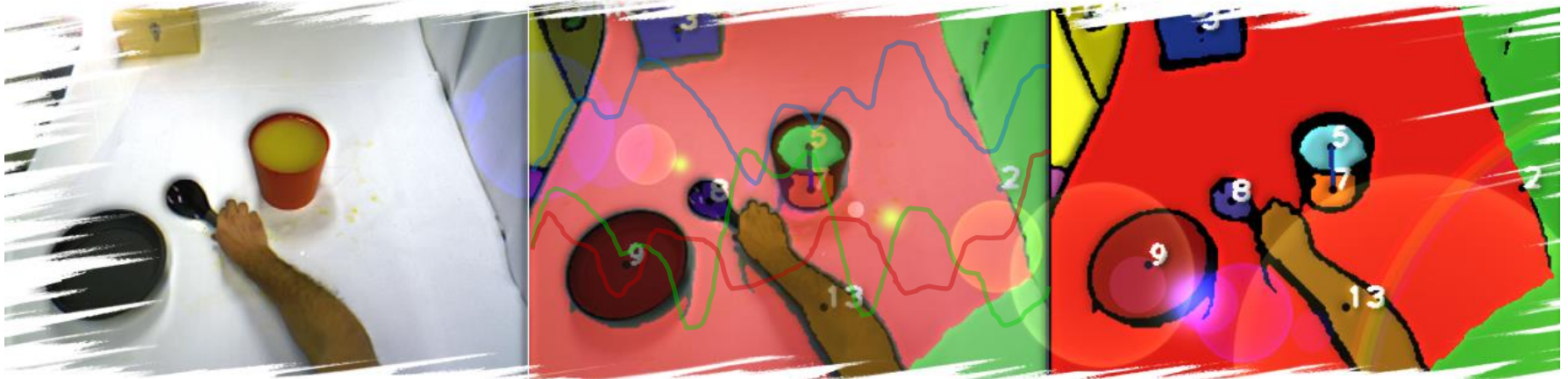


# Semantic Analysis of Manipulation Actions: Semantic Event Chains



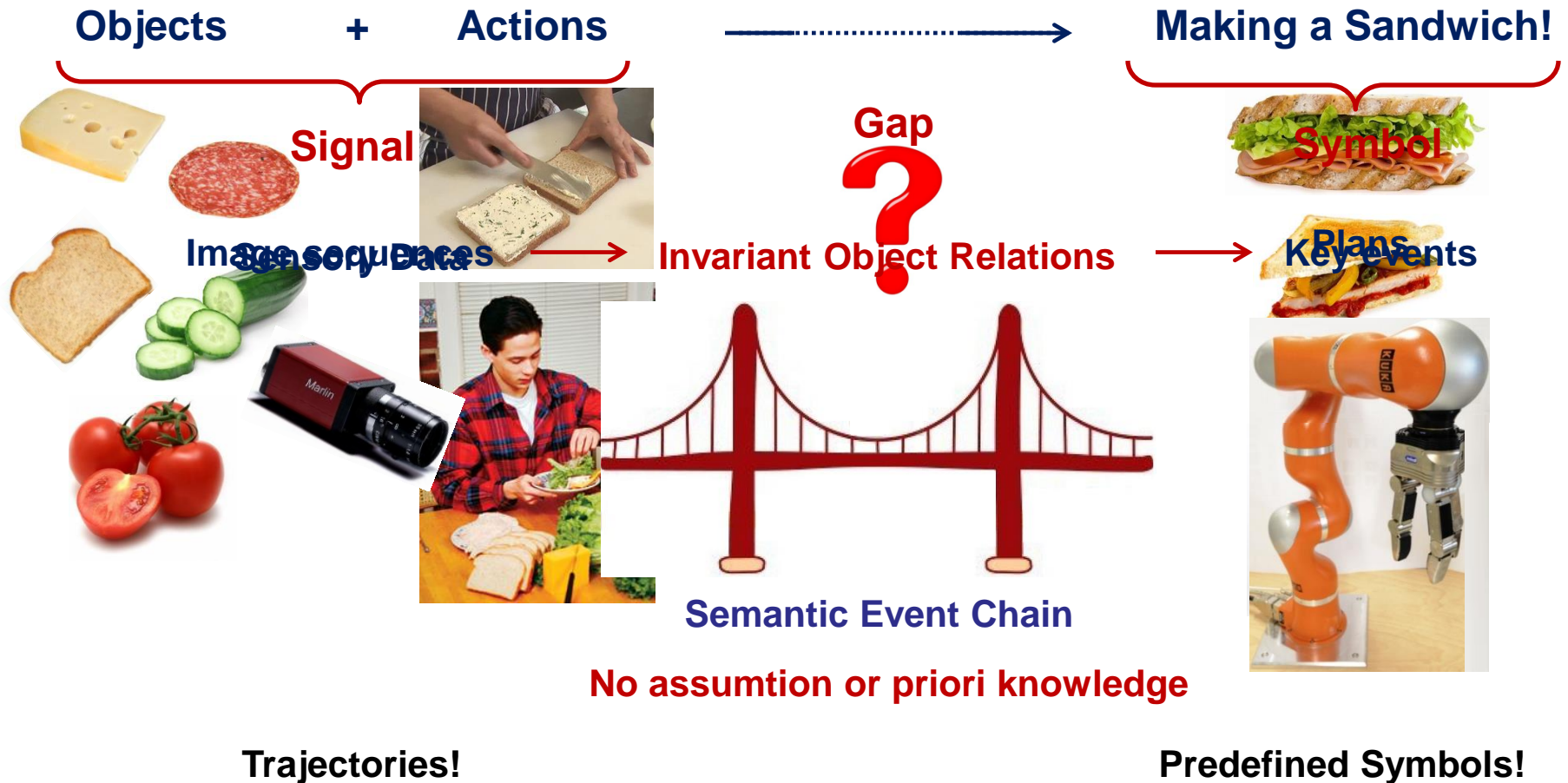
Eren Erdal Aksoy



Georg-August-Universität Göttingen

June 11, 2013

# Semantic Analysis? Meaning of Actions



---

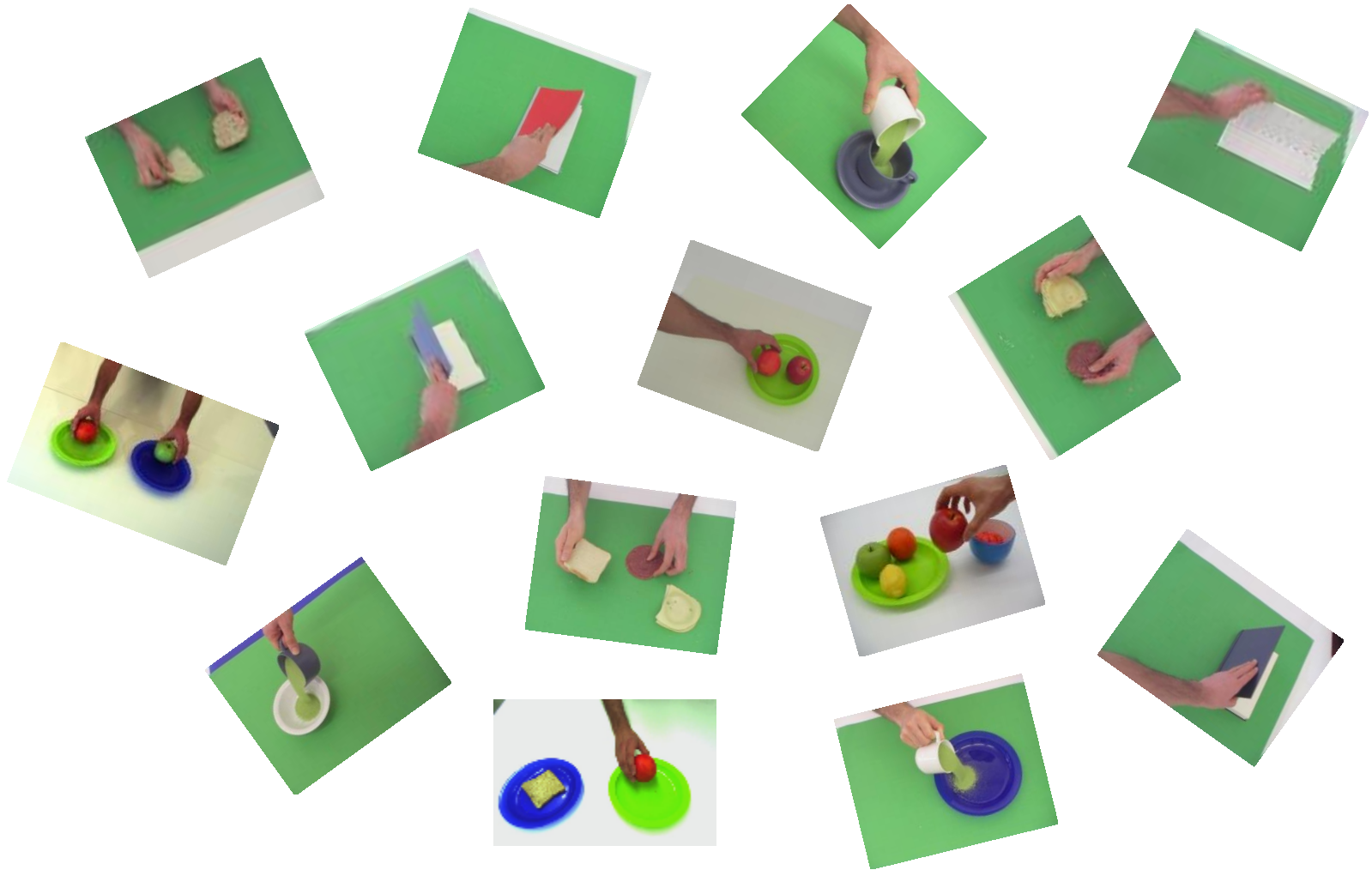
# Outline

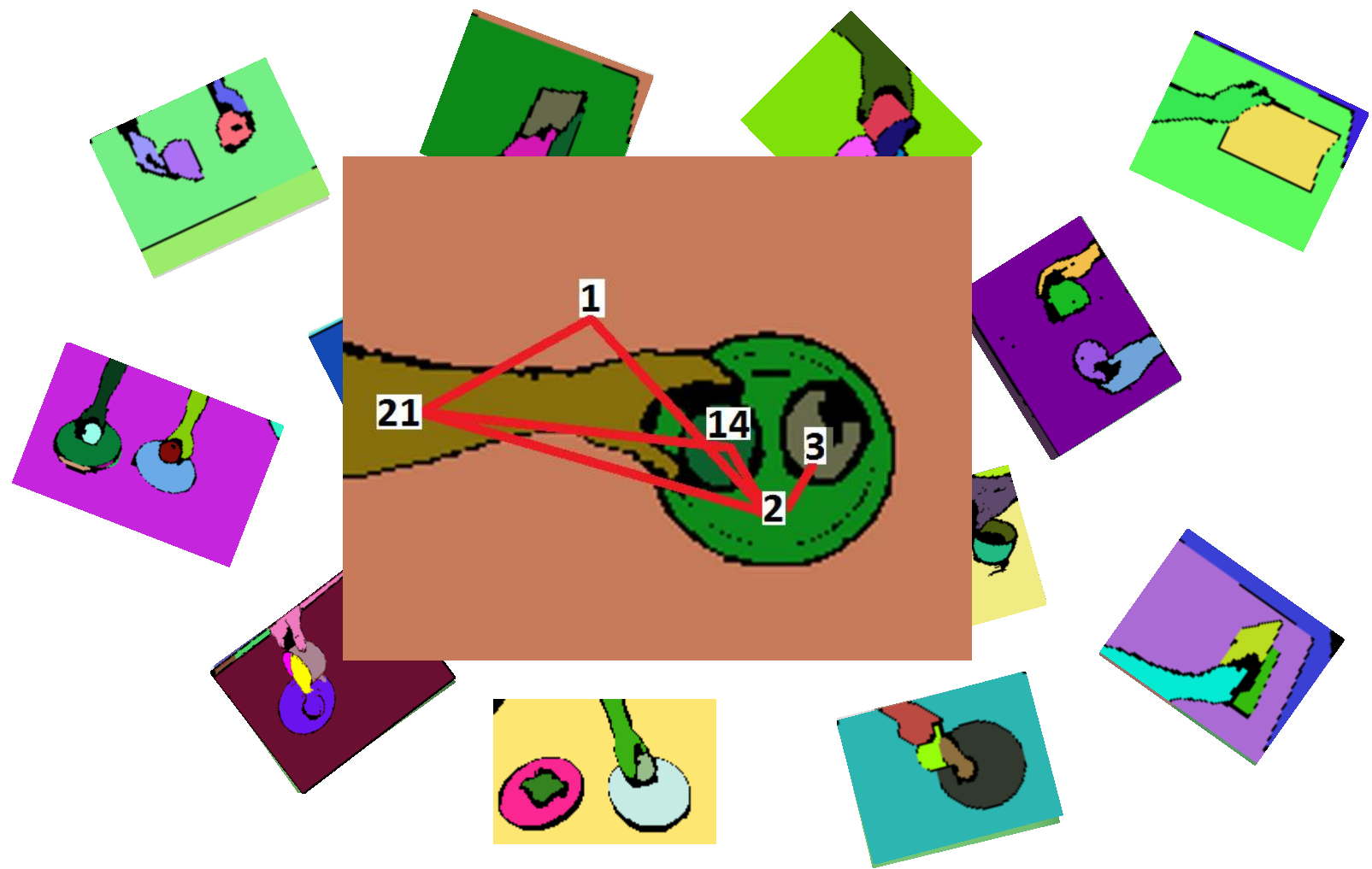
1. Introduction
2. Algorithm: SECs
3. Learning & Execution
4. Enriched SECs & Linguistics
5. Conclusion
6. Questions & Comments

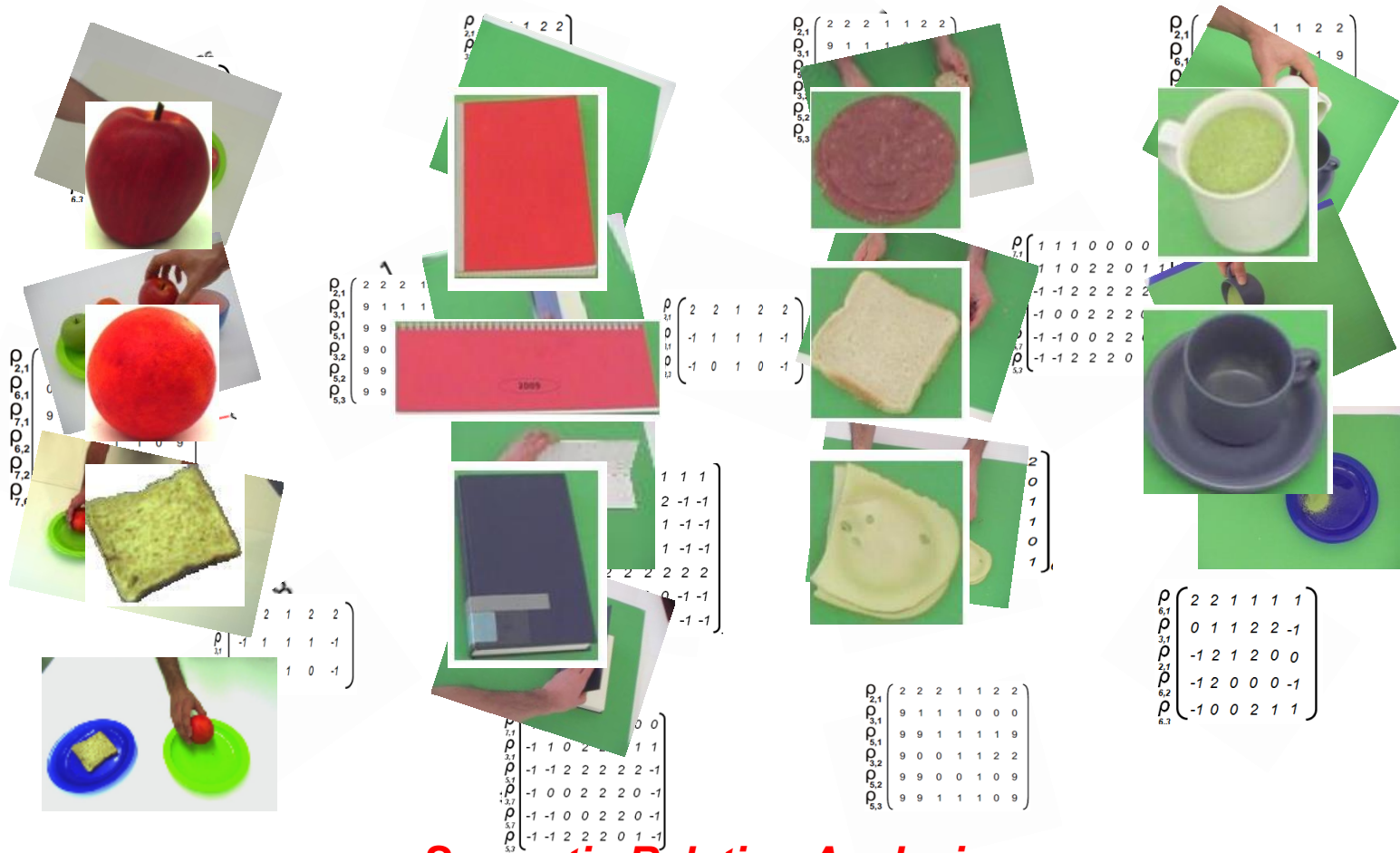
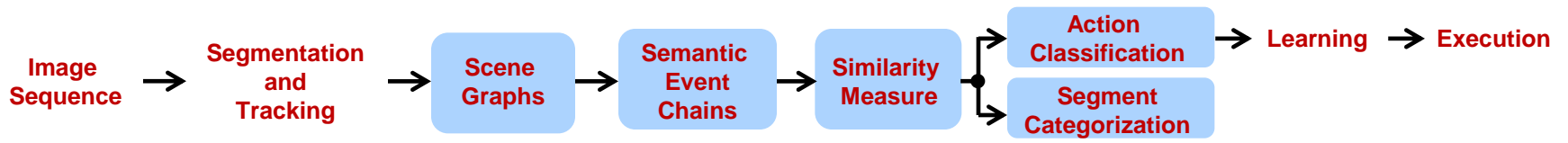
---

# Outline

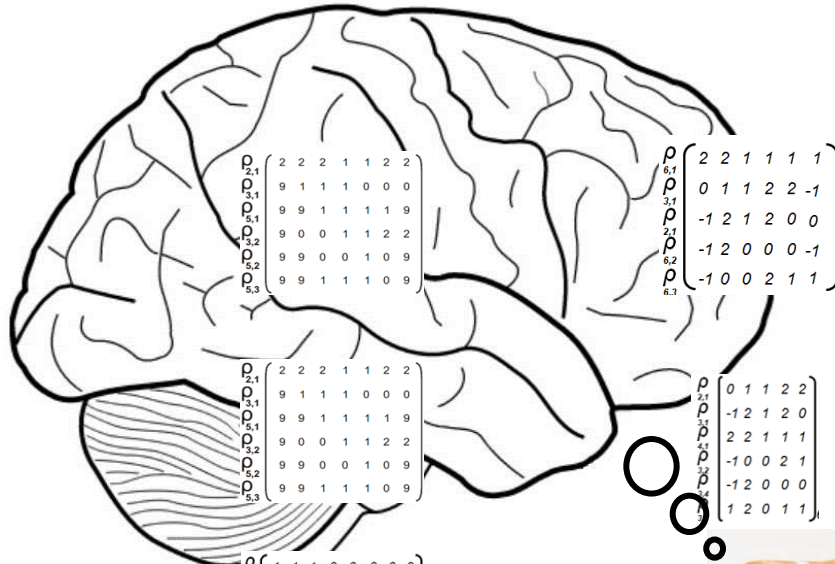
1. Introduction
2. Algorithm: SECs
3. Learning & Execution
4. Enriched SECs & Linguistics
5. Conclusion
6. Questions & Comments







## Semantic Relation Analysis



Defining a generic method for manipulations that can be used by agents to learn by observation not only to distinguish between different manipulations but also to categorize objects and to execute manipulations.

$$\rho_{1,1} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 2 & 2 & 0 & 1 & 1 \\ -1 & -1 & 2 & 2 & 2 & 2 & -1 & -1 \\ -1 & 0 & 0 & 2 & 2 & 2 & 0 & -1 \\ -1 & -1 & 0 & 0 & 2 & 2 & 0 & -1 \\ -1 & -1 & 2 & 2 & 2 & 0 & 1 & -1 \end{pmatrix}$$

$$\rho_{2,1} \begin{pmatrix} 2 & 2 & 2 & 1 & 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 9 \\ 9 & 1 & 1 & 1 & 1 & 1 & 1 & 9 \\ 2 & 2 & 1 & 1 & 1 & 1 & 0 & 9 \\ 9 & 0 & 1 & 0 & 0 & 0 & 0 & 9 \\ 9 & 0 & 1 & 1 & 1 & 1 & 1 & 9 \end{pmatrix}$$



Semantic Event Chain Models

$$\rho_{3,1} \begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}$$

$$\rho_{1,1} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 2 & 2 & 0 & 1 & 1 \\ -1 & -1 & 2 & 2 & 2 & 2 & -1 & -1 \\ -1 & 0 & 0 & 2 & 2 & 2 & 0 & -1 \\ -1 & -1 & 0 & 0 & 2 & 2 & 0 & -1 \\ -1 & -1 & 2 & 2 & 2 & 0 & 1 & -1 \end{pmatrix}$$





---

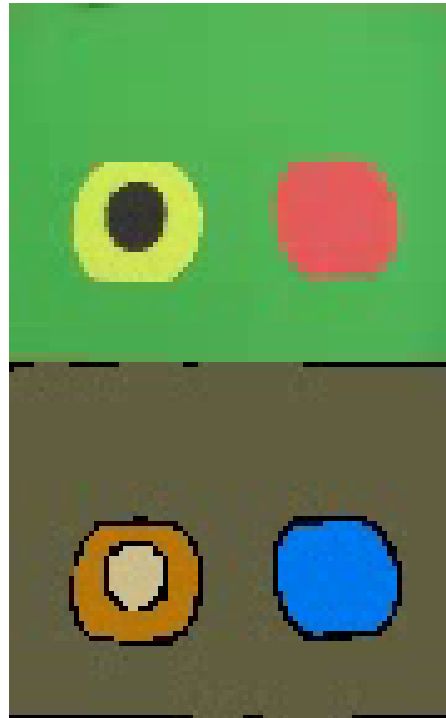
# Outline

1. Introduction
2. Algorithm: SECs
3. Learning & Execution
4. Enriched SECs & Linguistics
5. Conclusion
6. Questions & Comments

# Algorithm – Artificial Scenario

---

*Original  
Images*



*Segmentation  
&  
Tracking*

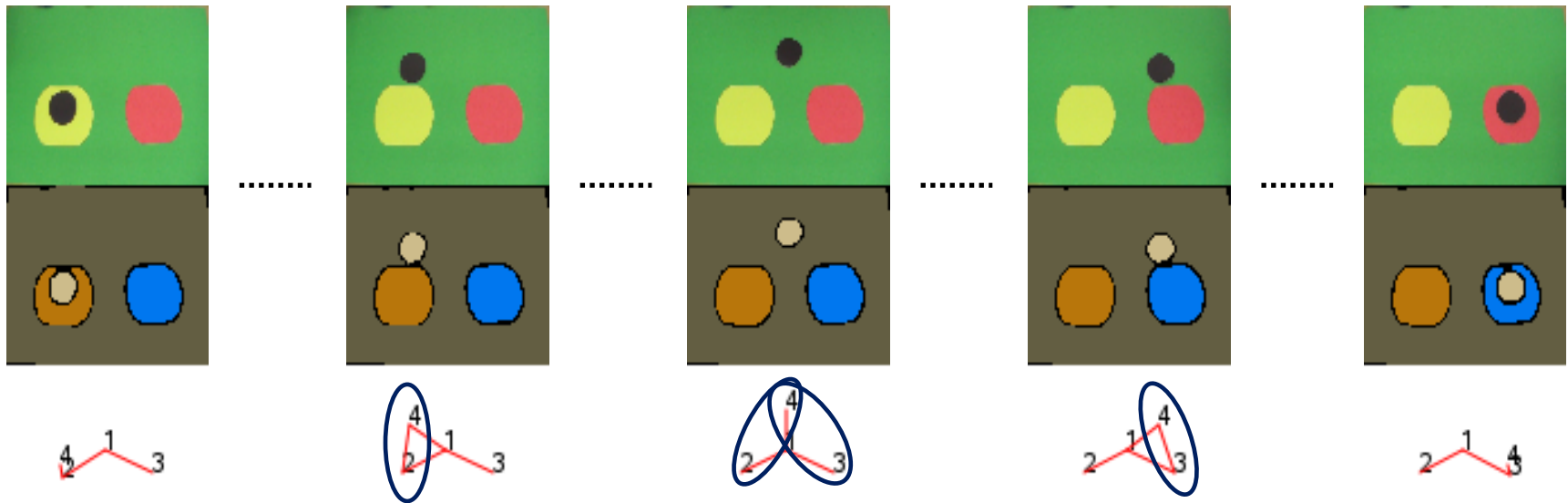


*2D Continuous  
Graphs*



*Aksoy et al. ICRA 2010  
Abramov et al. 3DPVT 2010  
Dellen et al. Sensors 2009*

# Algorithm – Artificial Scenario



-----> 1 : Background

The essence is to extract the Topological Changes on the linked graphs as these are indicative of a "Change"

-----> 2 : Full-to-Empty (yellow) Vessel

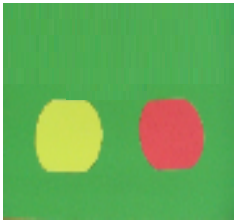
-----> 3 : Empty-to-Full (red) Vessel

-----> 4 : Content

2,1 " 3,1 4,1 3,2 4,2 4,3

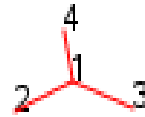
Object Relations

# Algorithm – 2D Object Relations


 $\rho_{4,2}$ 

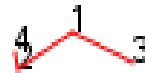

= Absence

= -1


 $\rho_{4,2}$ 


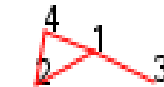
= No Connection

= 0 (no edges)


 $\rho_{4,2}$ 


= Overlapping

= 1 (one edge)

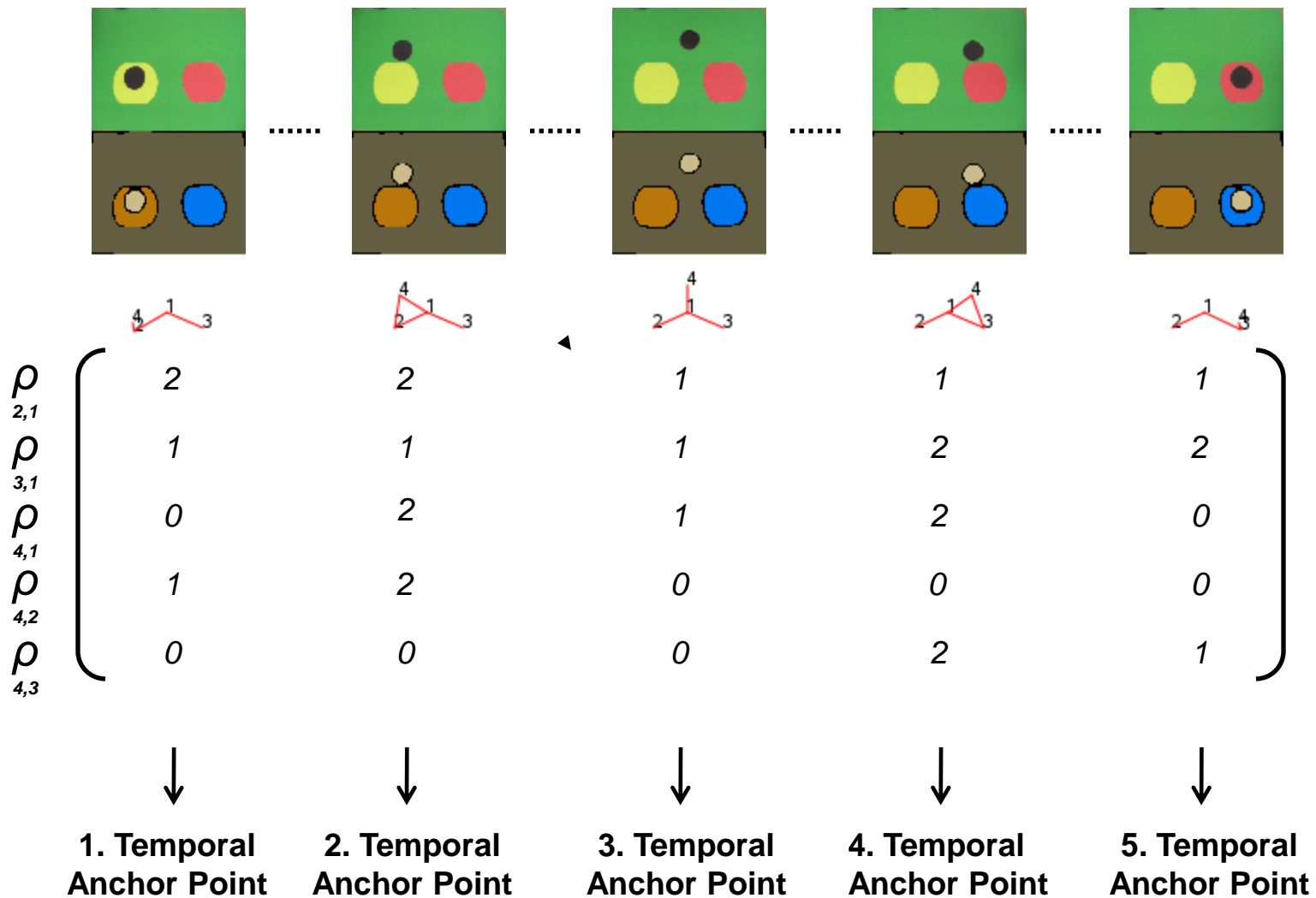

 $\rho_{4,2}$ 


= Touching

= 2 (two edges)

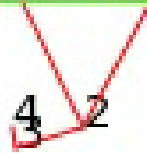
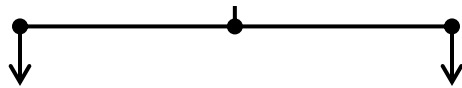
$\rho_{4,2}$  } Object No 2 : Full-to-Empty (yellow) Vessel  
 Object No 4 : Content

# Algorithm – Semantic Event Chain (SEC)

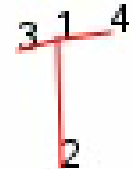
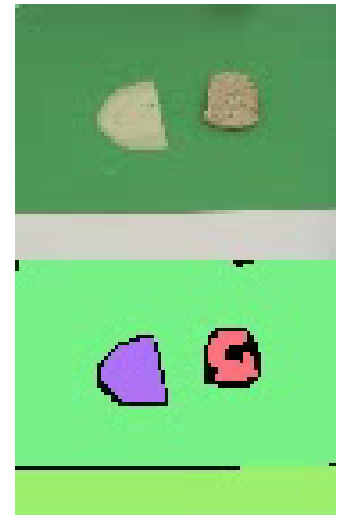
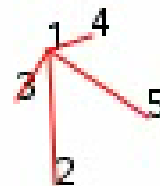
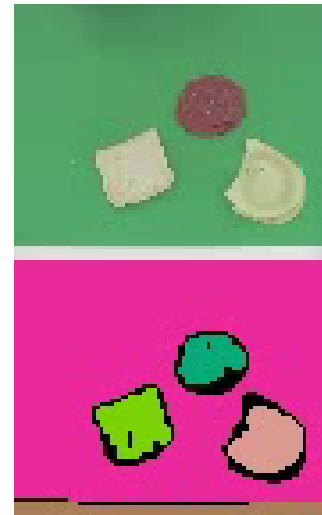
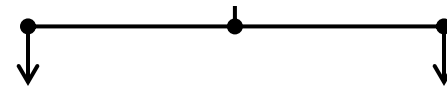


# Algorithm – Similarity Measure

*“Moving Object”*



*“Making Sandwich”*



# Algorithm – Similarity Measure

“Moving Object”



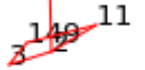
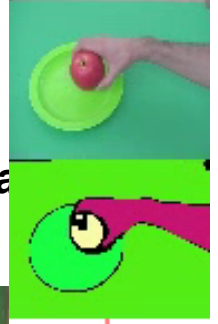
“Moving Object”

“Moving Object”

“Making Sandwich”

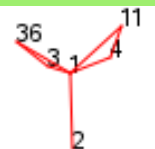
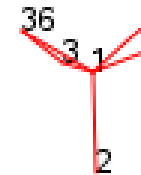
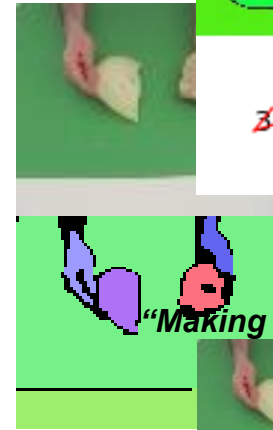
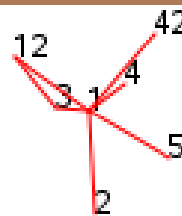
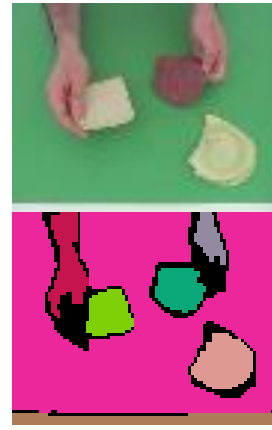
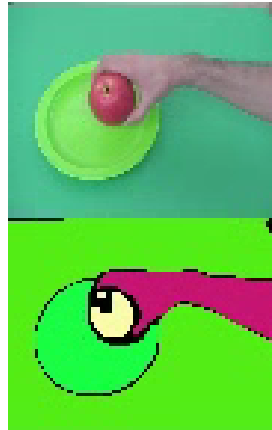
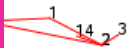
“Making Sa

“Moving Object”



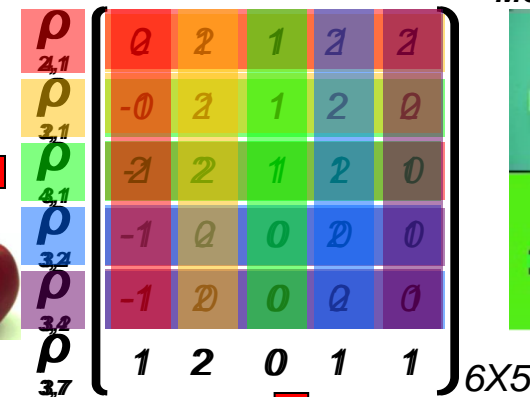
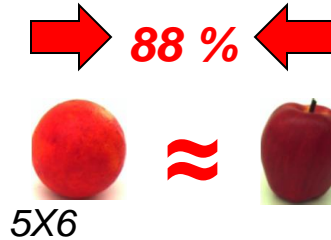
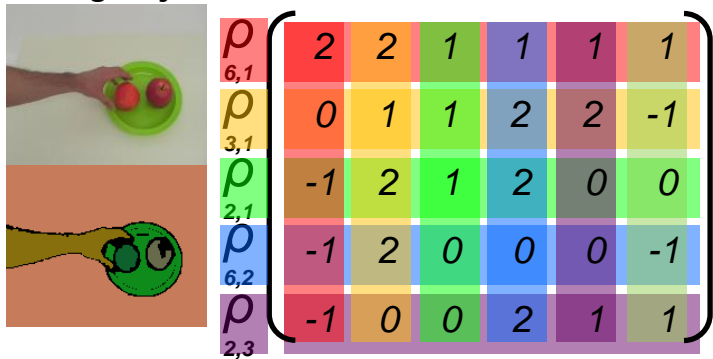
“Making Sandwich”

“Making Sandwich”



# Algorithm – Similarity Measure

“Moving Object”

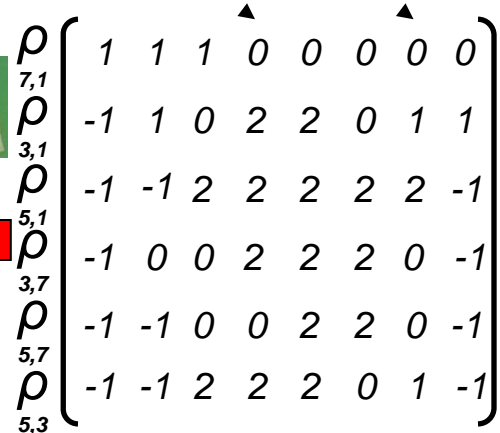
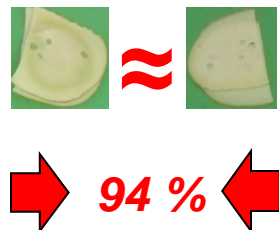
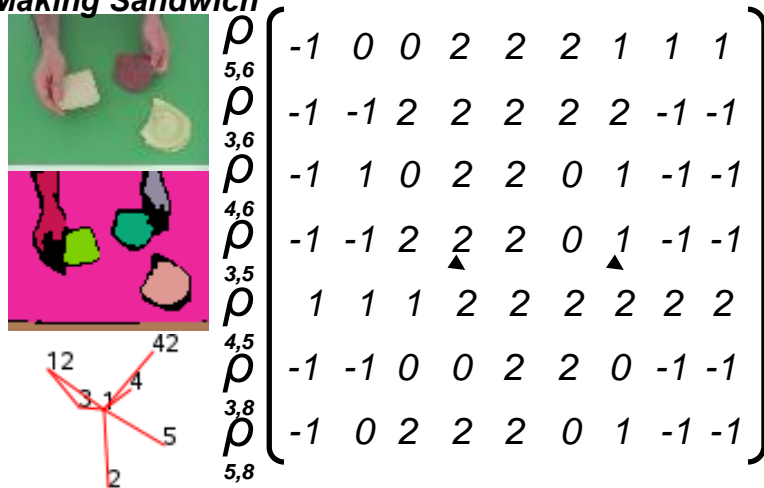


“Moving Object”

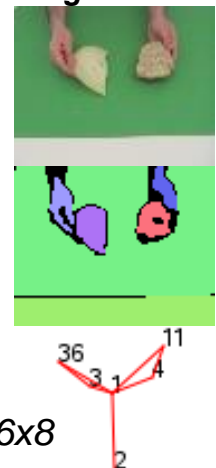


18% sub-string search in spatiotemporal domain

“Making Sandwich”



“Making Sandwich”



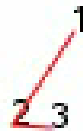
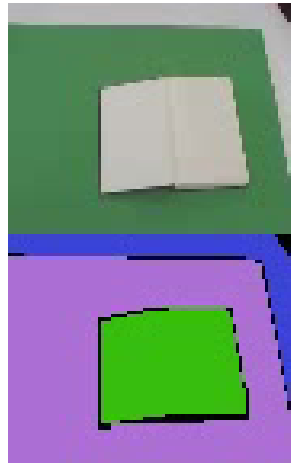


# Real Test Actions

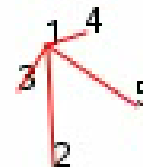
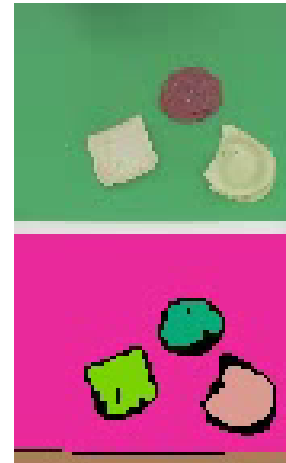
**“Moving Object”**  
x 4



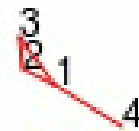
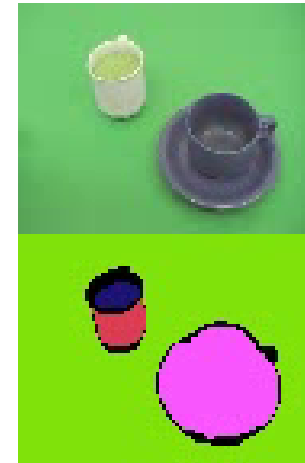
**“Opening Book”**  
x 4



**“Making Sandwich”**  
x 4

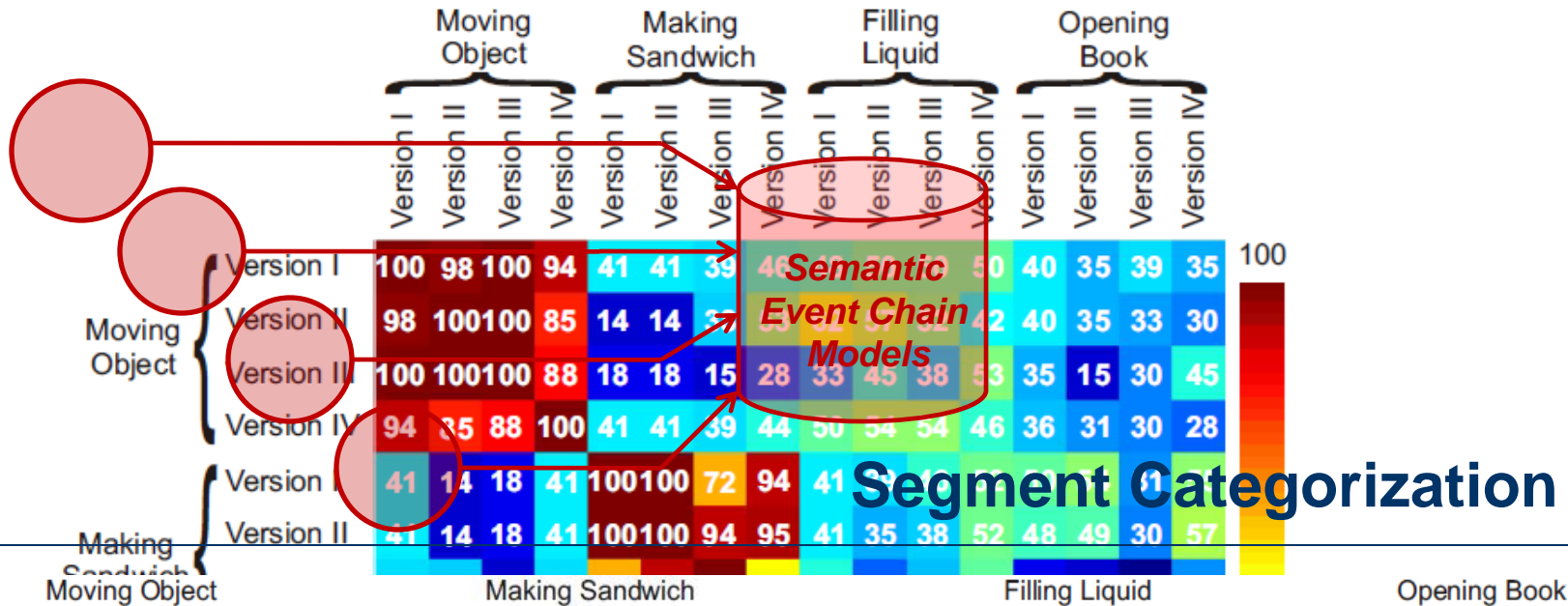


**“Filling Liquid”**  
x 4

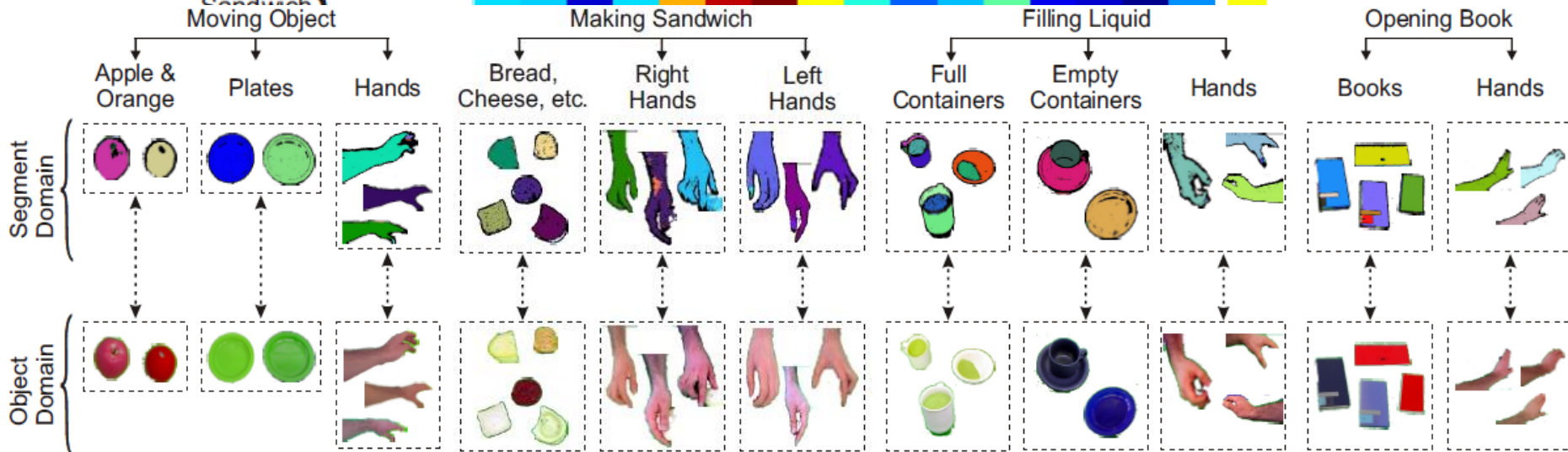


*Aksoy et al. IJRR 2011*  
*Aksoy et al. ICRA 2010*

# Action Classification



**Segment Categorization**



---

# Outline

1. Introduction
2. Algorithm: SECs
- 3. Learning & Execution**
4. Enriched SECs & Linguistics
5. Conclusion
6. Questions & Comments

# Learning SEC Models - Training

## MANipulation ACTION Dataset (**MANIAC**)

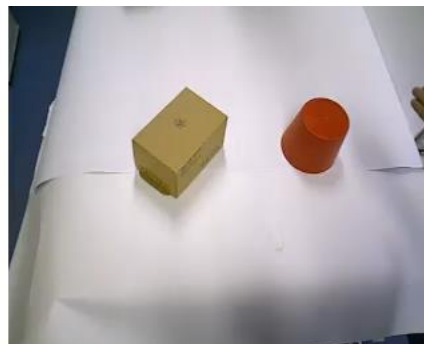
PUSH



HIDE



PUT ON TOP



STIR

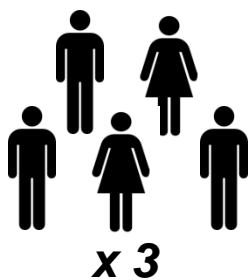


= 15

15

15

15



**15 Different Objects!...**

# Learning SEC Models - Training

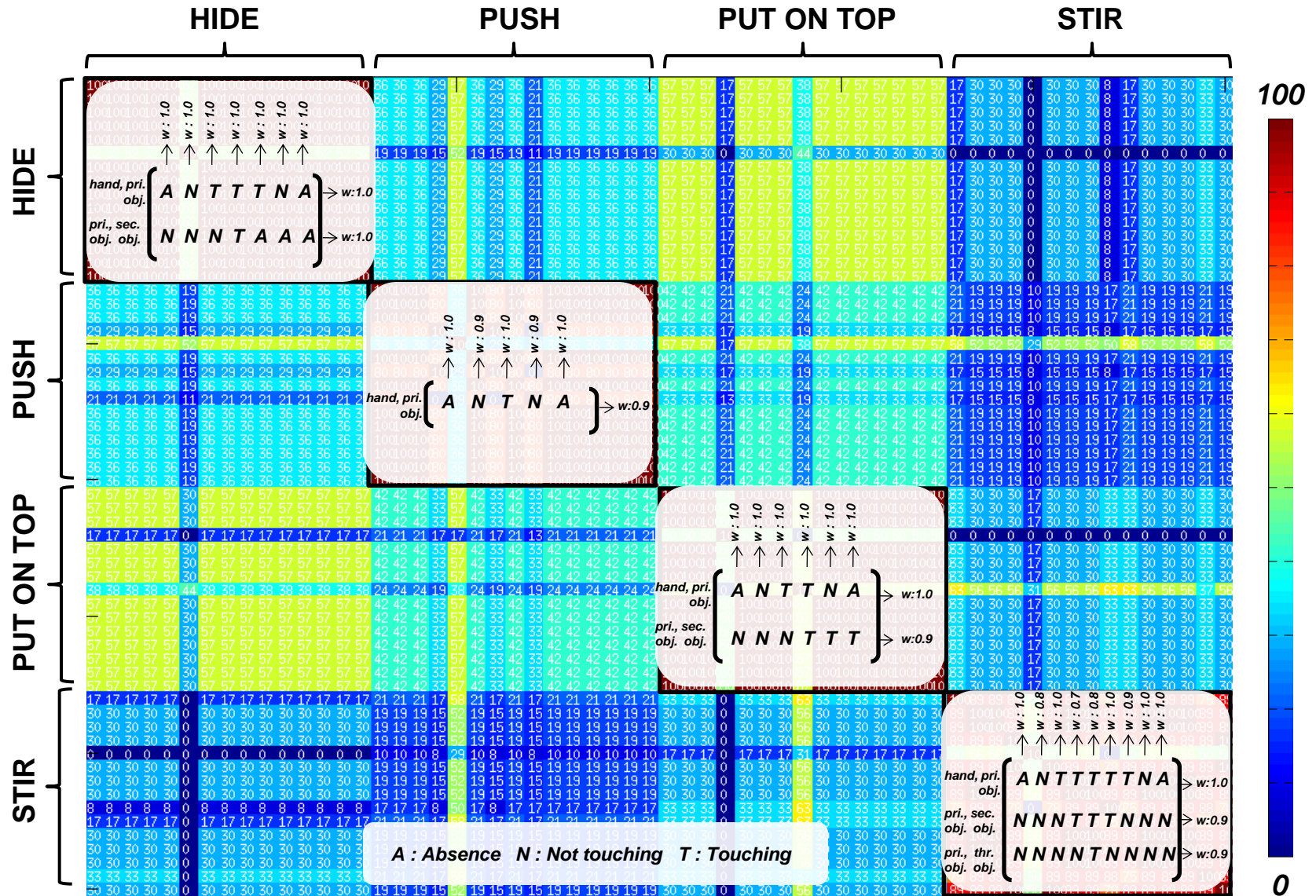
---

## MANipulation ACtion Dataset (**MANIAC**)

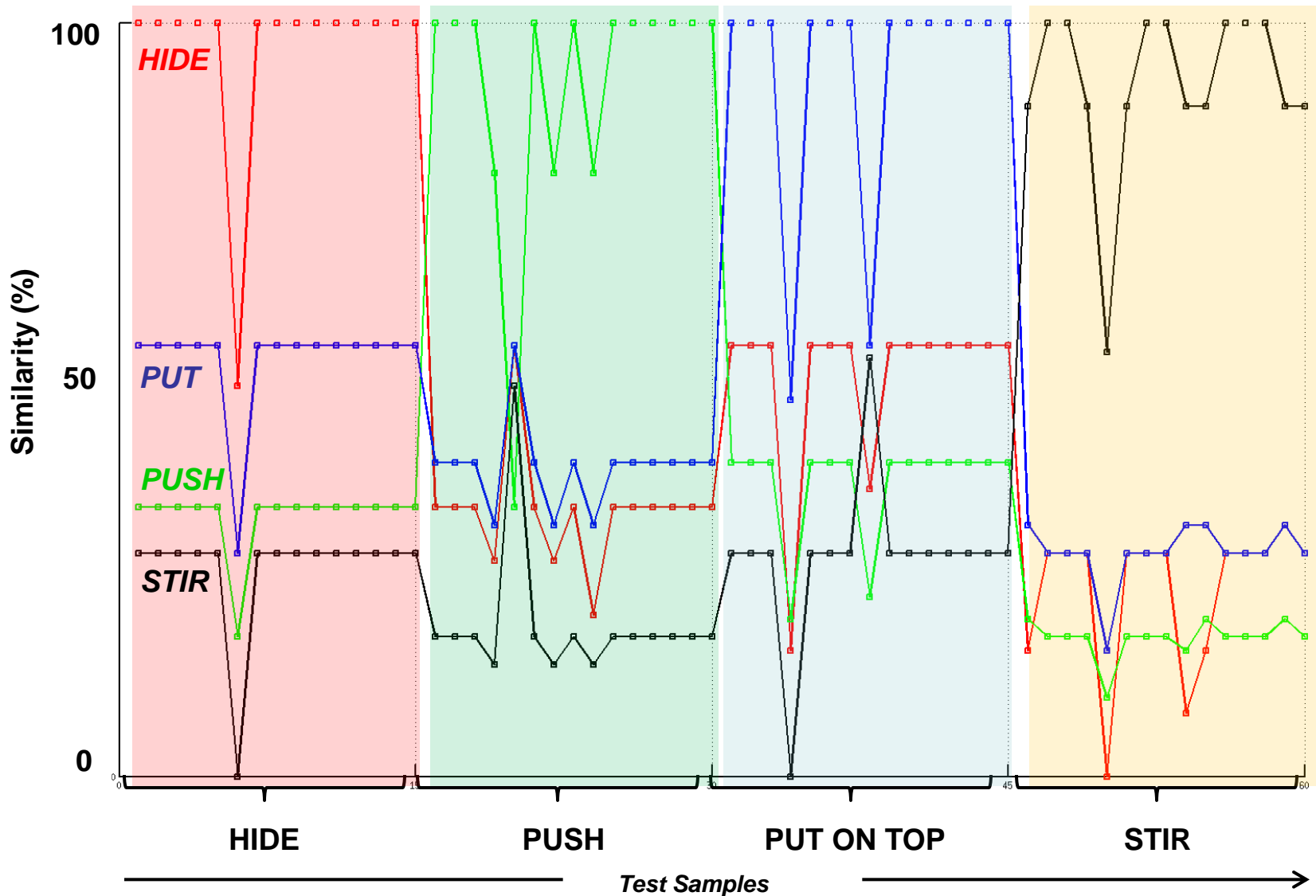
PUT ON TOP



# Learning SEC Models - Training



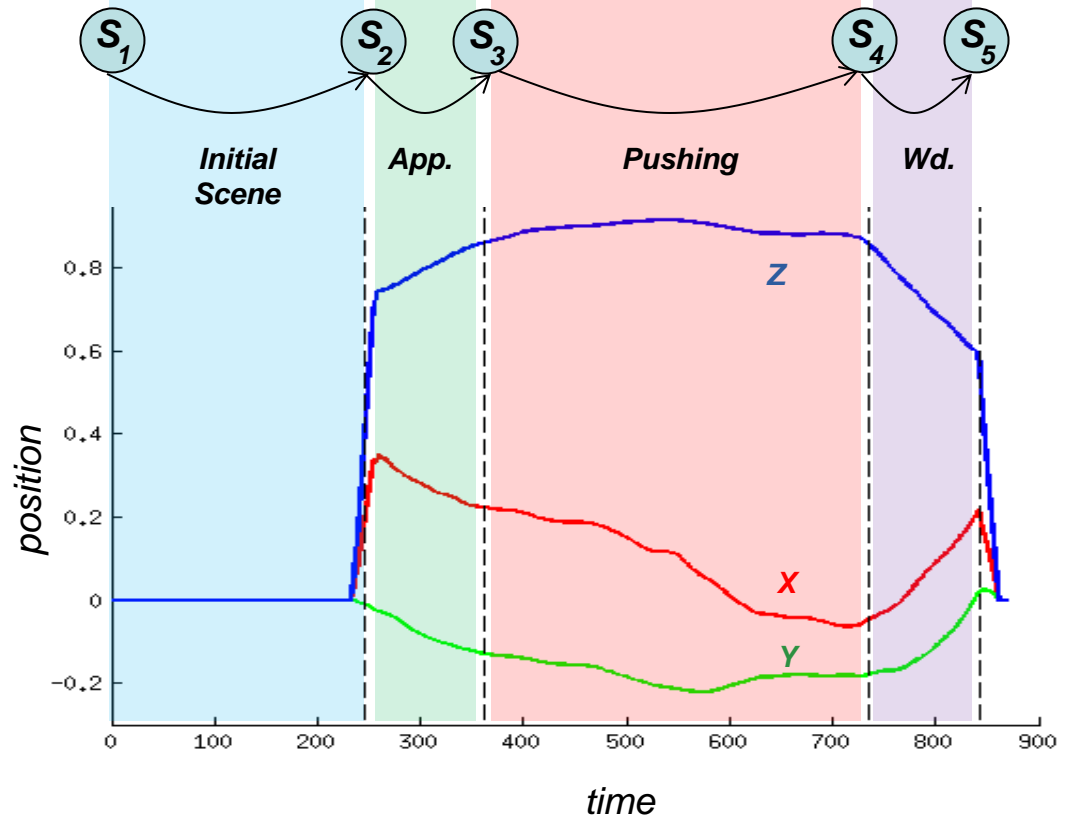
# Learning SEC Models - Testing



# Executing SEC Models

**PUSH**

hand, primary object [ A N T N A ]





# Executing SEC Models

**PUSH**

*hand, primary object* ( A N T N A )

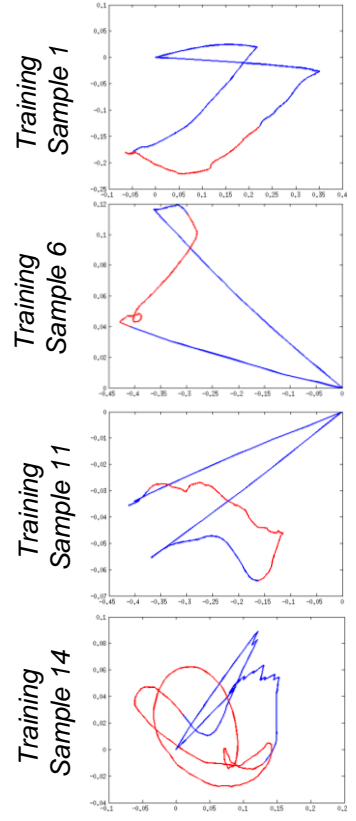


**Initial Scene**

**App.**

**Pushing**

**Wd.**



# Executing SEC Models

**PUSH**

*hand, primary object* ( A N T N A )

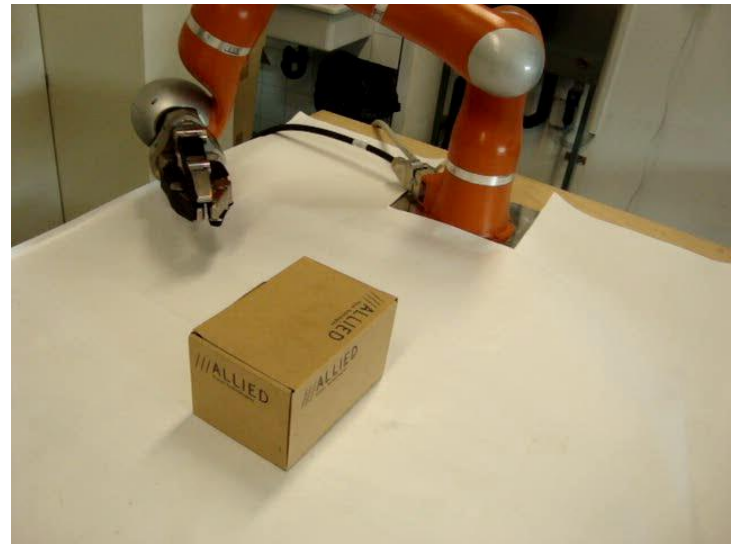
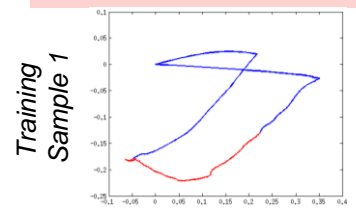
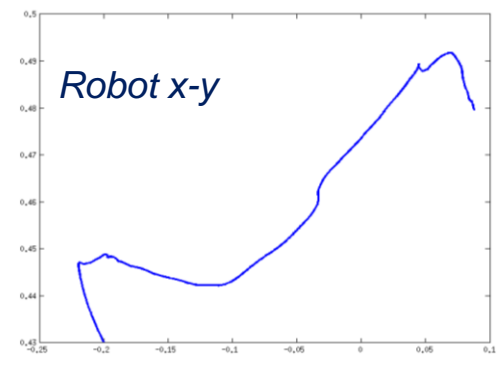
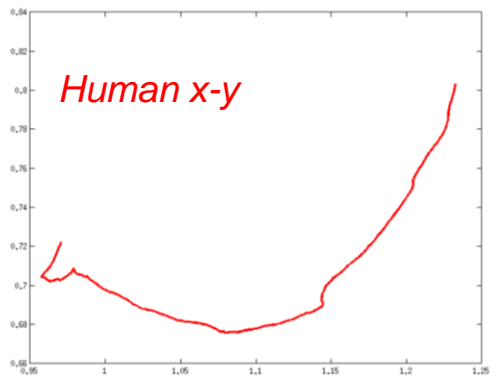


**Initial Scene**

**App.**

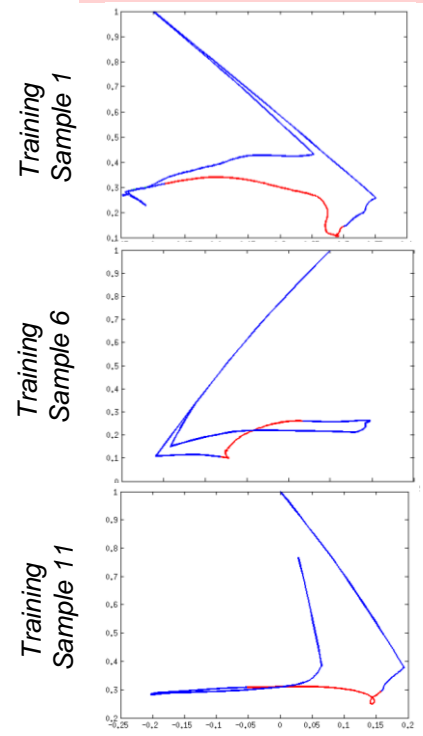
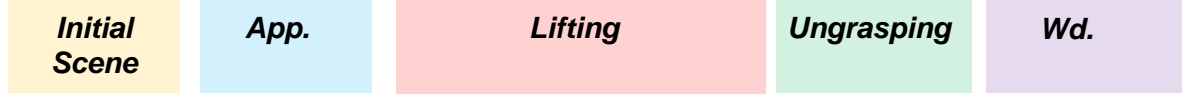
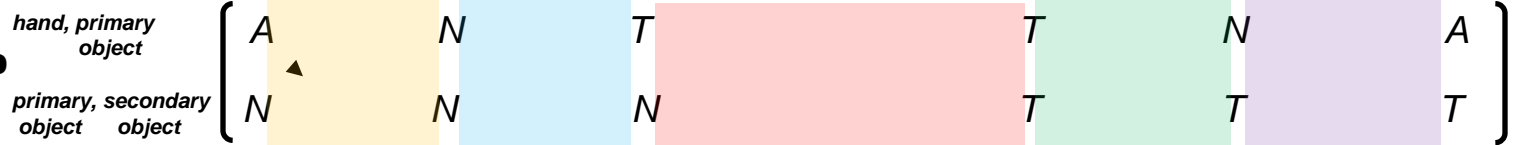
**Pushing**

**Wd.**



# Executing SEC Models

**PUT ON TOP**

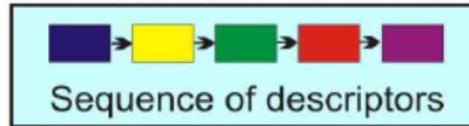


# SECs : Piaget & Robots

In Memory:

SEC with syntactic elements

Action Schema for Action  $\alpha$



“PUT ON TOP”

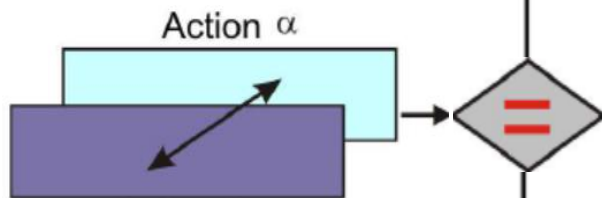
“PUSH”

“HIDE”

“STIR”

SEC-LEVEL

**Semantic Comparison**  
at the level of the “whole” Action



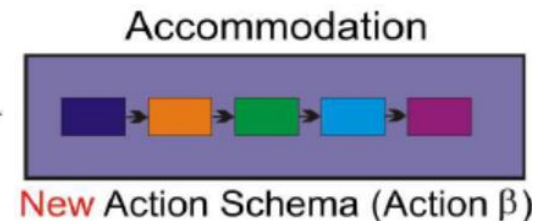
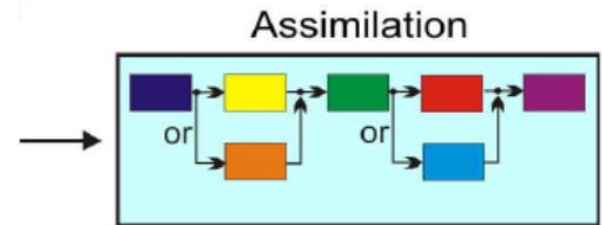
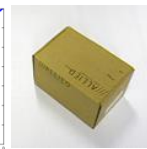
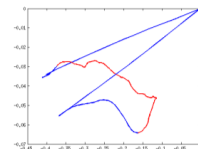
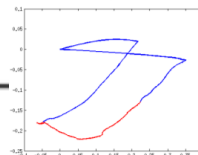
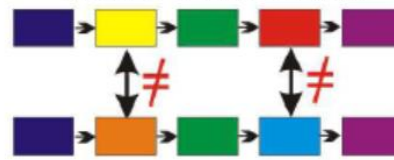
“PUSH”

Yes

No

**Syntactic Comparison**

“inside” each Action



Aksoy et al. ICDL 2013

---

# Outline

1. Introduction
2. Algorithm: SECs
3. Learning & Execution
- 4. Enriched SECs & Linguistics**
5. Conclusion
6. Questions & Comments

# Enriched SECs

---



## SECs with:

- Multiple views
- Occlusion handling
- Object Information
- Pose Estimation
- Trajectory Information

***Bottom-up Approach!...***

# Enriched SECs & Linguistics

Grammar:

**Subject** + **Verb** + **Direct Object** + **Indirect Object**

Manipulation:

**Manipulator** + **Action** + **Primary Object** + **Secondary Object**

Example:

The hand puts a cup on top of a box

SEC:

**Subject, Indir. Obj., Dir. Obj.**  $\left( \begin{array}{cccccc} A & N & T & T & N & A \\ N & N & N & T & T & T \end{array} \right)$   
**Verb (pose, push, speed, etc.)**



*Aein et al. (Submitted) IROS 2013*

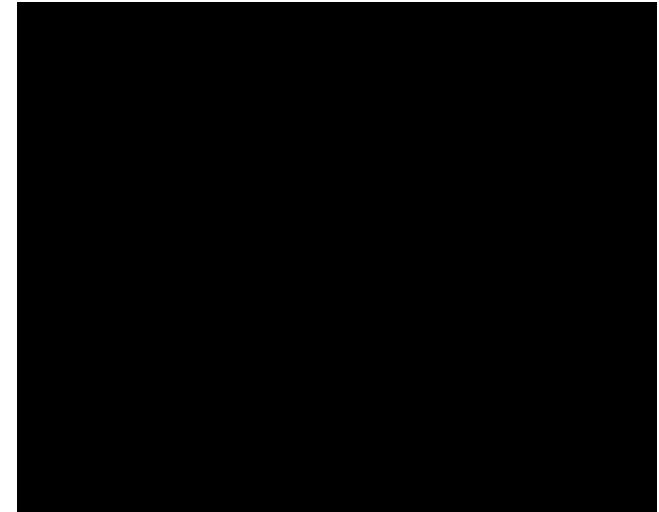
# Enriched SECs & Linguistics

Grammar:

**Subject** + **Verb** + **Direct Object** + **Indirect Object**

Example:

**The hand** first **pushes** a **box** then a **ball**.  
**The hand** afterwards **hides** the **ball**  
 with a **bucket** and **puts** the **box**  
 on top of **the bucket**.



SEC:

	<i>Pushing</i>		<i>Pushing</i>		<i>Hiding</i>					<i>Putting</i>					
<i>hand</i> , <i>bucket</i>	N	N	N	N	N	T	T	T	N	N	N	N	N	A	
<i>hand</i> , <i>box</i>	N	T	N	N	N	N	N	N	N	N	N	T	T	N	A
<i>hand</i> , <i>ball</i>	N	N	N	T	N	N	N	A	A	A	A	A	A	A	A
<i>hand</i> , <i>noise</i>	A	A	A	A	A	A	A	A	A	T	N	N	A	A	A
<i>box</i> , <i>bucket</i>	N	N	N	N	N	N	N	N	N	N	N	N	T	T	T
<i>ball</i> , <i>bucket</i>	N	N	N	N	N	N	T	A	A	A	A	A	A	A	A

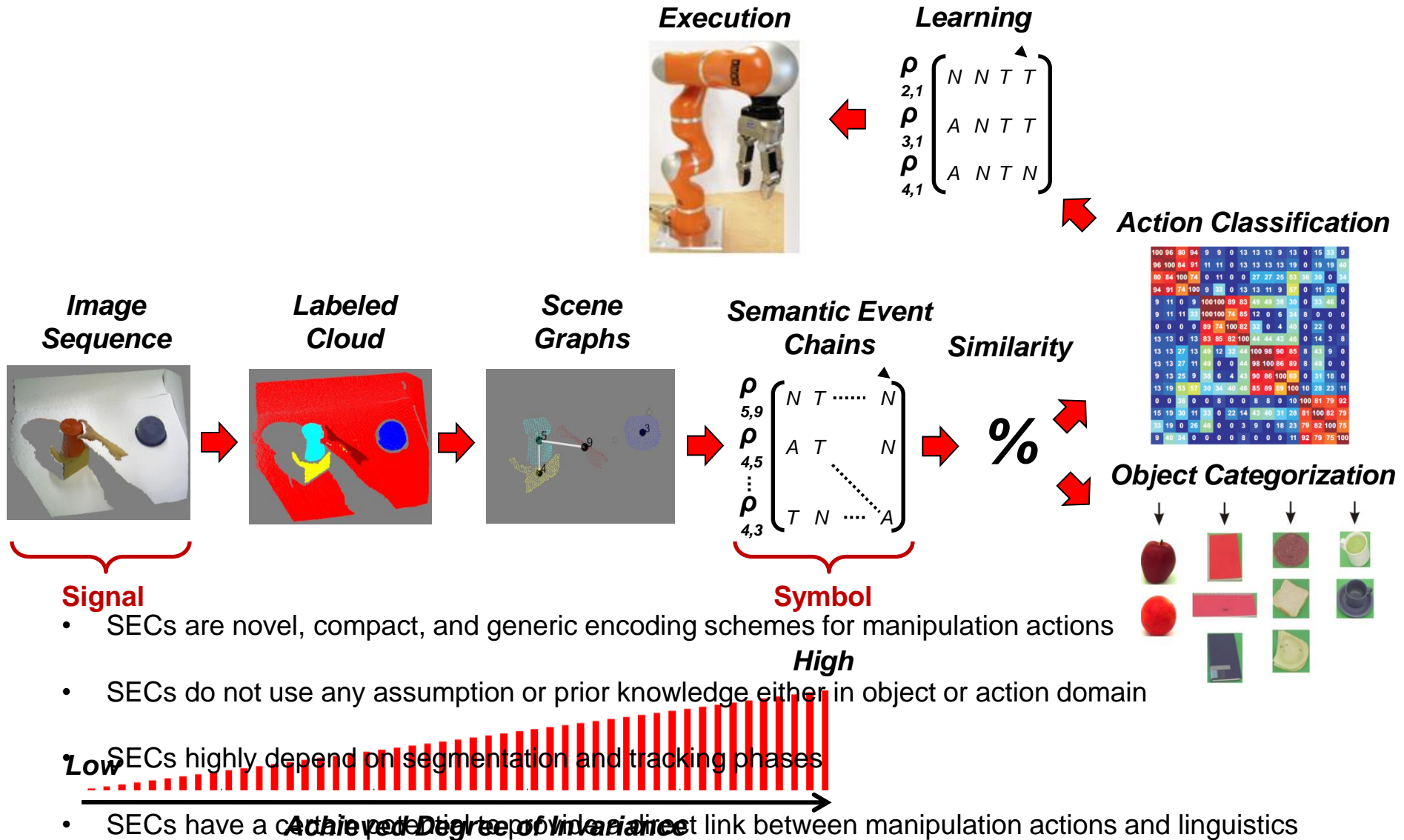


---

# Outline

1. Introduction
2. Algorithm: SECs
3. Learning & Execution
4. Enriched SECs & Linguistics
- 5. Conclusion**
6. Questions & Comments

# Conclusion



- SECs are novel, compact, and generic encoding schemes for manipulation actions

- SECs do not use any assumption or prior knowledge either in object or action domain

SECs highly depend on segmentation and tracking phases

- SECs have a compact link between manipulation actions and linguistics

---

# Outline

1. Introduction
2. Algorithm: SECs
3. Learning & Execution
4. Enriched SECs & Linguistics
5. Conclusion
6. Questions & Comments