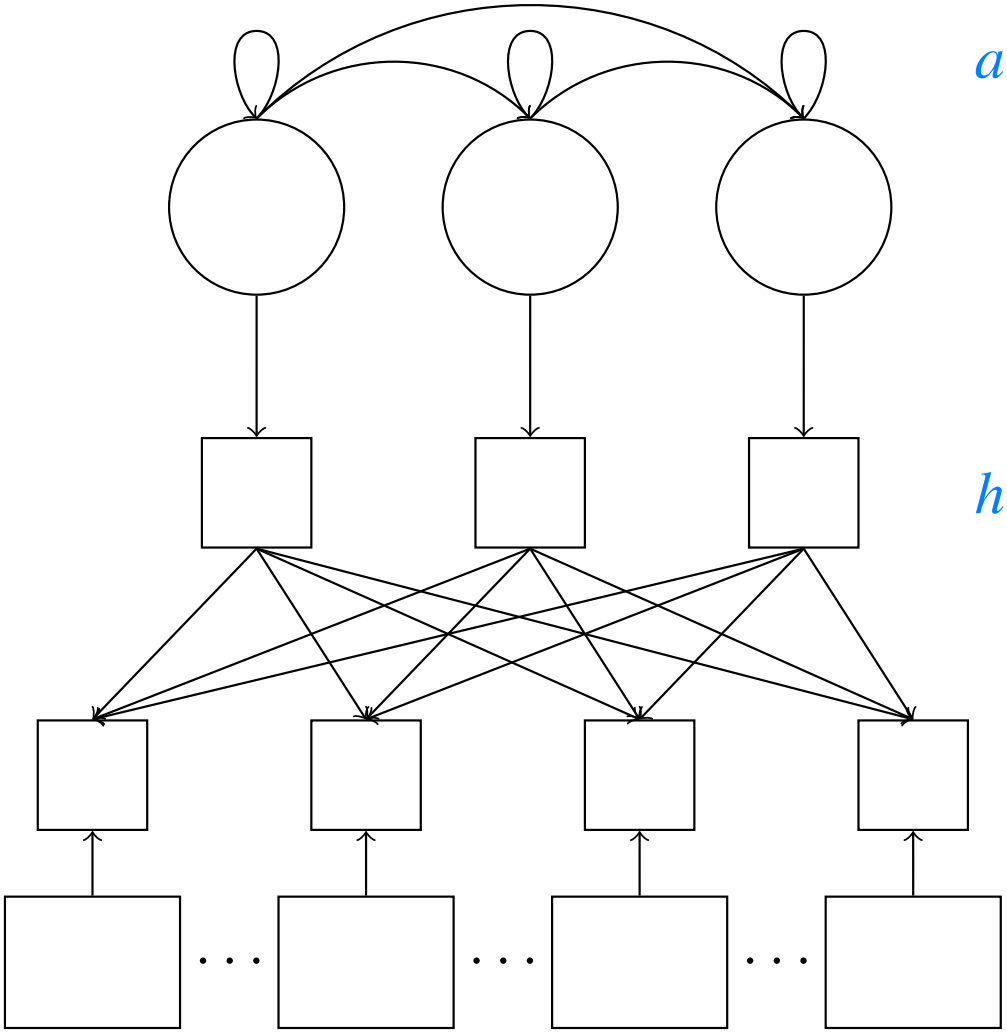
Event recognition



- HMMs

  Baum and Petrie (1966)
- One HMM per event class

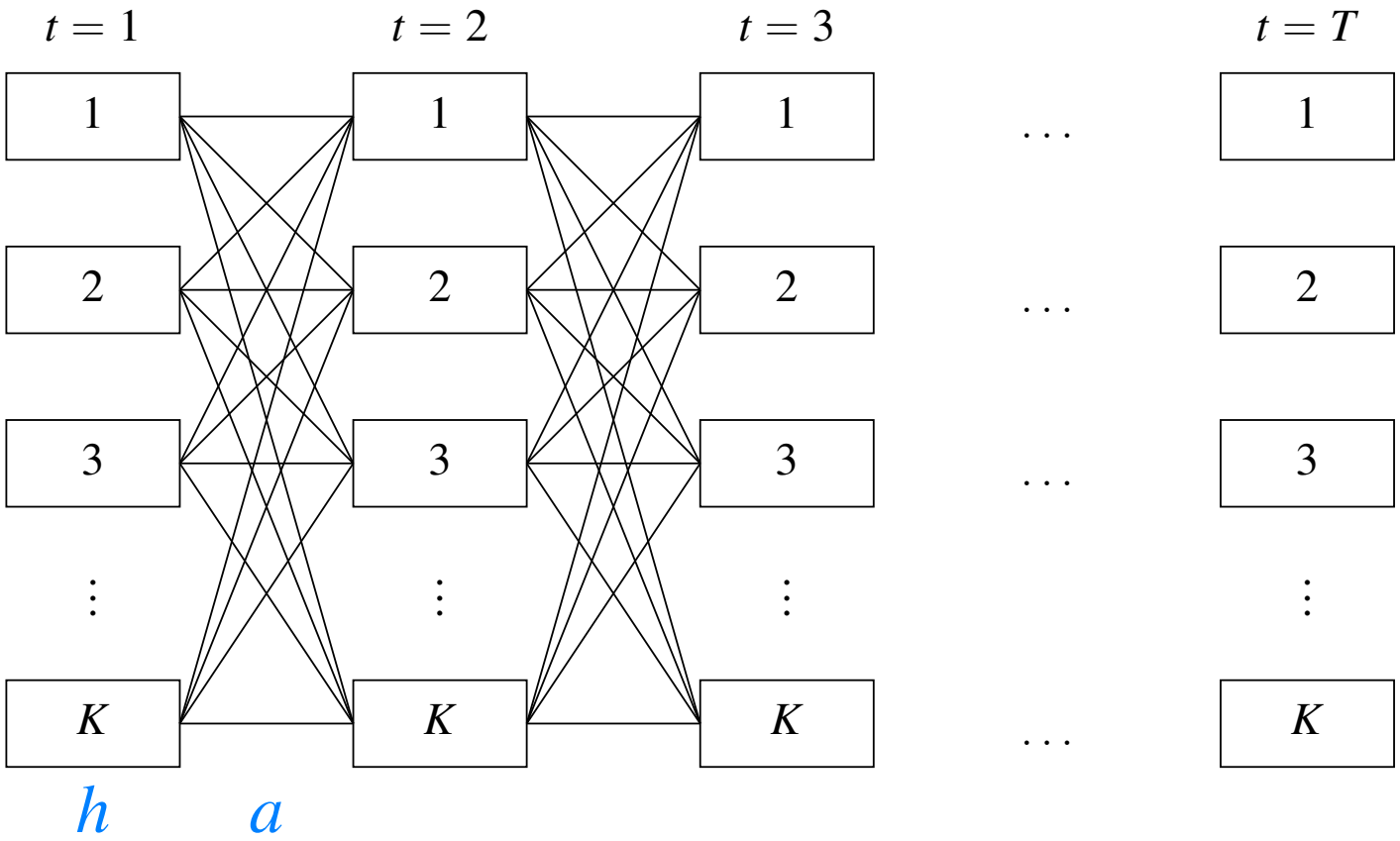
Event recognition



- HMMs

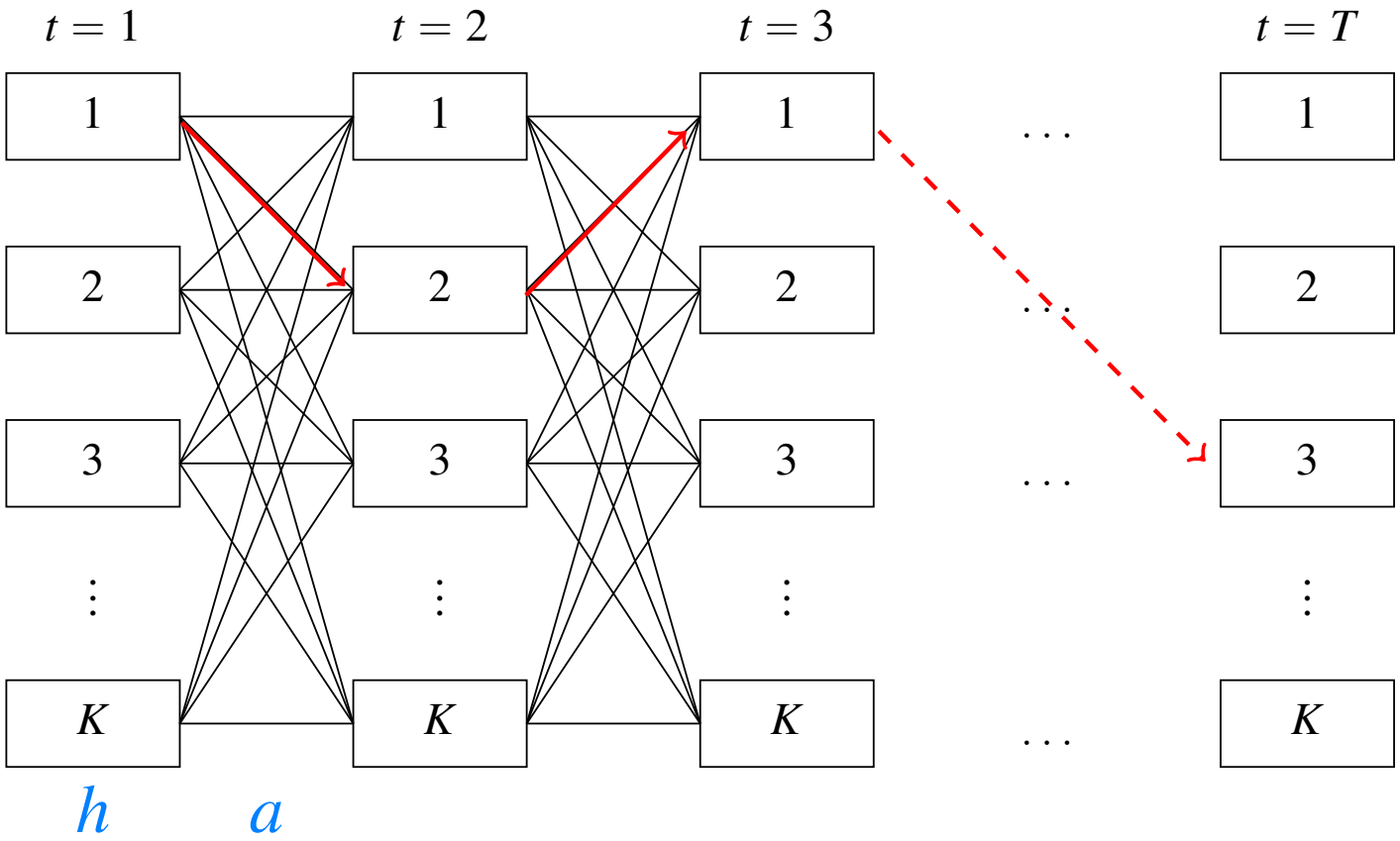
  Baum and Petrie (1966)
- One HMM per event class
- Try each HMM

Event recognition



$$\max_{k^1,\ldots,k^T} \sum_{t=1}^{T} h(k^t, b^t_{\hat{\jmath}^t}) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

Event recognition

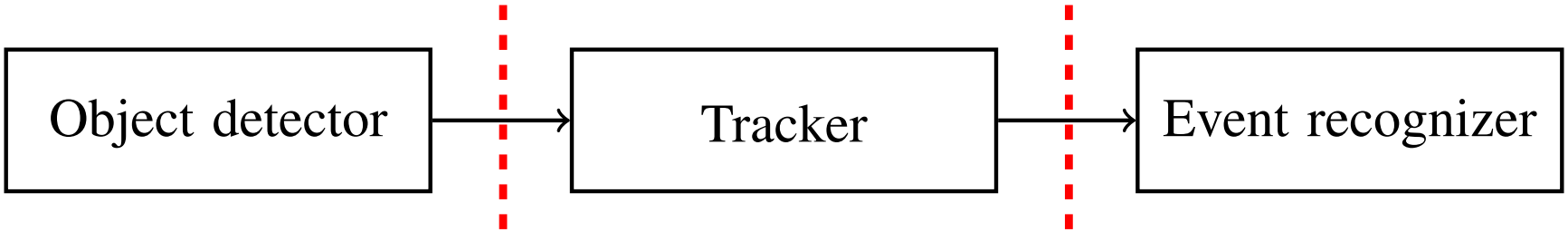


$$\max_{k^1,\ldots,k^T} \sum_{t=1}^{T} h(k^t, b^t_{\hat{\jmath}^t}) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

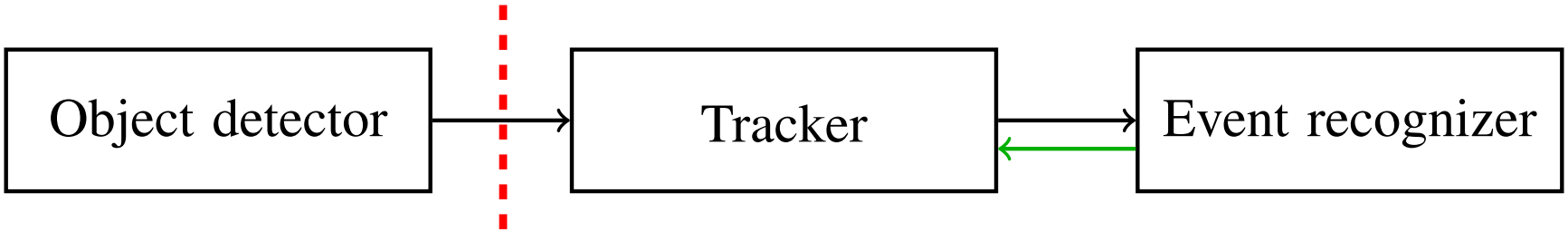
Examples

Examples

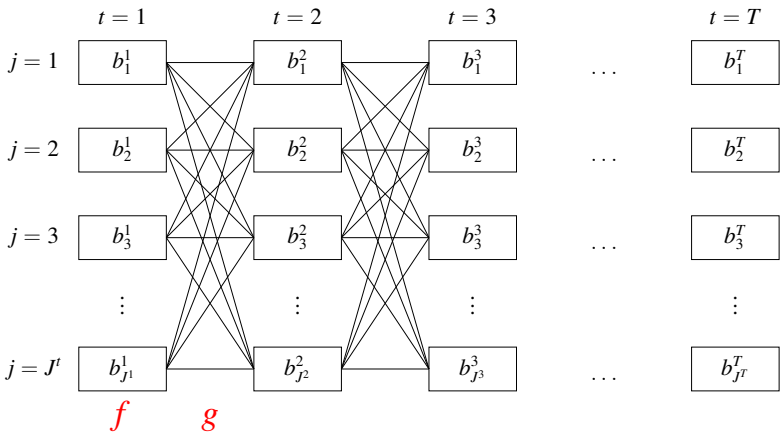
What we're going to do



What we're going to do

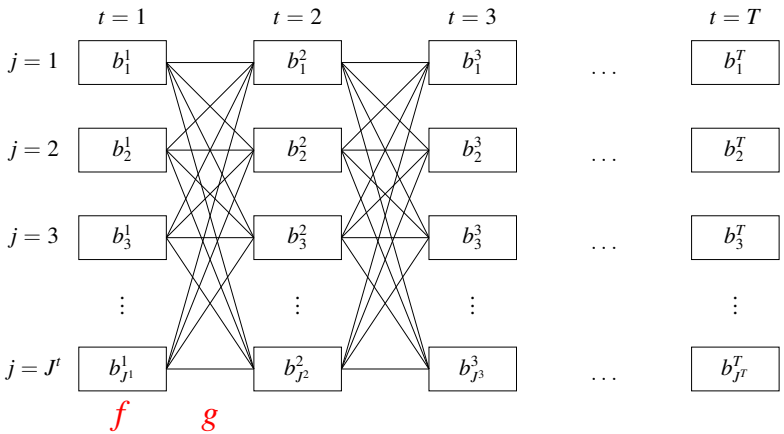


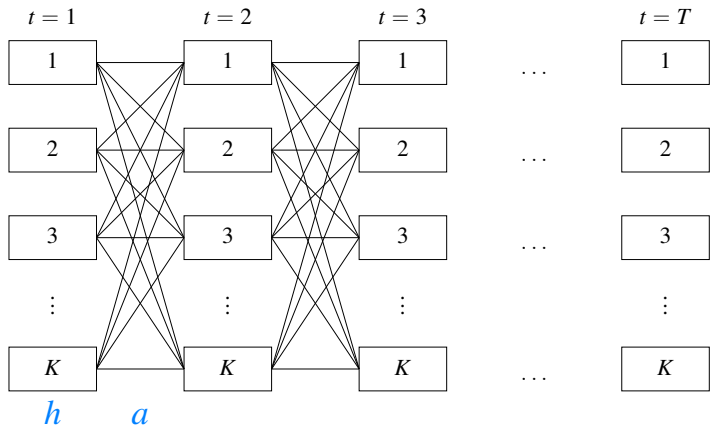
Tracking in the context of event recognition



$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t)$$

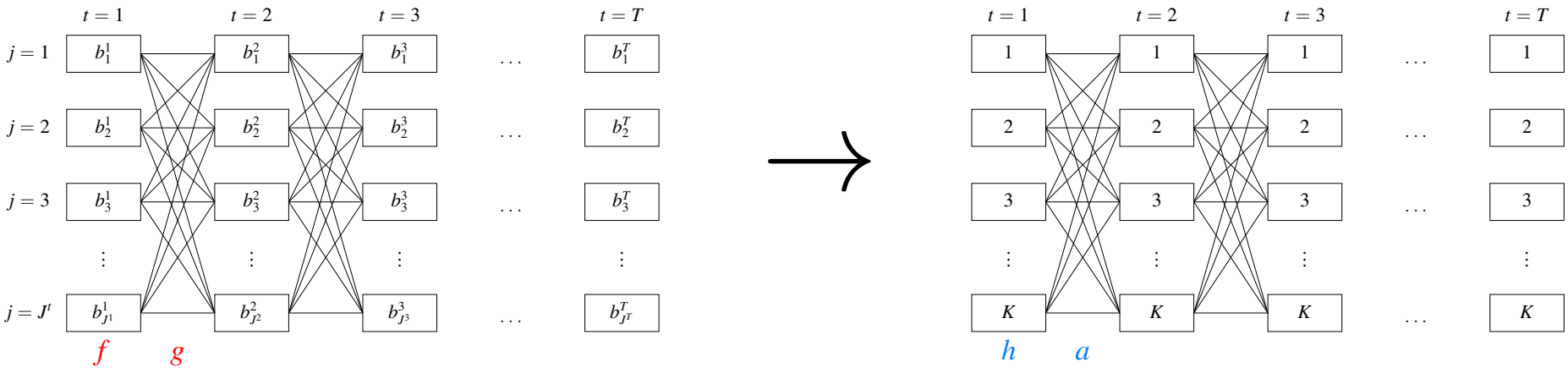
Tracking in the context of event recognition





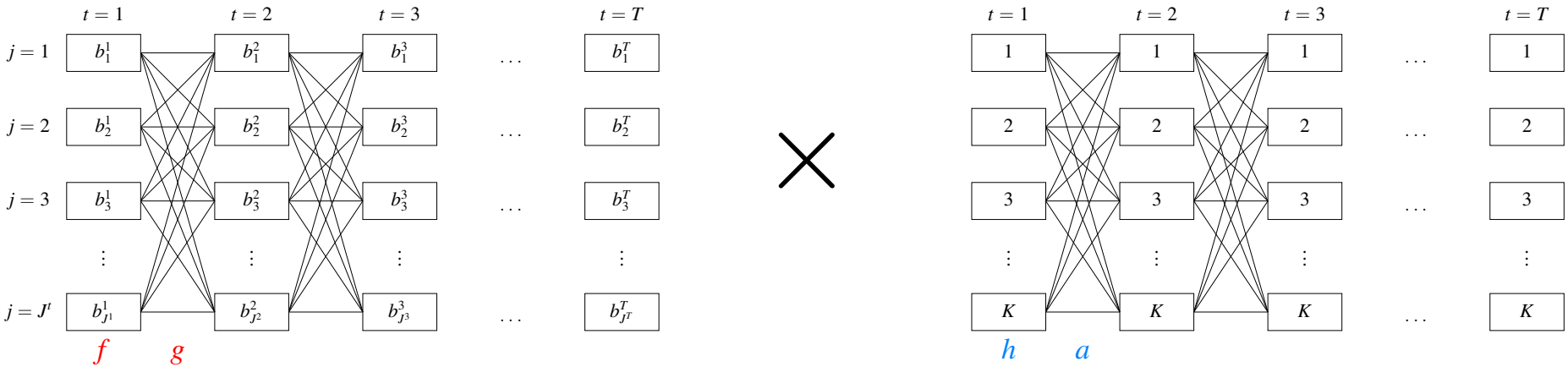
$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b^t_{j^t}) + \sum_{t=2}^{T} g(b^{t-1}_{j^{t-1}}, b^t_{j^t}) + \max_{k^1,\ldots,k^T} \sum_{t=1}^{T} h(k^t, b^t_{\hat{j}^t}) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



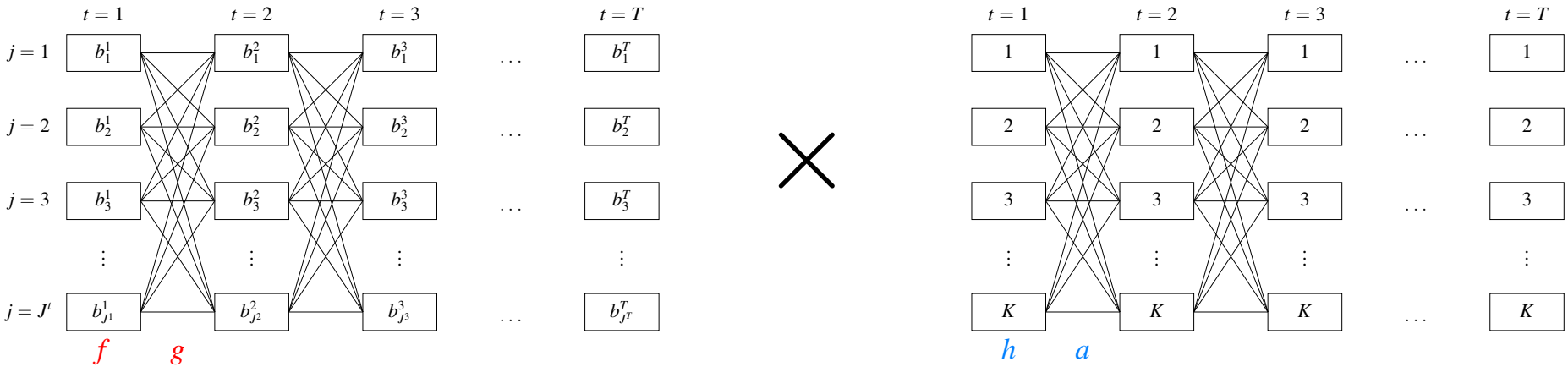
$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b^t_{j^t}) + \sum_{t=2}^{T} g(b^{t-1}_{j^{t-1}}, b^t_{j^t}) + \max_{k^1,\ldots,k^T} \sum_{t=1}^{T} h(k^t, b^t_{\hat{j}^t}) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



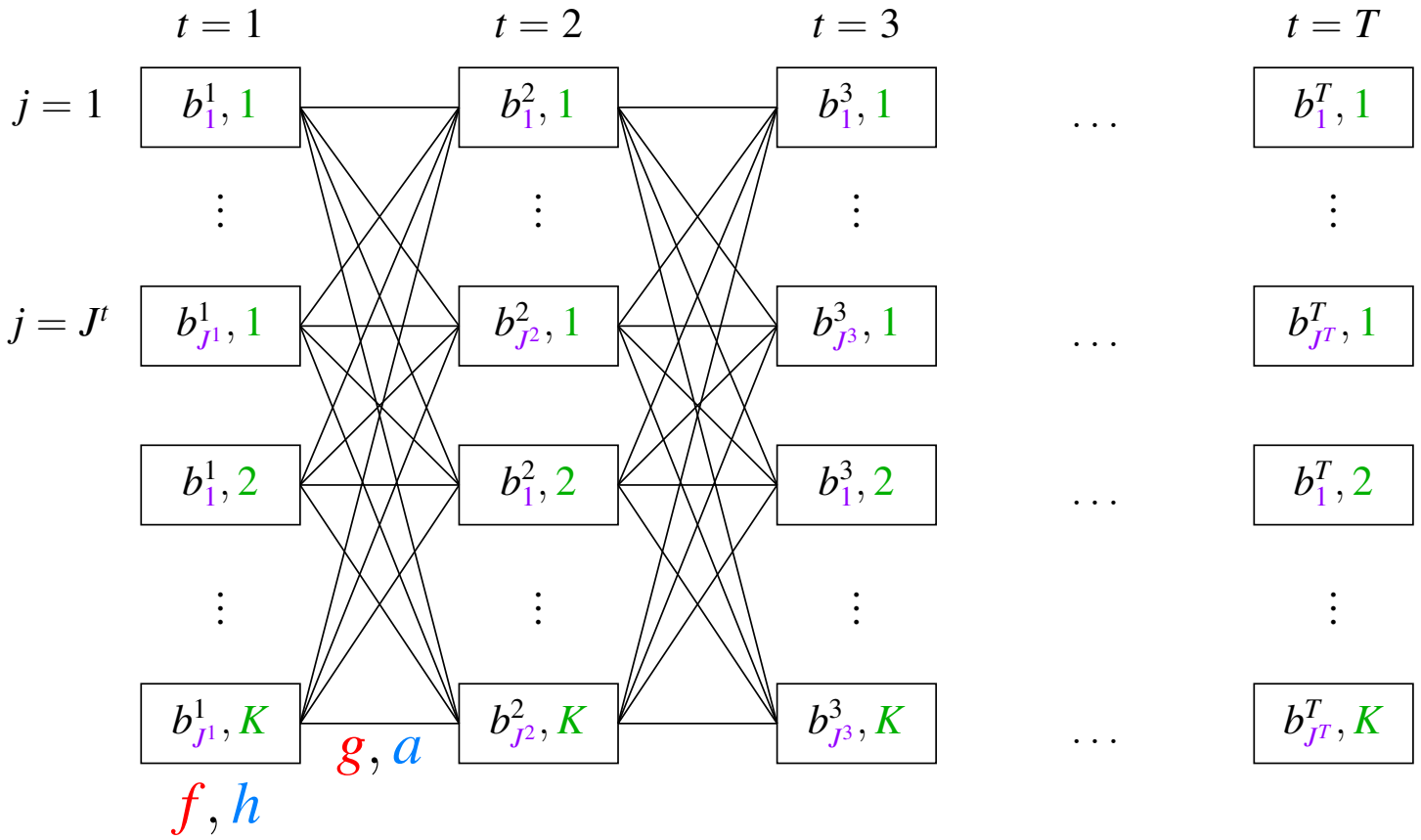
$$\max_{j^1,\ldots,j^T} \sum_{t=1}^{T} f(b^t_{j^t}) + \sum_{t=2}^{T} g(b^{t-1}_{j^{t-1}}, b^t_{j^t}) + \max_{k^1,\ldots,k^T} \sum_{t=1}^{T} h(k^t, b^t_{\hat{j}^t}) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



$$\max_{j^1,\ldots,j^T} \max_{k^1,\ldots,k^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t) + \sum_{t=1}^{T} h(k^t, b_{j^t}^t) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



$$\max_{j^1,\ldots,j^T} \max_{k^1,\ldots,k^T} \sum_{t=1}^{T} f(b_{j^t}^t) + \sum_{t=2}^{T} g(b_{j^{t-1}}^{t-1}, b_{j^t}^t) + \sum_{t=1}^{T} h(k^t, b_{j^t}^t) + \sum_{t=2}^{T} a(k^{t-1}, k^t)$$

Tracking in the context of event recognition in action

tracking

tracking and event recognition

Tracking in the context of event recognition in action

tracking

tracking and event recognition

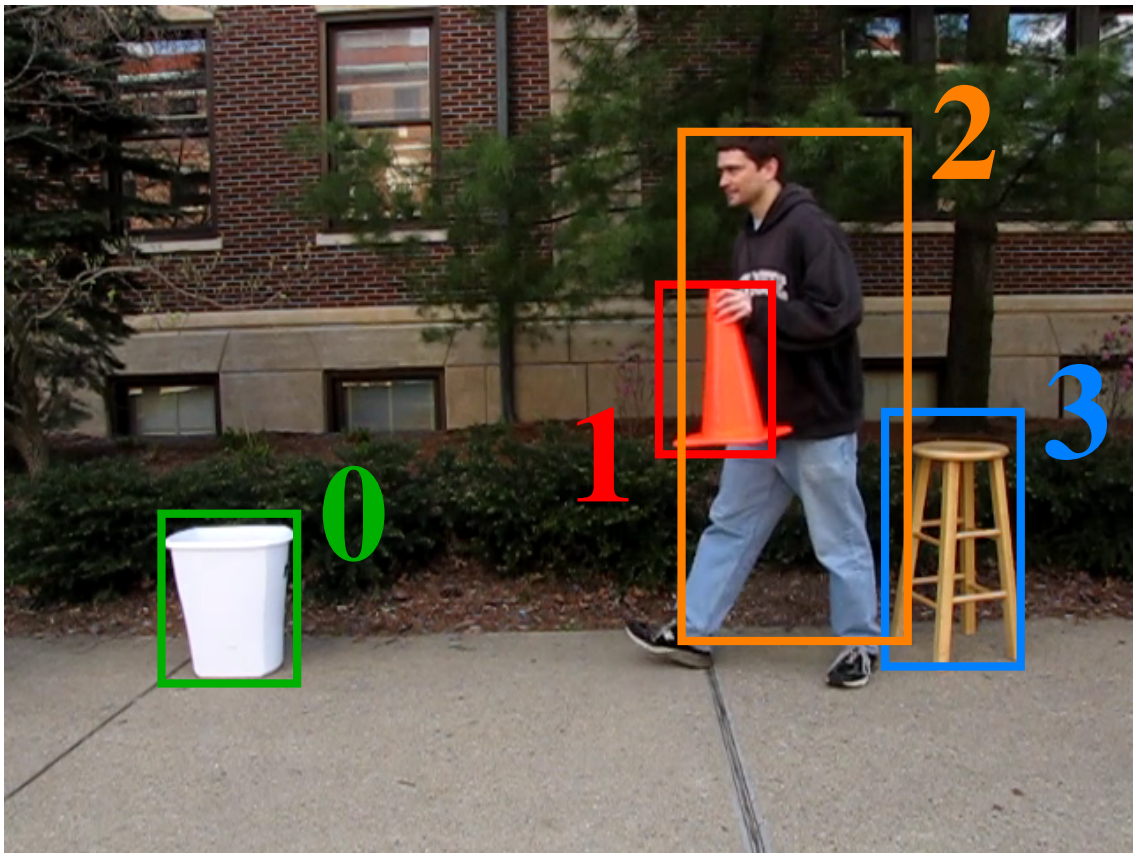
Sentence Tracker

The person to the left of the stool carried the traffic-cone towards the trash-can.

Sentence Tracker

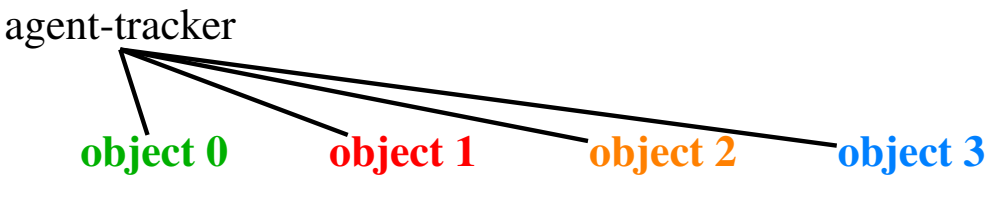
The person to the left of the stool carried the traffic-cone towards the trash-can.

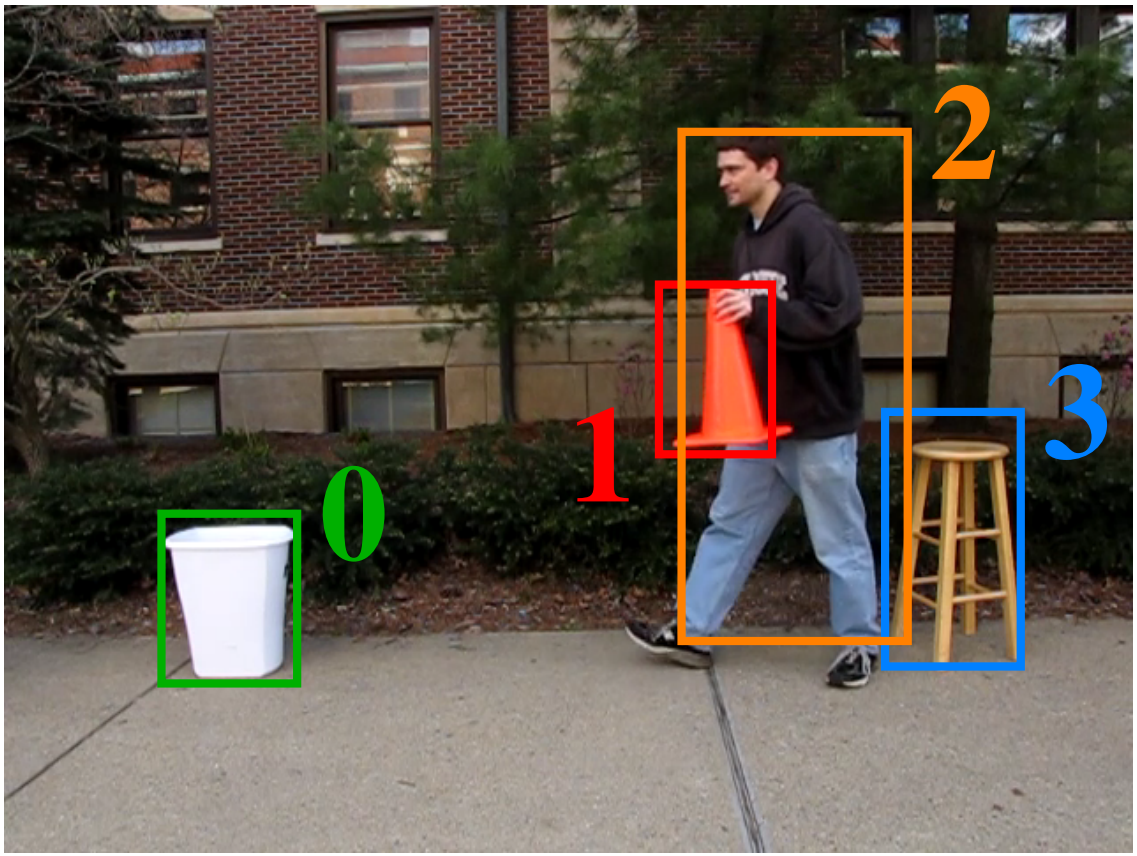
object 0 **object 1** **object 2** **object 3**



Sentence Tracker

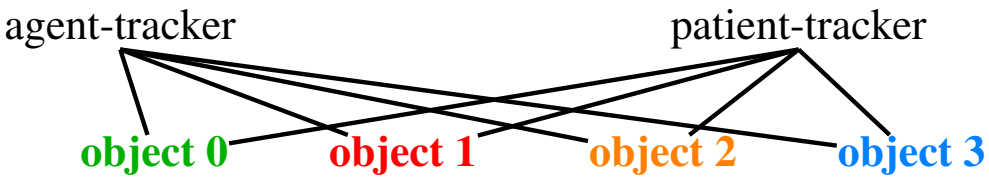
The person to the left of the stool carried the traffic-cone towards the trash-can.

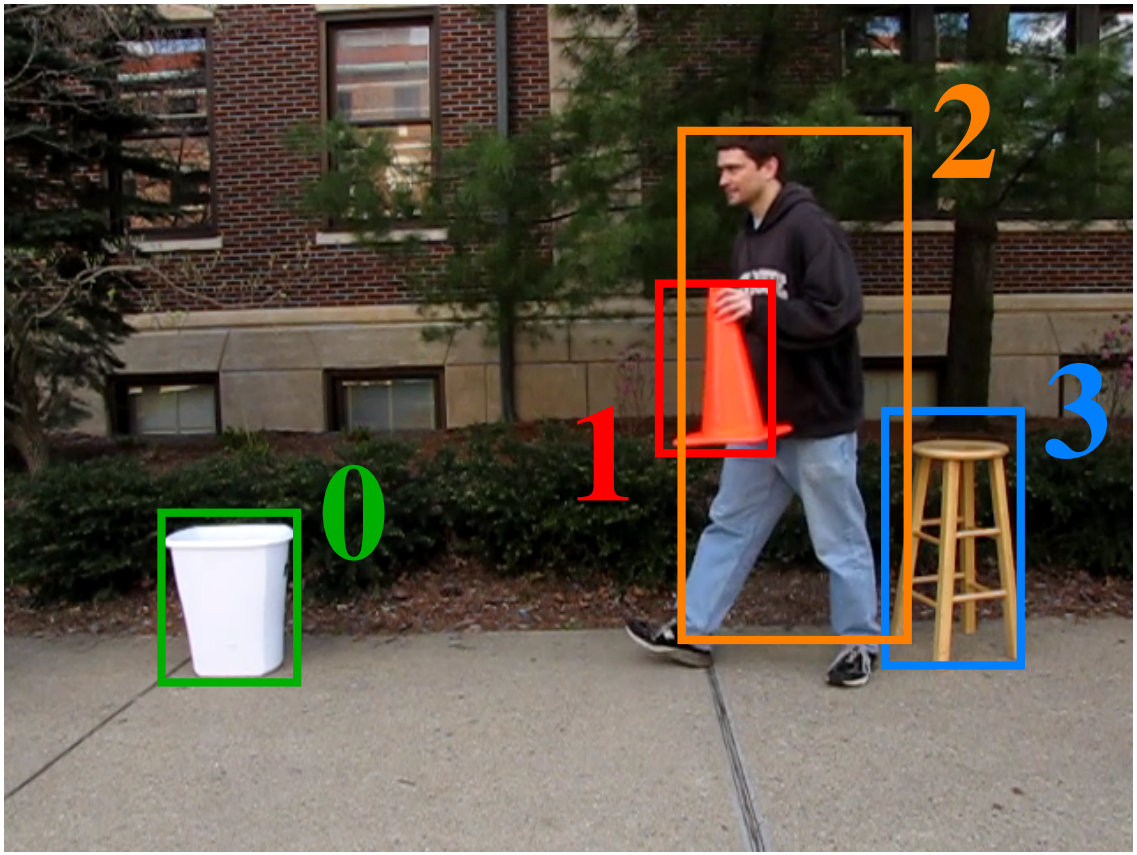




Sentence Tracker

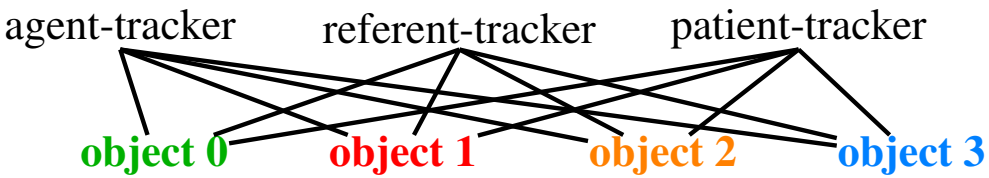
The person to the left of the stool carried the traffic-cone towards the trash-can.

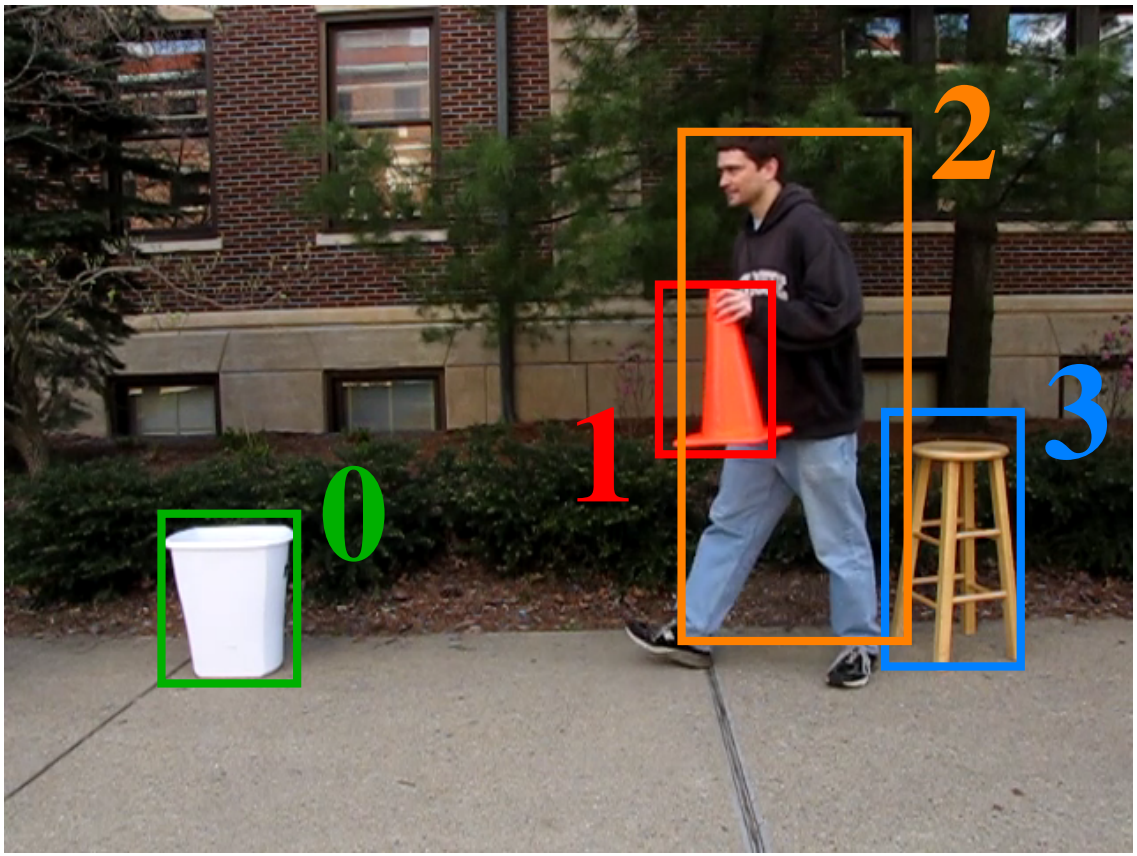




Sentence Tracker

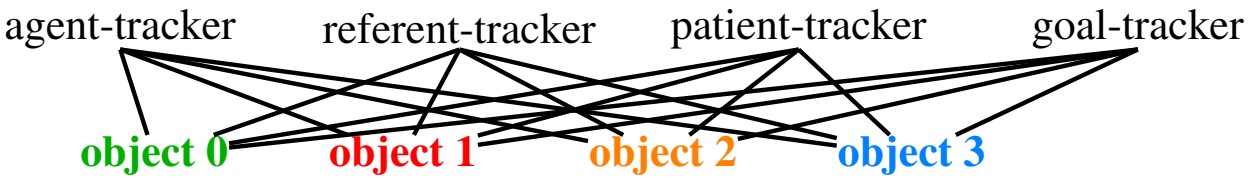
The person to the left of the stool carried the traffic-cone towards the trash-can.

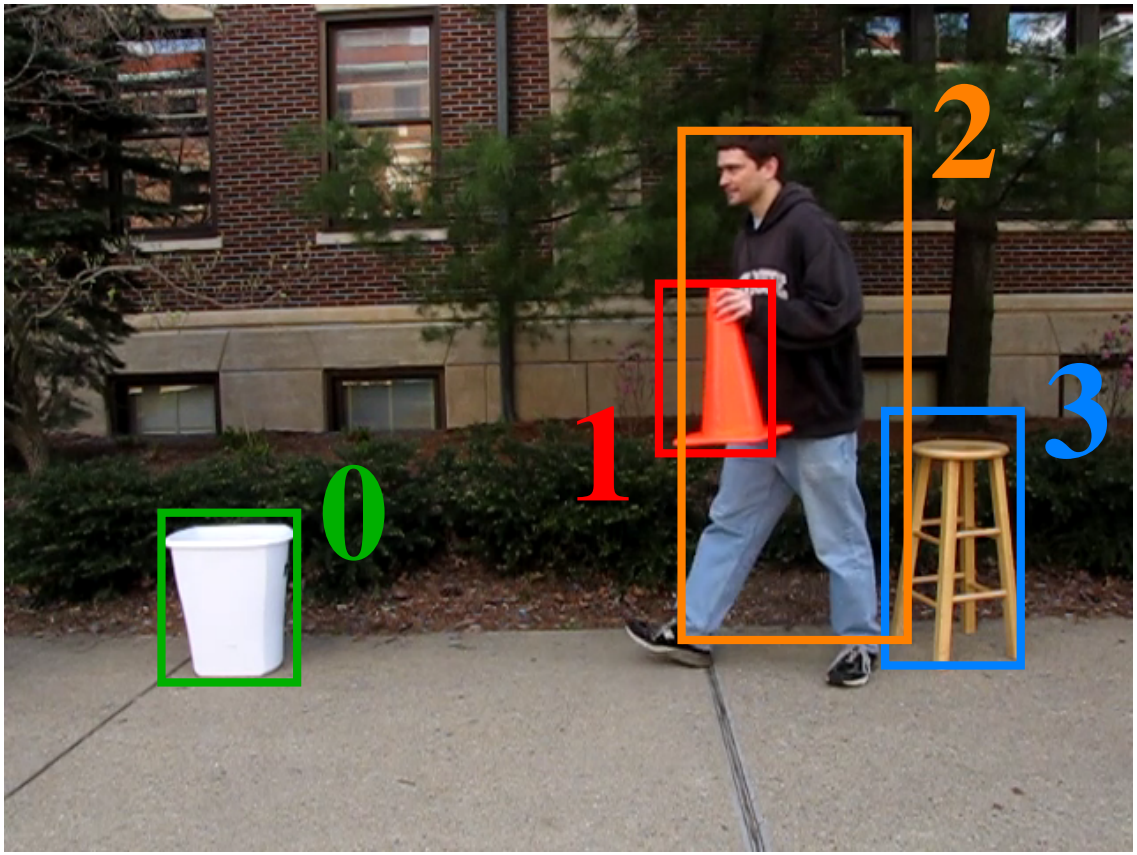




Sentence Tracker

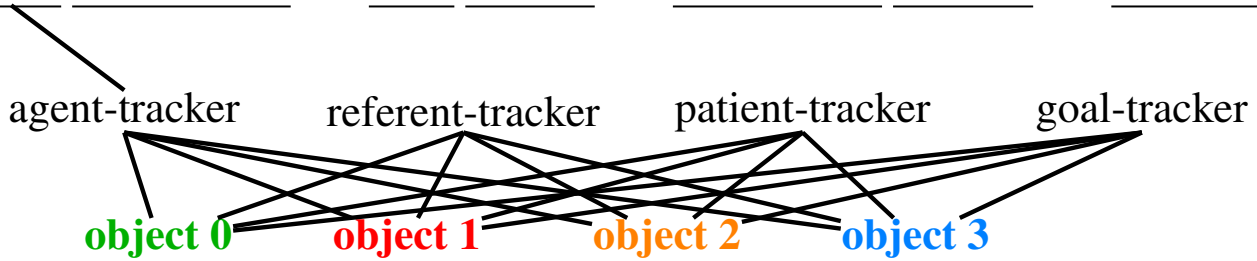
The person to the left of the stool carried the traffic-cone towards the trash-can.

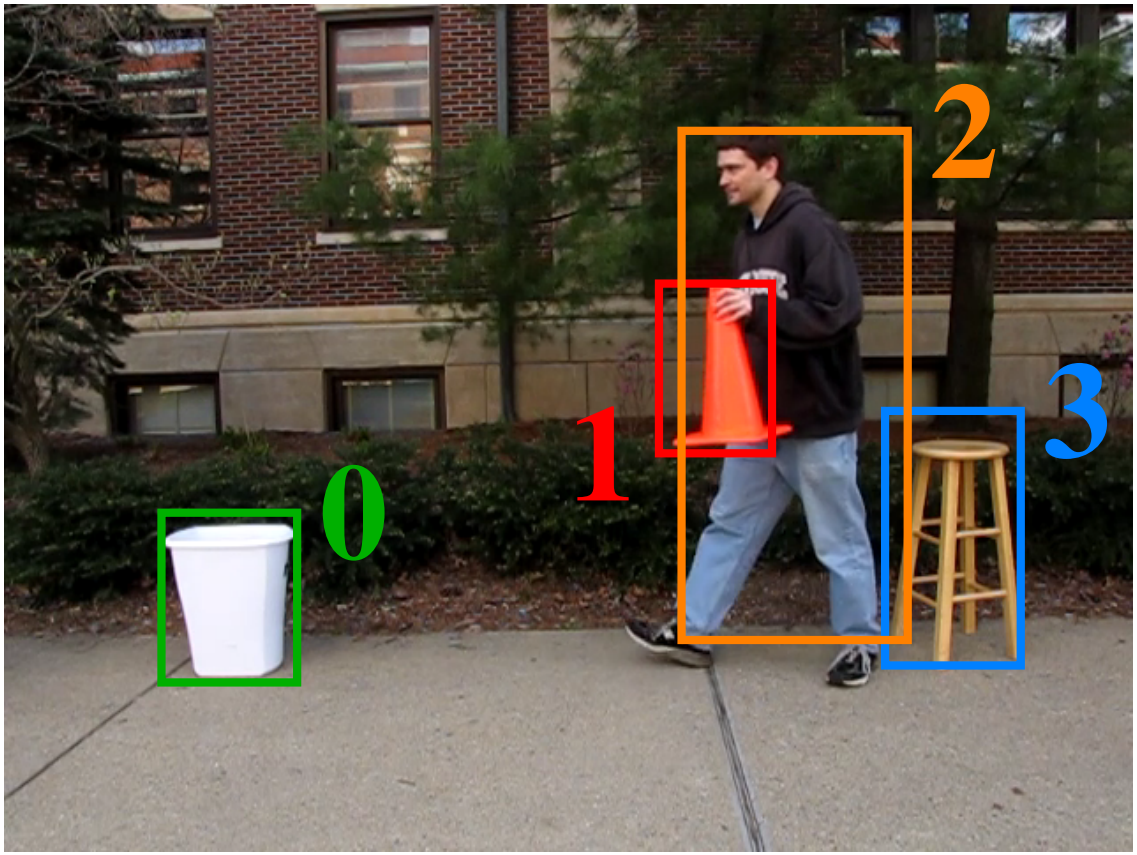




Sentence Tracker

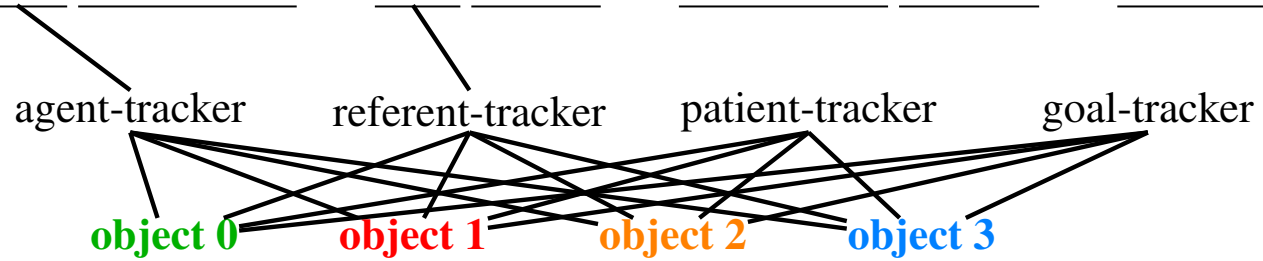
The person to the left of the stool carried the traffic-cone towards the trash-can.

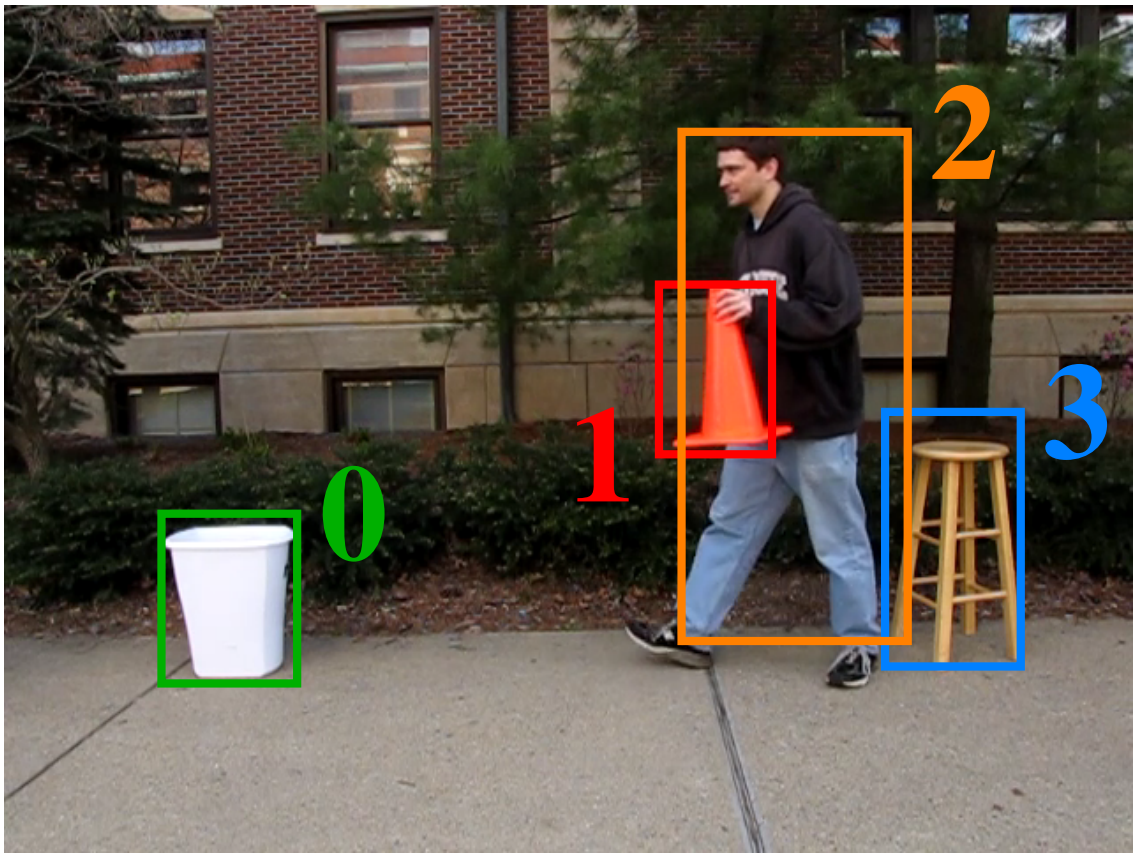




Sentence Tracker

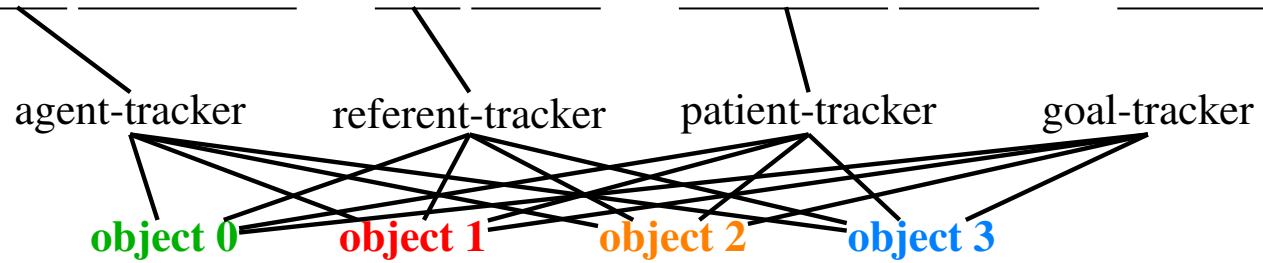
The person to the left of the stool carried the traffic-cone towards the trash-can.

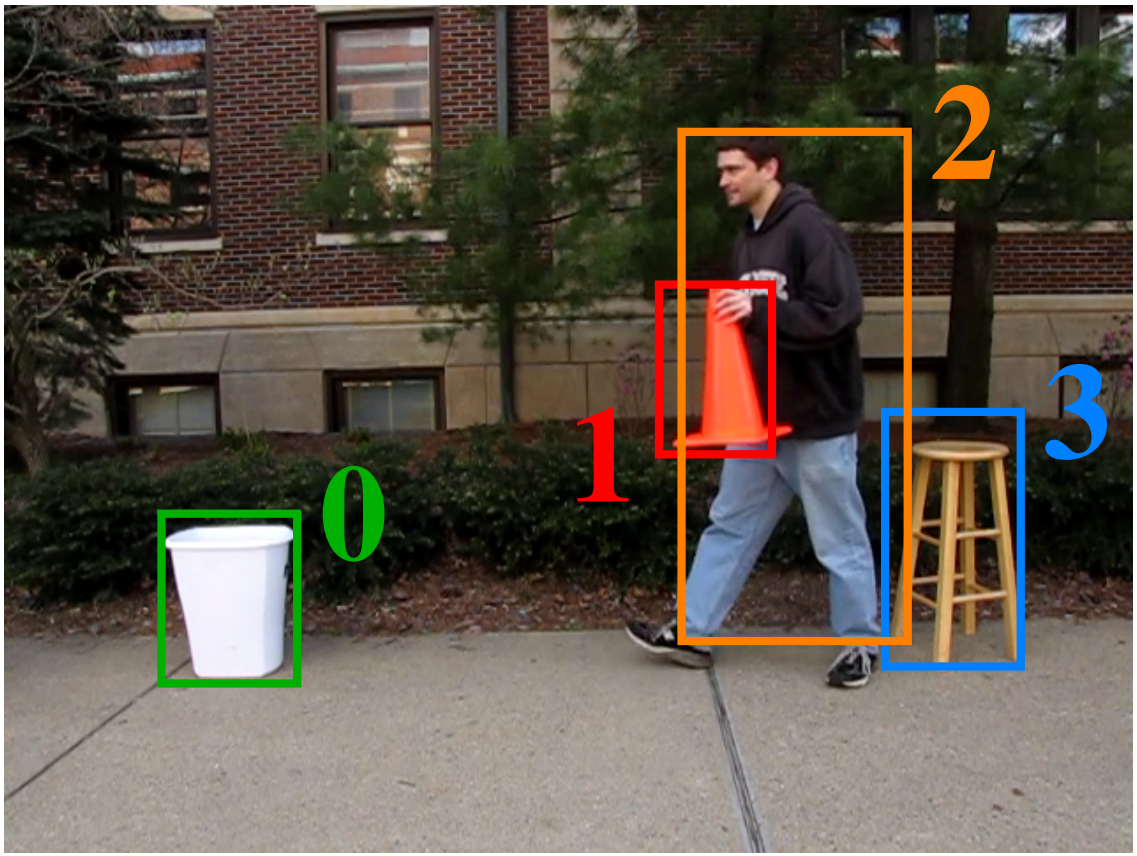




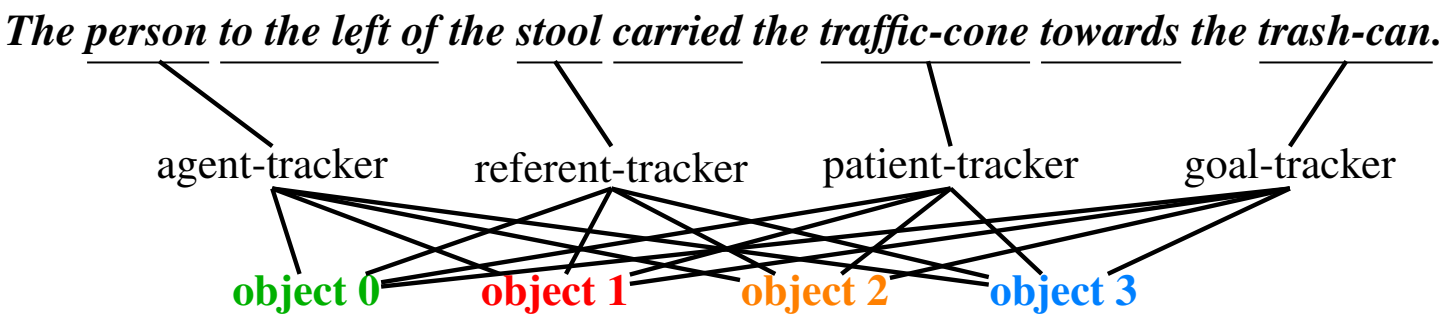
Sentence Tracker

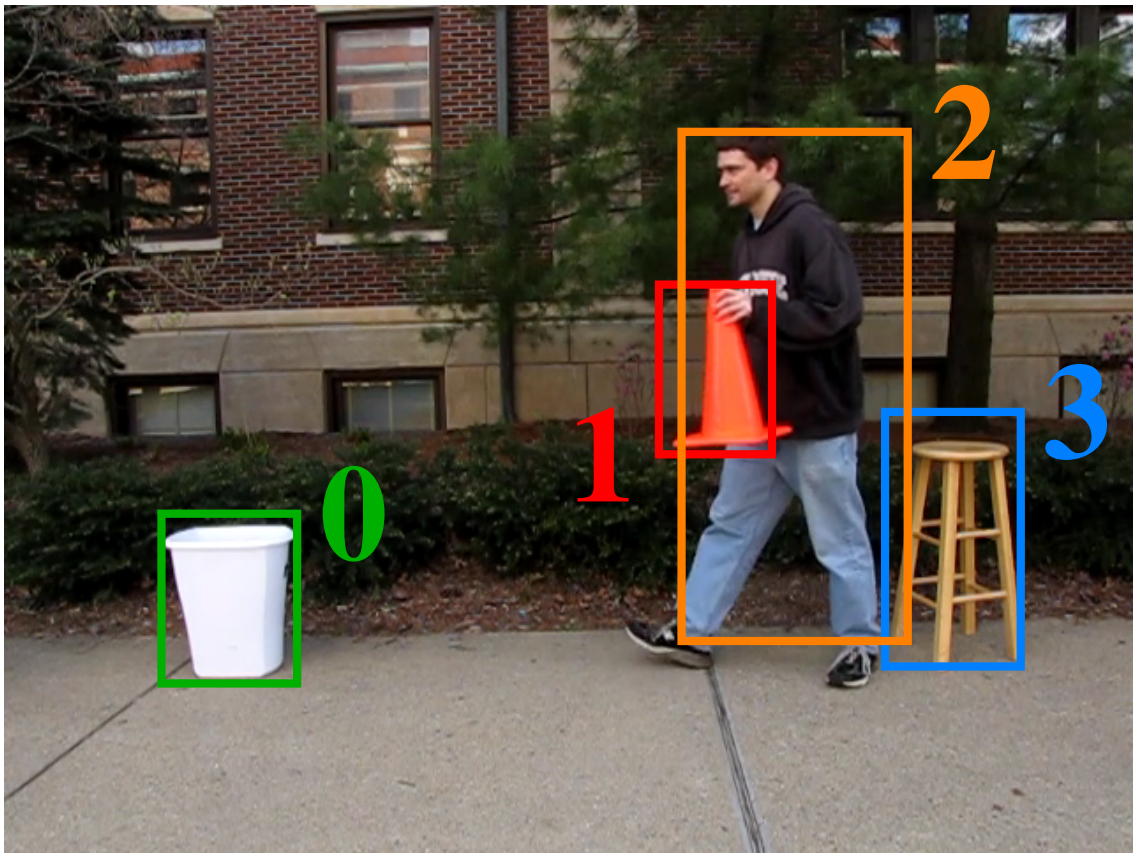
The person to the left of the stool carried the traffic-cone towards the trash-can.



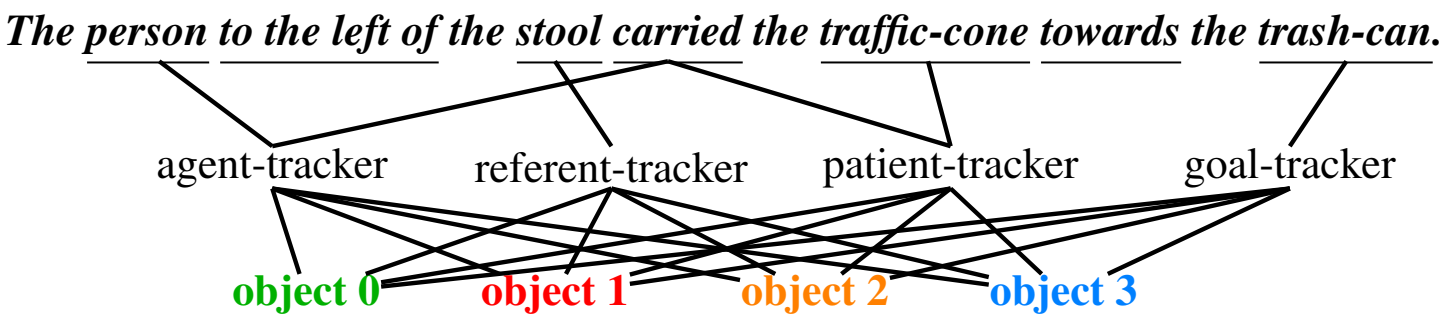


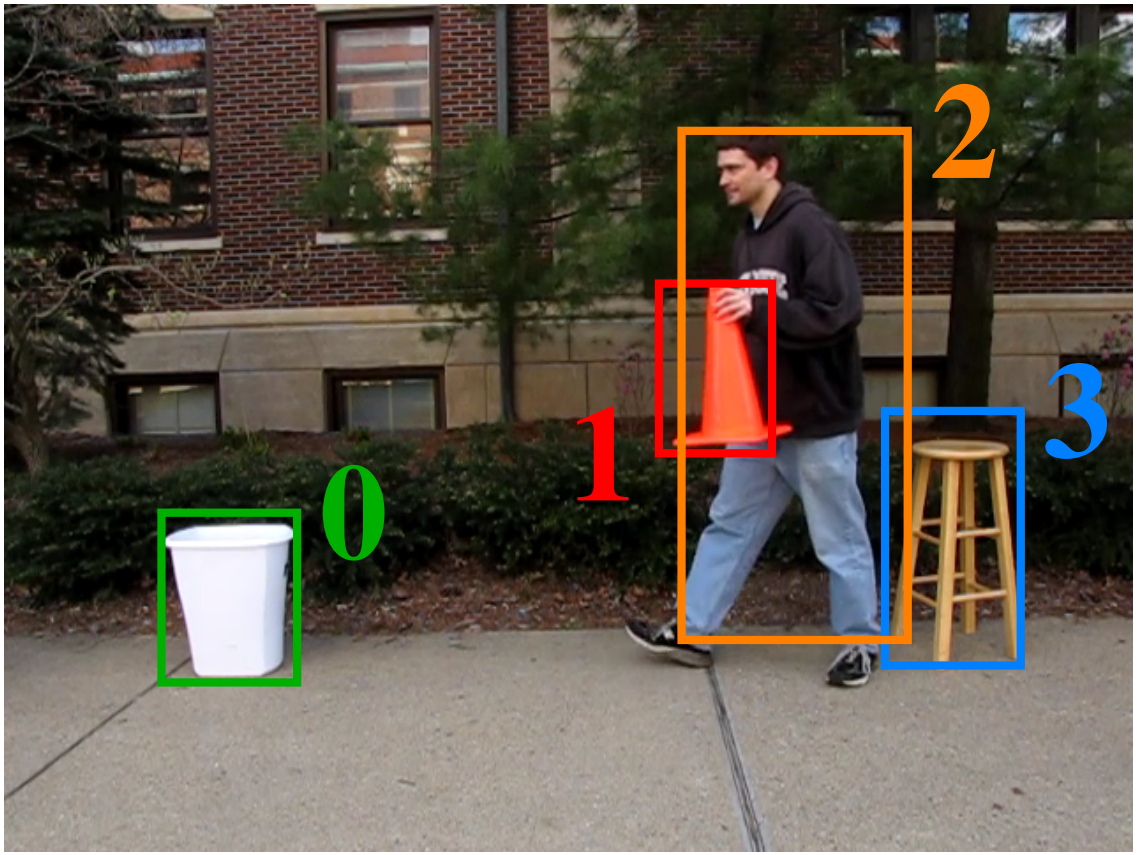
Sentence Tracker



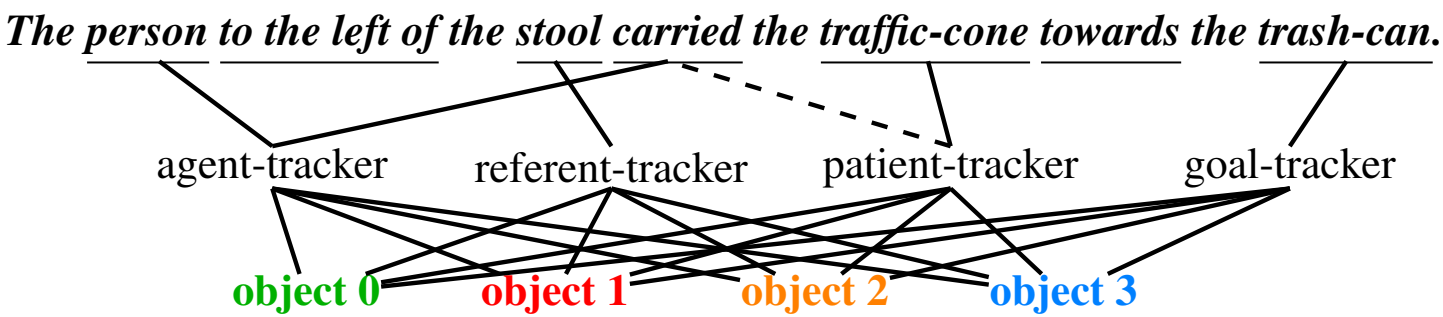


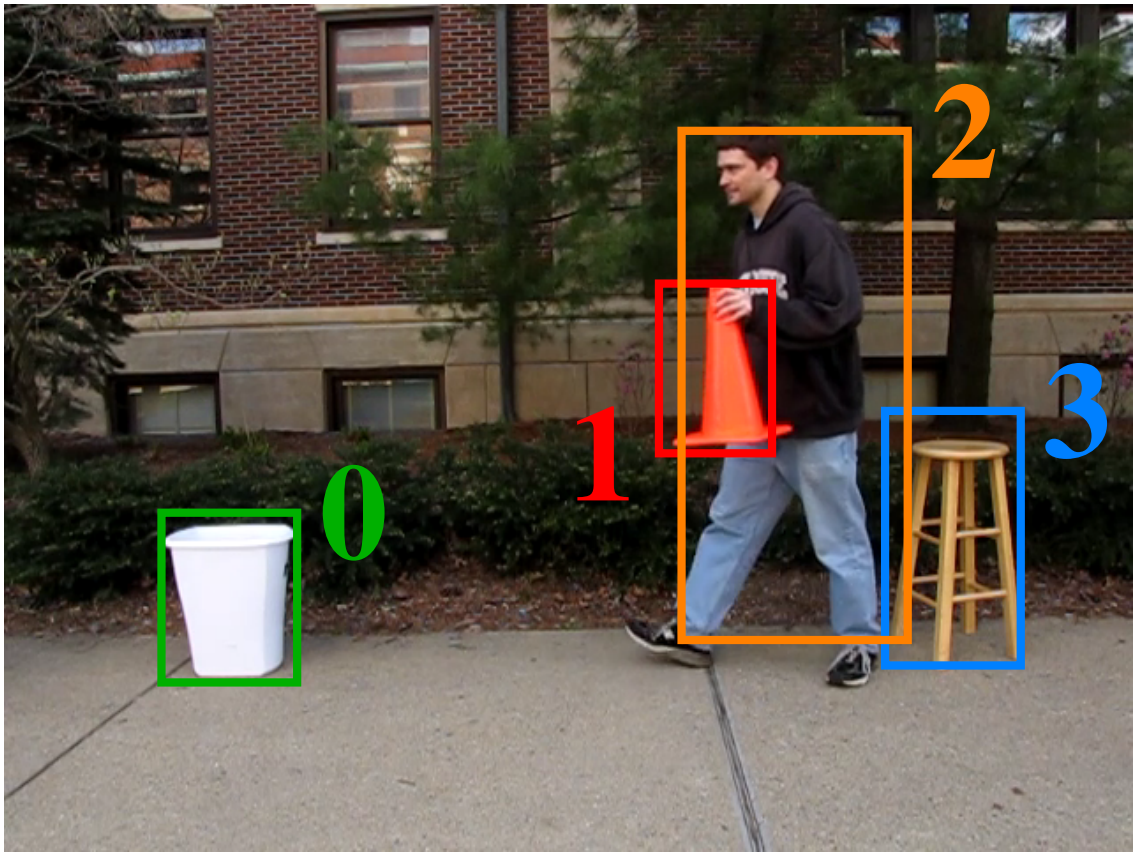
Sentence Tracker



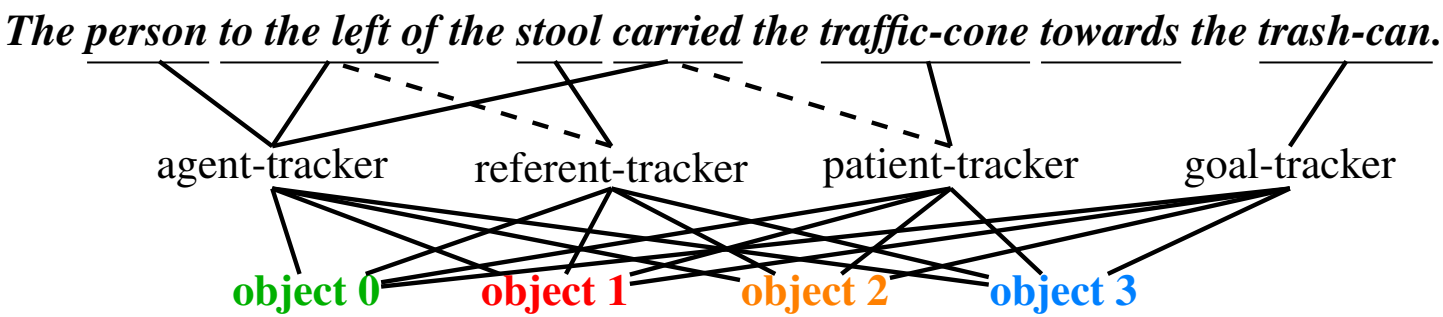


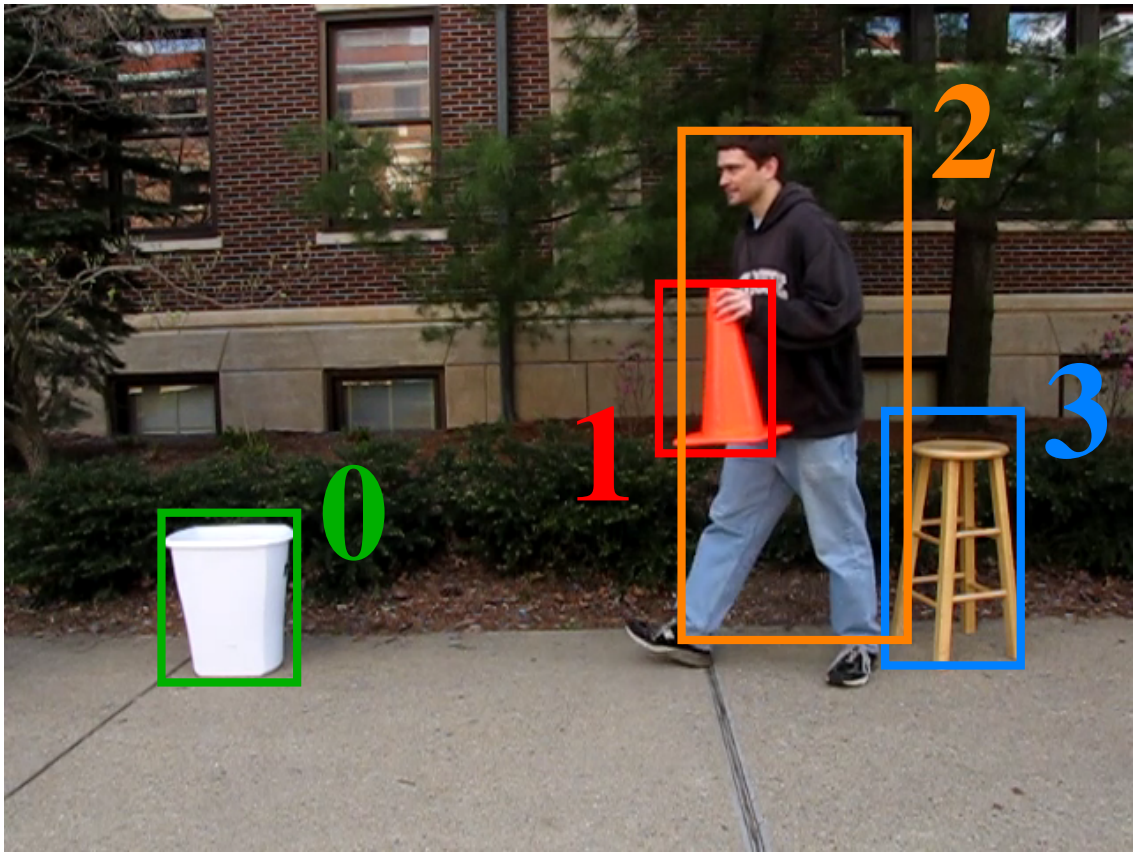
Sentence Tracker



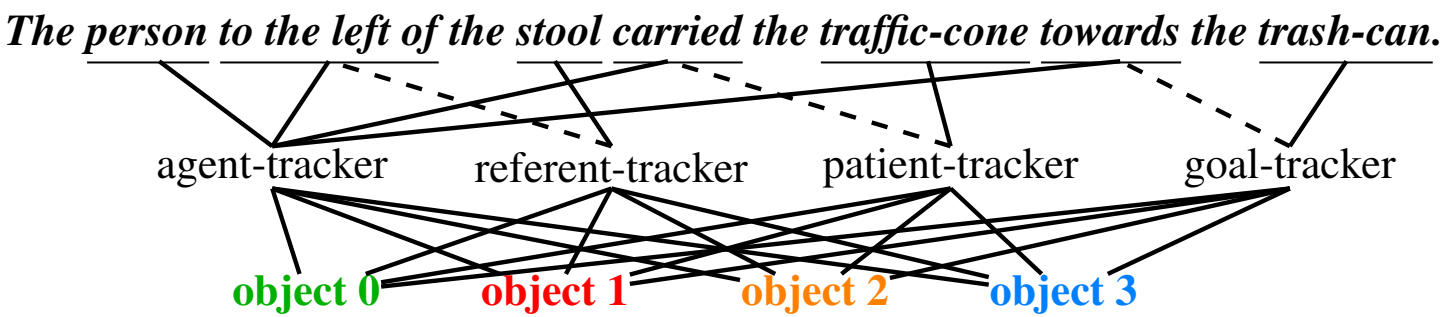


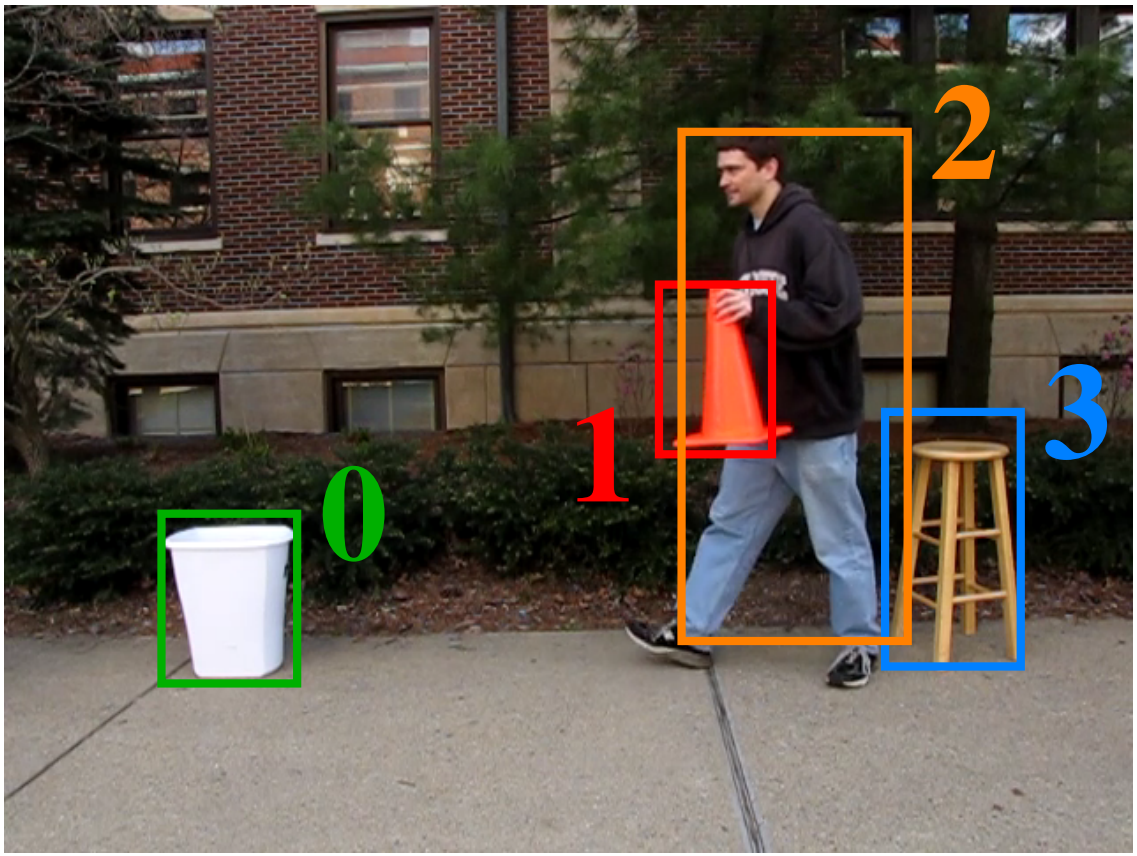
Sentence Tracker



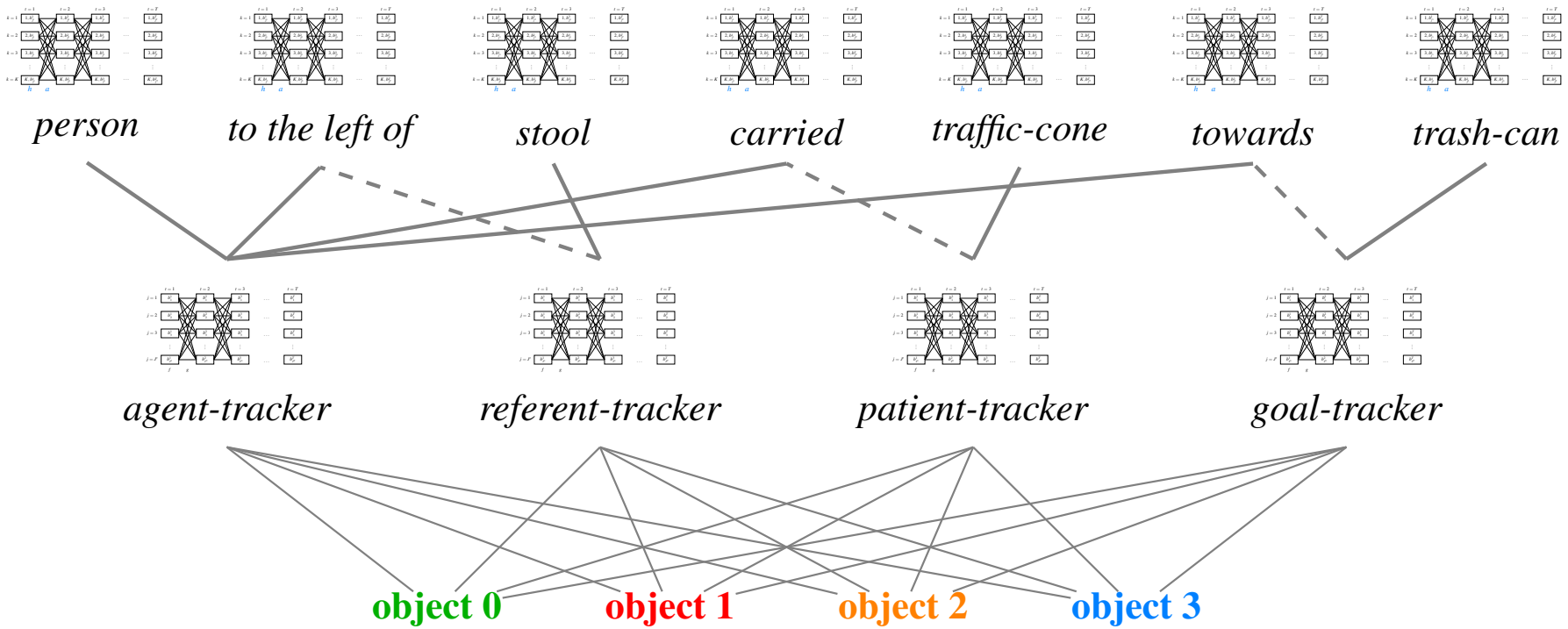


Sentence Tracker

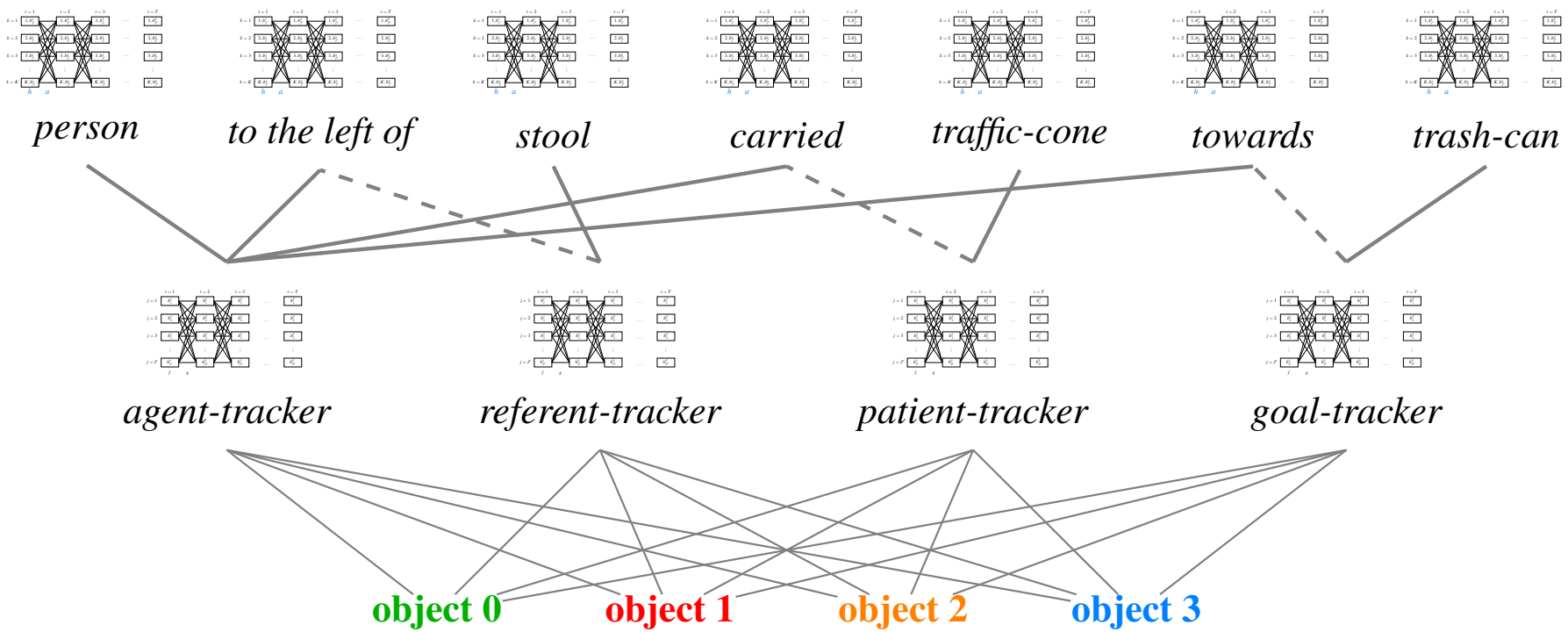




Sentence Tracker

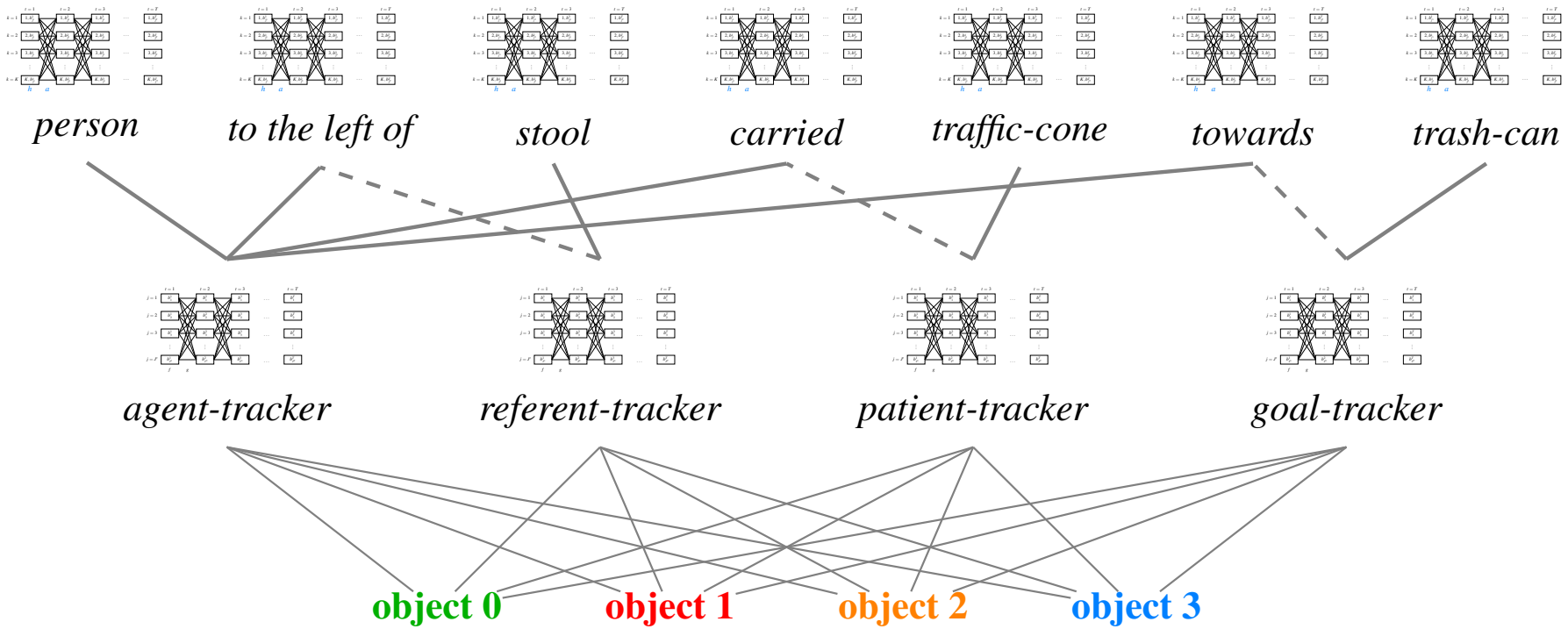


Sentence Tracker



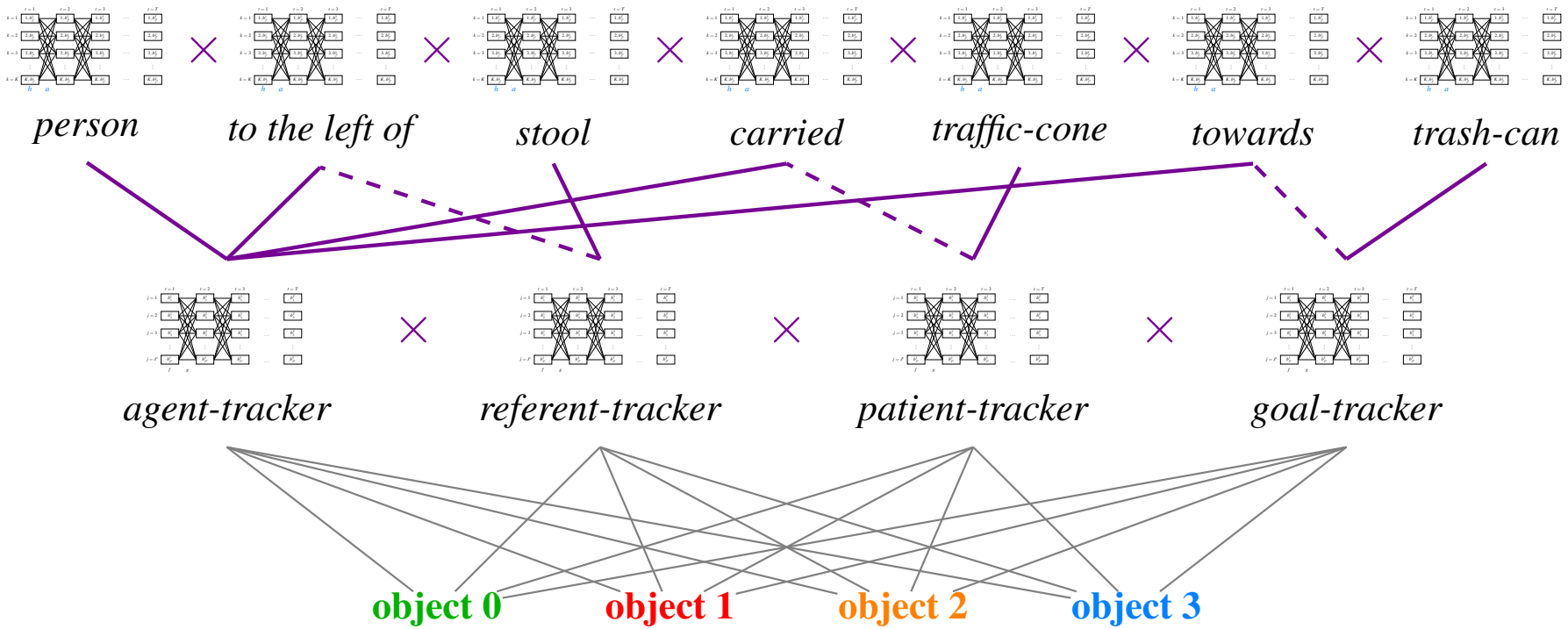
$$\max_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \left(\sum_{l=1}^{L} \sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_{l-1}^{t-1}}^{t-1}, b_{j_l^t}^t) \right) + \max_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \left(\sum_{w=1}^{W} \sum_{t=1}^{T} h_w(k_w^t, b_{w,1}^t, b_{w,2}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t) \right)$$

Sentence Tracker



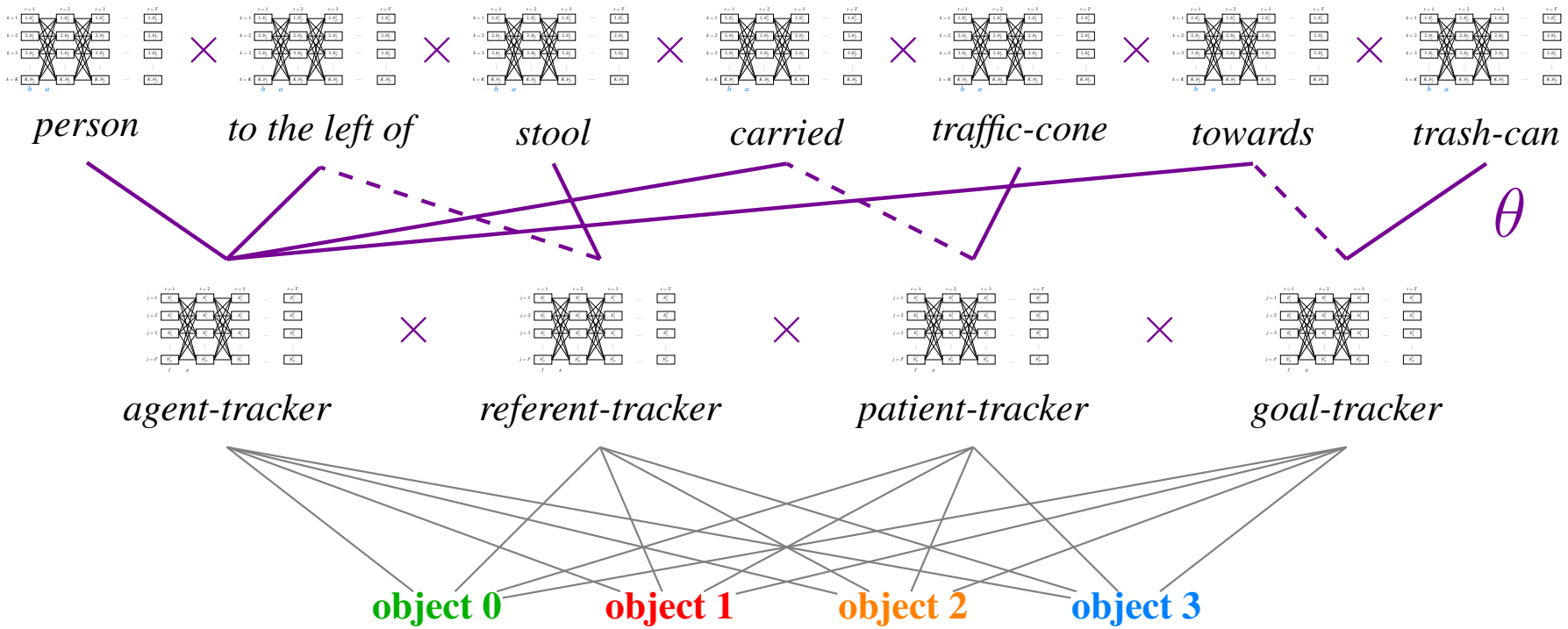
$$\max_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \max_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \sum_{l=1}^{L}\sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_{l-1}^{t-1}}^{t-1}, b_{j_l^t}^t) + \sum_{w=1}^{W}\sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t)$$

Sentence Tracker



$$\max_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \max_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \sum_{l=1}^{L}\sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_{l-1}^{t-1}}^{t-1}, b_{j_l^t}^t) + \sum_{w=1}^{W}\sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t)$$

Sentence Tracker



$$\max_{j_1^1,\ldots,j_L^T} \max_{k_1^1,\ldots,k_W^T} \sum_{l=1}^{L}\sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_{l-1}^{t-1}}^{t-1}, b_{j_l^t}^t) + \sum_{w=1}^{W}\sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t)$$

Unified Cost Function

$$\mathcal{S} : (\mathbf{B}, \mathbf{s}, \Lambda) \mapsto (\tau, \mathbf{J})$$

- $\mathbf{B}$: video
- $\mathbf{s}$: sentence
- Λ: lexcion
- τ: score
- $\mathbf{J}$: tracks

Four Uses of Unified Cost Function

Four Uses of Unified Cost Function

- **Focus of Attention**: video $\times$ sentence $\times$ lexicon $\rightarrow$ track collection

$$\begin{aligned} \mathbf{J}_1 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_1, \Lambda) \\ \mathbf{J}_2 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_2, \Lambda) \end{aligned}$$

Four Uses of Unified Cost Function

- **Focus of Attention**: video $\times$ sentence $\times$ lexicon $\rightarrow$ track collection

$$\begin{aligned} \mathbf{J}_1 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_1, \Lambda) \\ \mathbf{J}_2 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_2, \Lambda) \end{aligned}$$

- **Generation**: video $\times$ lexicon $\rightarrow$ sentence

$$\arg \max_{\mathbf{s}} \mathcal{S}_{\tau}(\mathbf{B}, \mathbf{s}, \Lambda)$$

Four Uses of Unified Cost Function

- **Focus of Attention**: video $\times$ sentence $\times$ lexicon $\rightarrow$ track collection

$$\begin{aligned} \mathbf{J}_1 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_1, \Lambda) \\ \mathbf{J}_2 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_2, \Lambda) \end{aligned}$$

- **Generation**: video $\times$ lexicon $\rightarrow$ sentence

$$\arg \max_{\mathbf{s}} \mathcal{S}_{\tau}(\mathbf{B}, \mathbf{s}, \Lambda)$$

- **Retrieval**: sentence $\times$ lexicon $\rightarrow$ video

$$\arg \max_{i} \mathcal{S}_{\tau}(\mathbf{B}_i, \mathbf{s}, \Lambda)$$

Four Uses of Unified Cost Function

- **Focus of Attention**: video $\times$ sentence $\times$ lexicon $\rightarrow$ track collection

$$\begin{aligned} \mathbf{J}_1 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_1, \Lambda) \\ \mathbf{J}_2 &= \mathcal{S}_{\mathbf{J}}(\mathbf{B}, \mathbf{s}_2, \Lambda) \end{aligned}$$

- **Generation**: video $\times$ lexicon $\rightarrow$ sentence

$$\arg \max_{\mathbf{s}} \mathcal{S}_{\tau}(\mathbf{B}, \mathbf{s}, \Lambda)$$

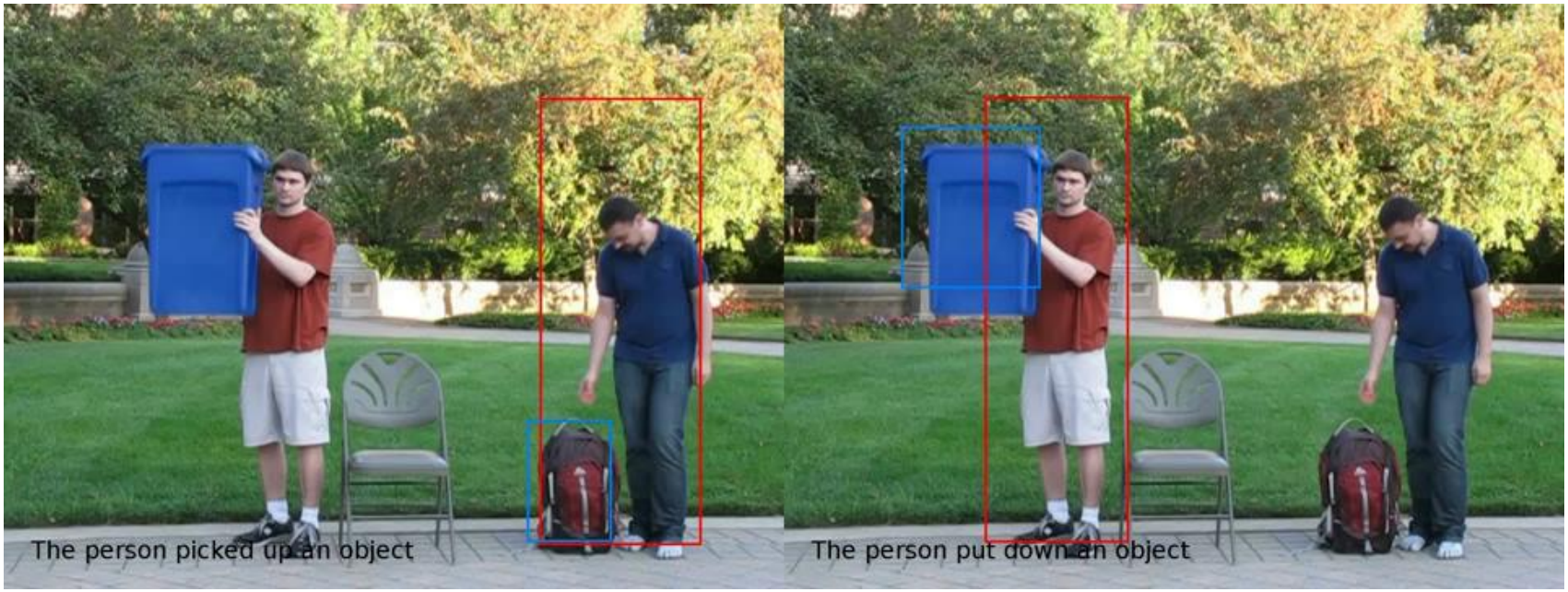
- **Retrieval**: sentence $\times$ lexicon $\rightarrow$ video

$$\arg \max_{i} \mathcal{S}_{\tau}(\mathbf{B}_i, \mathbf{s}, \Lambda)$$

- **Acquisition**: video $\times$ sentence $\rightarrow$ lexicon

$$\arg \max_{\Lambda} \sum_{m=1}^{M} \mathcal{S}_{\tau}(\mathbf{B}_m, \mathbf{s}_m, \Lambda)$$

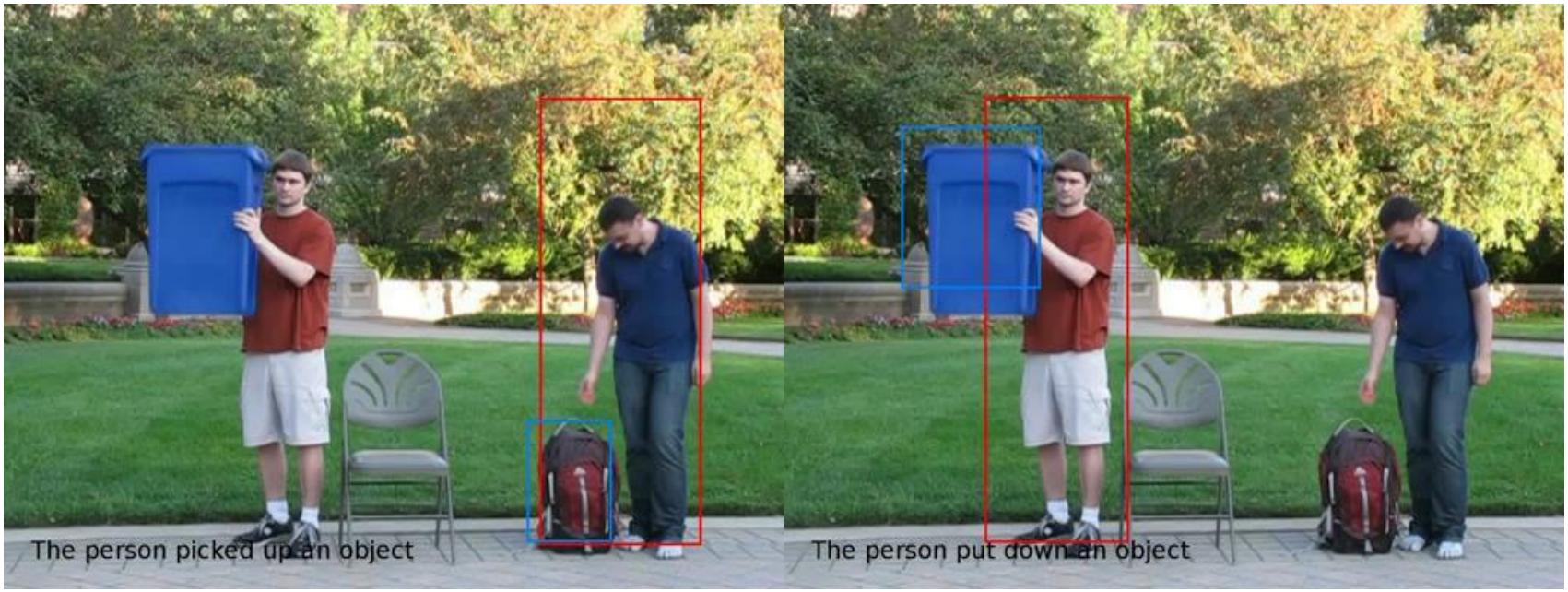
Using the Sentence Tracker to Focus Attention



We can differentiate events based on the verb:

The person picked up an object.
The person put down an object.

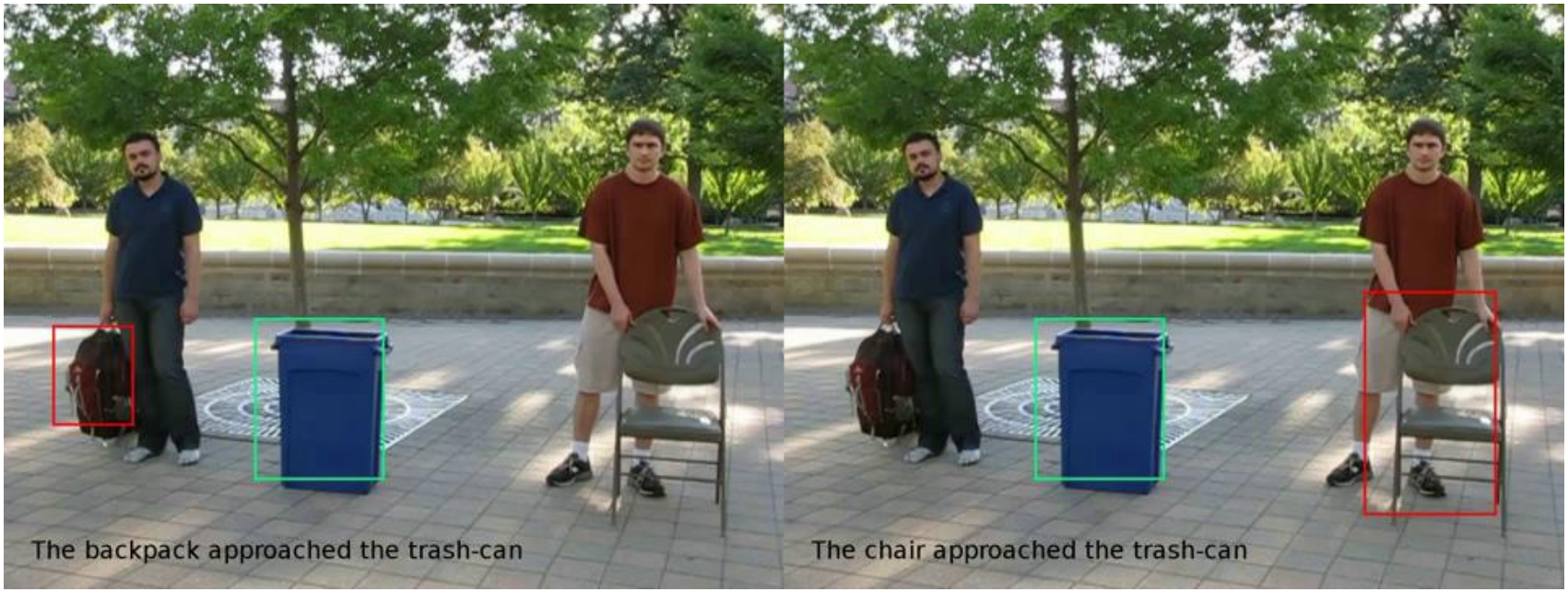
Using the Sentence Tracker to Focus Attention



We can differentiate events based on the verb:

The person picked up an object.
The person put down an object.

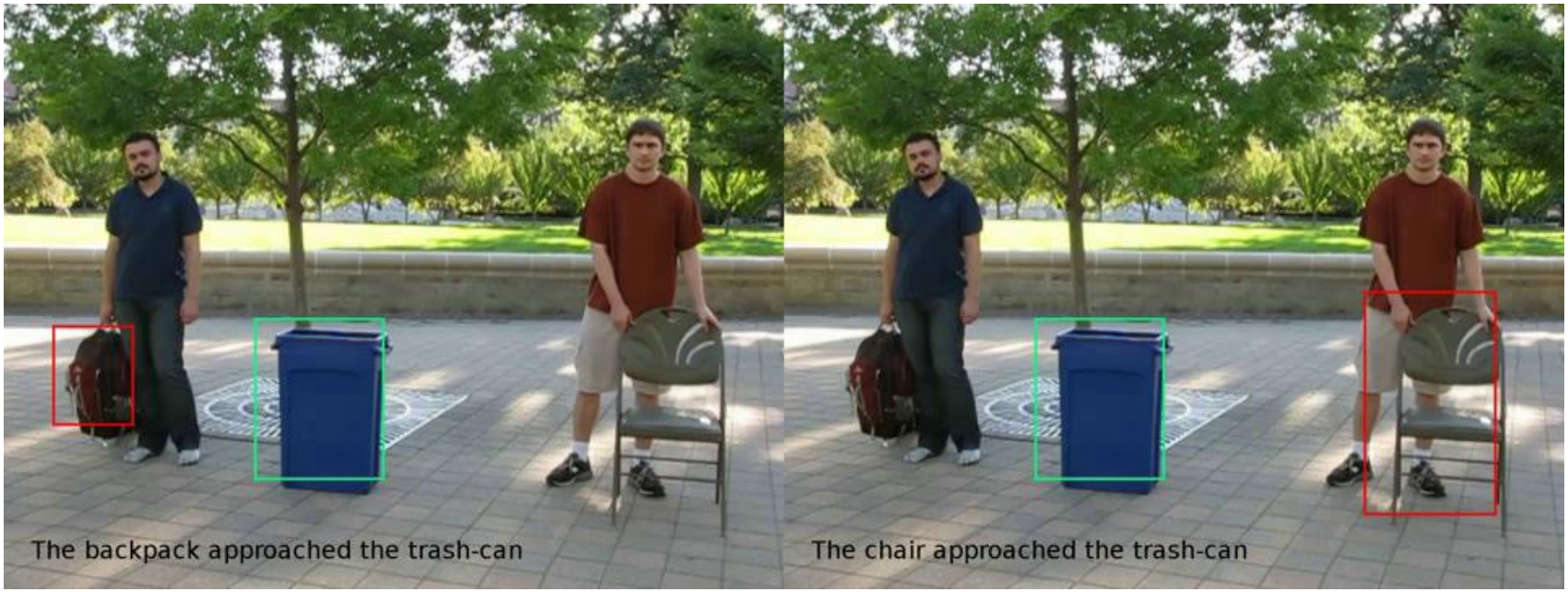
Using the Sentence Tracker to Focus Attention



We can differentiate events based on the subject noun:

The backpack approached the trash can.
The chair approached the trash can.

Using the Sentence Tracker to Focus Attention



We can differentiate events based on the subject noun:

The backpack approached the trash can.
The chair approached the trash can.

Using the Sentence Tracker to Focus Attention

We can differentiate events based on an adjective in the subject NP:

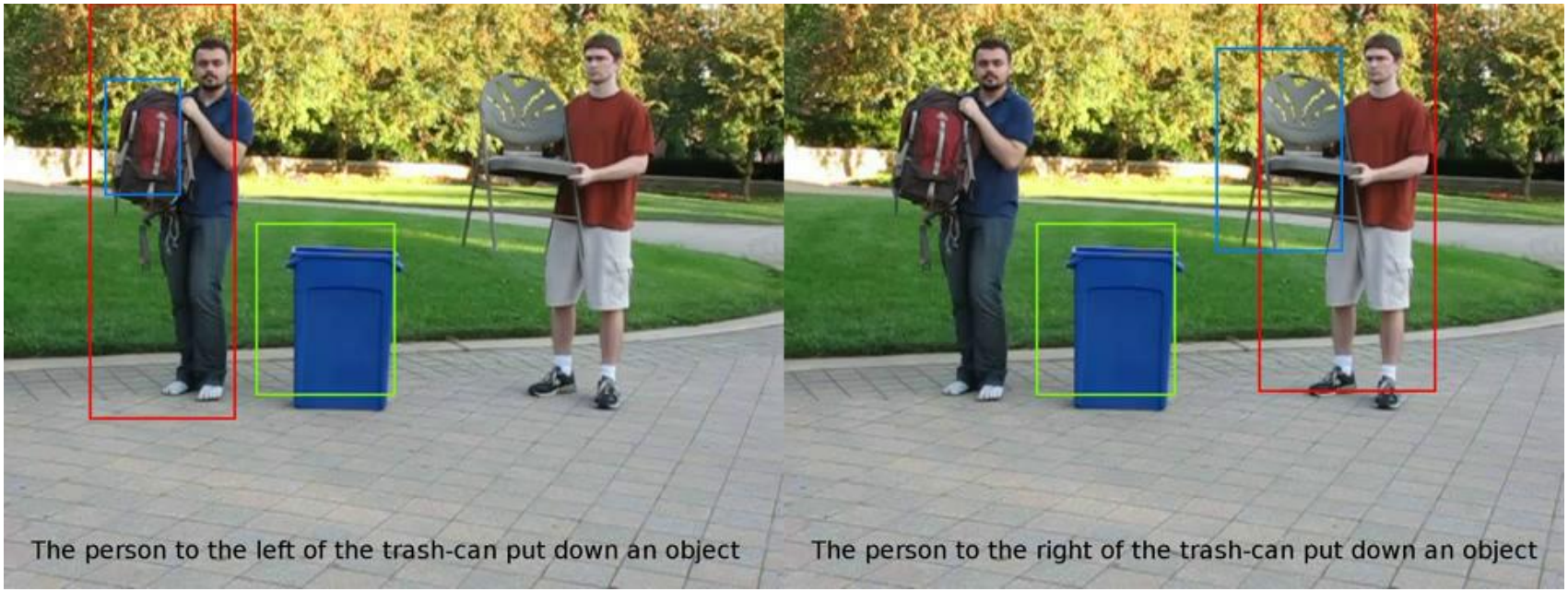
The red object approached the chair.
The blue object approached the chair.

Using the Sentence Tracker to Focus Attention

We can differentiate events based on an adjective in the subject NP:

The red object approached the chair.
The blue object approached the chair.

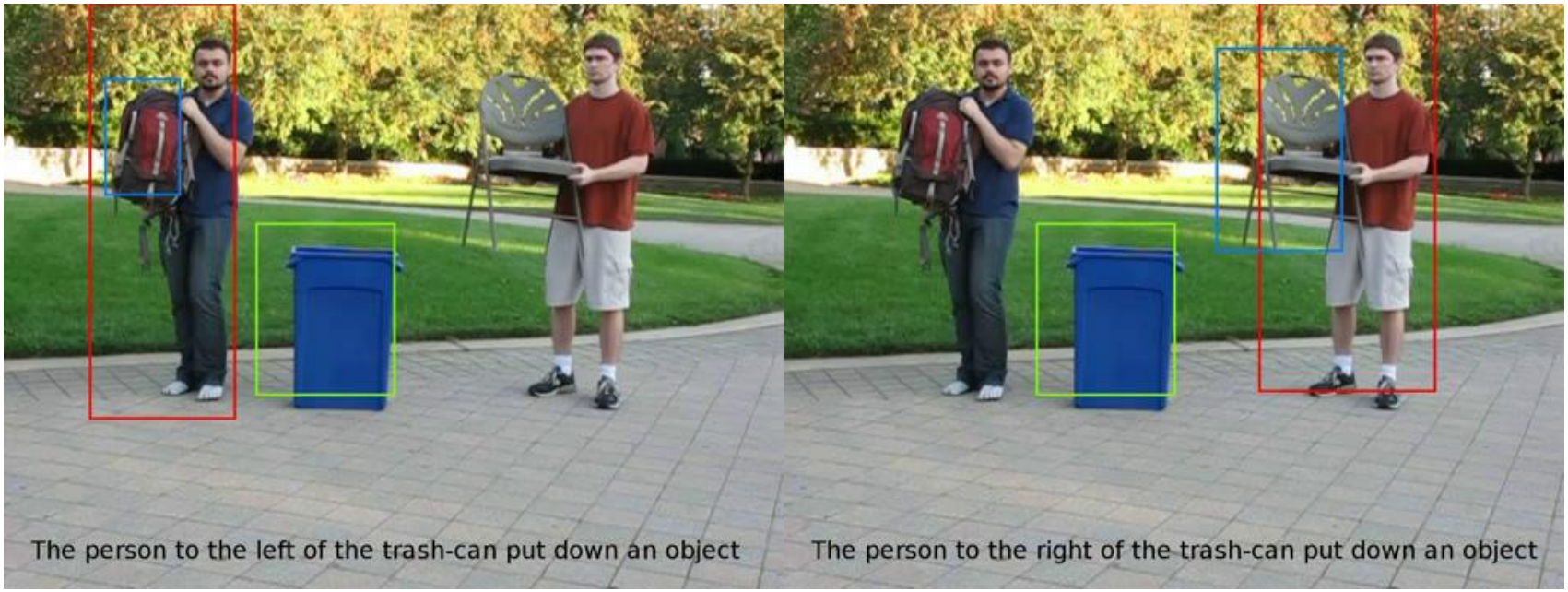
Using the Sentence Tracker to Focus Attention



We can differentiate events based on a preposition in the subject NP:

The person to the left of the trash can put down an object.
The person to the right of the trash can put down an object.

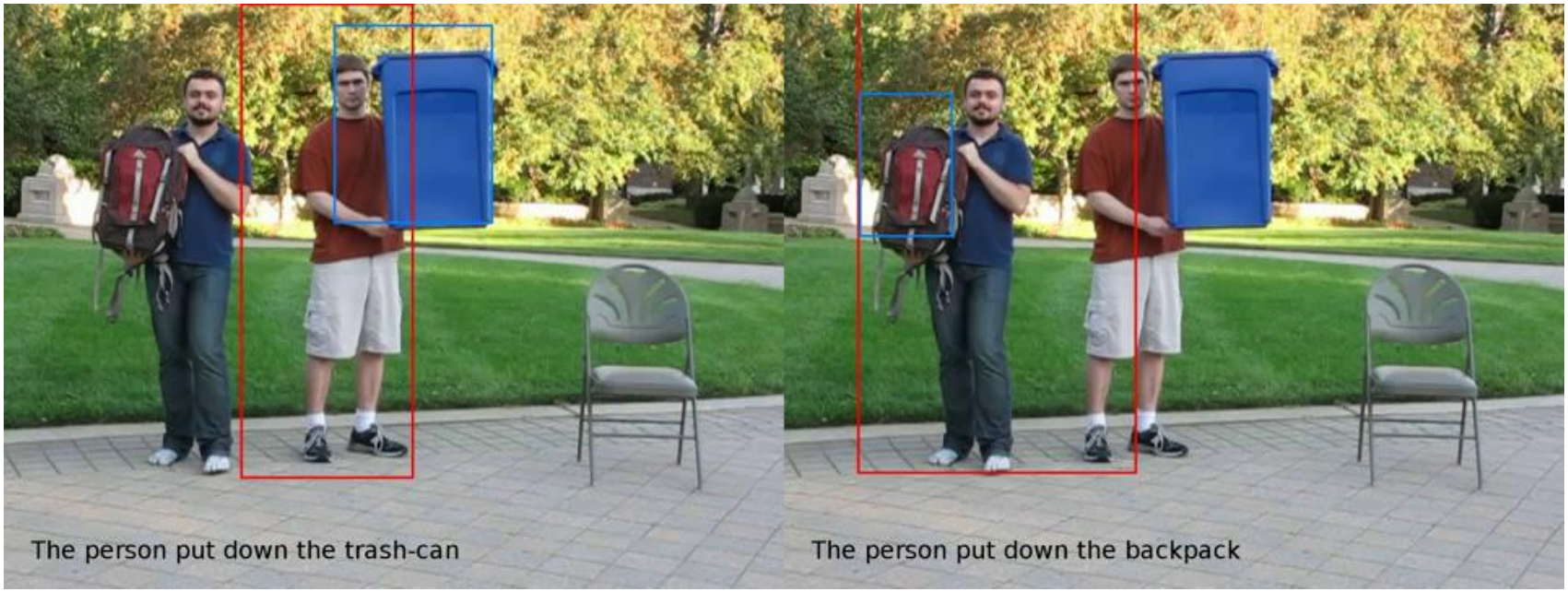
Using the Sentence Tracker to Focus Attention



We can differentiate events based on a preposition in the subject NP:

The person to the left of the trash can put down an object.
The person to the right of the trash can put down an object.

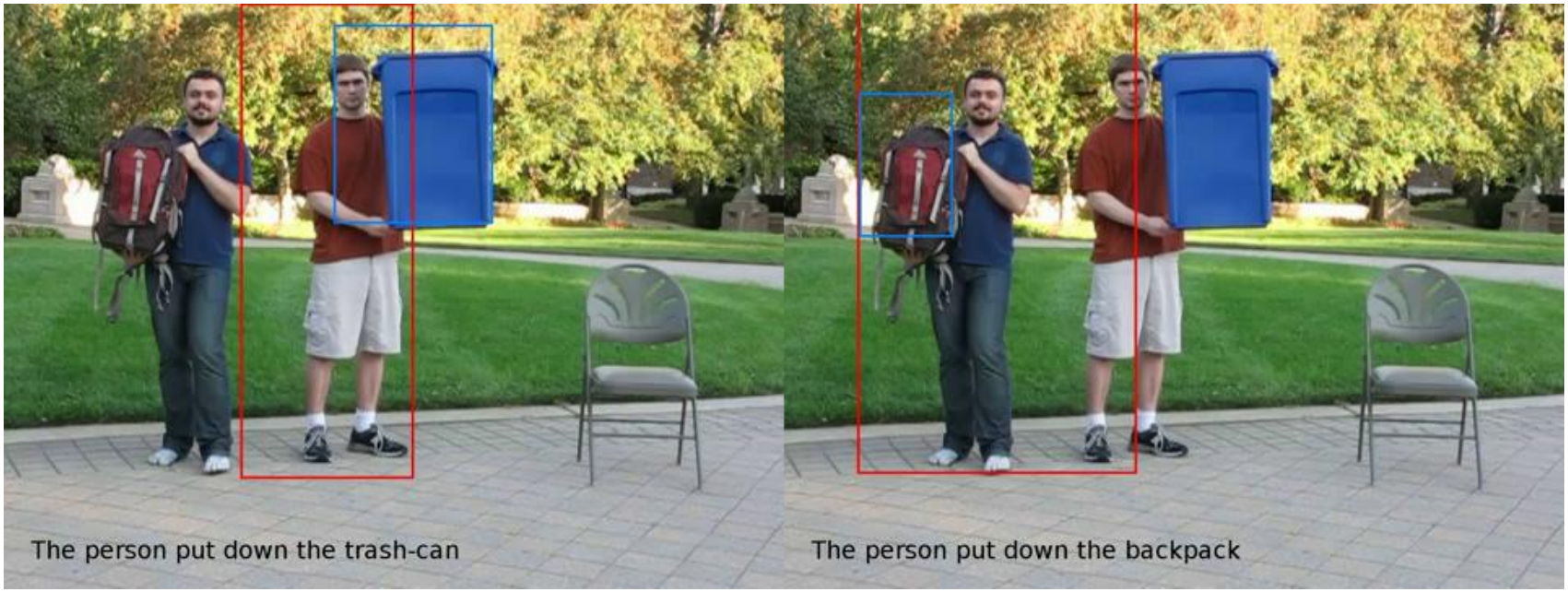
Using the Sentence Tracker to Focus Attention



We can differentiate events based on the object noun:

The person put down the trash can.
The person put down the backpack.

Using the Sentence Tracker to Focus Attention



We can differentiate events based on the object noun:

The person put down the trash can.
The person put down the backpack.

Using the Sentence Tracker to Focus Attention

We can differentiate events based on an adjective in the object NP:

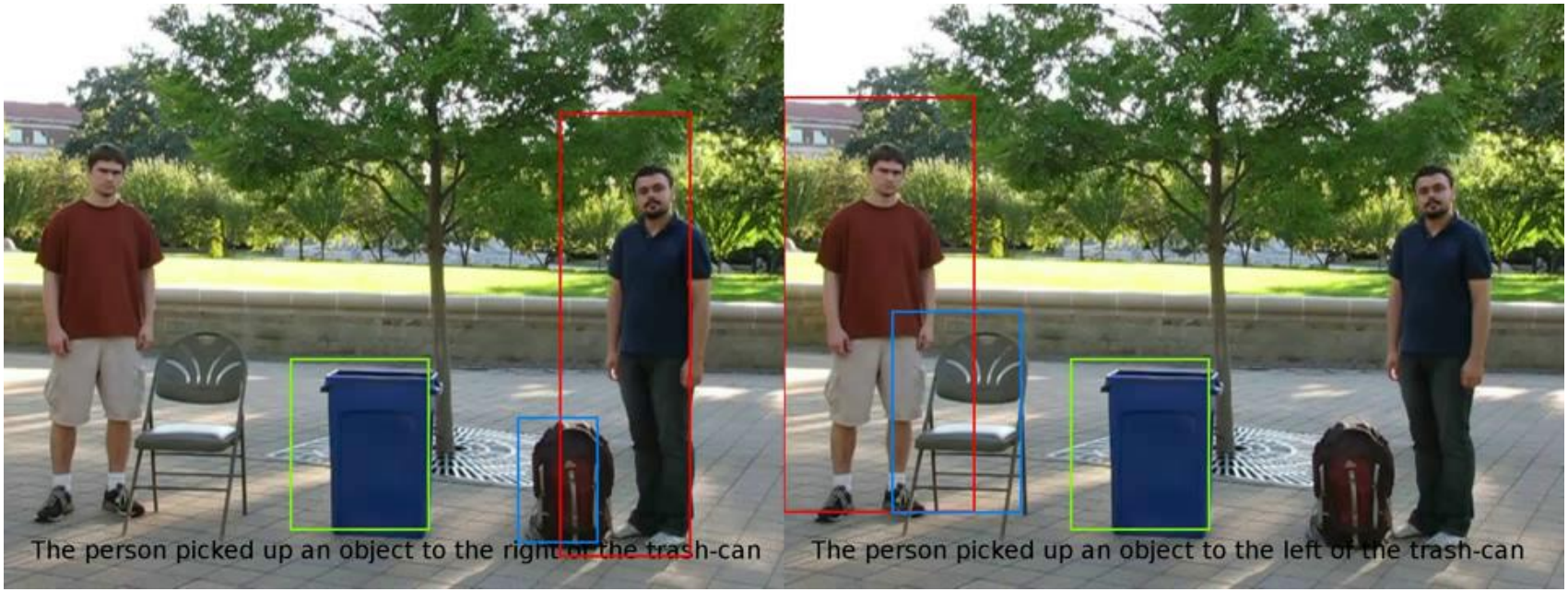
The person carried the blue object.
The person carried the red object.

Using the Sentence Tracker to Focus Attention

We can differentiate events based on an adjective in the object NP:

The person carried the blue object.
The person carried the red object.

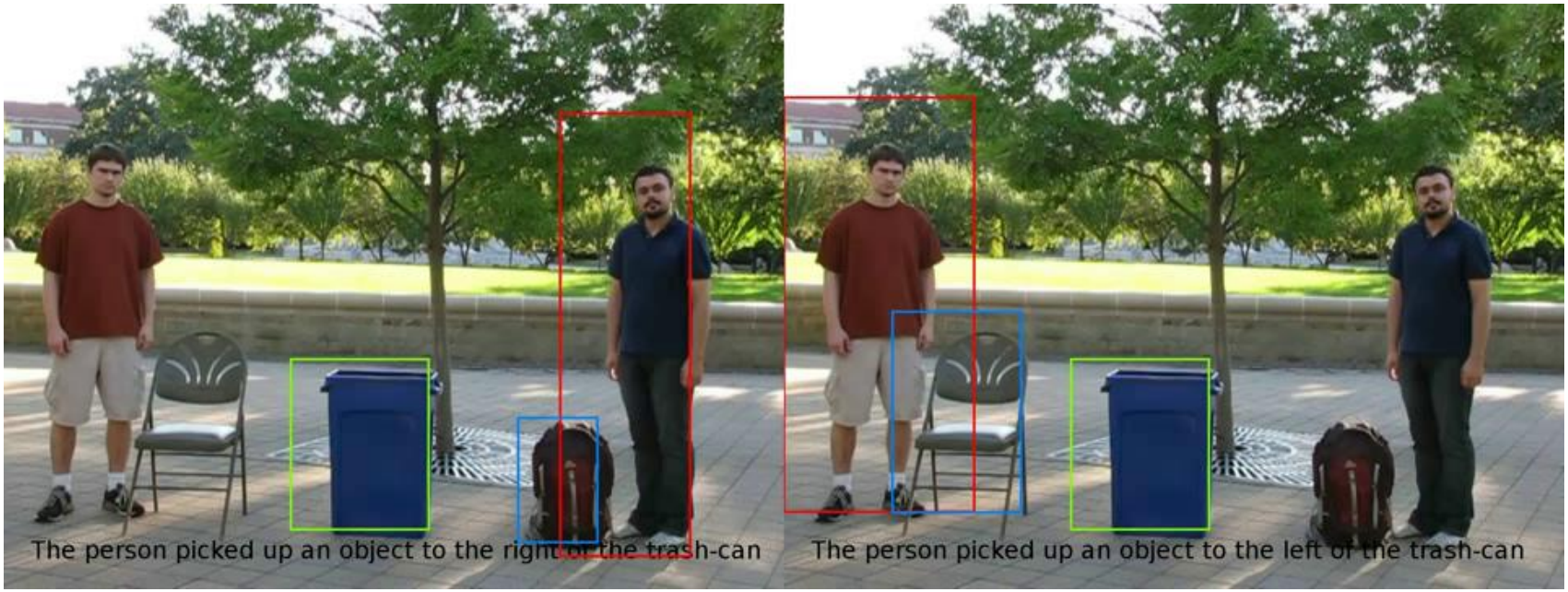
Using the Sentence Tracker to Focus Attention



We can differentiate events based on a preposition in the object NP:

The person picked up an object to the right of the trash can.
The person picked up an object to the left of the trash can.

Using the Sentence Tracker to Focus Attention



We can differentiate events based on a preposition in the object NP:

The person picked up an object to the right of the trash can.
The person picked up an object to the left of the trash can.

Using the Sentence Tracker to Focus Attention

We can differentiate events based on a preposition in an adjunct:

The person carried an object towards the trash can.
The person carried an object away from the trash can.

Using the Sentence Tracker to Focus Attention

We can differentiate events based on a preposition in an adjunct:

The person carried an object towards the trash can.
The person carried an object away from the trash can.

Using the Sentence Tracker to Perform Generation

Beam Search

Using the Sentence Tracker to Perform Generation

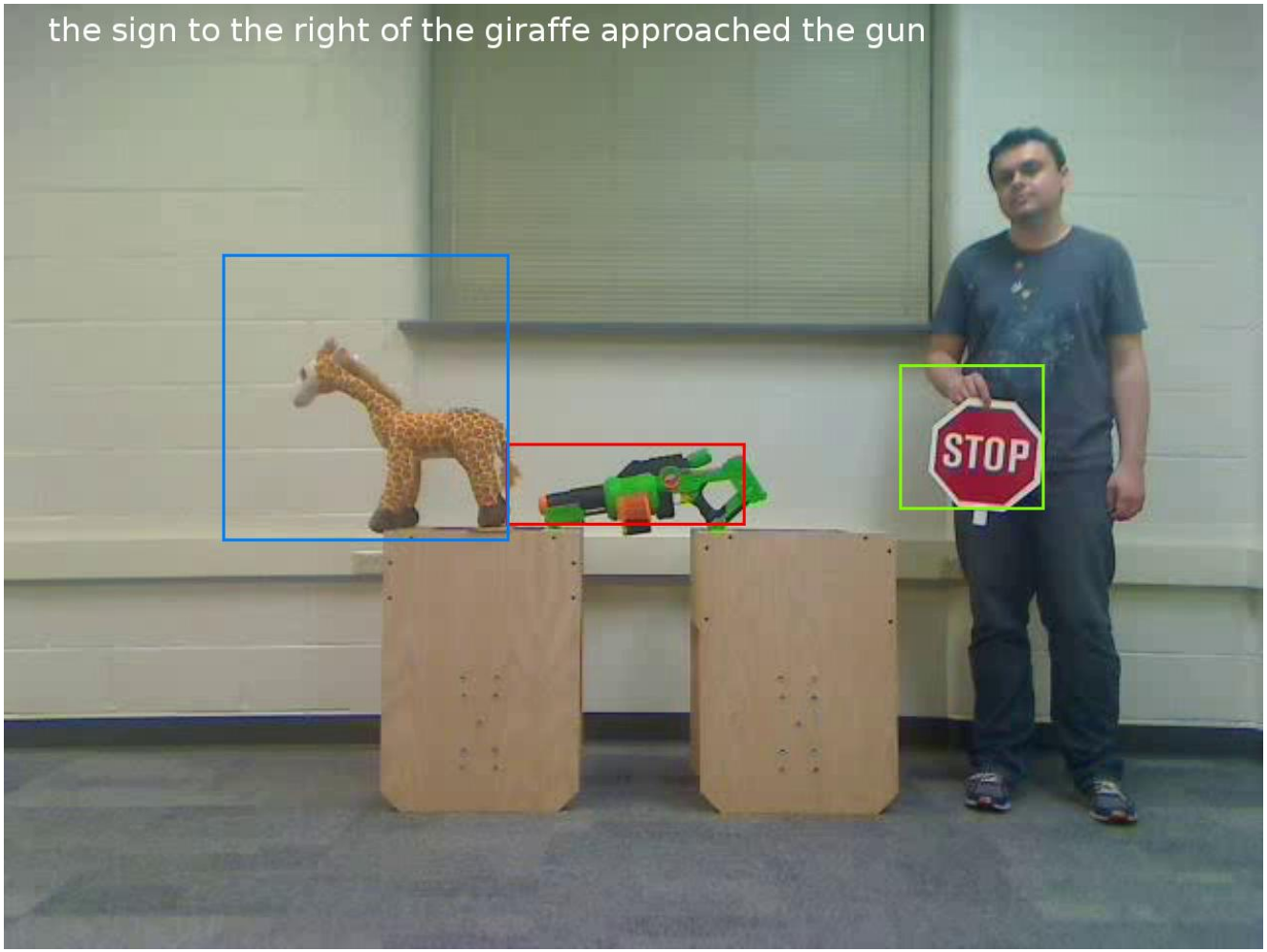
Beam Search

base case Find the k top-scoring 1-word phrases.

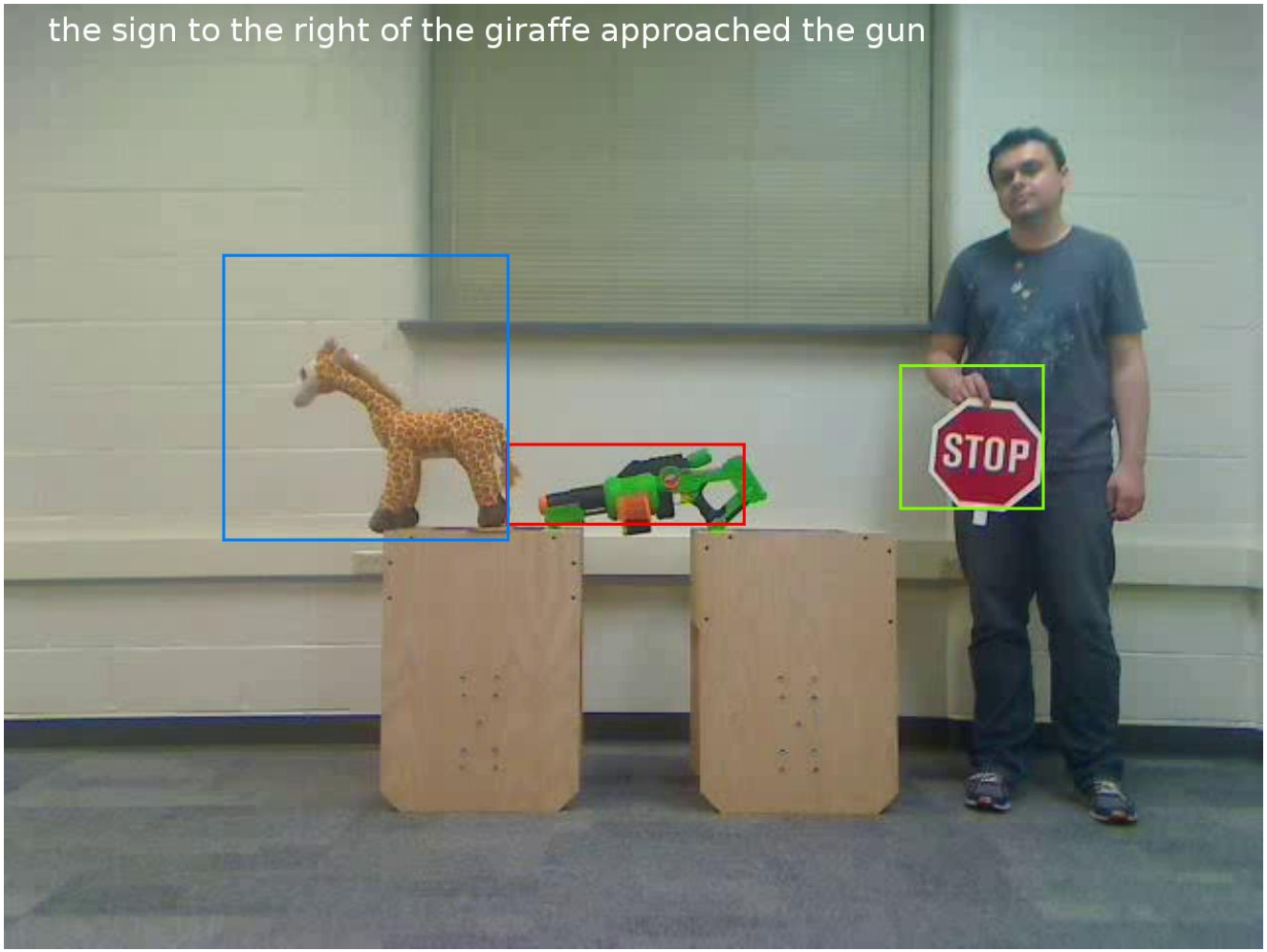
base case Find the k top-scoring 1-word phrases.

inductive case Given the k top-scoring n-word phrases, find the k top-scoring $n+1$-word phrases.

Using the Sentence Tracker to Perform Generation

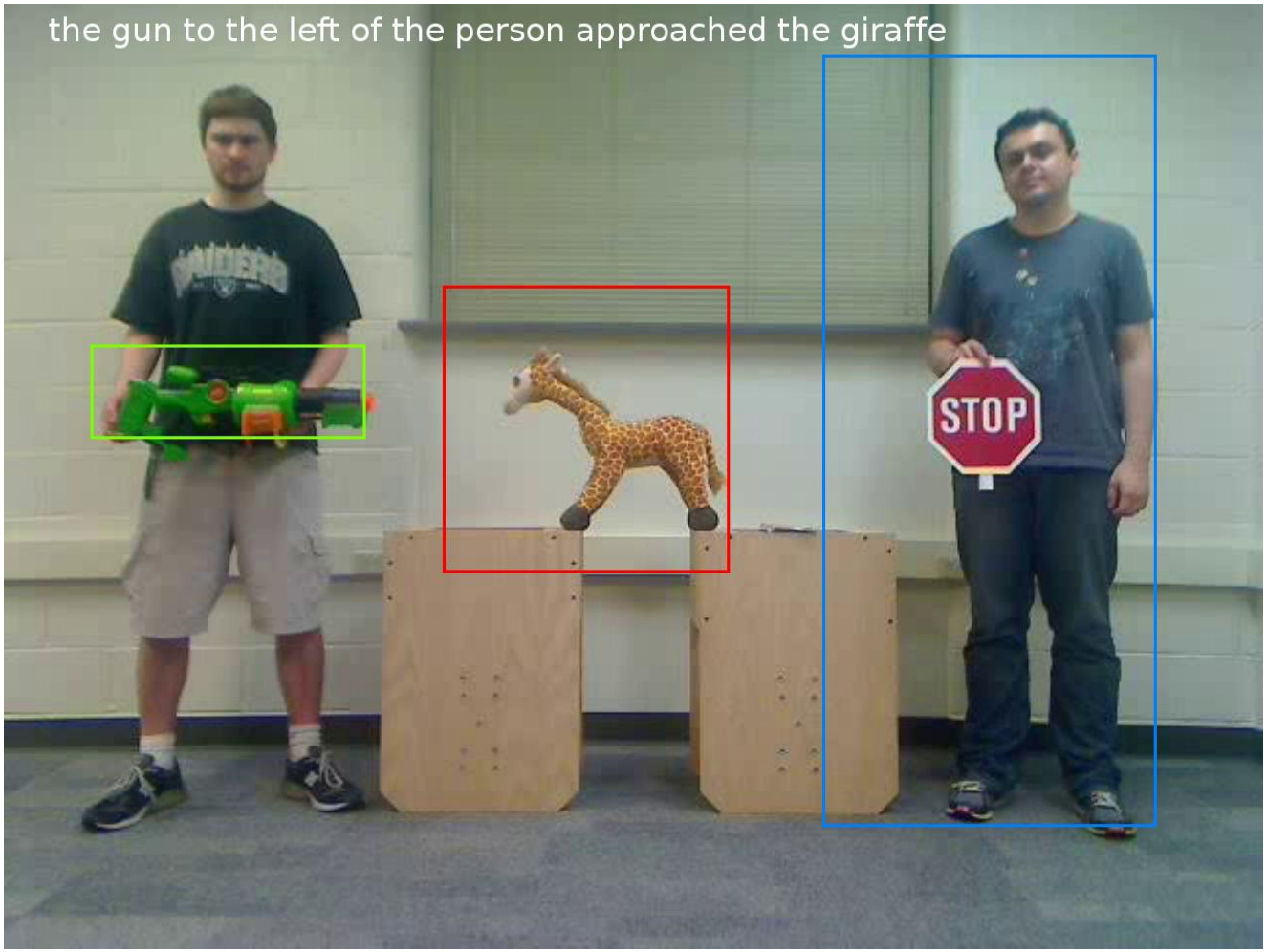


Using the Sentence Tracker to Perform Generation

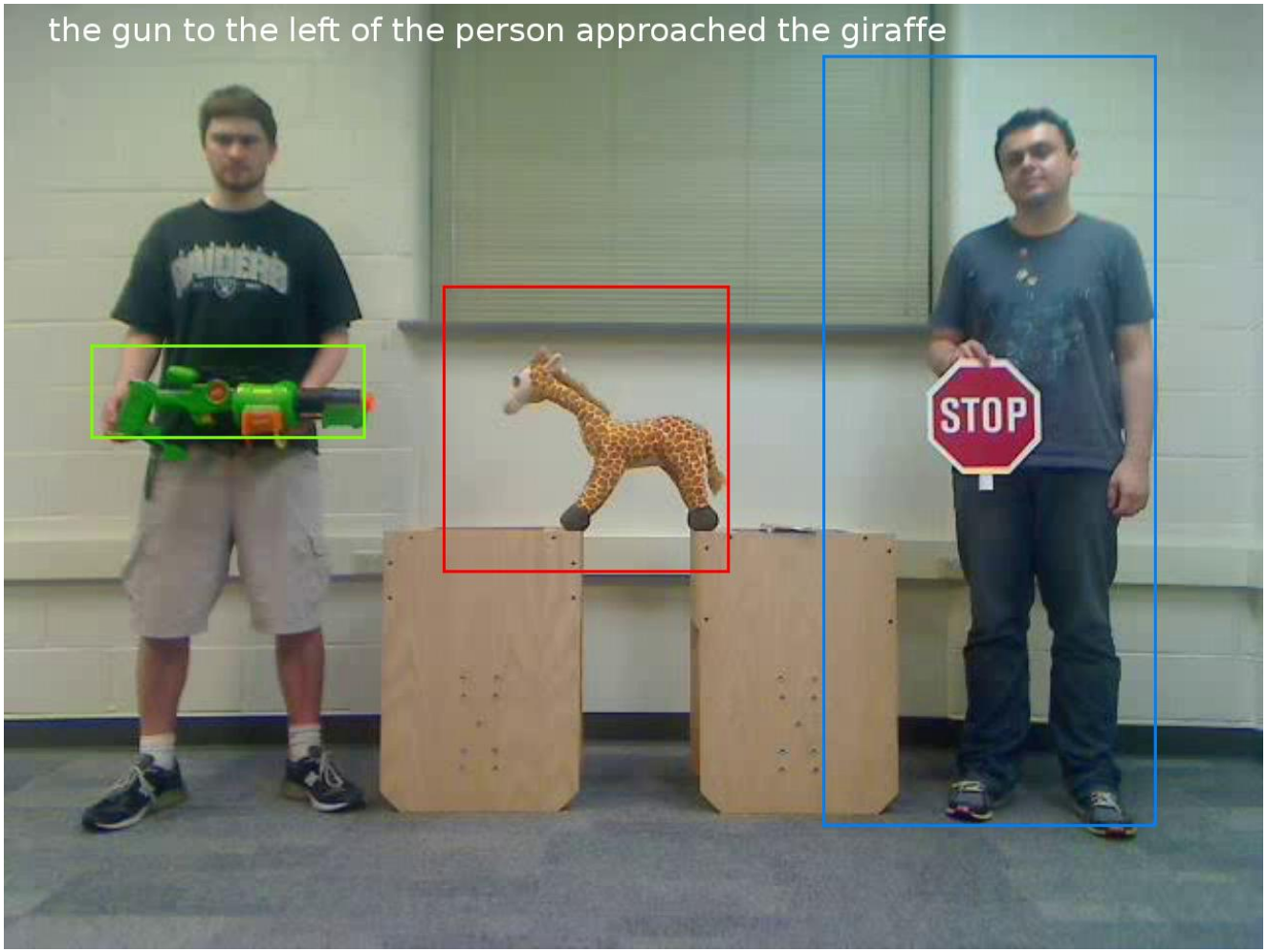


The sign to the right of the giraffe approached the gun.

Using the Sentence Tracker to Perform Generation

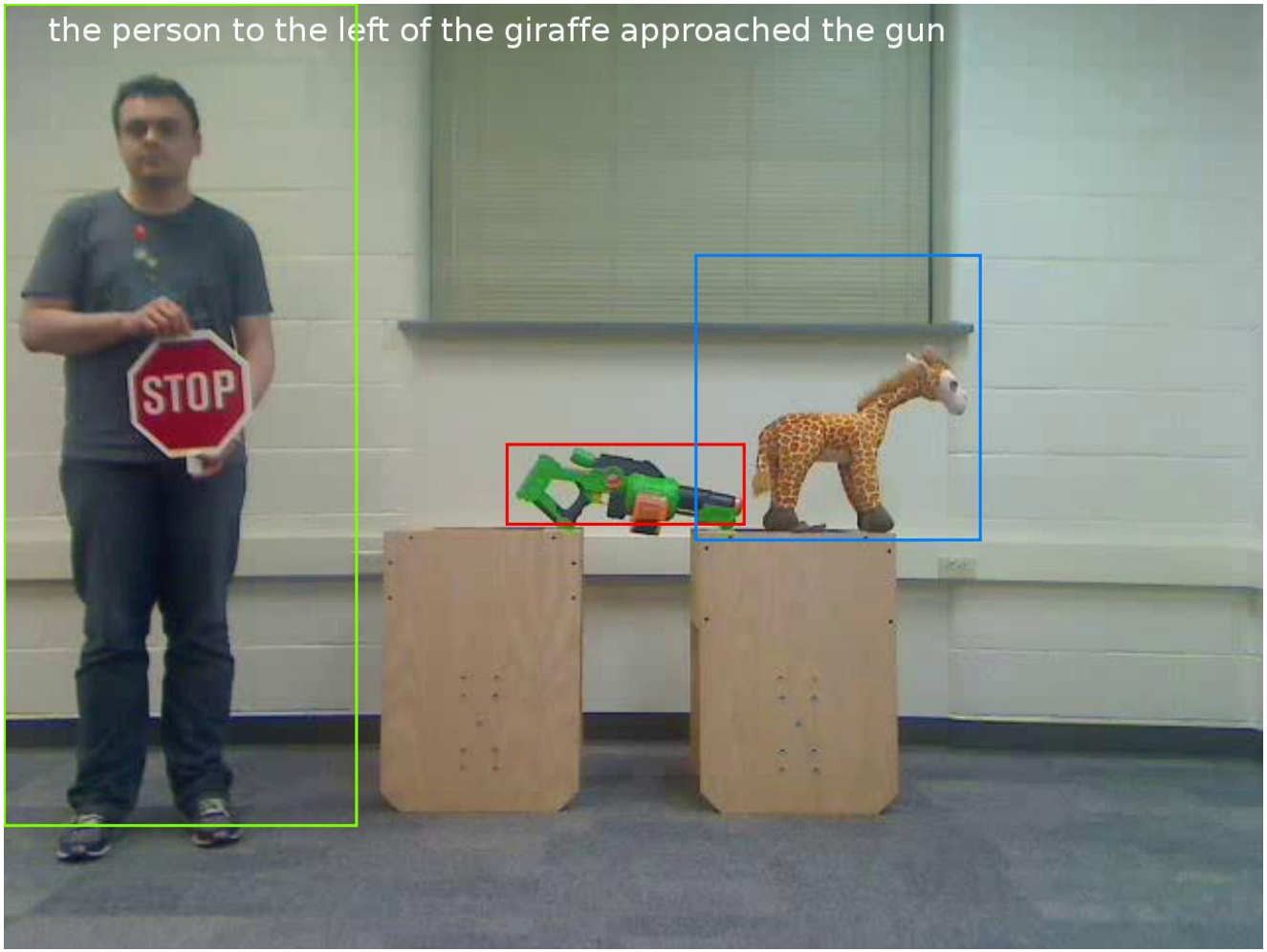


Using the Sentence Tracker to Perform Generation

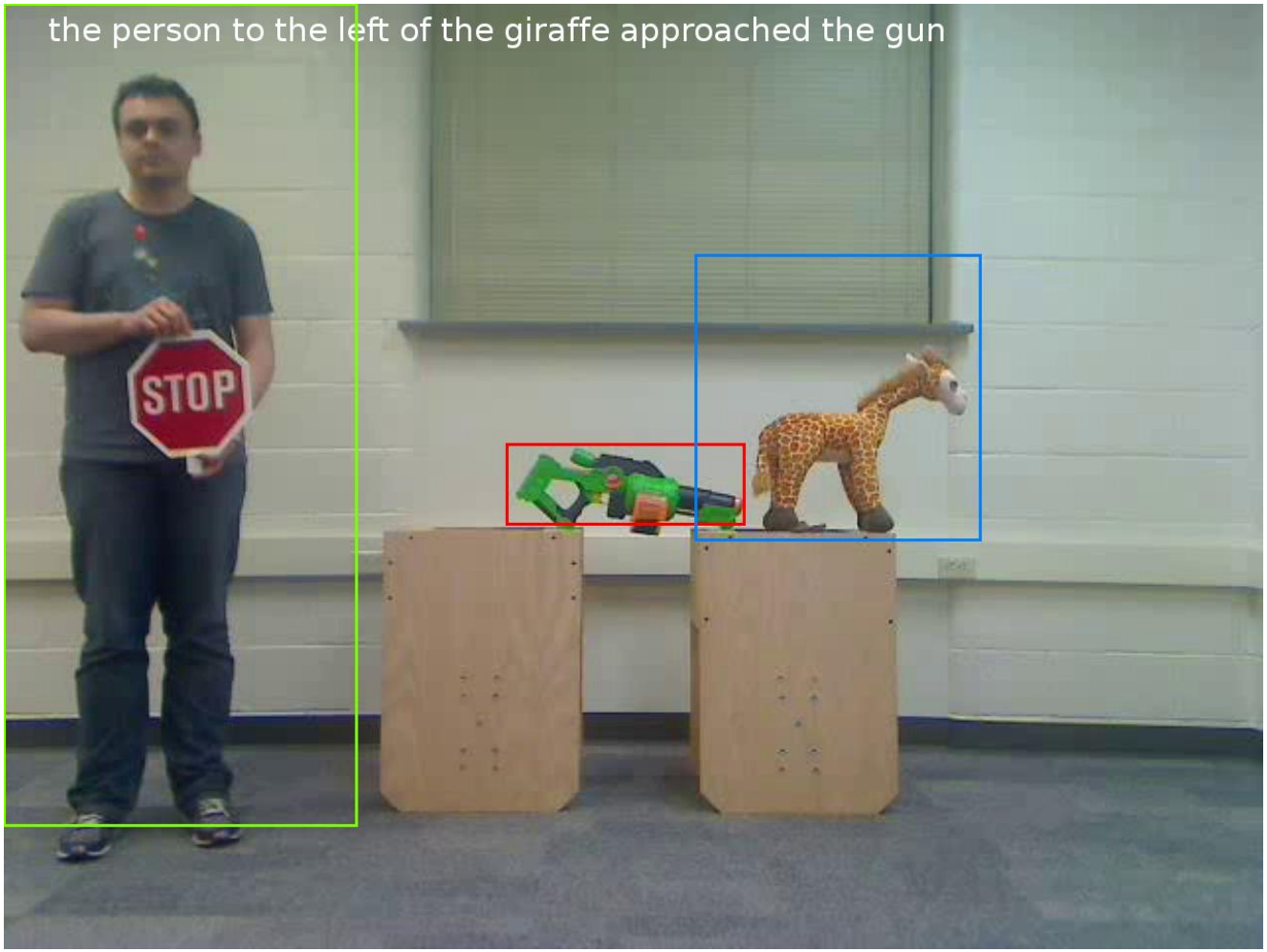


The gun to the left of the person approached the giraffe.

Using the Sentence Tracker to Perform Generation

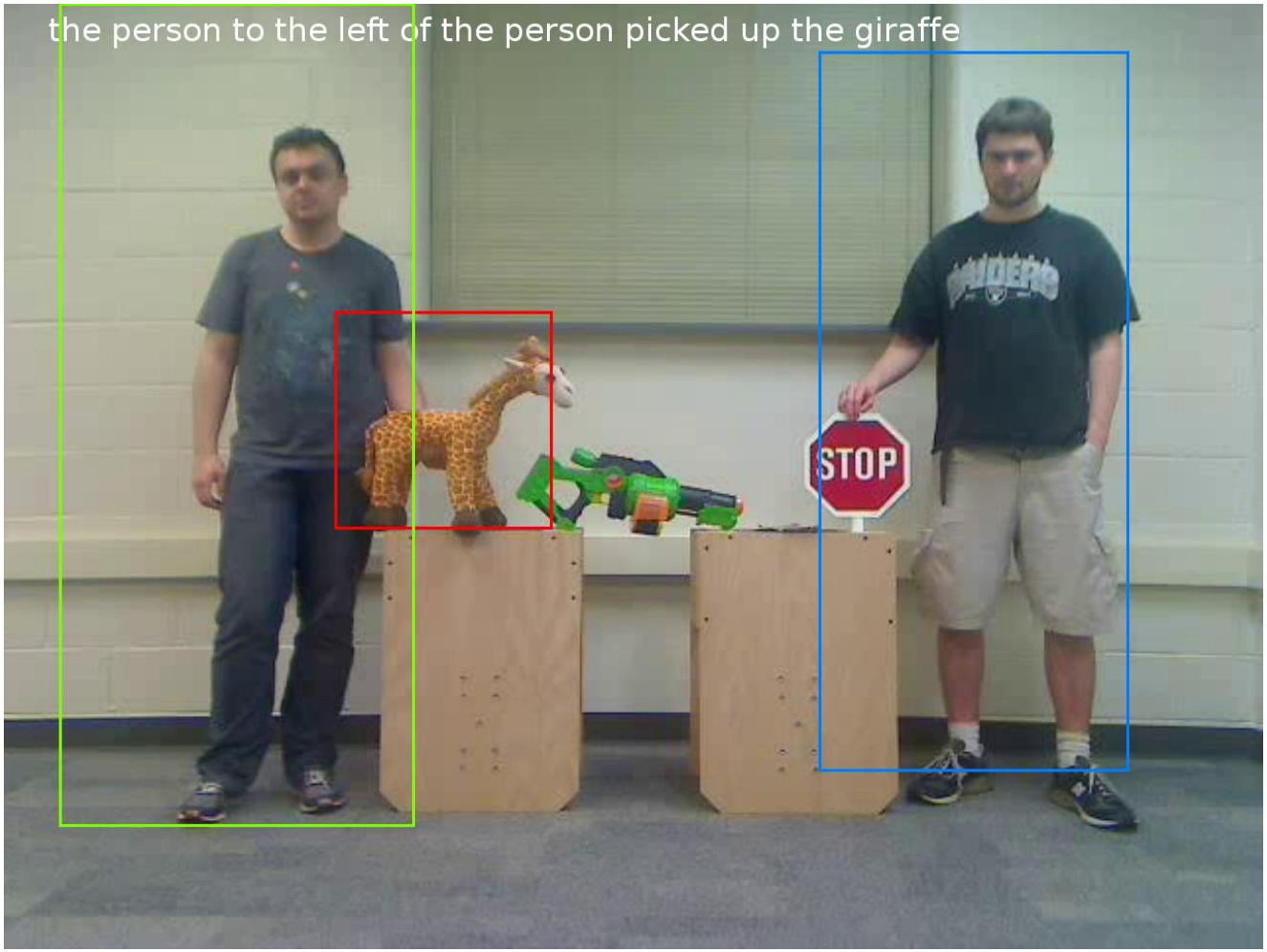


Using the Sentence Tracker to Perform Generation

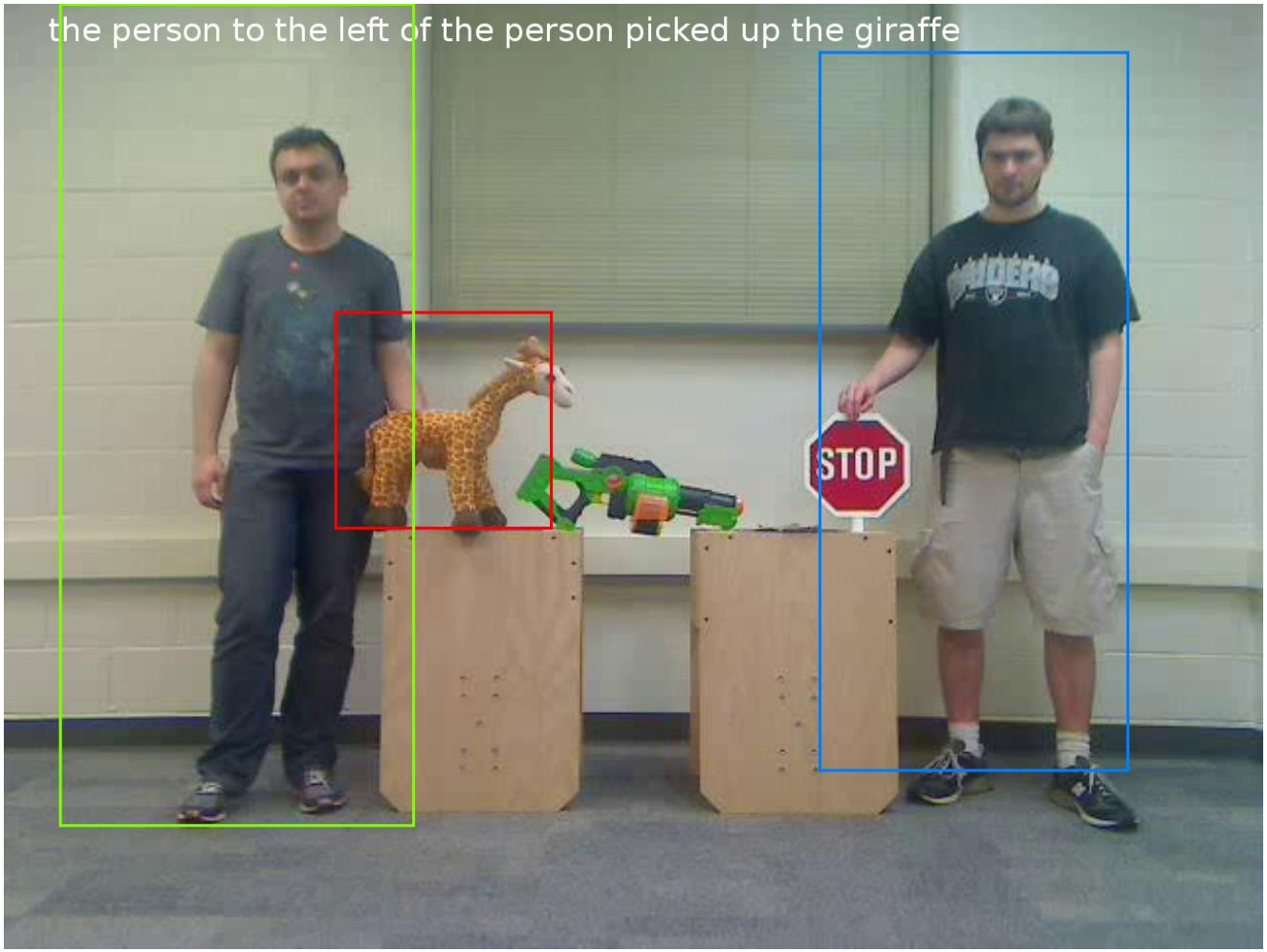


The person to the left of the giraffe approached the gun.

Using the Sentence Tracker to Perform Generation

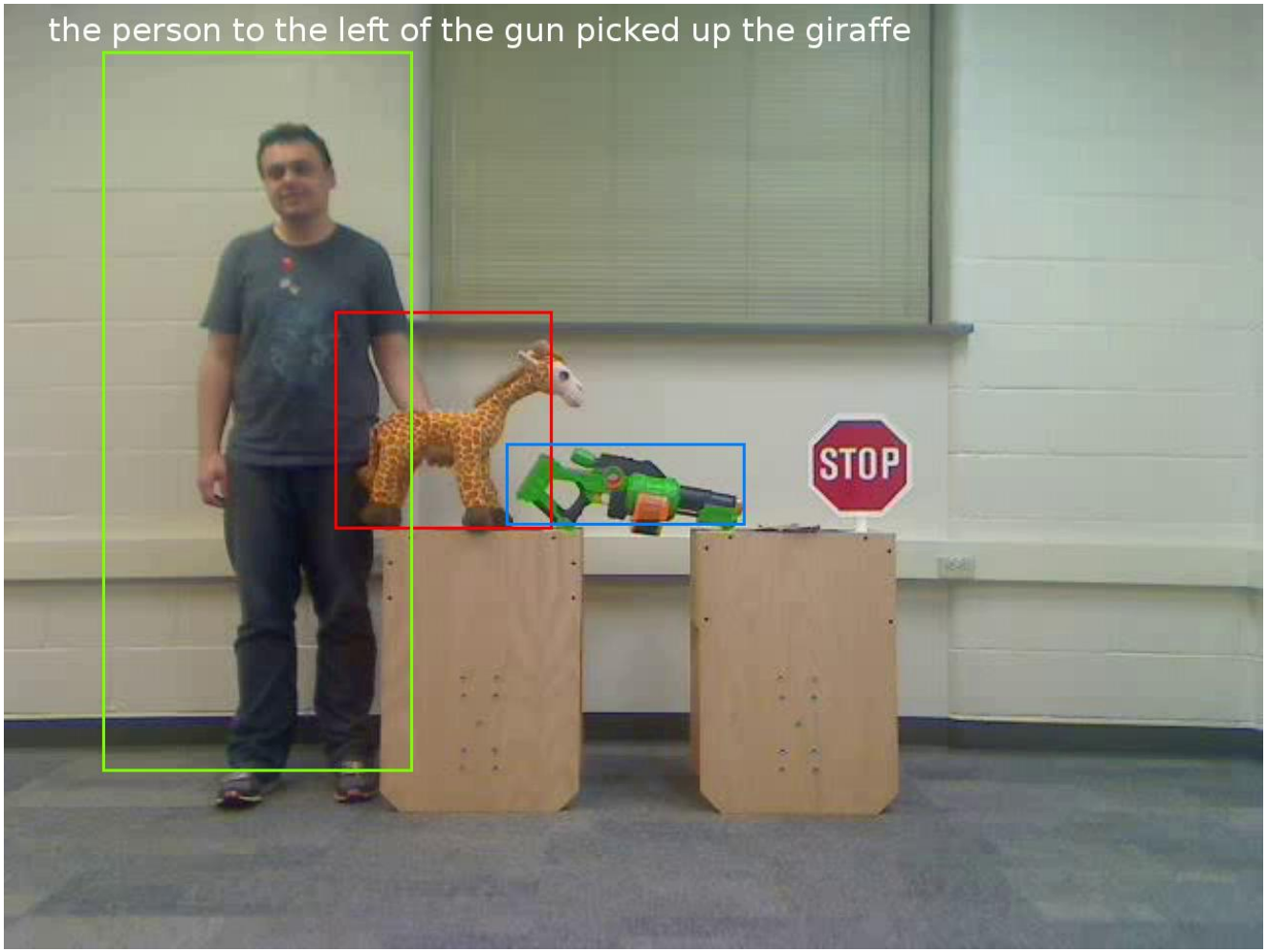


Using the Sentence Tracker to Perform Generation

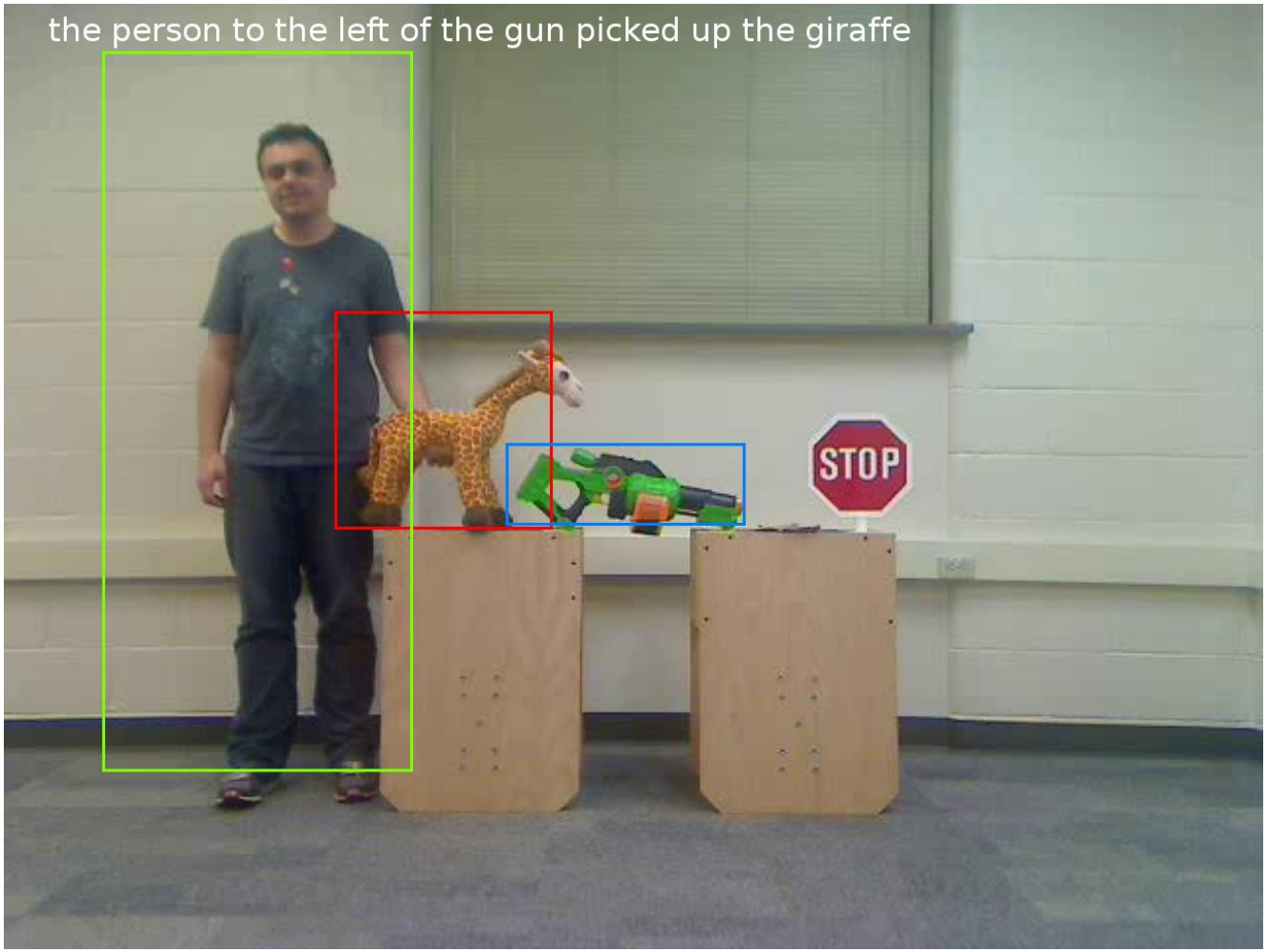


The person to the left of the person picked up the giraffe.

Using the Sentence Tracker to Perform Generation



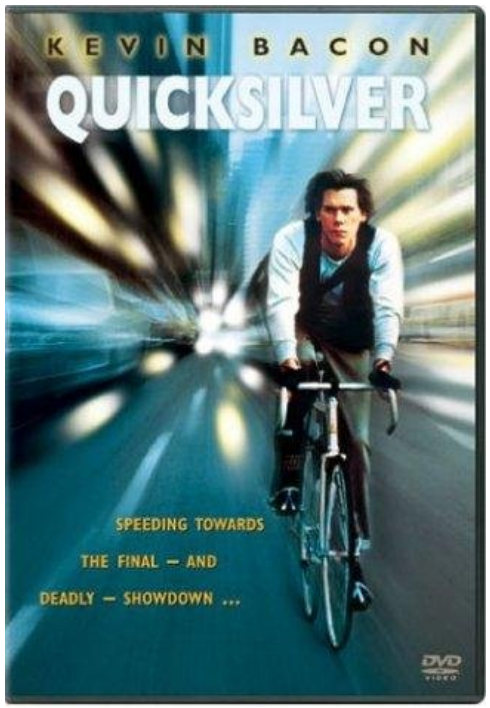
Using the Sentence Tracker to Perform Generation



The person to the left of the gun picked up the giraffe.

Using the Sentence Tracker to Perform Retrieval

- Quicksilver (Columbia Pictures, 1986)
- original
  - 106:17
  - 153080 frames
  - 23.976 fps
- downsampled
  - 19135 frames
  - 3.0 fps
  - 2125 clips
  - 12 frames/clip
  - 4 sec/clip



Using the Sentence Tracker to Perform Retrieval

The person approached the car from the left.

Using the Sentence Tracker to Perform Retrieval

The person approached the car from the left.

Using the Sentence Tracker to Perform Retrieval

The person approached the car from the left.

Using the Sentence Tracker to Perform Retrieval

The person rode the bicycle leftward.

Using the Sentence Tracker to Perform Retrieval

The person rode the bicycle leftward.

Using the Sentence Tracker to Perform Retrieval

The person rode the bicycle leftward.

Using the Sentence Tracker to Perform Retrieval

The person rode the bicycle rightward.

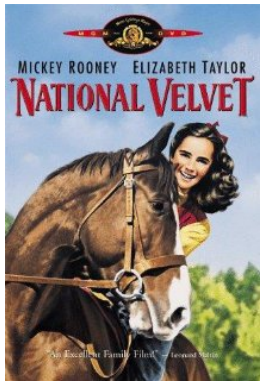
Using the Sentence Tracker to Perform Retrieval

The person rode the bicycle rightward.

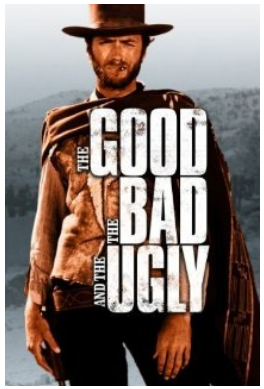
Using the Sentence Tracker to Perform Retrieval

The person rode the bicycle rightward.

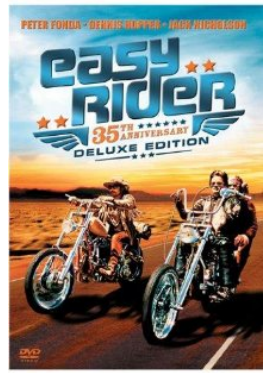
Using the Sentence Tracker to Perform Retrieval



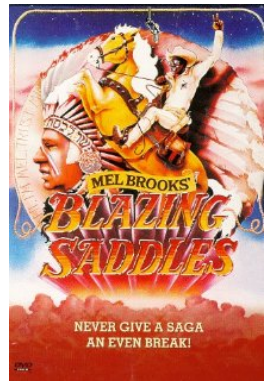
National Velvet
(MGM, 1944)



The Good The Bad And The Ugly
(PEA, 1966)



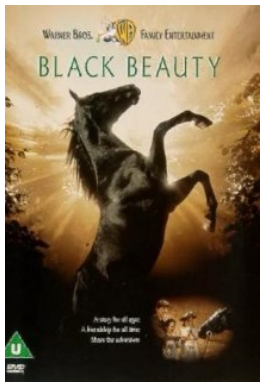
Easy Rider
(Columbia Pictures, 1969)



Blazing Saddles
(Warner Bros. Pictures, 1974)

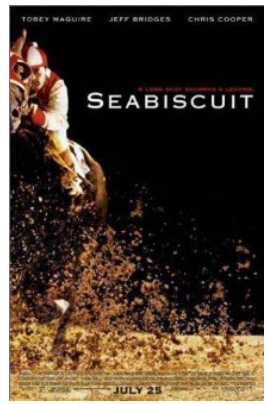
Black Stallion
(Omni Zoetrope, 1979)

Unforgiven
(Warner Bros. Pictures, 1992)

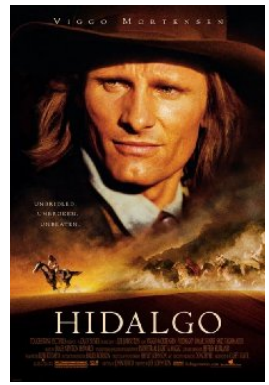


Black Beauty
(Warner Bros. Pictures, 1994)

Once Upon a Time in Mexico
(Columbia Pictures, 2003)



Seabiscuit
(Universal Pictures, 2003)



Hidalgo
(Touchstone Pictures, 2004)

Queries for the Westerns Corpus

$$\textit{The}\ \left\{\begin{array}{l}\textit{person}\\ \textit{horse}\end{array}\right\}\left\{\begin{array}{l}\textit{approached the}\ \left\{\begin{array}{l}\textit{person}\\ \textit{horse}\end{array}\right\}\left\{\begin{array}{l}\epsilon\\ \textit{quickly}\\ \textit{slowly}\end{array}\right\}\left\{\begin{array}{l}\epsilon\\ \textit{from the left}\\ \textit{from the right}\end{array}\right\}\\ \left\{\begin{array}{l}\textit{lead}\\ \textit{rode}\end{array}\right\}\textit{the}\ \left\{\begin{array}{l}\textit{person}\\ \textit{horse}\end{array}\right\}\left\{\begin{array}{l}\epsilon\\ \textit{quickly}\\ \textit{slowly}\end{array}\right\}\left\{\begin{array}{l}\epsilon\\ \textit{leftward}\\ \textit{rightward}\\ \left\{\begin{array}{l}\textit{towards}\\ \textit{away from}\end{array}\right\}\textit{the}\ \left\{\begin{array}{l}\textit{person}\\ \textit{horse}\end{array}\right\}\end{array}\right\}\end{array}\right\}.$$

- 204 sentences
- 63 involve people riding people, horses riding people and horses
- 141 remain

Live Demo

Live Demo

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.

Using the Sentence Tracker to Perform Acquisition

The person carried the chair away from the backpack.

Using the Sentence Tracker to Perform Acquisition

The person carried the chair away from the backpack.

Using the Sentence Tracker to Perform Acquisition








The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition


The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition


The person picked up the traffic-cone.

The person picked up the traffic-cone to the left of the stool.

The person put down the trash-can quickly.

The person carried the chair.

The person carried the backpack.

The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition


The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition








The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

training sentences → learned words

person	*traffic-cone*	*chair*	*stool*	*backpack*	*trash-can*
picked up	*carried*	*put down*	*approached*	*to the left of*	*quickly*

Using the Sentence Tracker to Perform Acquisition

The person picked up the traffic-cone.
The person picked up the traffic-cone to the left of the stool.
The person put down the trash-can quickly.

The person carried the chair.
The person carried the backpack.
The chair approached the backpack.

training sentences → learned words

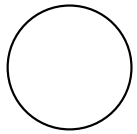
person	*traffic-cone*	*chair*	*stool*	*backpack*	*trash-can*
picked up	*carried*	*put down*	*approached*	*to the left of*	*quickly*

learned words → new sentences

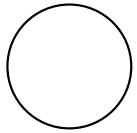
The backpack approached the person.
The person carried the trash-can.

The person put down the chair quickly.
. . .

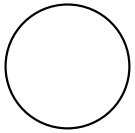
How Can We Do This?



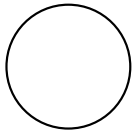
chair



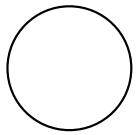
picked-up



person



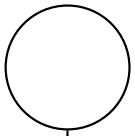
approached

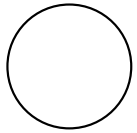


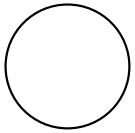
backpack

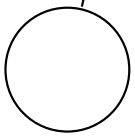
How Can We Do This?

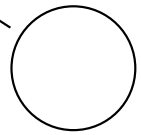
The chair approached the backpack.









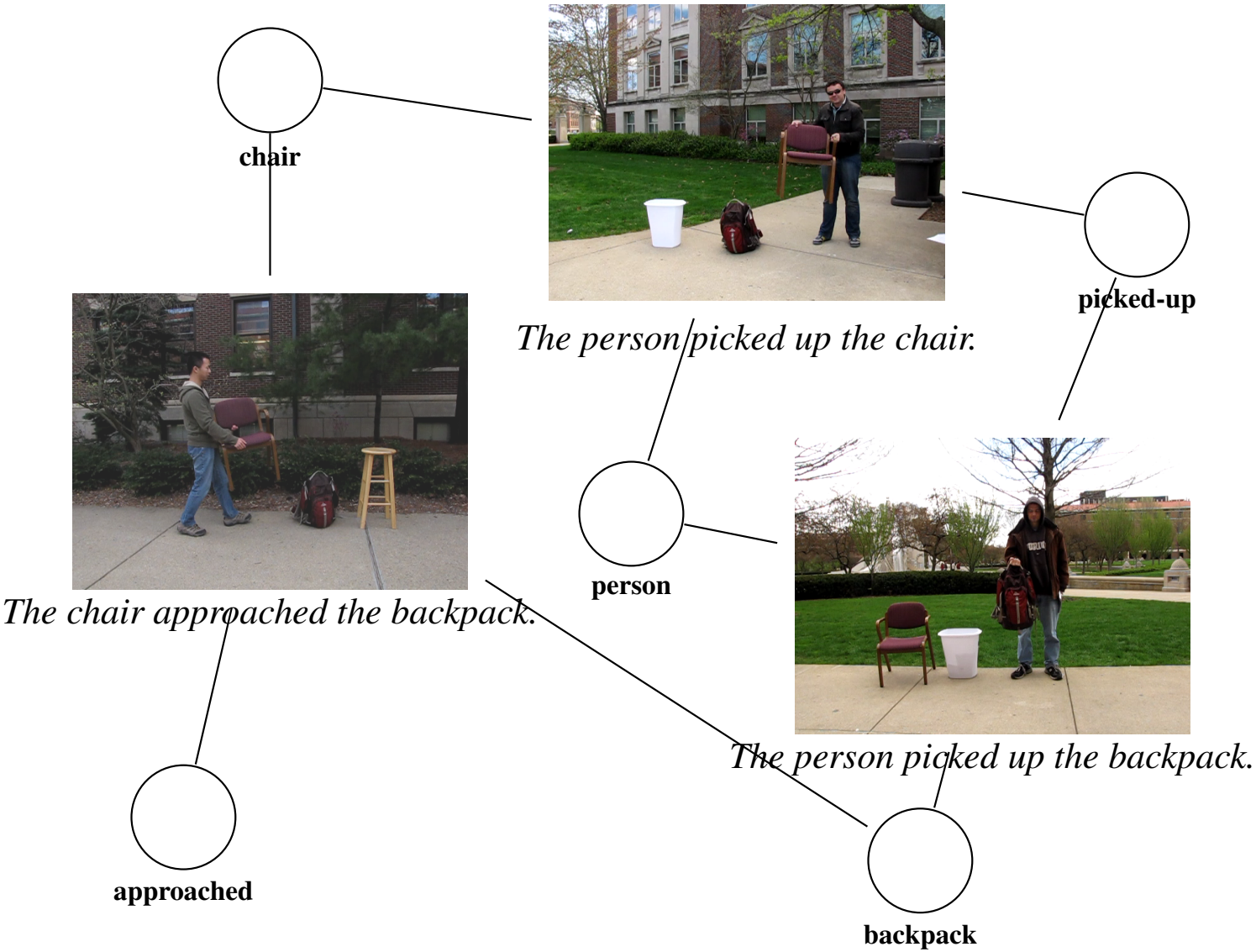


How Can We Do This?

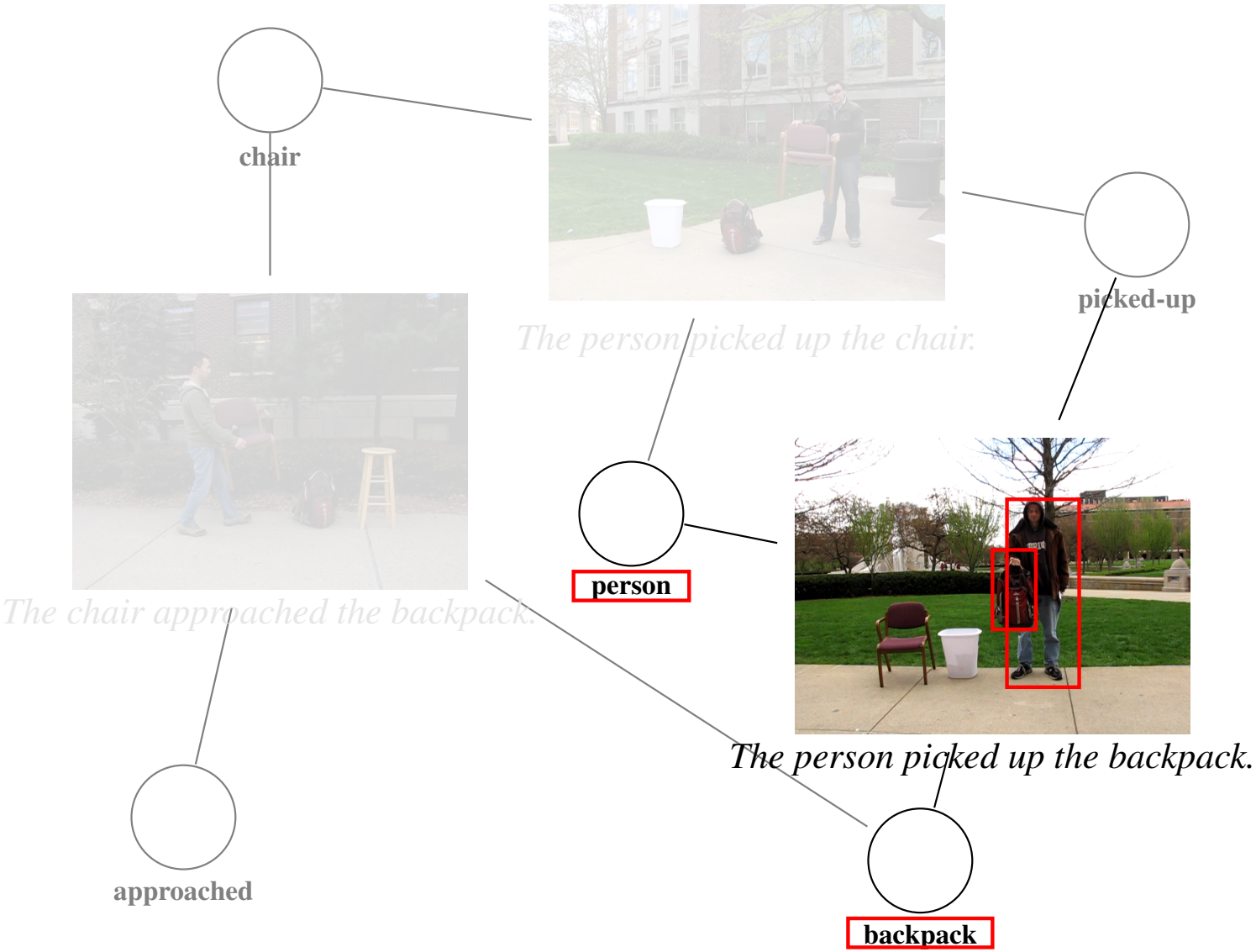
The chair approached the backpack.

The person picked up the backpack.

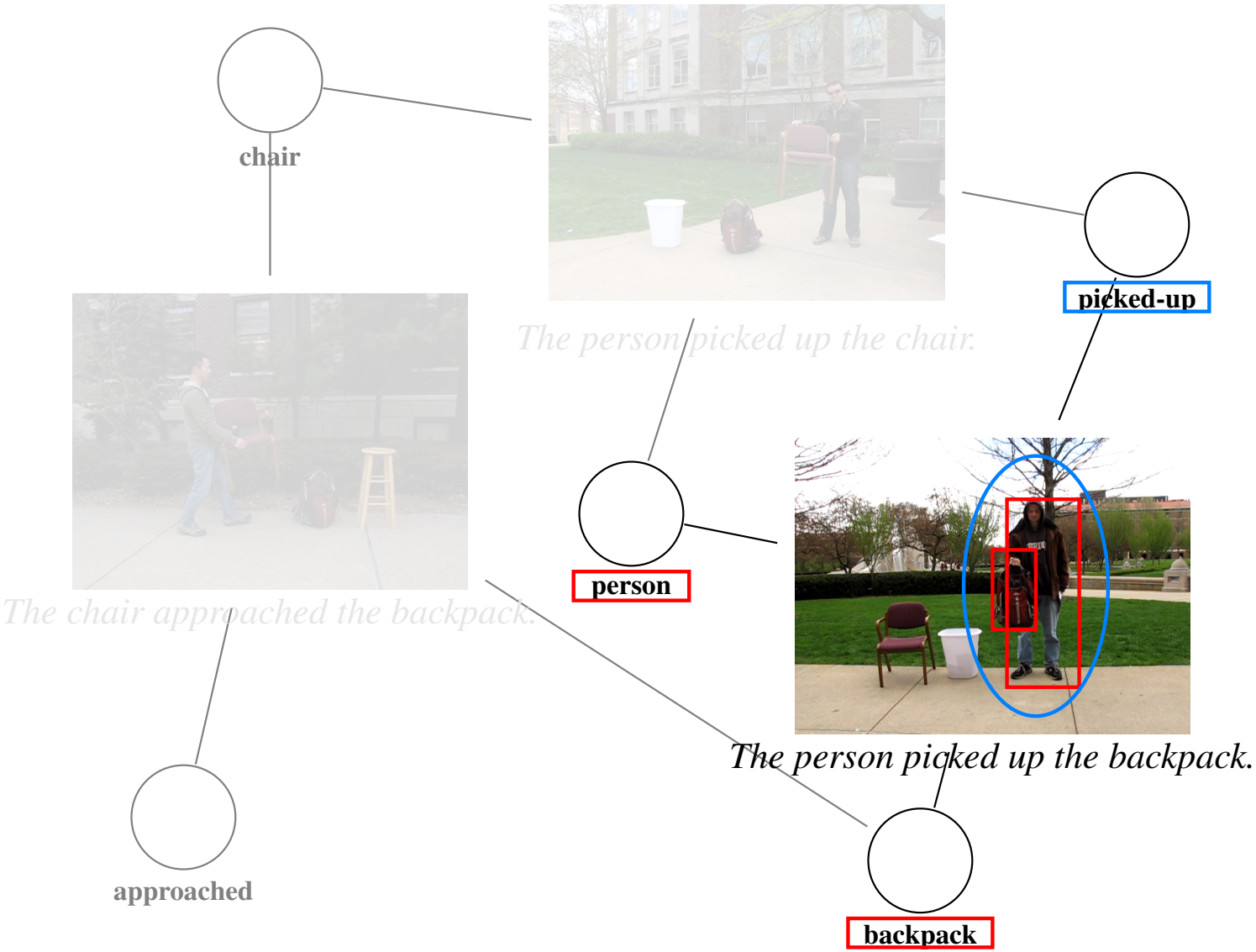
How Can We Do This?



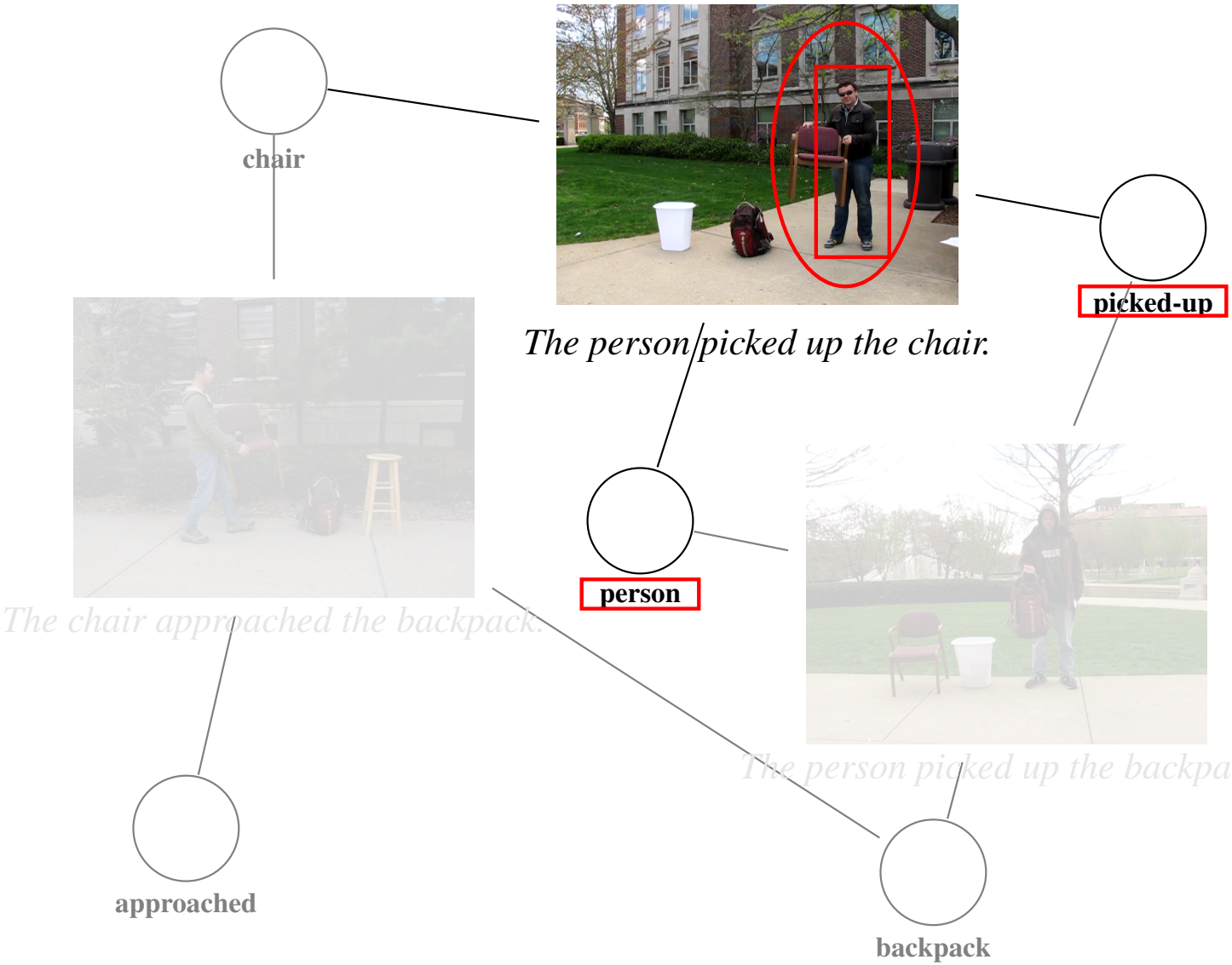
Inference Across *Different* Words in the *Same* Sentence



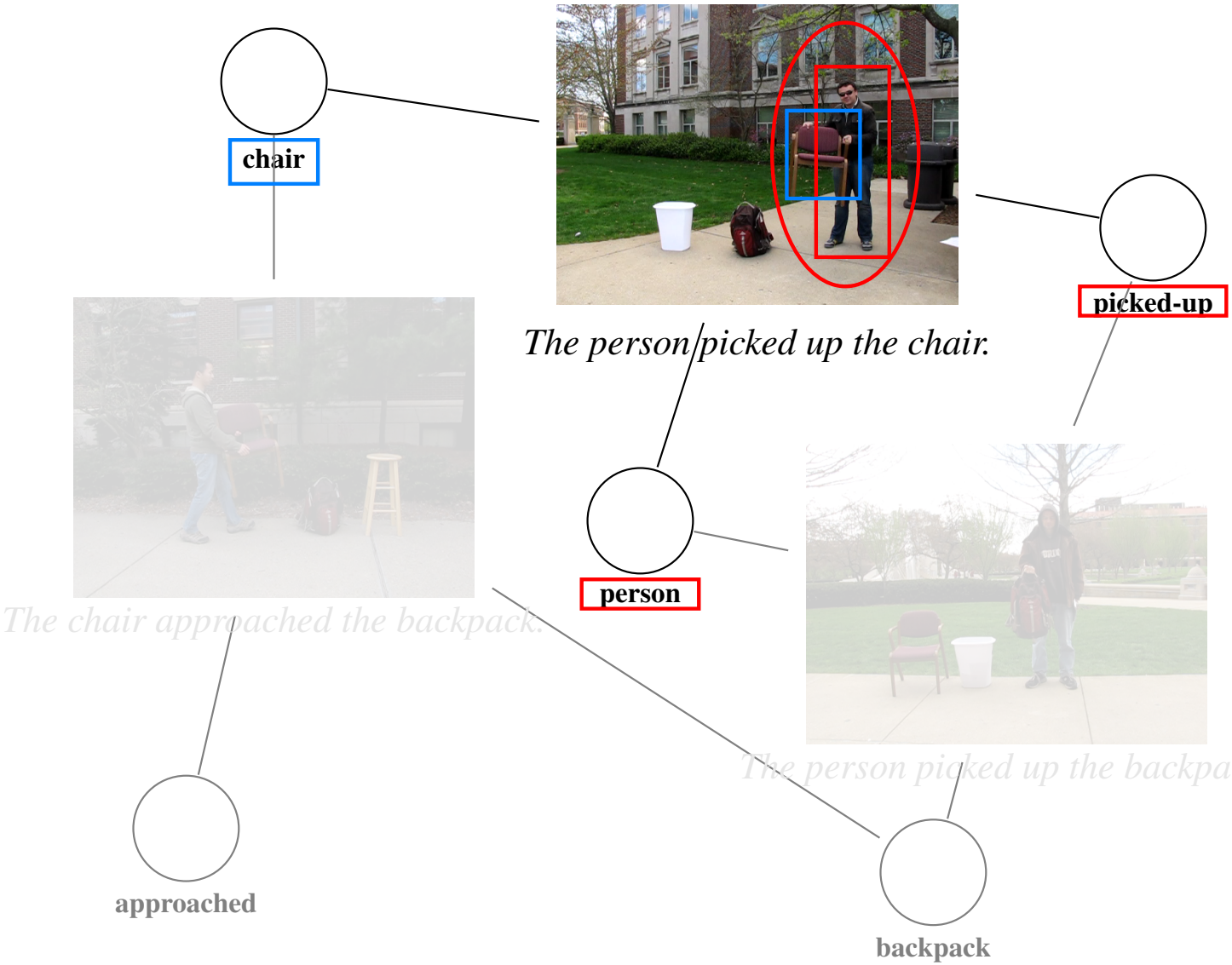
Inference Across *Different* Words in the *Same* Sentence



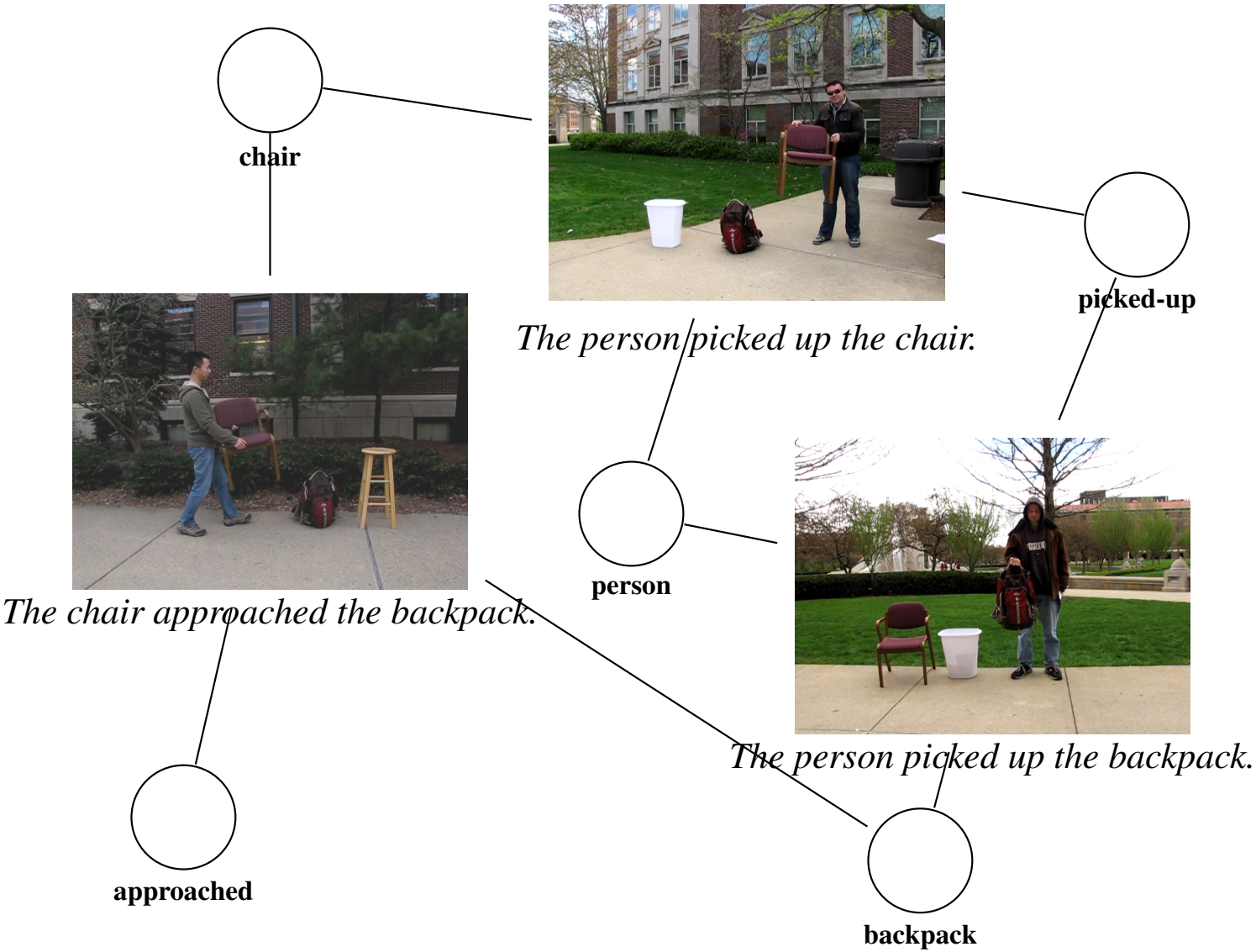
Inference Across the *Same* Word in *Different* Sentences



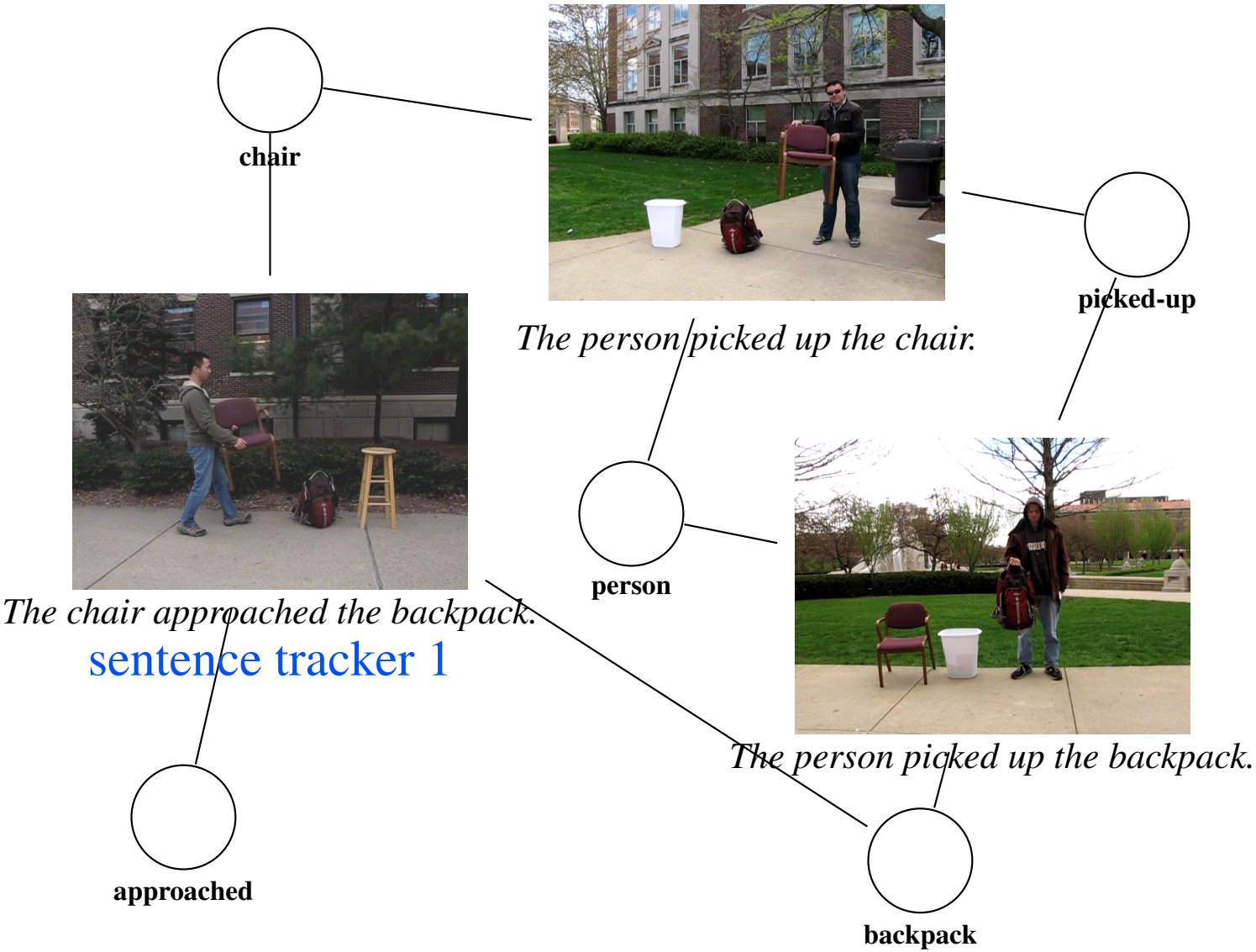
Inference Across the *Same* Word in *Different* Sentences



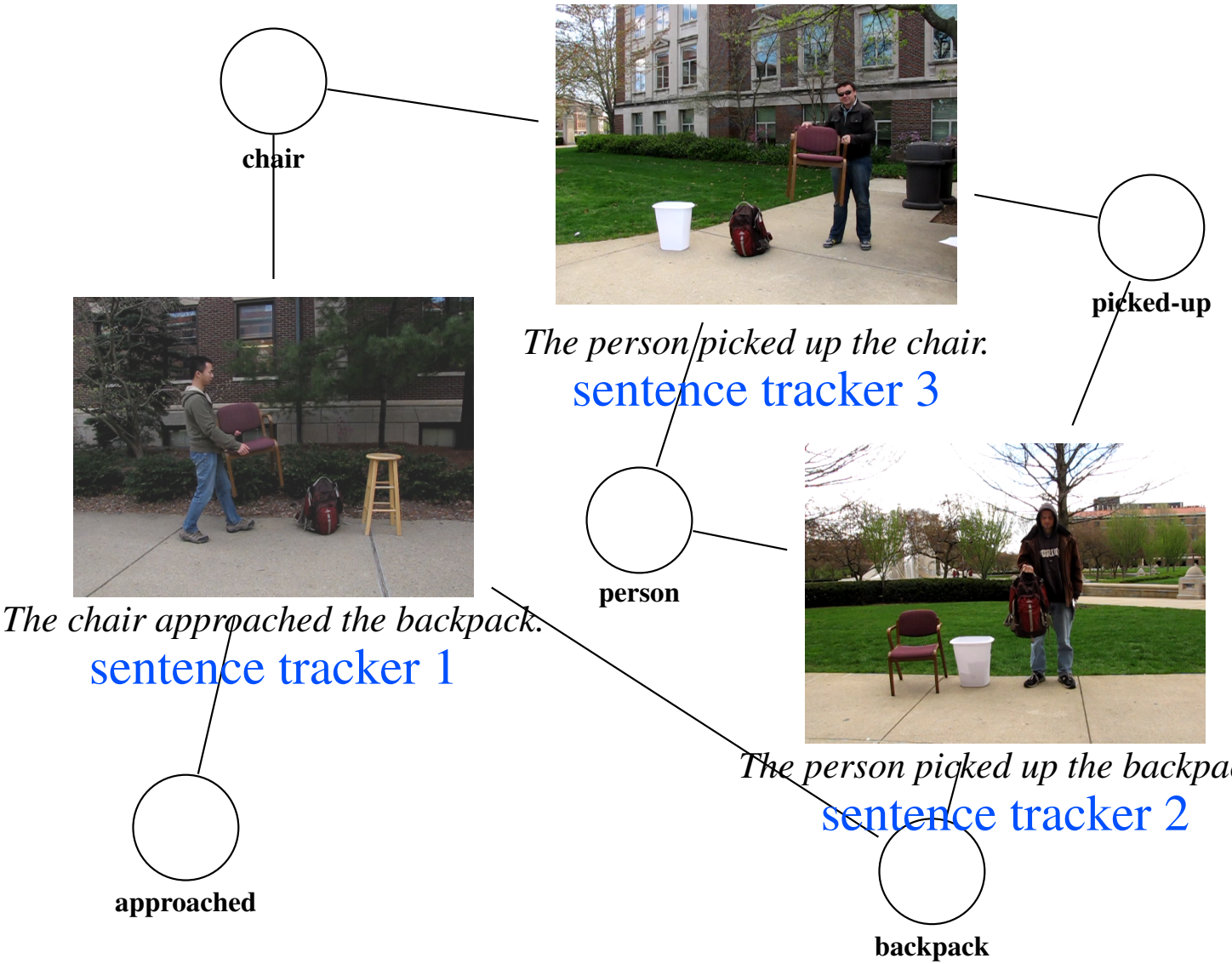
Sentence Tracker in EM



Sentence Tracker in EM



Sentence Tracker in EM



Sentence Tracker in EM

$$\max_{j_1^1,\ldots,j_1^T \; j_L^1,\ldots,j_L^T} \; \max_{k_1^1,\ldots,k_1^T \; k_W^1,\ldots,k_W^T} \; \sum_{l=1}^{L} \sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \sum_{w=1}^{W} \sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t)$$

Sentence Tracker in EM

$$\log \sum_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \sum_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \exp\left[\begin{array}{l} \sum_{l=1}^{L} \sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \\ \sum_{w=1}^{W} \sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t) \end{array} \right]$$

Sentence Tracker in EM

$$\log \sum_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \sum_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \exp\left[\begin{array}{l} \sum_{l=1}^{L} \sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \\ \sum_{w=1}^{W} \sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t) \end{array} \right]$$

- Wrap the sum of log likelihoods of all video-sentence pairs in EM.

Sentence Tracker in EM

$$\log \sum_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \sum_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \exp \left[\begin{array}{l} \sum_{l=1}^{L} \sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \\ \sum_{w=1}^{W} \sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t) \end{array} \right]$$

- Wrap the sum of log likelihoods of all video-sentence pairs in EM.
- In the E-step, compute probability for tracks, HMM states, and outputs.

Sentence Tracker in EM

$$\log \sum_{\substack{j_1^1,\ldots,j_1^T \\ j_L^1,\ldots,j_L^T}} \sum_{\substack{k_1^1,\ldots,k_1^T \\ k_W^1,\ldots,k_W^T}} \exp \left[\begin{array}{l} \sum_{l=1}^{L} \sum_{t=1}^{T} f(b_{j_l^t}^t) + \sum_{t=2}^{T} g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \\ \sum_{w=1}^{W} \sum_{t=1}^{T} h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t) + \sum_{t=2}^{T} a_w(k_w^{t-1}, k_w^t) \end{array} \right]$$

- Wrap the sum of log likelihoods of all video-sentence pairs in EM.
- In the E-step, compute probability for tracks, HMM states, and outputs.
- In the M-step, the transition matrix $a_w(k_w^{t-1}, k_w^t)$ and output distribution $h_w(k_w^t, b_{j_{\theta_w^1}^t}^t, b_{j_{\theta_w^2}^t}^t)$ are re-estimated.

Experiment Results

Experiment Results

- learn all content words in the lexicon

Experiment Results

$$
\begin{aligned}
\text{N} &\rightarrow \textit{person} \mid \textit{backpack} \mid \textit{trash-can} \mid \textit{chair} \mid \textit{traffic-cone} \mid \textit{stool} \\
\text{P} &\rightarrow \textit{to the left of} \mid \textit{to the right of} \\
\text{V} &\rightarrow \textit{picked up} \mid \textit{put down} \mid \textit{carried} \mid \textit{approached} \\
\text{ADV} &\rightarrow \textit{quickly} \mid \textit{slowly} \\
\text{PM} &\rightarrow \textit{towards} \mid \textit{away from}
\end{aligned}
$$

Experiment Results

- learn all content words in the lexicon
- 95 video clips, each video clip contains 1 person + 2 or 3 objects

Experiment Results

- learn all content words in the lexicon
- 95 video clips, each video clip contains 1 person + 2 or 3 objects
- about 200 training video-sentence pairs + 240 test video-sentence pairs

Experiment Results

- learn all content words in the lexicon
- 95 video clips, each video clip contains 1 person + 2 or 3 objects
- about 200 training video-sentence pairs + 240 test video-sentence pairs
- test on videos/sentences *never seen* in training set

Experiment Results

- learn all content words in the lexicon
- 95 video clips, each video clip contains 1 person + 2 or 3 objects
- about 200 training video-sentence pairs + 240 test video-sentence pairs
- test on videos/sentences *never seen* in training set

The person to the left of the stool picked up the chair.

Experiment Results

- learn all content words in the lexicon
- 95 video clips, each video clip contains 1 person + 2 or 3 objects
- about 200 training video-sentence pairs + 240 test video-sentence pairs
- test on videos/sentences *never seen* in training set

The person to the left of the stool picked up the chair.

The person carried the backpack towards the stool.

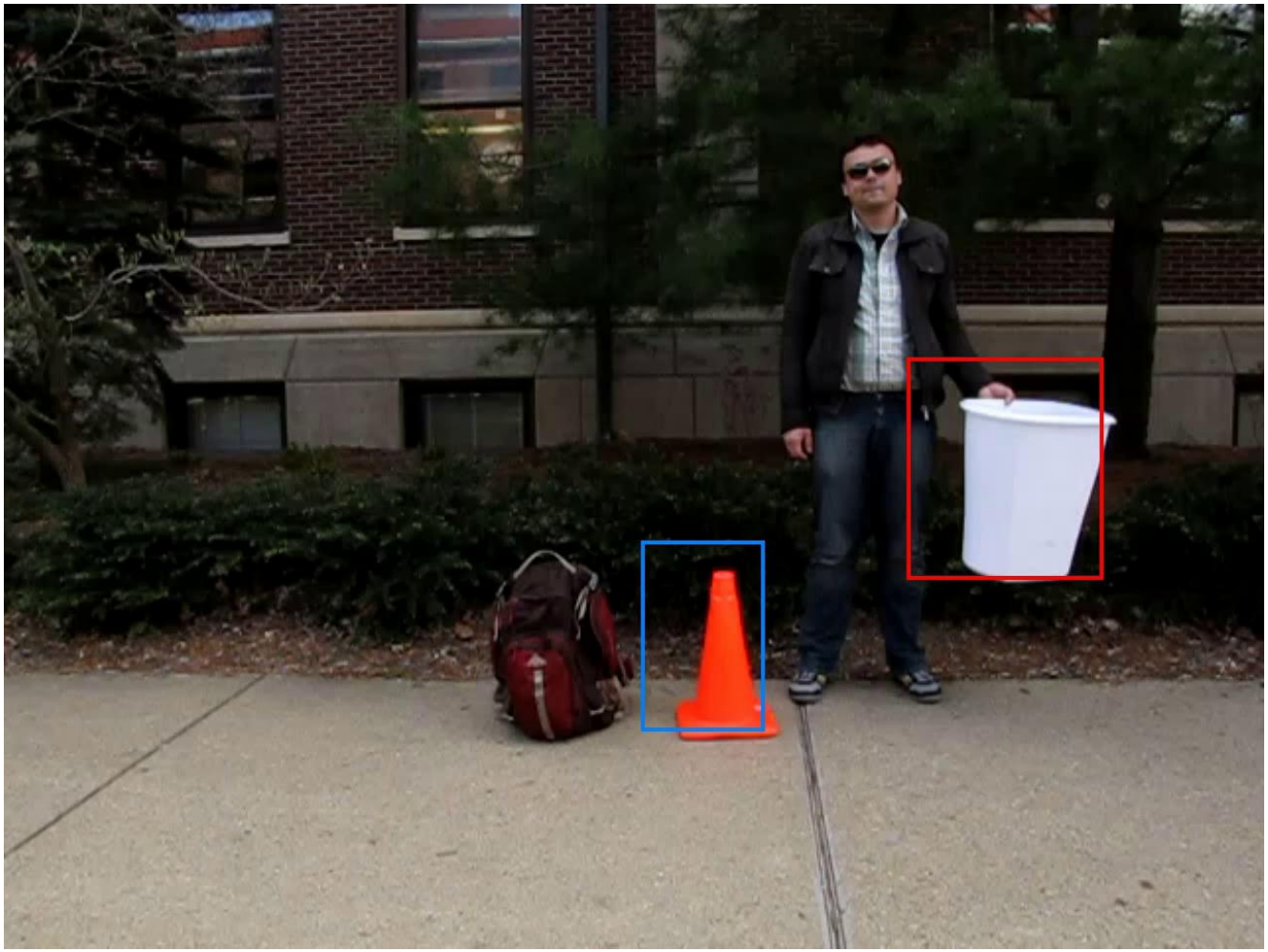
Experiment Results

- learn all content words in the lexicon
- 95 video clips, each video clip contains 1 person + 2 or 3 objects
- about 200 training video-sentence pairs + 240 test video-sentence pairs
- test on videos/sentences *never seen* in training set

The person to the left of the stool picked up the chair.

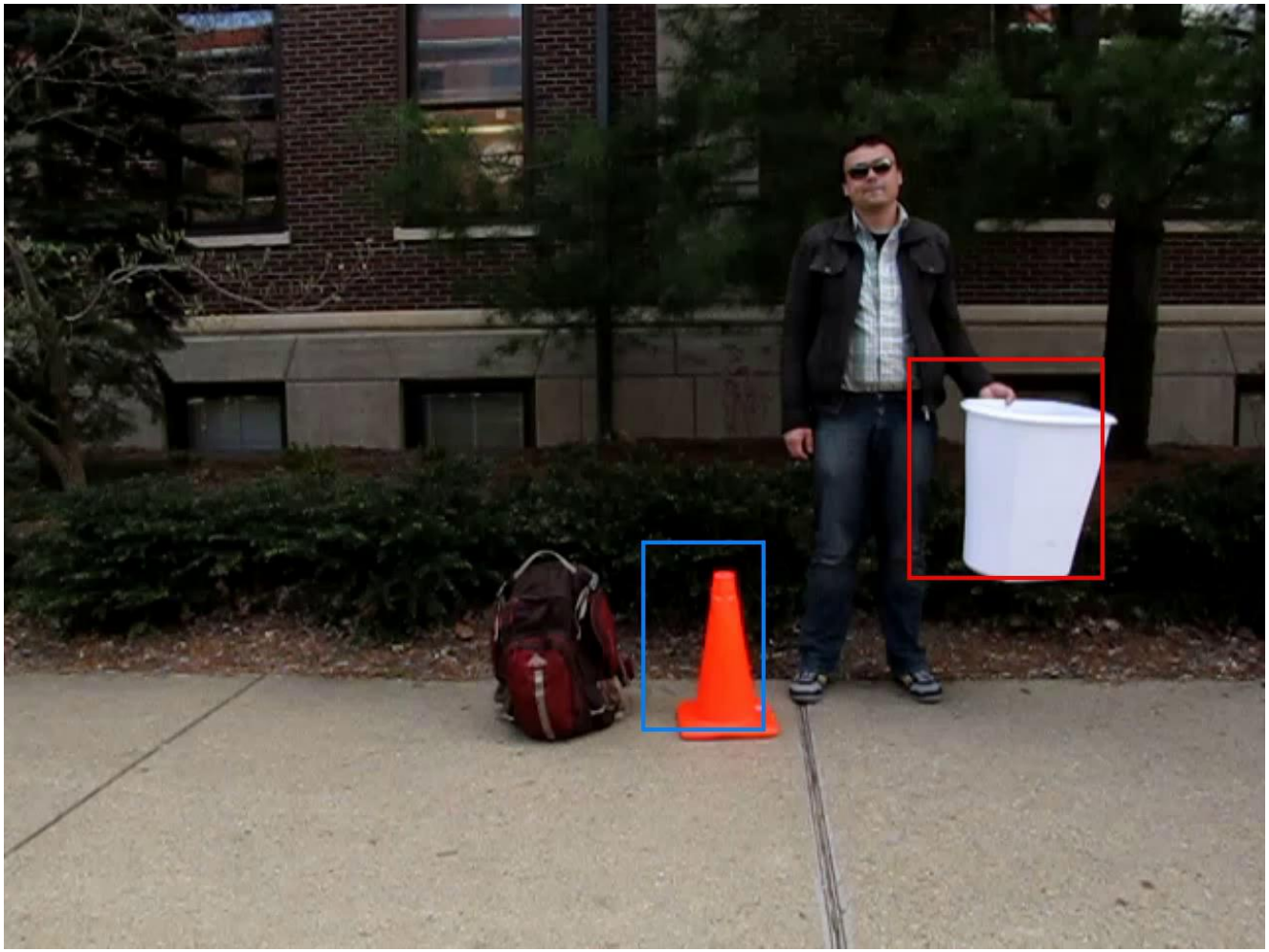
The person carried the backpack towards the stool.

New Sentence Generation with Trained Models



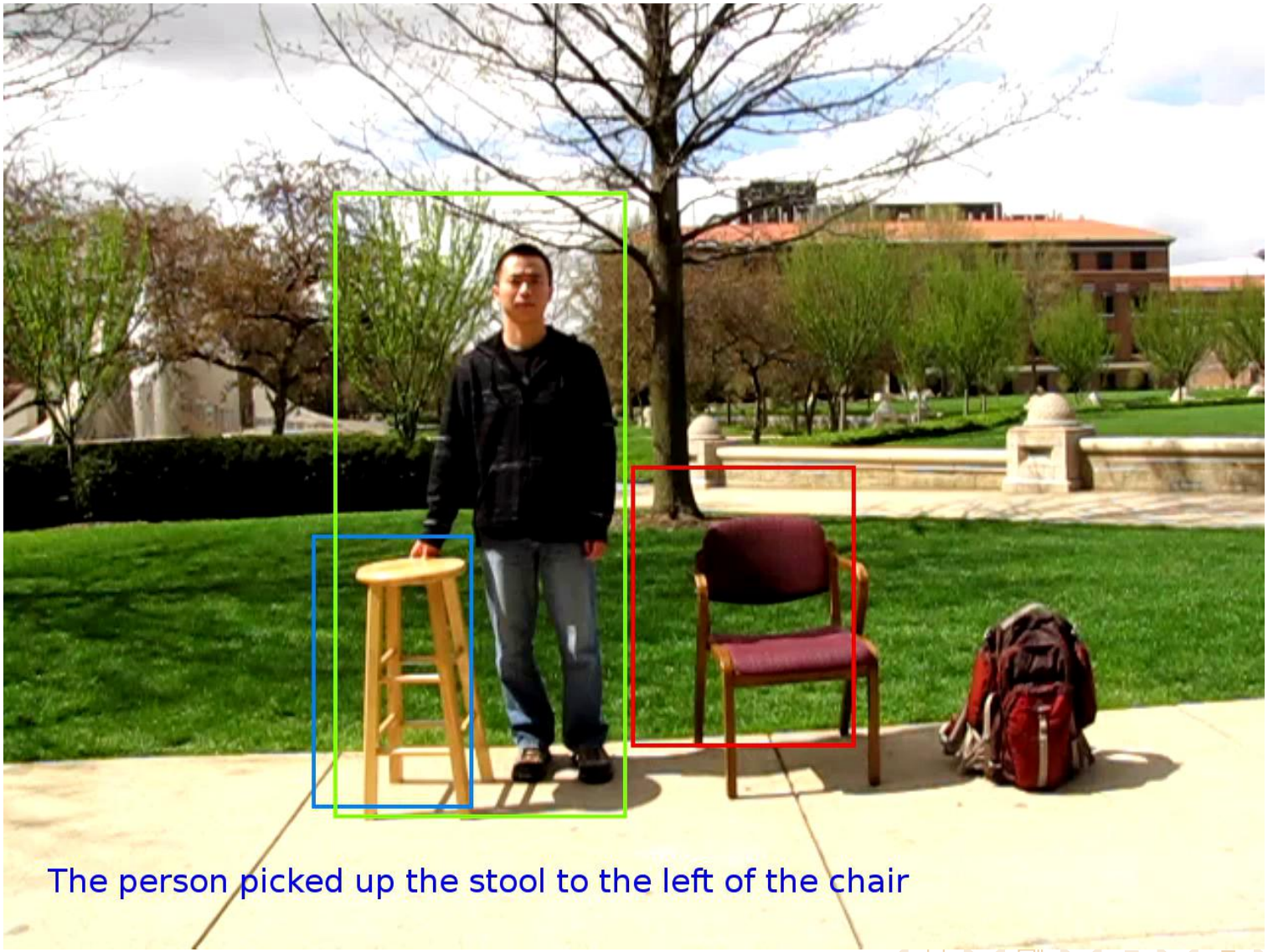
The trash-can approached the traffic-cone.

New Sentence Generation with Trained Models

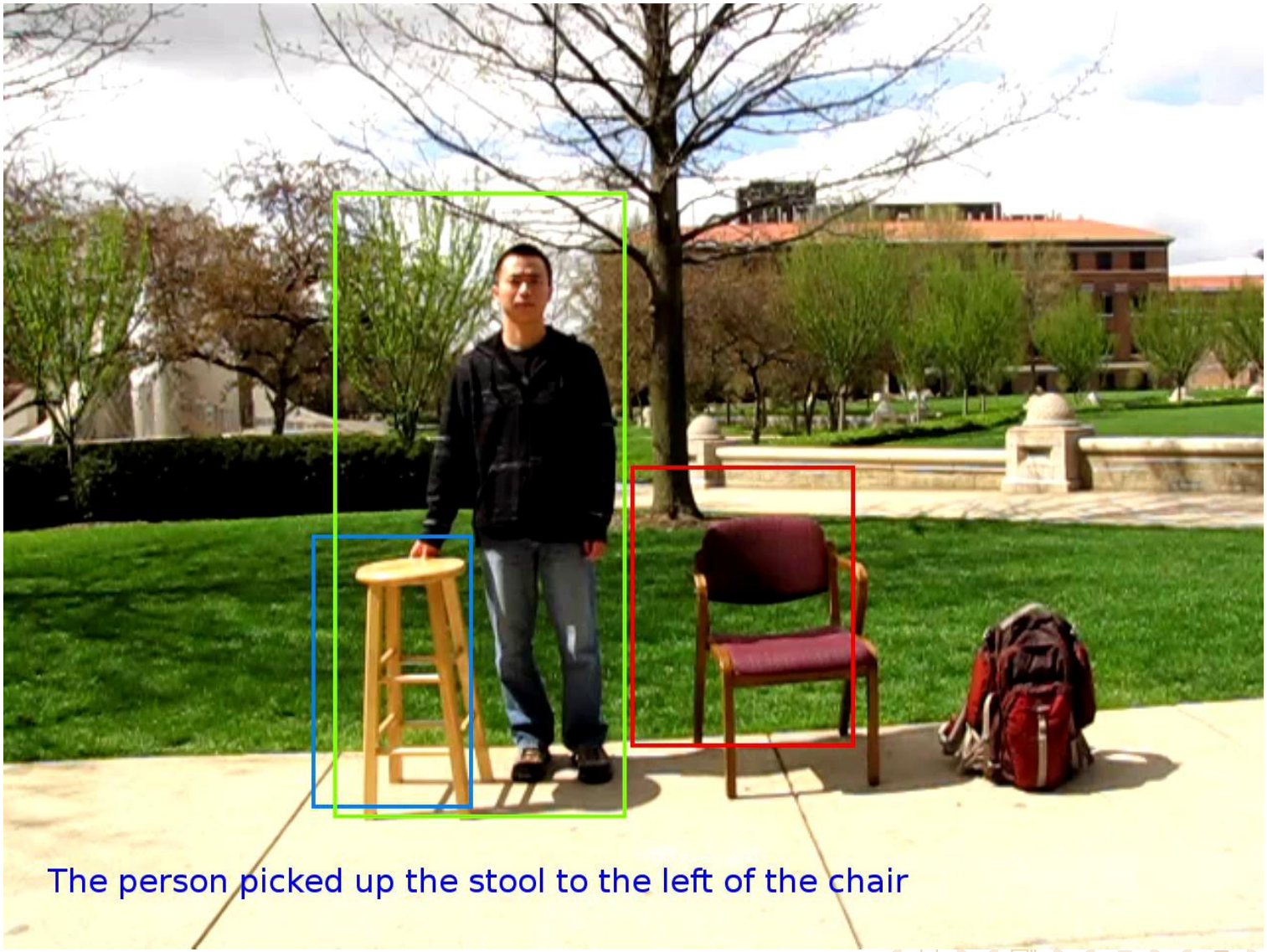


The trash-can approached the traffic-cone.

New Sentence Generation with Trained Models



New Sentence Generation with Trained Models



Outline

1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

Joint work

Haonan Yu

Sentence-Directed Video Object Codetection

Sentence-Directed Video Object Codetection

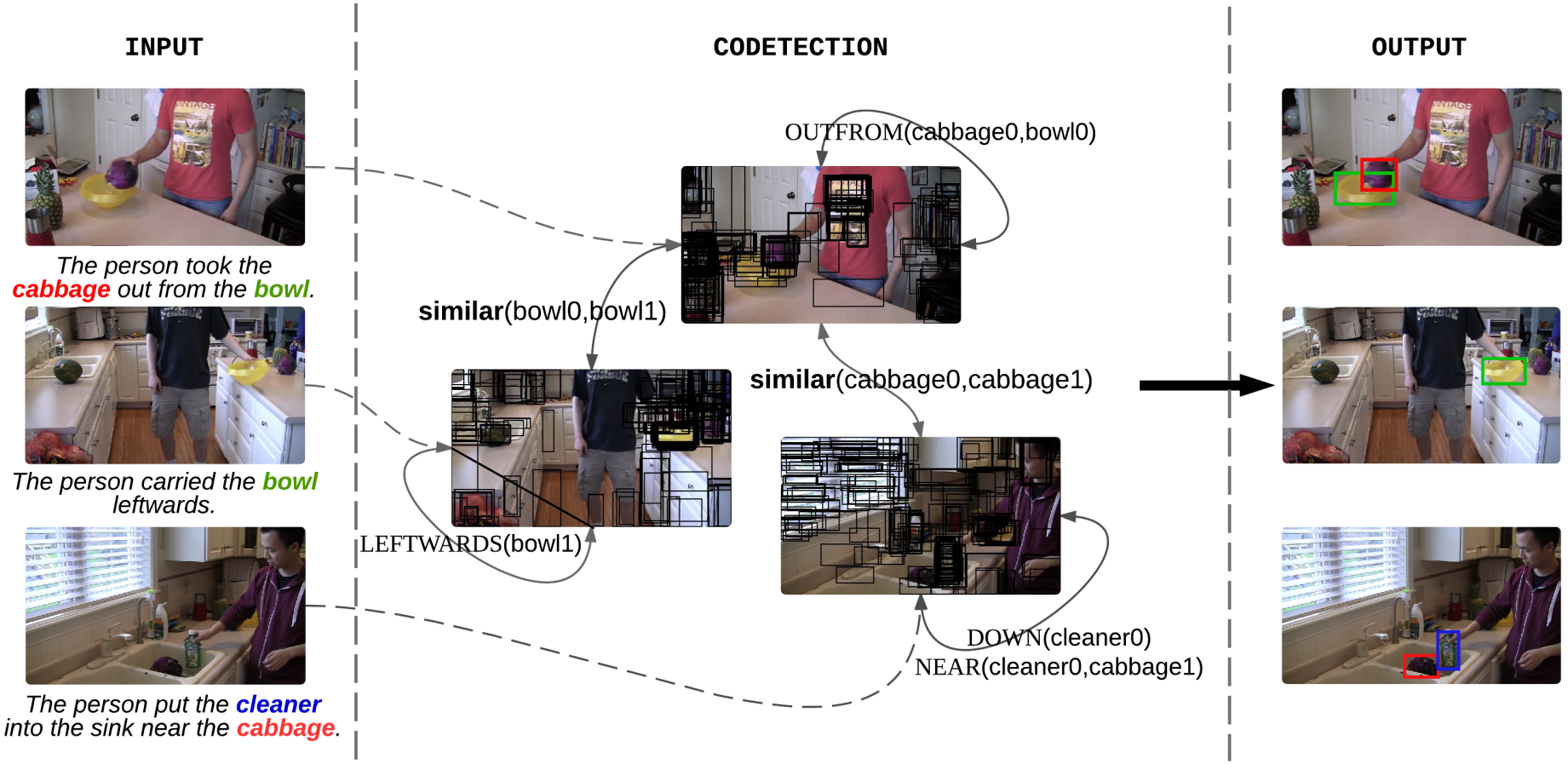
video captioning: video+detections→sentences

Sentence-Directed Video Object Codetection

video captioning: video+detections→sentences

inverse video captioning: video+sentences→detections

Overview


INPUT
CODETECTION
OUTPUT
The person took the cabbage out from the bowl.
The person carried the bowl leftwards.
The person put the cleaner into the sink near the cabbage.
OUTFROM(cabbage0,bowl0)
similar(bowl0,bowl1)
similar(cabbage0,cabbage1)
LEFTWARDS(bowl1)
DOWN(cleaner0)
NEAR(cleaner0,cabbage1)

The 7 ‘No’s

The 7 ‘No’s

- no background subtraction

The 7 ‘No’s

- no background subtraction
- no object detector

The 7 ‘No’s

- no background subtraction
- no object detector
- no object models

The 7 ‘No’s

- no background subtraction
- no object detector
- no object models
- no per-object-class parameters

The 7 ‘No’s

- no background subtraction
- no object detector
- no object models
- no per-object-class parameters
- no learning

The 7 'No's

- no background subtraction
- no object detector
- no object models
- no per-object-class parameters
- no learning
- no training data

The 7 'No's

- no background subtraction
- no object detector
- no object models
- no per-object-class parameters
- no learning
- no training data
- no human-annotated bounding boxes

The person put the cleaner into the sink near the cabbage.

The person put the cleaner into the sink near the cabbage.

$$\text{DOWN}(\textit{cleaner}) \wedge \text{NEAR}(\textit{cleaner}, \textit{cabbage})$$

The person put the cleaner into the sink near the cabbage.

$$\text{DOWN}(cleaner) \wedge \text{NEAR}(cleaner, cabbage)$$

Generated using Stanford parser (Socher et al. ACL 2013) and methods of Lin et al. (CVPR 2014).

The person put the cleaner into the sink near the cabbage.

$$\text{DOWN}(\textit{cleaner}) \wedge \text{NEAR}(\textit{cleaner}, \textit{cabbage})$$

Generated using Stanford parser (Socher et al. ACL 2013) and methods of Lin et al. (CVPR 2014).
Predicates are soft.

The person put the cleaner into the sink near the cabbage.

$$\text{DOWN}(\textit{cleaner}) \wedge \text{NEAR}(\textit{cleaner}, \textit{cabbage})$$

Generated using Stanford parser (Socher et al. ACL 2013) and methods of Lin et al. (CVPR 2014).
Predicates are soft.
Some are unary, some are binary.

Our Predicates

Predicates	Constants
$\textsc{move}(p) \triangleq \mathsf{medFlMg}(p)$	$\Delta\textsc{distLarge} \triangleq 0.25$
$\textsc{moveUp}(p) \triangleq \textsc{move}(p) + \mathsf{distLessThan}\left(\mathsf{y}(p^{(T)}) - \mathsf{y}(p^{(1)}), -\Delta\textsc{distLarge}\right)$	$\Delta\textsc{distSmall} \triangleq 0.05$
$\textsc{moveDown}(p) \triangleq \textsc{move}(p) + \mathsf{distGreaterThan}\left(\mathsf{y}(p^{(T)}) - \mathsf{y}(p^{(1)}), \Delta\textsc{distLarge}\right)$	$\Delta\textsc{angle} \triangleq \pi/2$
$\textsc{moveHorizontal}(p) \triangleq \textsc{move}(p) + \mathsf{distGreaterThan}\left(\left\lvert\mathsf{x}(p^{(T)}) - \mathsf{x}(p^{(1)})\right\rvert, \Delta\textsc{distLarge}\right)$	
$\textsc{moveLeftwards}(p) \triangleq \textsc{move}(p) + \mathsf{distLessThan}\left(\mathsf{x}(p^{(T)}) - \mathsf{x}(p^{(1)}), -\Delta\textsc{distLarge}\right)$	
$\textsc{moveRightwards}(p) \triangleq \textsc{move}(p) + \mathsf{distGreaterThan}\left(\mathsf{x}(p^{(T)}) - \mathsf{x}(p^{(1)}), \Delta\textsc{distLarge}\right)$	
$\textsc{rotate}(p) \triangleq \textsc{move}(p) + \max_t \mathsf{hasRotation}\left(\mathsf{rotAngle}(p^{(t)}), \Delta\textsc{angle}\right)$	
$\textsc{towards}(p_1,p_2) \triangleq \textsc{move}(p_1) + \mathsf{distLessThan}\left(\mathsf{dist}(p_1^{(T)},p_2^{(T)}) - \mathsf{dist}(p_1^{(1)},p_2^{(1)}), -\Delta\textsc{distLarge}\right)$	
$\textsc{awayFrom}(p_1,p_2) \triangleq \textsc{move}(p_1) + \mathsf{distGreaterThan}\left(\mathsf{dist}(p_1^{(T)},p_2^{(T)}) - \mathsf{dist}(p_1^{(1)},p_2^{(1)}), \Delta\textsc{distLarge}\right)$	
$\textsc{leftOfStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{x}(p_1^{(1)}) - \mathsf{x}(p_2^{(1)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{leftOfEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{x}(p_1^{(T)}) - \mathsf{x}(p_2^{(T)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{rightOfStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{x}(p_1^{(1)}) - \mathsf{x}(p_2^{(1)}), \Delta\textsc{distSmall}\right)$	
$\textsc{rightOfEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{x}(p_1^{(T)}) - \mathsf{x}(p_2^{(T)}), \Delta\textsc{distSmall}\right)$	
$\textsc{onTopOfStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2)$	
$+\mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), -2\Delta\textsc{distLarge}\right)$	
$+\mathsf{distLessThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), 0\right)$	
$+\mathsf{distLessThan}\left(\left\lvert\mathsf{x}(p_1^{(1)}) - \mathsf{x}(p_2^{(1)})\right\rvert, 2\Delta\textsc{distSmall}\right)$	
$\textsc{onTopOfEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2)$	
$+\mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), -2\Delta\textsc{distLarge}\right)$	
$+\mathsf{distLessThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), 0\right)$	
$+\mathsf{distLessThan}\left(\left\lvert\mathsf{x}(p_1^{(T)}) - \mathsf{x}(p_2^{(T)})\right\rvert, 2\Delta\textsc{distSmall}\right)$	
$\textsc{nearStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{dist}(p_1^{(1)},p_2^{(1)}), 2\Delta\textsc{distSmall}\right)$	
$\textsc{nearEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{dist}(p_1^{(T)},p_2^{(T)}), 2\Delta\textsc{distSmall}\right)$	
$\textsc{inStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \textsc{nearStart}(p_1,p_2) + \mathsf{smaller}(p_1^{(1)},p_2^{(1)})$	
$\textsc{inEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \textsc{nearEnd}(p_1,p_2) + \mathsf{smaller}(p_1^{(T)},p_2^{(T)})$	
$\textsc{belowStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), \Delta\textsc{distSmall}\right)$	
$\textsc{belowEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), \Delta\textsc{distSmall}\right)$	
$\textsc{aboveStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{aboveEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{over}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2)$	
$+\max_t \begin{pmatrix} \mathsf{distLessThan}\left(\mathsf{y}(p_1^{(t)}) - \mathsf{y}(p_2^{(t)}), -\Delta\textsc{distSmall}\right) \\ \mathsf{distLessThan}\left(\left\lvert\mathsf{x}(p_1^{(t)}) - \mathsf{x}(p_2^{(t)})\right\rvert, \Delta\textsc{distLarge}\right) \end{pmatrix}$	

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model
   - object instances appearing in the sentences as vertices

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model
   - object instances appearing in the sentences as vertices
   - tracks in the video associated with the sentence as vertex labels

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model
   - object instances appearing in the sentences as vertices
   - tracks in the video associated with the sentence as vertex labels
   - edge between two object instances in the same sentence
     clique for all object instances for the same noun

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model
   - object instances appearing in the sentences as vertices
   - tracks in the video associated with the sentence as vertex labels
   - edge between two object instances in the same sentence
     clique for all object instances for the same noun
   - unary predicate score from sentences as vertex score

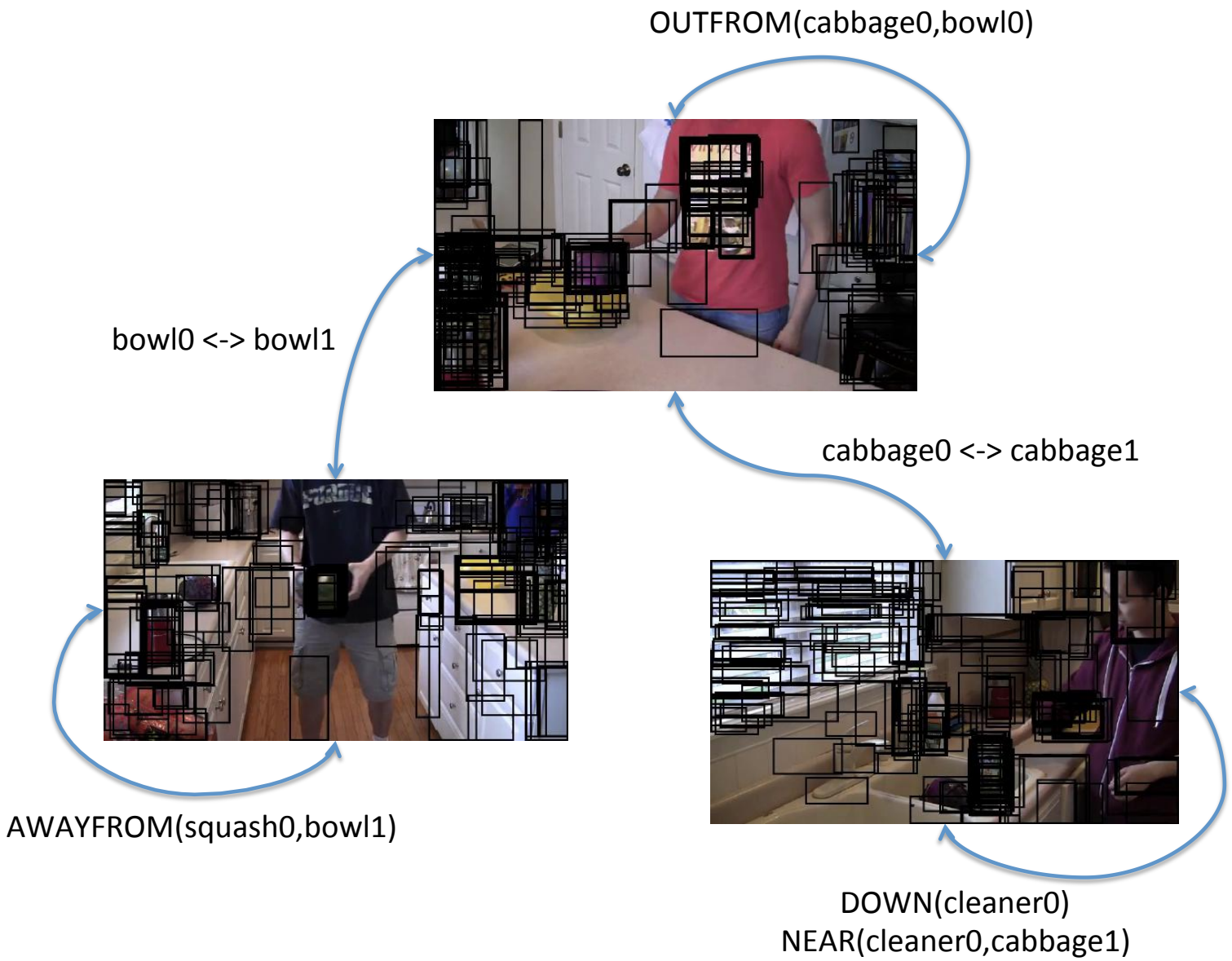
Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model
   - object instances appearing in the sentences as vertices
   - tracks in the video associated with the sentence as vertex labels
   - edge between two object instances in the same sentence
     clique for all object instances for the same noun
   - unary predicate score from sentences as vertex score
   - binary predicate score from sentences as edge score
     similarity score as edge score

Method

1. generate proposals with EdgeBoxes (Zitnick et al. ECCV 2014) and MCG (Arbelaez et al. CVPR 2014)
2. sample MOVING and STATIONARY proposals from sampled frames
3. track sampled MOVING proposal with CamShift (Bradski 1998) in HSV and STATIONARY proposals with MeanShift (Comaniciu et al. 2000) in RGB forward and backward over whole clip
4. rotate proposal multiples of 90°
5. graphical model
   - object instances appearing in the sentences as vertices
   - tracks in the video associated with the sentence as vertex labels
   - edge between two object instances in the same sentence
     clique for all object instances for the same noun
   - unary predicate score from sentences as vertex score
   - binary predicate score from sentences as edge score
     similarity score as edge score
     - χ^2 of PHOW (Bosch et al. ICCV 2007) and L_2 HOG (Dalal & Triggs CVPR 2005) to measure similarity

Method


OUTFROM(cabbage0,bowl0)
bowl0 <-> bowl1
cabbage0 <-> cabbage1
AWAYFROM(squash0,bowl1)
DOWN(cleaner0)
NEAR(cleaner0,cabbage1)

Four Variants

	SIM (variant 1)	**FLOW** (variant 2)	**SIM+FLOW** (variant 3)	**SENT** (variant 4)	**SIM+SENT** (our full method)
Similarity score?	yes	no	yes	no	yes
Sentence score?	no	partial	partial	yes	yes

partial: motion and temporal coherence but no other components of sentence semantics

Our Predicates

Predicates	Constants
$\textsc{move}(p) \triangleq \mathsf{medFlMg}(p)$	$\Delta\textsc{distLarge} \triangleq 0.25$
$\textsc{moveUp}(p) \triangleq \textsc{move}(p) + \mathsf{distLessThan}\left(\mathsf{y}(p^{(T)}) - \mathsf{y}(p^{(1)}), -\Delta\textsc{distLarge}\right)$	$\Delta\textsc{distSmall} \triangleq 0.05$
$\textsc{moveDown}(p) \triangleq \textsc{move}(p) + \mathsf{distGreaterThan}\left(\mathsf{y}(p^{(T)}) - \mathsf{y}(p^{(1)}), \Delta\textsc{distLarge}\right)$	$\Delta\textsc{angle} \triangleq \pi/2$
$\textsc{moveHorizontal}(p) \triangleq \textsc{move}(p) + \mathsf{distGreaterThan}\left(\left\lvert\mathsf{x}(p^{(T)}) - \mathsf{x}(p^{(1)})\right\rvert, \Delta\textsc{distLarge}\right)$	
$\textsc{moveLeftwards}(p) \triangleq \textsc{move}(p) + \mathsf{distLessThan}\left(\mathsf{x}(p^{(T)}) - \mathsf{x}(p^{(1)}), -\Delta\textsc{distLarge}\right)$	
$\textsc{moveRightwards}(p) \triangleq \textsc{move}(p) + \mathsf{distGreaterThan}\left(\mathsf{x}(p^{(T)}) - \mathsf{x}(p^{(1)}), \Delta\textsc{distLarge}\right)$	
$\textsc{rotate}(p) \triangleq \textsc{move}(p) + \max_t \mathsf{hasRotation}\left(\mathsf{rotAngle}(p^{(t)}), \Delta\textsc{angle}\right)$	
$\textsc{towards}(p_1,p_2) \triangleq \textsc{move}(p_1) + \mathsf{distLessThan}\left(\mathsf{dist}(p_1^{(T)},p_2^{(T)}) - \mathsf{dist}(p_1^{(1)},p_2^{(1)}), -\Delta\textsc{distLarge}\right)$	
$\textsc{awayFrom}(p_1,p_2) \triangleq \textsc{move}(p_1) + \mathsf{distGreaterThan}\left(\mathsf{dist}(p_1^{(T)},p_2^{(T)}) - \mathsf{dist}(p_1^{(1)},p_2^{(1)}), \Delta\textsc{distLarge}\right)$	
$\textsc{leftOfStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{x}(p_1^{(1)}) - \mathsf{x}(p_2^{(1)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{leftOfEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{x}(p_1^{(T)}) - \mathsf{x}(p_2^{(T)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{rightOfStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{x}(p_1^{(1)}) - \mathsf{x}(p_2^{(1)}), \Delta\textsc{distSmall}\right)$	
$\textsc{rightOfEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{x}(p_1^{(T)}) - \mathsf{x}(p_2^{(T)}), \Delta\textsc{distSmall}\right)$	
$\textsc{onTopOfStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2)$	
$+\mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), -2\Delta\textsc{distLarge}\right)$	
$+\mathsf{distLessThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), 0\right)$	
$+\mathsf{distLessThan}\left(\left\lvert\mathsf{x}(p_1^{(1)}) - \mathsf{x}(p_2^{(1)})\right\rvert, 2\Delta\textsc{distSmall}\right)$	
$\textsc{onTopOfEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2)$	
$+\mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), -2\Delta\textsc{distLarge}\right)$	
$+\mathsf{distLessThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), 0\right)$	
$+\mathsf{distLessThan}\left(\left\lvert\mathsf{x}(p_1^{(T)}) - \mathsf{x}(p_2^{(T)})\right\rvert, 2\Delta\textsc{distSmall}\right)$	
$\textsc{nearStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{dist}(p_1^{(1)},p_2^{(1)}), 2\Delta\textsc{distSmall}\right)$	
$\textsc{nearEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{dist}(p_1^{(T)},p_2^{(T)}), 2\Delta\textsc{distSmall}\right)$	
$\textsc{inStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \textsc{nearStart}(p_1,p_2) + \mathsf{smaller}(p_1^{(1)},p_2^{(1)})$	
$\textsc{inEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \textsc{nearEnd}(p_1,p_2) + \mathsf{smaller}(p_1^{(T)},p_2^{(T)})$	
$\textsc{belowStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), \Delta\textsc{distSmall}\right)$	
$\textsc{belowEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distGreaterThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), \Delta\textsc{distSmall}\right)$	
$\textsc{aboveStart}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{y}(p_1^{(1)}) - \mathsf{y}(p_2^{(1)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{aboveEnd}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2) + \mathsf{distLessThan}\left(\mathsf{y}(p_1^{(T)}) - \mathsf{y}(p_2^{(T)}), -\Delta\textsc{distSmall}\right)$	
$\textsc{over}(p_1,p_2) \triangleq \mathsf{tempCoher}(p_2)$	
$+\max_t \begin{pmatrix} \mathsf{distLessThan}\left(\mathsf{y}(p_1^{(t)}) - \mathsf{y}(p_2^{(t)}), -\Delta\textsc{distSmall}\right) \\ \mathsf{distLessThan}\left(\left\lvert\mathsf{x}(p_1^{(t)}) - \mathsf{x}(p_2^{(t)})\right\rvert, \Delta\textsc{distLarge}\right) \end{pmatrix}$	

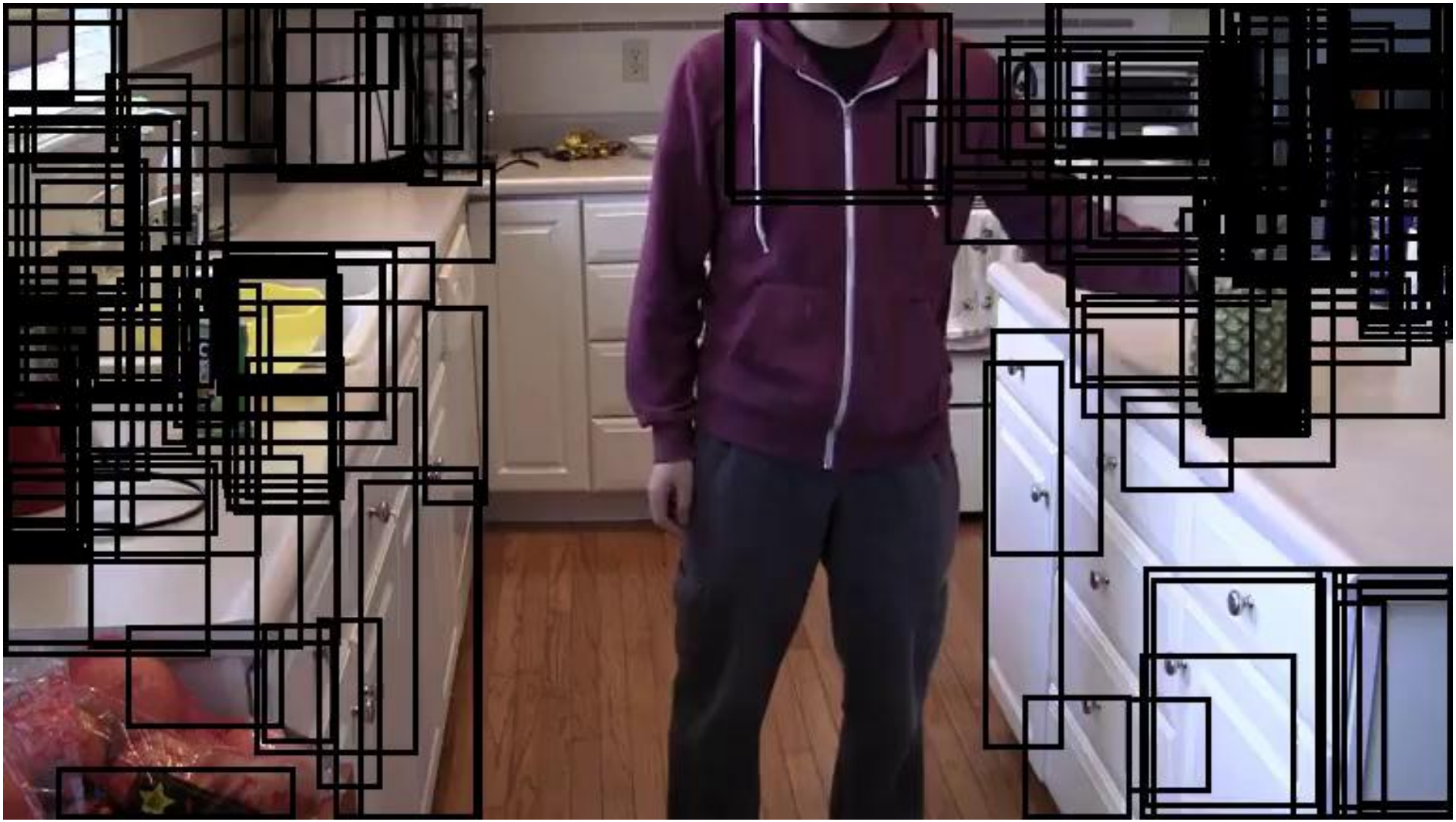
Proposals

The person put the cleaner into the sink near the cabbage.

Proposals

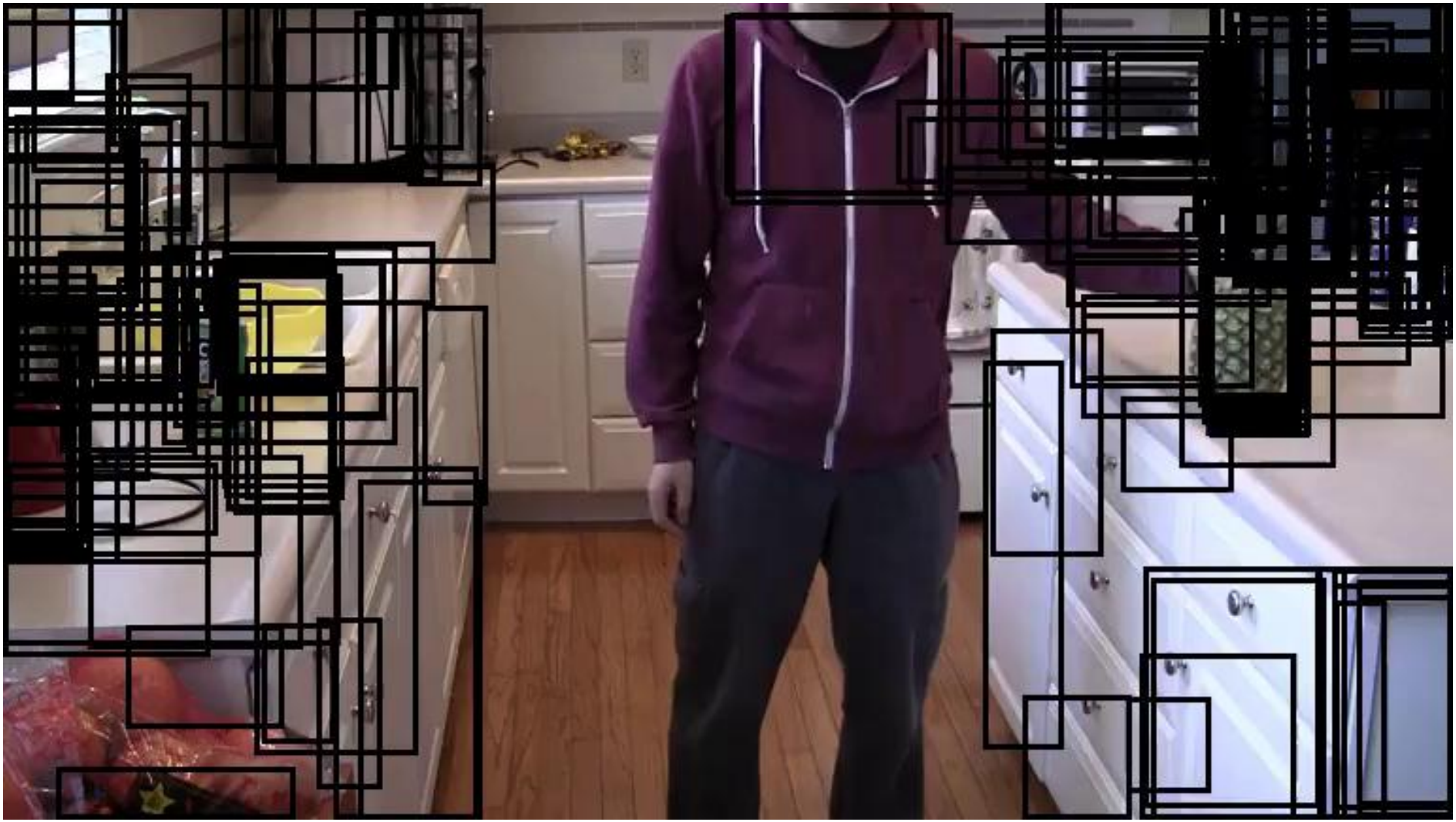
The person put the cleaner into the sink near the cabbage.

Proposals



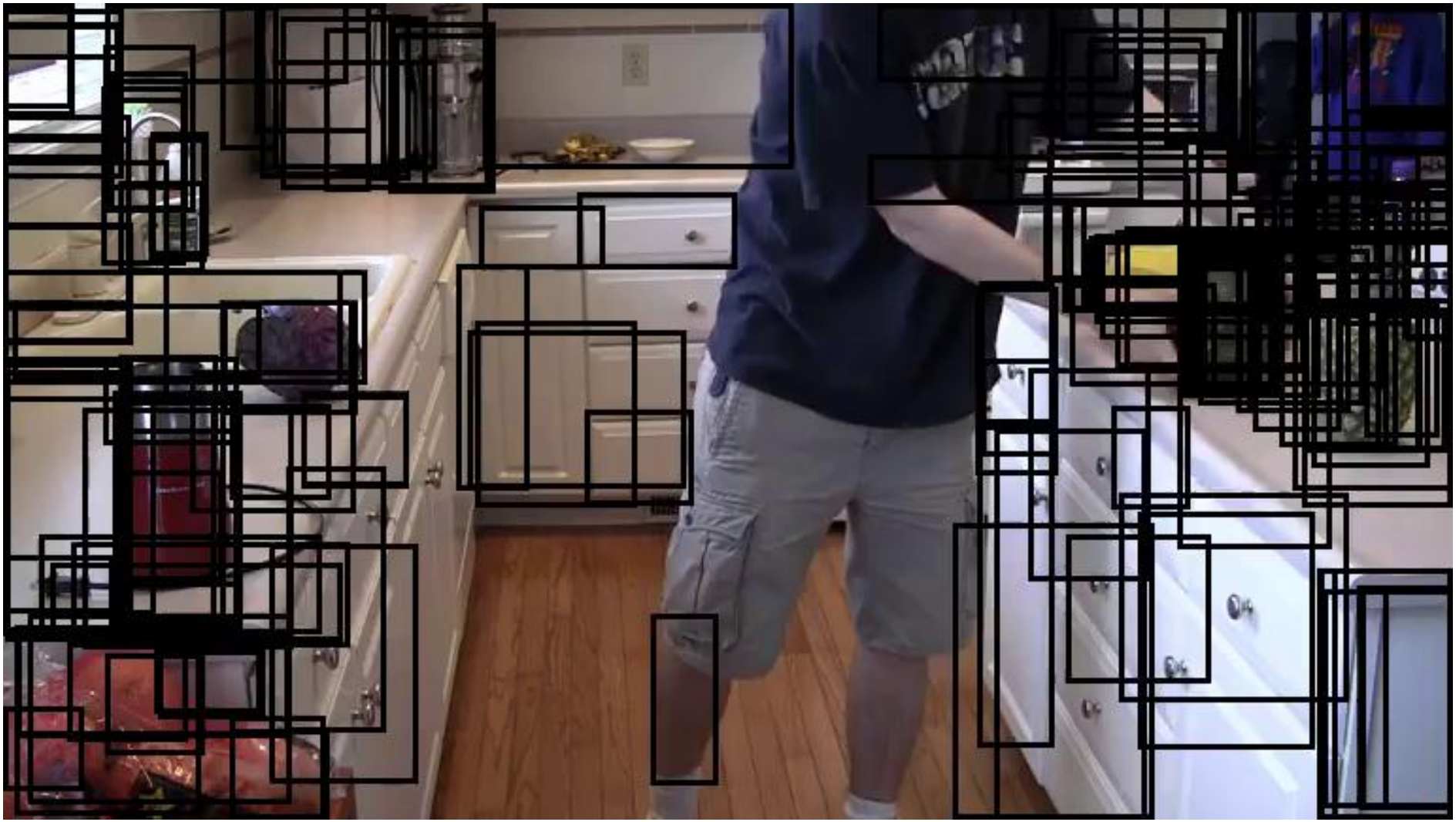
The person carried the pineapple towards the cleaner.

Proposals



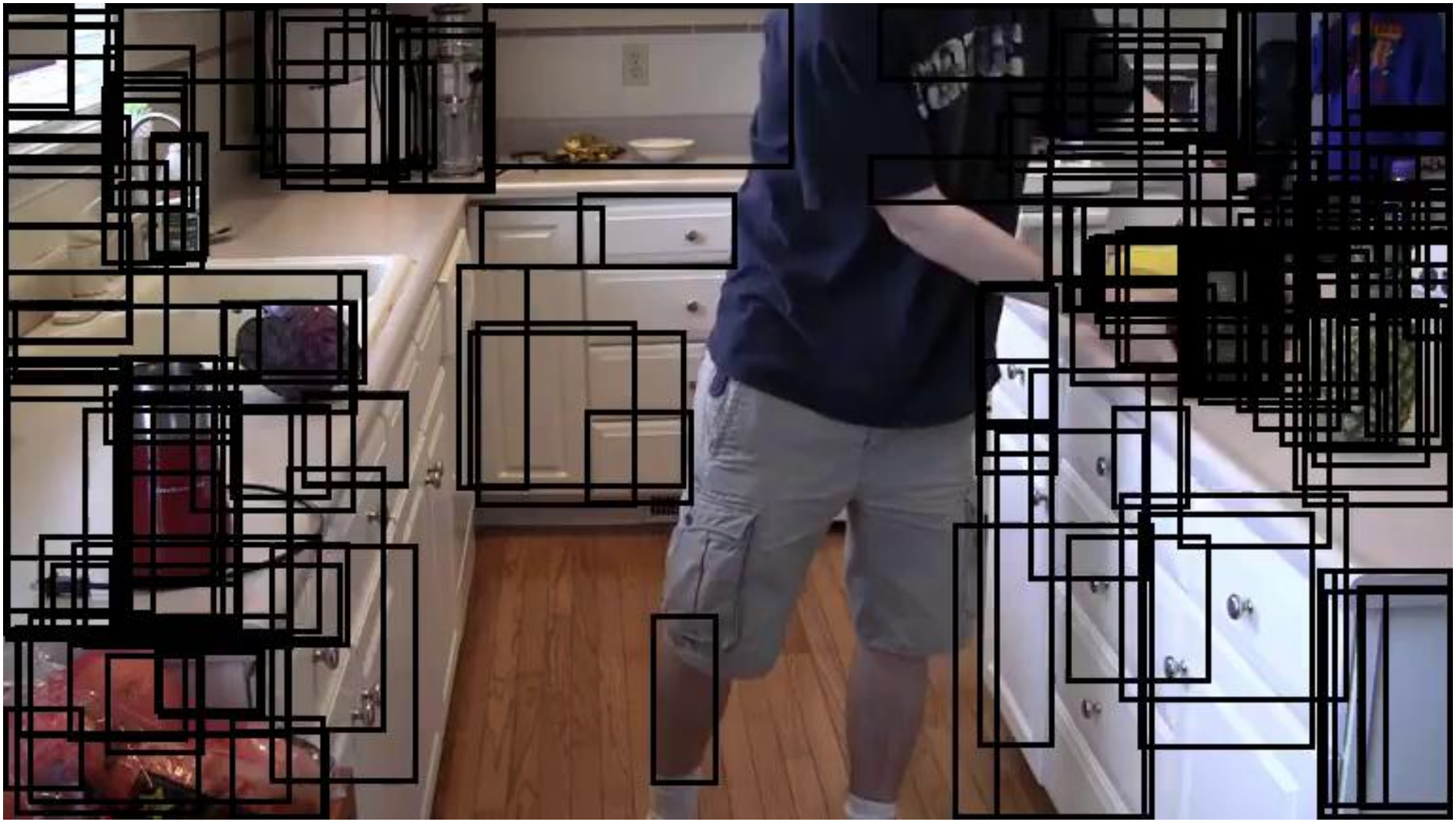
The person carried the pineapple towards the cleaner.

Proposals



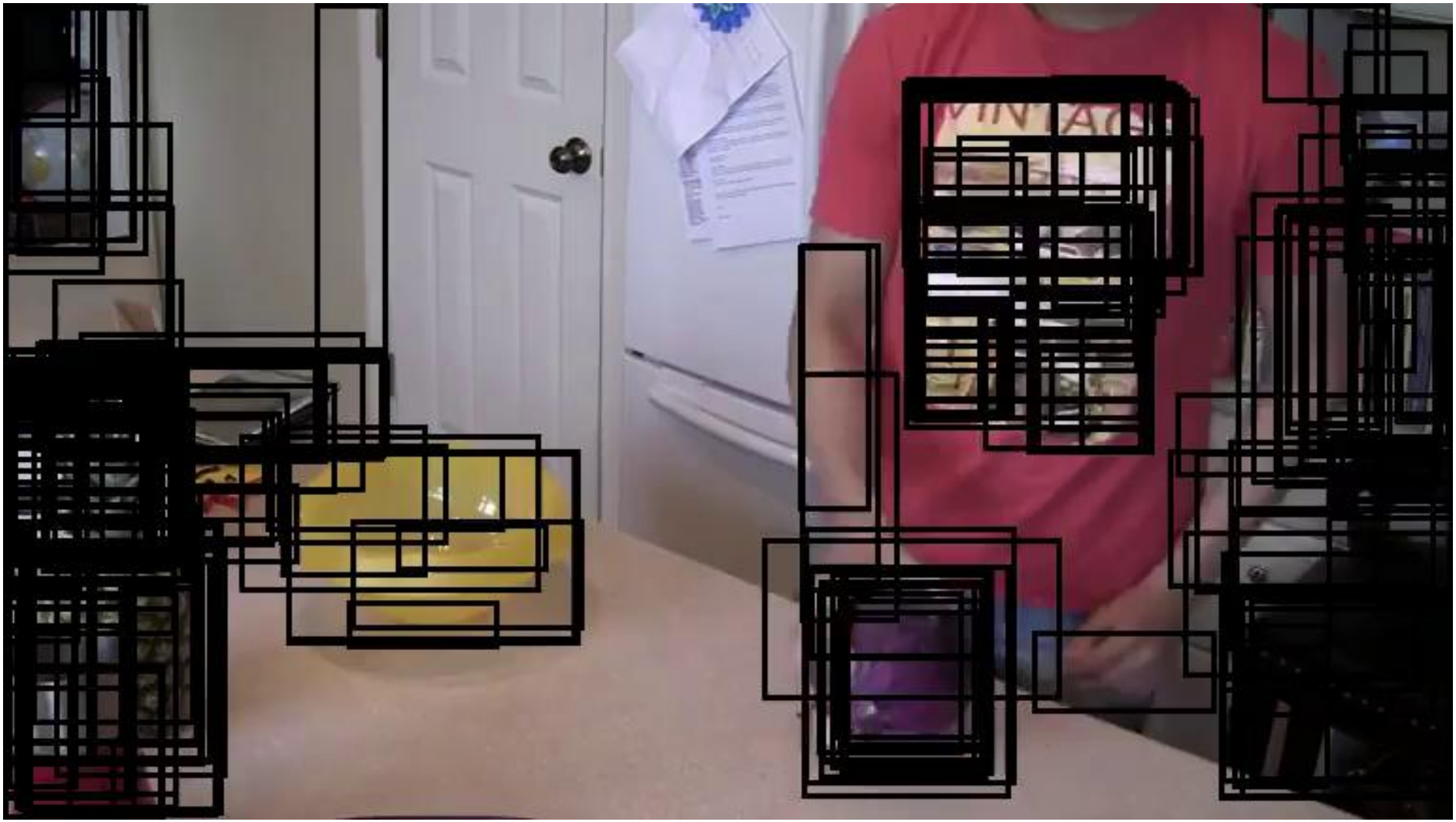
The person too the squash away from the pineapple and put it near the coffee.

Proposals



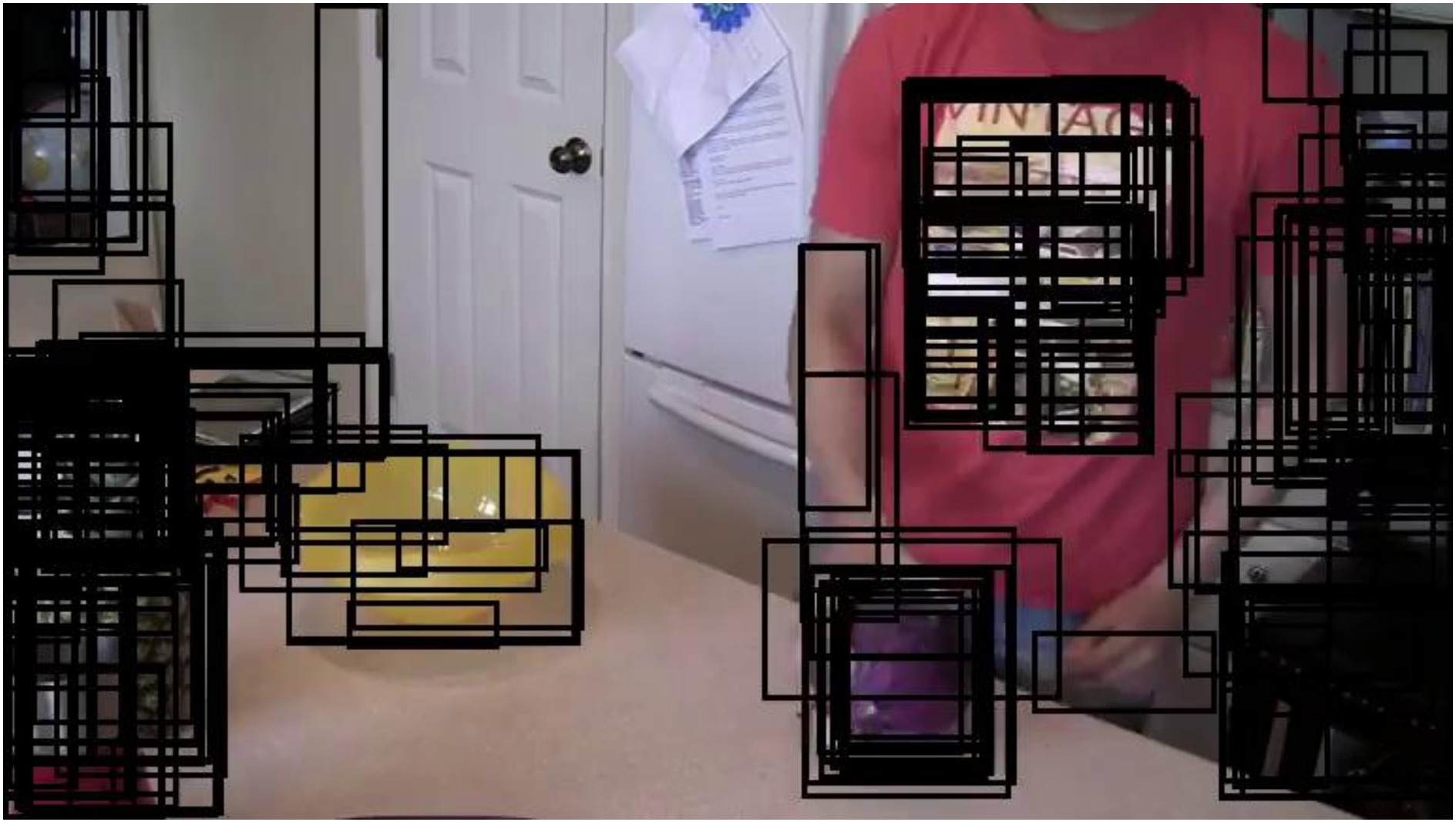
The person too the squash away from the pineapple and put it near the coffee.

Proposals



The person put the cabbage into the bowl.

Proposals



The person put the cabbage into the bowl.

SIM

The person put the cleaner into the sink near the cabbage.

SIM

The person put the cleaner into the sink near the cabbage.

SIM

The person carried the pineapple towards the cleaner.

SIM

The person carried the pineapple towards the cleaner.

SIM

The person too the squash away from the pineapple and put it near the coffee.

SIM

The person too the squash away from the pineapple and put it near the coffee.

The person put the cabbage into the bowl.

The person put the cabbage into the bowl.

FLOW

The person put the cleaner into the sink near the cabbage.

FLOW

The person put the cleaner into the sink near the cabbage.

FLOW

The person carried the pineapple towards the cleaner.

FLOW

The person carried the pineapple towards the cleaner.

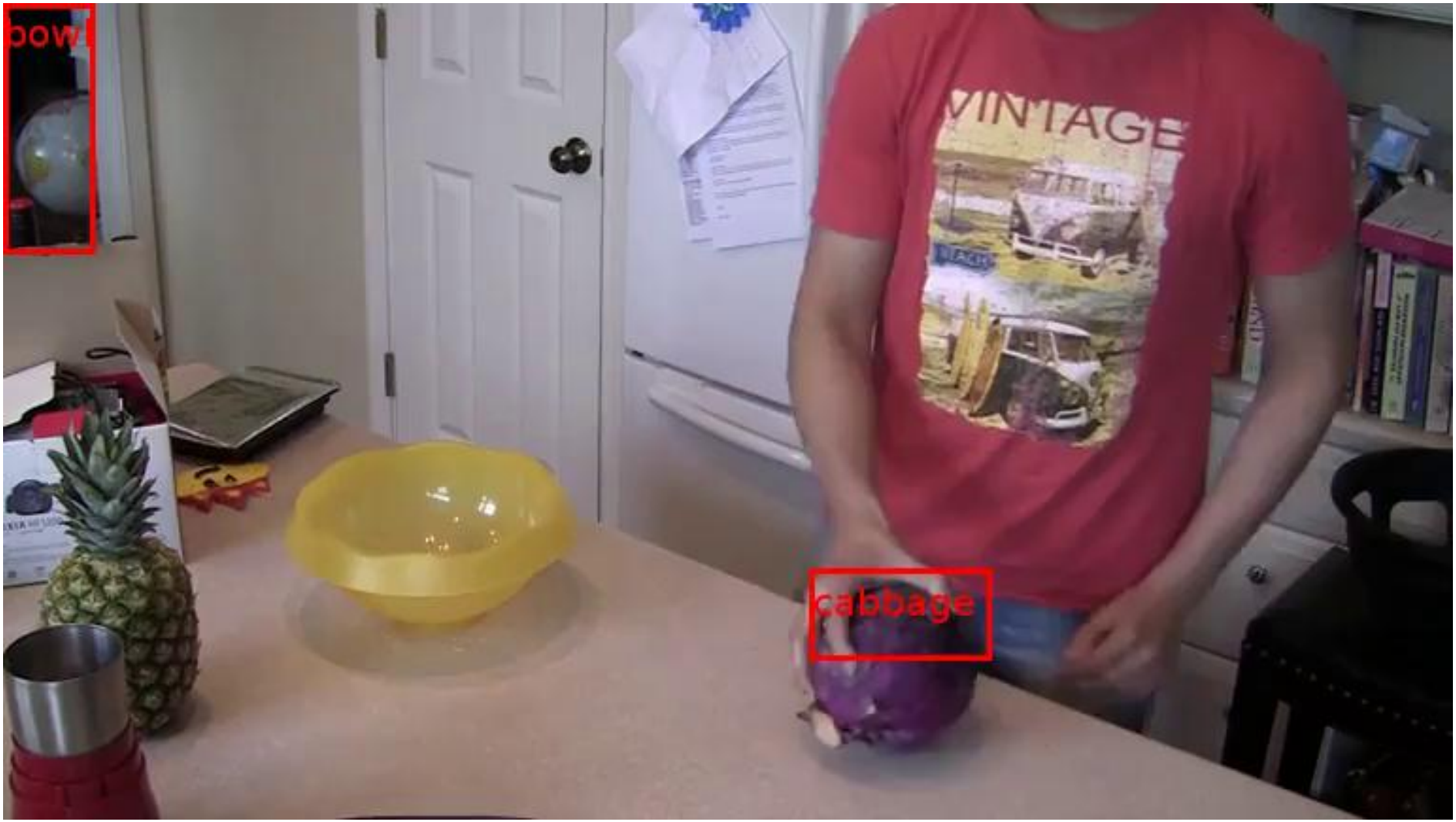
FLOW

The person too the squash away from the pineapple and put it near the coffee.

FLOW

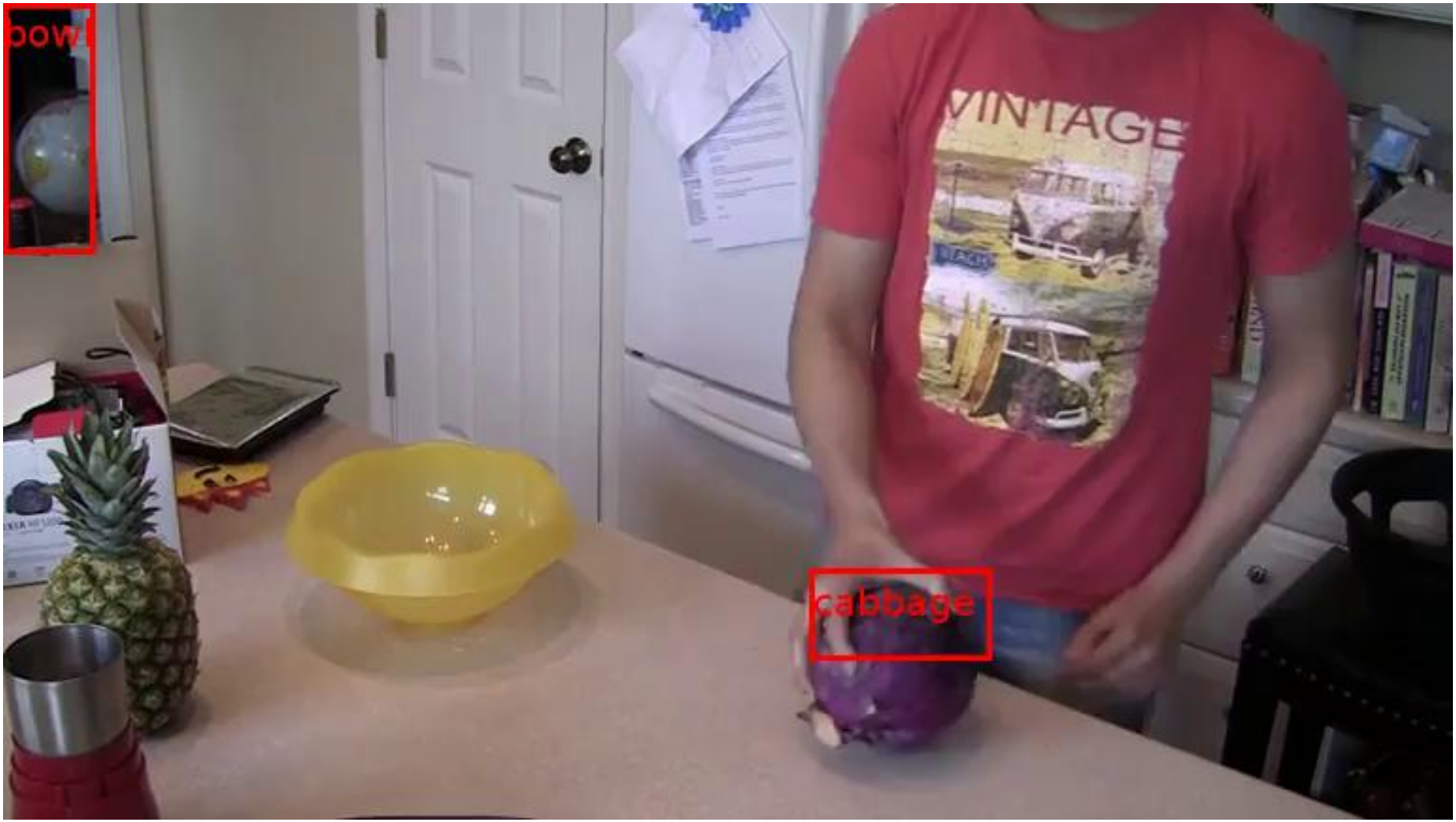
The person too the squash away from the pineapple and put it near the coffee.

FLOW



The person put the cabbage into the bowl.

FLOW



The person put the cabbage into the bowl.

SIM+FLOW

The person put the cleaner into the sink near the cabbage.

SIM+FLOW

The person put the cleaner into the sink near the cabbage.

SIM+FLOW

The person carried the pineapple towards the cleaner.

SIM+FLOW

The person carried the pineapple towards the cleaner.

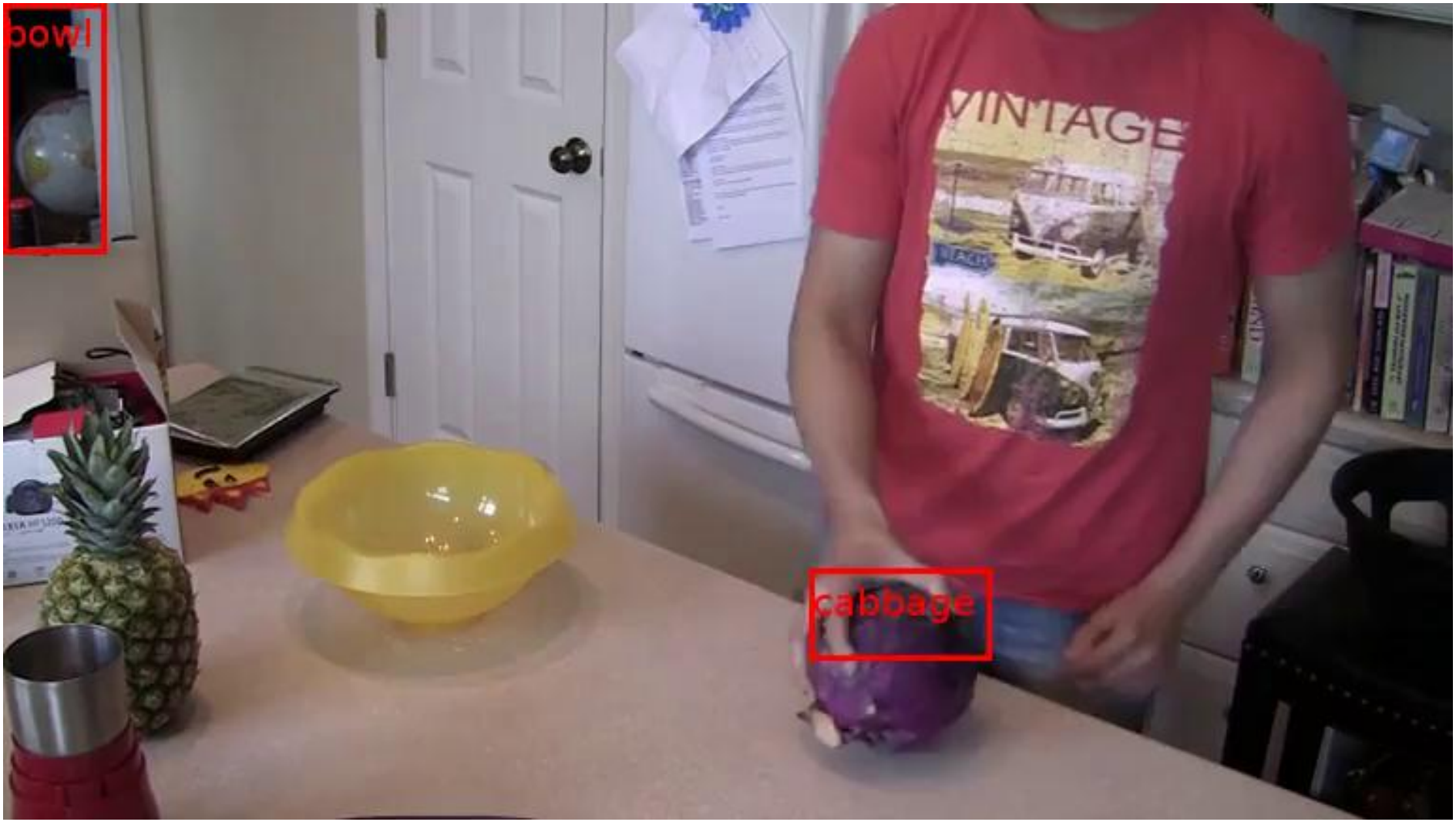
SIM+FLOW

The person too the squash away from the pineapple and put it near the coffee.

SIM+FLOW

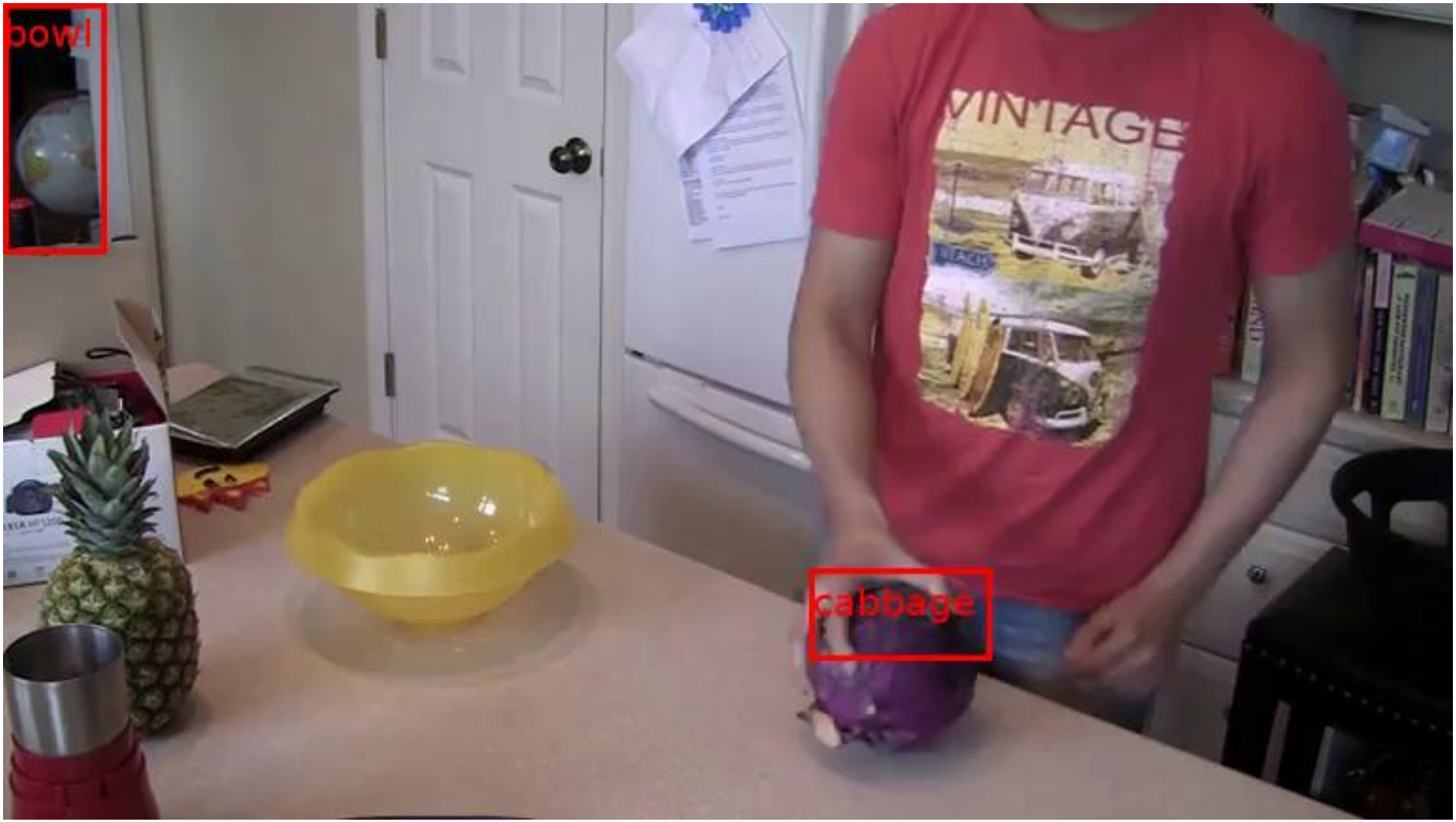
The person too the squash away from the pineapple and put it near the coffee.

SIM+FLOW



The person put the cabbage into the bowl.

SIM+FLOW



The person put the cabbage into the bowl.

SENT

The person put the cleaner into the sink near the cabbage.

SENT

The person put the cleaner into the sink near the cabbage.

SENT

The person carried the pineapple towards the cleaner.

SENT

The person carried the pineapple towards the cleaner.

SENT

The person too the squash away from the pineapple and put it near the coffee.

SENT

The person too the squash away from the pineapple and put it near the coffee.

SENT

The person put the cabbage into the bowl.

SENT

The person put the cabbage into the bowl.

SIM+SENT

The person put the cleaner into the sink near the cabbage.

SIM+SENT

The person put the cleaner into the sink near the cabbage.

SIM+SENT

The person carried the pineapple towards the cleaner.

SIM+SENT

The person carried the pineapple towards the cleaner.

SIM+SENT

The person too the squash away from the pineapple and put it near the coffee.

SIM+SENT

The person too the squash away from the pineapple and put it near the coffee.

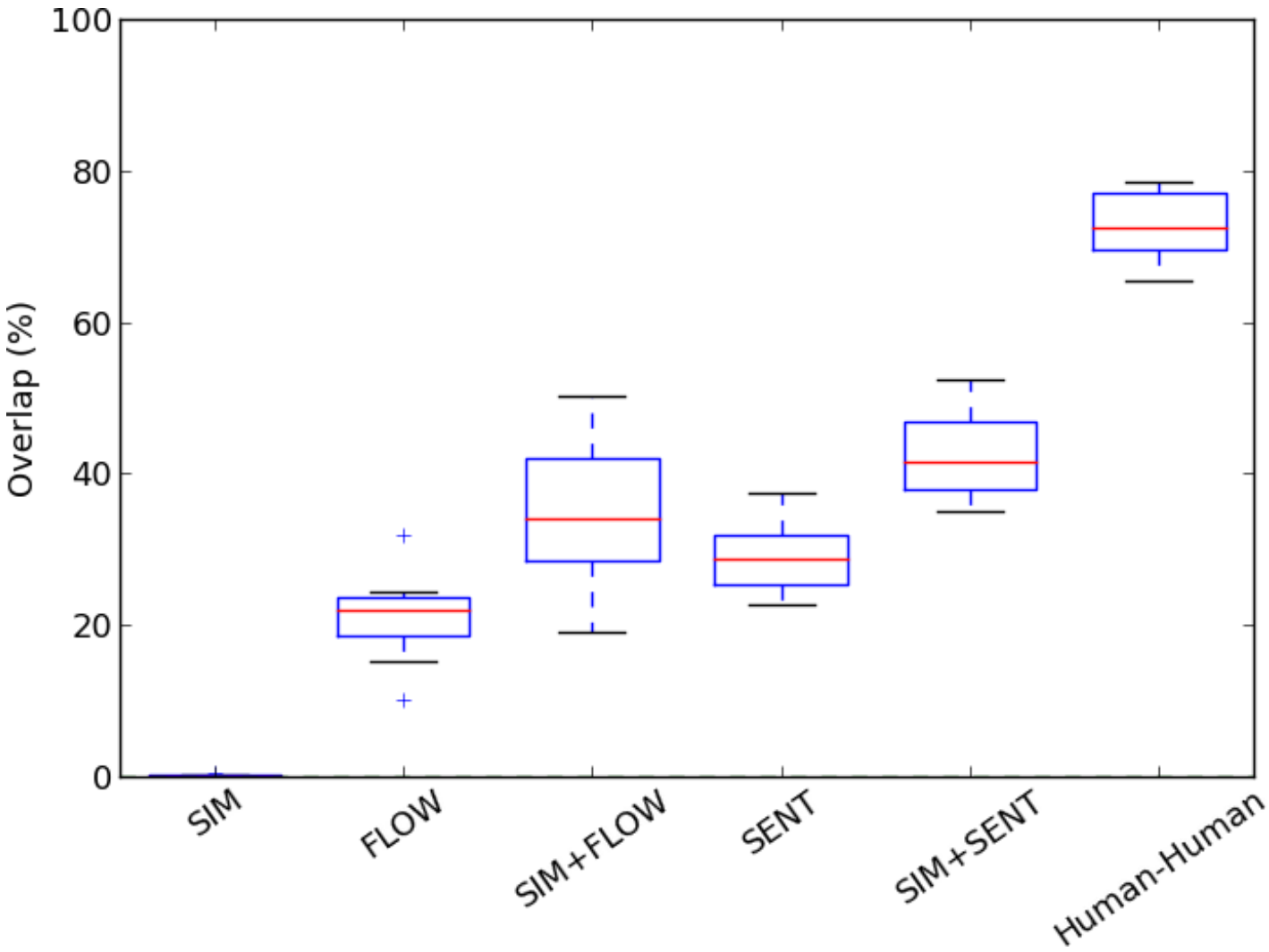
SIM+SENT

The person put the cabbage into the bowl.

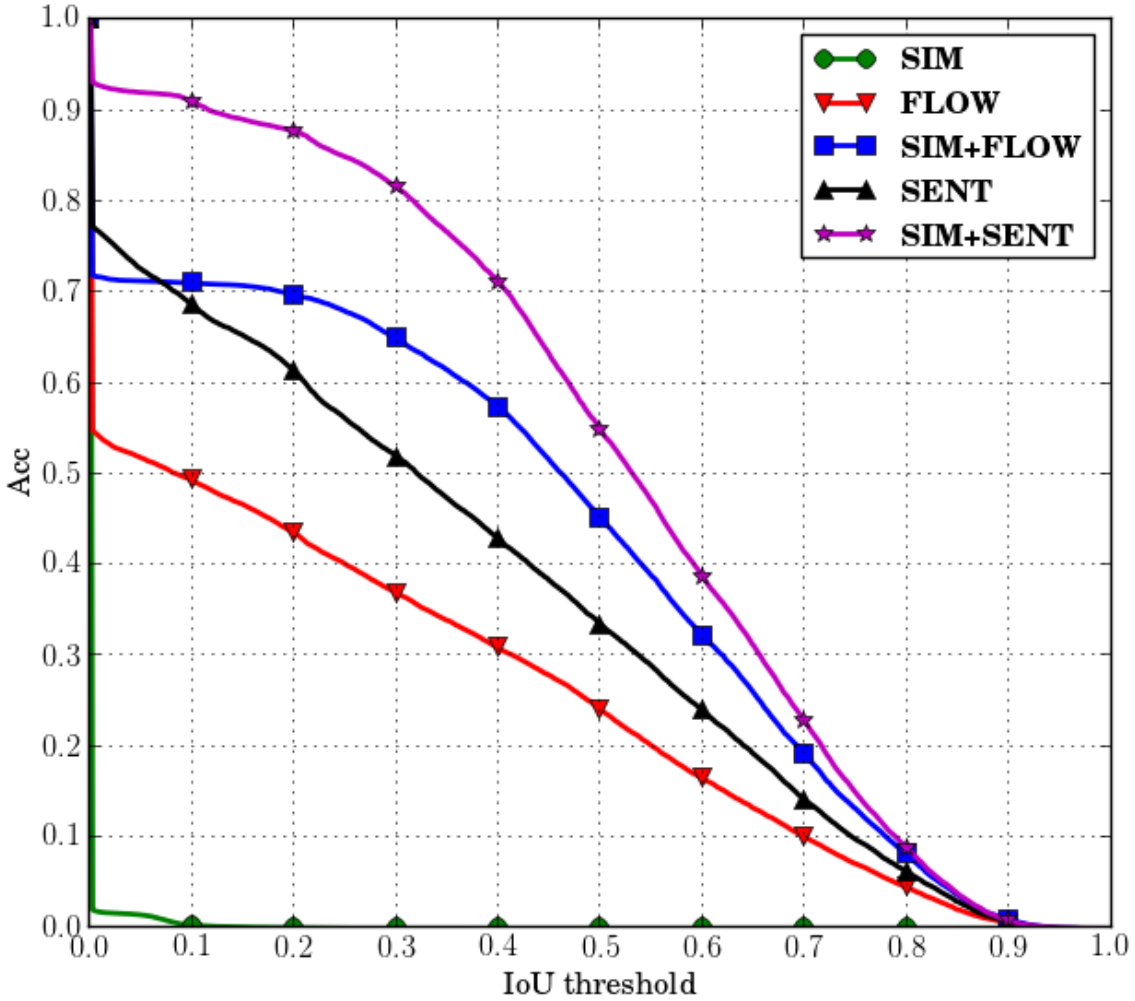
SIM+SENT

The person put the cabbage into the bowl.

IoU Scores


100
80
60
40
20
0
Overlap (%)
SIM
FLOW
SIM+FLOW
SENT
SIM+SENT
Human-Human

Codetection Accuracy



More Examples

More Examples

Outline

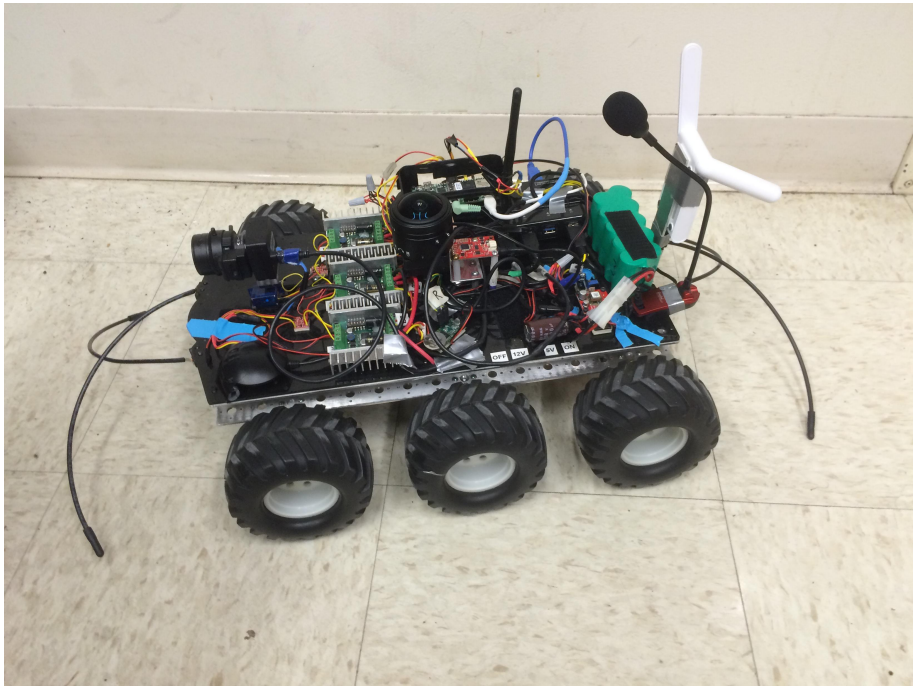
1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

Joint work

Daniel Paul Barrett Scott Alan Bronikowski Haonan Yu

Our Custom Mobile Robot

- IMU (3-axis accelerometers, gyros, and magnetometers)
- GPS
- 6 independently controllable wheel motors
- 2 shaft encoders with Teensy controller
- Gumstix Overo FireSTORM + Summit running Linux
- Bluetooth, WiFi, and 4G LTE
- front and rear bump sensors
- ultrasonic rangefinder
- pan-tilt front-facing camera (Point Grey)
- omnidirectional camera (Point Grey)
- audio input and output
- touchscreen
- Logitech Wireless Gamepad
- custom firmware on IMU and Teensy
- synchronized timestamped logging of sensor and control data



Outline

1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

Grounding Language Semantics in Robotics

$$\mathcal{R} : (\mathbf{s}, \mathbf{p}, \mathbf{f}, \Lambda) \mapsto \tau$$

- **s**: sentence
- **p**: path
- **f**: floorplan
- Λ: lexicon
- τ: score

Three Uses of the Unified Scoring Function

Three Uses of the Unified Scoring Function

- **Language Acquisition**: sentence $\times$ path $\rightarrow$ lexicon

$$\arg \max_{\Lambda} \sum_{i=1} \mathcal{R}(\mathbf{s}_i, \mathbf{p}_i, \mathbf{f}_i, \Lambda)$$

Three Uses of the Unified Scoring Function

- **Language Acquisition**: sentence $\times$ path $\rightarrow$ lexicon

$$\arg \max_{\Lambda} \sum_{i=1} \mathcal{R}(\mathbf{s}_i, \mathbf{p}_i, \mathbf{f}_i, \Lambda)$$

- **Language Generation**: path $\times$ lexicon $\rightarrow$ sentence

$$\arg \max_{\mathbf{s}} \mathcal{R}(\mathbf{s}, \mathbf{p}, \mathbf{f}, \Lambda)$$

Three Uses of the Unified Scoring Function

- **Language Acquisition**: sentence $\times$ path $\rightarrow$ lexicon

$$\arg \max_{\Lambda} \sum_{i=1} \mathcal{R}(\mathbf{s}_i, \mathbf{p}_i, \mathbf{f}_i, \Lambda)$$

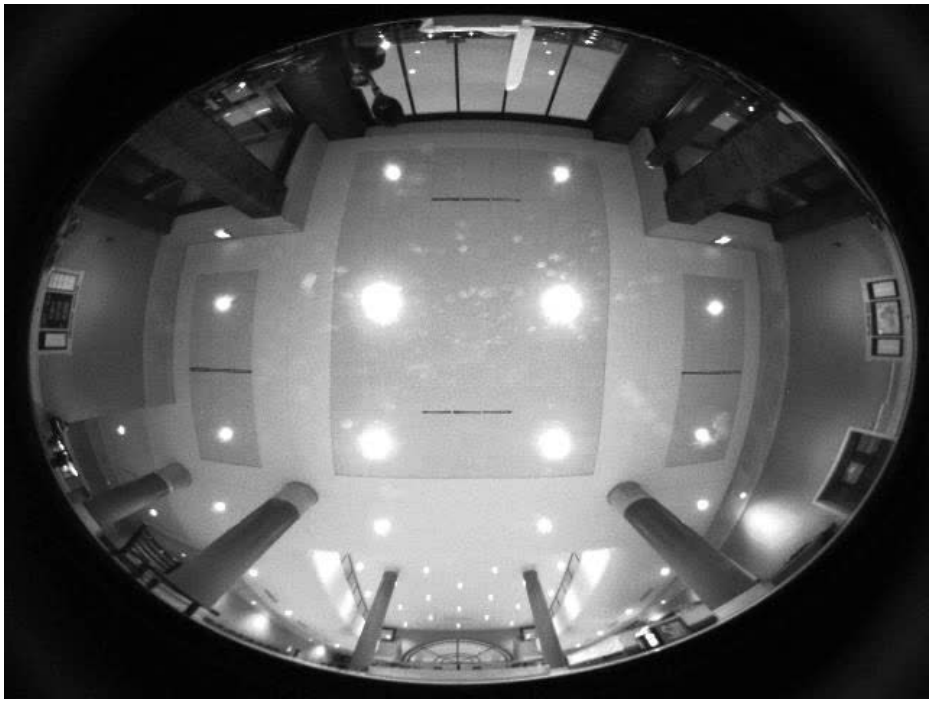
- **Language Generation**: path $\times$ lexicon $\rightarrow$ sentence

$$\arg \max_{\mathbf{s}} \mathcal{R}(\mathbf{s}, \mathbf{p}, \mathbf{f}, \Lambda)$$

- **Language Comprehension**: sentence $\times$ lexicon $\rightarrow$ path

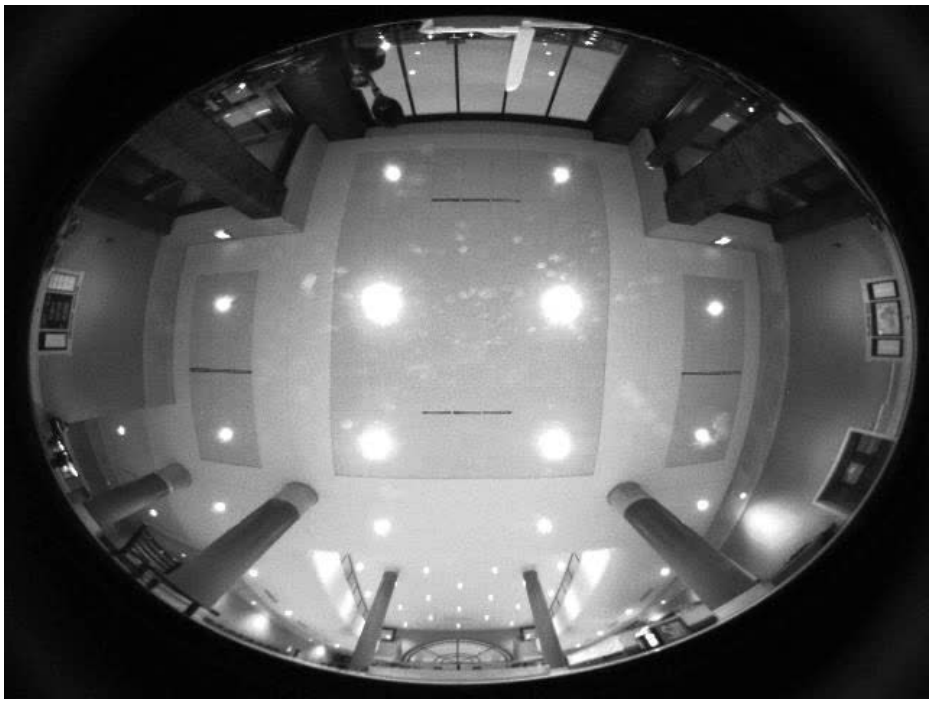
$$\arg \max_{\mathbf{p}} \mathcal{R}(\mathbf{s}, \mathbf{p}, \mathbf{f}, \Lambda)$$

Language Acquisition



The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

Language Acquisition



The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

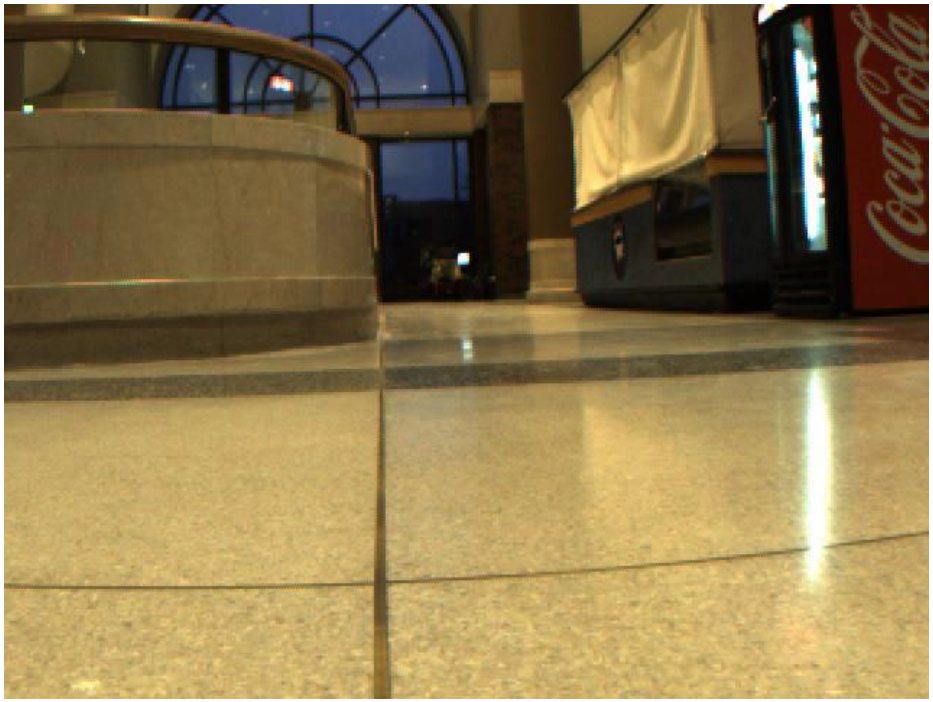
Language Acquisition

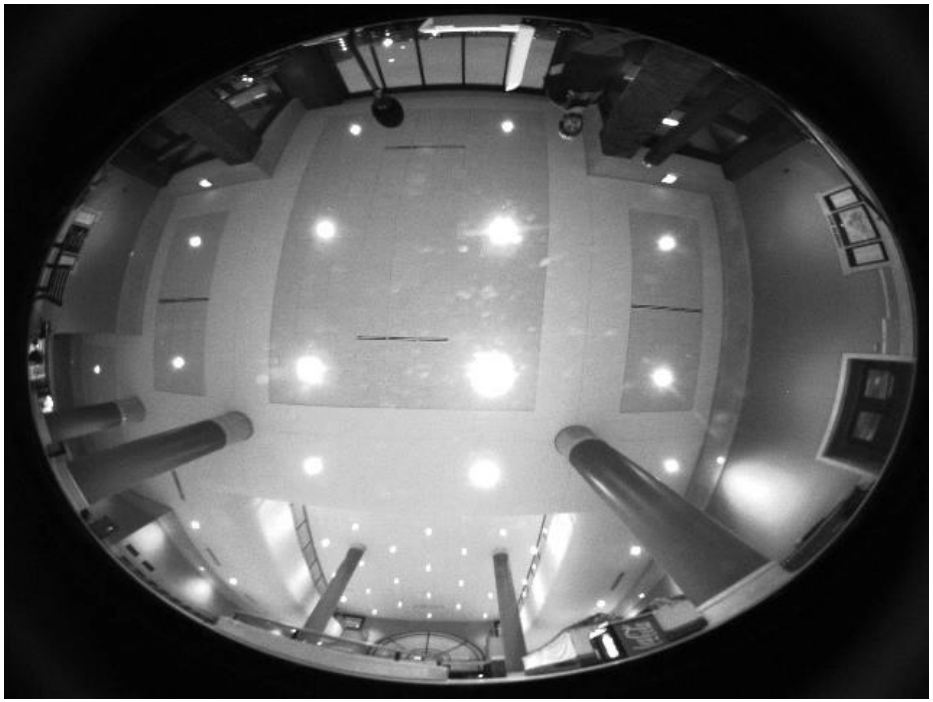
The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

Language Acquisition

The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

Language Acquisition

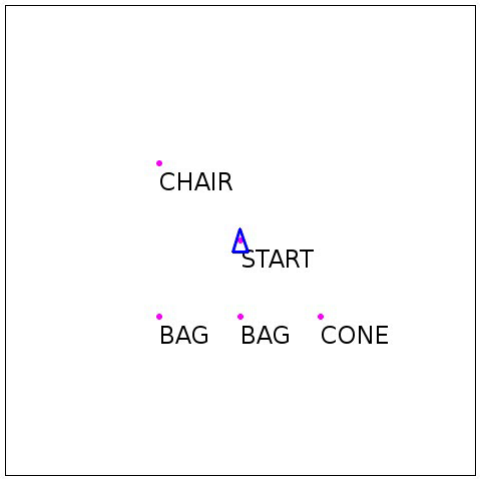




The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

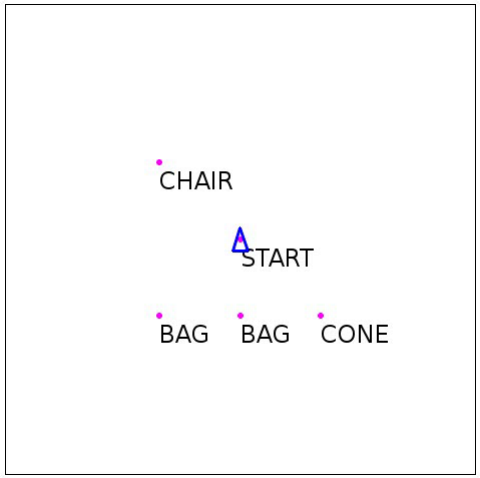
Language Acquisition

input: *The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.*



Language Acquisition

input: *The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.*



Language Acquisition

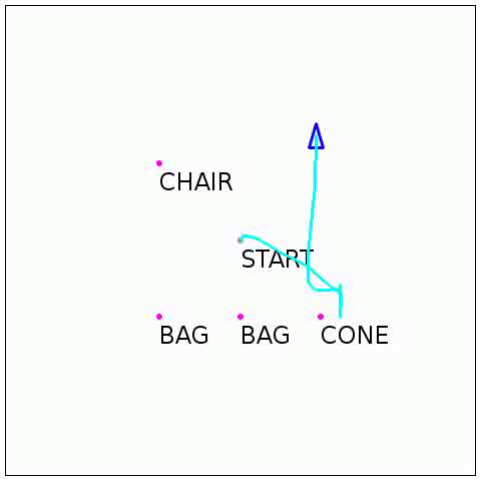
input: *The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.*

Language Acquisition

input: *The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.*

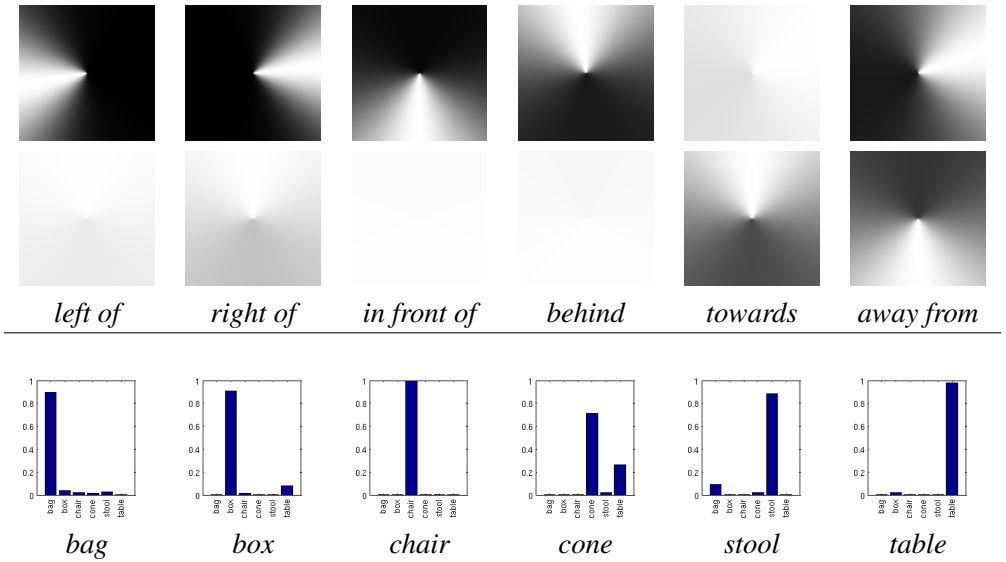
Language Acquisition

input: *The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.*

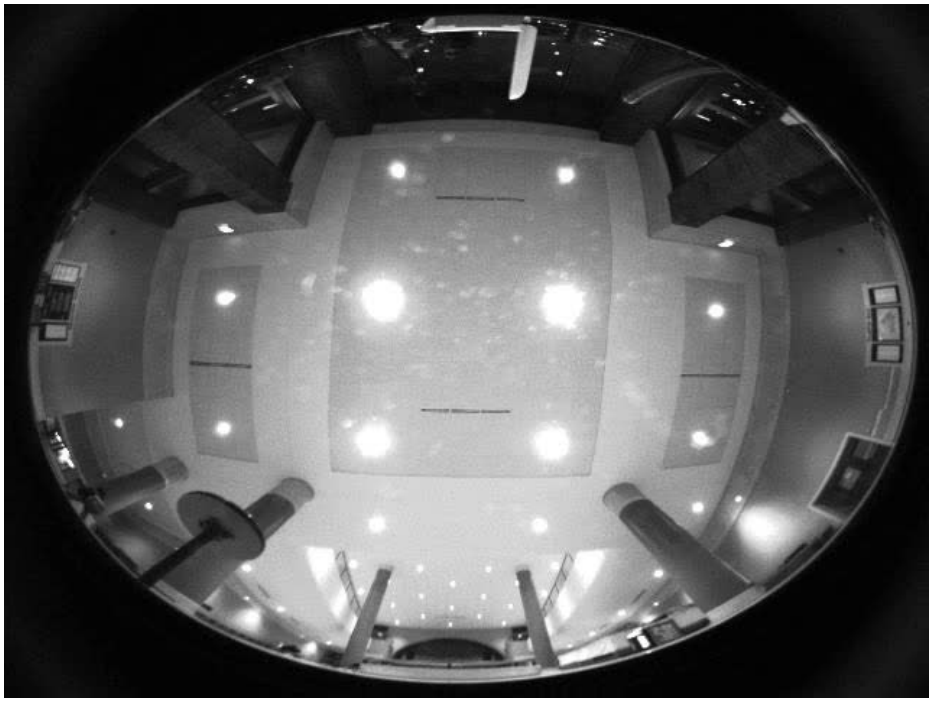


. . . **plus 599 more**

output:

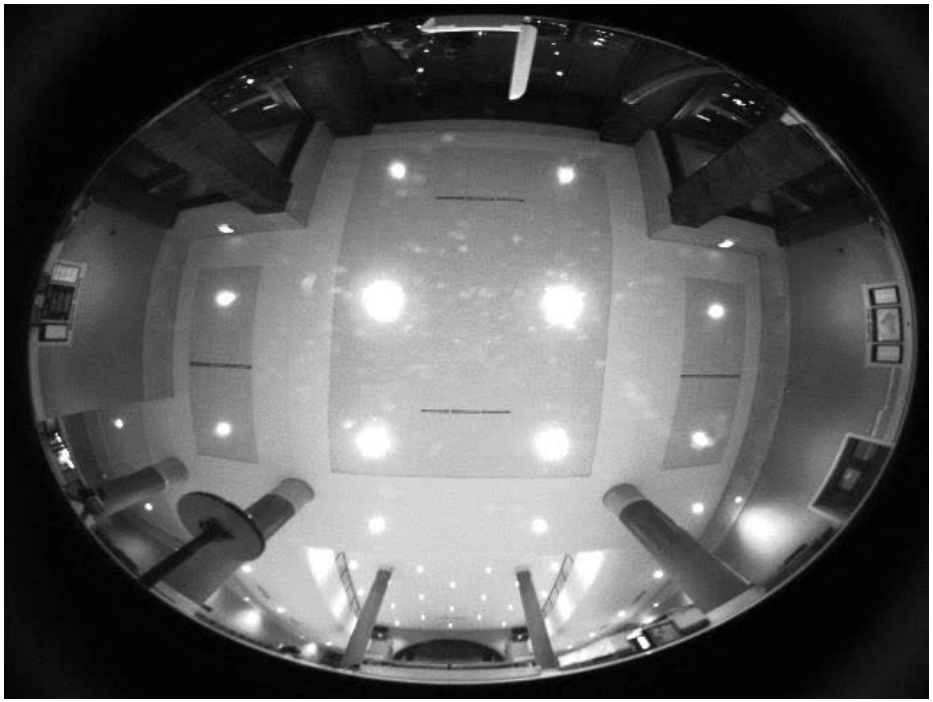


Language Generation



The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation



The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

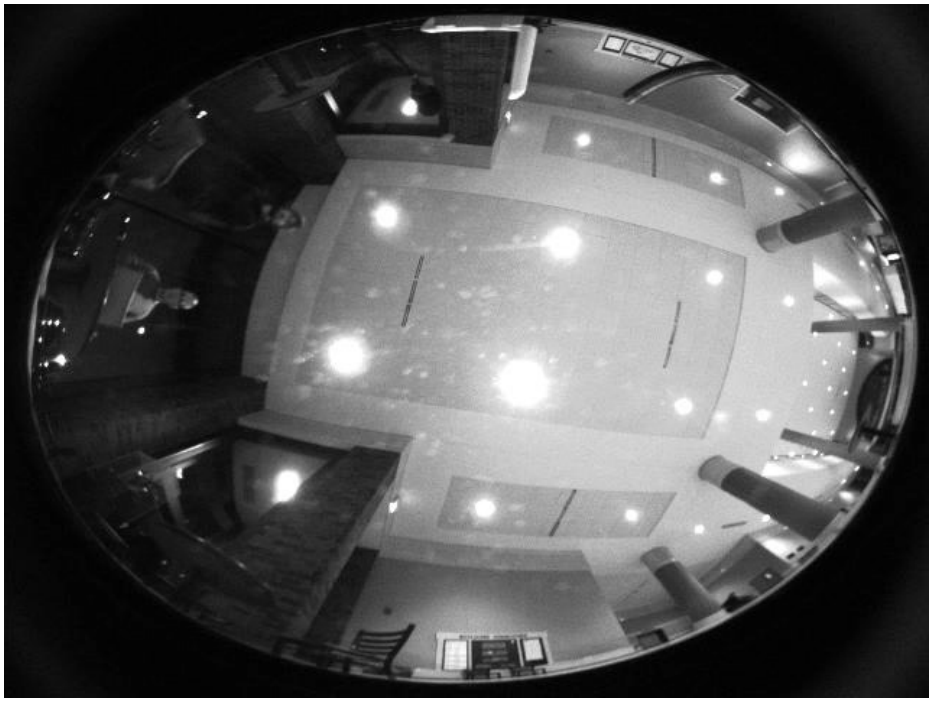
Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

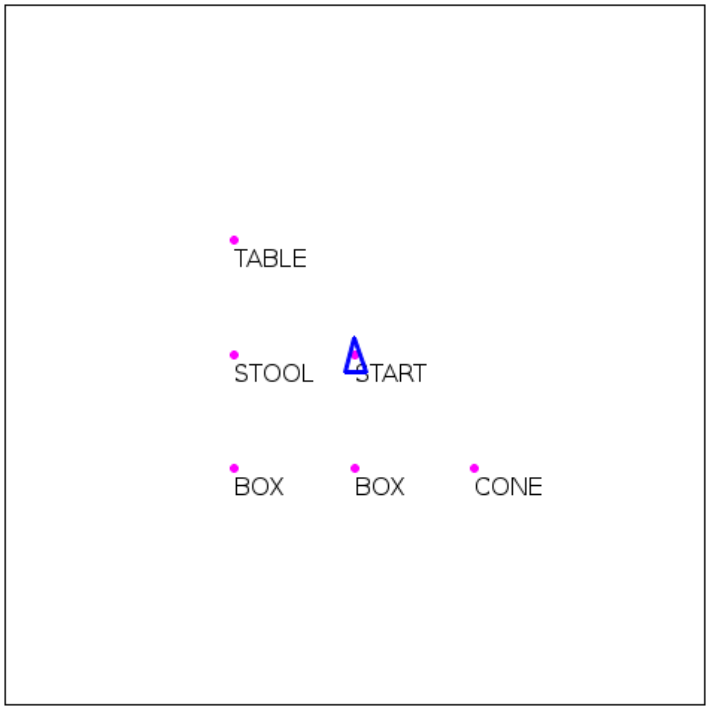
Language Generation



The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.

Language Generation

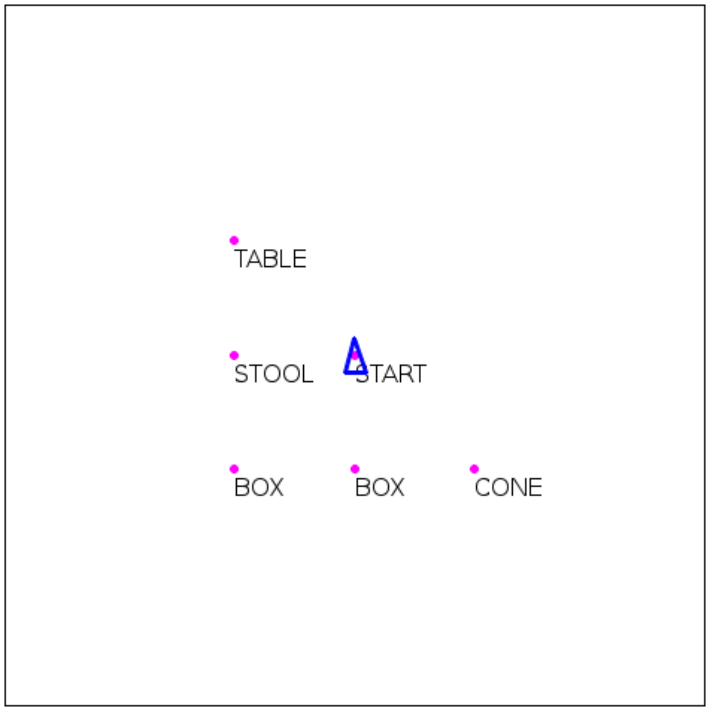
input:



output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:



output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

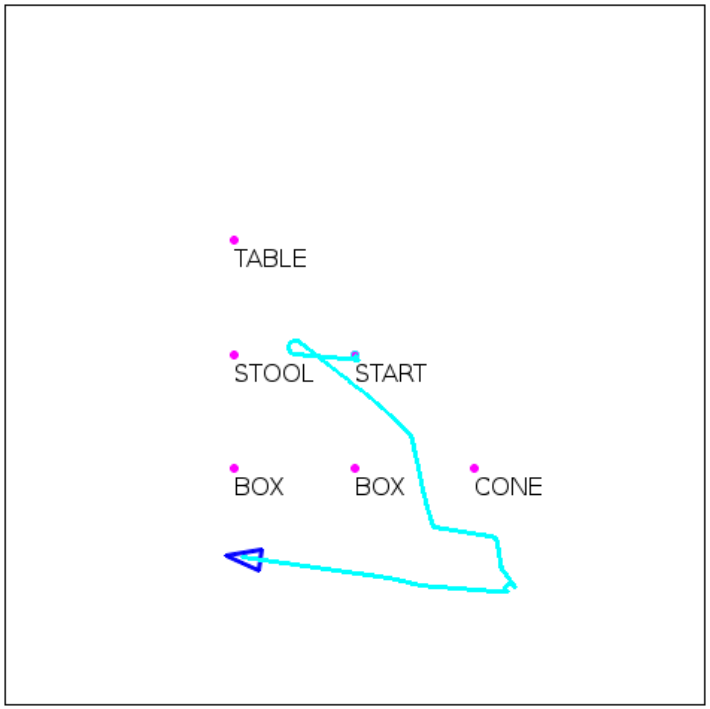
Language Generation

input:

output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Generation

input:

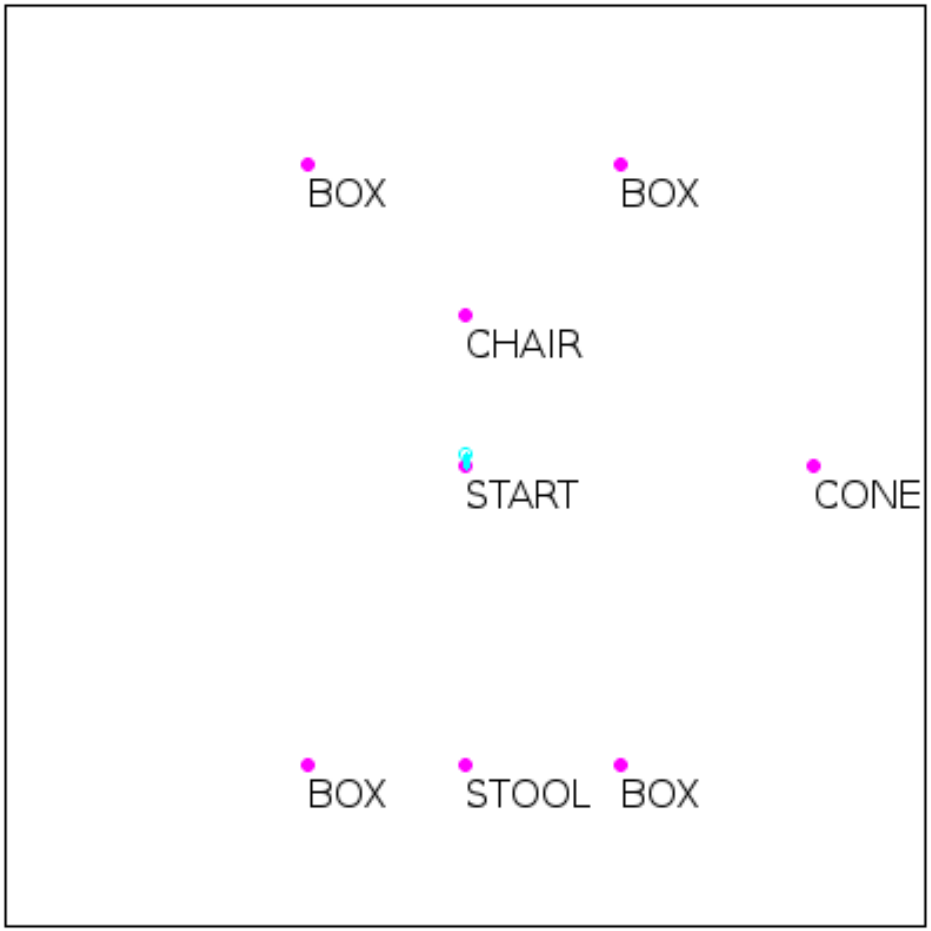


output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which is right of the box then went in front of the box which is left of the box.*

Language Comprehension

input: *The robot went away from the cone then went behind the box which is right of the chair and which is behind the cone then went towards the stool.*

output:

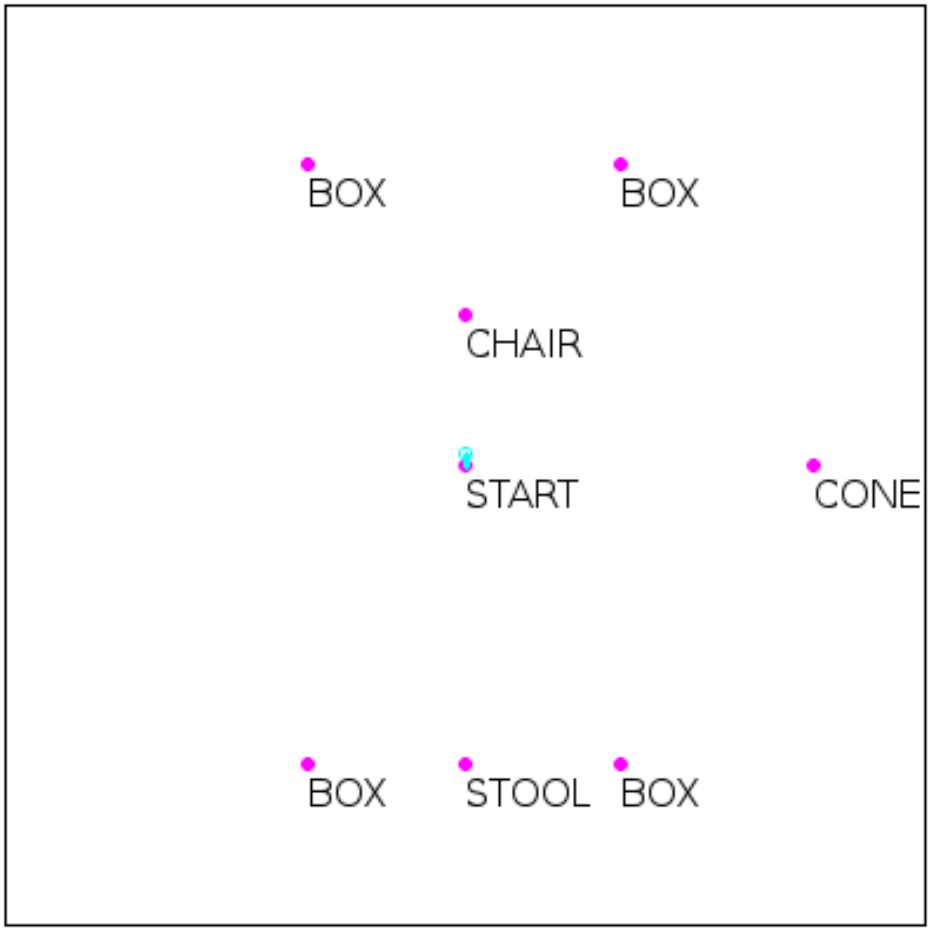


- Determine path waypoints that satisfy sentence
- Add intermediate points to avoid obstacles

Language Comprehension

input: *The robot went away from the cone then went behind the box which is right of the chair and which is behind the cone then went towards the stool.*

output:

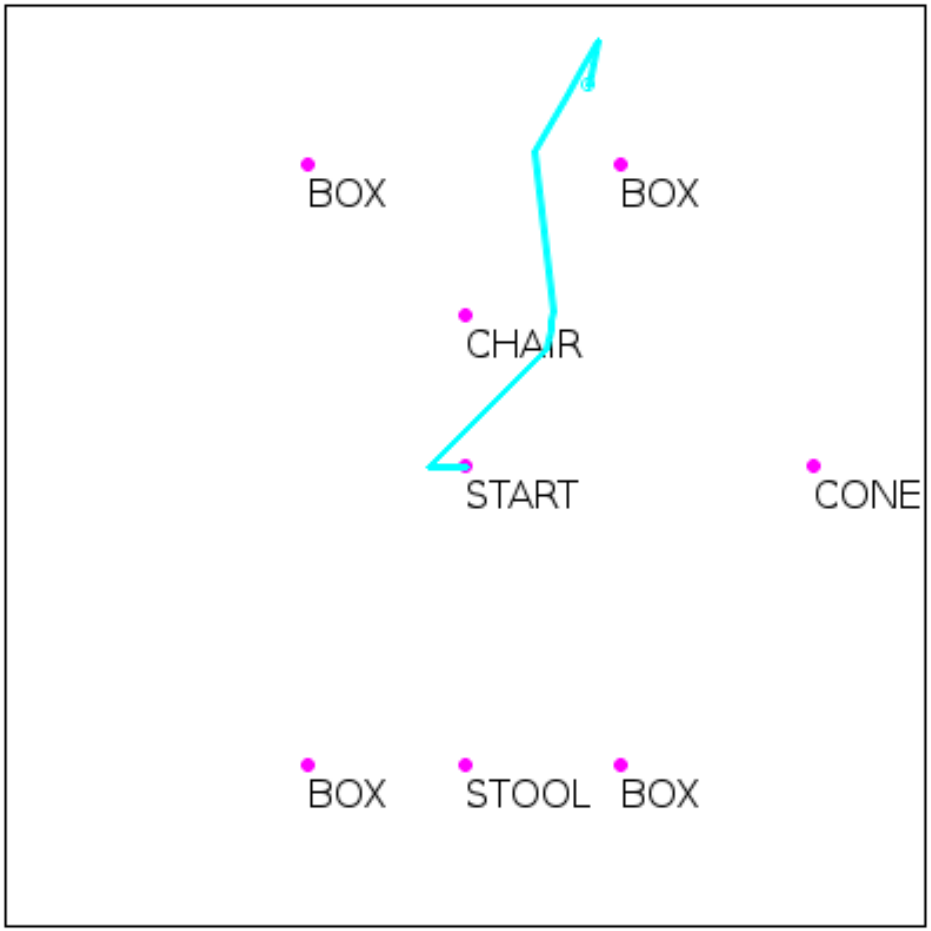


- Determine path waypoints that satisfy sentence
- Add intermediate points to avoid obstacles

Language Comprehension

input: *The robot went away from the cone then went behind the box which is right of the chair and which is behind the cone then went towards the stool.*

output:

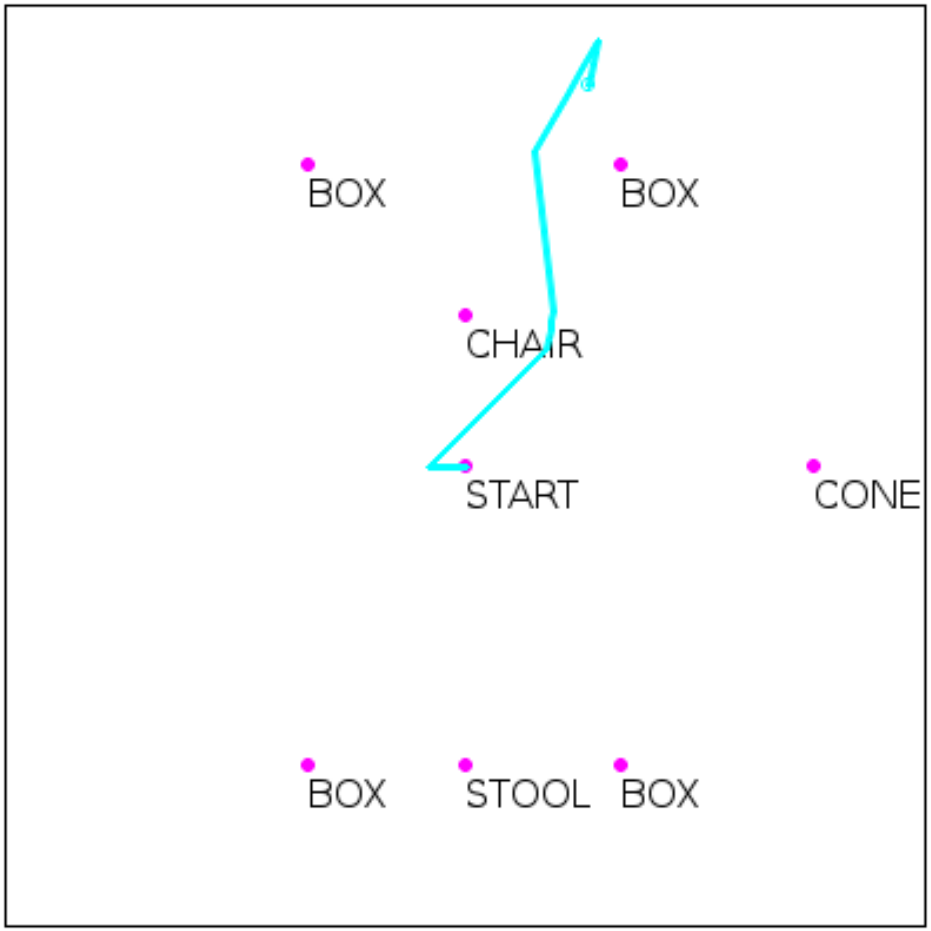


- Determine path waypoints that satisfy sentence
- Add intermediate points to avoid obstacles

Language Comprehension

input: *The robot went away from the cone then went behind the box which is right of the chair and which is behind the cone then went towards the stool.*

output:

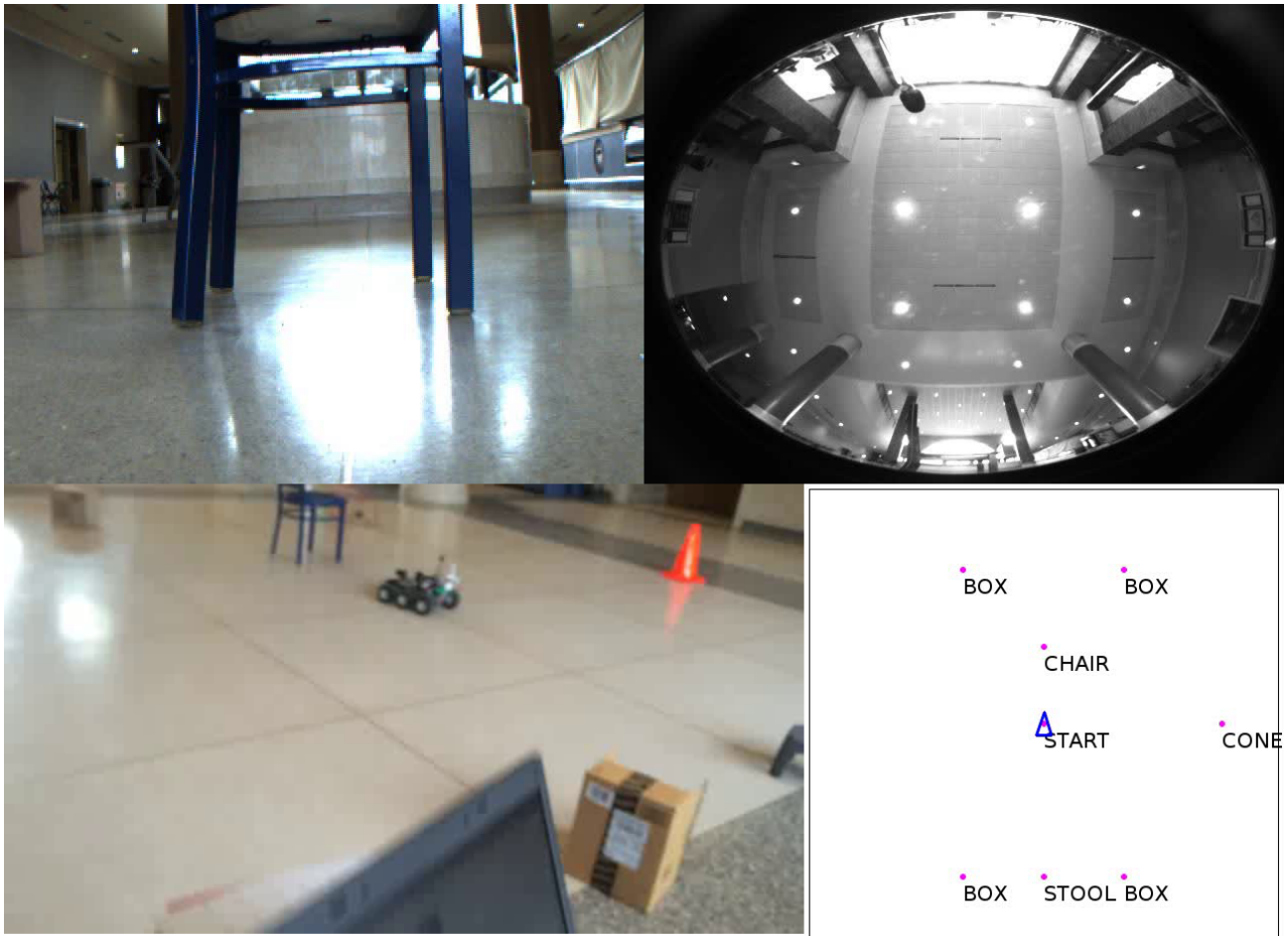


- Determine path waypoints that satisfy sentence
- Add intermediate points to avoid obstacles

Language Comprehension

The Effect of Different Prepositions (1)

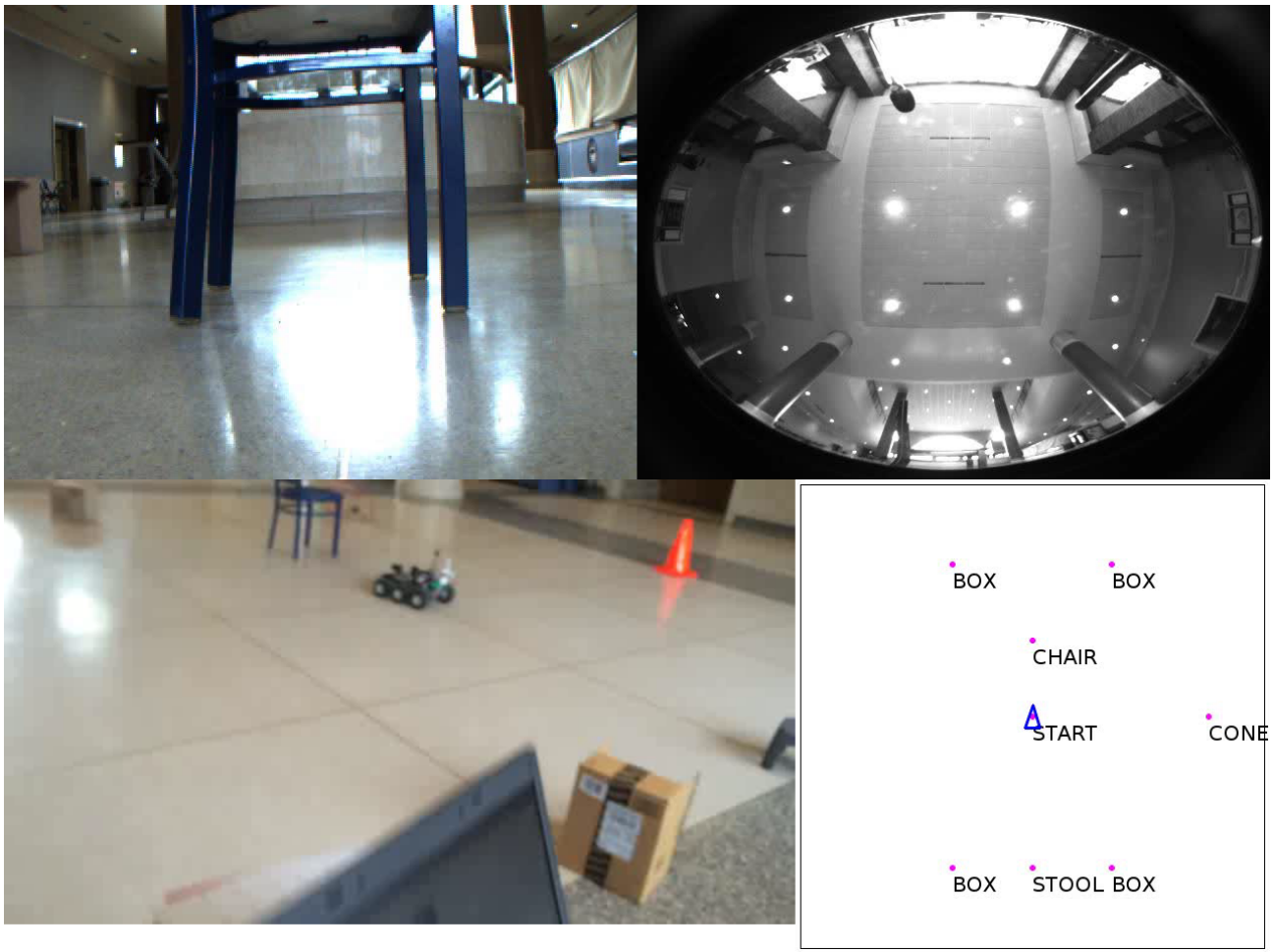
The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **behind** *the cone then went towards the stool.*



Language Comprehension

The Effect of Different Prepositions (1)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **behind** *the cone then went towards the stool.*



Language Comprehension

The Effect of Different Prepositions (1)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **behind** *the cone then went towards the stool.*

Language Comprehension

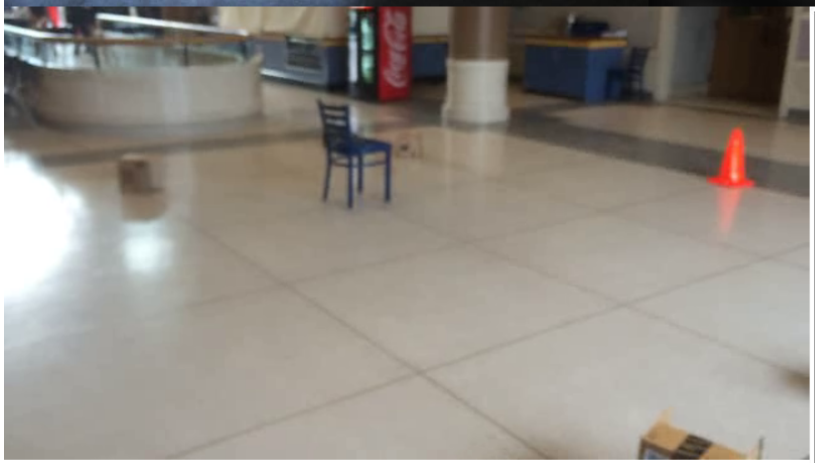
The Effect of Different Prepositions (1)

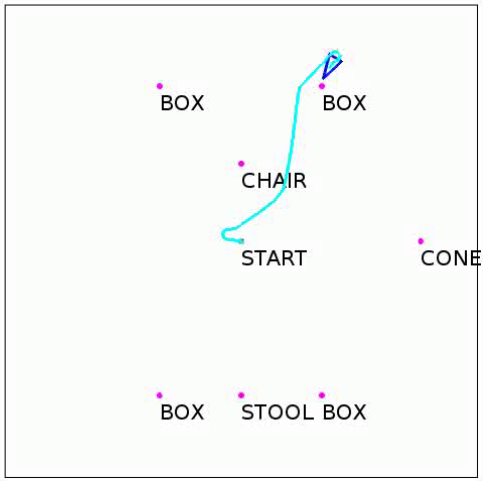
The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **behind** *the cone then went towards the stool.*

Language Comprehension

The Effect of Different Prepositions (1)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **behind** *the cone then went towards the stool.*

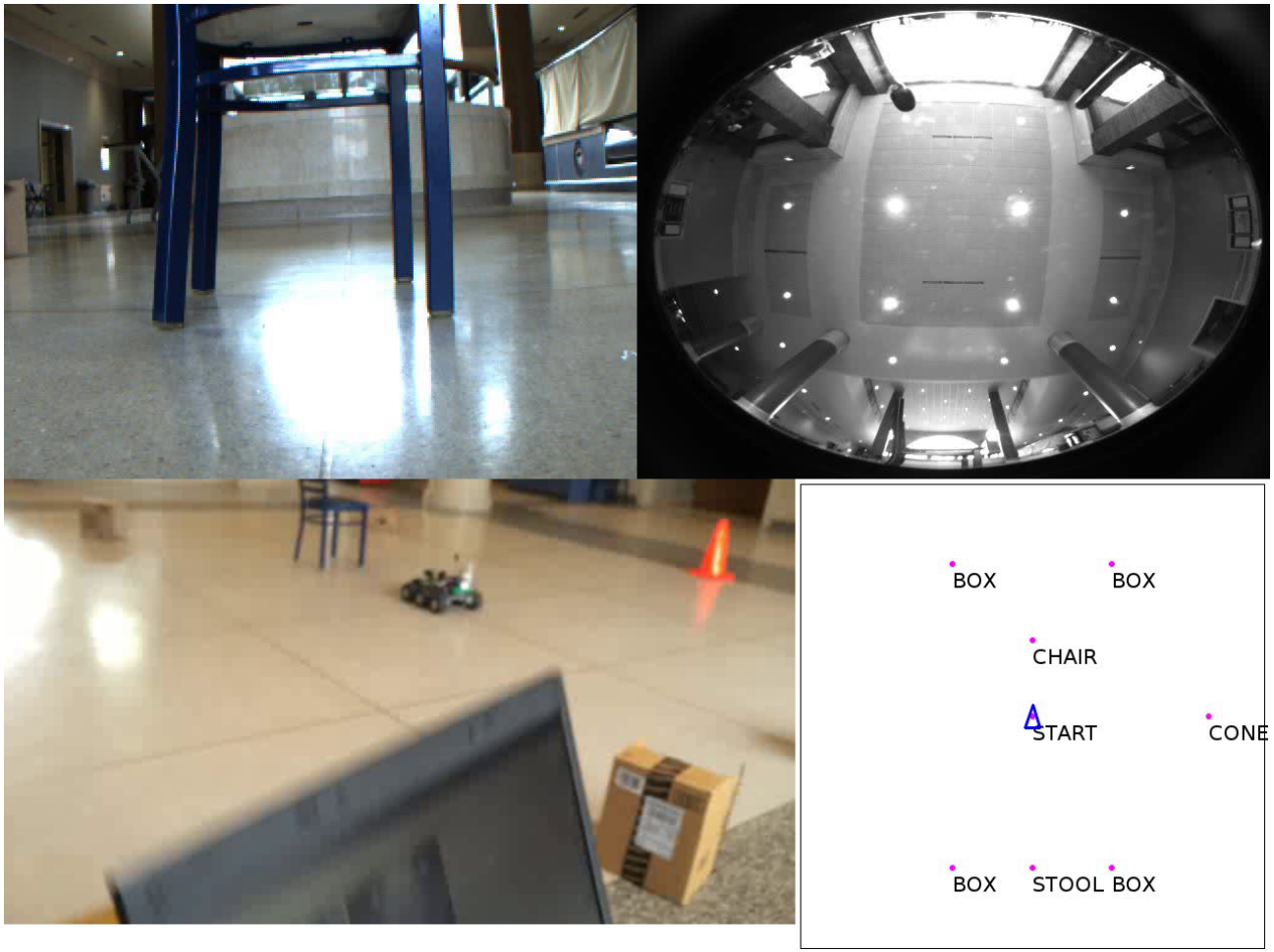




Language Comprehension

The Effect of Different Prepositions (2)

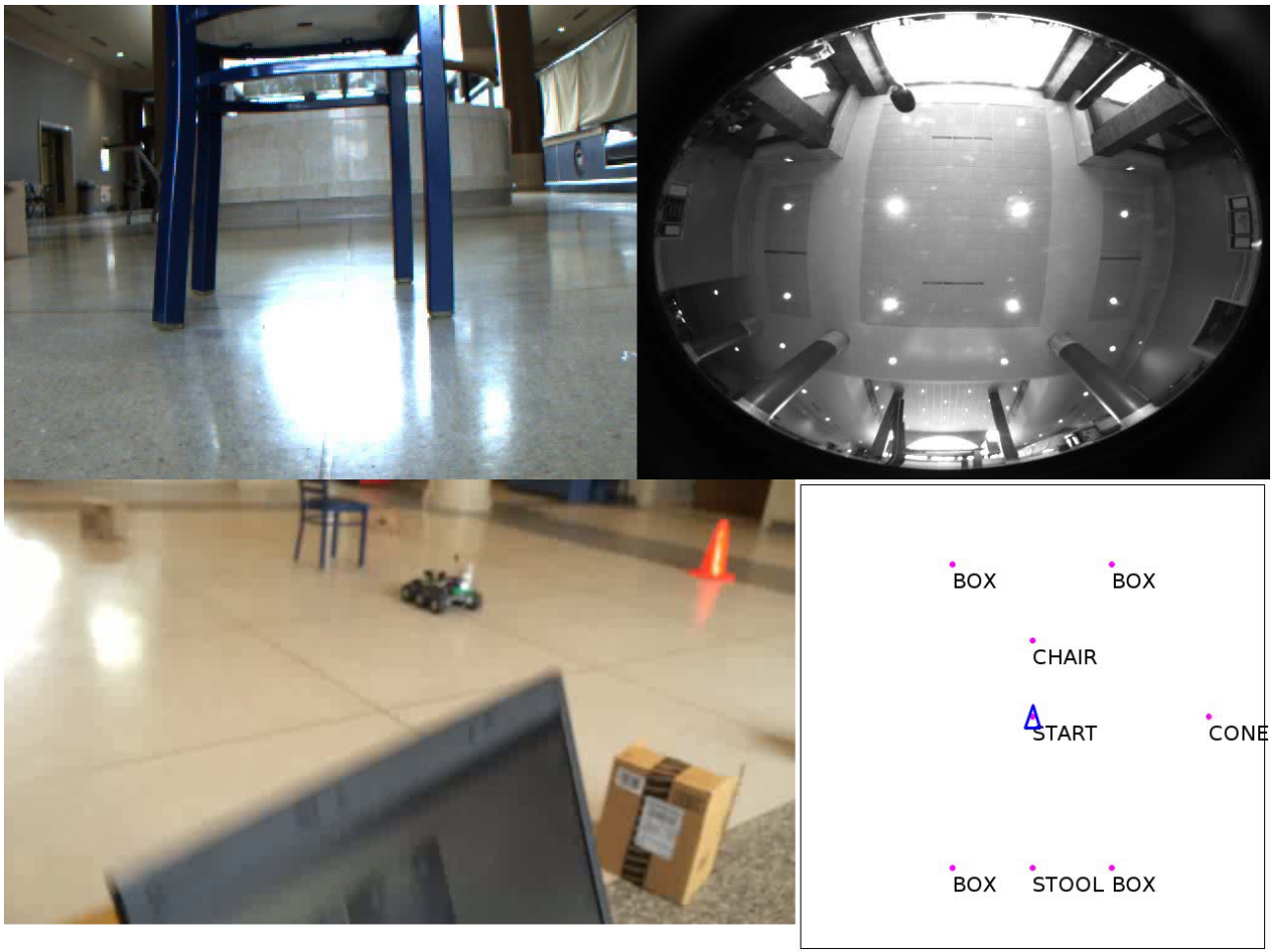
The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **in front of** *the cone then went towards the stool.*



Language Comprehension

The Effect of Different Prepositions (2)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **in front of** *the cone then went towards the stool.*



Language Comprehension

The Effect of Different Prepositions (2)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **in front of** *the cone then went towards the stool.*

Language Comprehension

The Effect of Different Prepositions (2)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **in front of** *the cone then went towards the stool.*

Language Comprehension

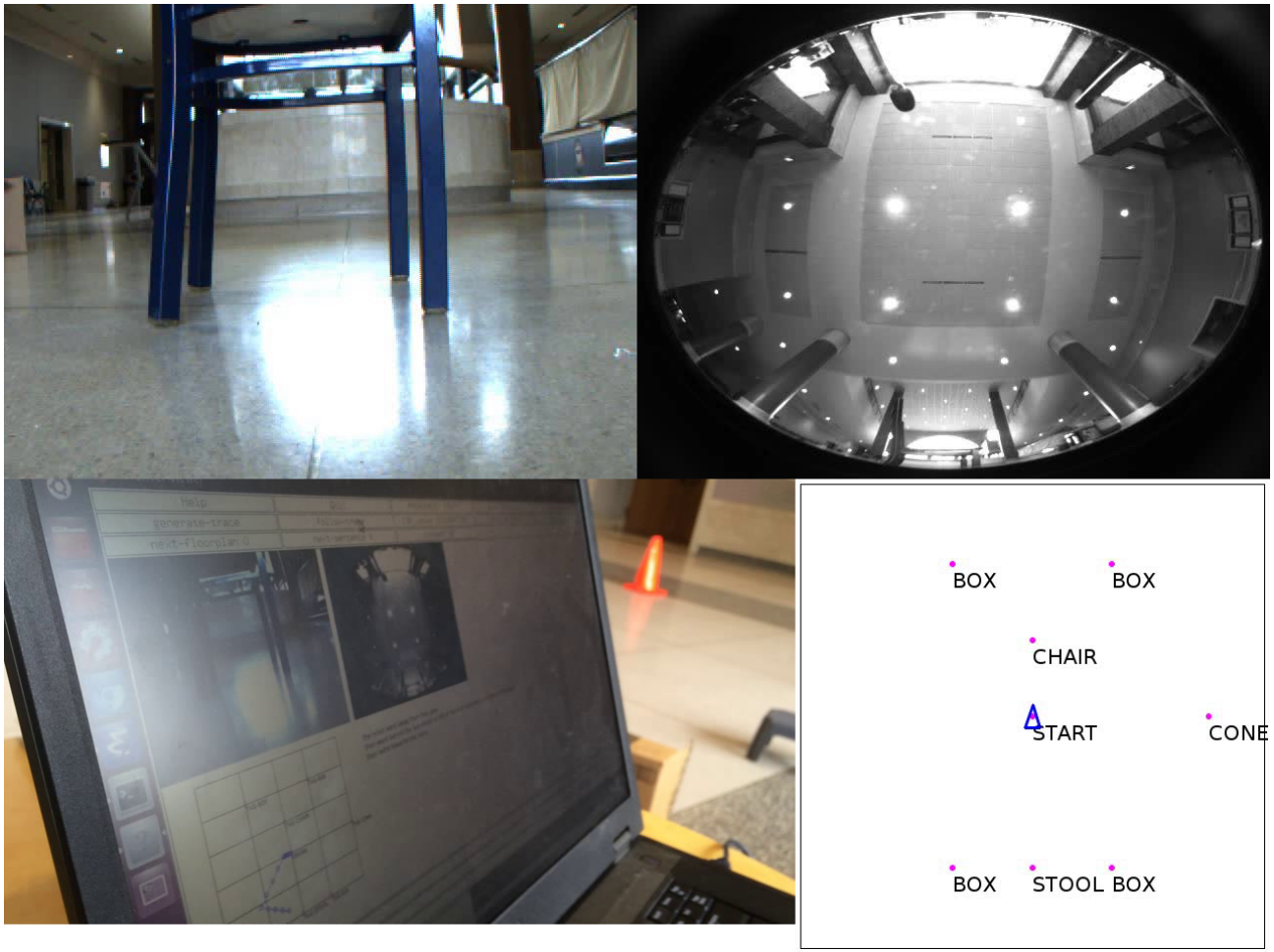
The Effect of Different Prepositions (2)

The robot went away from the cone then went behind the box which is **right of** *the chair and which is* **in front of** *the cone then went towards the stool.*

Language Comprehension

The Effect of Different Prepositions (3)

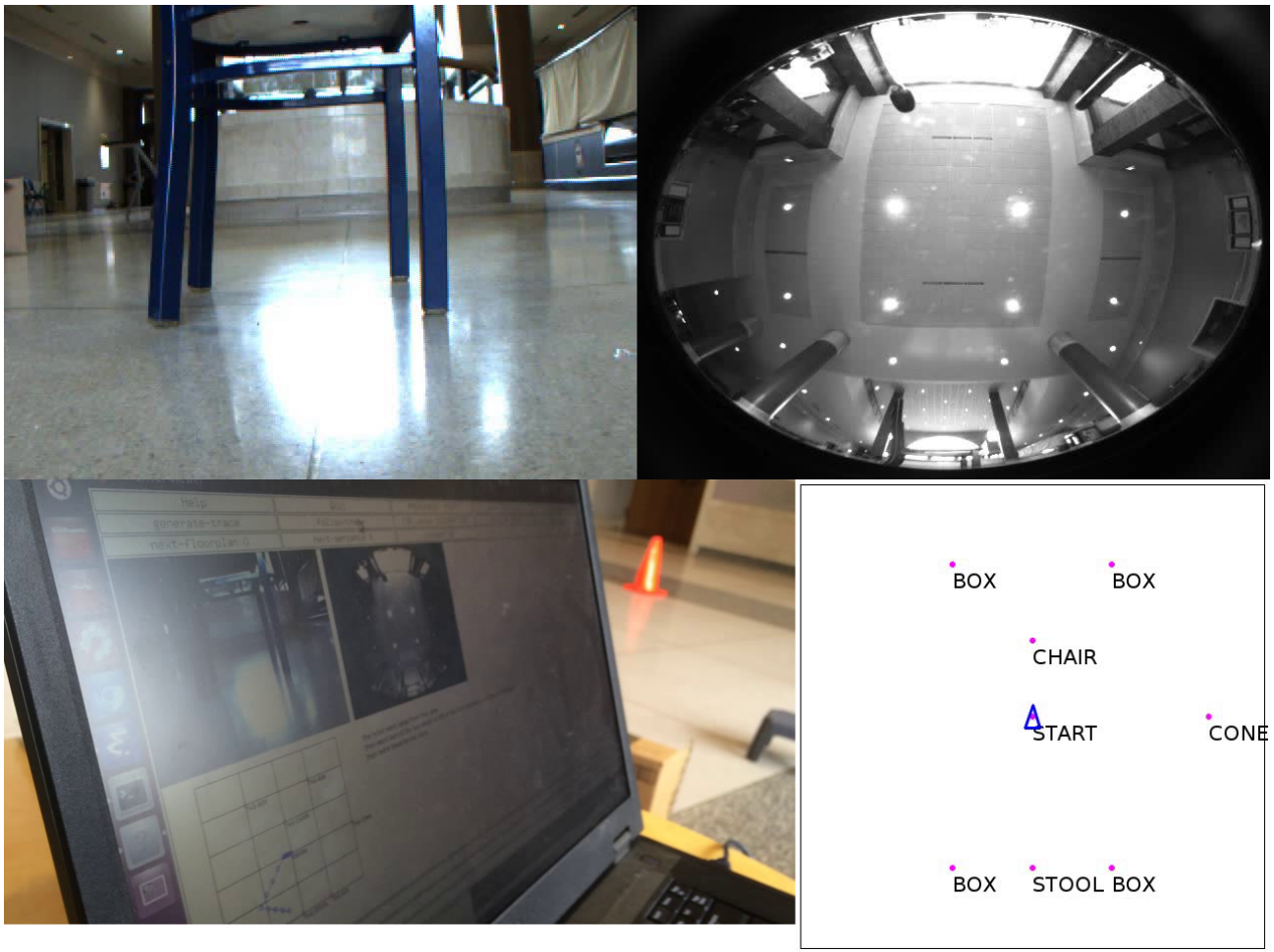
The robot went away from the cone then went behind the box which is **left of** *the chair and which is* **in front of** *the cone then went towards the stool.*



Language Comprehension

The Effect of Different Prepositions (3)

The robot went away from the cone then went behind the box which is **left of** *the chair and which is* **in front of** *the cone then went towards the stool.*



Language Comprehension

The Effect of Different Prepositions (3)

The robot went away from the cone then went behind the box which is **left of** *the chair and which is* **in front of** *the cone then went towards the stool.*

Language Comprehension

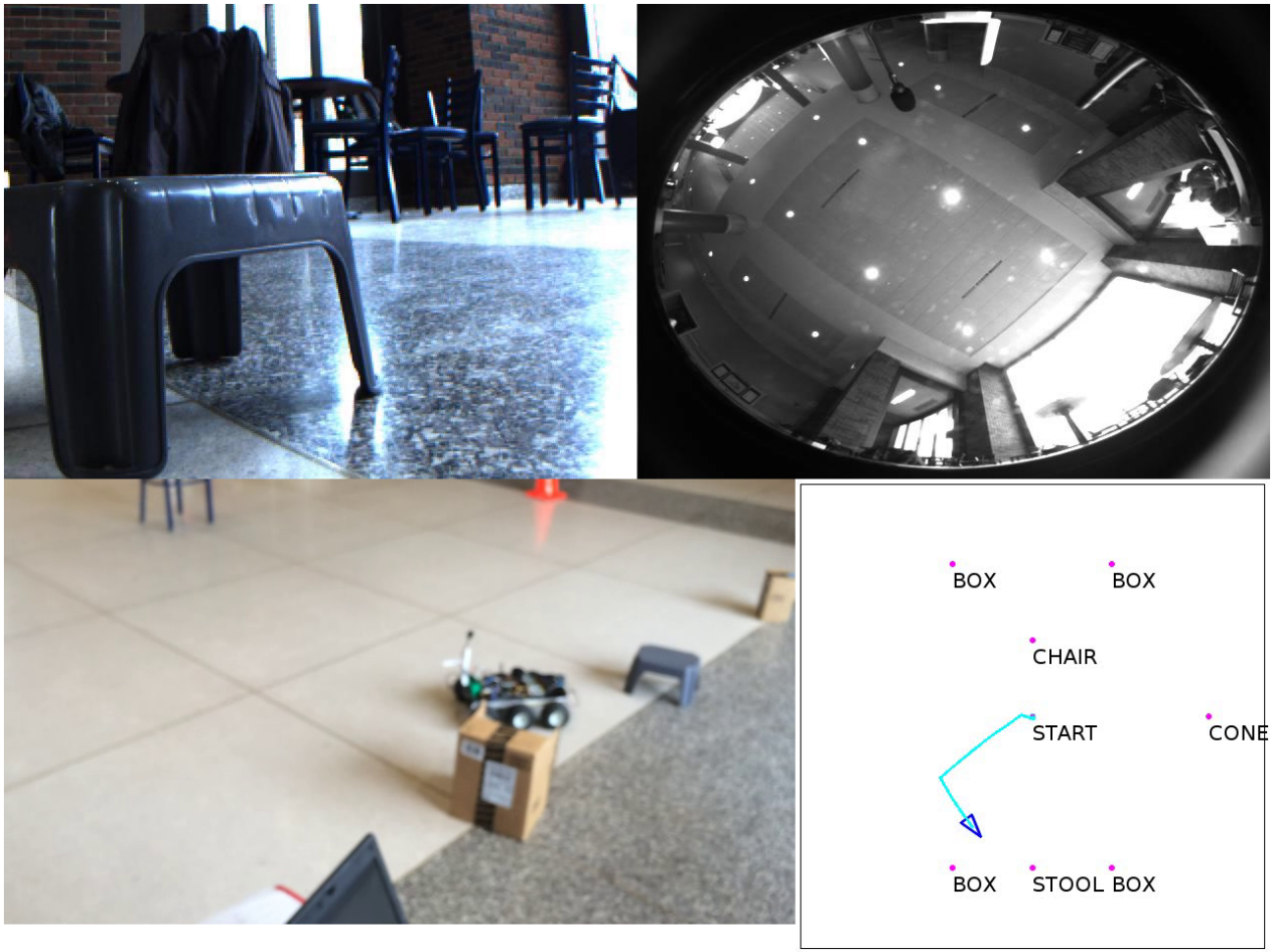
The Effect of Different Prepositions (3)

The robot went away from the cone then went behind the box which is **left of** *the chair and which is* **in front of** *the cone then went towards the stool.*

Language Comprehension

The Effect of Different Prepositions (3)

The robot went away from the cone then went behind the box which is **left of** *the chair and which is* **in front of** *the cone then went towards the stool.*



Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

- all sentences naturally elicited from humans through AMT

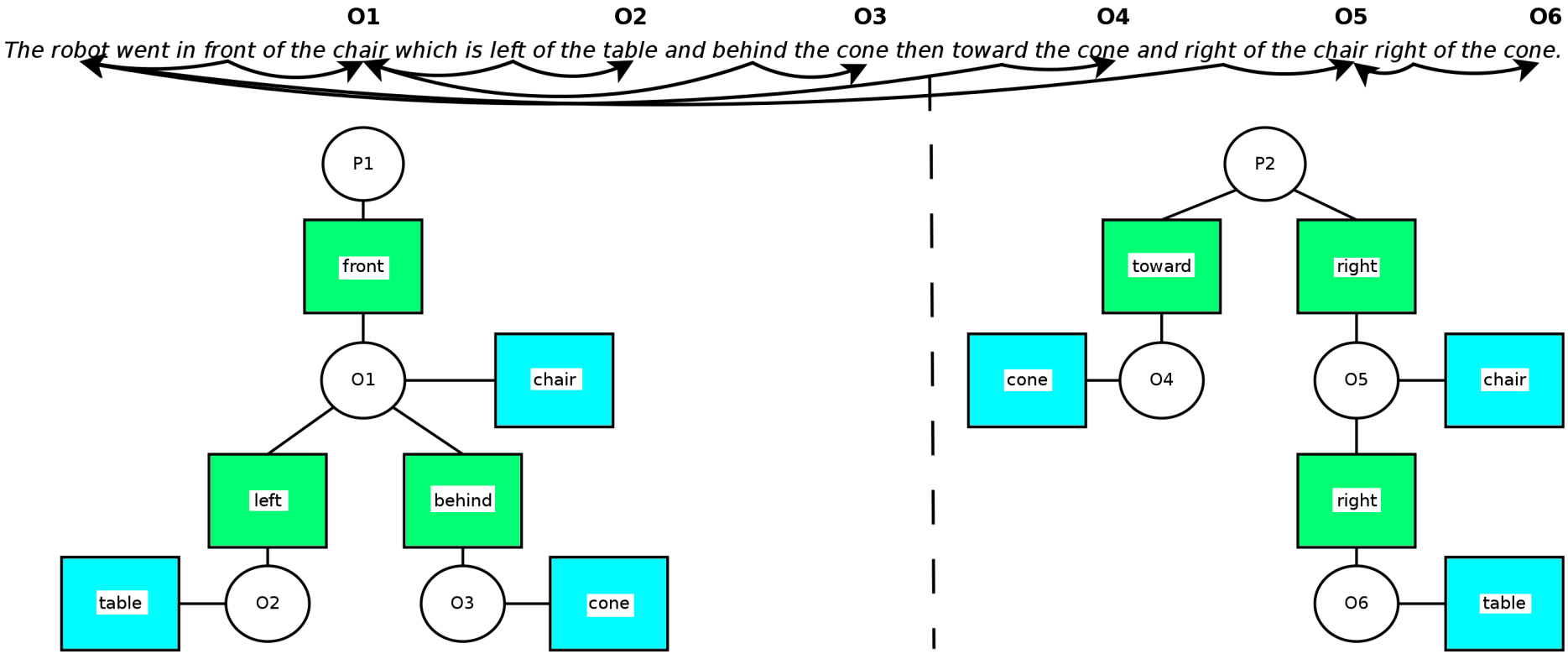
Logical Form

The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

$$[\alpha, \beta, \gamma, \delta]\{t, u, v, w, x, y, z\} \left(\begin{array}{l} \text{LEFT}(\alpha, t) \wedge \text{STOOL}(t) \wedge \\ \text{TOWARD}(\beta, u) \wedge \text{CONE}(u) \wedge \text{BEHIND}(u, v) \wedge \text{STOOL}(v) \wedge \\ \text{TOWARD}(\gamma, w) \wedge \text{TABLE}(w) \wedge \text{LEFT}(w, x) \wedge \text{CONE}(x) \wedge \\ \text{TOWARD}(\delta, y) \wedge \text{LEFT}(\delta, z) \wedge \text{STOOL}(y) \wedge \text{STOOL}(z) \end{array} \right)$$

- all sentences naturally elicited from humans through AMT
- no grammar or parse trees at all

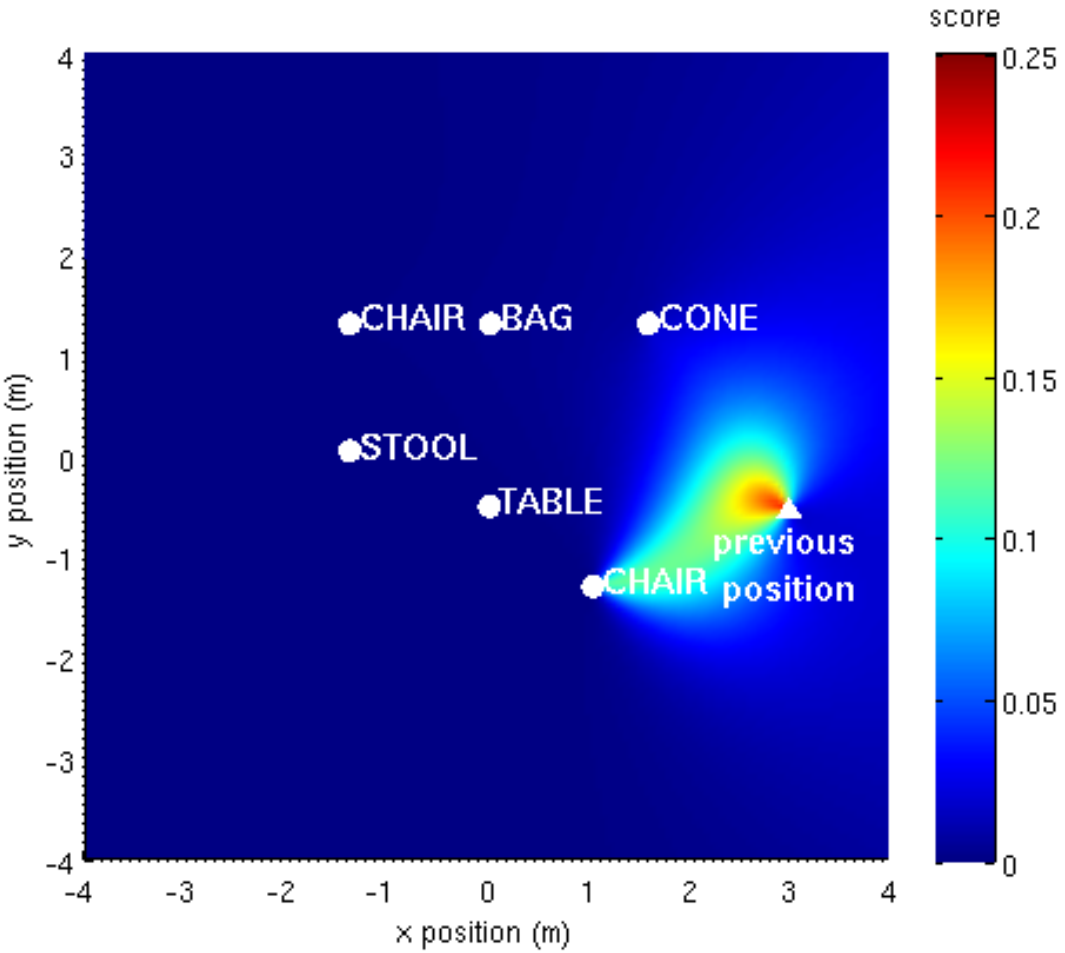
Parsing


O1
O2
O3
O4
O5
O6
The robot went in front of the chair which is left of the table and behind the cone then toward the cone and right of the chair right of the cone.
P1
front
O1
chair
left
behind
table
O2
O3
cone
P2
toward
right
cone
O4
O5
chair
right
O6
table

Semantics as a Soft Context-Sensitive Scoring Function

O4 **O5** **O6**

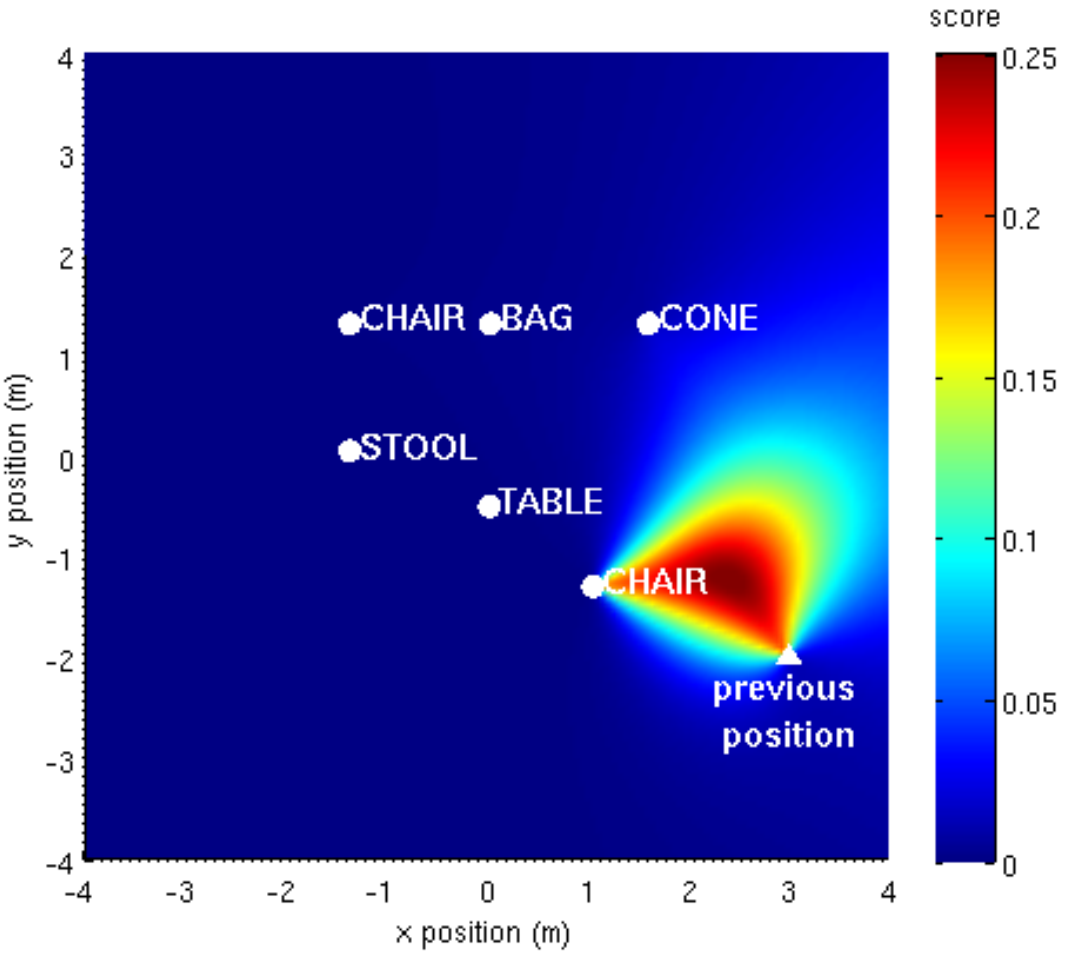
toward the cone and right of the chair right of the table.



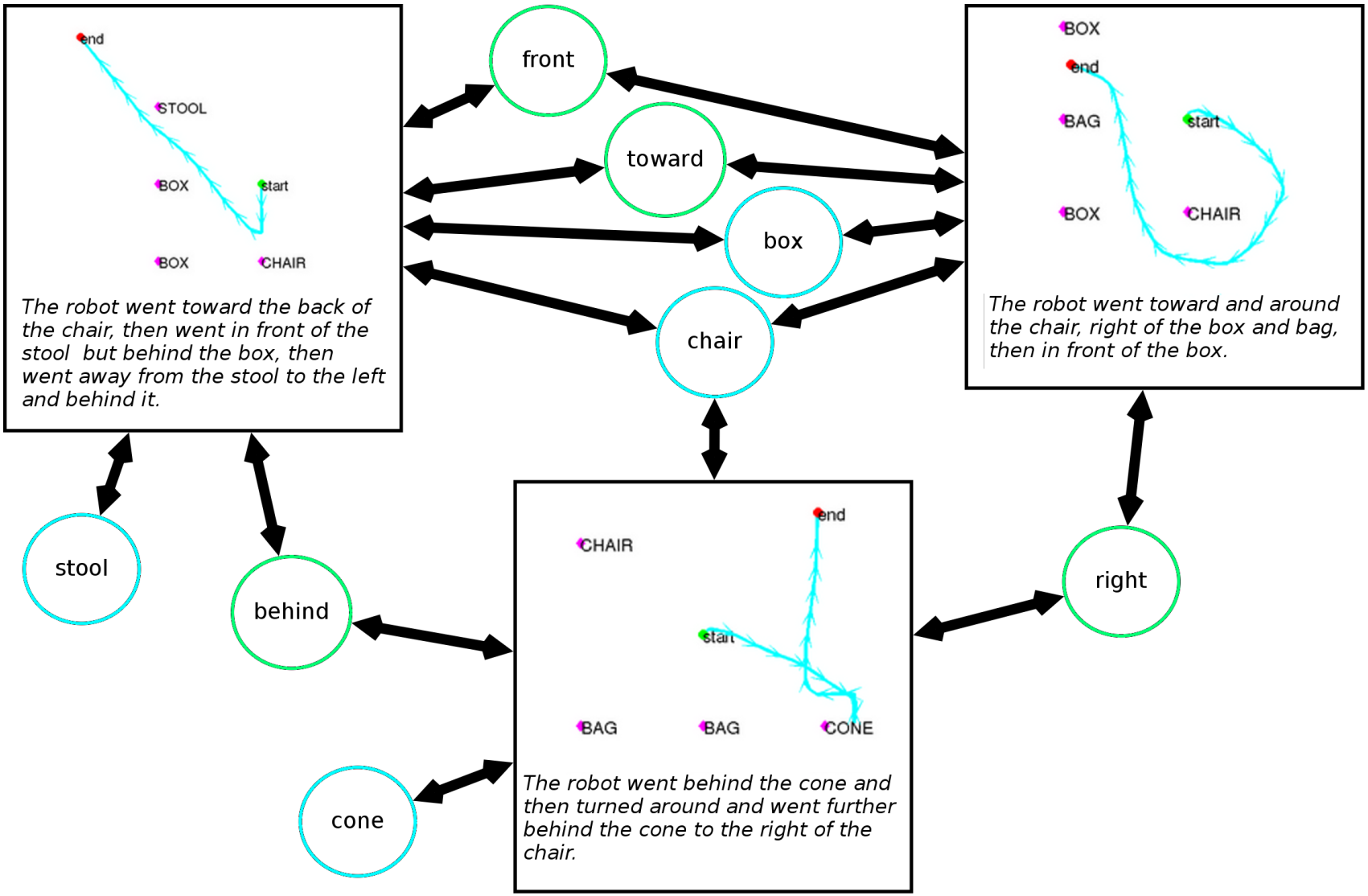
Semantics as a Soft Context-Sensitive Scoring Function

O4 **O5** **O6**

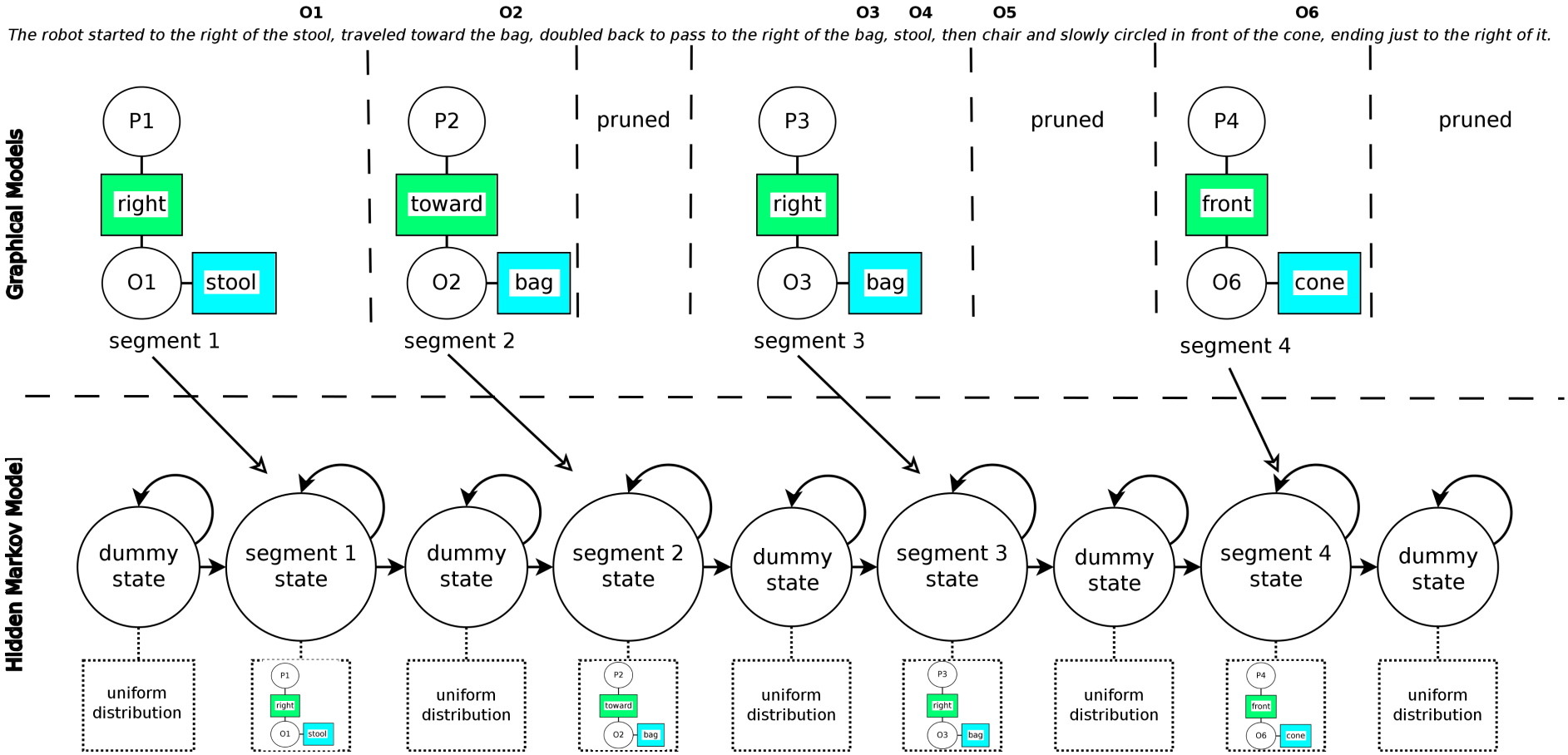
toward the cone and right of the chair right of the table.



Acquisition Method

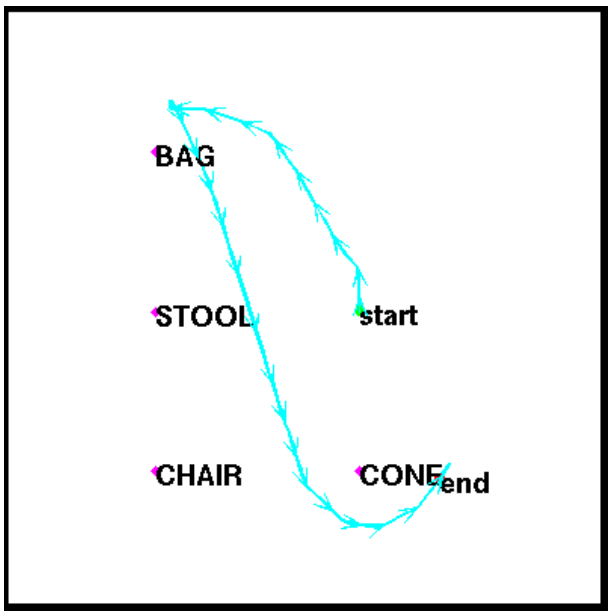


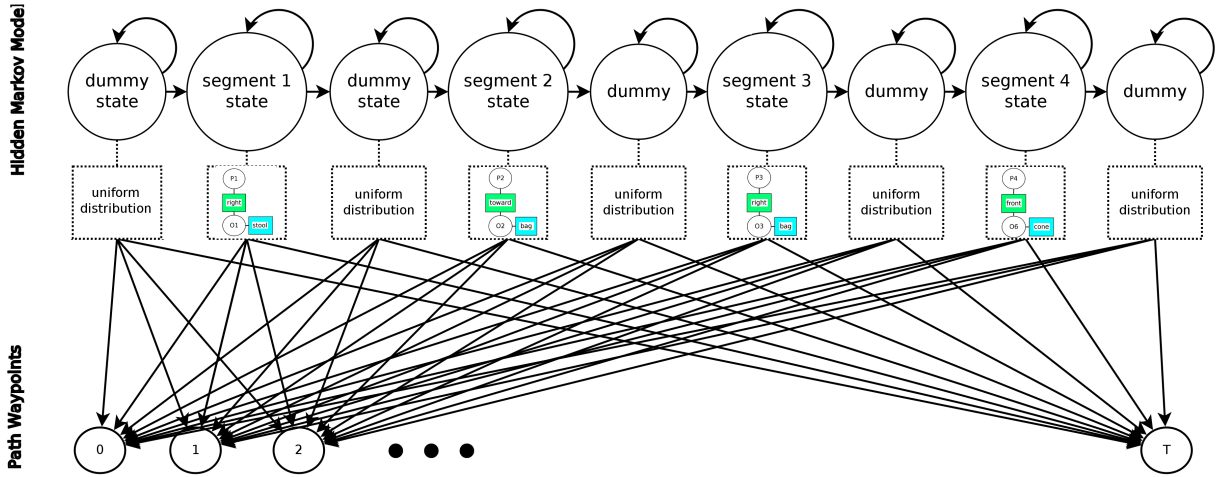
Acquisition Method



Acquisition Method

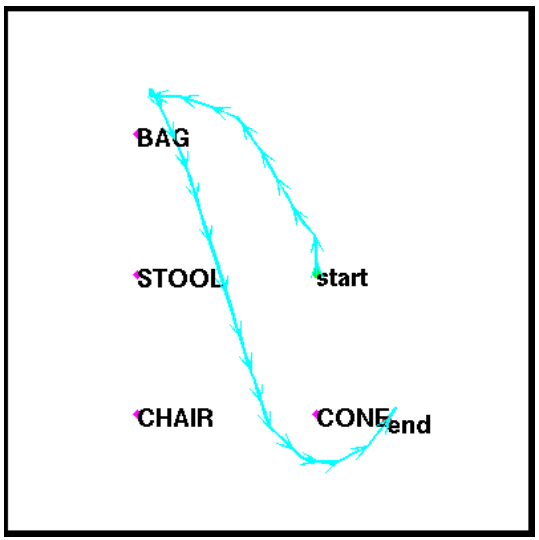
The robot started to the right of the stool, traveled toward the bag, doubled back to pass to the right of the bag, stool, then chair and slowly circled in front of the cone, ending just to the right of it.

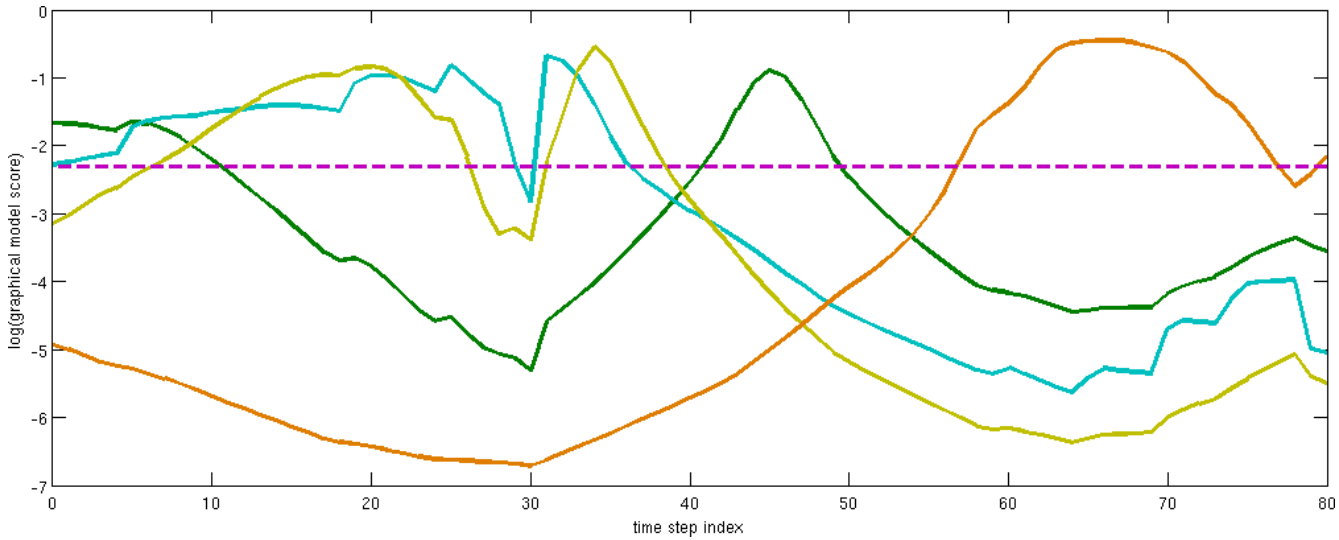




Acquisition Method

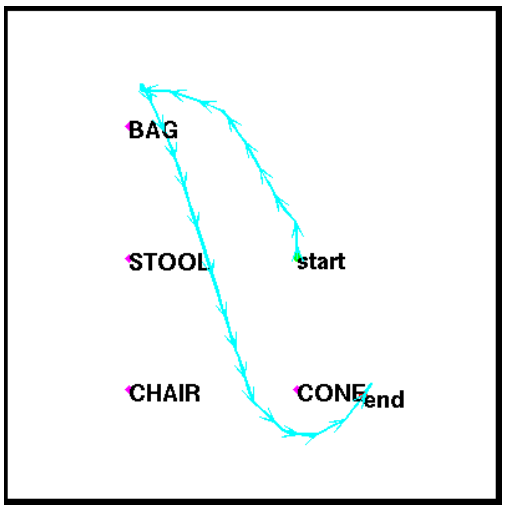
The robot started to the right of the stool, traveled toward the bag, doubled back to pass to the right of the bag, stool, then chair and slowly circled in front of the cone, ending just to the right of it.

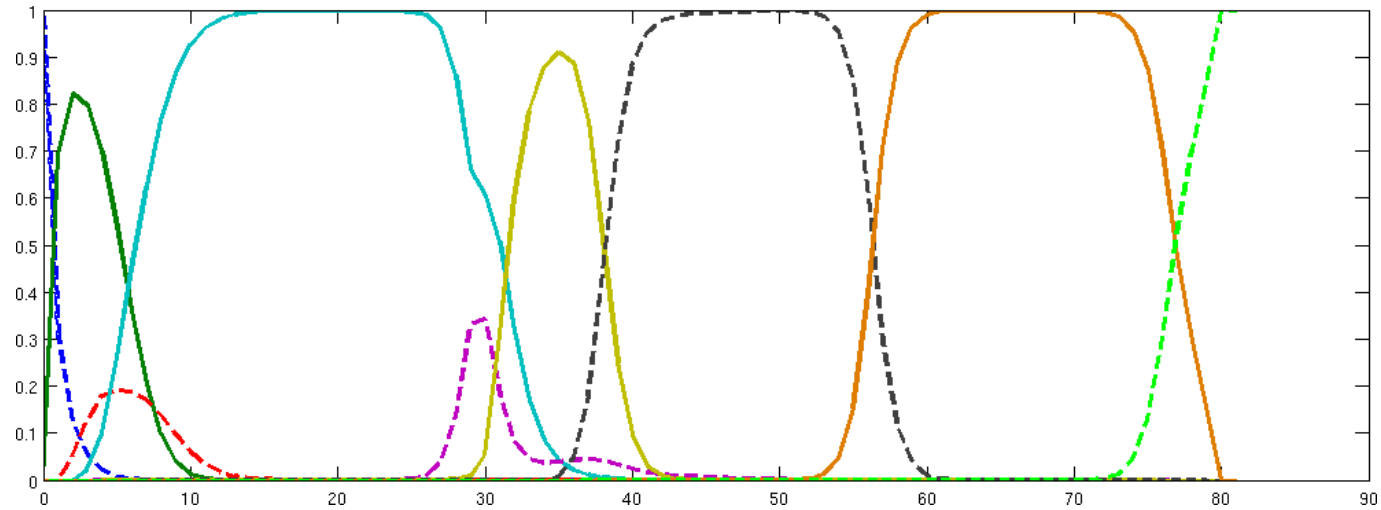




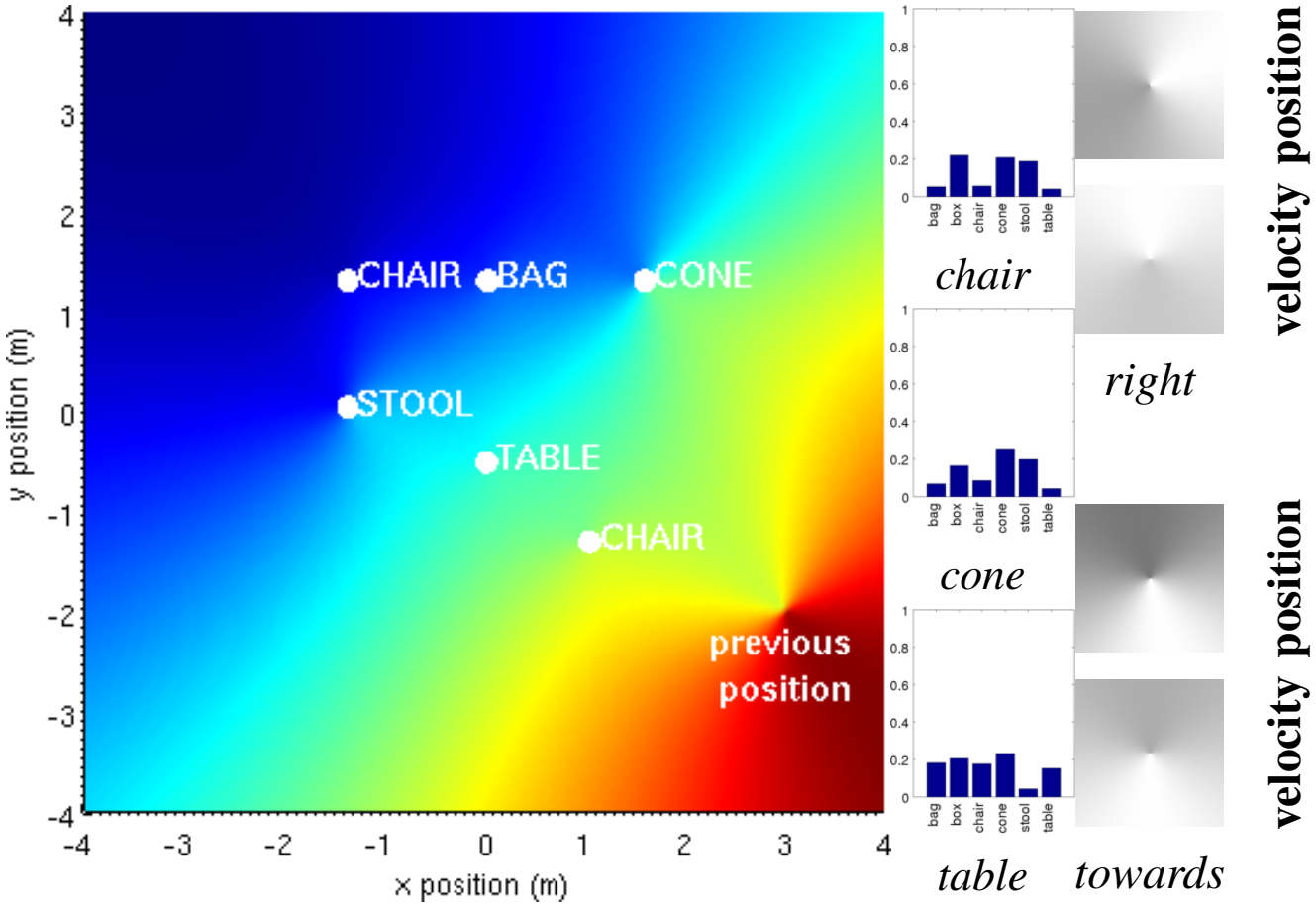
Acquisition Method

The robot started to the right of the stool, traveled toward the bag, doubled back to pass to the right of the bag, stool, then chair and slowly circled in front of the cone, ending just to the right of it.





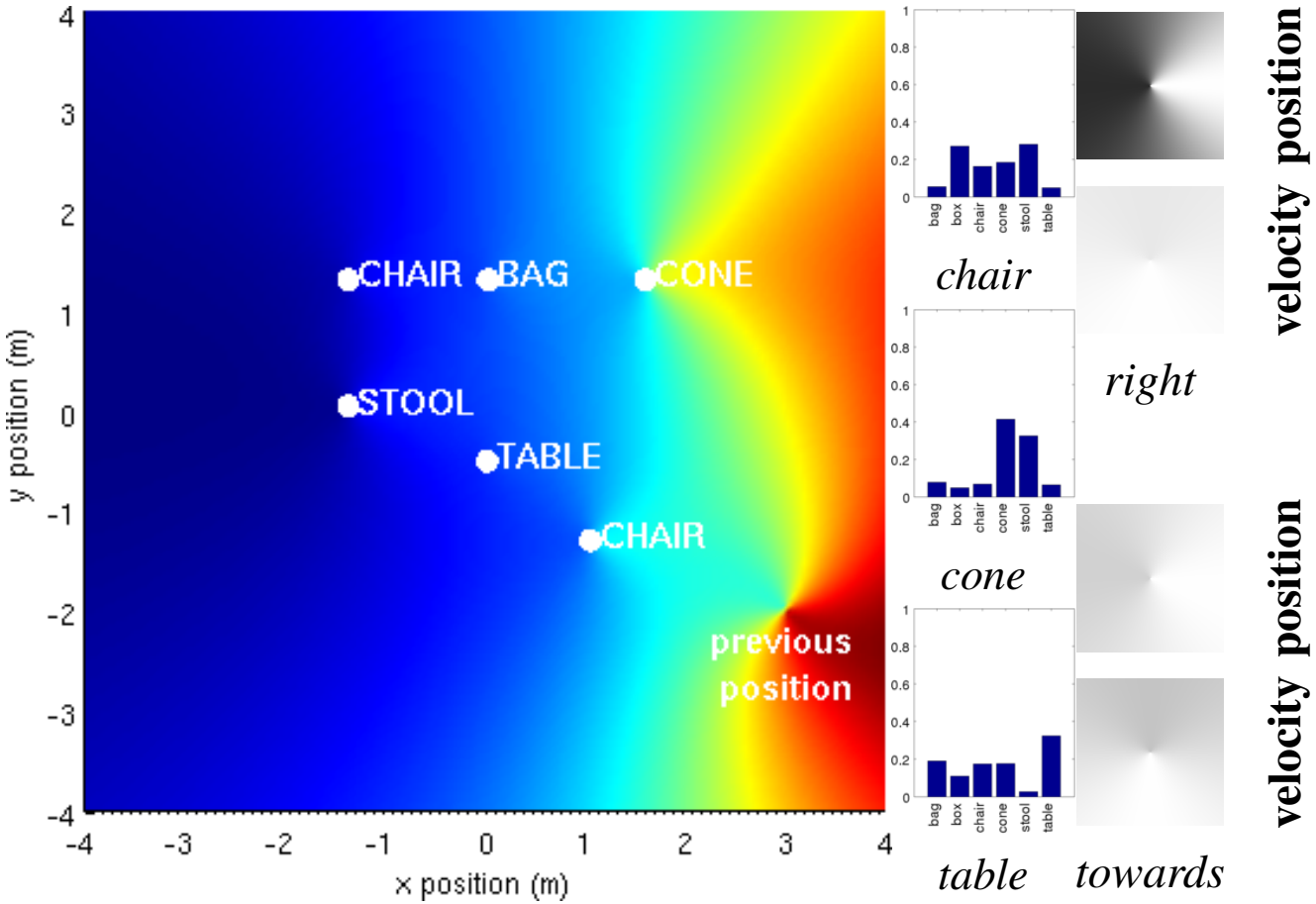
toward the cone and right of the chair right of the table



Acquisition Method

Iteration 1

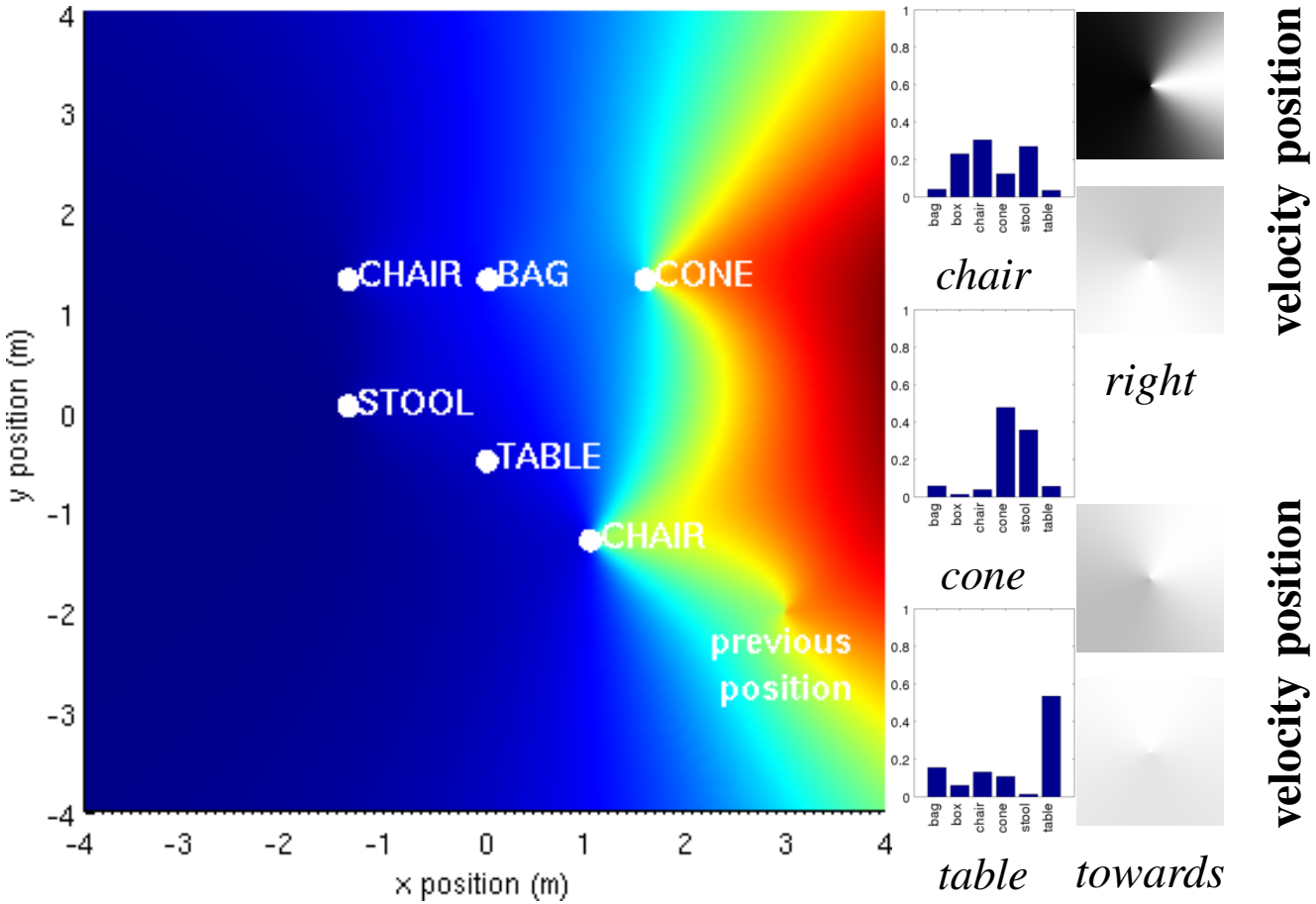
toward the cone and right of the chair right of the table



Acquisition Method

Iteration 2

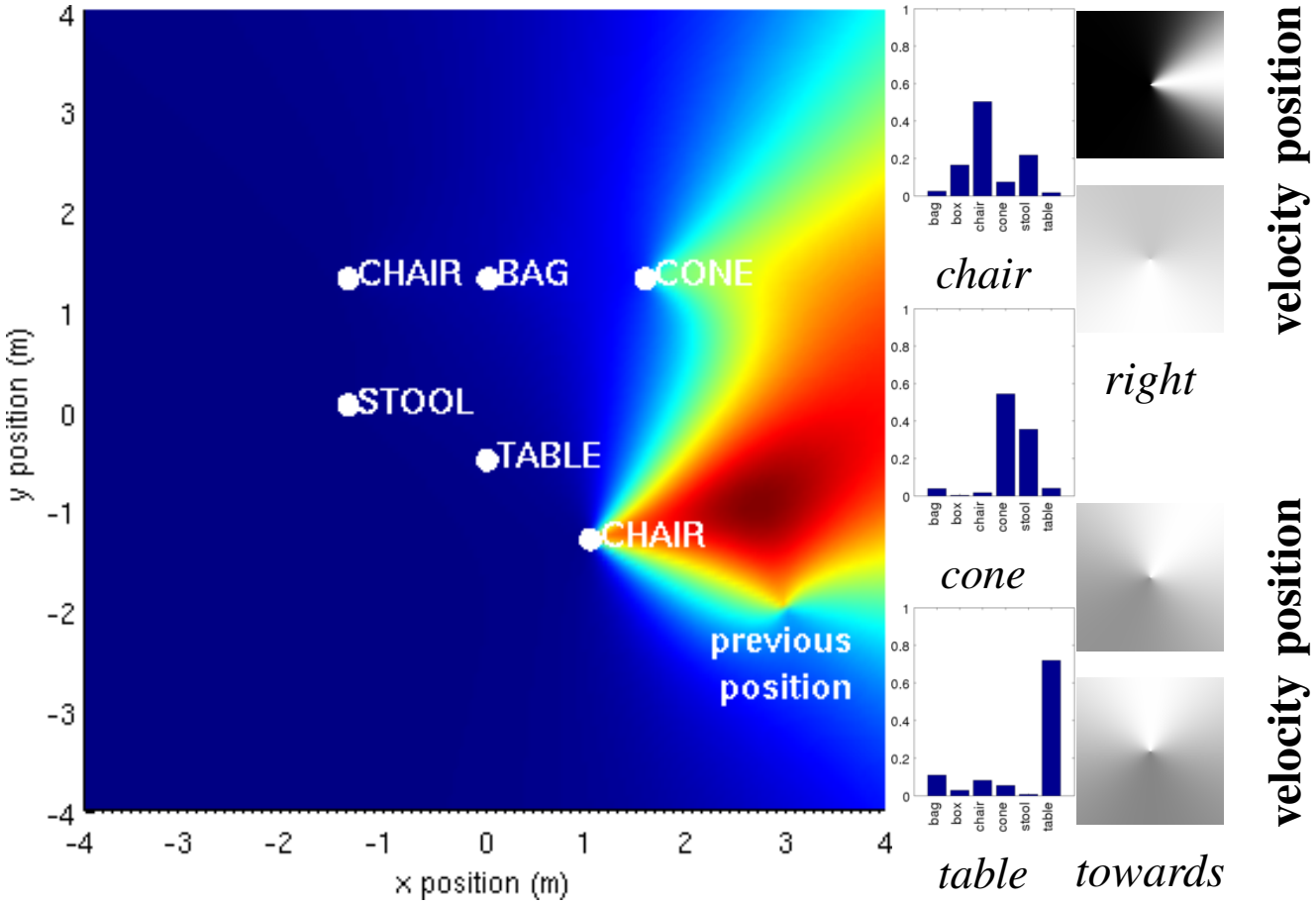
toward the cone and right of the chair right of the table



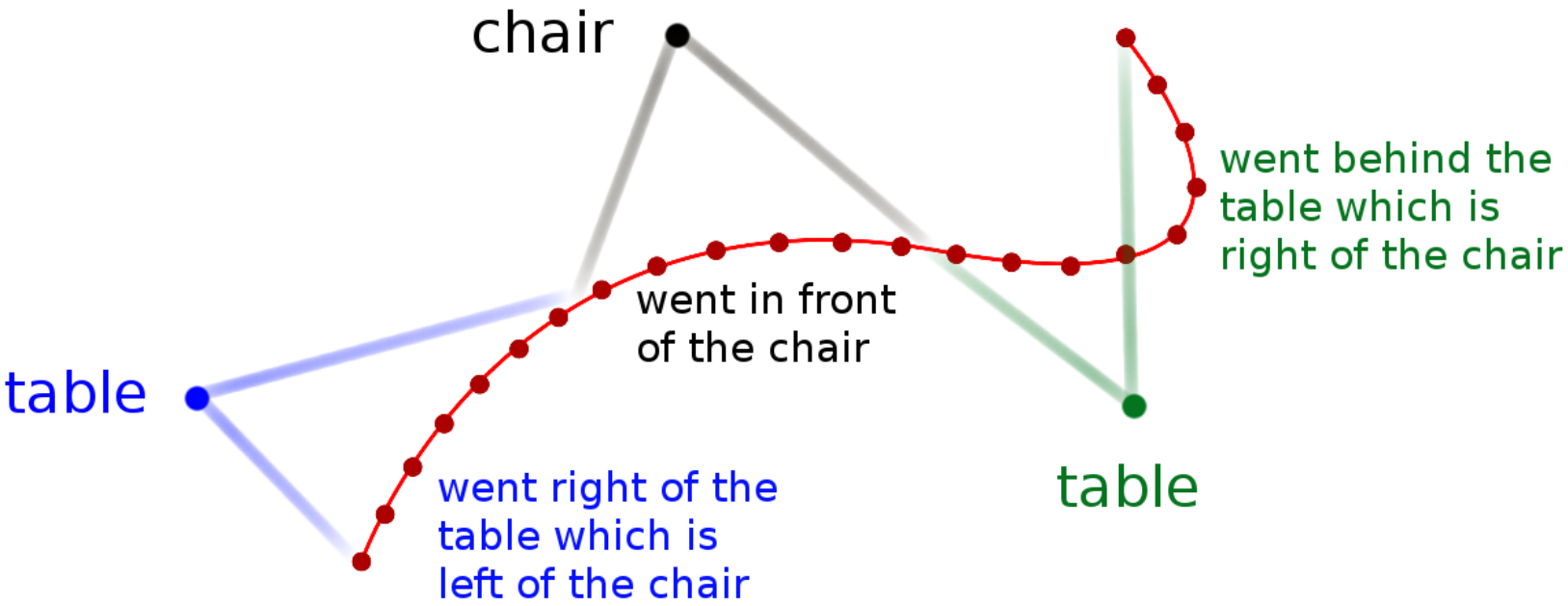
Acquisition Method

Iteration 3

toward the cone and right of the chair right of the table



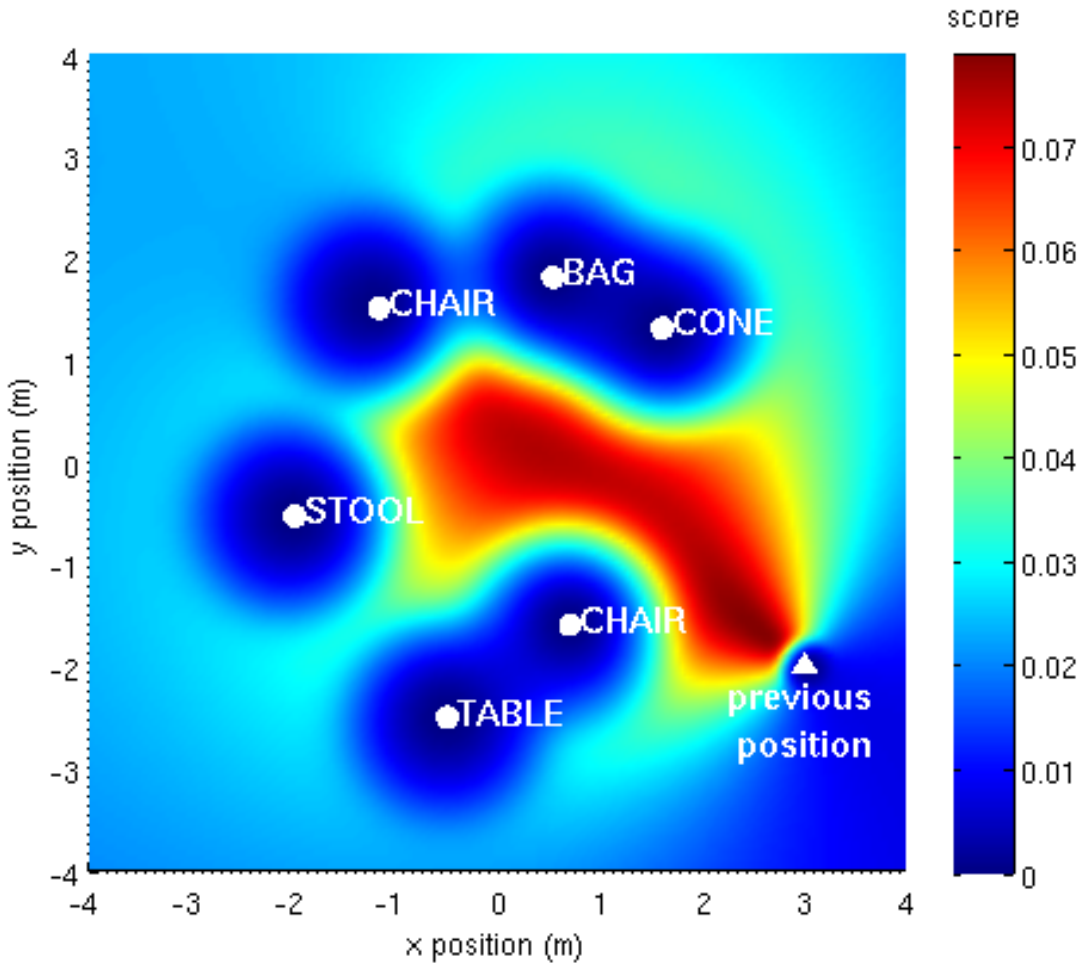
Generation Method



The robot went right of the table which is left of the chair, then went in front of the chair, then went behind the table which is right of the chair.

Comprehension Method

toward the chair left of the bag



Dataset

1. 10 random floorplans
   - 3 or 4 objects, at most one duplicate
   - tile corners (not perimeter)
2. 25 random sentences per floorplan
3. manually drive 250 paths; recover paths from odometry
4. get 3 AMT sentences for each path, 750 total
5. get AMT judgments for each sentence-path pair

■ acquisition (human sentences vs. human-driven paths)
tests human performance because testing human sentences produced after paths

Dataset

1. 10 random floorplans
   - 4 or 5 objects, at most one duplicate
   - tile corners, centers, or edge centers (not perimeter)
2. 10 random sentences per floorplan
3. automatically drive 100 paths; recover paths from odometry
4. get 3 AMT sentences for each path, 300 total
5. get AMT judgments for each sentence-path pair

comprehension (human sentences vs. machine-driven paths)
tests human performance because testing human sentences produced after paths

Dataset

1. 10 random floorplans
   - 4 or 5 objects, at most one duplicate
   - tile corners, centers, or edge centers (not perimeter)
2. 10 random sentences per floorplan
3. automatically drive 100 paths; recover paths from odometry
4. get 3 AMT sentences for each path, 300 total
5. automatically drive 300 paths; recover paths from odometry
6. get AMT judgments for each sentence-path pair

comprehension (human sentences vs. machine-driven paths)
tests machine performance because testing machine production of paths after human production of sentences

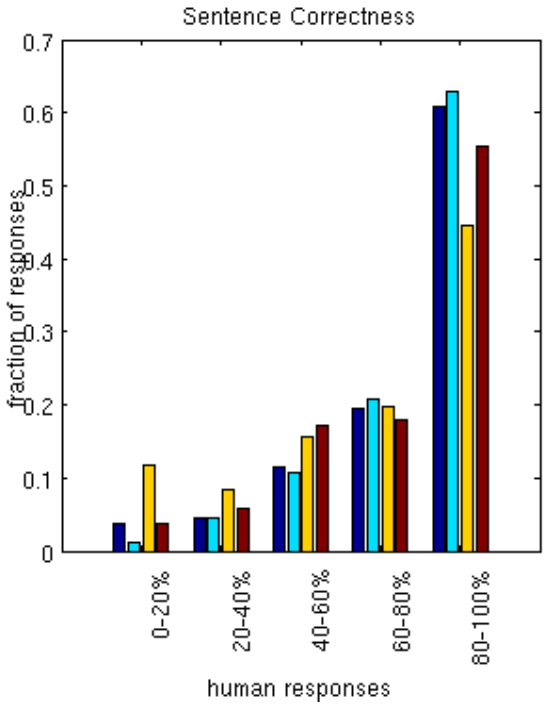
Dataset

1. 10 random floorplans
   - 4 or 5 objects, at most one duplicate
   - tile corners, centers, or edge centers (not perimeter)
2. 10 random sentences per floorplan
3. manually drive 100 paths; recover paths from odometry
4. generate 100 sentences
5. get AMT judgments for each sentence-path pair

■ generation (machine sentences vs. human-driven paths)
tests machine performance because testing machine production of sentences from paths

Sentence Correctness

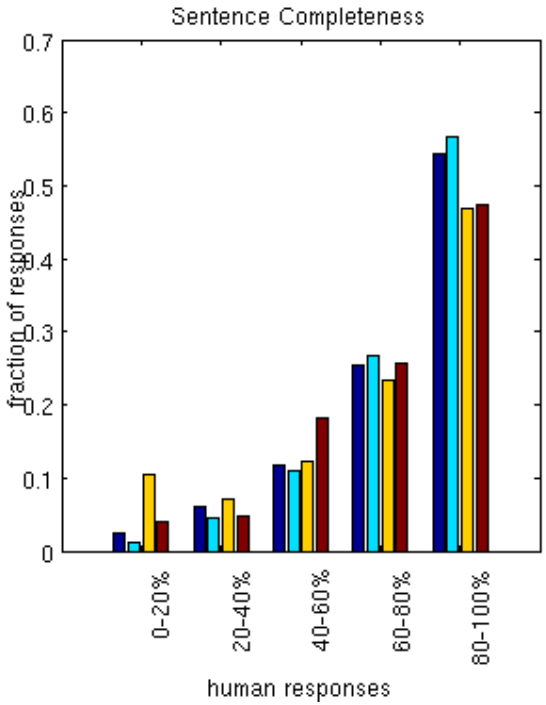
Approximately how much of the sentence is true of the path?



- acquisition (human sentences vs. human-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- generation (machine sentences vs. human-driven paths)

Sentence Completeness

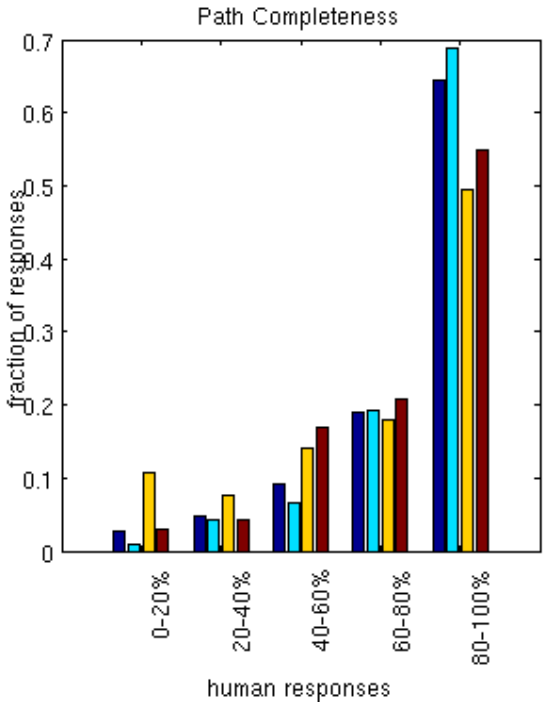
Approximately how much of the path is described by the sentence?



- acquisition (human sentences vs. human-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- generation (machine sentences vs. human-driven paths)

Path Completeness

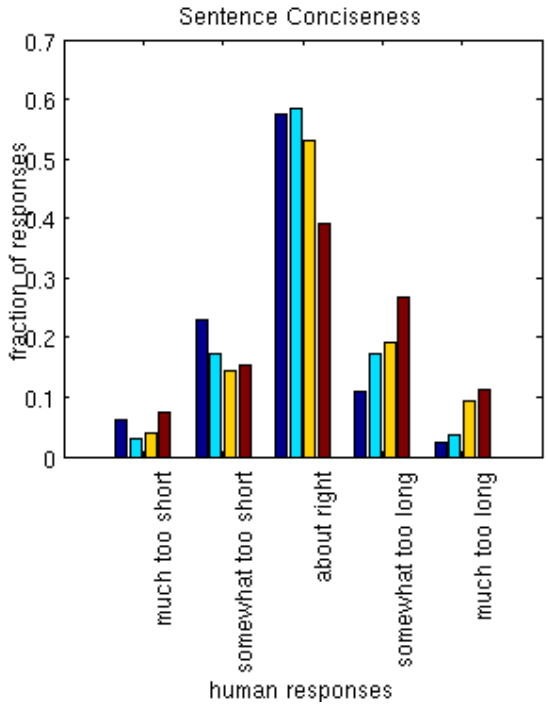
Approximately how much of the sentence is depicted by the path?



- acquisition (human sentences vs. human-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- generation (machine sentences vs. human-driven paths)

Sentence Conciseness

Rate the length of the sentence.

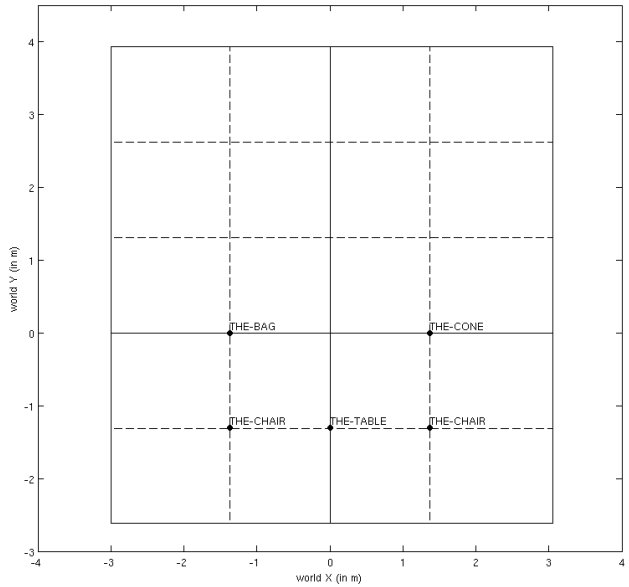


- acquisition (human sentences vs. human-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- comprehension (human sentences vs. machine-driven paths)
- generation (machine sentences vs. human-driven paths)

Outline

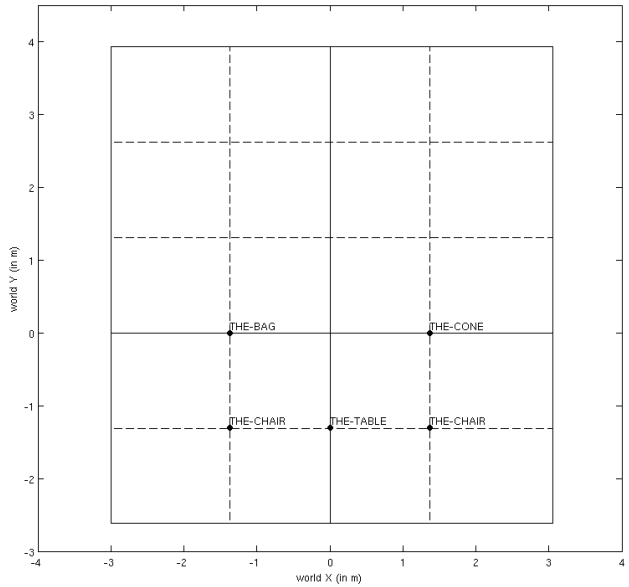
1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

Floorplans



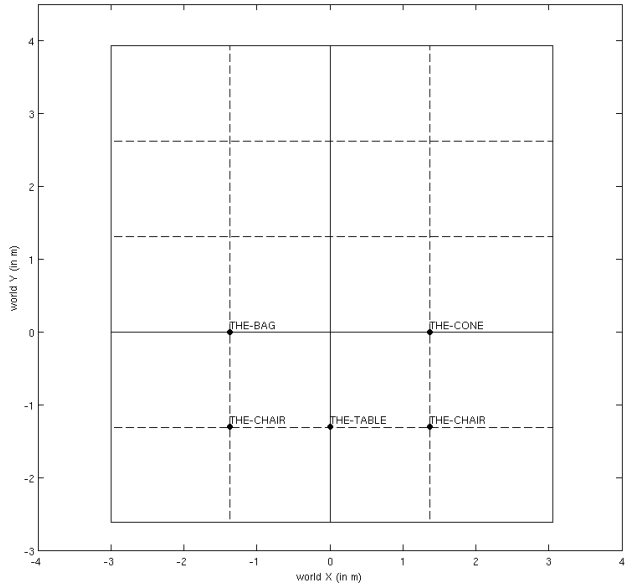
- So far, acquisition, generation, and comprehension all required floorplan as input.

Floorplans



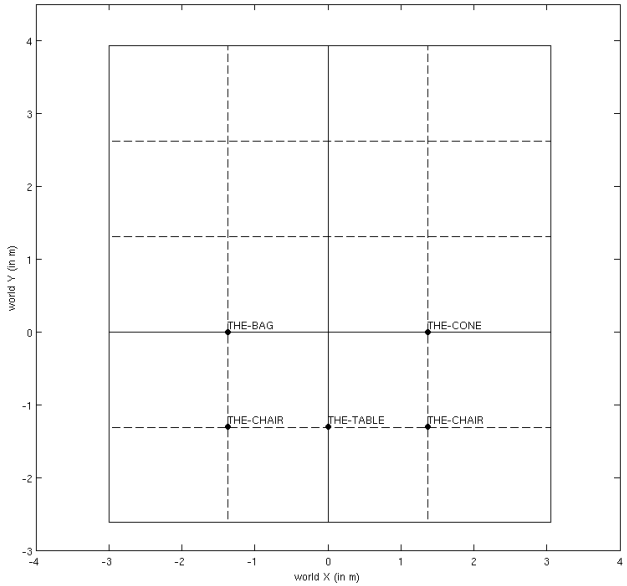
- So far, acquisition, generation, and comprehension all required floorplan as input.
- The floorplan took the form of a set of 2D points labeled with abstract classes.
  $\{(5, -3) : \textbf{foo}, (-7, 3) : \textbf{bar}\}$

Floorplans



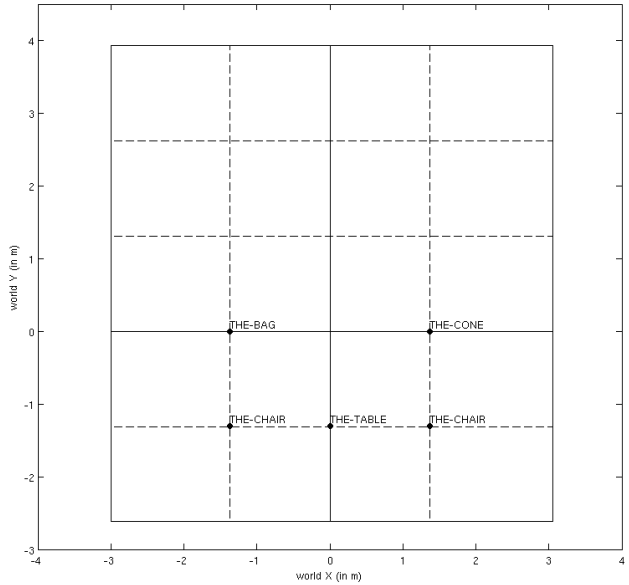
- The class labeling across different floorplans must be consistent; two instances of the same object class (in the same or different floorplans) should have the same label. This is what allows acquisition to work.

Floorplans



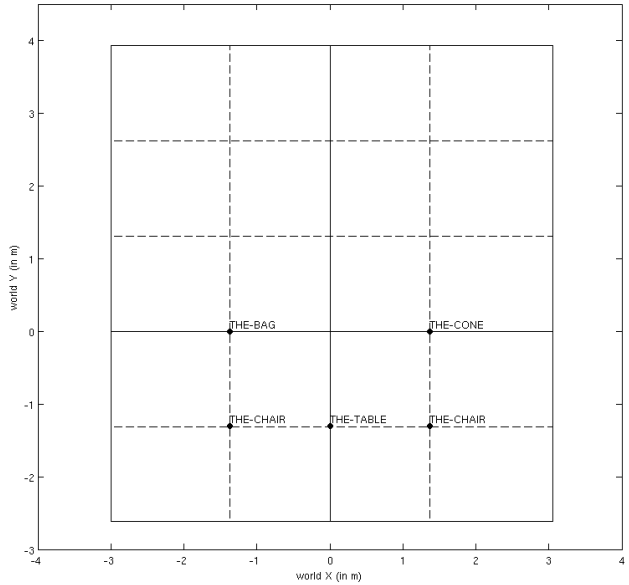
- The class labeling across different floorplans must be consistent; two instances of the same object class (in the same or different floorplans) should have the same label. This is what allows acquisition to work.
- The mapping from nouns to abstract class labels is ***learned*** (by the acquisition process).

Floorplans



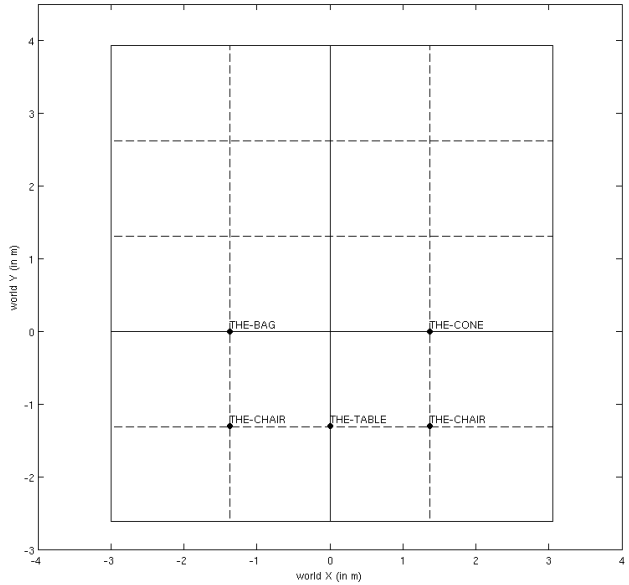
- It need not be a bijection.

Floorplans



- It need not be a bijection.
  - A noun can correspond to more than one abstract class label (homonymy).

Floorplans



- It need not be a bijection.
  - A noun can correspond to more than one abstract class label (homonymy).
  - An abstract class label can correspond to more than one noun (synonymy).

Using Codetection to Recover Floorplans

- Compute floorplan automatically from video stream and odometry using codetection

Using Codetection to Recover Floorplans

- Compute floorplan automatically from video stream and odometry using codetection
- Different from prior work on codetection

Using Codetection to Recover Floorplans

- Compute floorplan automatically from video stream and odometry using codetection
- Different from prior work on codetection
  1. egocentric video from a moving camera (changing position and orientation)

Using Codetection to Recover Floorplans

- Compute floorplan automatically from video stream and odometry using codetection
- Different from prior work on codetection
  1. egocentric video from a moving camera (changing position and orientation)
  2. integrates video stream with odometry and inertial guidance

Using Codetection to Recover Floorplans

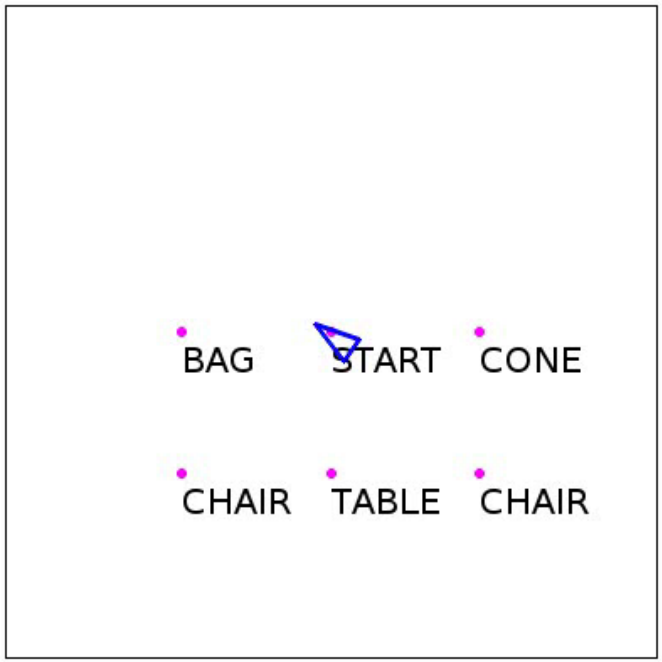
- Compute floorplan automatically from video stream and odometry using codetection
- Different from prior work on codetection
  1. egocentric video from a moving camera (changing position and orientation)
  2. integrates video stream with odometry and inertial guidance
  3. localizes in 3D, not just 2D

Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

Trace

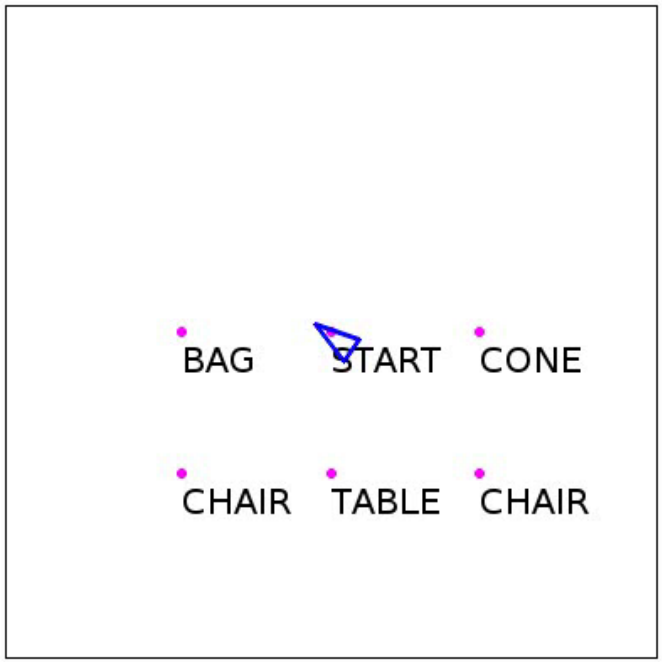


Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

Trace



Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

Trace

Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

Trace

Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

Trace

Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

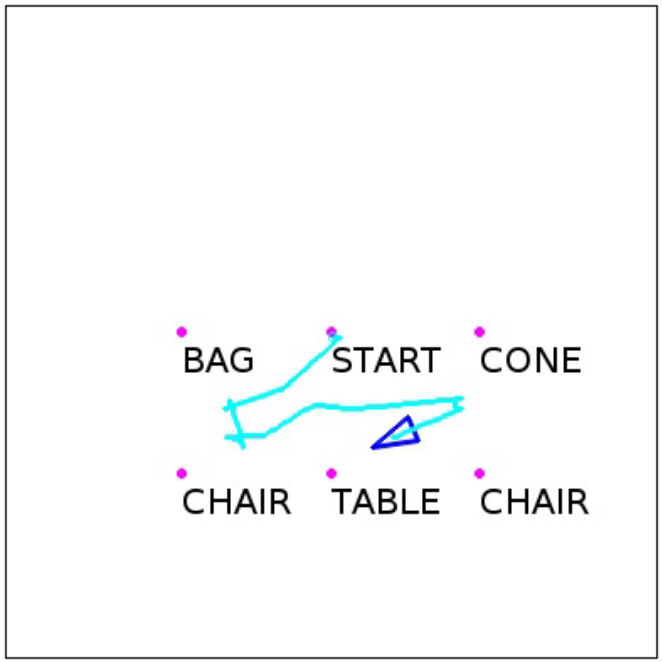
Trace

Data Collection

The robot went towards the chair which is left of the table then went away from the cone then went away from the bag then went behind the chair which is right of the table then went towards the table.

Video

Trace



. . . plus 59 more

Method

Codetection approach is a five-step process:

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
2. **Proposal Localization** Find 3D world location for each proposal box.

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
2. **Proposal Localization** Find 3D world location for each proposal box.
3. **Proposal Selection** Select at most a single proposal for each frame that denotes the prominent object in the field of view. Some frames will not have a selected proposal because there may not be a prominent object in the field of view.

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
2. **Proposal Localization** Find 3D world location for each proposal box.
3. **Proposal Selection** Select at most a single proposal for each frame that denotes the prominent object in the field of view. Some frames will not have a selected proposal because there may not be a prominent object in the field of view.
4. **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
2. **Proposal Localization** Find 3D world location for each proposal box.
3. **Proposal Selection** Select at most a single proposal for each frame that denotes the prominent object in the field of view. Some frames will not have a selected proposal because there may not be a prominent object in the field of view.
4. **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.
5. **Labeling** Assign an abstract class label to each localized object instance.

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
2. **Proposal Localization** Find 3D world location for each proposal box.
3. **Proposal Selection** Select at most a single proposal for each frame that denotes the prominent object in the field of view. Some frames will not have a selected proposal because there may not be a prominent object in the field of view.
4. **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.
5. **Labeling** Assign an abstract class label to each localized object instance.

Steps 1–4 are done independently for each floorplan, but jointly across all paths driven in that floorplan.

Method

Codetection approach is a five-step process:

1. **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
2. **Proposal Localization** Find 3D world location for each proposal box.
3. **Proposal Selection** Select at most a single proposal for each frame that denotes the prominent object in the field of view. Some frames will not have a selected proposal because there may not be a prominent object in the field of view.
4. **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.
5. **Labeling** Assign an abstract class label to each localized object instance.

Steps 1–4 are done independently for each floorplan, but jointly across all paths driven in that floorplan.
Step 5 is done jointly across all floorplans.

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries
     - close to any single image boundary and exceed specified height or width

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries
     - close to any single image boundary and exceed specified height or width
     - exceed both specified height and width

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries
     - close to any single image boundary and exceed specified height or width
     - exceed both specified height and width
     - outside floorplan

Proposal Generation and Selection

1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries
     - close to any single image boundary and exceed specified height or width
     - exceed both specified height and width
     - outside floorplan
   - edge score is weighted sum of

Proposal Generation and Selection

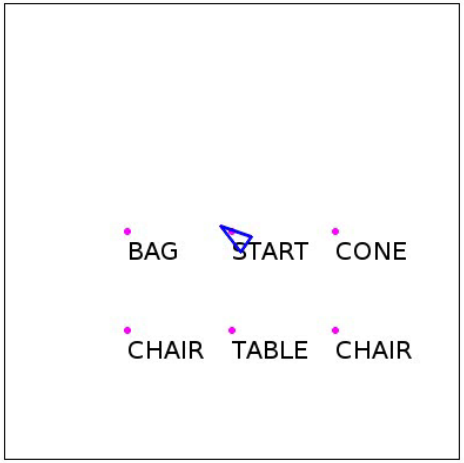
1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries
     - close to any single image boundary and exceed specified height or width
     - exceed both specified height and width
     - outside floorplan
   - edge score is weighted sum of
     - similarity of SIFT descriptors

Proposal Generation and Selection

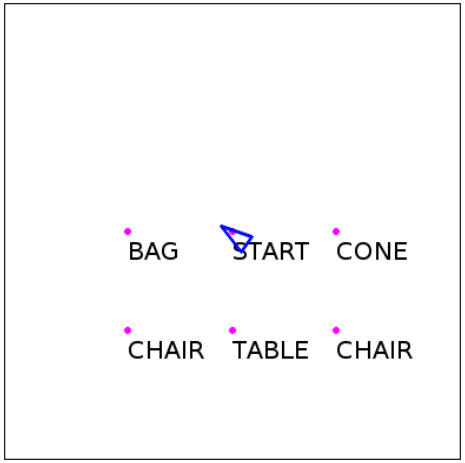
1. Generate proposals with MCG (Arbelaez et al. 2014)
2. Graphical model
   - vertex for each frame to denote the most prominent object in that frame
   - vertex labels range over proposals plus dummy
   - complete graph except no self edges
   - proposal score as vertex score, penalized by implausibility of world size and position recovered with projective geometry
     - behind camera (bottom edge above horizon)
     - close to any two image boundaries
     - close to any single image boundary and exceed specified height or width
     - exceed both specified height and width
     - outside floorplan
   - edge score is weighted sum of
     - similarity of SIFT descriptors
     - similarity of world size and position as determined by projective geometry

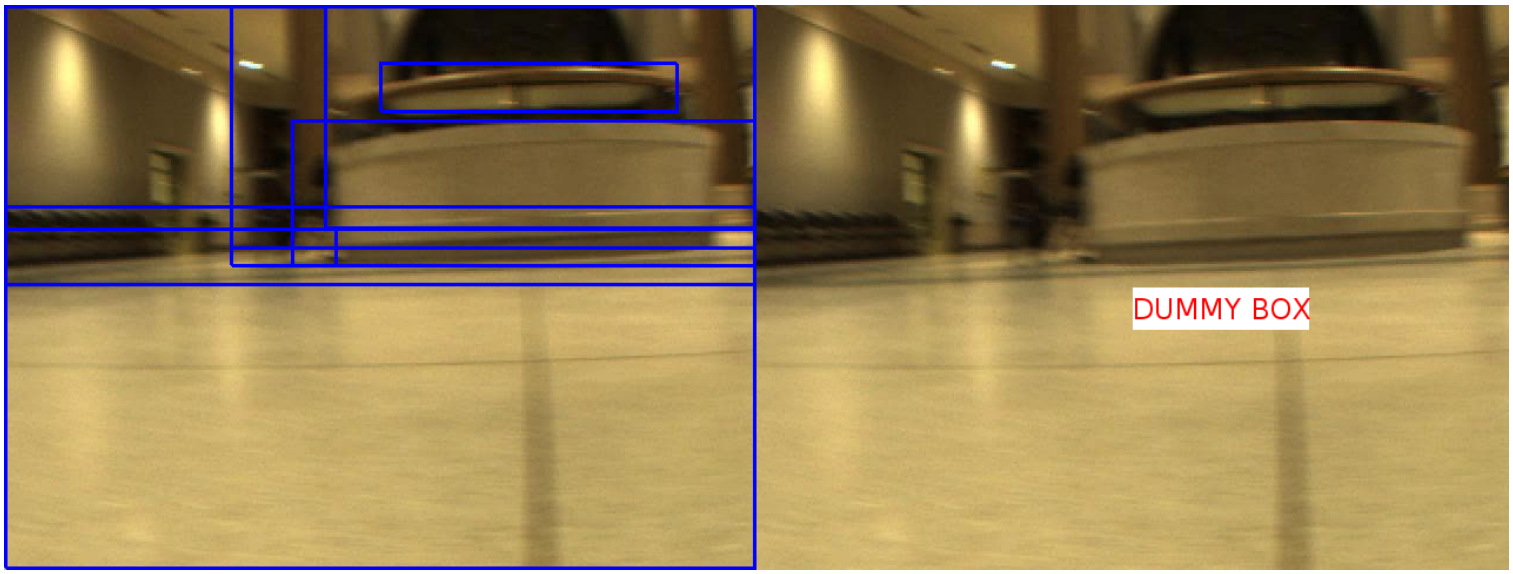
Proposal Generation, Selection, and Localization

raw proposals



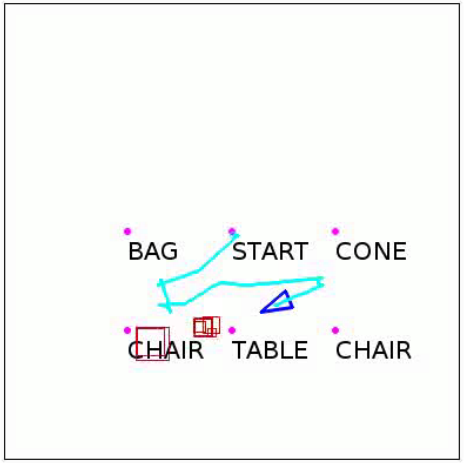
selected proposal (one per frame)



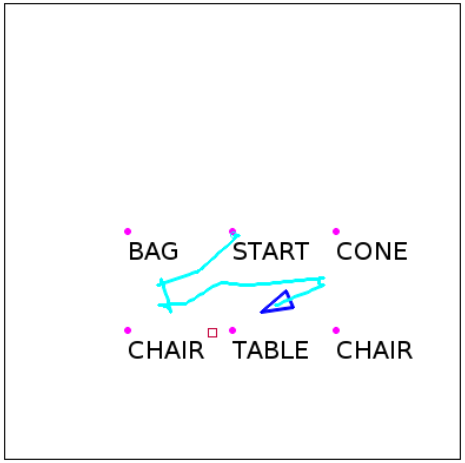


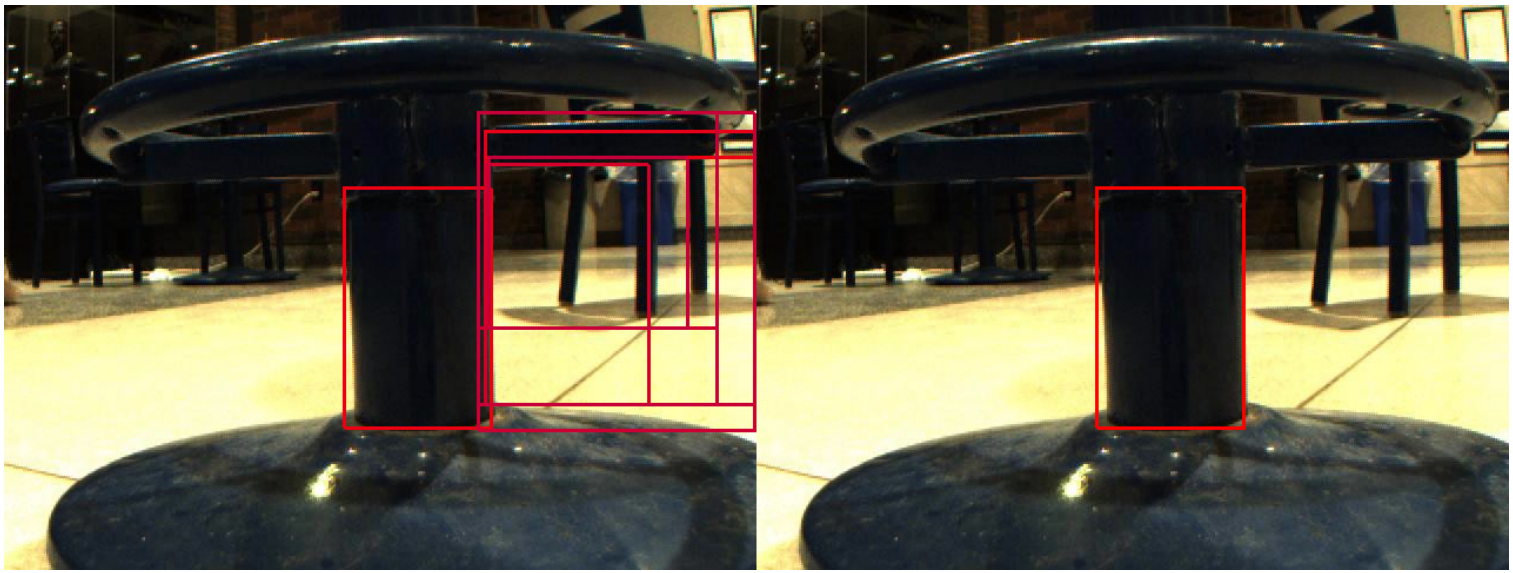
Proposal Generation, Selection, and Localization

raw proposals



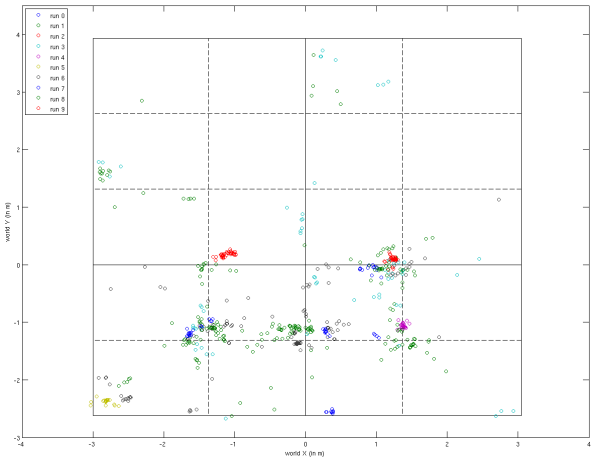
selected proposal (one per frame)





Clustering

Take selected proposal locations for all navigational paths on a floor plan

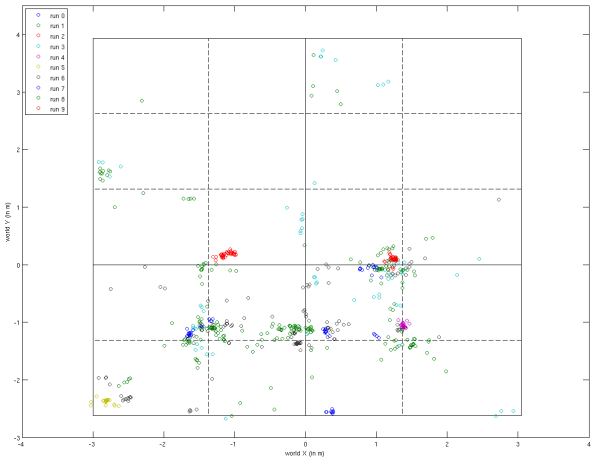


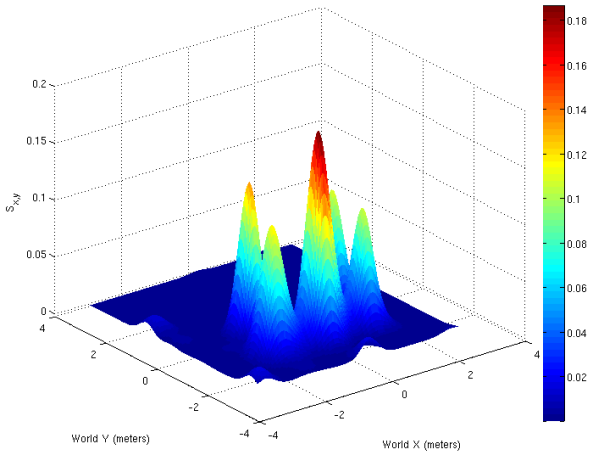
Clustering

Take selected proposal locations for all navigational paths on a floor plan

Compute density function

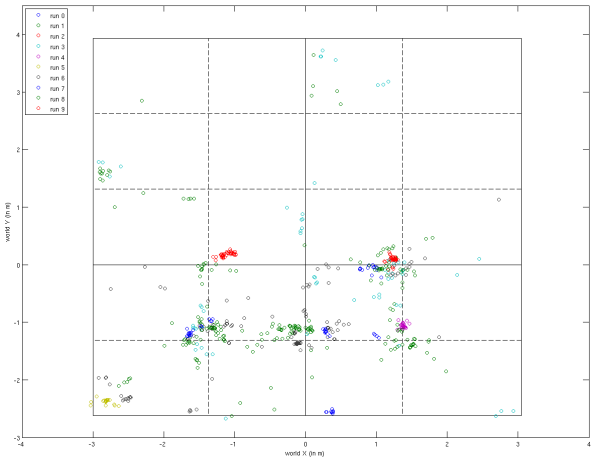
$$S_{x,y} = \sum_{n=1}^{N} f_n \frac{\|(x,y) - (x_n, y_n)\|}{v_n}$$





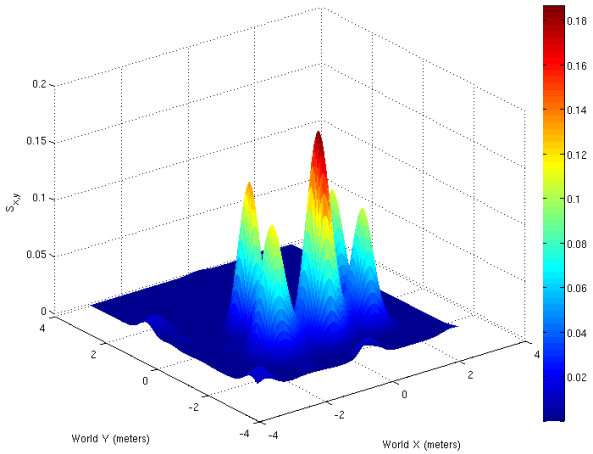
Clustering

Take selected proposal locations for all navigational paths on a floor plan

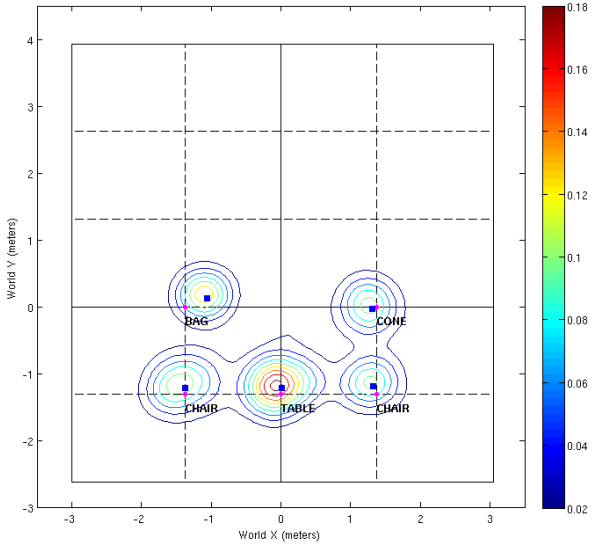


Compute density function

$$S_{x,y} = \sum_{n=1}^{N} f_n \frac{\|(x,y) - (x_n, y_n)\|}{v_n}$$



Find peaks to locate objects



Labeling

1. Assign proposals to closest peak, rejecting outliers.

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

5. Form a graphical model

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

5. Form a graphical model
   - vertex for each peak

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

5. Form a graphical model
   - vertex for each peak
   - vertex labels range over abstract object classes

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

5. Form a graphical model
   - vertex for each peak
   - vertex labels range over abstract object classes
   - complete graph except no self edges

Labeling

1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

5. Form a graphical model
   - vertex for each peak
   - vertex labels range over abstract object classes
   - complete graph except no self edges
   - no vertex score

Labeling

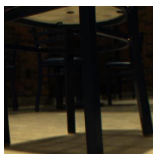
1. Assign proposals to closest peak, rejecting outliers.
2. Let C_i denote the proposals assigned to peak i.
3. Let U_{ab} denote the similarity between proposals a and b.
4. Compute similarity Q_{ij} between peaks i and j.

$$Q_{ij} = \frac{\sum_{a \in C_i} \max_{b \in C_j} U_{ab} + \sum_{b \in C_j} \max_{a \in C_i} U_{ab}}{|C_i| + |C_j|}$$

5. Form a graphical model
   - vertex for each peak
   - vertex labels range over abstract object classes
   - complete graph except no self edges
   - no vertex score
   - edge score high if same label and high similarity or different label and low similarity

Results

Labeling output:

BAG	**BOX**	**CHAIR**	**CONE**	**STOOL**	**TABLE**
class labels: 1, 2	class labels: 3, 4, 5	class labels: 6, 7, 8, 9	class label: 10	class label: 11	class label: 12
					
					
					

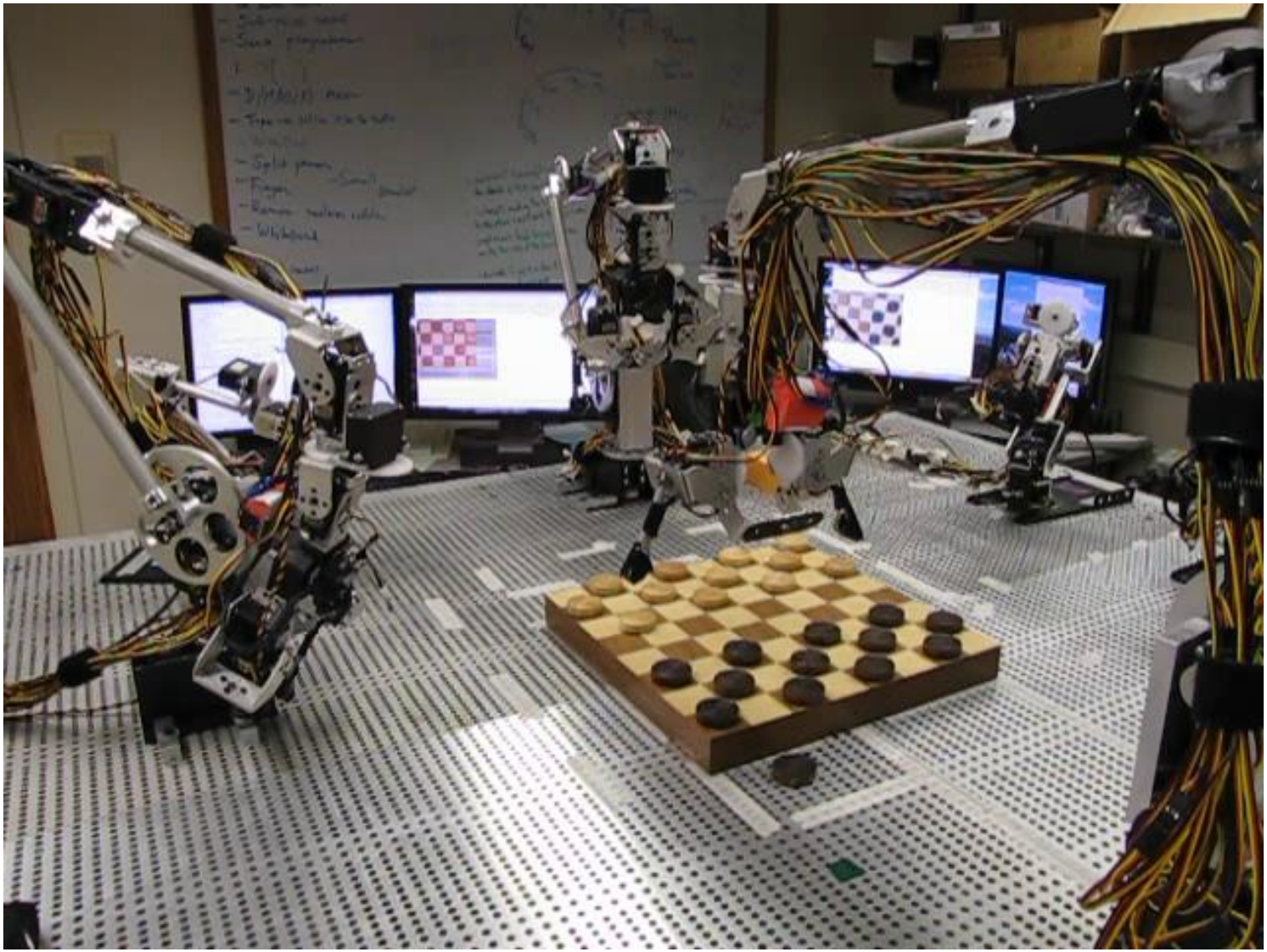
Outline

1. The Sentence Tracker
2. Sentence Directed Video Object Codetection
3. Driving Under the Influence (of Language)
   - Grounding Language Semantics in Robotics
   - Object Codetection from Mobile Robot Video
4. Playing Checkers from English

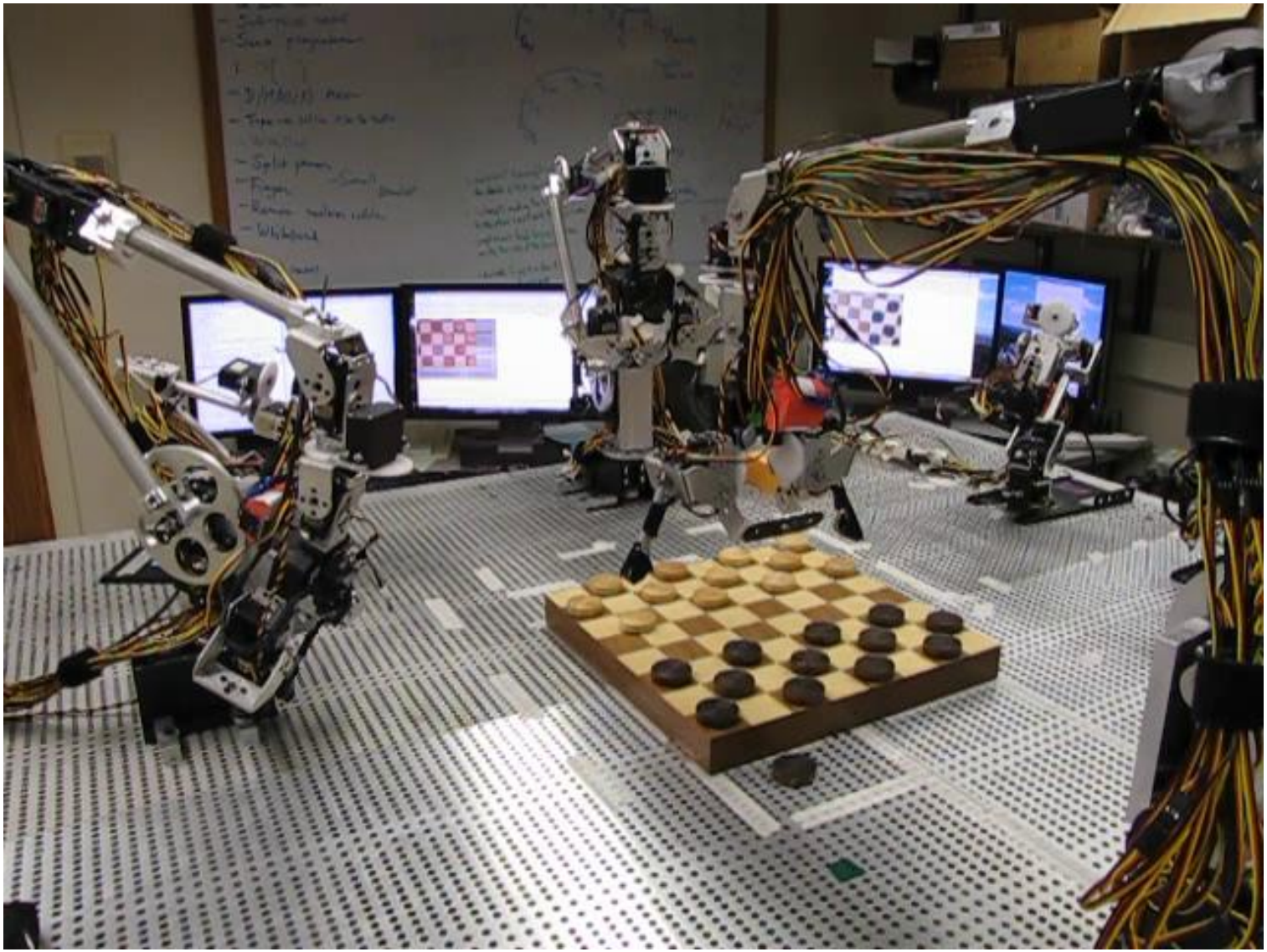
Joint work

Daniel Paul Barrett Seth Benjamin Zachary Burchill

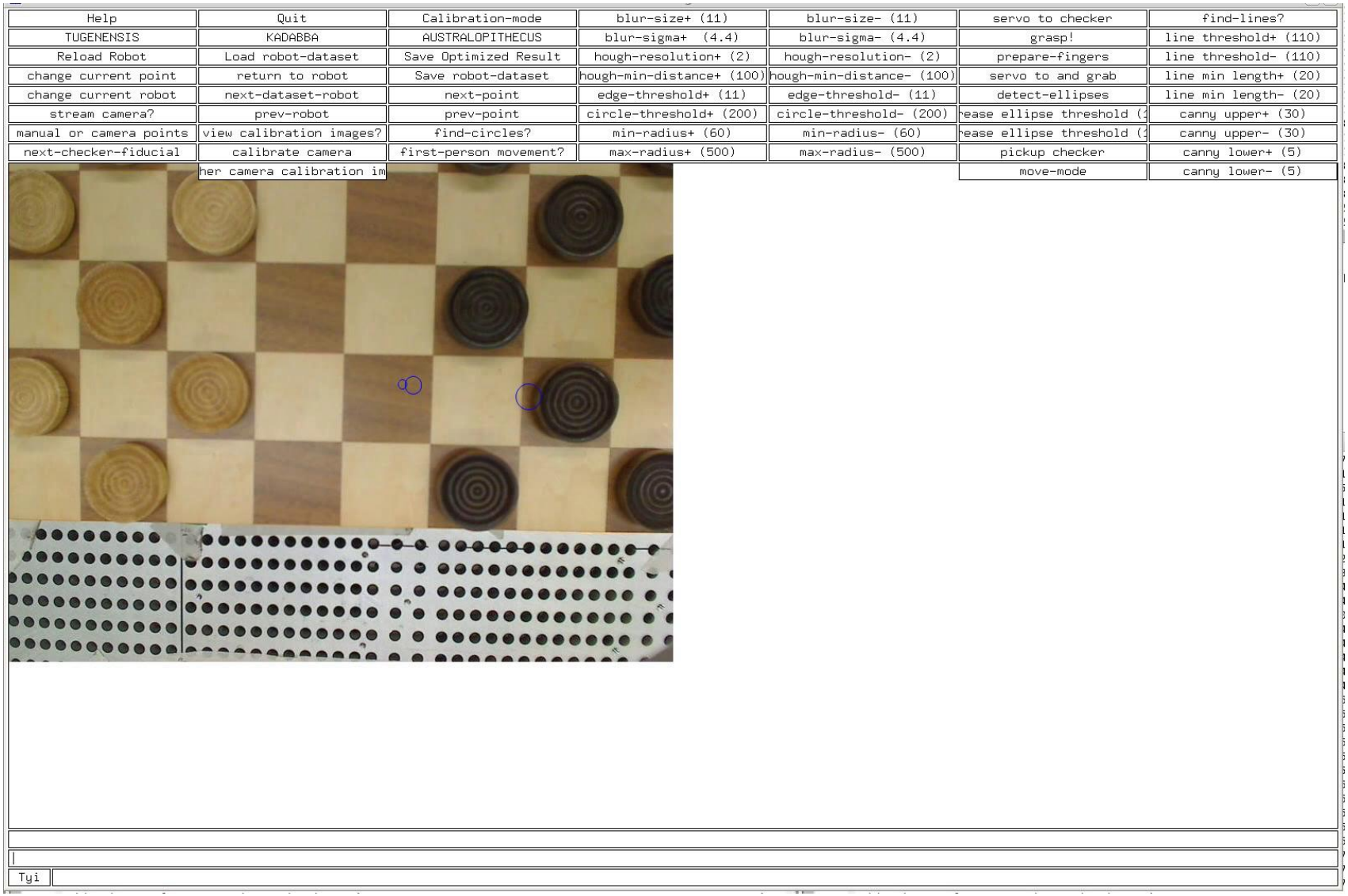
Two Robots Playing Checkers



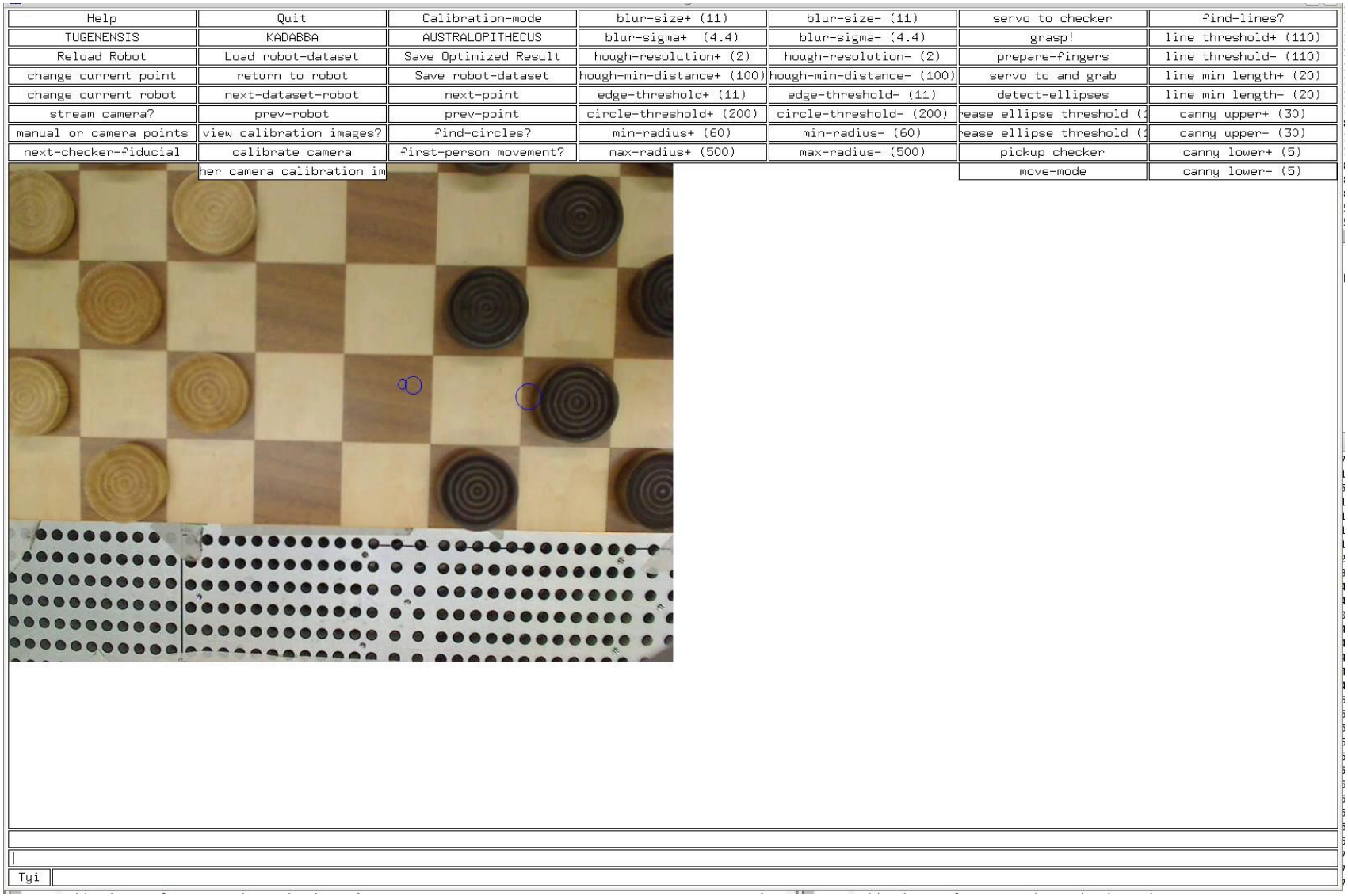
Two Robots Playing Checkers



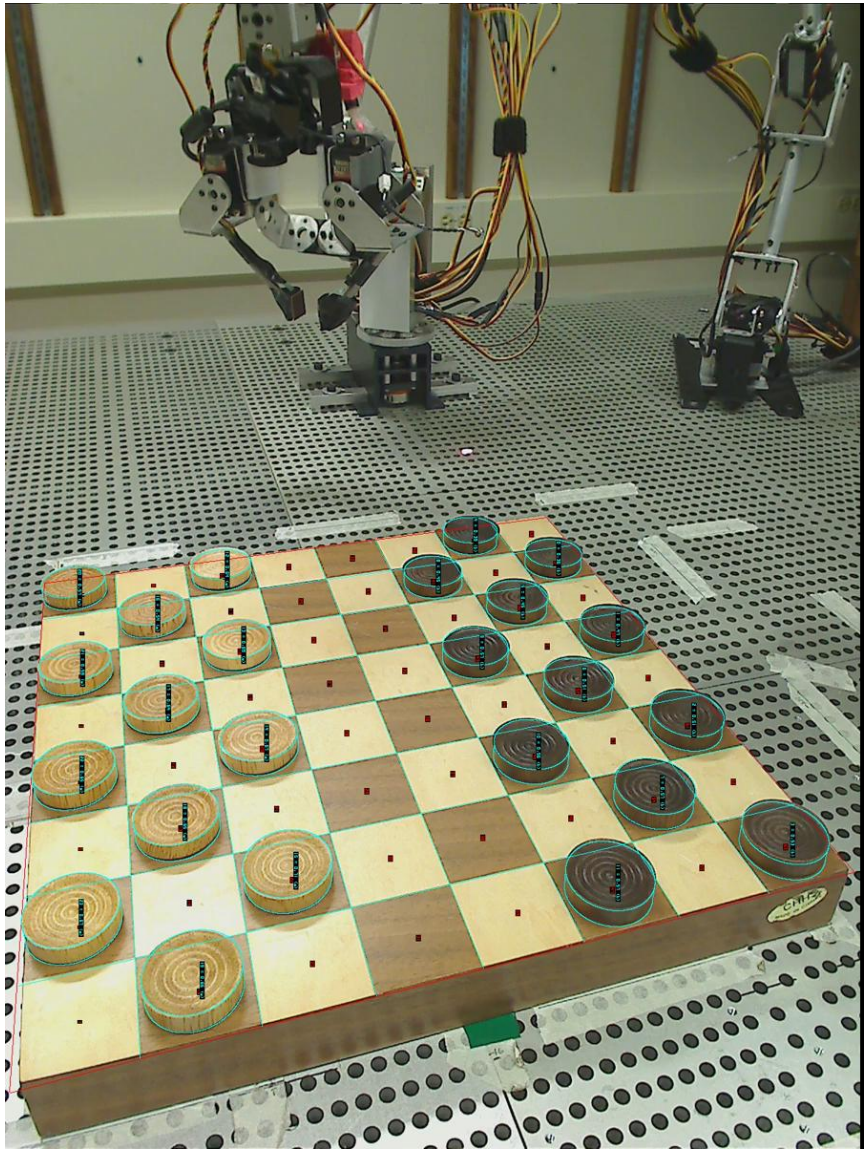
View from the Palm Camera



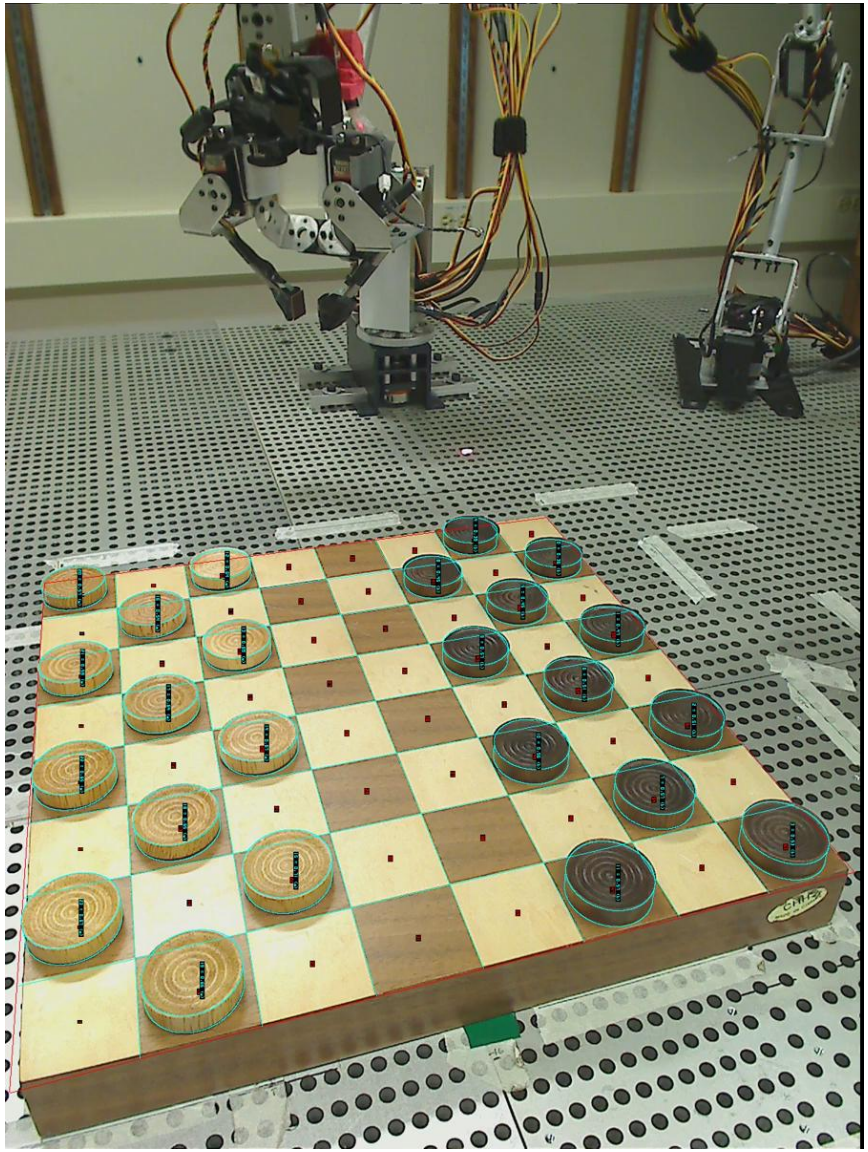
View from the Palm Camera



Recovering Checkers Game State from Computer Vision



Recovering Checkers Game State from Computer Vision



Sources of English Rules for Checkers

1. http://boardgames.about.com/cs/checkersdraughts/ht/play_checkers.htm
2. http://simple.wikipedia.org/wiki/Checkers
3. http://www.darkfish.com/checkers/rules.html
4. http://www.ducksters.com/games/checkers_rules.php
5. http://www.7is7.com/software/games/checkers/rules-en-us.html
6. http://www.chesslab.com/rules/checkersbasics.html
   http://www.chesslab.com/rules/checkersrules.html
7. http://www.learnplaywin.net/checkers/checkers-rules.htm
8. http://www.gametableonline.com/pop_rules.php?gid=20
9. http://www.indepthinfo.com/checkers/setup.shtml
   http://www.indepthinfo.com/checkers/play.shtml
   http://www.indepthinfo.com/checkers/crowning.shtml
10. http://winning-moves.com/images/kingmerulesv2.pdf
11. http://www.itsyourturn.com/t_helptopic2030.html
12. http://www.wikihow.com/Play-Checkers
13. http://www.yourturnmyturn.com/rules/checkers.php
14. http://www.flyordie.com/games/help/checkers/en/games_rules_checkers.html
15. http://brainking.com/en/GameRules?tp=7
16. http://www.netintellgames.com/checkersrules.htm
17. http://www.pcmag.com/article2/0,2817,1161217,00.asp
18. http://www.howcast.com/videos/297-how-to-play-checkers/
19. http://www.gamblingsites.com/skill-games/checkers/
20. http://www.mundigames.com/multiplayer/checkers/rules/

Rule Set #1

Part 1

Checkers is played by two players. Each player begins the game with 12 colored discs. (Typically, one set of pieces is black and the other red.)
The board consists of 64 squares, alternating between 32 dark and 32 light squares. It is positioned so that each player has a light square on the right side corner closest to him or her.
Each player places his or her pieces on the 12 dark squares closest to him or her.
Black moves first. Players then alternate moves.
Moves are allowed only on the dark squares, so pieces always move diagonally.
Single pieces are always limited to forward moves (toward the opponent).
A piece making a non-capturing move (not involving a jump) may move only one square.
A piece making a capturing move (a jump) leaps over one of the opponent's pieces, landing in a straight diagonal line on the other side. Only one piece may be captured in a single jump; however, multiple jumps are allowed on a single turn.
When a piece is captured, it is removed from the board.
If a player is able to make a capture, there is no option -- the jump must be made. If more than one capture is available, the player is free to choose whichever he or she prefers.
When a piece reaches the furthest row from the player who controls that piece, it is crowned and becomes a king. One of the pieces which had been captured is placed on top of the king so that it is twice as high as a single piece.
Kings are limited to moving diagonally, but may move both forward and backward. (Remember that single pieces, i.e. non-kings, are always limited to forward moves.)

Rule Set #1

Part 2

```
Kings may combine jumps in several directions -- forward and backward -- on
the same turn.  Single pieces may shift direction diagonally during a multiple
capture turn, but must always jump forward (toward the opponent).
A player wins the game when the opponent cannot make a move.  In most cases,
this is because all of the opponent's pieces have been captured, but it could
also be because all of his pieces are blocked in.
```

Rule Set #2

Part 1

In most games of checkers, there are two players. The players are at opposite ends of the board. One player has dark pieces, and one player has light pieces. They take turns moving their pieces. Players move their pieces diagonally from one square to another square. When a player jumps over their opponent's (the other player's) piece, you take that piece from the board.

English checkers.
Most English-speaking people call English checkers "draughts". English 'checkers' is played on an 8x8 chess board. Only the dark squares are used (the light squares are never used). For that reason, good players play differently in the left and right corners.

Pieces.
The pieces are flat and round. They are referred to as "men". They are usually colored red and white. For this reason, the darker pieces are usually called "Red" and the lighter pieces are always called "White." Some checkers sets have red and black pieces. Then the red pieces are called "White" and the black pieces "Red." And many sets simply use black and white draughts. There are two kinds of pieces: plain (single) pieces and "kings". A king is made by putting one plain piece on top of another.

Starting position.
Each player starts with 12 pieces on the three rows closest to their own side. The row closest to each player is called the "King␣Row". The darker colour moves first.

Rule Set #2

Part 2

```
How to move.
A player can move in two ways.  A piece can be moved forward, diagonally, to
the very next dark square.  In some variants, if one player's piece, the other
player's piece, and an empty square are lined up, then the first player must
"jump" the other player's piece.  In this case, the first player jumps over
the other player's piece onto the empty square and takes the other player's
piece off the board.  However, this is an uncommon ruleset not commonly
observed in the Americas.  A player can also use one piece to make multiple
jumps in any one single turn, provided each jump continues to lead immediately
into the next jump and in a straight line.  Sometimes a player may have the
option or a choice of which opponent piece he must jump.  In such cases, he
may then choose which to jump.  If you keep your hand on any piece when you're
moving, you have the choice to put it back and move another piece.
```

Rule Set #2

Part 3

```
Kings.
If a player's piece moves into the King Row on the other player's side, it
becomes a king.  It can move forward and backward.  (Regular pieces can only
move forward.) A king cannot jump out of the King Row until the next turn.
Unlike Regular pieces, Kings can "jump" various empty boxes at a time to
capture a regular piece.  These "King Jumps" may only occur in diagonally
aligned boxes.  Neither Kings nor regular pieces may move in any direction
that is not diagonal.

How the game ends.
The first player to lose all of his or her pieces loses the game.  If no
players are able to move, the player with the most amount of pieces wins.  If
the players have the same amount of pieces, the player with the most kings
wins.  If the players have an equal number of pieces and the same number of
kings the game is a draw.
```

Generated ZRF for Rule Set #1

Part 1

```
(define PIECE-MOVE      ($1 (verify empty?)
        (if (in-zone? KING-transition)
                (add KING)
        else
                add
        )
))

(define PIECE-JUMP      ($1 (verify empty?)
        (if (in-zone? KING-transition)
                (add KING)
        else
                add
        )
))

(define KING-MOVE       ($1 (verify empty?)
                add

))

(define KING-JUMP       ($1 (verify empty?)
                add

))
```

Generated ZRF for Rule Set #1

Part 2

```
(game
        (title "checkers1")
        (players P1 P2)
        (turn-order P1 P2)
        (move-priorities MOVE JUMP)

        (board
                (grid
                        (dimensions
                                ("a/b/c/d/e/f/g/h") ; columns
                                ("8/7/6/5/4/3/2/1") ; rows
                        )
                        (directions
                                (n 0 -1) (w -1 0) (s 0 1) (e 1 0)
                                (ne 1 -1) (nw -1 -1) (se 1 1) (sw -1 1)
                        )
                )
                (symmetry P2 (n s) (s n) (ne sw) (sw ne) (nw se) (se nw))
                (zone (name KING-transition) (players P1)
                        (positions h8 g8 f8 e8 d8 c8 b8 a8)
                )
                (zone (name KING-transition) (players P2)
                        (positions h1 g1 f1 e1 d1 c1 b1 a1)
                )
        )
```

Generated ZRF for Rule Set #1

Part 3

```
(board-setup
        (P1 (PIECE g1 e1 c1 a1 h2 f2 d2 b2 g3 e3 c3 a3) )
        (P2 (PIECE h6 f6 d6 b6 g7 e7 c7 a7 h8 f8 d8 b8) )
)

(piece
        (name PIECE)
        (moves
                (move-type MOVE)
                (PIECE-MOVE nw)
                (PIECE-MOVE ne)

                (move-type JUMP)
                (PIECE-JUMP nw)
                (PIECE-JUMP ne)

        )
)
(piece
        (name KING)
        (moves
                (move-type MOVE)
                (KING-MOVE nw)
                (KING-MOVE ne)
                (KING-MOVE sw)
                (KING-MOVE se)
```

Generated ZRF for Rule Set #1

Part 4

```
                    (move-type JUMP)
                    (KING-JUMP nw)
                    (KING-JUMP ne)
                    (KING-JUMP sw)
                    (KING-JUMP se)

            )
        )

        (loss-condition (P1 P2 ) stalemated )
        (loss-condition (P1 P2 ) (pieces-remaining 0) )

)
```

Generated ZRF for Rule Set #2

Part 1

```
(define PIECE-MOVE      ($1 (verify empty?)
        (if (in-zone? KING-transition)
                (add KING)
        else
                add
        )
))

(define PIECE-JUMP      ($1 (verify empty?)
        (if (in-zone? KING-transition)
                (add KING)
        else
                add
        )
))

(define KING-MOVE       ($1 (verify empty?)
                add

))

(define KING-JUMP       ($1 (verify empty?)
                add

))
```

Generated ZRF for Rule Set #2

Part 2

```
(game
        (title "checkers2")
        (players P1 P2)
        (turn-order P1 P2)
        (move-priorities MOVE JUMP)

        (board
                (grid
                        (dimensions
                                ("a/b/c/d/e/f/g/h") ; columns
                                ("8/7/6/5/4/3/2/1") ; rows
                        )
                        (directions
                                (n 0 -1) (w -1 0) (s 0 1) (e 1 0)
                                (ne 1 -1) (nw -1 -1) (se 1 1) (sw -1 1)
                        )
                )
                (symmetry P2 (n s) (s n) (ne sw) (sw ne) (nw se) (se nw))
                (zone (name KING-transition) (players P1)
                        (positions h8 g8 f8 e8 d8 c8 b8 a8)
                )
                (zone (name KING-transition) (players P2)
                        (positions h1 g1 f1 e1 d1 c1 b1 a1)
                )
        )
```

Generated ZRF for Rule Set #2

Part 3

```
(board-setup
        (P1 (PIECE g1 e1 c1 a1 h2 f2 d2 b2 g3 e3 c3 a3) )
        (P2 (PIECE h6 f6 d6 b6 g7 e7 c7 a7 h8 f8 d8 b8) )
)

(piece
        (name PIECE)
        (moves
                (move-type MOVE)
                (PIECE-MOVE nw)
                (PIECE-MOVE ne)

                (move-type JUMP)
                (PIECE-JUMP nw)
                (PIECE-JUMP ne)

        )
)
(piece
        (name KING)
        (moves
                (move-type MOVE)
                (KING-MOVE nw)
                (KING-MOVE ne)
                (KING-MOVE sw)
                (KING-MOVE se)
```

Generated ZRF for Rule Set #2

Part 4

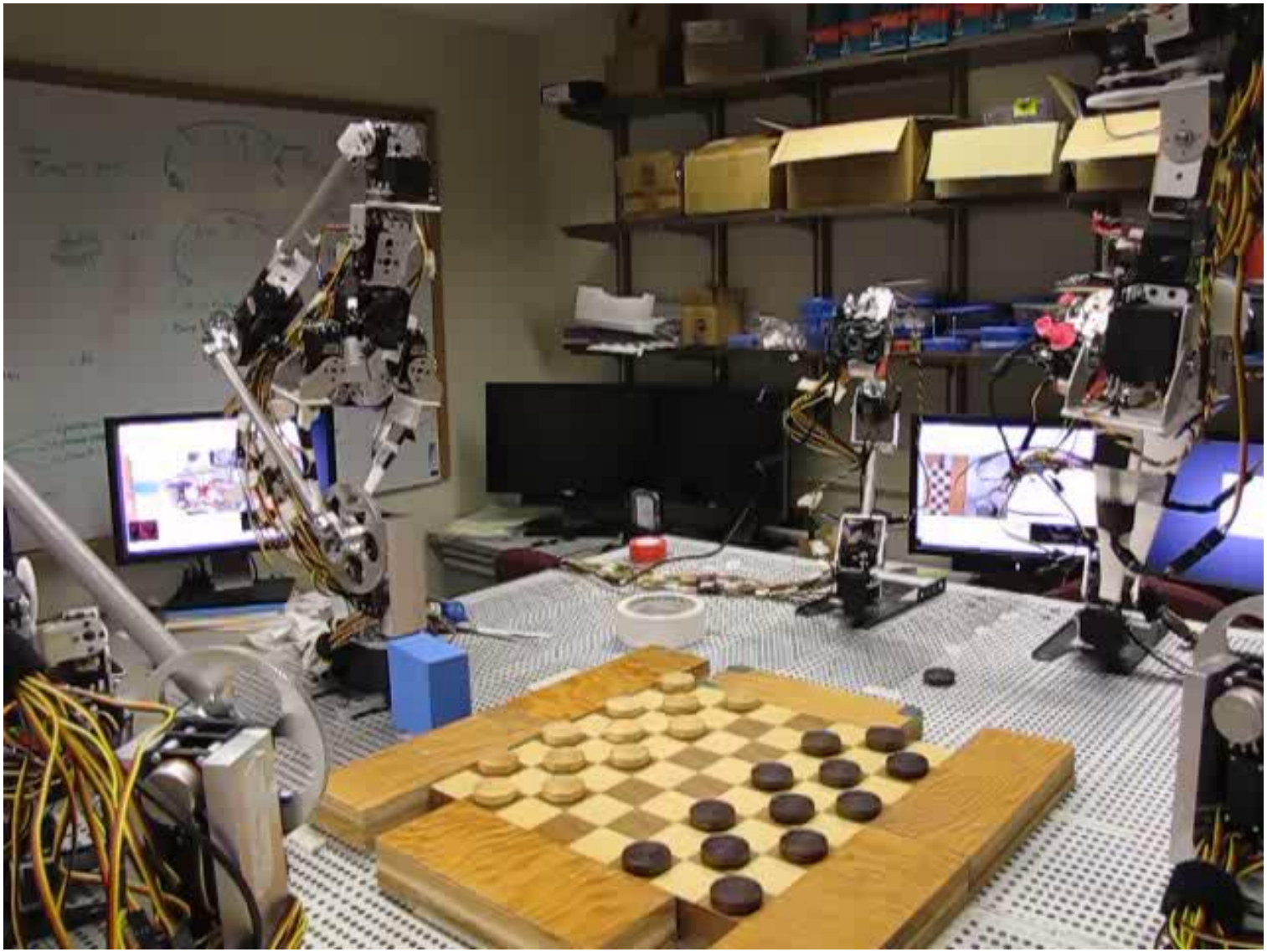
```
                    (move-type JUMP)
                    (KING-JUMP nw)
                    (KING-JUMP ne)
                    (KING-JUMP sw)
                    (KING-JUMP se)

            )
        )

        (loss-condition (P1 P2 ) (pieces-remaining 0) )

)
```

Two Robots Playing from Rule Sets #1 and #2



Two Robots Playing from Rule Sets #1 and #2