

Decoding the Brain to Help Build Machines

Jeffrey Mark Siskind, `qobi@purdue.edu`



Columbia, Friday 16 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



- 1 Verbs, Arguments, and Predication in the Human Brain
 - Experiment 1: ME-Y2
 - Experiment 2: hollywood2-text-speech
 - Experiment 3: compositionality-noninterleaved
- 2 The Sentence Tracker
- 3 Sentence Directed Video Object Codetection
- 4 Driving Under the Influence (of Language)
 - Grounding Language Semantics in Robotics
 - Object Codetection from Mobile Robot Video
- 5 Playing Checkers from English

1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

5 Playing Checkers from English

Joint work

Andrei Barbu
Scott Alan Bronikowski
Christiane D. Fellbaum
Sébastien Hélie
N. Siddharth
Caiming Xiong

Daniel Paul Barrett
Zachary Burchill
Catherine Hanson
Evguenia Malaia
Thomas Michael Talavage
Haonan Yu

Charles Roger Bradley
Jason J. Corso
Stephen José Hanson
Barak A. Pearlmutter
Ronnie B. Wilbur

- 1 Verbs, Arguments, and Predication in the Human Brain
 - Experiment 1: ME-Y2
 - Experiment 2: hollywood2-text-speech
 - Experiment 3: compositionality-noninterleaved
- 2 The Sentence Tracker
- 3 Sentence Directed Video Object Codetection
- 4 Driving Under the Influence (of Language)
 - Grounding Language Semantics in Robotics
 - Object Codetection from Mobile Robot Video
- 5 Playing Checkers from English

One Slide Tells it All



video

One Slide Tells it All



video



computer

One Slide Tells it All



video



computer

$\approx 50\%$

accuracy

One Slide Tells it All



video



computer

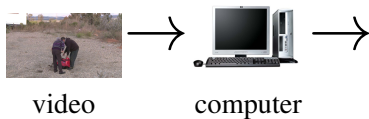
$\approx 50\%$

accuracy

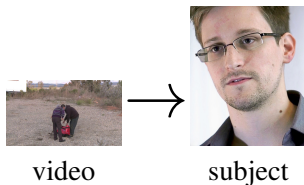


video

One Slide Tells it All



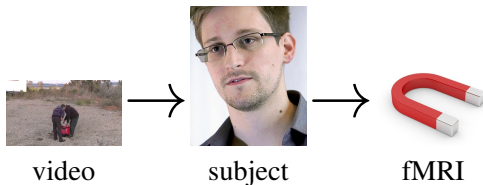
$\approx 50\%$
accuracy



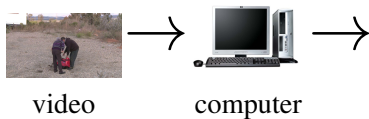
One Slide Tells it All



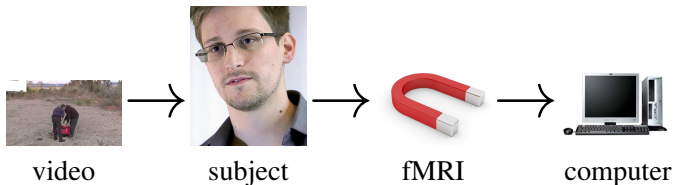
≈ 50%
accuracy



One Slide Tells it All



$\approx 50\%$
accuracy



One Slide Tells it All



video



computer



$\approx 50\%$

accuracy



video



subject



fMRI



computer



$\approx 70\%$

accuracy



Barbu, A., Barrett, D.P., Chen, W., Siddharth, N., Xiong, C., Corso, J.J., Fellbaum, C.D., Hanson, C., Hanson, S.J., Hélie, S., Malaia, E., Pearlmutter, B.A., Siskind, J.M., Talavage, T.M., and Wilbur, R.B., 'Seeing is Worse than Believing: Reading People's Minds Better than Computer-Vision Methods Recognize Actions,' ECCV 2014.



Barbu, A., Barrett, D.P., Chen, W., Siddharth, N., Xiong, C., Corso, J.J., Fellbaum, C.D., Hanson, C., Hanson, S.J., Hélie, S., Malaia, E., Pearlmutter, B.A., Siskind, J.M., Talavage, T.M., and Wilbur, R.B., 'Seeing is Worse than Believing: Reading People's Minds Better than Computer-Vision Methods Recognize Actions,' ECCV 2014.



⇒ $\left\{ \begin{array}{l} \textit{carry}, \\ \textit{dig}, \\ \textit{hold}, \\ \textit{pick up}, \\ \textit{put down}, \\ \textit{walk} \end{array} \right.$

Barbu, A., Barrett, D.P., Chen, W., Siddharth, N., Xiong, C., Corso, J.J., Fellbaum, C.D., Hanson, C., Hanson, S.J., Hélie, S., Malaia, E., Pearlmutter, B.A., Siskind, J.M., Talavage, T.M., and Wilbur, R.B., ‘Seeing is Worse than Believing: Reading People’s Minds Better than Computer-Vision Methods Recognize Actions,’ ECCV 2014.

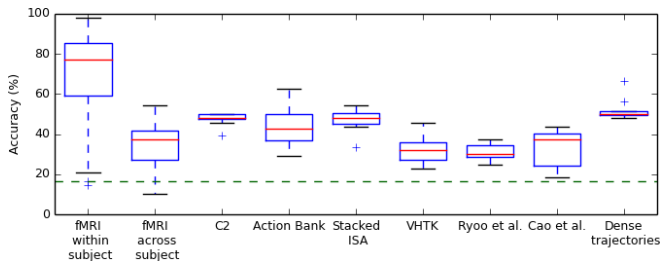


⇒ $\left\{ \begin{array}{l} \textit{carry}, \\ \textit{dig}, \\ \textit{hold}, \\ \textit{pick up}, \\ \textit{put down}, \\ \textit{walk} \end{array} \right.$

Barbu, A., Barrett, D.P., Chen, W., Siddharth, N., Xiong, C., Corso, J.J., Fellbaum, C.D., Hanson, C., Hanson, S.J., Hélie, S., Malaia, E., Pearlmutter, B.A., Siskind, J.M., Talavage, T.M., and Wilbur, R.B., ‘Seeing is Worse than Believing: Reading People’s Minds Better than Computer-Vision Methods Recognize Actions,’ ECCV 2014.

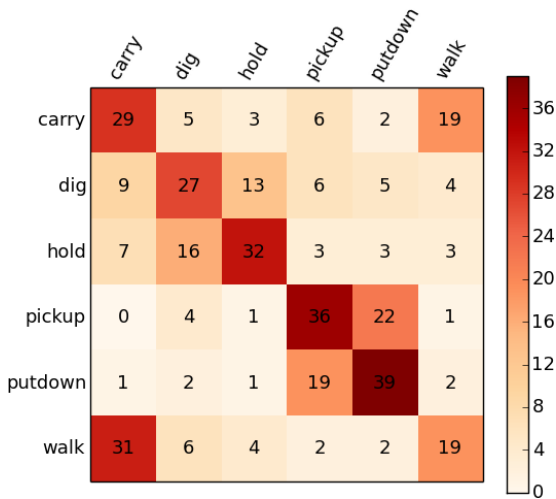
Classification Accuracies

fMRI within subject	69.7%
fMRI cross subject	34.8%
C2	47.4%
Action Bank	44.2%
Stacked ISA	46.8%
VHTK	32.5%
Ryoo et al.	31.2%
Cao et al.	33.3%
Dense Trajectories	52.3%
chance	16.6%



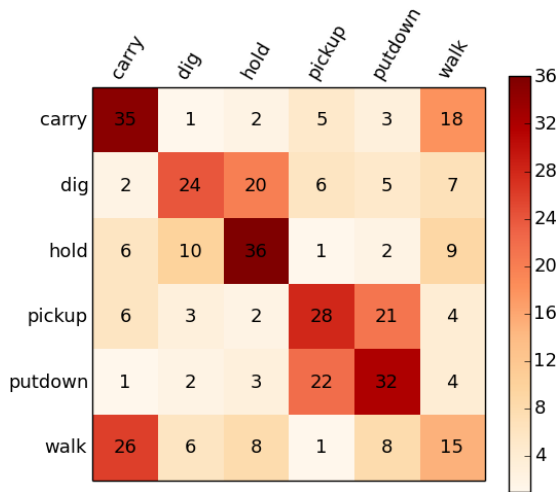
Confusion Matrices

C2



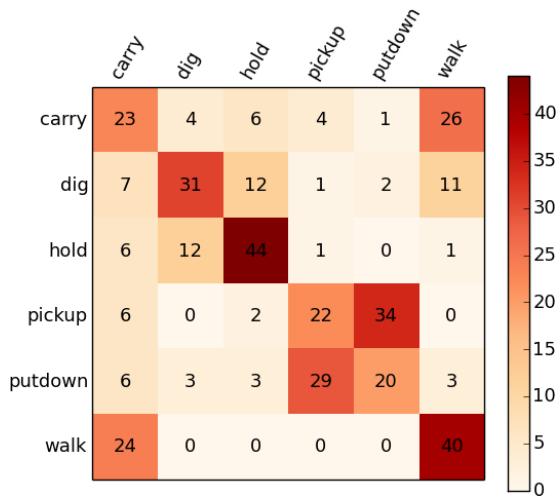
Confusion Matrices

Action Bank



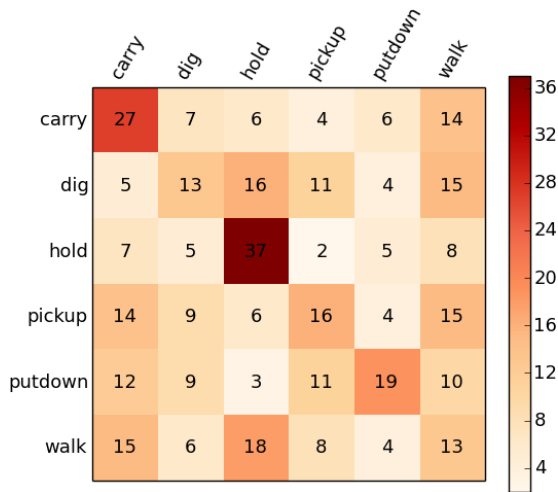
Confusion Matrices

Stacked ISA



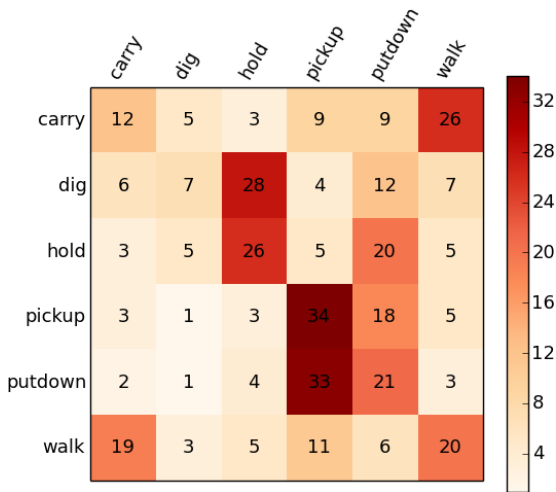
Confusion Matrices

VHTK



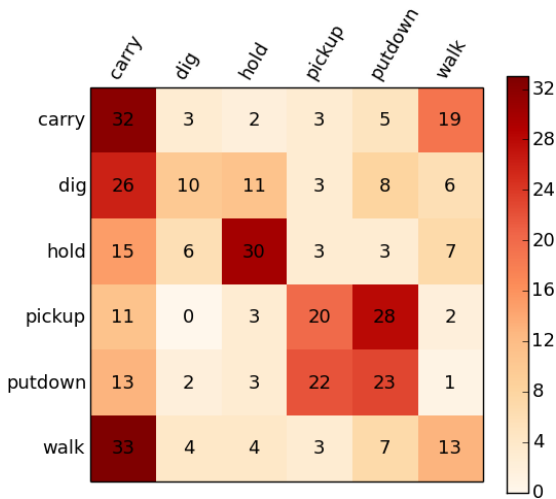
Confusion Matrices

Ryoo et al.



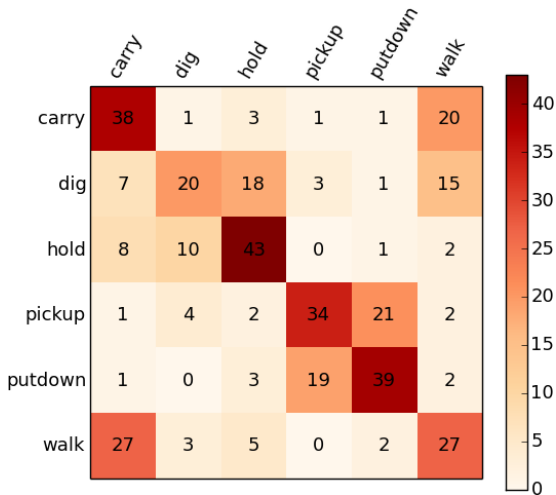
Confusion Matrices

Cao et al.



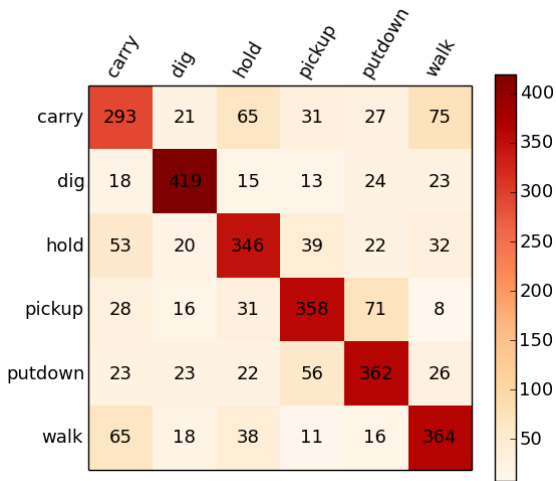
Confusion Matrices

Dense Trajectories



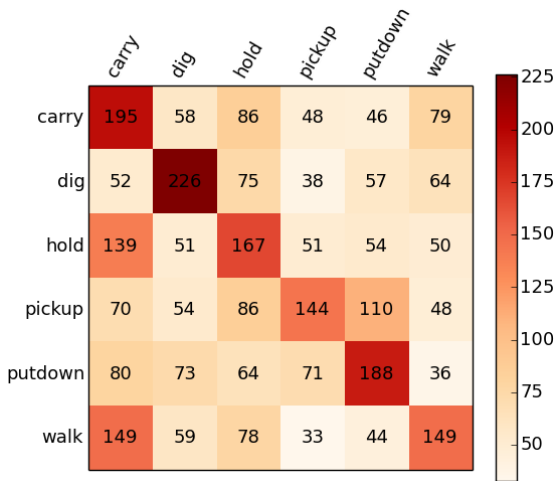
Confusion Matrices

fMRI within subject



Confusion Matrices

fMRI cross subject



Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*

Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*
 - ▶ recognize by horizontal motion

Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*
 - ▶ recognize by horizontal motion
- ▶ Computer-vision methods confuse *pick up* and *put down*

Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*
 - ▶ recognize by horizontal motion
- ▶ Computer-vision methods confuse *pick up* and *put down*
 - ▶ recognize by vertical motion

Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*
 - ▶ recognize by horizontal motion
- ▶ Computer-vision methods confuse *pick up* and *put down*
 - ▶ recognize by vertical motion
- ▶ Computer-vision methods confuse *dig* and *hold*

Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*
 - ▶ recognize by horizontal motion
- ▶ Computer-vision methods confuse *pick up* and *put down*
 - ▶ recognize by vertical motion
- ▶ Computer-vision methods confuse *dig* and *hold*
 - ▶ recognize by lack of motion

Confusion Matrices

- ▶ Computer-vision methods confuse *carry* and *walk*
 - ▶ recognize by horizontal motion
- ▶ Computer-vision methods confuse *pick up* and *put down*
 - ▶ recognize by vertical motion
- ▶ Computer-vision methods confuse *dig* and *hold*
 - ▶ recognize by lack of motion
- ▶ Humans (via fMRI) exhibit no such confusion

1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

5 Playing Checkers from English



HOLLYWOOD-2 Marszałek et al. (2009)



HOLLYWOOD-2 Marszałek et al. (2009)



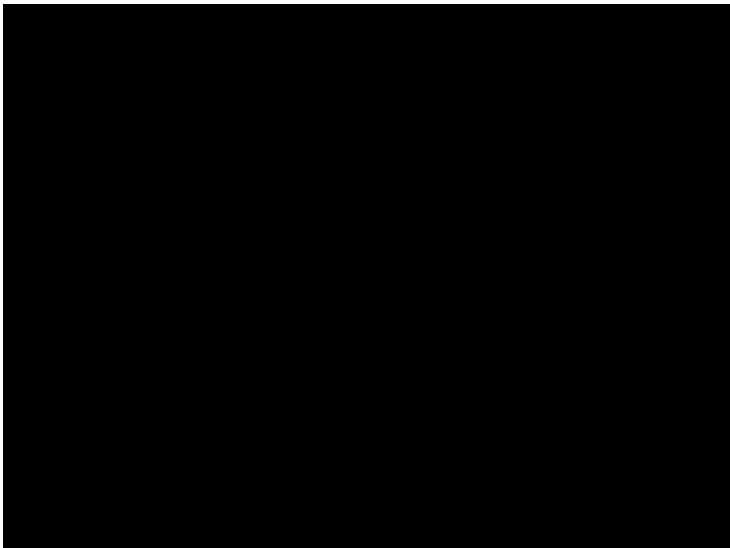
AnswerPhone,
DriveCar,
Eat,
FightPerson,
GetOutCar,
HandShake,
HugPerson,
Kiss,
Run,
SitDown,
SitUp,
StandUp

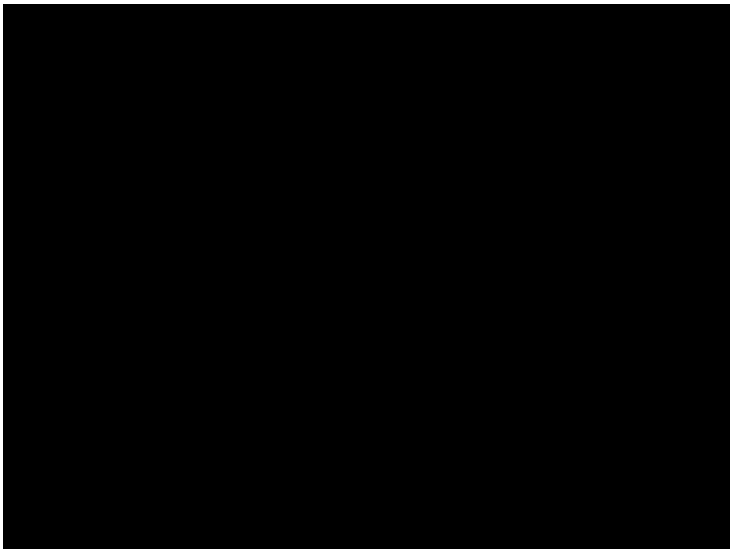
HOLLYWOOD-2 Marszałek et al. (2009)



AnswerPhone,
DriveCar,
Eat,
FightPerson,
GetOutCar,
HandShake,
HugPerson,
Kiss,
Run,
SitDown,
SitUp,
StandUp

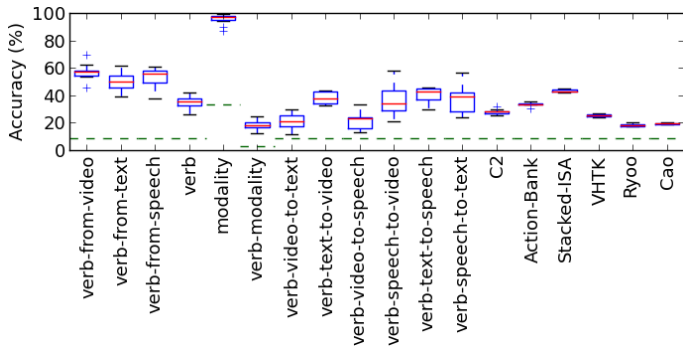
HOLLYWOOD-2 Marszałek et al. (2009)





Classification Accuracies

fMRI from-video	57.0%
fMRI from-text	50.0%
fMRI from-speech	52.7%
fMRI video-to-text	21.1%
fMRI text-to-video	38.0%
fMRI video-to-speech	21.6%
fMRI speech-to-video	36.0%
fMRI text-to-speech	40.0%
fMRI speech-to-text	37.1%
C2	28.0%
Action Bank	33.6%
Stacked ISA	43.3%
VHTK	25.3%
Ryoo et al.	18.3%
Cao et al.	19.2%
chance	8.4%



Confusion Matrices

fMRI verb-from-video

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	159	13	32	3	10	26	17	10	5	19	13	13
DriveCar	8	189	16	4	14	21	8	7	22	9	7	15
Eat	27	20	176	5	5	18	19	20	3	7	10	10
FightPerson	1	6	2	237	8	5	14	11	9	8	8	11
GetOutCar	6	15	4	10	196	5	4	3	23	14	12	28
HandShake	22	5	20	7	5	193	17	14	3	12	5	17
HugPerson	18	16	15	11	10	16	155	44	4	8	11	12
Kiss	13	9	12	11	3	21	38	178	2	13	12	8
Run	3	28	1	7	15	2	7	2	225	12	3	15
SitDown	13	9	5	7	17	13	6	7	6	155	29	53
SitUp	12	14	7	13	9	5	9	8	4	21	191	27
StandUp	6	13	1	12	25	24	8	9	14	52	21	135

Confusion Matrices

fMRI verb-from-text

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	151	16	13	25	10	25	14	10	13	17	9	17
DriveCar	26	149	10	22	9	15	11	10	15	16	14	23
Eat	16	12	171	7	8	9	10	17	30	12	13	15
FightPerson	15	18	10	161	18	23	12	11	8	14	8	22
GetOutCar	19	20	15	15	157	15	18	7	12	14	7	21
HandShake	18	18	8	22	12	157	14	10	13	11	14	23
HugPerson	20	16	14	19	13	17	168	12	10	13	8	10
Kiss	2	14	27	8	5	12	7	177	20	19	13	16
Run	13	10	24	9	13	8	8	19	181	11	12	12
SitDown	11	12	13	15	13	18	11	20	15	151	21	20
SitUp	16	15	10	9	15	15	10	13	13	30	159	15
StandUp	16	25	8	22	15	20	8	18	16	17	15	140

Confusion Matrices

fMRI verb-from-speech

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	161	15	11	14	15	14	24	12	4	17	14	19
DriveCar	18	174	5	11	14	13	18	8	9	20	15	15
Eat	6	11	177	10	5	17	8	27	21	9	13	16
FightPerson	20	18	12	157	19	11	20	11	10	12	14	16
GetOutCar	22	18	4	15	189	15	11	3	13	10	9	11
HandShake	14	13	11	16	17	157	28	17	12	15	13	7
HugPerson	12	14	6	20	16	24	165	14	16	6	10	17
Kiss	6	5	24	1	2	19	14	190	17	16	10	16
Run	9	8	15	11	6	10	16	24	176	13	17	15
SitDown	14	31	15	12	14	17	11	10	16	165	7	8
SitUp	12	15	22	16	11	11	11	12	17	27	154	12
StandUp	16	18	14	18	15	12	11	20	14	11	10	161

Confusion Matrices

C2

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	49	7	99	2	13	29	8	30	20	29	15	19
DriveCar	12	75	30	28	41	17	13	17	61	7	16	3
Eat	25	15	169	1	18	13	19	12	11	22	10	5
FightPerson	3	21	7	186	14	2	10	8	17	36	9	7
GetOutCar	8	44	34	23	67	25	8	9	49	14	21	18
HandShake	16	27	30	19	7	82	17	41	19	28	12	22
HugPerson	20	29	24	43	11	27	28	27	27	29	24	31
Kiss	35	30	11	16	21	38	23	60	10	50	16	10
Run	11	4	15	86	29	17	6	13	87	34	5	13
SitDown	20	3	21	16	10	25	4	21	15	157	10	18
SitUp	24	21	46	48	24	12	19	16	13	26	32	39
StandUp	22	3	32	17	30	21	25	18	10	26	32	84

Confusion Matrices

Action Bank

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	97	19	52	10	12	25	31	11	0	39	24	0
DriveCar	1	142	11	21	31	42	10	12	39	10	0	1
Eat	42	27	122	10	14	32	14	31	9	13	5	1
FightPerson	0	28	0	193	16	10	15	4	28	1	11	14
GetOutCar	25	29	27	43	104	9	1	0	15	0	39	28
HandShake	30	27	13	6	12	145	27	22	20	11	0	7
HugPerson	35	11	7	57	0	38	55	53	0	26	23	15
Kiss	44	0	23	34	21	25	26	89	4	31	3	20
Run	21	36	6	89	30	11	7	0	91	18	0	11
SitDown	27	16	58	14	0	10	23	34	19	91	1	27
SitUp	30	10	30	16	21	10	11	40	2	0	73	77
StandUp	13	2	11	20	41	16	11	1	21	21	74	89

Confusion Matrices

Stacked ISA

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	65	0	72	3	3	60	4	29	0	49	21	14
DriveCar	0	258	10	15	12	0	0	0	22	0	0	3
Eat	17	24	174	0	27	16	17	7	4	20	14	0
FightPerson	8	6	8	187	11	12	10	0	36	19	13	10
GetOutCar	2	39	22	29	147	5	0	5	22	10	4	35
HandShake	55	11	12	0	17	133	4	42	11	12	14	9
HugPerson	31	1	10	18	7	39	42	73	9	21	52	17
Kiss	20	2	10	6	7	53	37	122	0	27	34	2
Run	8	45	3	71	14	0	3	0	132	21	10	13
SitDown	29	0	7	1	3	26	6	36	12	143	15	42
SitUp	40	2	16	13	11	25	29	21	7	11	82	63
StandUp	11	0	0	2	18	6	40	4	9	25	26	179

Confusion Matrices

VHTK

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	36	1	179	14	2	7	16	43	2	2	9	9
DriveCar	0	132	18	73	5	33	0	32	24	0	3	0
Eat	21	1	207	10	4	28	8	22	0	10	5	4
FightPerson	0	17	15	143	5	5	16	31	30	20	38	0
GetOutCar	17	7	53	63	5	34	0	42	31	12	28	28
HandShake	26	10	79	1	11	60	12	66	8	15	20	12
HugPerson	27	25	69	61	7	10	17	60	13	9	22	0
Kiss	32	24	42	32	5	36	13	103	8	0	25	0
Run	7	31	43	94	2	8	12	7	64	10	30	12
SitDown	9	2	81	44	5	19	0	32	10	100	12	6
SitUp	5	8	48	60	2	43	4	75	19	1	29	26
StandUp	9	10	59	50	5	15	4	11	47	10	24	76

Confusion Matrices

Ryoo

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	5	2	87	41	16	29	1	113	0	7	11	8
DriveCar	2	11	73	54	17	17	0	44	22	2	0	78
Eat	0	14	189	25	11	1	0	65	3	7	0	5
FightPerson	10	19	3	186	10	18	2	11	10	0	3	48
GetOutCar	0	9	55	58	42	19	0	38	10	10	31	48
HandShake	10	30	10	16	0	26	4	191	0	0	0	33
HugPerson	3	0	40	128	3	20	0	65	7	0	13	41
Kiss	2	13	46	44	19	30	2	112	1	4	4	43
Run	12	11	50	129	4	9	0	66	11	12	5	11
SitDown	0	0	43	87	24	14	0	99	16	9	3	25
SitUp	0	16	56	39	30	8	3	63	16	0	29	60
StandUp	10	11	20	77	15	1	12	47	1	0	43	83

Confusion Matrices

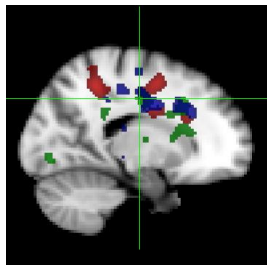
Cao

	AnswerPhone	DriveCar	Eat	FightPerson	GetOutCar	HandShake	HugPerson	Kiss	Run	SitDown	SitUp	StandUp
AnswerPhone	21	2	62	60	35	31	18	64	7	8	12	0
DriveCar	26	2	33	66	51	13	12	39	29	9	39	1
Eat	16	21	172	20	3	8	4	41	0	20	8	7
FightPerson	6	1	0	221	37	9	3	10	19	0	4	10
GetOutCar	10	7	38	63	100	18	17	25	0	3	22	17
HandShake	27	10	11	27	29	41	27	89	2	24	23	10
HugPerson	17	0	21	168	11	21	17	26	9	21	0	9
Kiss	33	0	20	64	32	44	11	55	5	35	5	16
Run	32	8	25	120	17	9	1	43	1	42	10	12
SitDown	15	0	30	74	10	24	37	44	0	76	10	0
SitUp	0	11	42	87	78	23	6	31	1	8	9	24
StandUp	11	15	18	126	48	22	1	20	8	9	16	26

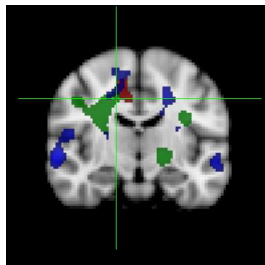
Hot Off The Press

	SVM	NN
fMRI from-video	57.0%	60.6%
fMRI from-text	50.0%	53.1%
fMRI from-speech	52.7%	56.0%
fMRI from-all	34.8%	51.8%
fMRI video-to-text	21.1%	26.6%
fMRI text-to-video	38.0%	44.6%
fMRI video-to-speech	21.6%	28.0%
fMRI speech-to-video	36.0%	49.4%
fMRI text-to-speech	40.0%	44.0%
fMRI speech-to-text	37.1%	45.6%

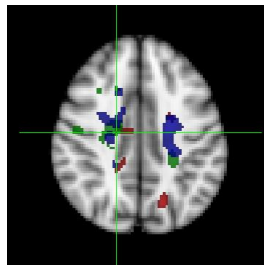
Brain Regions for Video, Text, and Speech



slice 36 (X = 18.00)



slice 56 (Y = -14.00)



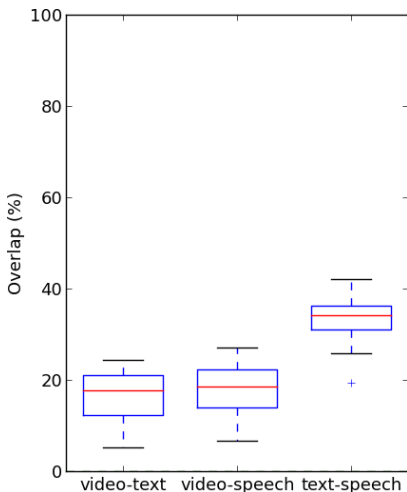
slice 56 (Z = 40.00)

- video
- text
- speech

Subject 05, MNI_152

Quantifying Overlap Between Video, Text, and Speech

$$\frac{|X \cap Y|}{|X \cup Y|}$$



In 4000 voxels ($\approx 10\%$ of brain volume) with highest magnitude of SVM weight.

1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

5 Playing Checkers from English

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{carried} \\ \textit{folded} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\} \times \left\{ \begin{array}{l} [\textit{on the}]\textit{left}[\textit{ward}] \\ [\textit{on the}]\textit{right}[\textit{ward}] \end{array} \right\}$$

⎧
Dan
Haonan
Jeff
Scott
⎨

actor

carried
folded
left

verb

the chair
the paper
the shirt

object

$$\left\{ \begin{array}{l} \textit{left}[\textit{ward}] \\ \textit{right}[\textit{ward}] \end{array} \right\}$$

direction

{ [*on the*]left
[*on the*]right }

location

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{carried} \\ \textit{folded} \\ \textit{left} \end{array} \right\}$$

actor-verb

$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\}$ $\times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\}$

actor-object

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{left}[\textit{ward}] \\ \textit{right}[\textit{ward}] \end{array} \right\}$$

actor-direction

$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\}$ $\times \left\{ \begin{array}{l} \textit{[on the]left} \\ \textit{[on the]right} \end{array} \right\}$

actor-location

$$\left\{ \begin{array}{l} \textit{carried} \\ \textit{folded} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\}$$

verb-object

$$\left\{ \begin{array}{l} \textit{carried} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{left}[\textit{ward}] \\ \textit{right}[\textit{ward}] \end{array} \right\}$$

verb-direction

$$\left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{left[ward]} \\ \textit{right[ward]} \end{array} \right\}$$

object-direction

$$\left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{[on the]left} \\ \textit{[on the]right} \end{array} \right\}$$

object-location

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{carried} \\ \textit{folded} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\}$$

actor-verb-object

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{carried} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{left}[\textit{ward}] \\ \textit{right}[\textit{ward}] \end{array} \right\}$$

actor-verb-direction

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{left}[\textit{ward}] \\ \textit{right}[\textit{ward}] \end{array} \right\}$$

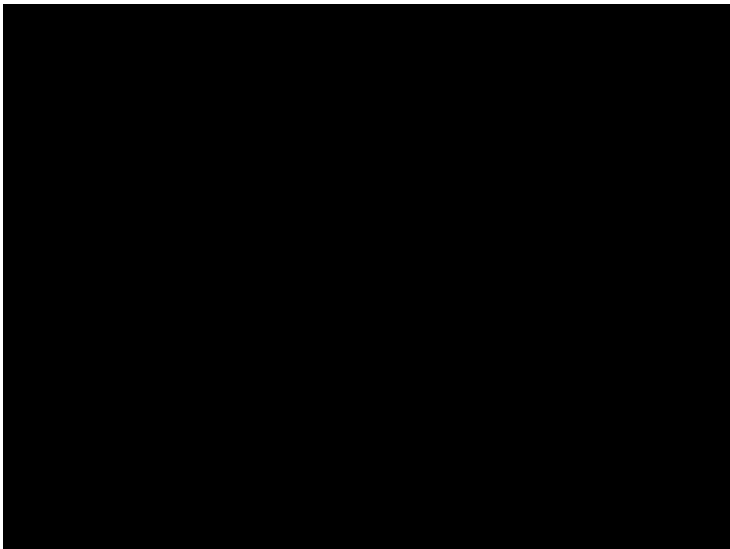
actor-object-direction

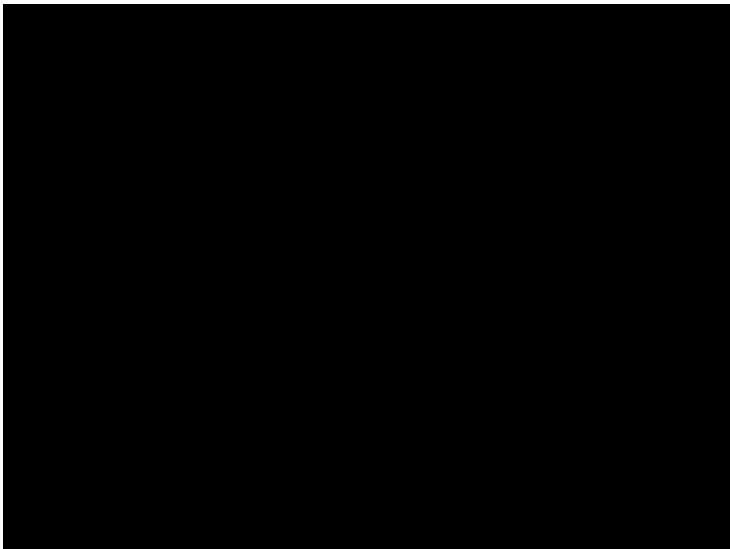
$$\left\{ \begin{array}{l} \textit{carried} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{left}[\textit{ward}] \\ \textit{right}[\textit{ward}] \end{array} \right\}$$

verb-object-direction

$$\left\{ \begin{array}{l} \textit{Dan} \\ \textit{Haonan} \\ \textit{Jeff} \\ \textit{Scott} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{carried} \\ \textit{folded} \\ \textit{left} \end{array} \right\} \times \left\{ \begin{array}{l} \textit{the chair} \\ \textit{the paper} \\ \textit{the shirt} \end{array} \right\} \times \left\{ \begin{array}{l} [\textit{on the}]\textit{left}[\textit{ward}] \\ [\textit{on the}]\textit{right}[\textit{ward}] \end{array} \right\}$$

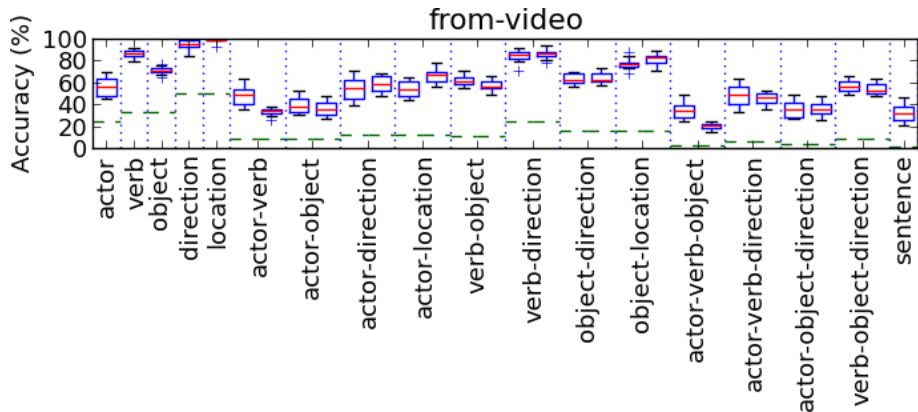
sentence





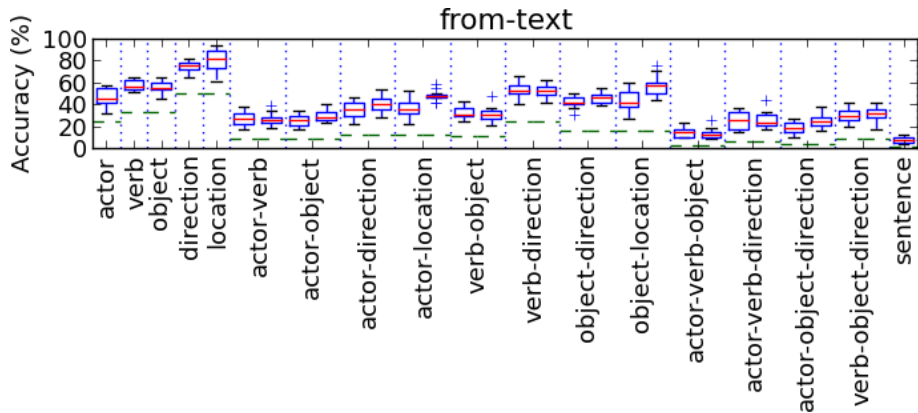
- independent** Train on individual constituents, test on constituent pairs and triples.
- joint** Train and test on constituent pairs and triples.

Classification Accuracies



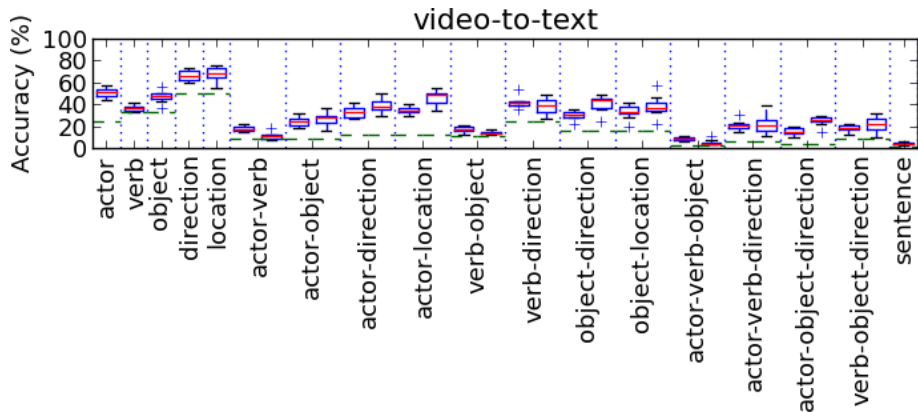
independent on left, joint on right of each pair and triple

Classification Accuracies



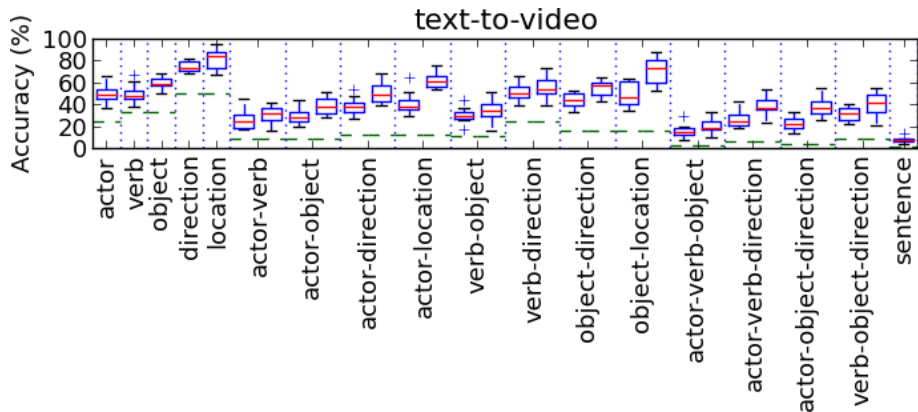
independent on left, joint on right of each pair and triple

Classification Accuracies



independent on left, joint on right of each pair and triple

Classification Accuracies



independent on left, joint on right of each pair and triple

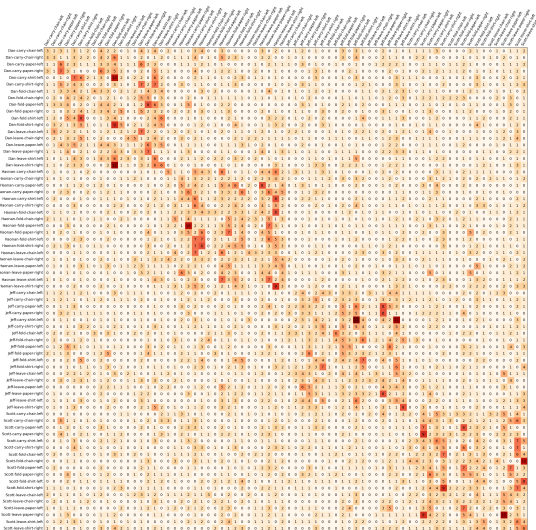
Classification Accuracies

decode whole sentence

fMRI from-video	32.2%
fMRI from-text	7.6%
fMRI video-to-text	4.4%
fMRI text-to-video	7.6%
chance	1.4%

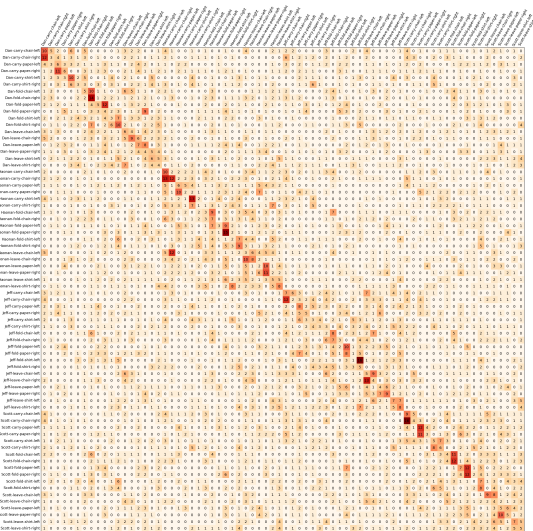
Confusion Matrices

sentence video-to-text



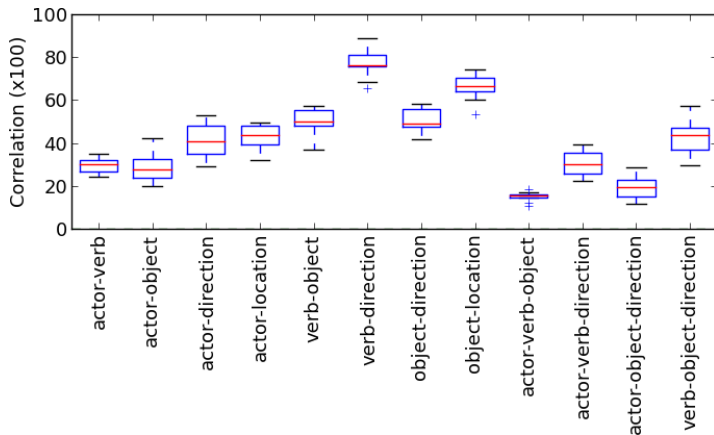
Confusion Matrices

sentence text-to-video



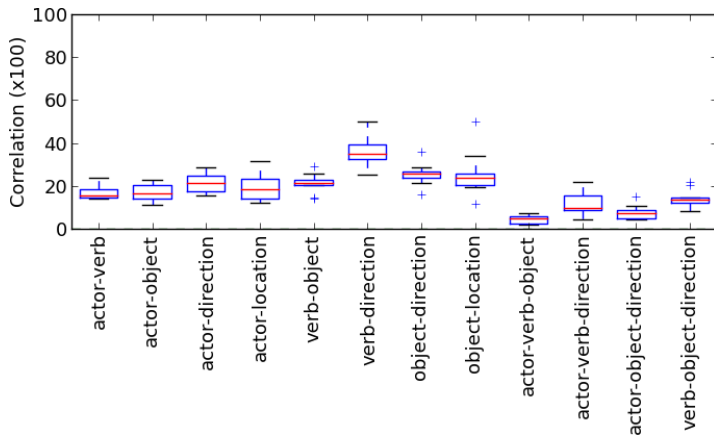
Correlation of Independent and Joint Classifier Judgments

from-video



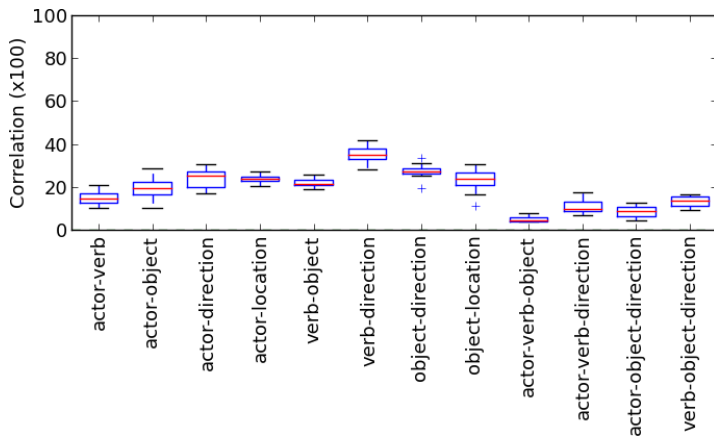
Correlation of Independent and Joint Classifier Judgments

from-text



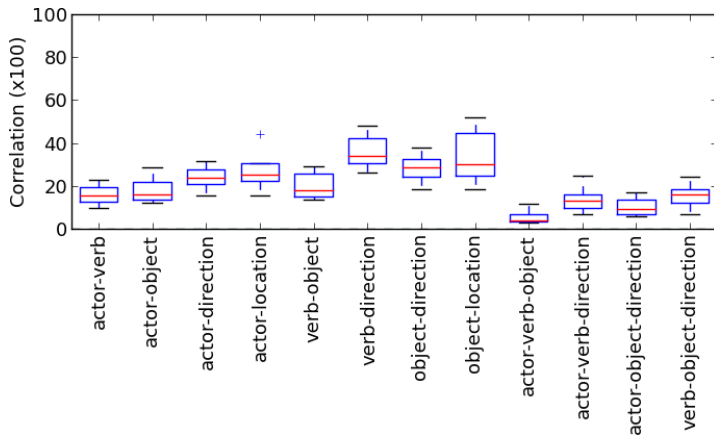
Correlation of Independent and Joint Classifier Judgments

video-to-text

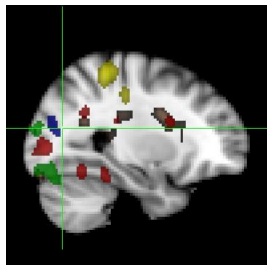


Correlation of Independent and Joint Classifier Judgments

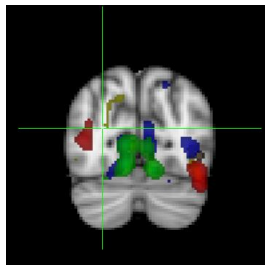
text-to-video



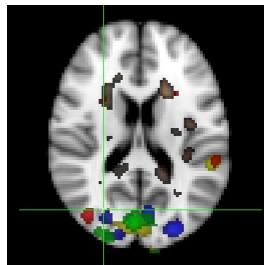
Brain Regions for Constituents



slice 31 (X = 28.00)



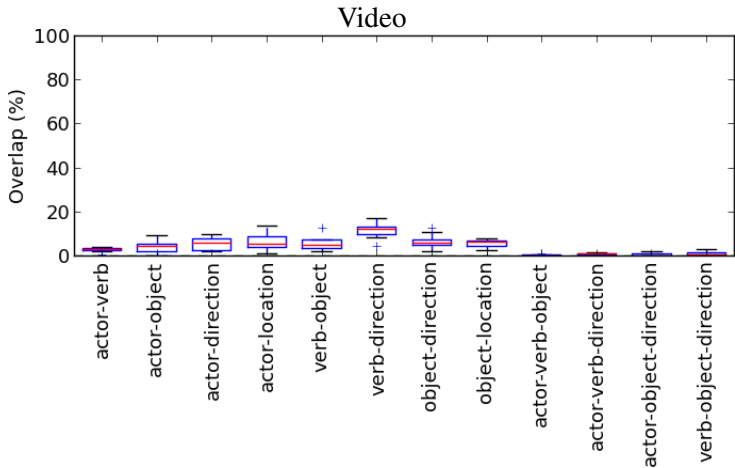
slice 27 (Y = -72.00)



slice 45 (Z = 18.00)

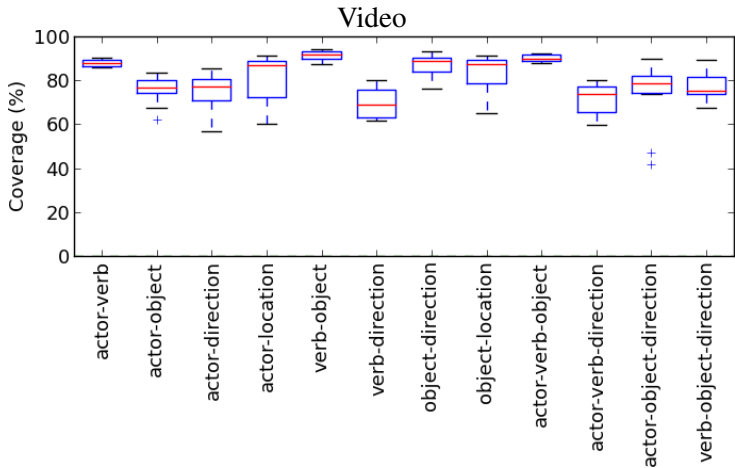
- actor
- verb
- object
- direction
- location

Subject 01, MNI_152, from-video



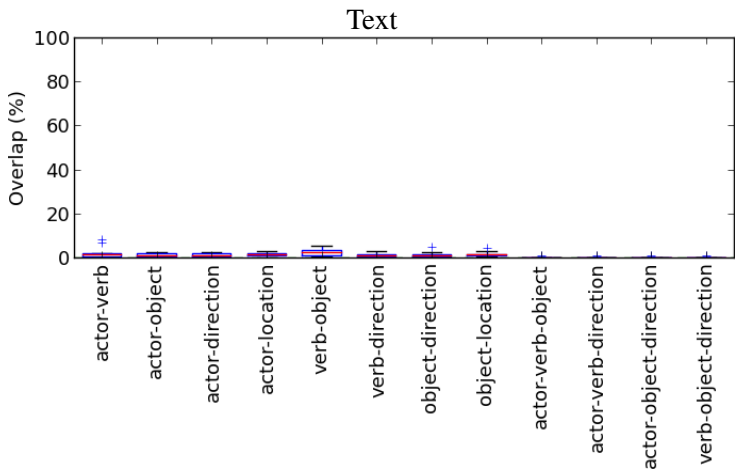
$$\frac{\bigcap_i \text{independent}_i}{\bigcup_i \text{independent}_i}$$

Fraction of overlap between all of the constituent classifier regions in the independent classifier



$$\frac{\left| \left(\bigcup_i \text{independent}_i \right) \cap \text{joint} \right|}{|\text{joint}|}$$

Fraction of the joint classifier region covered by the union of the independent classifier regions



$$\frac{\bigcap_i \text{independent}_i}{\bigcup_i \text{independent}_i}$$

Fraction of overlap between all of the constituent classifier regions in the independent classifier

Summary

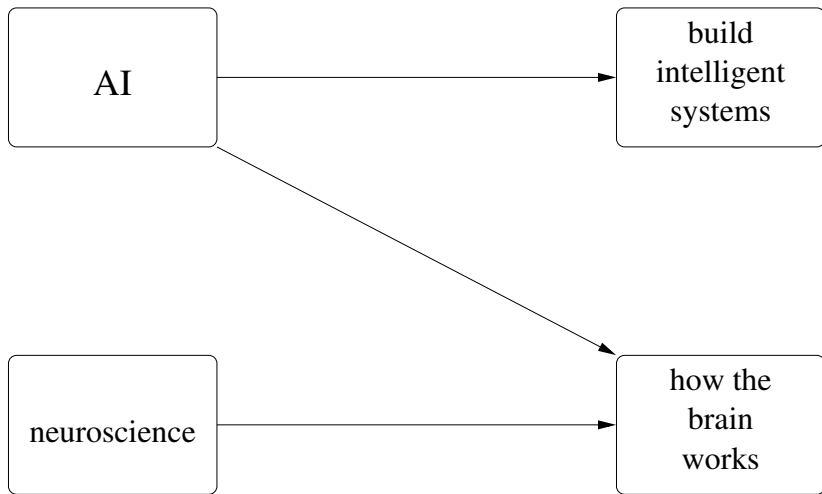
AI

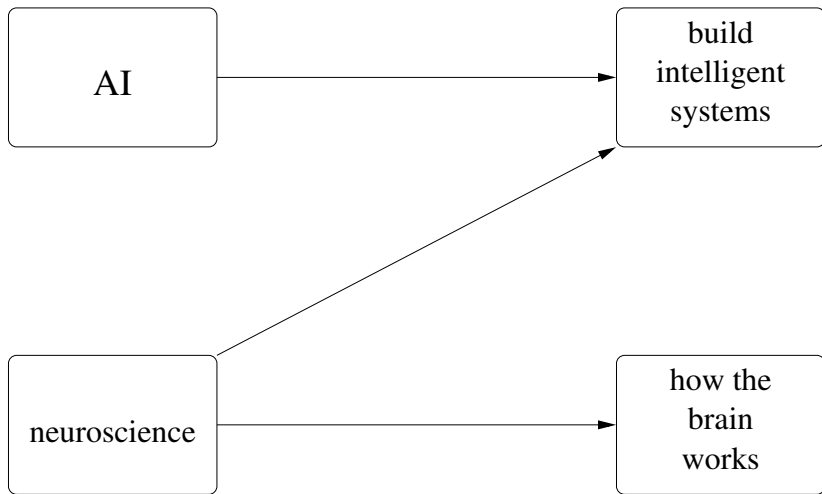


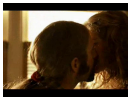


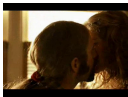
neuroscience

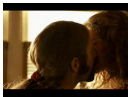




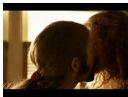




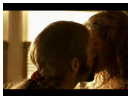


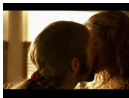


Kiss

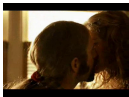


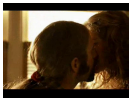
Kiss



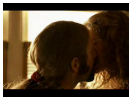


Kiss

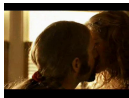




Kiss

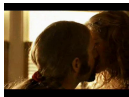


Kiss

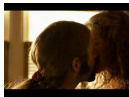


Kiss

=



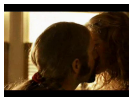
Kiss



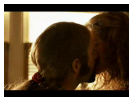
Kiss

=

=



Kiss

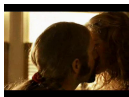


Kiss

=

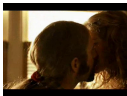
=

=



Kiss

```
(define (AI)
  (let loop (mental-state '())
    (let* ((percept (perceive eyes ears skin tongue nose))
           (action (cognition percepts)))
      (act! arms legs mouth action)
      (loop mental-state))))
```

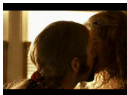


Kiss

=

=

=

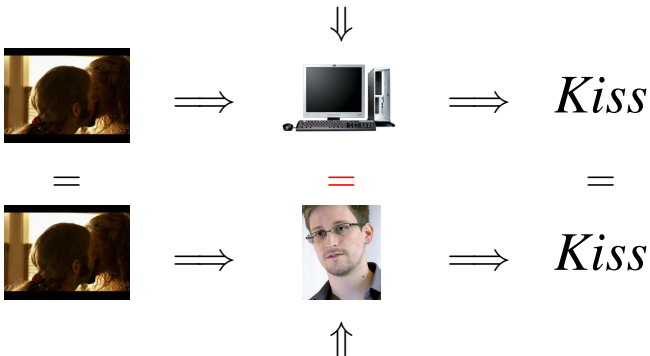


Kiss

```

(define (AI)
  (let loop (mental-state '())
    (let* ((percept (perceive eyes ears skin tongue nose))
           (action (cognition percepts)))
      (act! arms legs mouth action)
      (loop mental-state))))

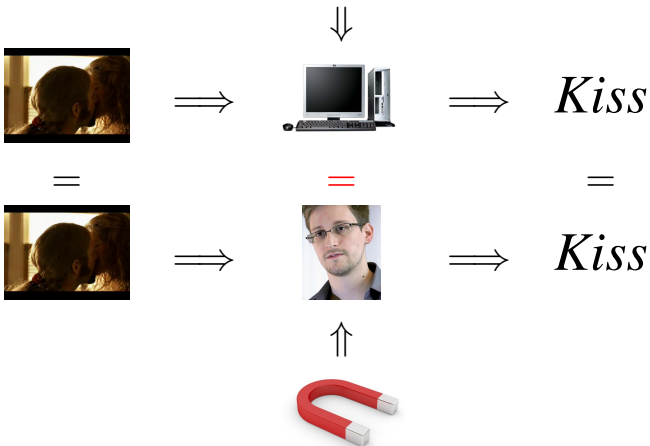
```



```

(define (AI)
  (let loop (mental-state '())
    (let* ((percept (perceive eyes ears skin tongue nose))
           (action (cognition percepts)))
      (act! arms legs mouth action)
      (loop mental-state))))

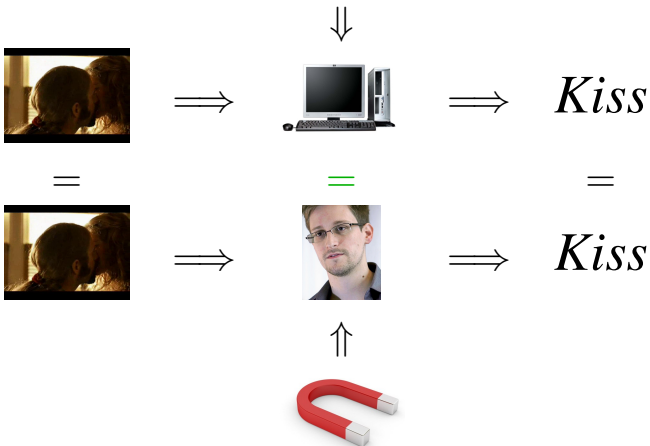
```



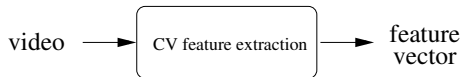
```

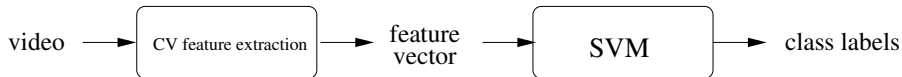
(define (AI)
  (let loop (mental-state '())
    (let* ((percept (perceive eyes ears skin tongue nose))
           (action (cognition percepts)))
      (act! arms legs mouth action)
      (loop mental-state))))

```

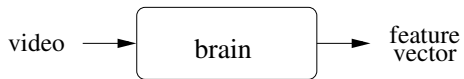


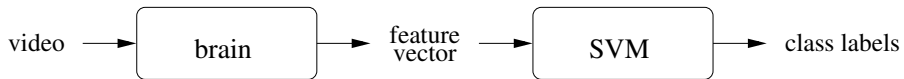
video

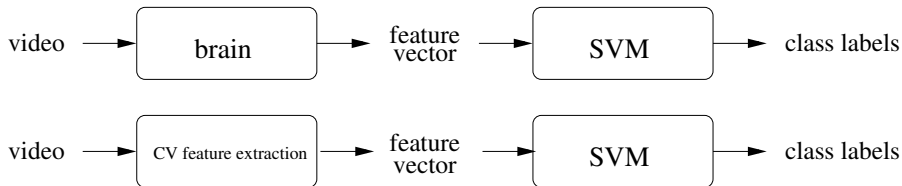


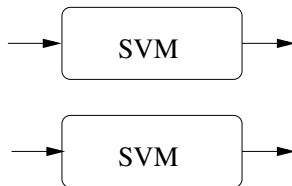


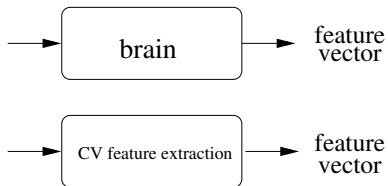
video

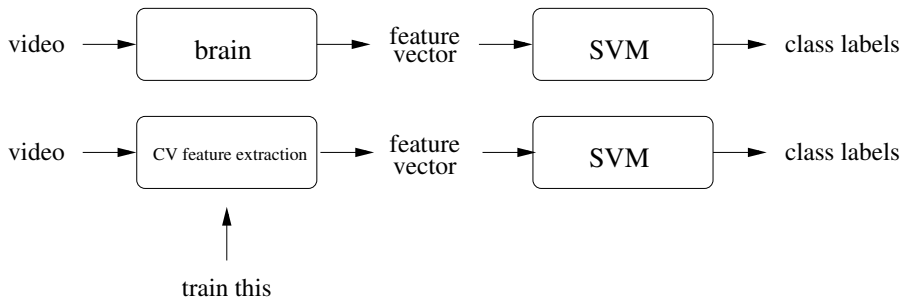




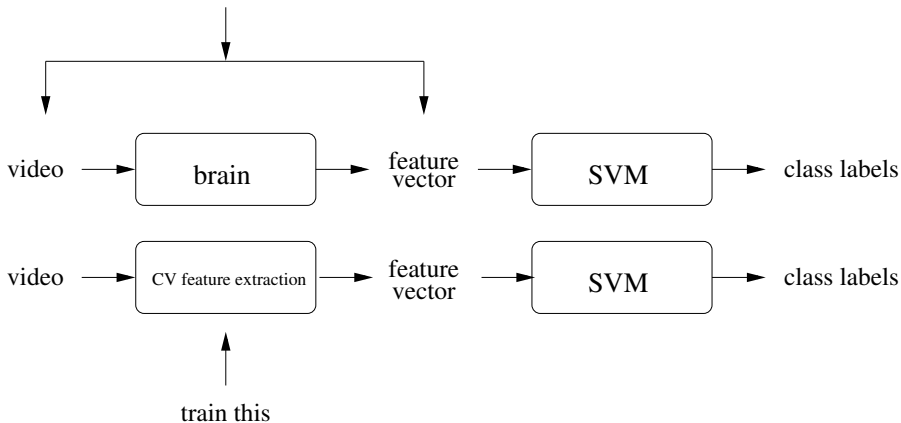


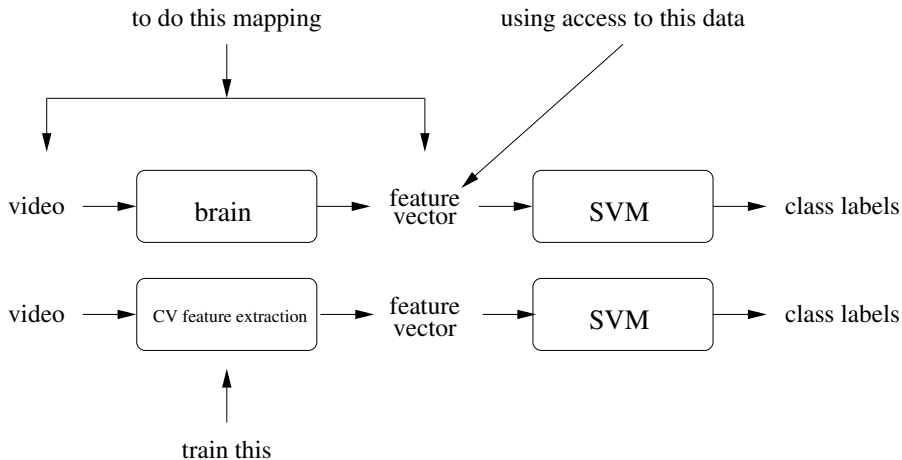






to do this mapping





1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

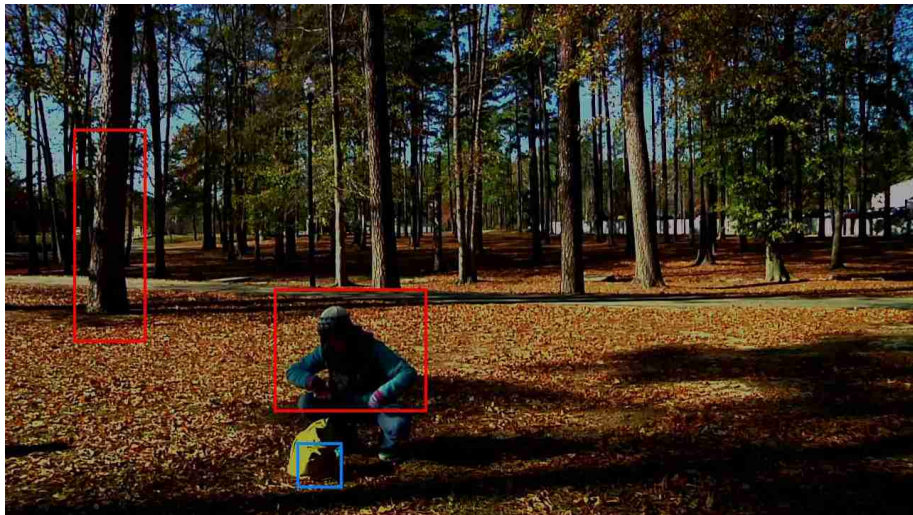
5 Playing Checkers from English

Andrei Barbu Daniel Paul Barrett N. Siddharth Haonan Yu

Object detection: Felzenszwalb et al. (2008)



Object detection: Felzenszwalb et al. (2008)



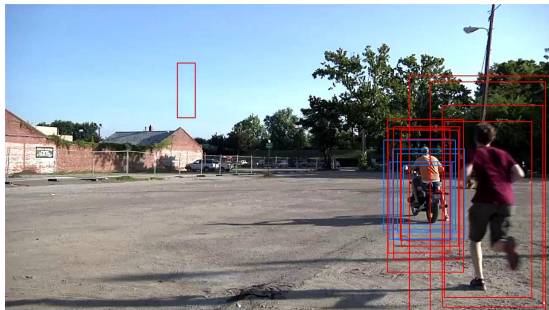
False **positives**

Object detection: Felzenszwalb et al. (2008)



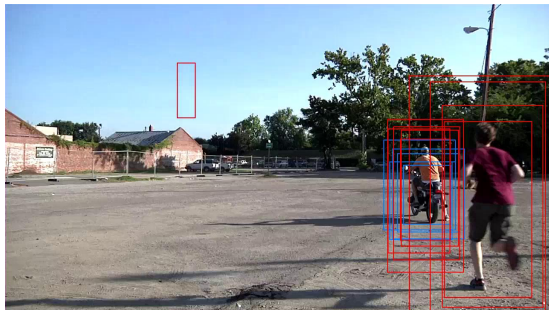
False **negatives**

From uncertain object detections to robust object tracks



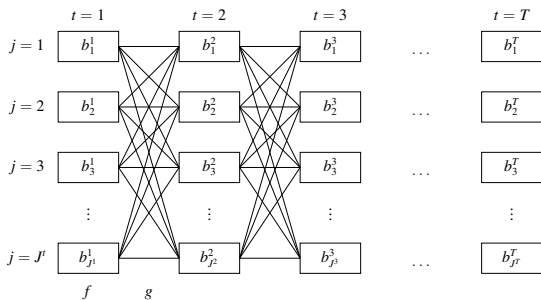
- ▶ for each object model in each frame
- ▶ overgenerate
 - ▶ many false positives
 - ▶ compensate for false negatives

From uncertain object detections to robust object tracks



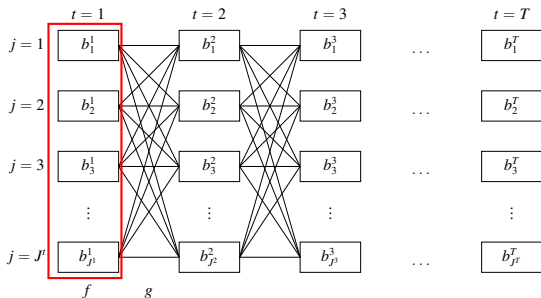
- ▶ for each object model in each frame
- ▶ overgenerate
 - ▶ many false positives
 - ▶ compensate for false negatives

From uncertain object detections to robust object tracks



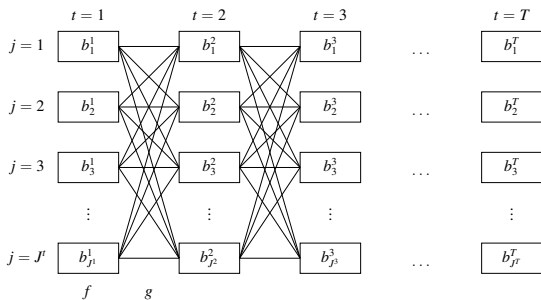
- find single detection for each object in each frame

From uncertain object detections to robust object tracks



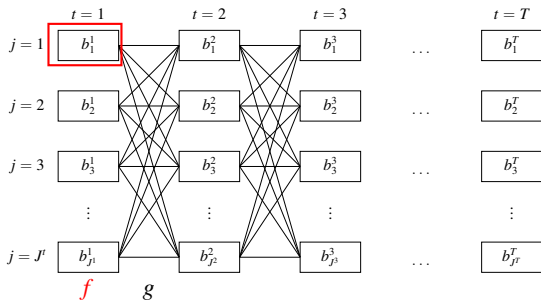
- find single detection for each object in each frame

From uncertain object detections to robust object tracks



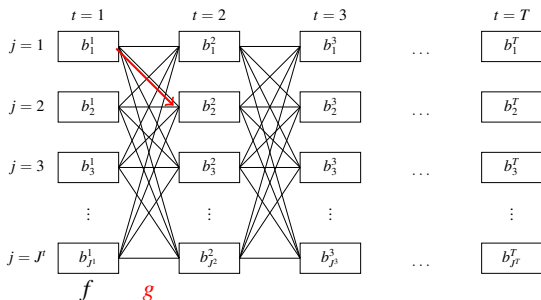
- ▶ find single detection for each object in each frame
- ▶ temporally coherent track

From uncertain object detections to robust object tracks



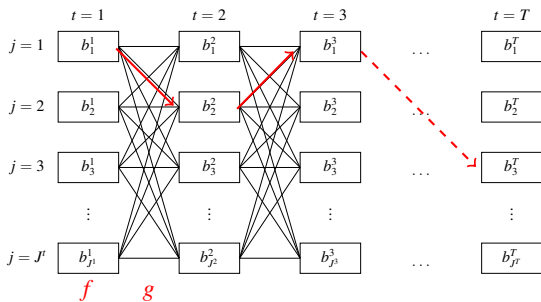
- ▶ find single detection for each object in each frame
- ▶ temporally coherent track
 - ▶ object detector confidence (f)

From uncertain object detections to robust object tracks



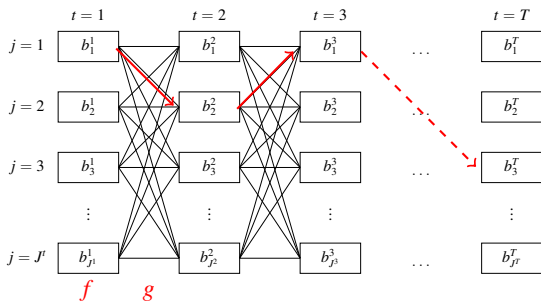
- ▶ find single detection for each object in each frame
- ▶ temporally coherent track
 - ▶ object detector confidence (f)
 - ▶ motion coherence (g)

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

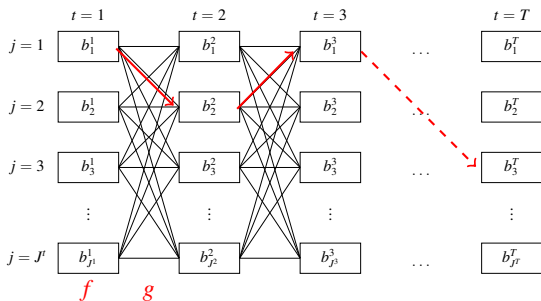
From uncertain object detections to robust object tracks



- ▶ 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

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

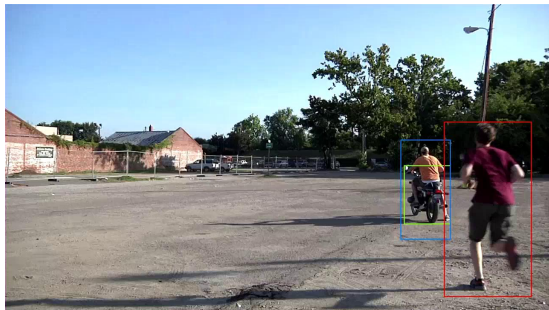
From uncertain object detections to robust object tracks



- ▶ 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)

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

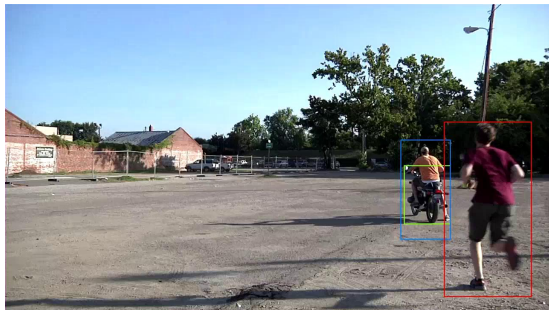
From uncertain object detections to robust object tracks



- ▶ 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)

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

From uncertain object detections to robust object tracks



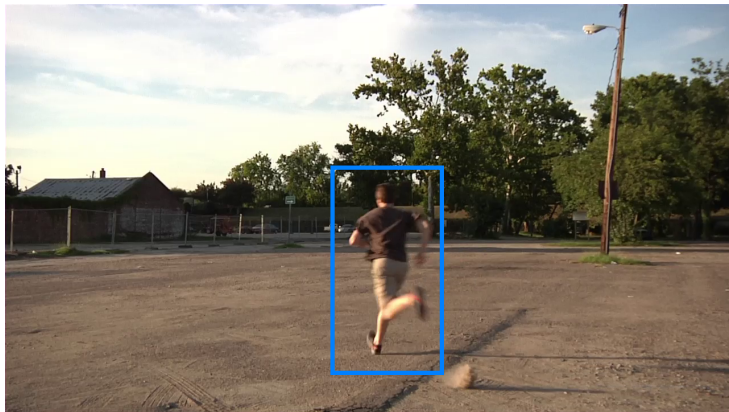
- ▶ find single detection for each object in each frame
- ▶ temporally coherent track
 - ▶ object detector confidence (f)
 - ▶ motion coherence (g)
- ▶ optimal path through lattice of detections
- ▶ dynamic programming
Bellman (1957), Viterbi (1967)

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

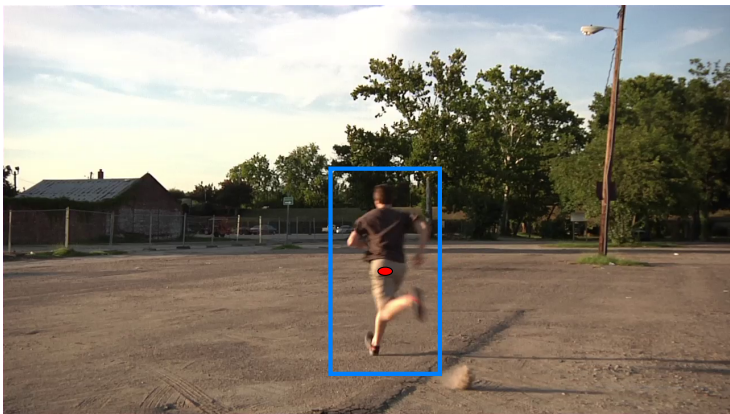
Feature vector—single participant



Feature vector—single participant

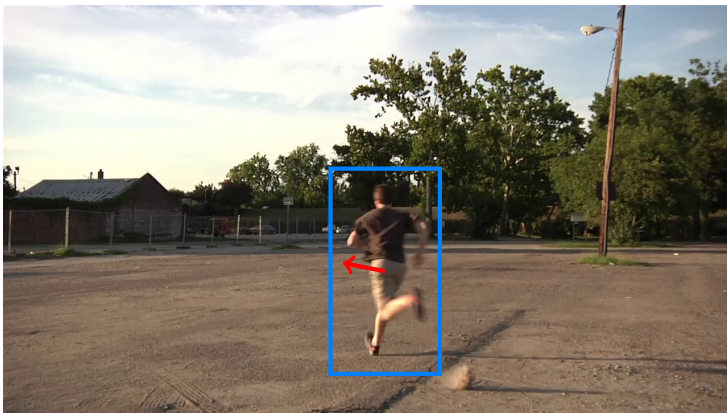


Feature vector—single participant



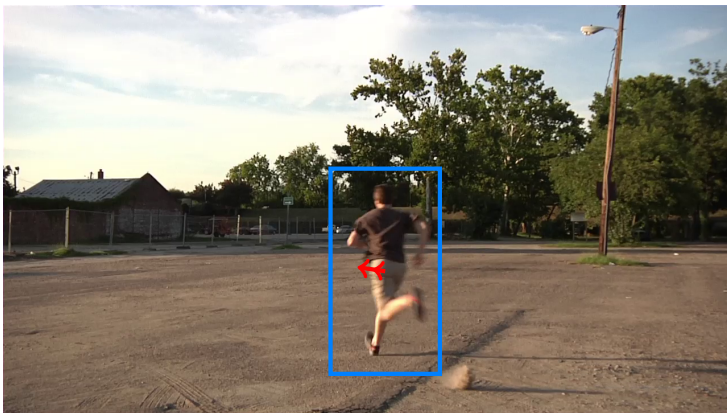
position	$\frac{d}{dt}$ position	$\frac{d^2}{dt^2}$ position	
aspect ratio	$\frac{d}{dt}$ aspect ratio	area	$\frac{d}{dt}$ area
object class	root-filter index		

Feature vector—single participant



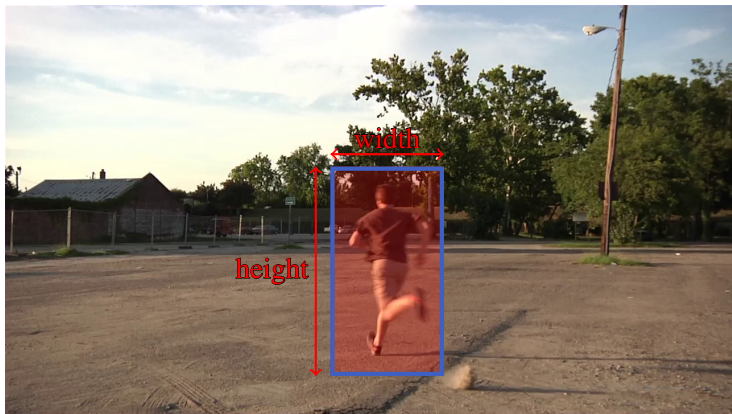
position	$\frac{d}{dt}$ position	$\frac{d^2}{dt^2}$ position	
aspect ratio	$\frac{d}{dt}$ aspect ratio	area	$\frac{d}{dt}$ area
object class	root-filter index		

Feature vector—single participant



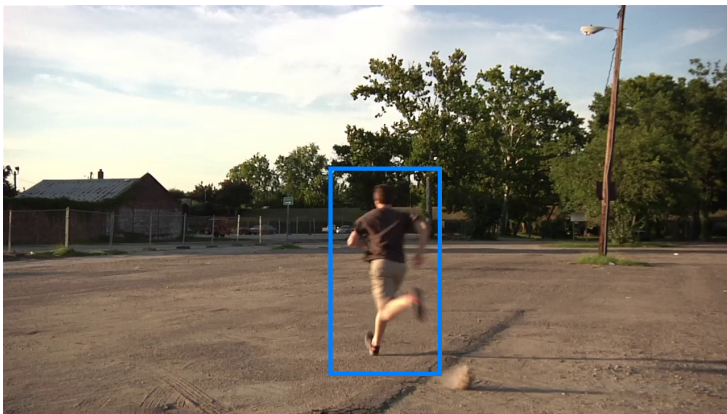
position	$\frac{d}{dt}$ position	$\frac{d^2}{dt^2}$ position	
aspect ratio	$\frac{d}{dt}$ aspect ratio	area	$\frac{d}{dt}$ area
object class	root-filter index		

Feature vector—single participant



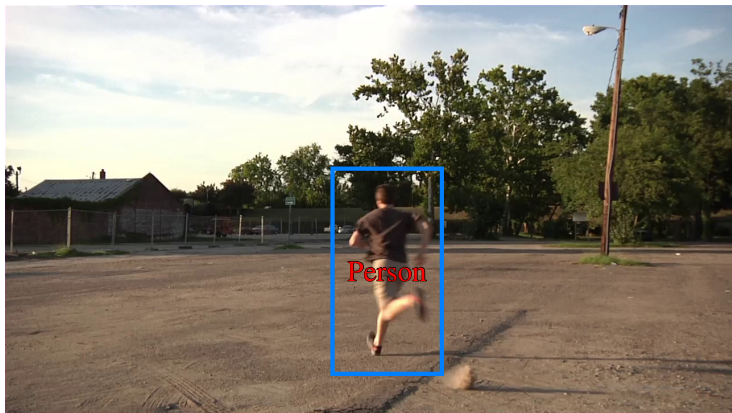
position	$\frac{d}{dt}$ position	$\frac{d^2}{dt^2}$ position	
aspect ratio	$\frac{d}{dt}$ aspect ratio	area	$\frac{d}{dt}$ area
object class	root-filter index		

Feature vector—single participant



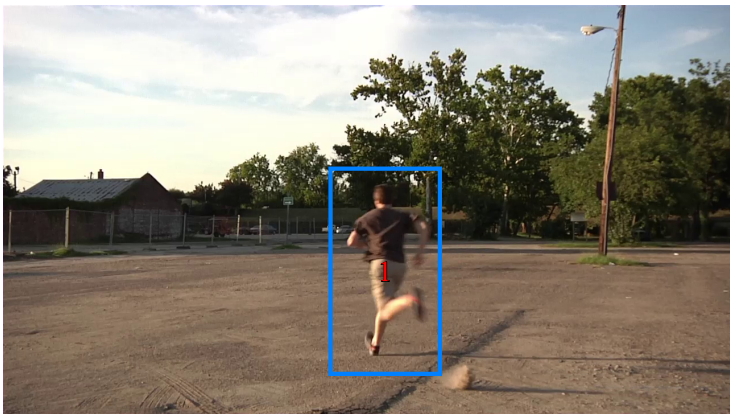
position	$\frac{d}{dt}$ position	$\frac{d^2}{dt^2}$ position	
aspect ratio	$\frac{d}{dt}$ aspect ratio	area	$\frac{d}{dt}$ area
object class	root-filter index		

Feature vector—single participant



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

Feature vector—single participant

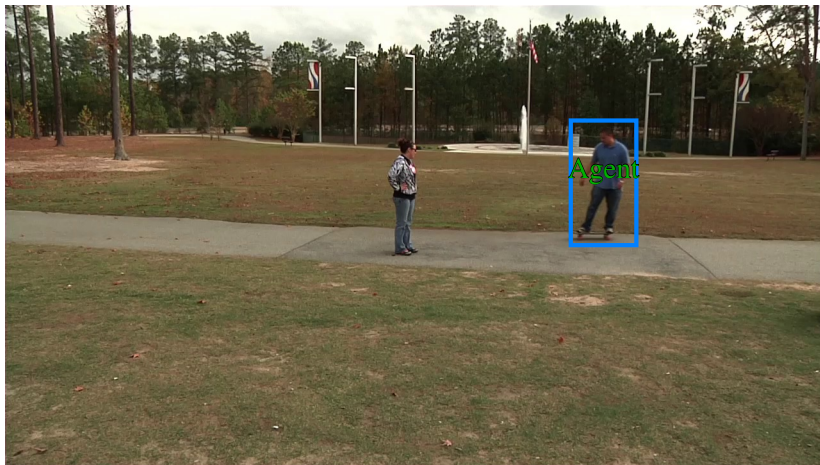


position	$\frac{d}{dt}$ position	$\frac{d^2}{dt^2}$ position	
aspect ratio	$\frac{d}{dt}$ aspect ratio	area	$\frac{d}{dt}$ area
object class	root-filter index		

Feature vector—dual participant



Feature vector—dual participant



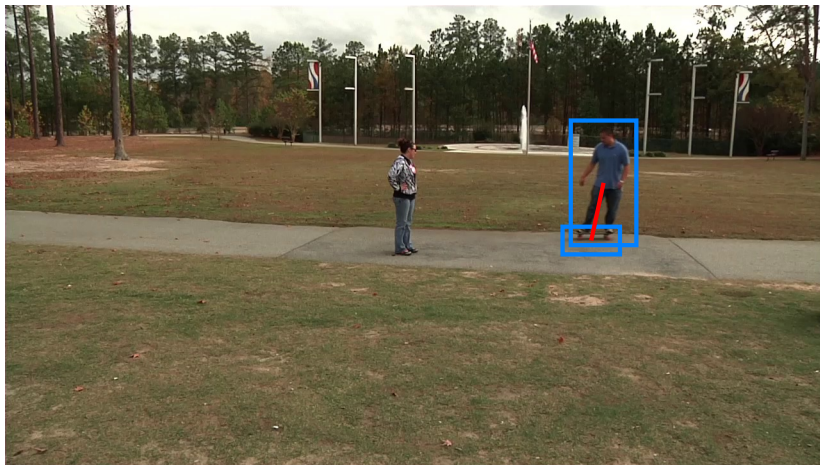
Feature vector—dual participant



Feature vector—dual participant

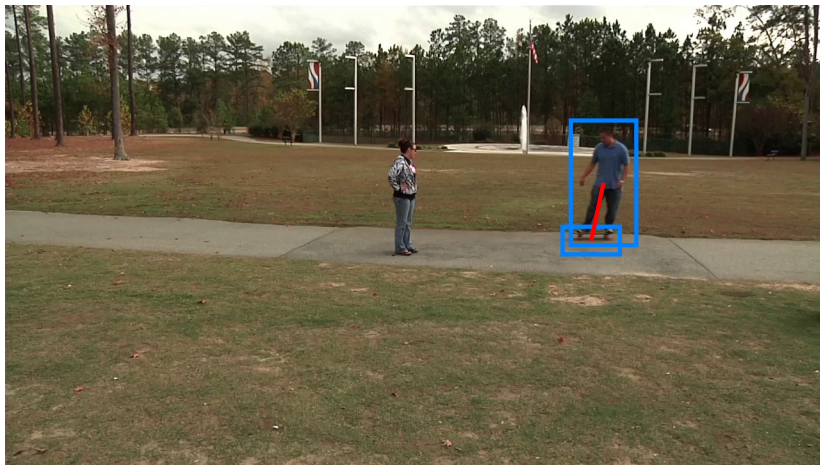


Feature vector—dual participant



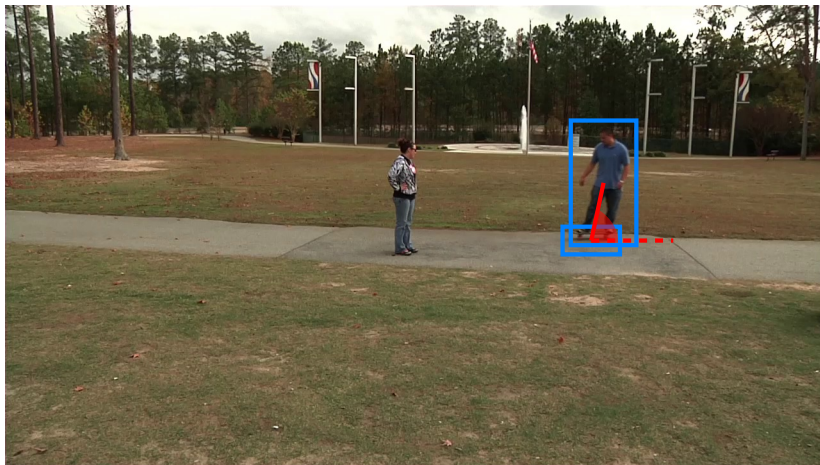
distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

Feature vector—dual participant



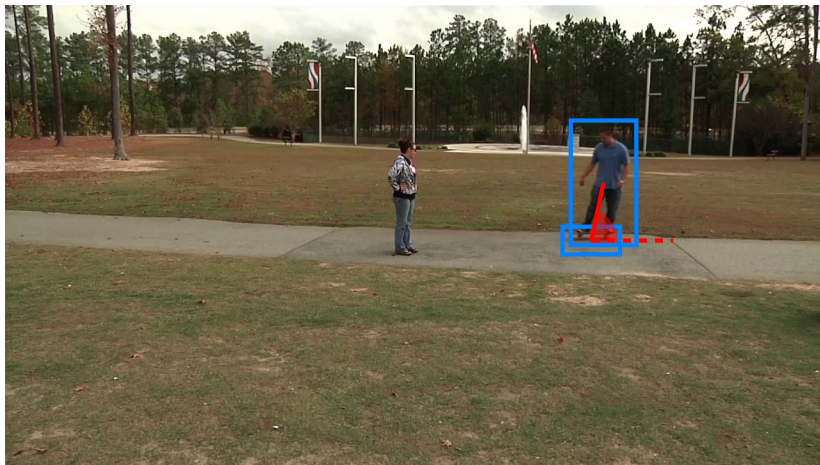
distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

Feature vector—dual participant



distance $\frac{d}{dt}$ distance **orientation** $\frac{d}{dt}$ orientation

Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

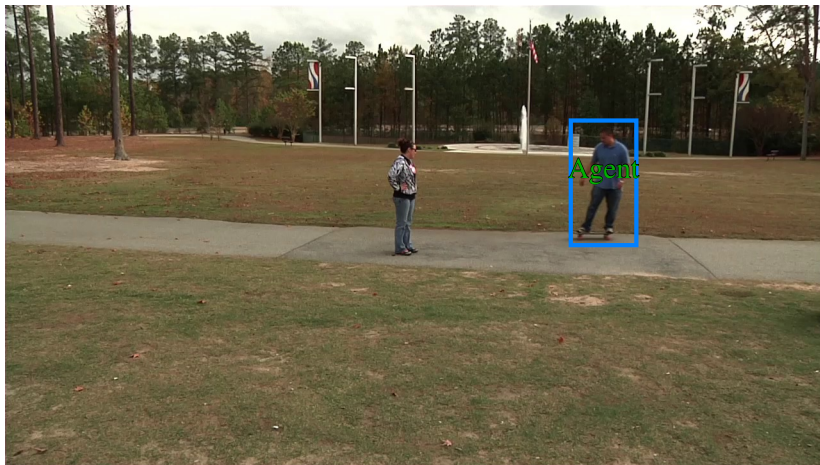
Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

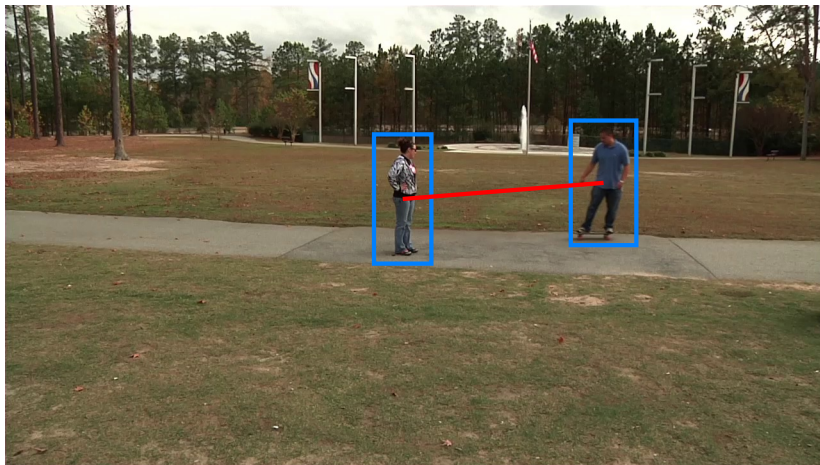
Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

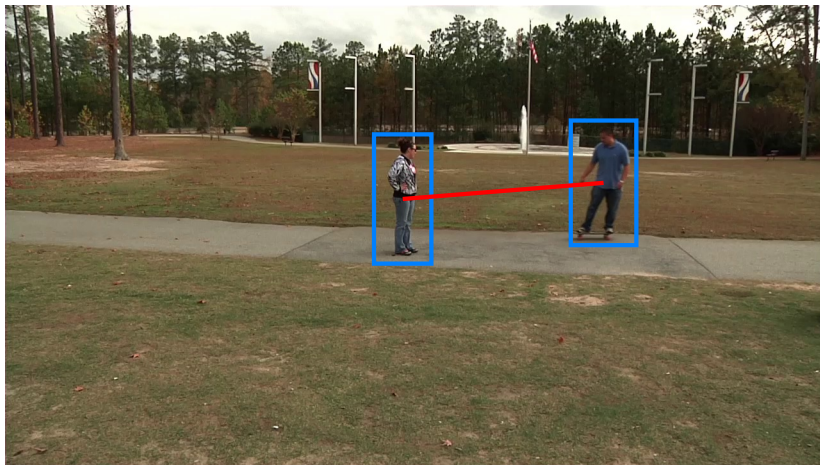
Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

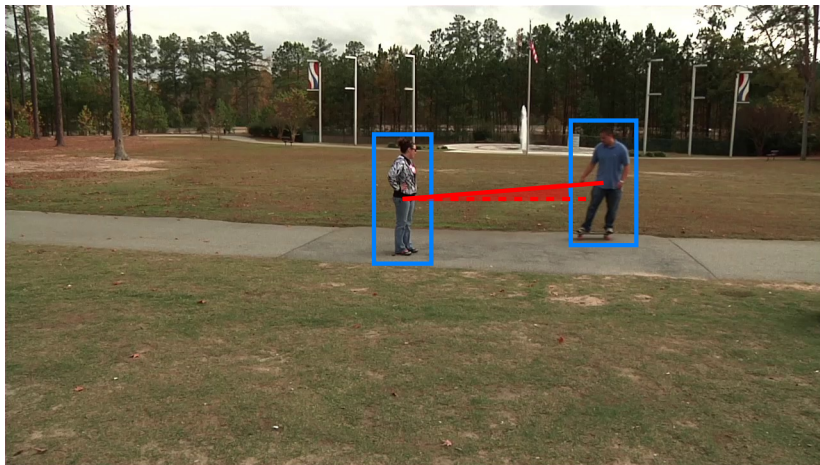
Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

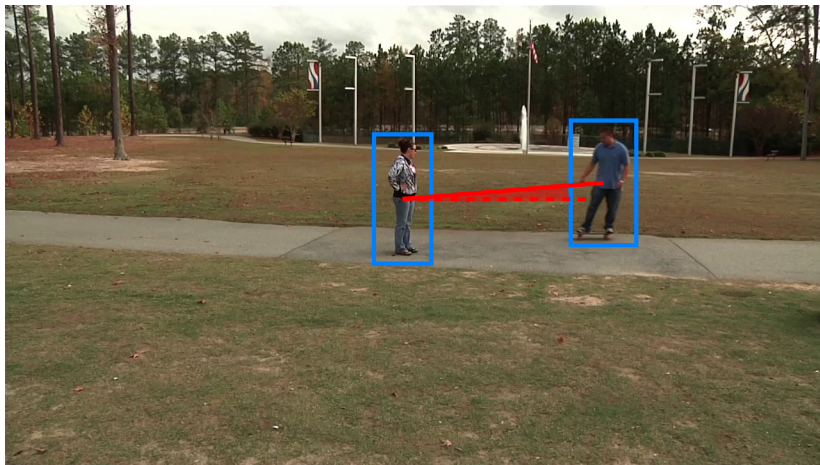
Feature vector—dual participant



distance $\frac{d}{dt}$ distance **orientation** $\frac{d}{dt}$ orientation

person riding skateboard

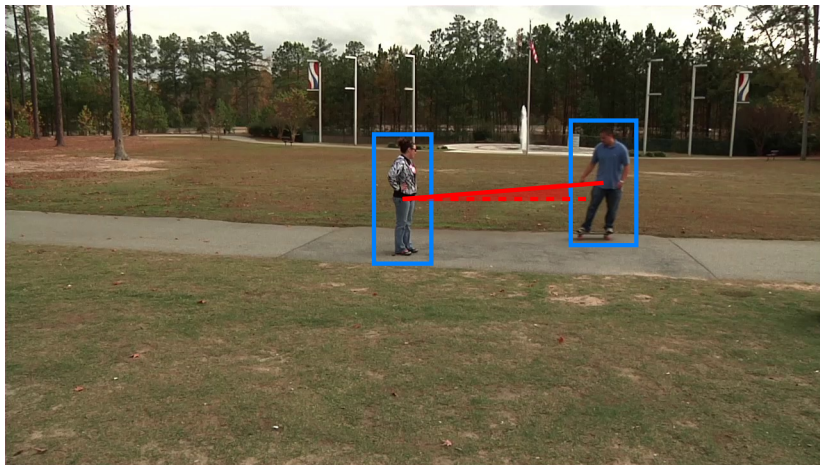
Feature vector—dual participant



distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

Feature vector—dual participant

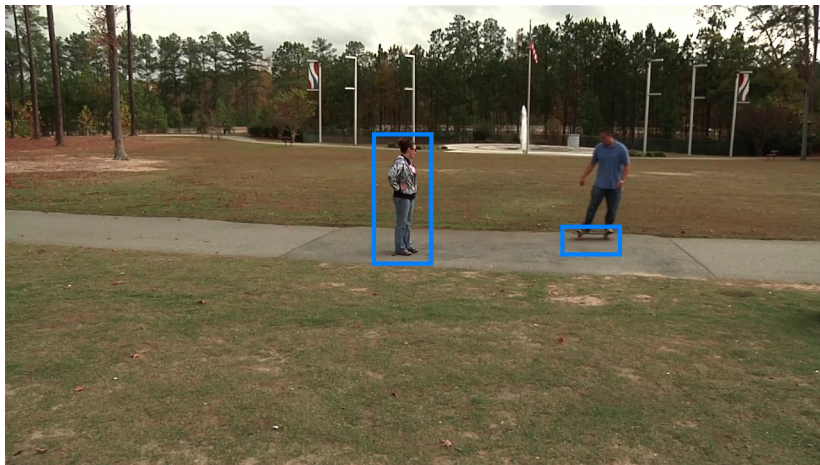


distance $\frac{d}{dt}$ distance orientation $\frac{d}{dt}$ orientation

person riding skateboard

person approaching person

Feature vector—dual participant



distance

$\frac{d}{dt}$ distance

orientation

$\frac{d}{dt}$ orientation

person riding skateboard

person approaching person

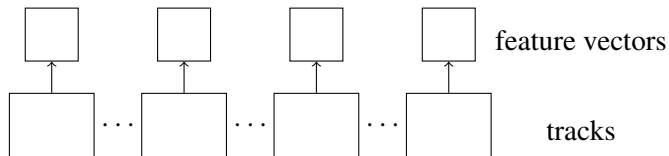
skateboard approaching person

Event recognition

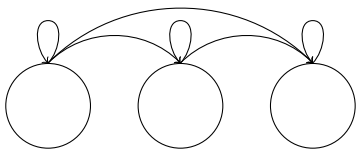


tracks

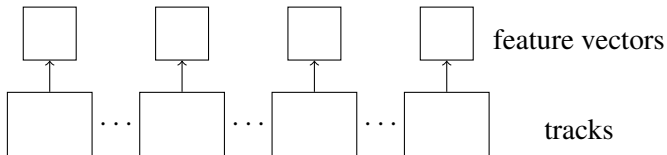
Event recognition



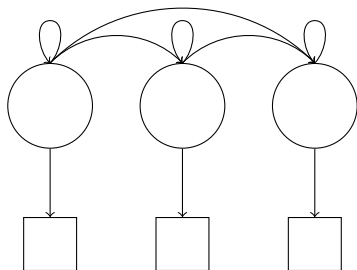
Event recognition



- a* ▶ HMMs
Baum and Petrie (1966)



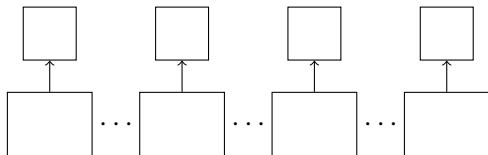
Event recognition



a ▶ HMMs
Baum and Petrie (1966)

h

feature vectors



tracks

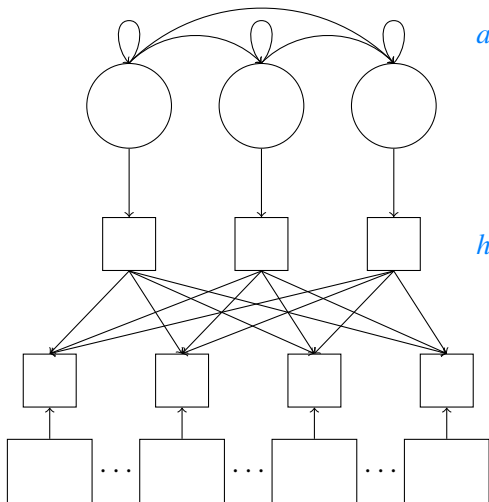
Event recognition

a ▶ HMMs
Baum and Petrie (1966)

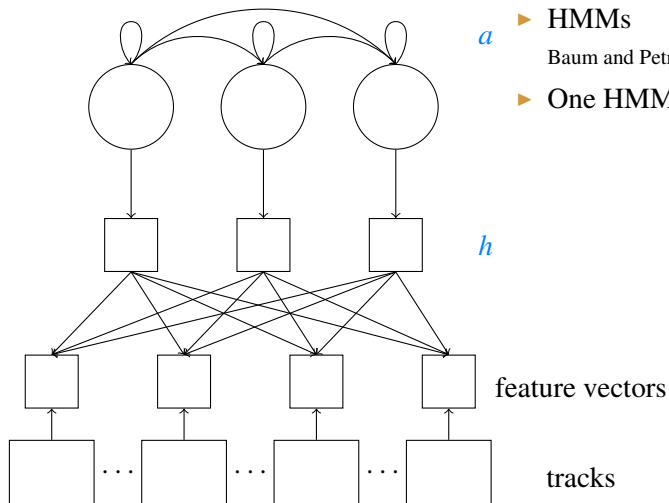
h

feature vectors

tracks



Event recognition

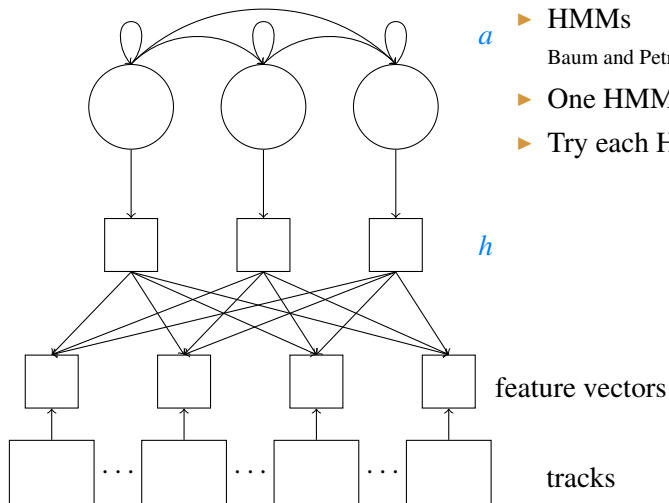


- a ▶ HMMs
Baum and Petrie (1966)
- ▶ One HMM per event class

feature vectors

tracks

Event recognition



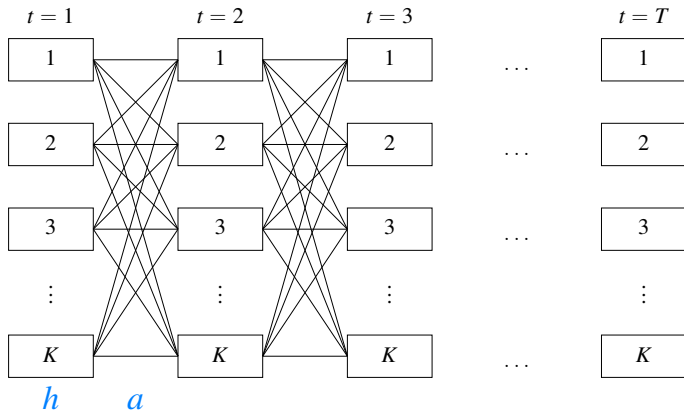
- a ▶ HMMs
Baum and Petrie (1966)
- ▶ One HMM per event class
- ▶ Try each HMM

h

feature vectors

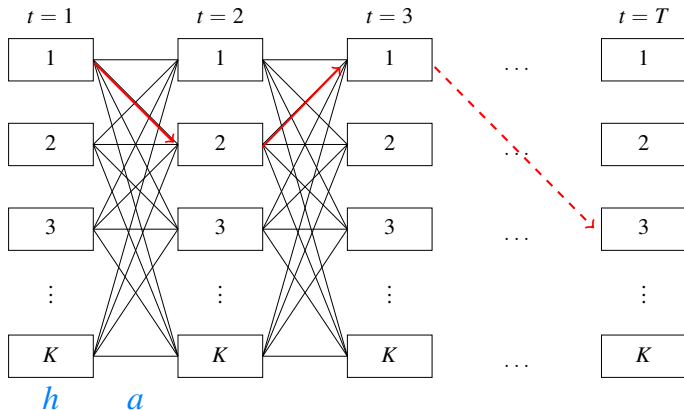
tracks

Event recognition



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

Event recognition



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

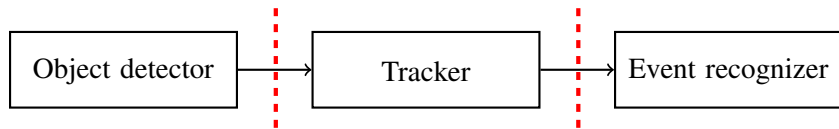
Examples



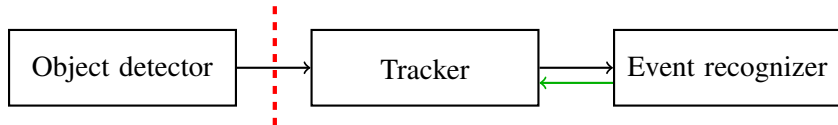
Examples



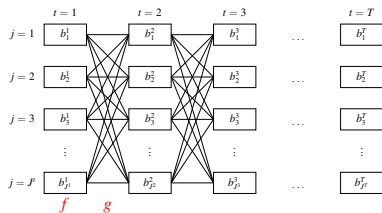
What we're going to do



What we're going to do

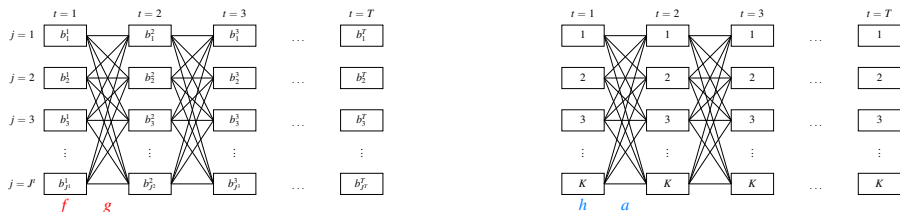


Tracking in the context of event recognition



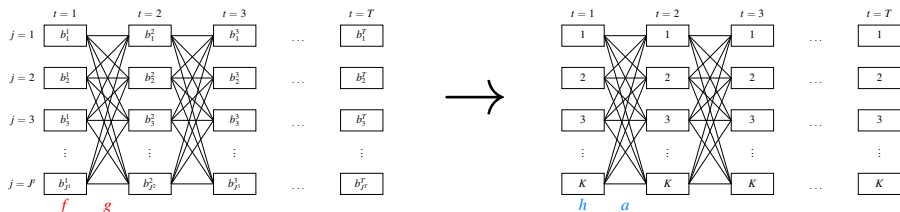
$$\max_{j^1, \dots, j^T} \sum_{t=1}^T f(b_{j^t}^t) + \sum_{t=2}^T g(b_{j^{t-1}}^{t-1}, b_{j^t}^t)$$

Tracking in the context of event recognition



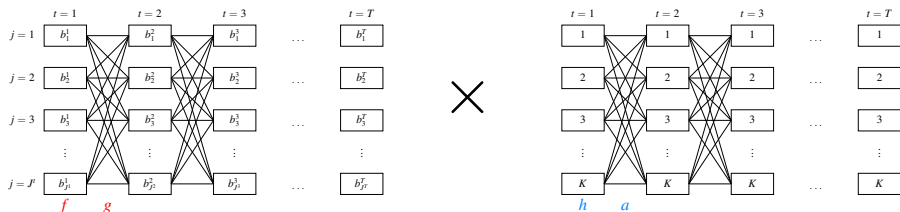
$$\max_{j^1, \dots, j^T} \sum_{t=1}^T f(b_{j^t}^t) + \sum_{t=2}^T g(b_{j^{t-1}}^{t-1}, b_{j^t}^t) + \max_{k^1, \dots, k^T} \sum_{t=1}^T h(k^t, b_{j^t}^t) + \sum_{t=2}^T a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



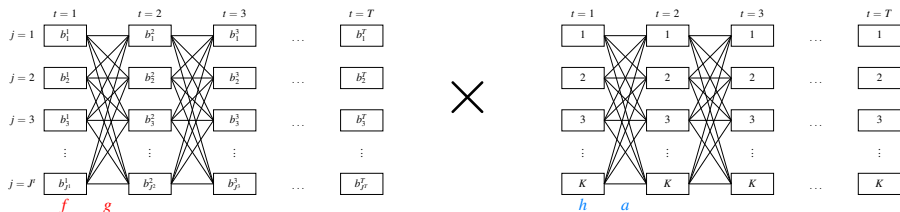
$$\max_{j^1, \dots, j^T} \sum_{t=1}^T f(b_{j^t}^t) + \sum_{t=2}^T g(b_{j^{t-1}}^{t-1}, b_{j^t}^t) + \max_{k^1, \dots, k^T} \sum_{t=1}^T h(k^t, b_{j^t}^t) + \sum_{t=2}^T a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



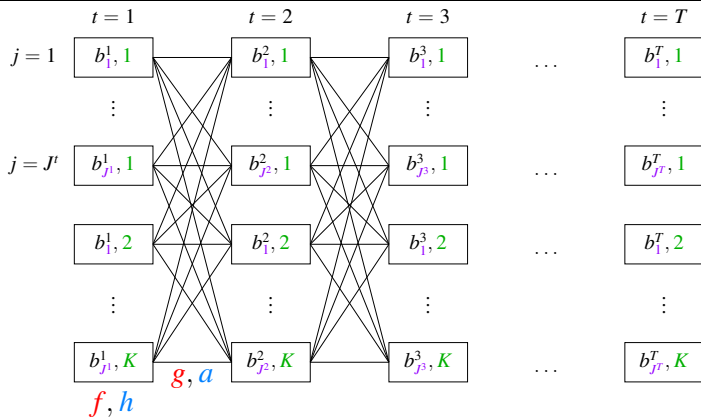
$$\max_{j^1, \dots, j^T} \sum_{t=1}^T f(b_{j^t}^t) + \sum_{t=2}^T g(b_{j^{t-1}}^{t-1}, b_{j^t}^t) + \max_{k^1, \dots, k^T} \sum_{t=1}^T h(k^t, b_{j^t}^t) + \sum_{t=2}^T a(k^{t-1}, k^t)$$

Tracking in the context of event recognition



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

Tracking in the context of event recognition



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

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

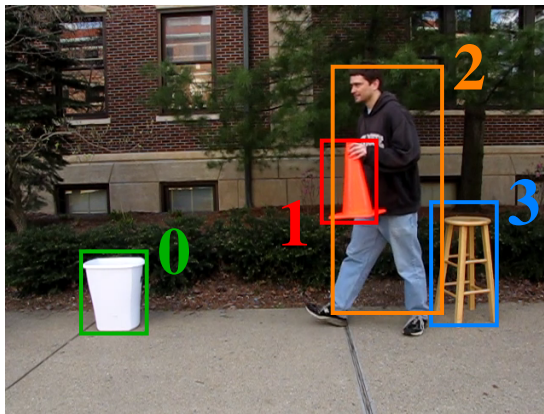
The person to the left of the stool carried the traffic-cone towards the trash-can.

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.

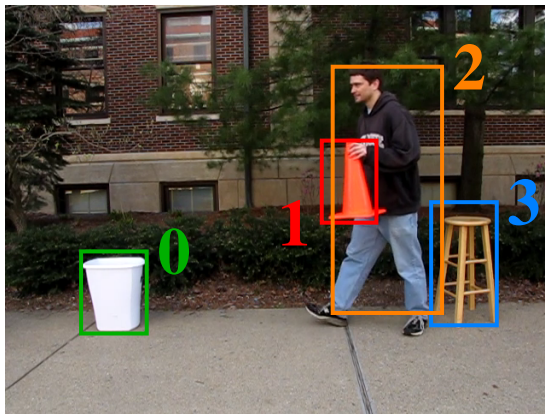
agent-tracker

object 0

object 1

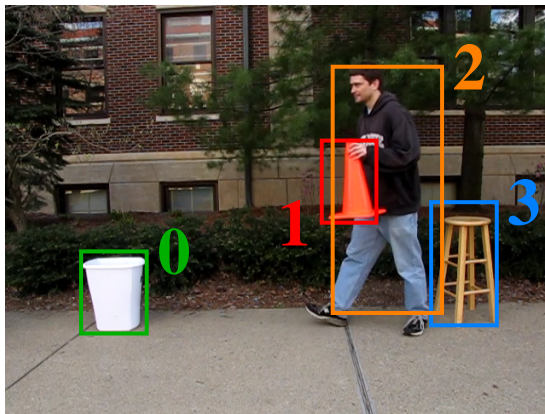
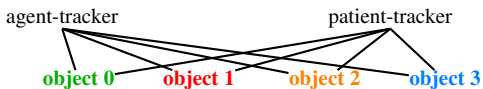
object 2

object 3



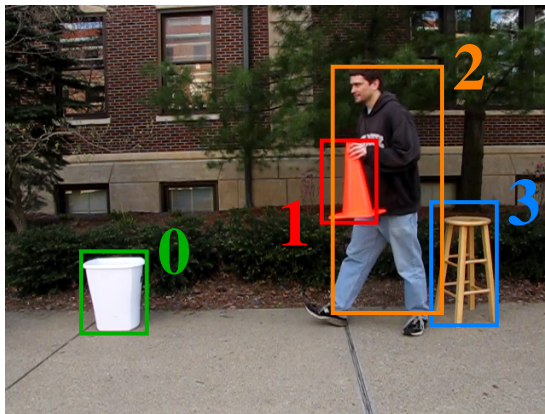
Sentence Tracker

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



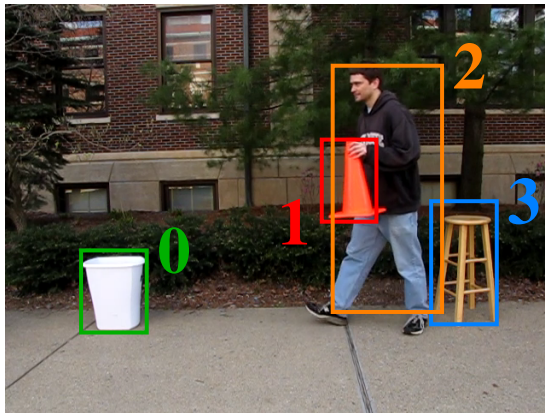
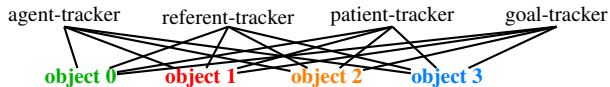
Sentence Tracker

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



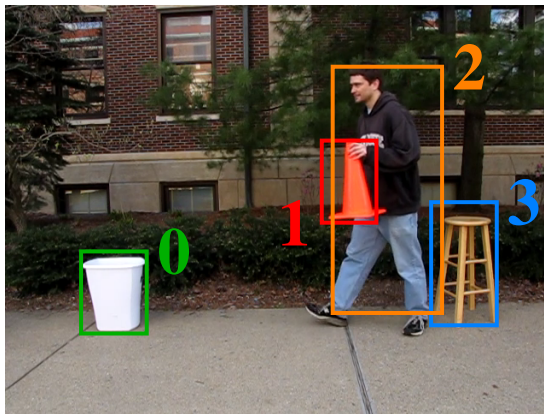
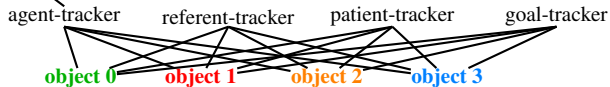
Sentence Tracker

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



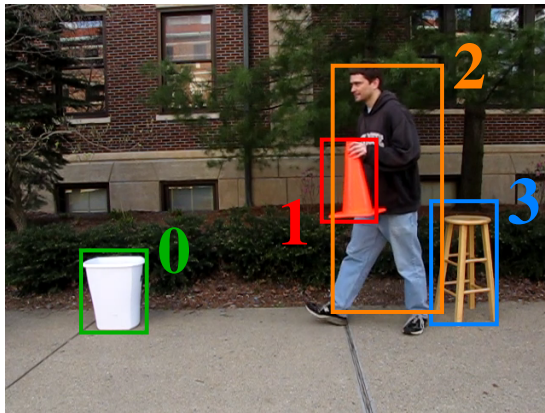
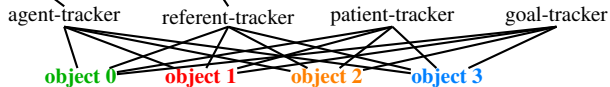
Sentence Tracker

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



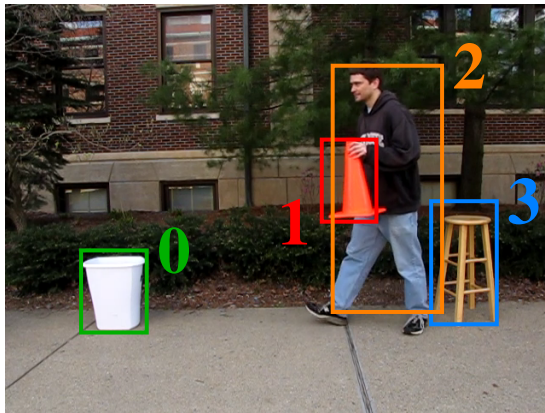
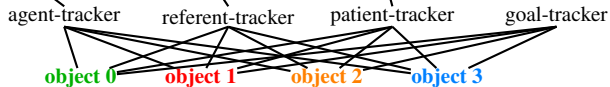
Sentence Tracker

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



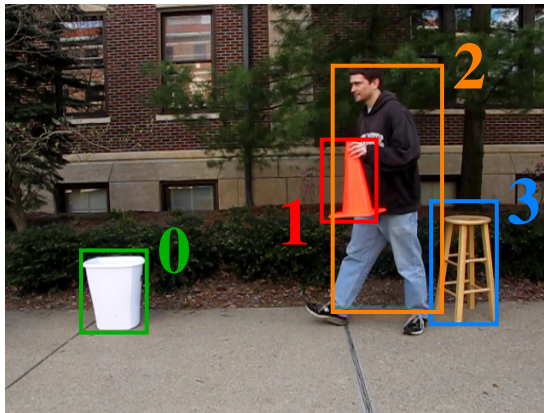
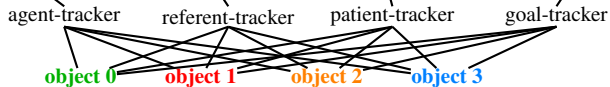
Sentence Tracker

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



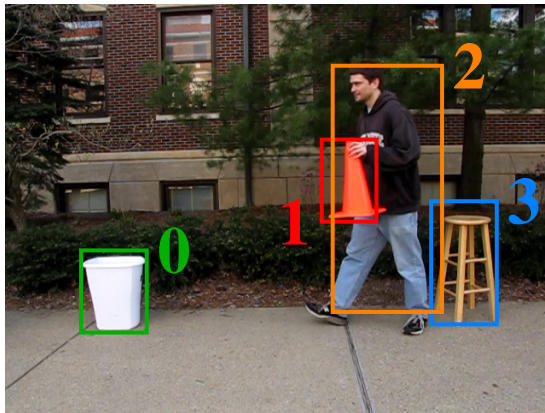
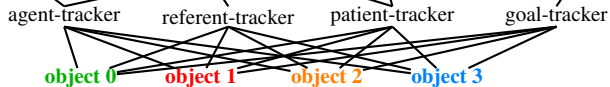
Sentence Tracker

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



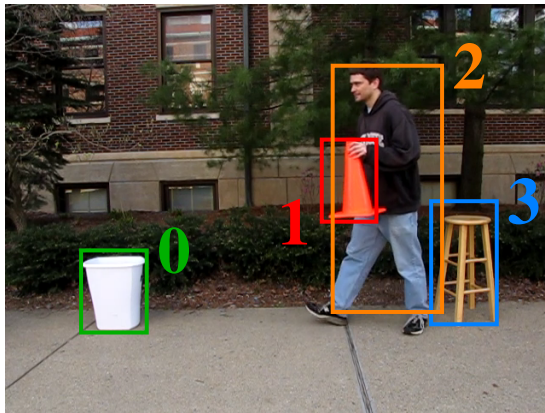
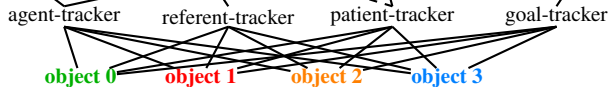
Sentence Tracker

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



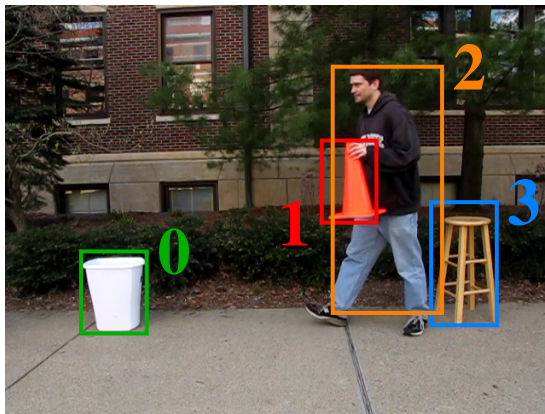
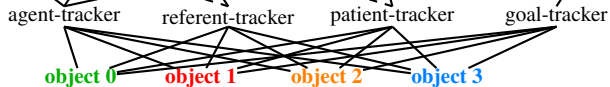
Sentence Tracker

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



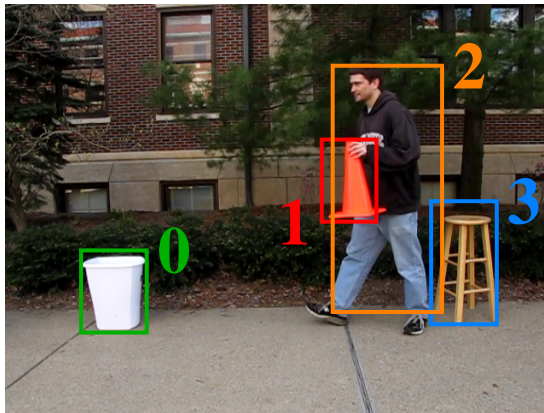
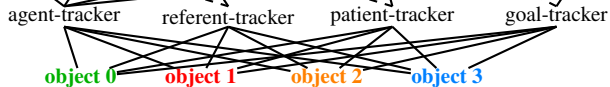
Sentence Tracker

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

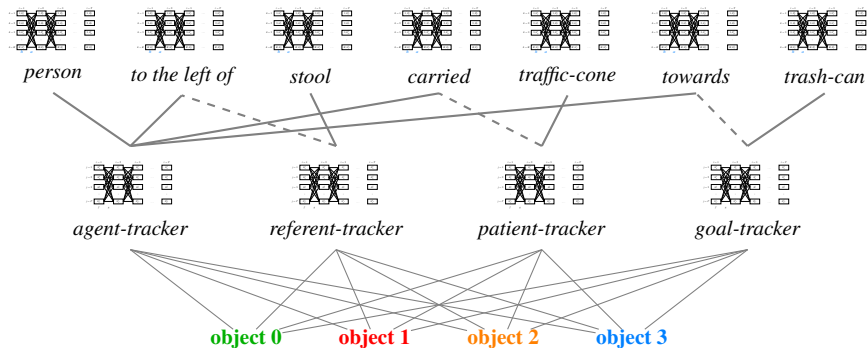


Sentence Tracker

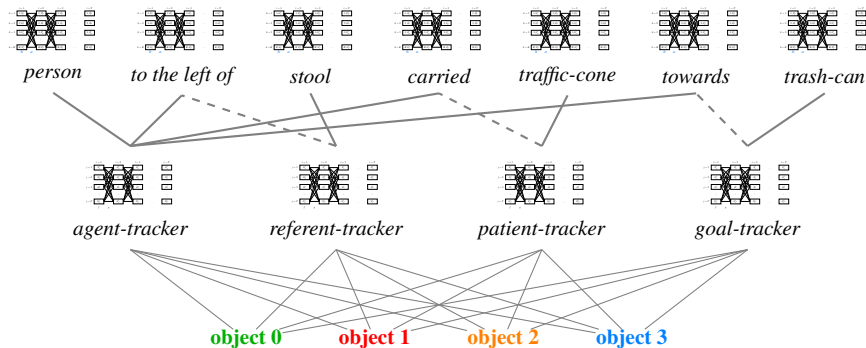
The person to the left of the stool carried the traffic-cone towards the trash-can.



Sentence Tracker

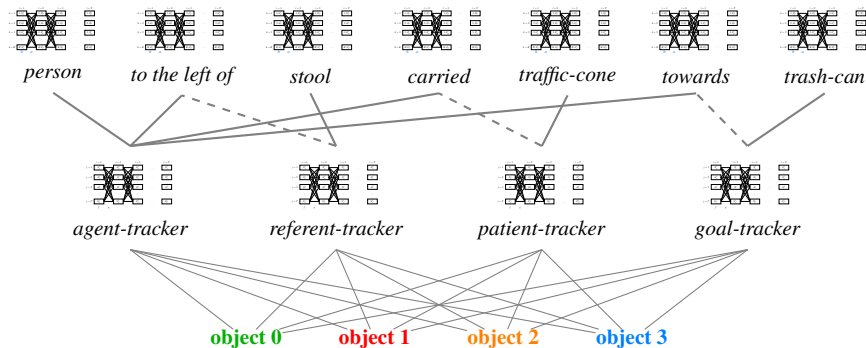


Sentence Tracker



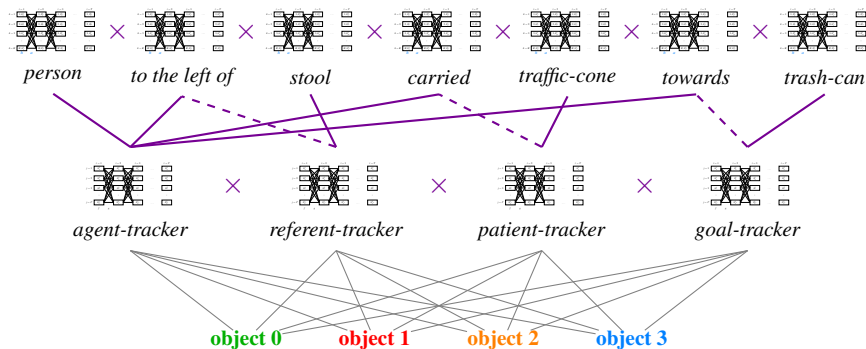
$$\max_{\substack{j_1^1, \dots, j_1^T \\ j_L^1, \dots, j_L^T}} \left(\sum_{l=1}^L \sum_{t=1}^T f(b_{j_l^t}^t) + \sum_{t=2}^T g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) \right) + \max_{\substack{k_1^1, \dots, k_1^T \\ k_W^1, \dots, k_W^T}} \left(\sum_{w=1}^W \sum_{t=1}^T h_w(k_w^t, b_{w,1}^t, b_{w,2}^t) + \sum_{t=2}^T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

Sentence Tracker



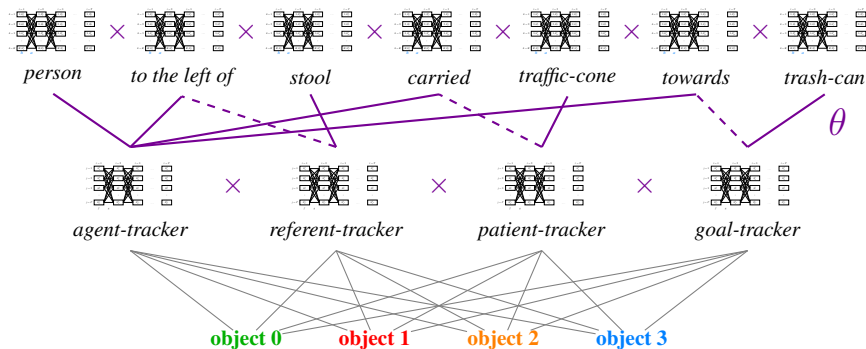
$$\max_{j_1^1, \dots, j_1^T} \max_{k_1^1, \dots, k_1^T} \sum_{l=1}^L \sum_{t=1}^T f(b_{j_l^t}^t) + \sum_{t=2}^T g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \sum_{w=1}^W \sum_{t=1}^T h_w(k_w^t, b_{j_{\theta_w^1}^t}, b_{j_{\theta_w^2}^t}) + \sum_{t=2}^T a_w(k_w^{t-1}, k_w^t)$$

Sentence Tracker



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

Sentence Tracker



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

Unified Cost Function

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

- ▶ **B**: video
- ▶ **s**: sentence
- ▶ Λ : lexicon
- ▶ τ : score
- ▶ **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

$$\mathbf{J}_1 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_1, \Lambda)$$

$$\mathbf{J}_2 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_2, \Lambda)$$

Four Uses of Unified Cost Function

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

$$\mathbf{J}_1 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_1, \Lambda)$$

$$\mathbf{J}_2 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_2, \Lambda)$$

- ▶ **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

$$\mathbf{J}_1 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_1, \Lambda)$$

$$\mathbf{J}_2 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_2, \Lambda)$$

- ▶ **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

$$\mathbf{J}_1 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_1, \Lambda)$$

$$\mathbf{J}_2 = \mathcal{S}_J(\mathbf{B}, \mathbf{s}_2, \Lambda)$$

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

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

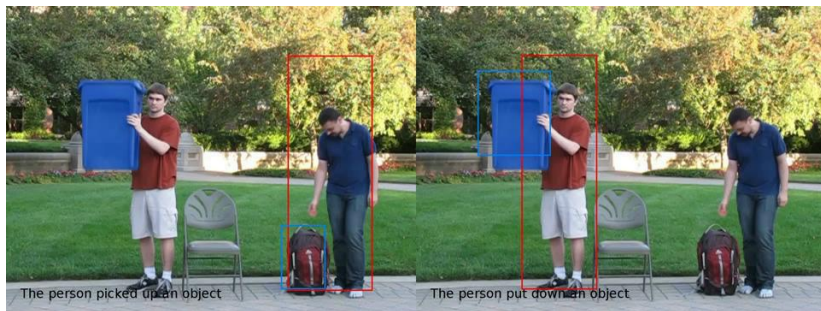
- ▶ **Retrieval:** sentence \times lexicon \rightarrow video

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

- ▶ **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

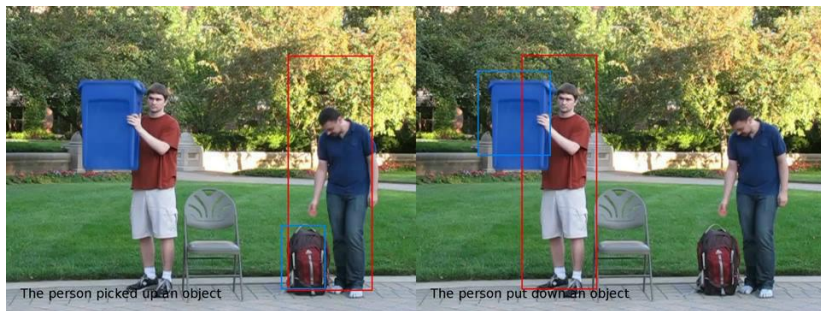


We can differentiate events based on the **verb**:

*The person **picked up** an object.*

*The person **put down** an object.*

Using the Sentence Tracker to Focus Attention

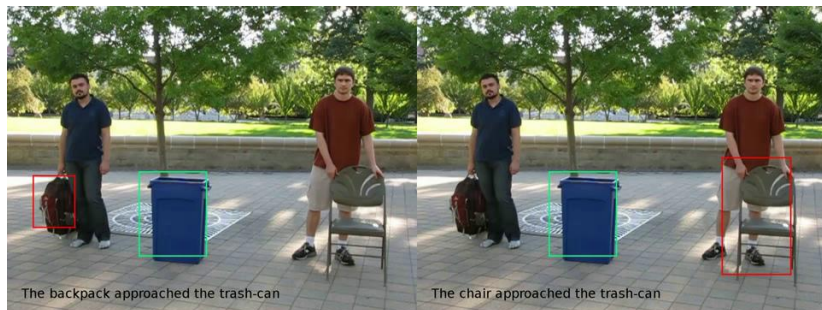


We can differentiate events based on the **verb**:

*The person **picked up** an object.*

*The person **put down** an object.*

Using the Sentence Tracker to Focus Attention

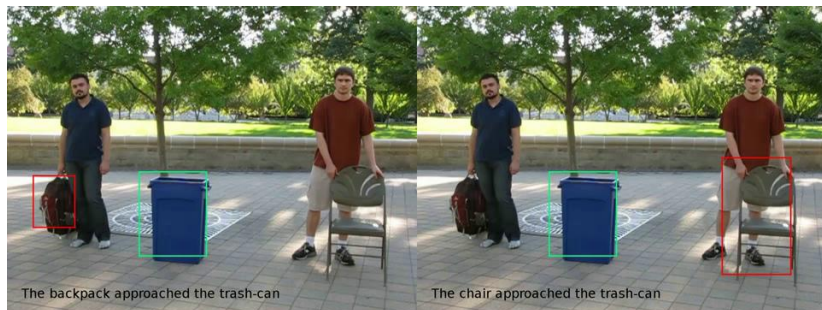


We can differentiate events based on the **subject noun**:

*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

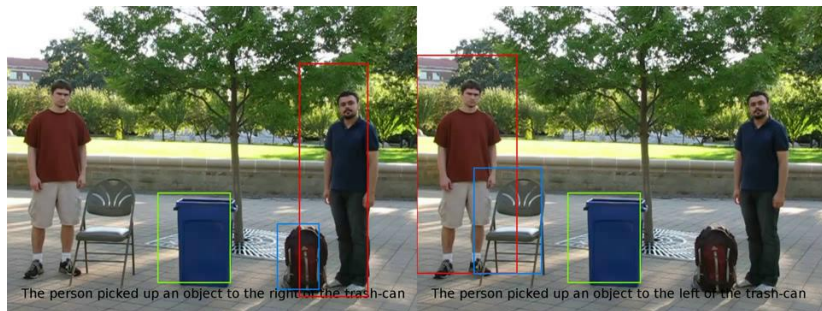


We can differentiate events based on an **adjective in the object NP**:

*The person carried the **blue** object.*

*The person carried the **red** object.*

Using the Sentence Tracker to Focus Attention

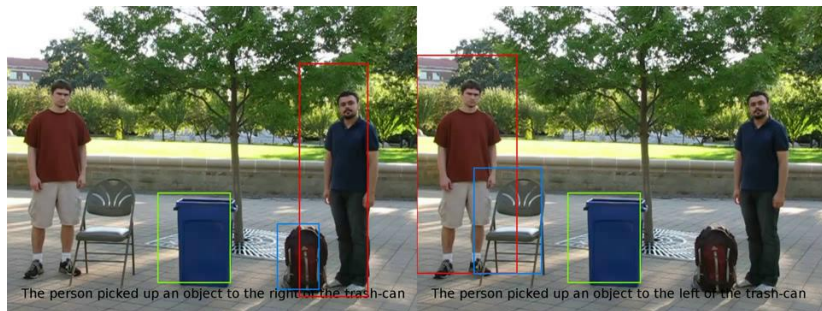


We can differentiate events based on a **preposition in the object NP**:

*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.

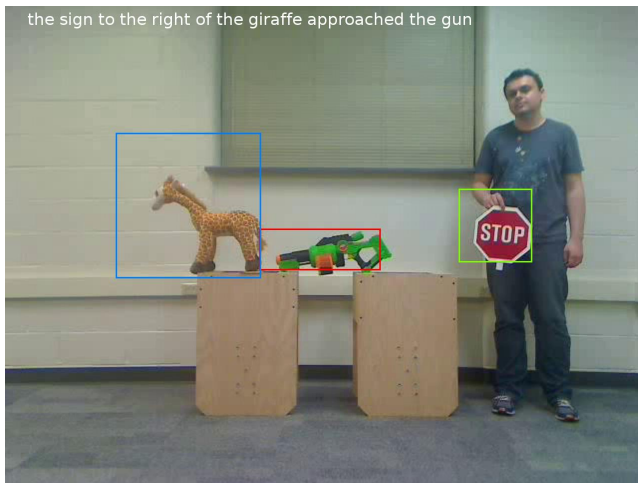
Using the Sentence Tracker to Perform Generation

Beam Search

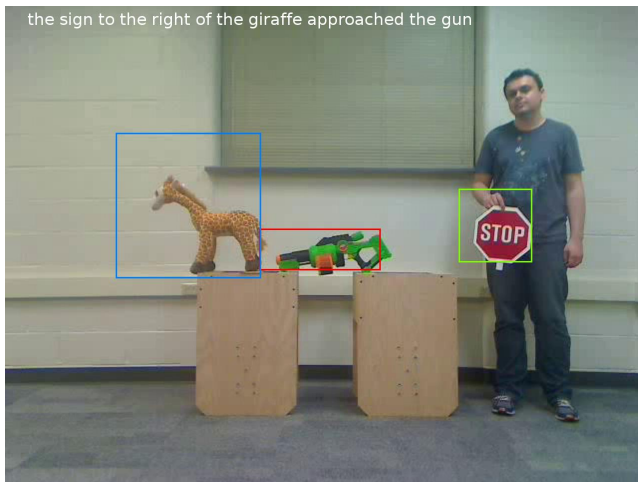
base case Find the k top-scoring 1-word phrases.

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

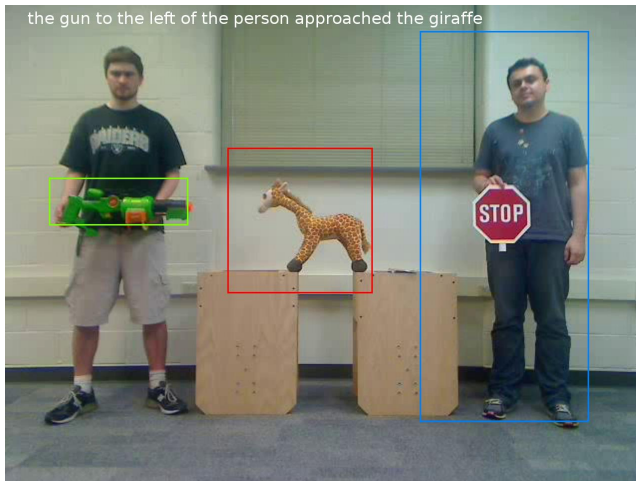


Using the Sentence Tracker to Perform Generation

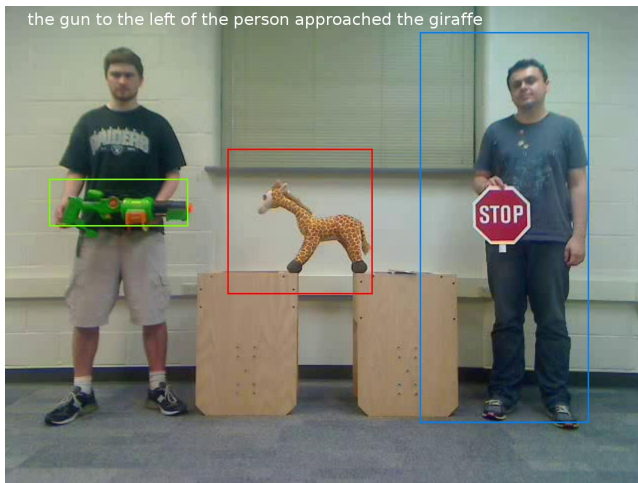


The sign to the right of the giraffe approached the gun.

Using the Sentence Tracker to Perform Generation

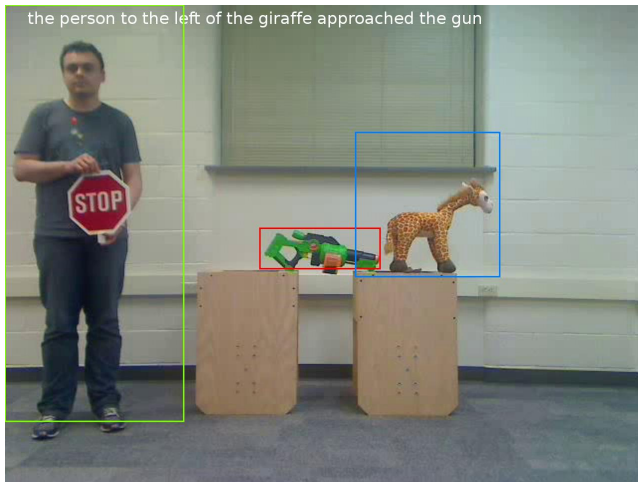


Using the Sentence Tracker to Perform Generation

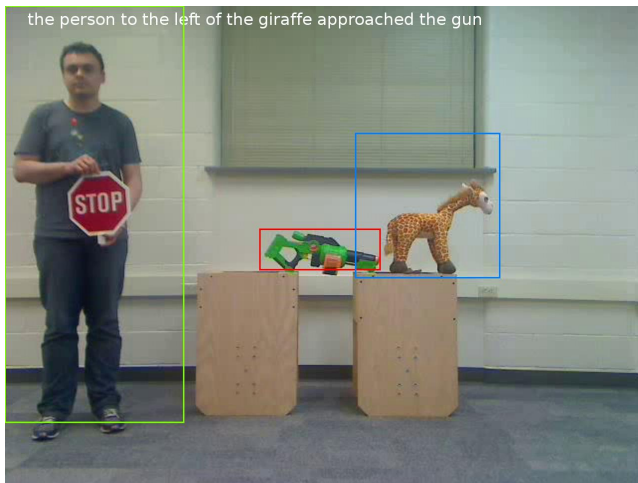


The gun to the left of the person approached the giraffe.

Using the Sentence Tracker to Perform Generation

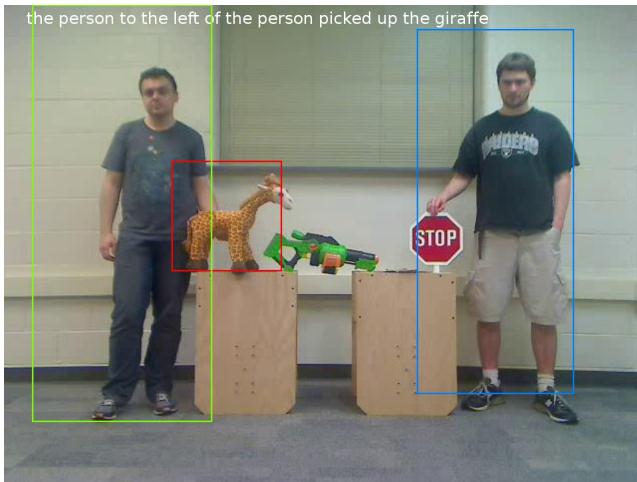


Using the Sentence Tracker to Perform Generation

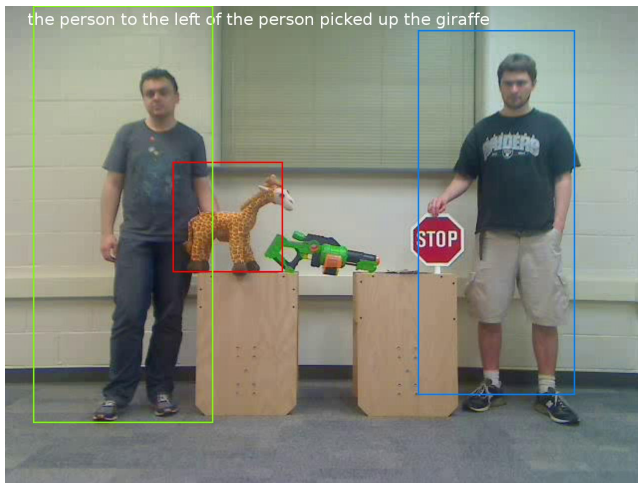


The person to the left of the giraffe approached the gun.

Using the Sentence Tracker to Perform Generation

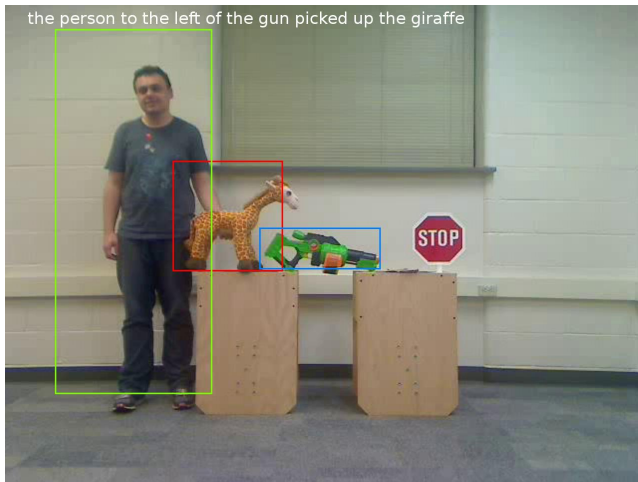


Using the Sentence Tracker to Perform Generation

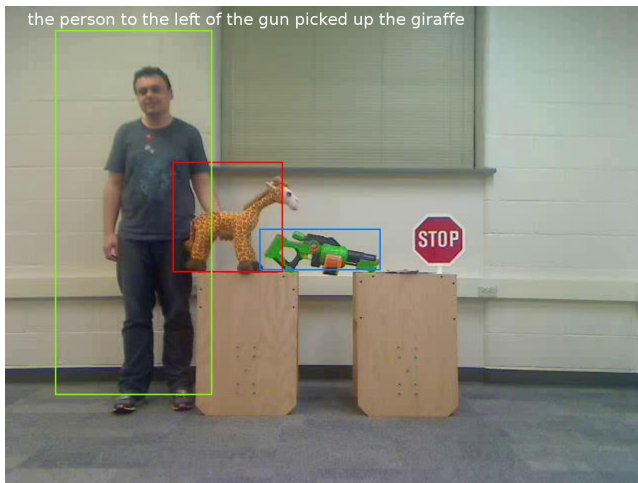


The person to the left of the person picked up the giraffe.

Using the Sentence Tracker to Perform Generation



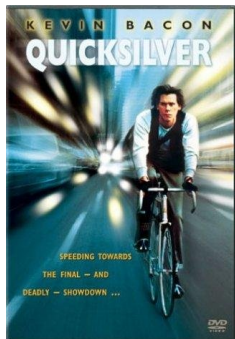
Using the Sentence Tracker to Perform Generation



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 rode the bicycle leftward.

Using the Sentence Tracker to Perform Retrieval

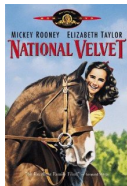
The person rode the bicycle rightward.

Using the Sentence Tracker to Perform Retrieval



The person rode the bicycle rightward.

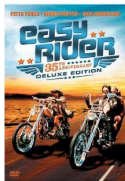
Using the Sentence Tracker to Perform Retrieval



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)



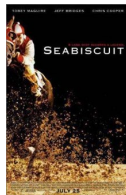
Unforgiven
(Warner Bros. Pictures, 1992)



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

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

- ▶ 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.

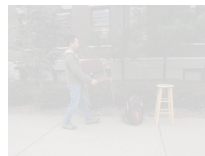
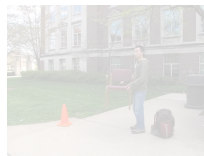
Using the Sentence Tracker to Perform Acquisition



*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.*

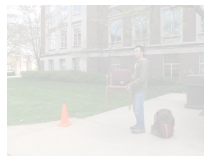
Using the Sentence Tracker to Perform Acquisition



*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.*

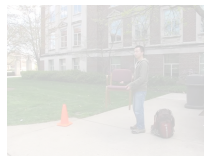
Using the Sentence Tracker to Perform Acquisition



*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.*

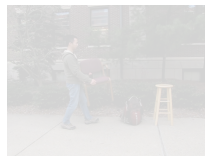
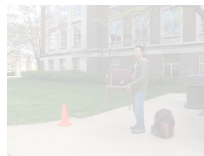
Using the Sentence Tracker to Perform Acquisition



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

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

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 put down the chair quickly.*
The person carried the trash-can. ...

How Can We Do This?



chair



picked-up



person

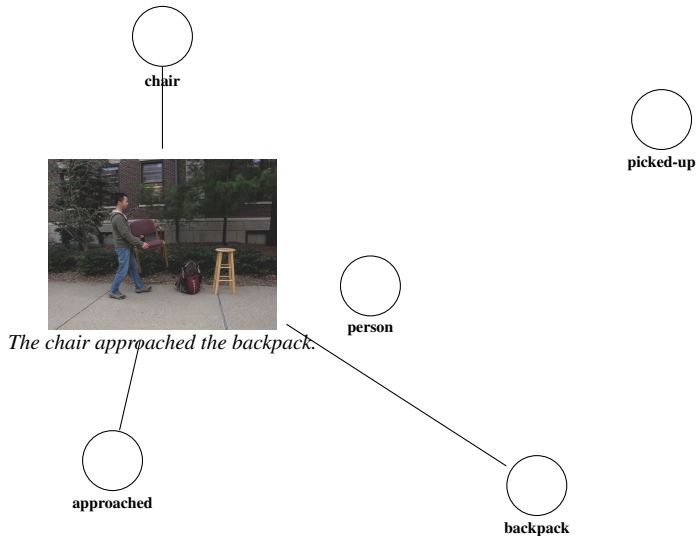


approached

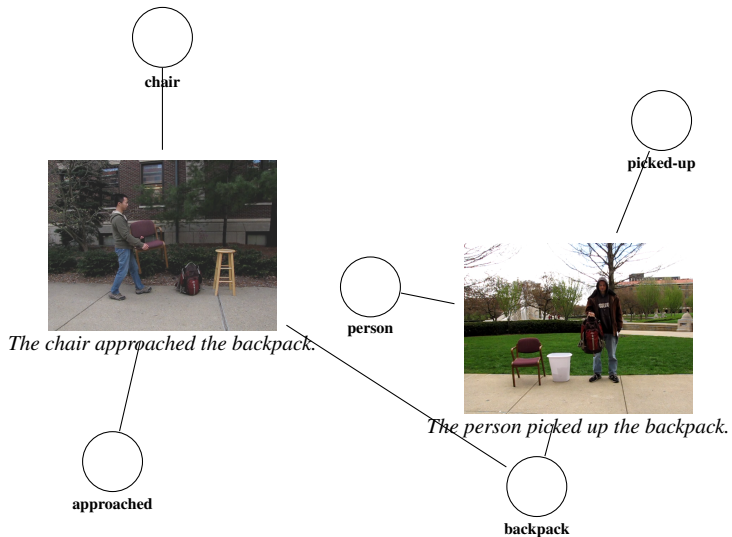


backpack

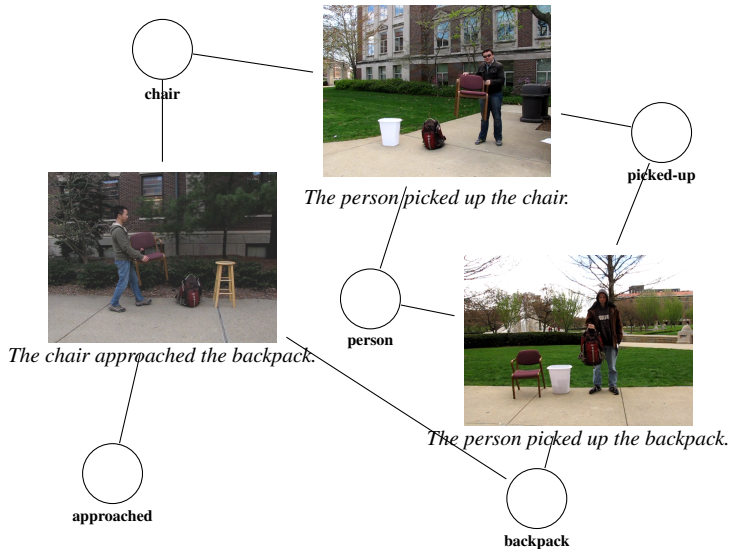
How Can We Do This?



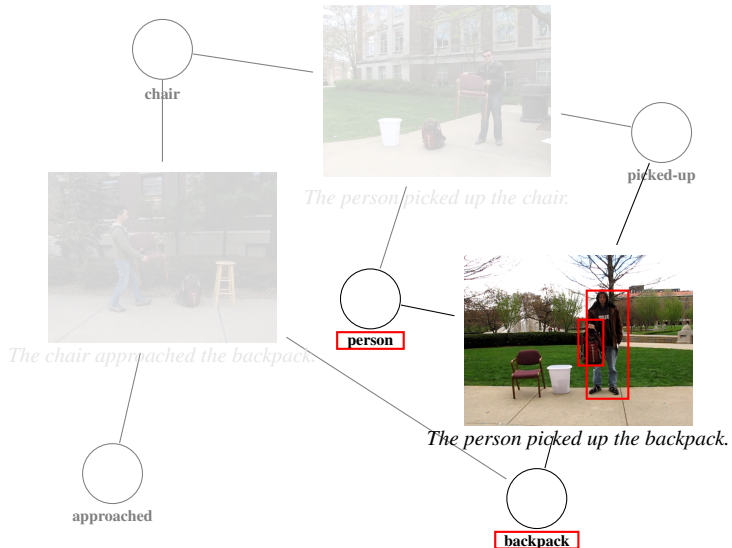
How Can We Do This?



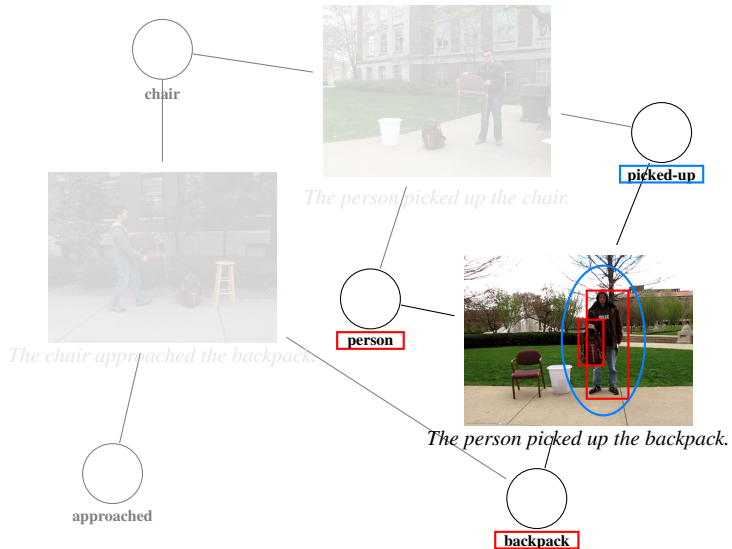
How Can We Do This?



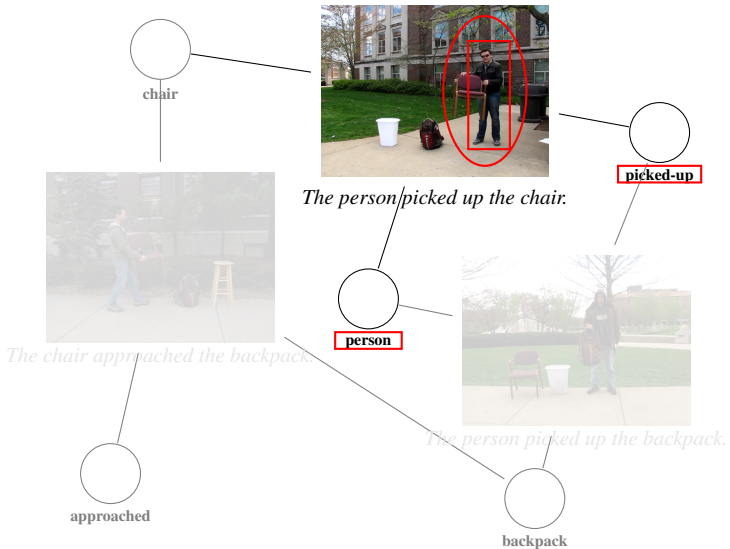
Inference Across *Different* Words in the *Same* Sentence



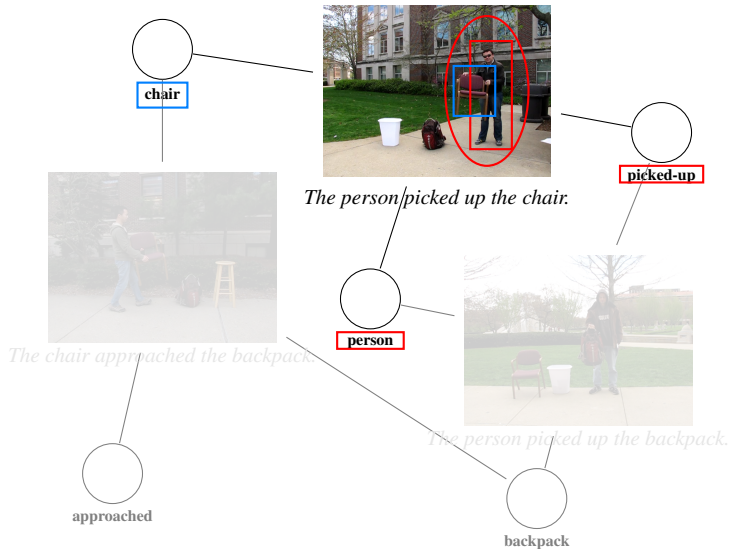
Inference Across *Different* Words in the *Same* Sentence



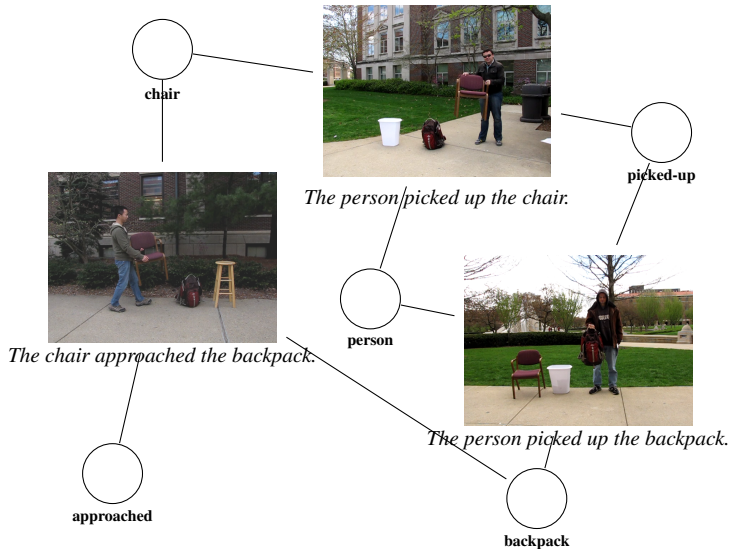
Inference Across the *Same* Word in *Different* Sentences



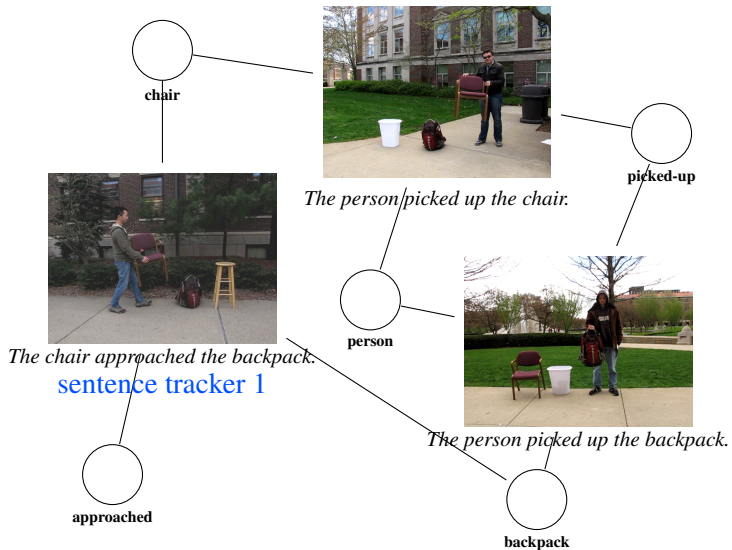
Inference Across the *Same* Word in *Different* Sentences



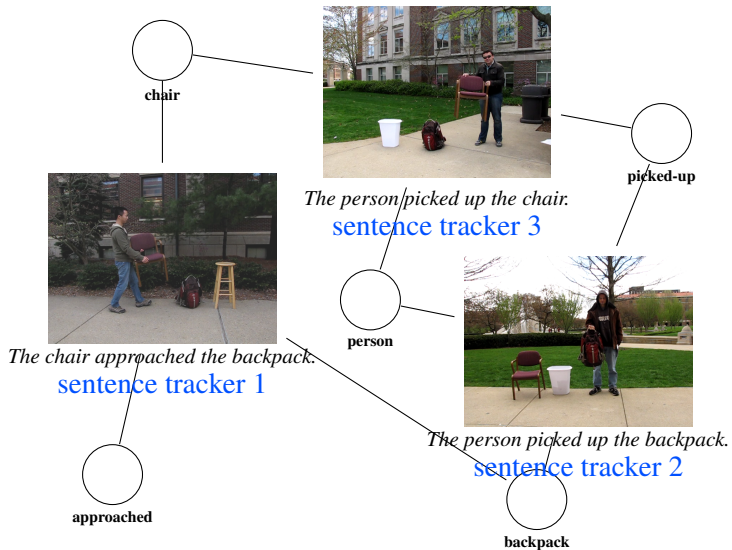
Sentence Tracker in EM



Sentence Tracker in EM



Sentence Tracker in EM



Sentence Tracker in EM

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

Sentence Tracker in EM

$$\log \sum_{\substack{j_1^1, \dots, j_1^T \\ j_L^1, \dots, j_L^T}} \sum_{\substack{k_1^1, \dots, k_1^T \\ k_W^1, \dots, k_W^T}} \exp \left[\begin{array}{l} \sum_{l=1}^L \sum_{t=1}^T f(b_{j_l^t}^t) + \sum_{t=2}^T g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \\ \sum_{w=1}^W \sum_{t=1}^T h_w(k_w^t, b_{\theta_w^1}^{j_1^t}, b_{\theta_w^2}^{j_2^t}) + \sum_{t=2}^T a_w(k_w^{t-1}, k_w^t) \end{array} \right]$$

Sentence Tracker in EM

$$\log \sum_{\substack{j_1^1, \dots, j_1^T \\ j_L^1, \dots, j_L^T}} \sum_{\substack{k_1^1, \dots, k_1^T \\ k_W^1, \dots, k_W^T}} \exp \left[\begin{array}{l} \sum_{l=1}^L \sum_{t=1}^T f(b_{j_l^t}^t) + \sum_{t=2}^T g(b_{j_l^{t-1}}^{t-1}, b_{j_l^t}^t) + \\ \sum_{w=1}^W \sum_{t=1}^T h_w(k_w^t, b_{\theta_w^1}^{j_1^t}, b_{\theta_w^2}^{j_2^t}) + \sum_{t=2}^T a_w(k_w^{t-1}, k_w^t) \end{array} \right]$$

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

Sentence Tracker in EM

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

- ▶ 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

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

- ▶ 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

N → *person* | *backpack* | *trash-can* | *chair* | *traffic-cone* | *stool*

P → *to the left of* | *to the right of*

V → *picked up* | *put down* | *carried* | *approached*

ADV → *quickly* | *slowly*

PM → *towards* | *away from*

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

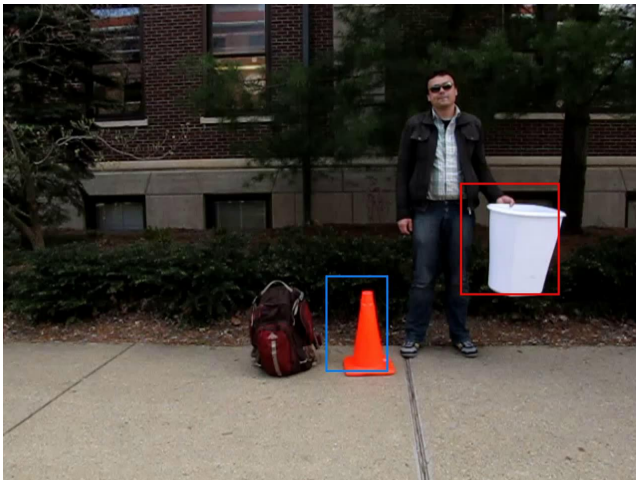


The person to the left of the stool picked up the chair.



The person carried the backpack towards the stool.

New Sentence Generation with Trained Models



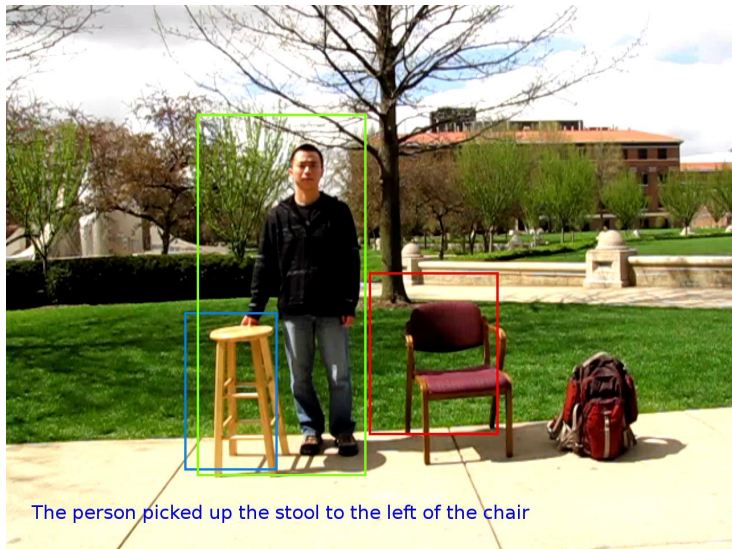
The trash-can approached the traffic-cone.

New Sentence Generation with Trained Models

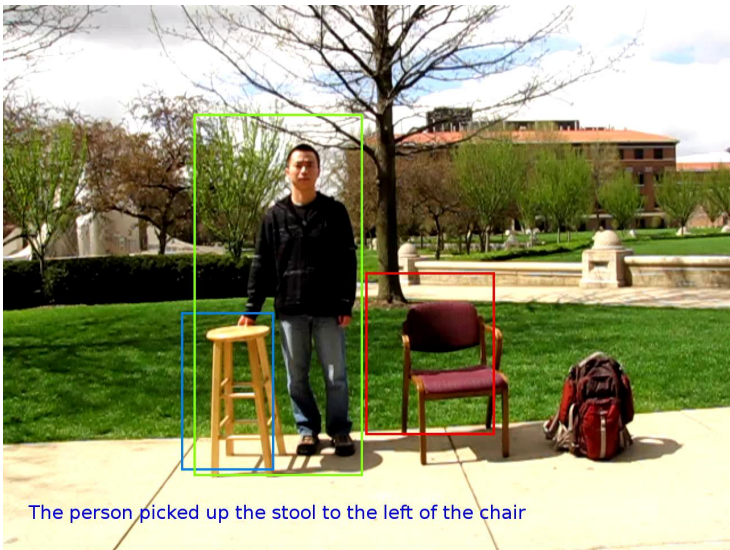


The trash-can approached the traffic-cone.

New Sentence Generation with Trained Models



New Sentence Generation with Trained Models



Outline

1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

5 Playing Checkers from English

Haonan Yu

Sentence-Directed Video Object Codetection

Sentence-Directed Video Object Codetection

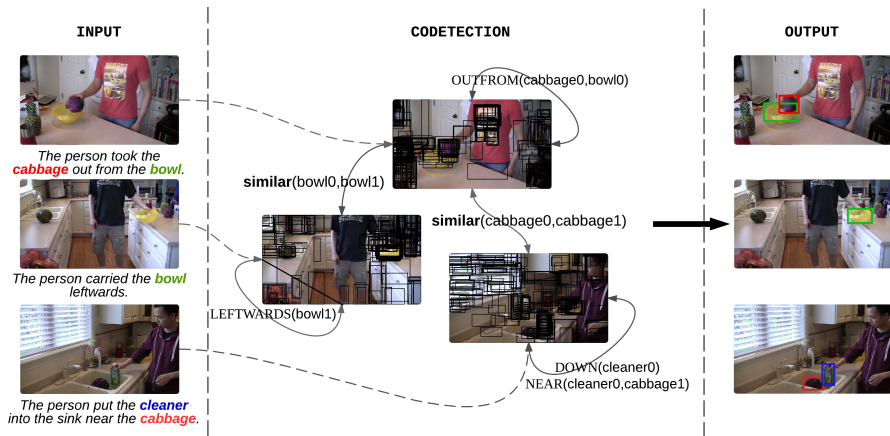
video captioning: video+detections→sentences

Sentence-Directed Video Object Codetection

video captioning: video+detections→sentences

inverse video captioning: video+sentences→detections

Overview



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.

DOWN(*cleaner*) \wedge 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.

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

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
$\text{MOVE}(p) \hat{=} \text{medFMg}(p)$	$\Delta\text{DISTLARGE} \hat{=} 0.25$
$\text{MOVEUP}(p) \hat{=} \text{MOVE}(p) + \text{distLessThan} \left(y(p^{(T)}) - y(p^{(1)}), -\Delta\text{DISTLARGE} \right)$	$\Delta\text{DISTSMALL} \hat{=} 0.05$
$\text{MOVEDOWN}(p) \hat{=} \text{MOVE}(p) + \text{distGreaterThan} \left(y(p^{(T)}) - y(p^{(1)}), \Delta\text{DISTLARGE} \right)$	$\Delta\text{ANGLE} \hat{=} \pi/2$
$\text{MOVEHORIZONTAL}(p) \hat{=} \text{MOVE}(p) + \text{distGreaterThan} \left(x(p^{(T)}) - x(p^{(1)}) , \Delta\text{DISTLARGE} \right)$	
$\text{MOVELEFTWARDS}(p) \hat{=} \text{MOVE}(p) + \text{distLessThan} \left(x(p^{(T)}) - x(p^{(1)}), -\Delta\text{DISTLARGE} \right)$	
$\text{MOVERIGHTWARDS}(p) \hat{=} \text{MOVE}(p) + \text{distGreaterThan} \left(x(p^{(T)}) - x(p^{(1)}), \Delta\text{DISTLARGE} \right)$	
$\text{ROTATE}(p) \hat{=} \text{MOVE}(p) + \max \text{ hasRotation} \left(\text{rotAngle}(p^{(0)}), \Delta\text{ANGLE} \right)$	
$\text{TOWARDS}(p_1, p_2) \hat{=} \text{MOVE}(p_1) + \text{distLessThan} \left(\text{dist}(p_1^{(T)}, p_2^{(T)}) - \text{dist}(p_1^{(1)}, p_2^{(1)}), -\Delta\text{DISTLARGE} \right)$	
$\text{AWAYFROM}(p_1, p_2) \hat{=} \text{MOVE}(p_1) + \text{distGreaterThan} \left(\text{dist}(p_1^{(T)}, p_2^{(T)}) - \text{dist}(p_1^{(1)}, p_2^{(1)}), \Delta\text{DISTLARGE} \right)$	
$\text{LEFTOFSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan} \left(x(p_1^{(1)}) - x(p_2^{(1)}), -\Delta\text{DISTSMALL} \right)$	
$\text{LEFTOFEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan} \left(x(p_1^{(T)}) - x(p_2^{(T)}), -\Delta\text{DISTSMALL} \right)$	
$\text{RIGHTOFSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan} \left(x(p_1^{(1)}) - x(p_2^{(1)}), \Delta\text{DISTSMALL} \right)$	
$\text{RIGHTOFEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan} \left(x(p_1^{(T)}) - x(p_2^{(T)}), \Delta\text{DISTSMALL} \right)$	
$\text{ONTOPOFSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2)$	
$\quad + \text{distGreaterThan} \left(y(p_1^{(1)}) - y(p_2^{(1)}), -2\Delta\text{DISTLARGE} \right)$	
$\quad + \text{distLessThan} \left(y(p_1^{(1)}) - y(p_2^{(1)}), 0 \right)$	
$\quad + \text{distLessThan} \left(x(p_1^{(1)}) - x(p_2^{(1)}) , 2\Delta\text{DISTSMALL} \right)$	
$\text{ONTOPOFEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2)$	
$\quad + \text{distGreaterThan} \left(y(p_1^{(T)}) - y(p_2^{(T)}), -2\Delta\text{DISTLARGE} \right)$	
$\quad + \text{distLessThan} \left(y(p_1^{(T)}) - y(p_2^{(T)}), 0 \right)$	
$\quad + \text{distLessThan} \left(x(p_1^{(T)}) - x(p_2^{(T)}) , 2\Delta\text{DISTSMALL} \right)$	
$\text{NEARSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan} \left(\text{dist}(p_1^{(1)}, p_2^{(1)}), 2\Delta\text{DISTSMALL} \right)$	
$\text{NEAREND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan} \left(\text{dist}(p_1^{(T)}, p_2^{(T)}), 2\Delta\text{DISTSMALL} \right)$	
$\text{INSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{NEARSTART}(p_1, p_2) + \text{smaller}(p_1^{(1)}, p_2^{(1)})$	
$\text{INEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{NEAREND}(p_1, p_2) + \text{smaller}(p_1^{(T)}, p_2^{(T)})$	
$\text{BELOWSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan} \left(y(p_1^{(1)}) - y(p_2^{(1)}), \Delta\text{DISTSMALL} \right)$	
$\text{BELOWEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan} \left(y(p_1^{(T)}) - y(p_2^{(T)}), \Delta\text{DISTSMALL} \right)$	
$\text{ABOVESTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan} \left(y(p_1^{(1)}) - y(p_2^{(1)}), -\Delta\text{DISTSMALL} \right)$	
$\text{ABOVEEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan} \left(y(p_1^{(T)}) - y(p_2^{(T)}), -\Delta\text{DISTSMALL} \right)$	
$\text{OVER}(p_1, p_2) \hat{=} \text{tempCoher}(p_2)$	
$\quad + \max \left(\begin{array}{l} \text{distLessThan} \left(y(p_1^{(0)}) - y(p_2^{(0)}), -\Delta\text{DISTSMALL} \right) \\ \text{distLessThan} \left(x(p_1^{(0)}) - x(p_2^{(0)}) , \Delta\text{DISTLARGE} \right) \end{array} \right)$	

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

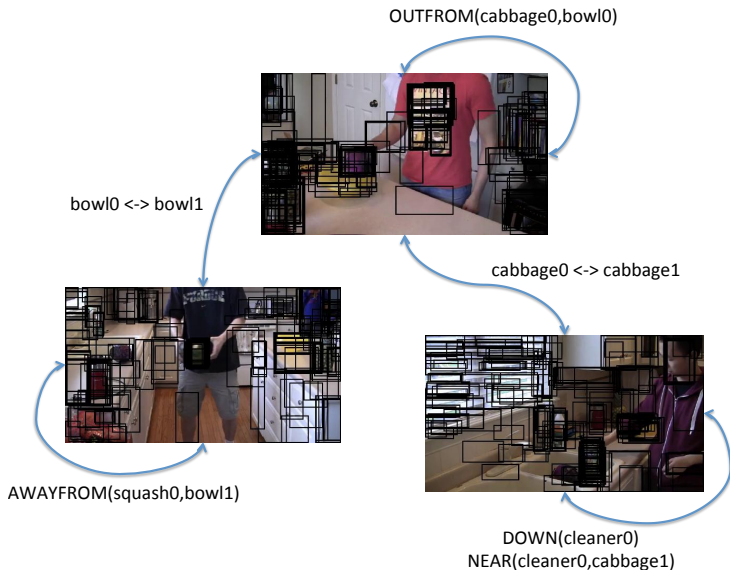
Method

- 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



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
$\text{MOVE}(p) \hat{=} \text{medFMg}(p)$	$\Delta\text{DISTLARGE} \hat{=} 0.25$
$\text{MOVEUP}(p) \hat{=} \text{MOVE}(p) + \text{distLessThan}(y(p^{(T)}) - y(p^{(1)}), -\Delta\text{DISTLARGE})$	$\Delta\text{DISTSMALL} \hat{=} 0.05$
$\text{MOVEDOWN}(p) \hat{=} \text{MOVE}(p) + \text{distGreaterThan}(y(p^{(T)}) - y(p^{(1)}), \Delta\text{DISTLARGE})$	$\Delta\text{ANGLE} \hat{=} \pi/2$
$\text{MOVEHORIZONTAL}(p) \hat{=} \text{MOVE}(p) + \text{distGreaterThan}(x(p^{(T)}) - x(p^{(1)}) , \Delta\text{DISTLARGE})$	
$\text{MOVELEFTWARDS}(p) \hat{=} \text{MOVE}(p) + \text{distLessThan}(x(p^{(T)}) - x(p^{(1)}), -\Delta\text{DISTLARGE})$	
$\text{MOVERIGHTWARDS}(p) \hat{=} \text{MOVE}(p) + \text{distGreaterThan}(x(p^{(T)}) - x(p^{(1)}), \Delta\text{DISTLARGE})$	
$\text{ROTATE}(p) \hat{=} \text{MOVE}(p) + \max \text{ hasRotation}(\text{rotAngle}(p^{(0)}), \Delta\text{ANGLE})$	
$\text{TOWARDS}(p_1, p_2) \hat{=} \text{MOVE}(p_1) + \text{distLessThan}(\text{dist}(p_1^{(T)}, p_2^{(T)}) - \text{dist}(p_1^{(1)}, p_2^{(1)}), -\Delta\text{DISTLARGE})$	
$\text{AWAYFROM}(p_1, p_2) \hat{=} \text{MOVE}(p_1) + \text{distGreaterThan}(\text{dist}(p_1^{(T)}, p_2^{(T)}) - \text{dist}(p_1^{(1)}, p_2^{(1)}), \Delta\text{DISTLARGE})$	
$\text{LEFTOFSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan}(x(p_1^{(1)}) - x(p_2^{(1)}), -\Delta\text{DISTSMALL})$	
$\text{LEFTOFEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan}(x(p_1^{(T)}) - x(p_2^{(T)}), -\Delta\text{DISTSMALL})$	
$\text{RIGHTOFSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan}(x(p_1^{(1)}) - x(p_2^{(1)}), \Delta\text{DISTSMALL})$	
$\text{RIGHTOFEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan}(x(p_1^{(T)}) - x(p_2^{(T)}), \Delta\text{DISTSMALL})$	
$\text{ONTOPOFSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2)$	
$\quad + \text{distGreaterThan}(y(p_1^{(1)}) - y(p_2^{(1)}), -2\Delta\text{DISTLARGE})$	
$\quad + \text{distLessThan}(y(p_1^{(1)}) - y(p_2^{(1)}), 0)$	
$\quad + \text{distLessThan}(x(p_1^{(1)}) - x(p_2^{(1)}) , 2\Delta\text{DISTSMALL})$	
$\text{ONTOPOFEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2)$	
$\quad + \text{distGreaterThan}(y(p_1^{(T)}) - y(p_2^{(T)}), -2\Delta\text{DISTLARGE})$	
$\quad + \text{distLessThan}(y(p_1^{(T)}) - y(p_2^{(T)}), 0)$	
$\quad + \text{distLessThan}(x(p_1^{(T)}) - x(p_2^{(T)}) , 2\Delta\text{DISTSMALL})$	
$\text{NEARSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan}(\text{dist}(p_1^{(1)}, p_2^{(1)}), 2\Delta\text{DISTSMALL})$	
$\text{NEAREND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan}(\text{dist}(p_1^{(T)}, p_2^{(T)}), 2\Delta\text{DISTSMALL})$	
$\text{INSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{NEARSTART}(p_1, p_2) + \text{smaller}(p_1^{(1)}, p_2^{(1)})$	
$\text{INEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{NEAREND}(p_1, p_2) + \text{smaller}(p_1^{(T)}, p_2^{(T)})$	
$\text{BELOWSTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan}(y(p_1^{(1)}) - y(p_2^{(1)}), \Delta\text{DISTSMALL})$	
$\text{BELOWEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distGreaterThan}(y(p_1^{(T)}) - y(p_2^{(T)}), \Delta\text{DISTSMALL})$	
$\text{ABOVESTART}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan}(y(p_1^{(1)}) - y(p_2^{(1)}), -\Delta\text{DISTSMALL})$	
$\text{ABOVEEND}(p_1, p_2) \hat{=} \text{tempCoher}(p_2) + \text{distLessThan}(y(p_1^{(T)}) - y(p_2^{(T)}), -\Delta\text{DISTSMALL})$	
$\text{OVER}(p_1, p_2) \hat{=} \text{tempCoher}(p_2)$	
$\quad + \max \left(\begin{array}{l} \text{distLessThan}(y(p_1^{(0)}) - y(p_2^{(0)}), -\Delta\text{DISTSMALL}) \\ \text{distLessThan}(x(p_1^{(0)}) - x(p_2^{(0)}) , \Delta\text{DISTLARGE}) \end{array} \right)$	

Proposals



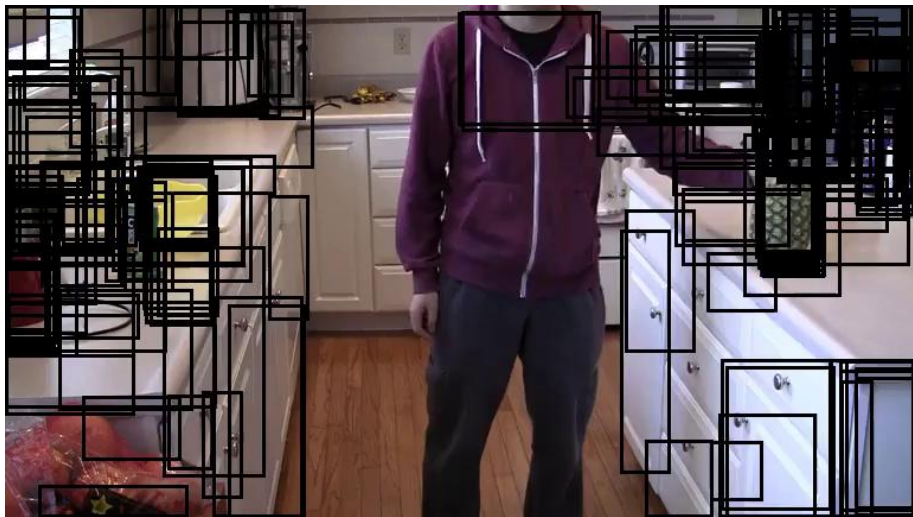
The person put the cleaner into the sink near the cabbage.

Proposals



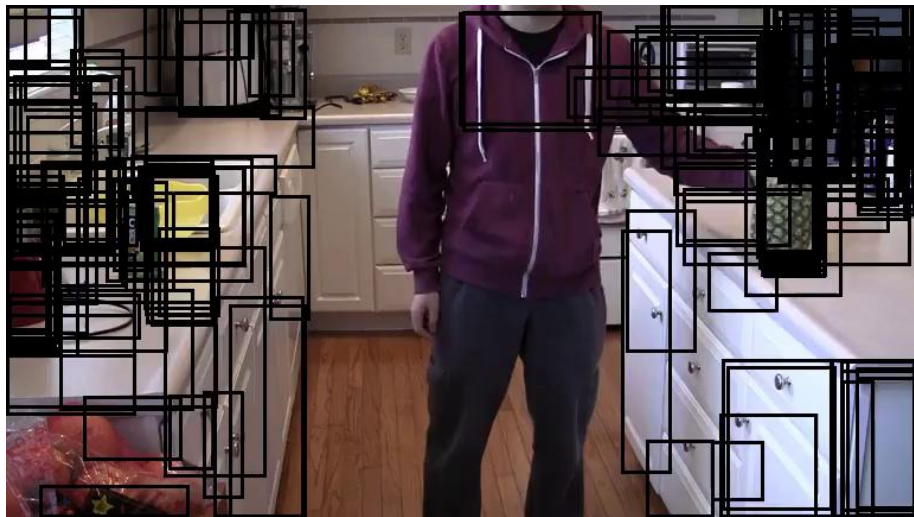
The person put the cleaner into the sink near the cabbage.

Proposals



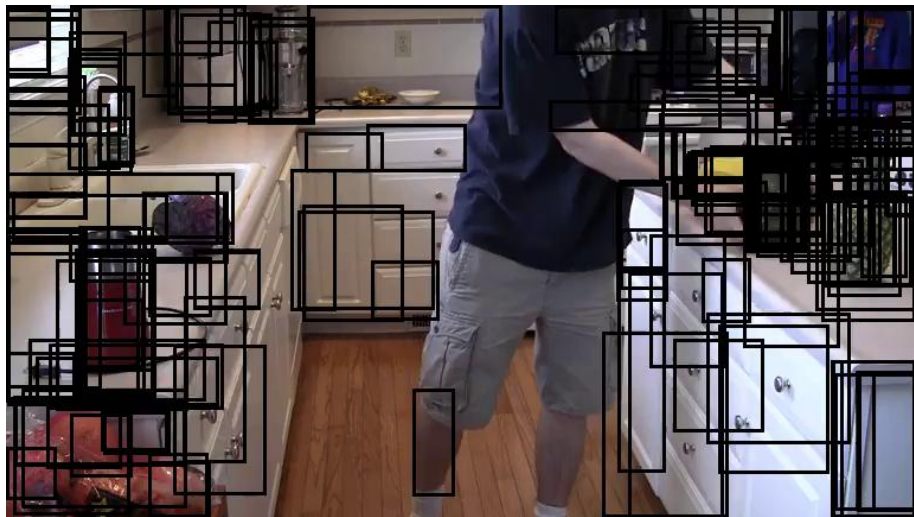
The person carried the pineapple towards the cleaner.

Proposals



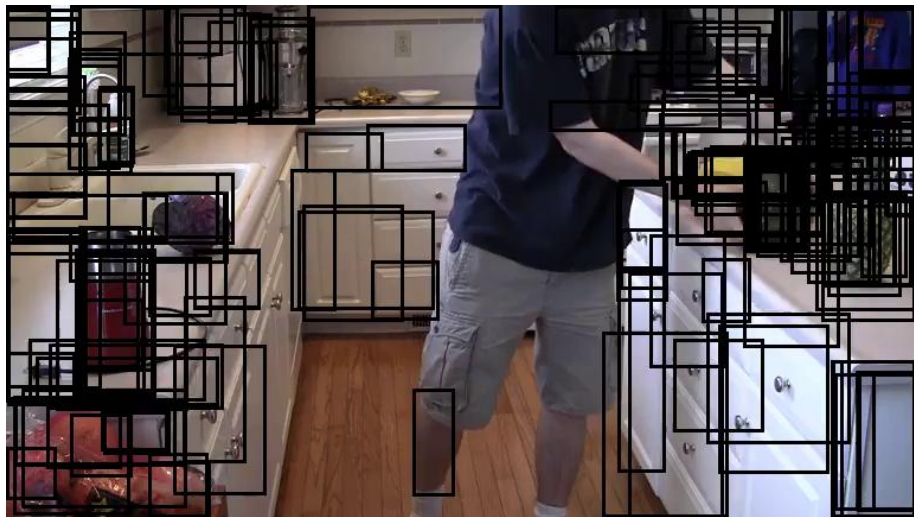
The person carried the pineapple towards the cleaner.

Proposals



The person took 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.



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



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



The person carried the pineapple towards the cleaner.



The person carried the pineapple towards the cleaner.



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



The person took 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.



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



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 took the squash away from the pineapple and put it near the coffee.

FLOW



The person took 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.



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



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



The person carried the pineapple towards the cleaner.



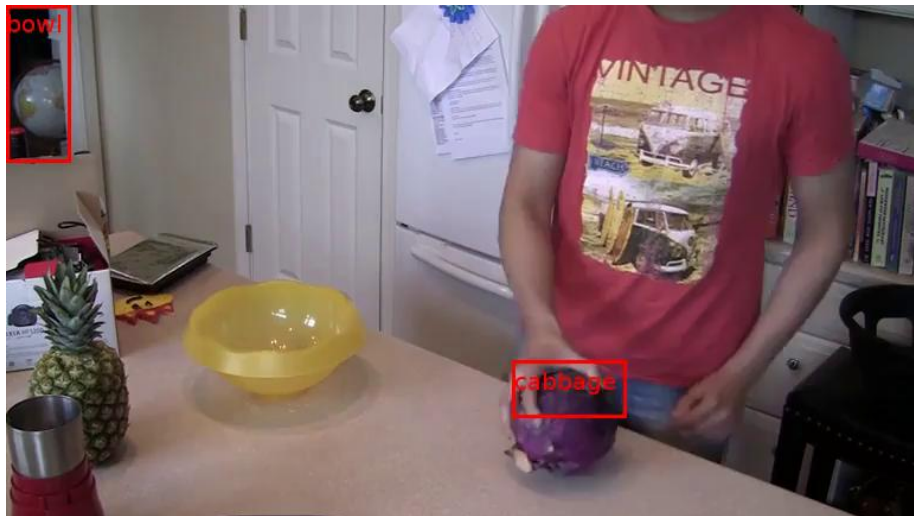
The person carried the pineapple towards the cleaner.



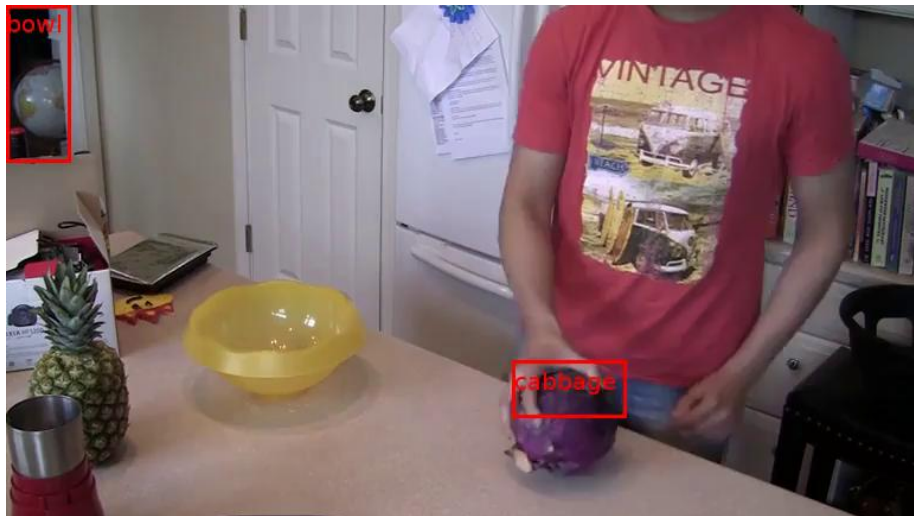
The person took the squash away from the pineapple and put it near the coffee.



The person took 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.



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



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



The person carried the pineapple towards the cleaner.



The person carried the pineapple towards the cleaner.



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



The person took 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.



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



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



The person carried the pineapple towards the cleaner.



The person carried the pineapple towards the cleaner.



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



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

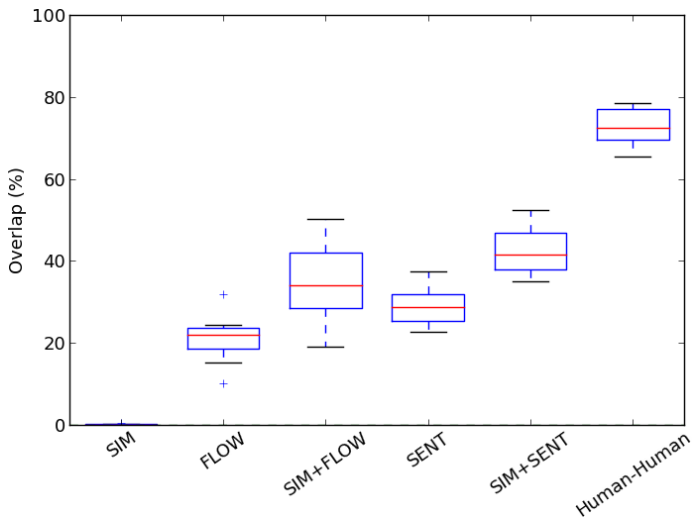


The person put the cabbage into the bowl.

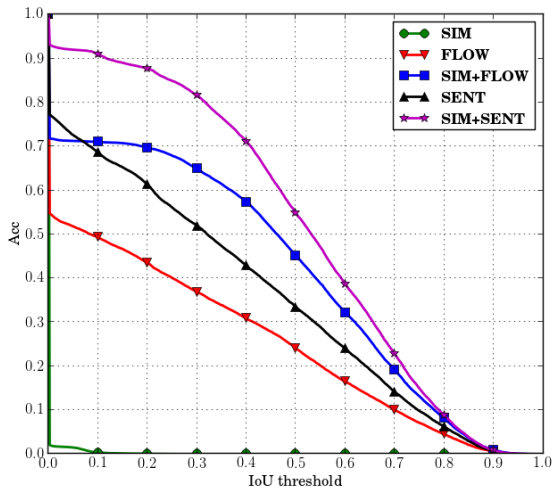


The person put the cabbage into the bowl.

IoU Scores



Codetection Accuracy



More Examples

the person put the bowl into the sink



More Examples

the person put the bowl into the sink



1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

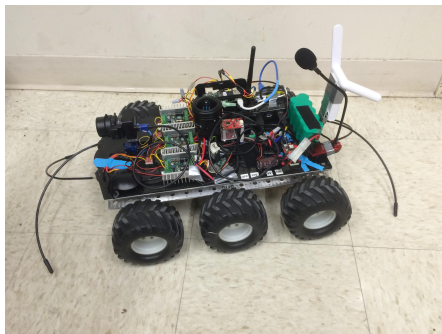
- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

5 Playing Checkers from English

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



1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- **Grounding Language Semantics in Robotics**
- Object Codetection from Mobile Robot Video

5 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

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

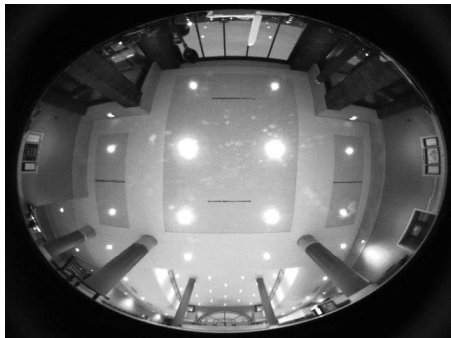
- ▶ **Language Generation:** path \times lexicon \rightarrow sentence

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

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

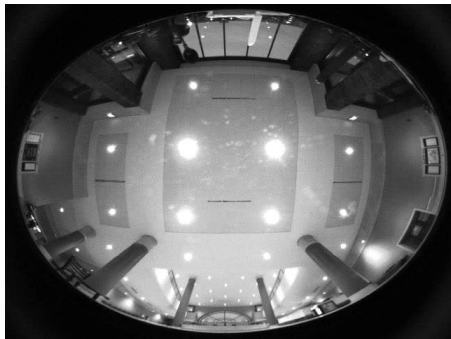
$$\arg \max_{\mathbf{p}} \mathcal{R}(\mathbf{s}, \mathbf{p}, \mathbf{f}, \Lambda)$$

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

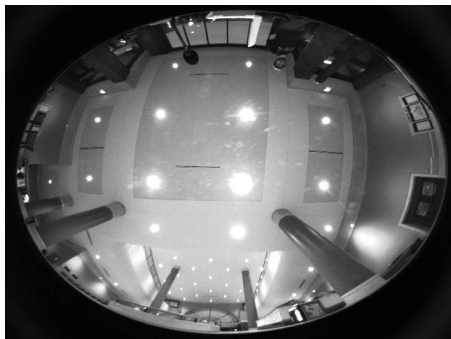
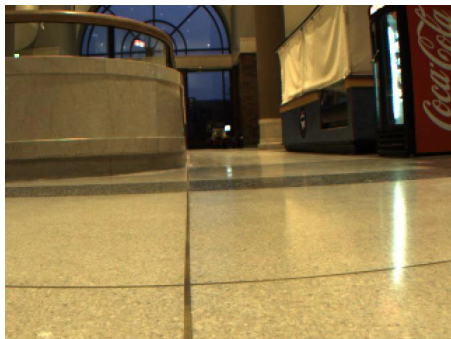


The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

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

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

Language Acquisition

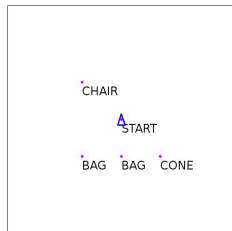


The robot went behind the cone and then turned around and went further behind the cone to the right of the chair.

Language Acquisition

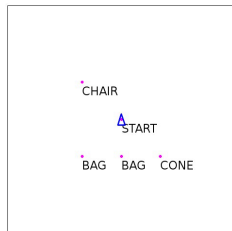
input:

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

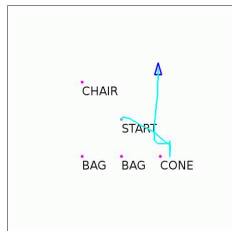
input:

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

Language Acquisition

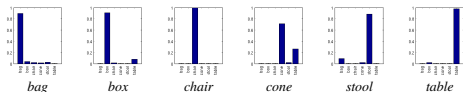
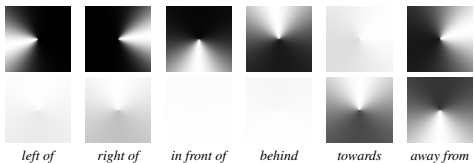
input:

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

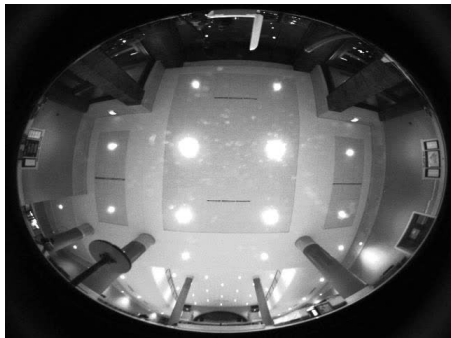


... plus 599 more

output:

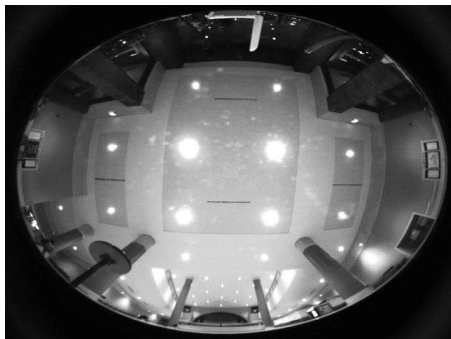


Language Generation



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.*

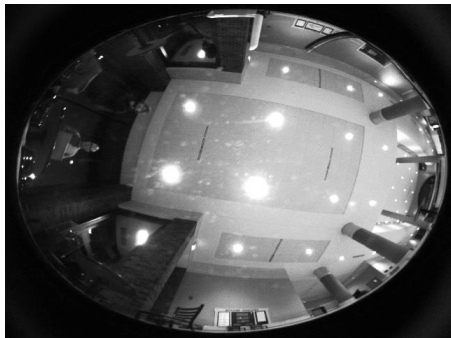
Language Generation

*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.***

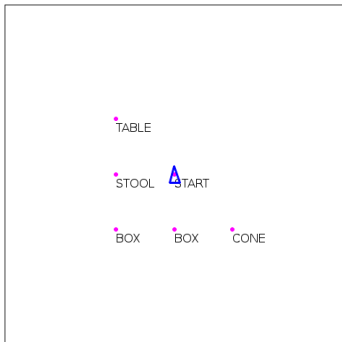
Language Generation



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:

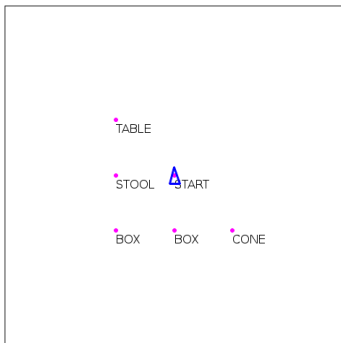


output:

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.

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.*

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.*

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.*

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.*

input:

*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.*

output:

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.*

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.*

input:

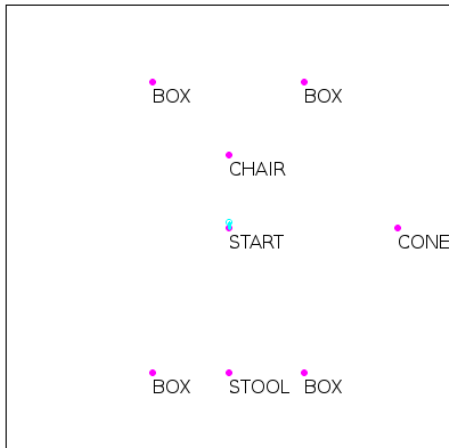
output: *The robot went behind the box which is right of the box then went right of the stool then went right of the box which is right of the box then went left of the cone then went in front of the cone then went away from the cone then went in front of the cone then went in front of the box which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

Language Comprehension

The robot went away from the cone then went behind the box

input: *which is right of the chair and which is behind the cone then went towards the stool.*

output:



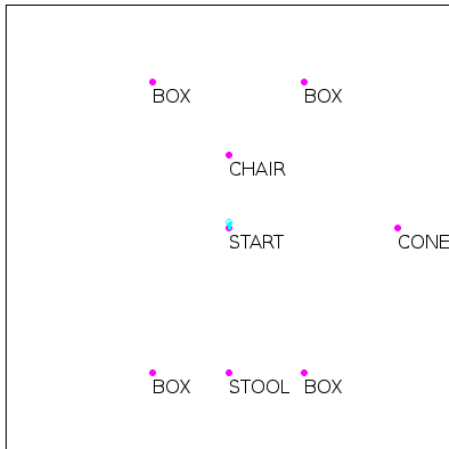
- ▶ Determine path waypoints that satisfy sentence
- ▶ Add intermediate points to avoid obstacles

Language Comprehension

The robot went away from the cone then went behind the box

input: *which is right of the chair and which is behind the cone then went towards the stool.*

output:



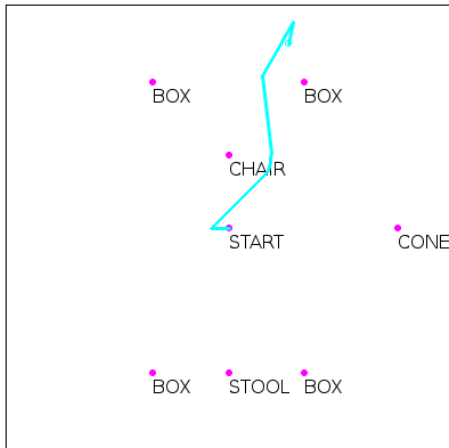
- ▶ Determine path waypoints that satisfy sentence
- ▶ Add intermediate points to avoid obstacles

Language Comprehension

The robot went away from the cone then went behind the box

input: *which is right of the chair and which is behind the cone then went towards the stool.*

output:



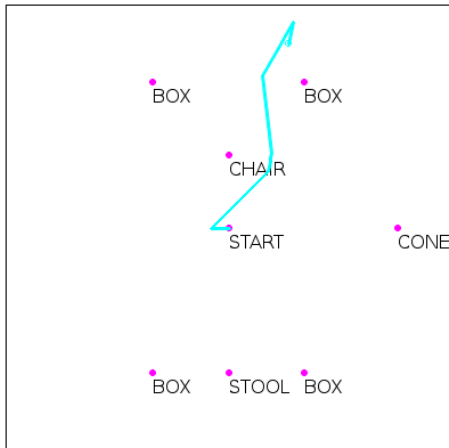
- ▶ Determine path waypoints that satisfy sentence
- ▶ Add intermediate points to avoid obstacles

Language Comprehension

The robot went away from the cone then went behind the box

input: *which is right of the chair and which is behind the cone then went towards the stool.*

output:

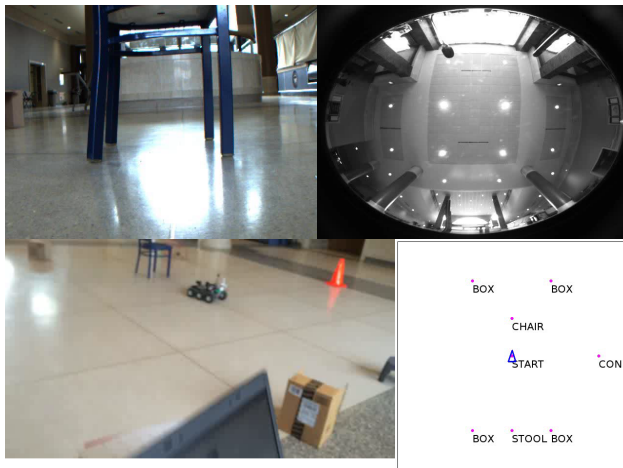


- ▶ Determine path waypoints that satisfy sentence
- ▶ Add intermediate points to avoid obstacles

Language Comprehension

The Effect of Different Prepositions (1)

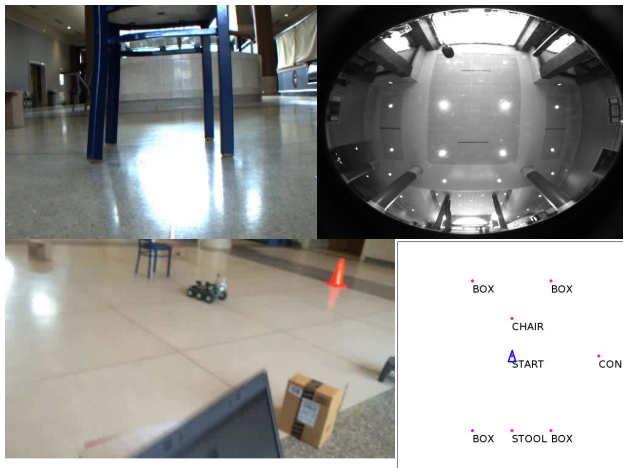
The robot went away from the cone then went behind the box which is right of the chair and which is behind the cone then went towards the stool.



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

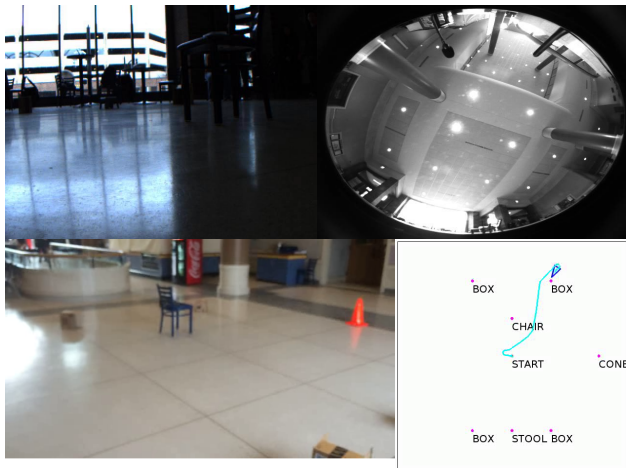
The Effect of Different Prepositions (1)

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)

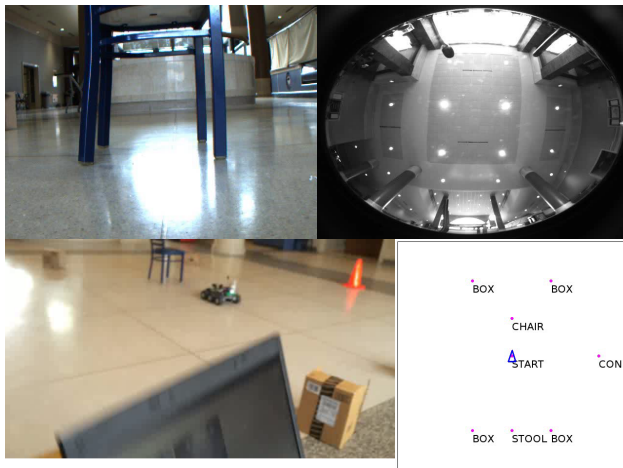
The robot went away from the cone then went behind the box which is right of the chair and which is behind the cone then went towards the stool.



Language Comprehension

The Effect of Different Prepositions (2)

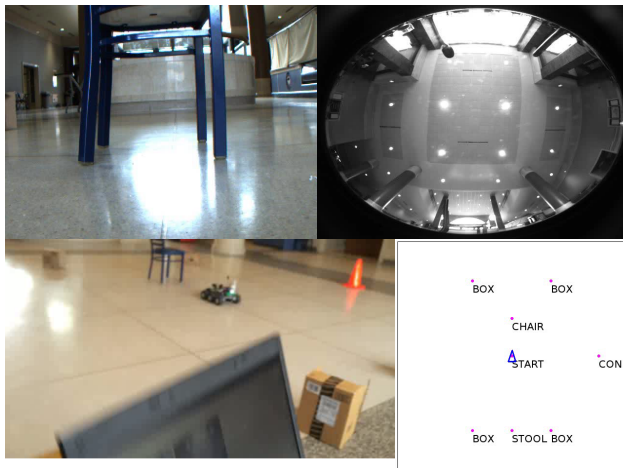
The robot went away from the cone then went behind the box which is right of the chair and which is in front of the cone then went towards the stool.



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

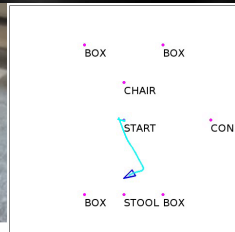
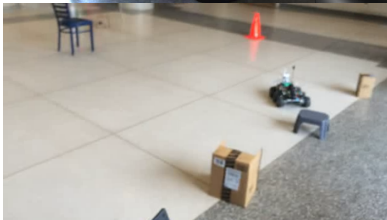
The Effect of Different Prepositions (2)

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)

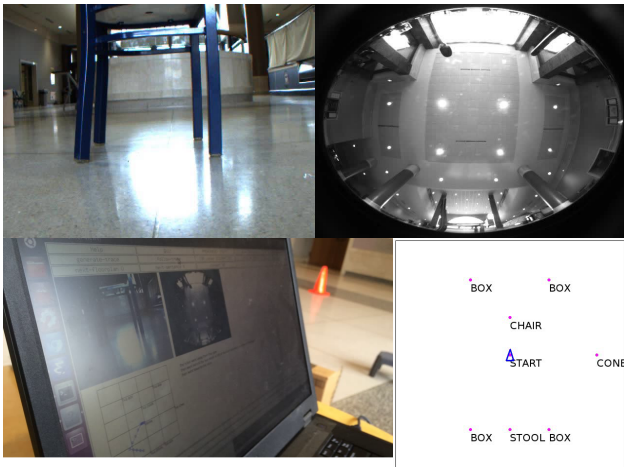
The robot went away from the cone then went behind the box which is right of the chair and which is in front of the cone then went towards the stool.



Language Comprehension

The Effect of Different Prepositions (3)

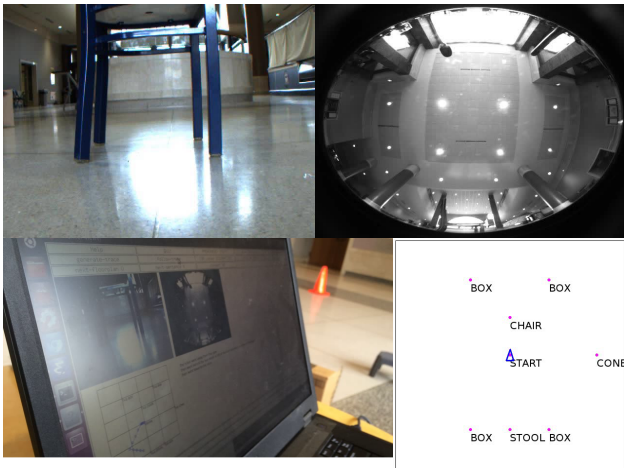
The robot went away from the cone then went behind the box which is left of the chair and which is in front of the cone then went towards the stool.



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

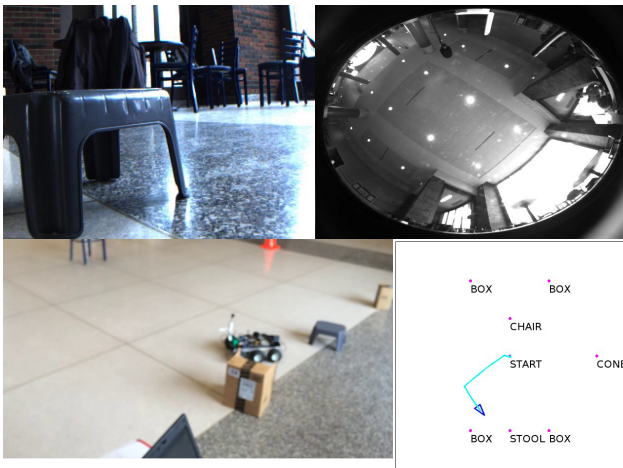
The Effect of Different Prepositions (3)

*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.

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

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

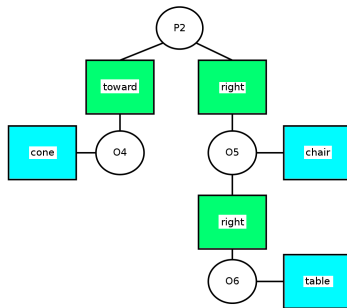
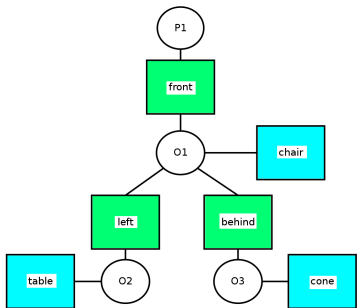
The robot went toward the left side of the stool, then toward the cone which is behind the stool, then toward the table which is left of the cone, then went back toward the stool and to the left of the stool.

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

- ▶ 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.



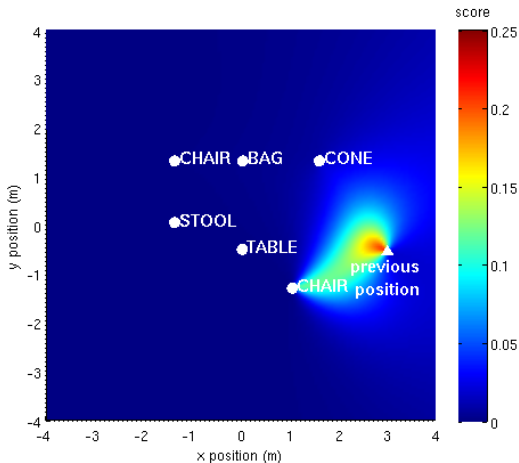
Semantics as a Soft Context-Sensitive Scoring Function

O4

O5

O6

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



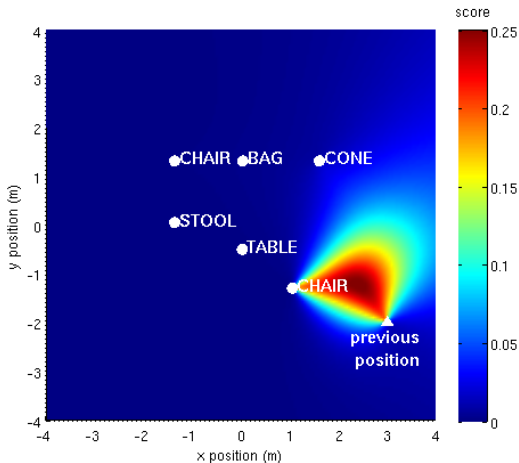
Semantics as a Soft Context-Sensitive Scoring Function

O4

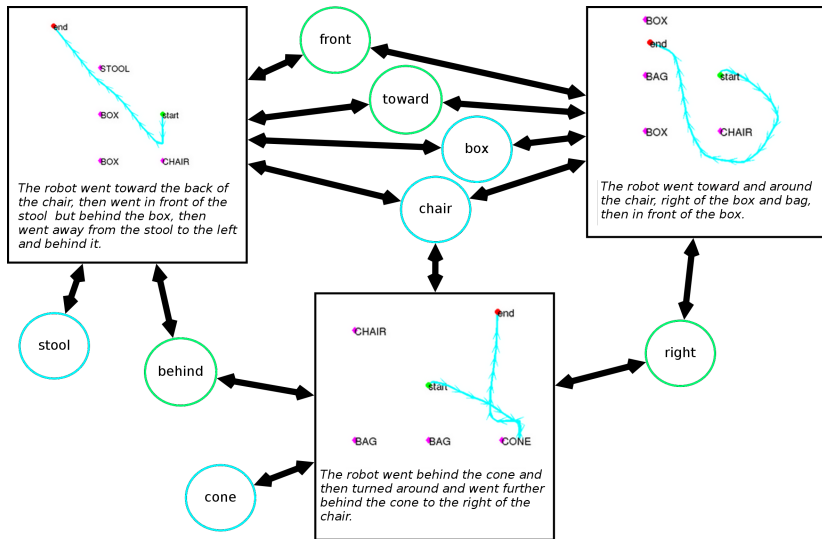
O5

O6

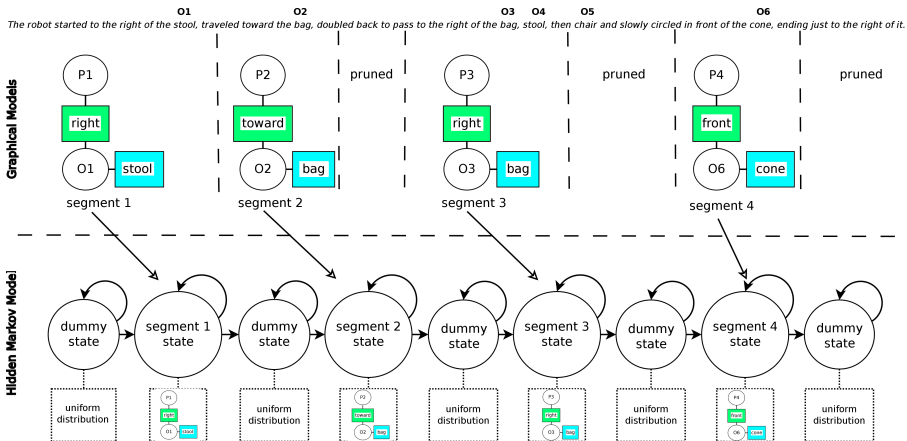
toward the cone and right of the chair right of the table.



Acquisition Method

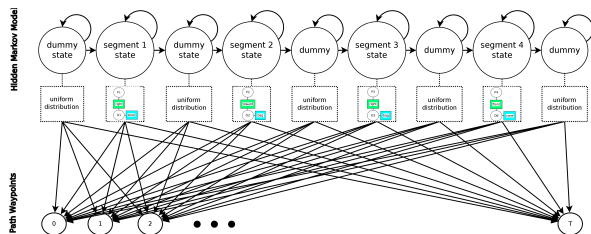
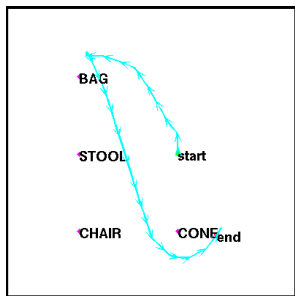


Acquisition Method



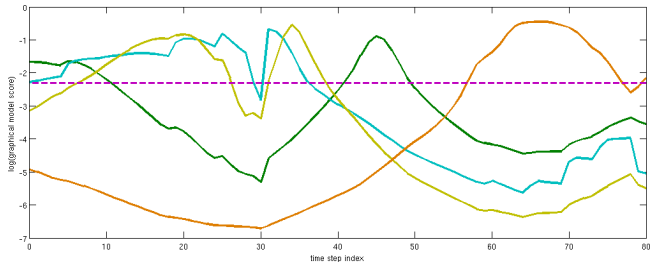
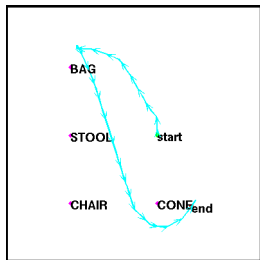
Acquisition Method

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.



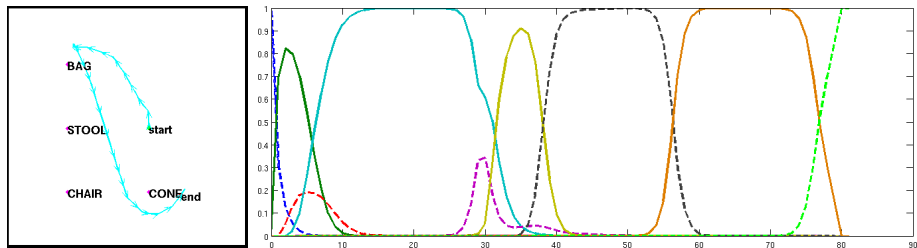
Acquisition Method

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

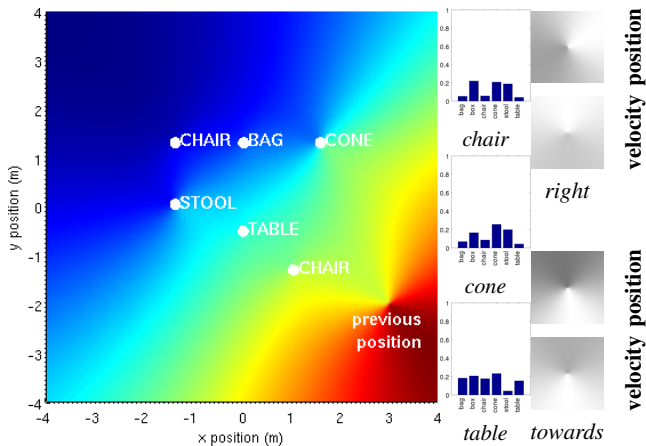
The robot started to the *right of the stool*, traveled *toward the bag*, doubled back to pass to the *right of the bag*, stool, then chair and slowly circled in *front of the cone*, ending just to the right of it.



Acquisition Method

Iteration 0

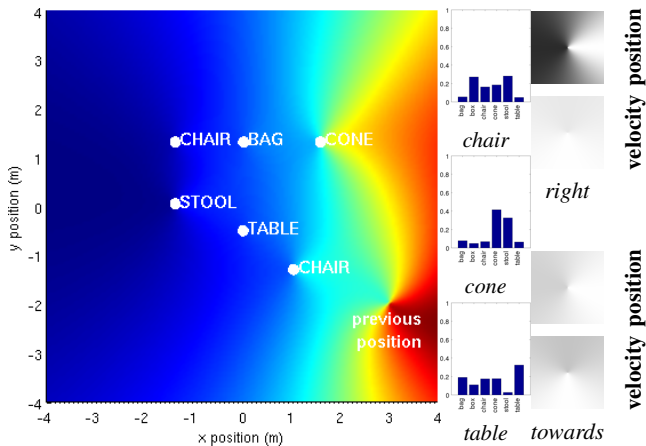
toward the cone and right of the chair right of the table



Acquisition Method

Iteration 1

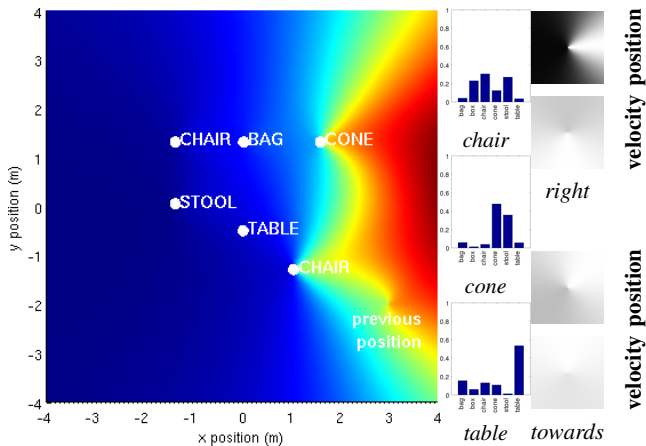
toward the cone and right of the chair right of the table



Acquisition Method

Iteration 2

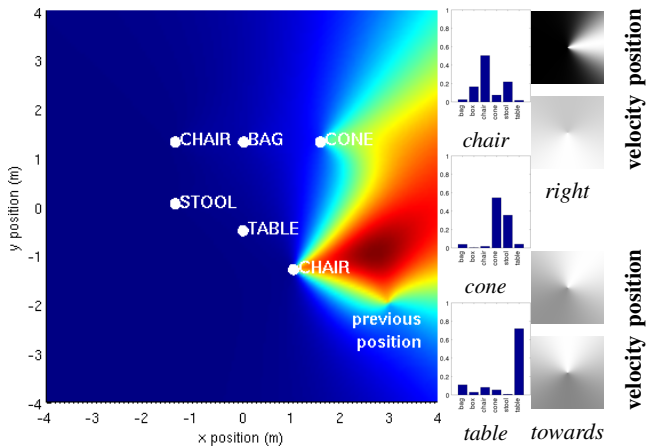
toward the cone and right of the chair right of the table



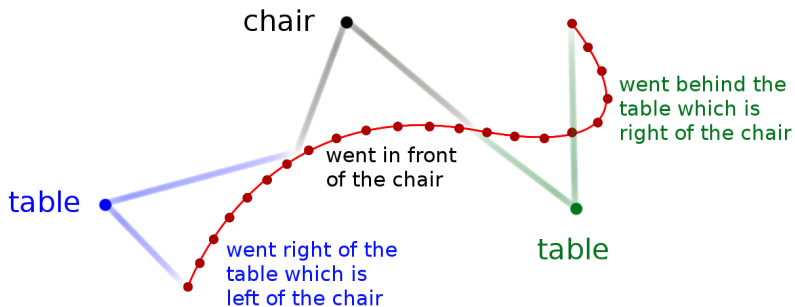
Acquisition Method

Iteration 3

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



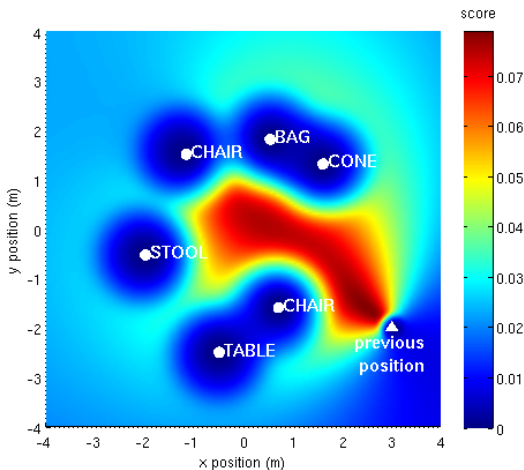
Generation Method



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



- 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**

- 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**

- 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

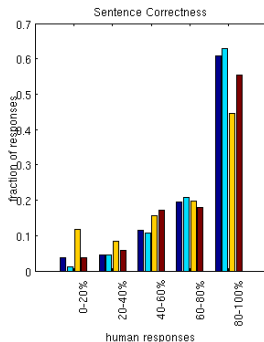
■ comprehension (human sentences vs. machine-driven paths)
tests machine performance because testing machine production of **paths** after human production of **sentences**

- 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

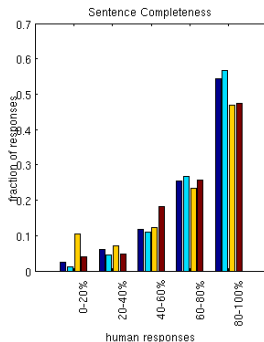
Approximately how much of the sentence is true of the path?



- 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

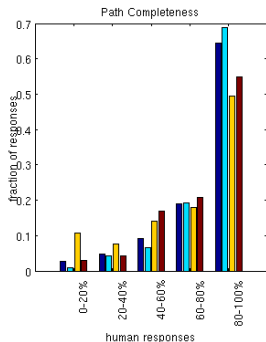
Approximately how much of the path is described by the sentence?



- 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)

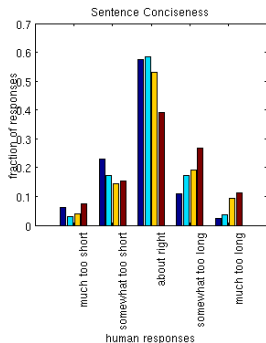
Path Completeness

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)

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 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

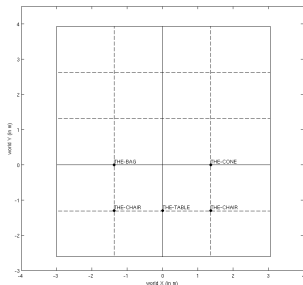
3 Sentence Directed Video Object Codetection

4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

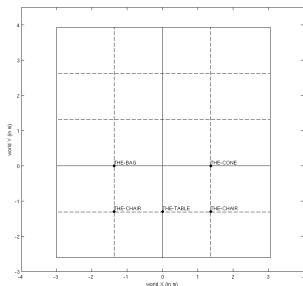
5 Playing Checkers from English

Floorplans



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

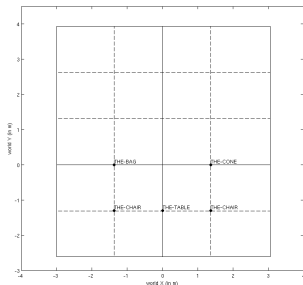
Floorplans



- ▶ 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.

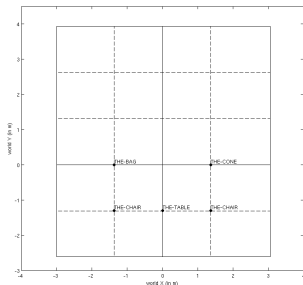
$\{(5, -3) : \mathbf{foo}, (-7, 3) : \mathbf{bar}\}$

Floorplans



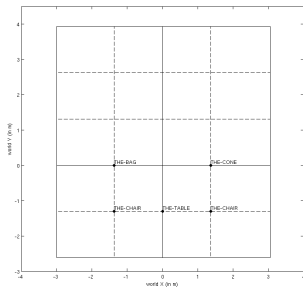
- ▶ The class labeling across different floorplans must be consistent; two instances of the same object class (in the same or different floorplans) should have the same label. This is what allows acquisition to work.

Floorplans



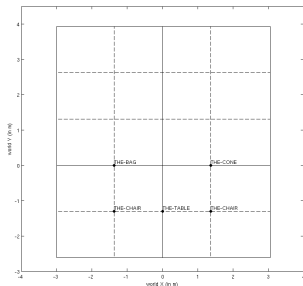
- ▶ The class labeling across different floorplans must be consistent; two instances of the same object class (in the same or different floorplans) should have the same label. This is what allows acquisition to work.
- ▶ The mapping from nouns to abstract class labels is *learned* (by the acquisition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

Floorplans



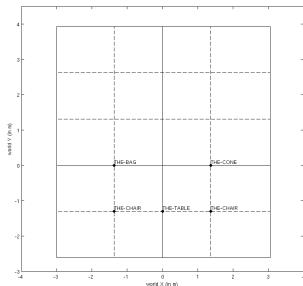
- It need not be a bijection.

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

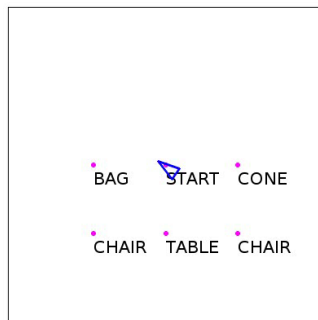
Data Collection

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



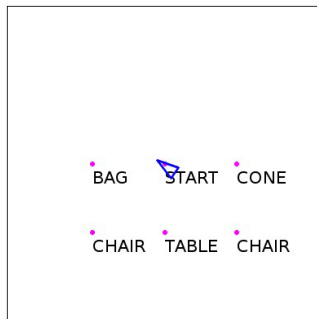
Data Collection

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

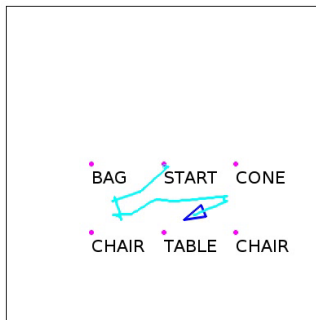
Data Collection

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:

Codetection approach is a five-step process:

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

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.

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.

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.

Codetection approach is a five-step process:

- ➊ **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
- ➋ **Proposal Localization** Find 3D world location for each proposal box.
- ➌ **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.
- ➍ **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.
- ➎ **Labeling** Assign an abstract class label to each localized object instance.

Codetection approach is a five-step process:

- ➊ **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
- ➋ **Proposal Localization** Find 3D world location for each proposal box.
- ➌ **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.
- ➍ **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.
- ➎ **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.

Codetection approach is a five-step process:

- ➊ **Proposal Generation** Generate a set of 2D proposal boxes for each frame.
- ➋ **Proposal Localization** Find 3D world location for each proposal box.
- ➌ **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.
- ➍ **Clustering** Cluster locations of selected proposals to find object locations on the floor plan.
- ➎ **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

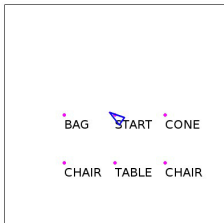
- 1 Generate proposals with MCG (Arbelaez et al. 2014)
- 2 Graphical model
 - ▶ vertex for each frame to denote the most prominent object in that frame
 - ▶ vertex labels range over proposals plus dummy
 - ▶ complete graph except no self edges
 - ▶ proposal score as vertex score, penalized by implausibility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

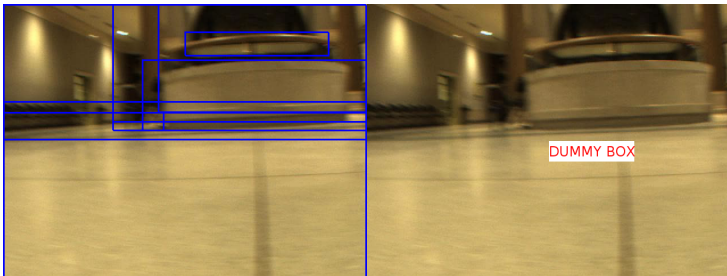
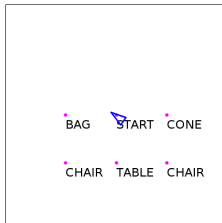
- 1 Generate proposals with MCG (Arbelaez et al. 2014)
- 2 Graphical model
 - ▶ vertex for each frame to denote the most prominent object in that frame
 - ▶ vertex labels range over proposals plus dummy
 - ▶ complete graph except no self edges
 - ▶ proposal score as vertex score, penalized by implausibility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

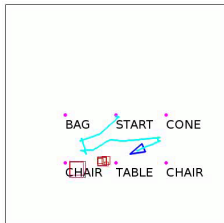


selected proposal (one per frame)

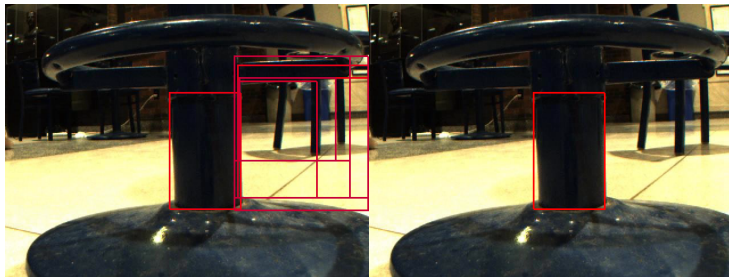
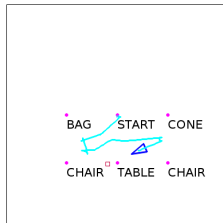


Proposal Generation, Selection, and Localization

raw proposals

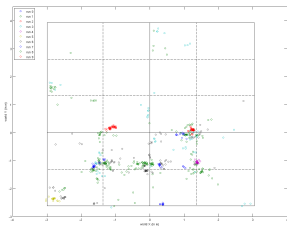


selected proposal (one per frame)



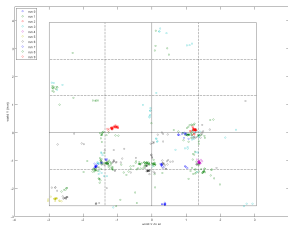
Clustering

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



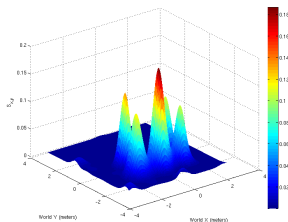
Clustering

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



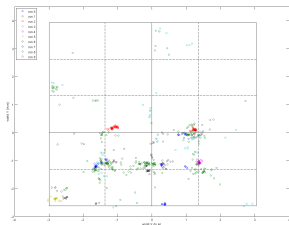
Compute density function

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



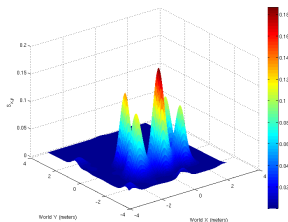
Clustering

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

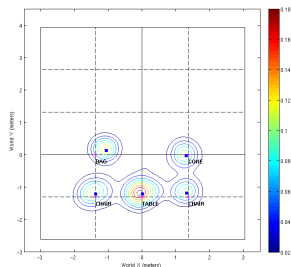


Compute density function

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



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

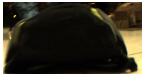

















- 1 Assign proposals to closest peak, rejecting outliers.
- 2 Let C_i denote the proposals assigned to peak i .
- 3 Let U_{ab} denote the similarity between proposals a and b .
- 4 Compute similarity Q_{ij} between peaks i and j .

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

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

Outline

1 Verbs, Arguments, and Predication in the Human Brain

- Experiment 1: ME-Y2
- Experiment 2: hollywood2-text-speech
- Experiment 3: compositionality-noninterleaved

2 The Sentence Tracker

3 Sentence Directed Video Object Codetection

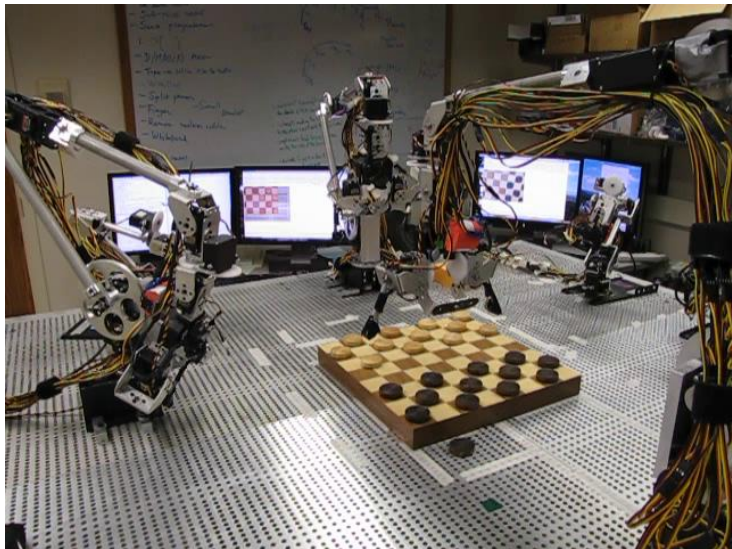
4 Driving Under the Influence (of Language)

- Grounding Language Semantics in Robotics
- Object Codetection from Mobile Robot Video

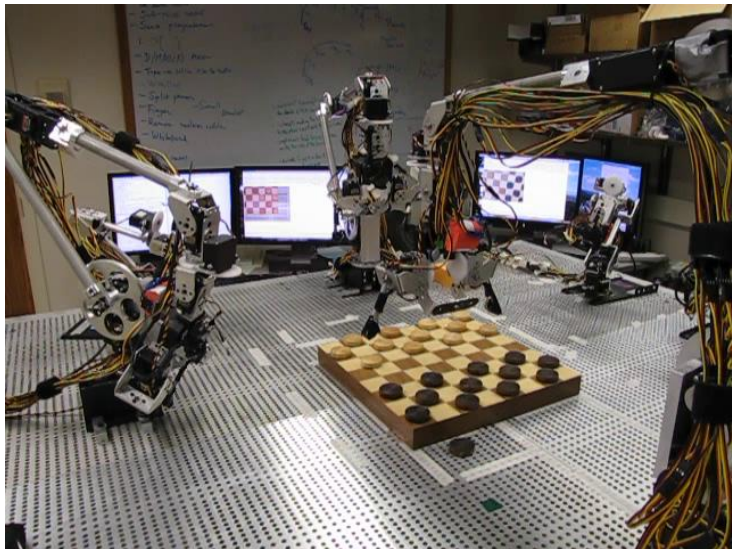
5 Playing Checkers from English

Daniel Paul Barrett Seth Benjamin Zachary Burchill

Two Robots Playing Checkers



Two Robots Playing Checkers



View from the Palm Camera

Help	Quit	Calibration-node	blur-size+ (11)	blur-size- (11)	servo to checker	find-lines?
TUGENHSIS	KADABBA	#AUSTRALOPITHECUS	blur-size+ (4.4)	blur-size- (4.4)	grasp!	line threshold+ (110)
Reload Robot	Load robot-dataset	Save Optimized Result	hough-resolution+ (2)	hough-resolution- (2)	prepare-fingers	line threshold- (110)
change current point	return to robot	Save robot-dataset	hough-min-distance+ (100)	hough-min-distance- (100)	servo to and grab	line min length+ (20)
change current robot	next-dataset-robot	next-point	edge-threshold+ (11)	edge-threshold- (11)	detect-ellipses	line min length- (20)
stream camera?	prev-robot	prev-point	circle-threshold+ (200)	circle-threshold- (200)	ease ellipse threshold (canny upper+ (30)
manual or camera points	view calibration images?	find-circles?	min-radius+ (60)	min-radius- (60)	ease ellipse threshold (canny upper- (30)
next-checker-fluclial	calibrate camera	first-person movement?	max-radius+ (500)	max-radius- (500)	pickup checker	canny lower+ (5)
	her camera calibration in				move-node	canny lower- (5)

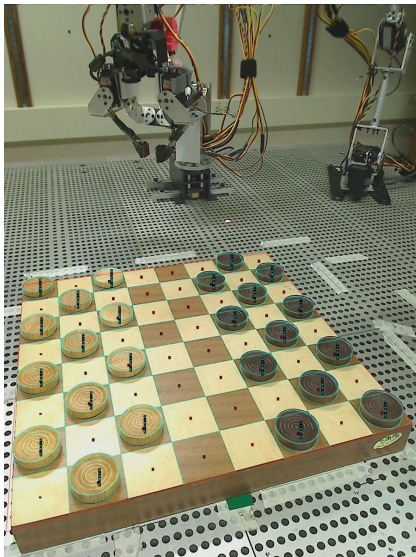
The image shows a top-down view of a chessboard. The board is divided into light and dark squares. Several pieces are visible, including pawns and knights. Two blue circles are drawn on the board, one on a dark square and one on a light square, likely indicating points of interest for the camera's vision system. The bottom of the image is partially obscured by a metal grate with a grid of holes.

View from the Palm Camera

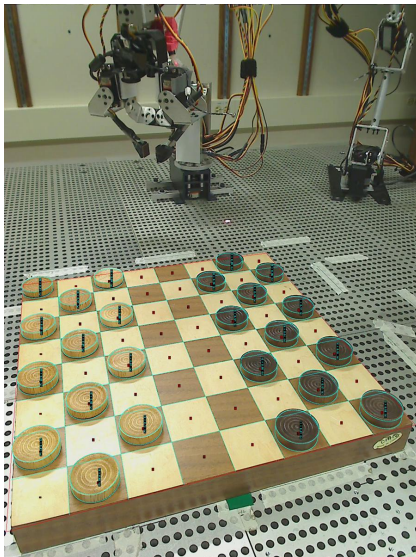
Help	Quit	Calibration-node	blur-size+ (11)	blur-size- (11)	servo to checker	find-lines?
TUGENHSIS	KADABBA	#AUSTRALOPITHECUS	blur-size+ (4.4)	blur-size- (4.4)	grasp!	line threshold+ (110)
Reload Robot	Load robot-dataset	Save Optimized Result	hough-resolution+ (2)	hough-resolution- (2)	prepare-fingers	line threshold- (110)
change current point	return to robot	Save robot-dataset	hough-min-distance+ (100)	hough-min-distance- (100)	servo to and grab	line min length+ (20)
change current robot	next-dataset-robot	next-point	edge-threshold+ (11)	edge-threshold- (11)	detect-ellipses	line min length- (20)
stream camera?	prev-robot	prev-point	circle-threshold+ (200)	circle-threshold- (200)	ease ellipse threshold (canny upper+ (30)
manual or camera points	view calibration images?	find-circles?	min-radius+ (60)	min-radius- (60)	ease ellipse threshold (canny upper- (30)
next-checker-fluclial	calibrate camera	first-person movement?	max-radius+ (500)	max-radius- (500)	pickup checker	canny lower+ (5)
	her camera calibration in				move-node	canny lower- (5)

The image shows a top-down view of a chessboard. The board is divided into light and dark squares. Several pieces are visible, including pawns and knights. In the foreground, there is a metal grate with a grid of circular holes. Two blue circles are drawn on the board, highlighting specific pieces or positions.

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

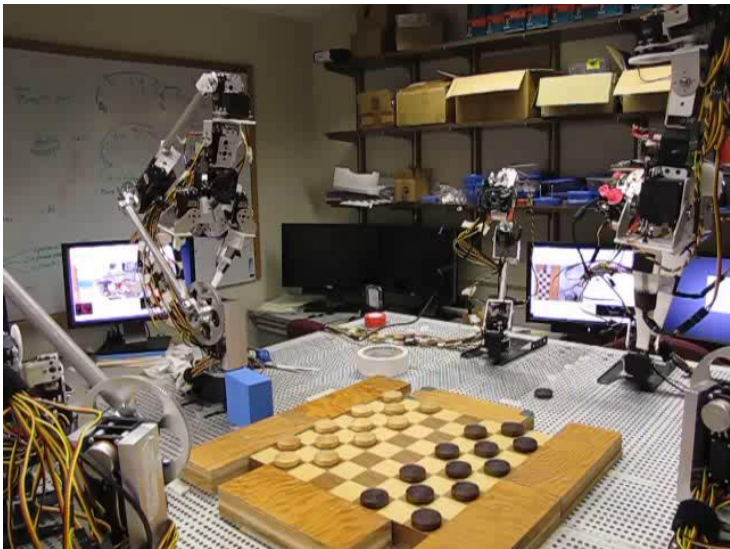
```
(move-type JUMP)
(KING-JUMP nw)
(KING-JUMP ne)
(KING-JUMP sw)
(KING-JUMP se)
```

```
)
```

```
(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

