Scene Representation for Robots: From Elementary Behaviors to Grasping

Stephan Zibner

RUB

INI

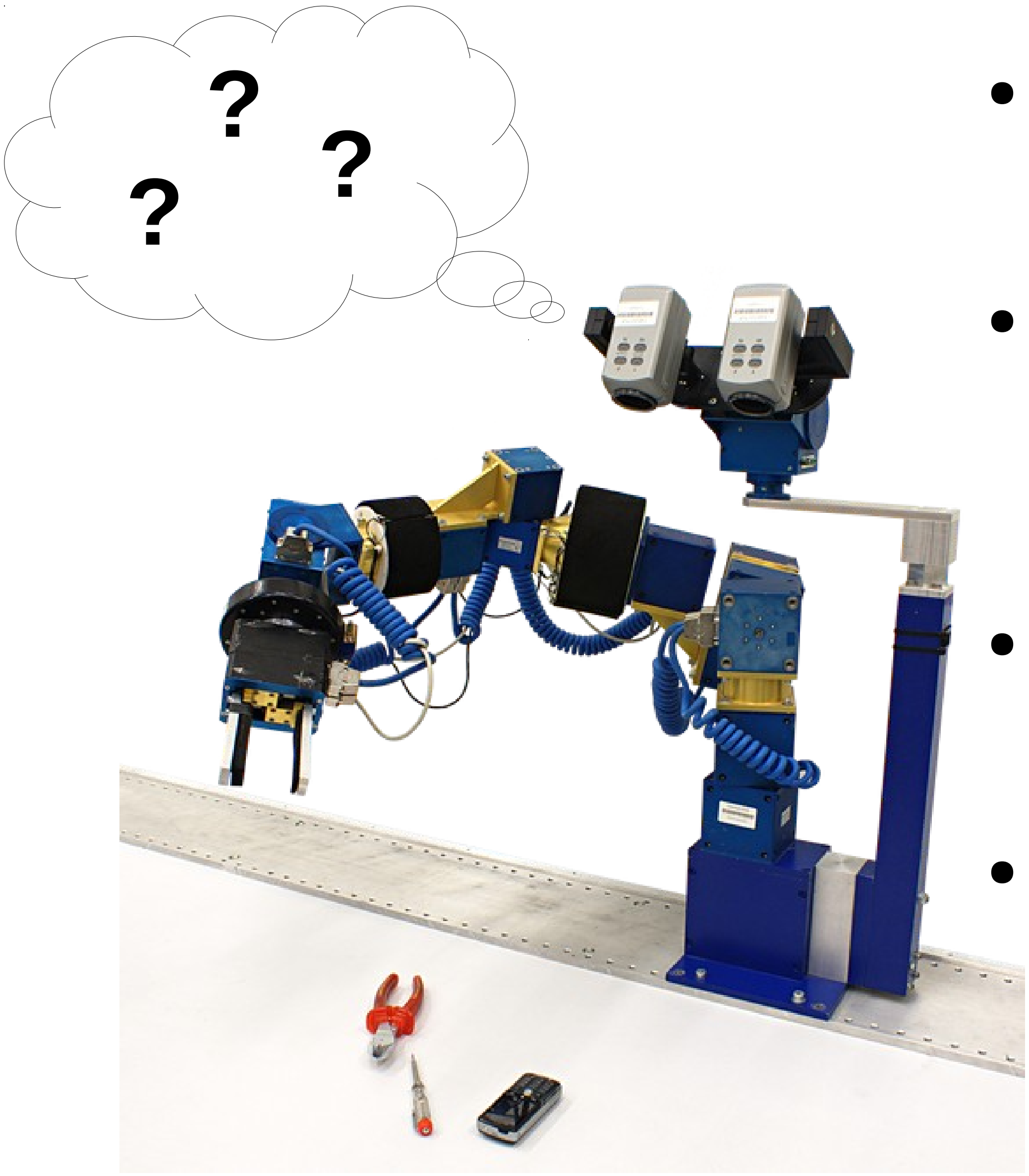
SEVENTH FRAMEWORK PROGRAMME

Outline

- Representing Scenes
  - In Humans
  - Robotic Scenario
- Building Blocks
- Elementary Behaviors
  - Exploration
  - Maintenance
  - Query
- Reaching and Grasping

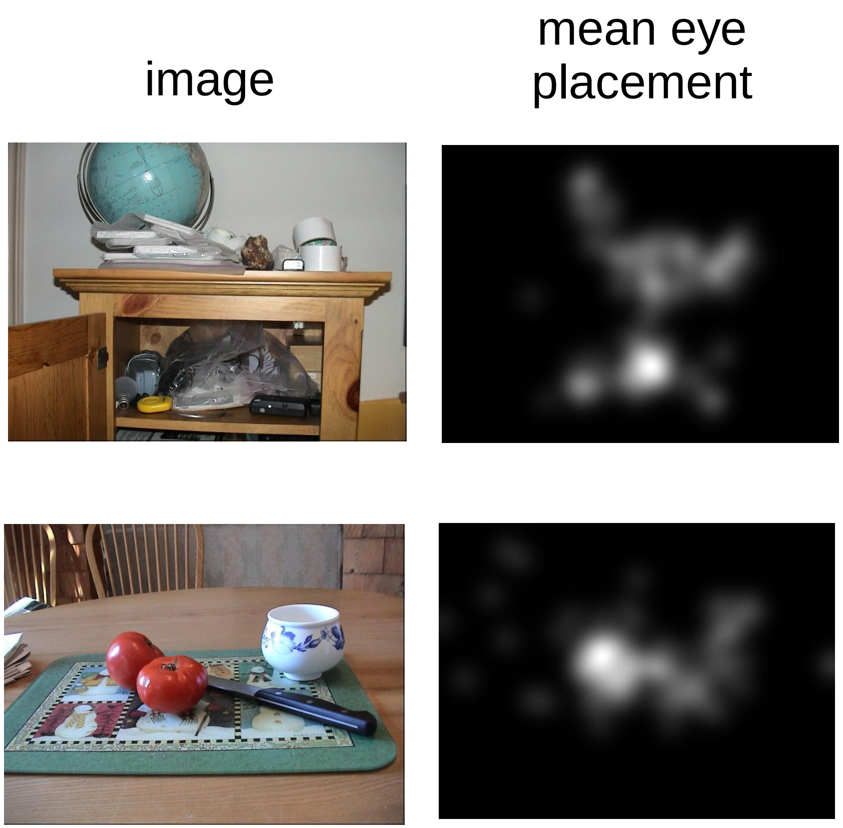
Representing Scenes

Scene Representation



- internal representation of environment
- foundation for every higher cognitive operation and action
- stable despite eye and body movements
- limited capacity, link to long-term memory

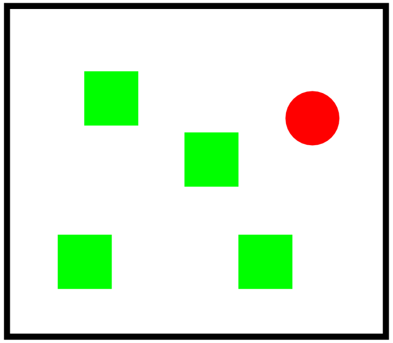
Scene Representation in Humans



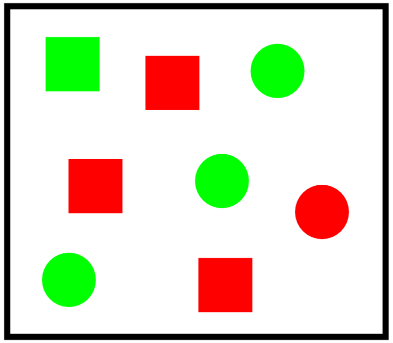
Bruce, Tsotsos (2009)

- human eye movement is not random (saliency)

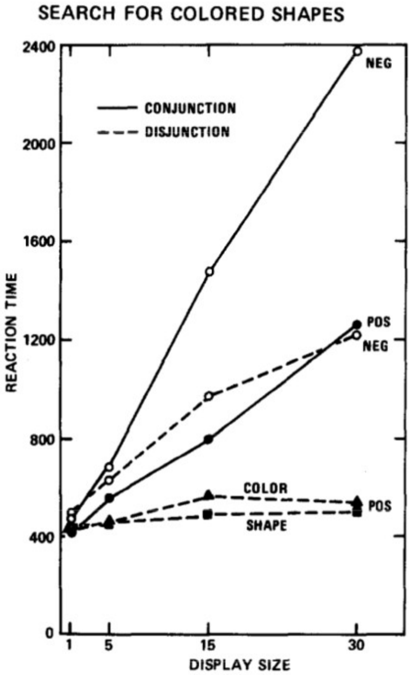
Scene Representation in Humans



feature search



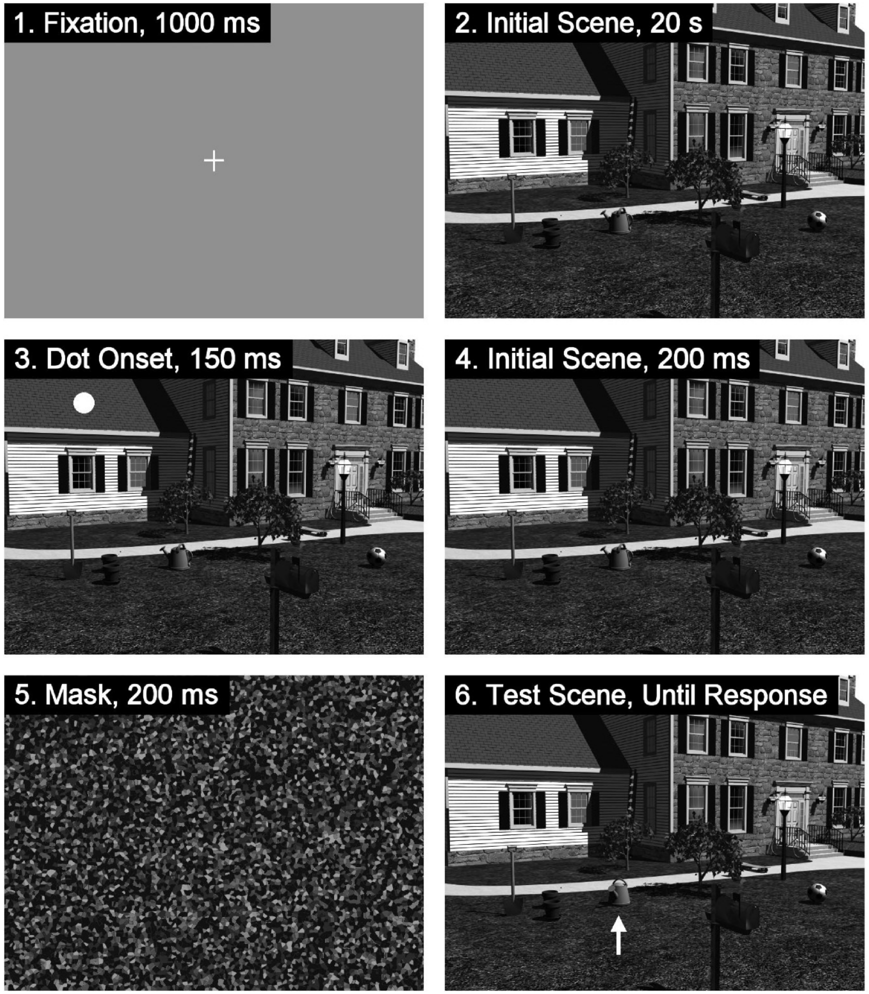
conjunctive search



Bottom: Treisman, Gelade (1980)

- human eye movement is not random (saliency)
- visual search highlights target objects

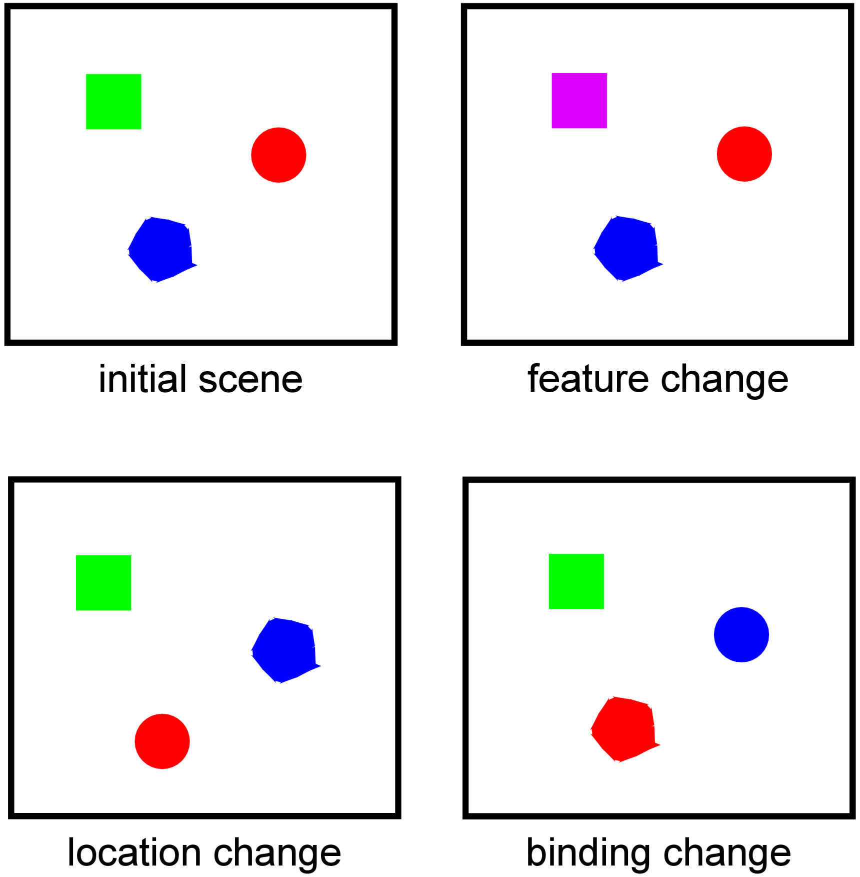
Scene Representation in Humans



Hollingworth (2005)

- human eye movement is not random (saliency)
- visual search highlights target objects
- details of a scene are kept in memory

Scene Representation in Humans


initial scene

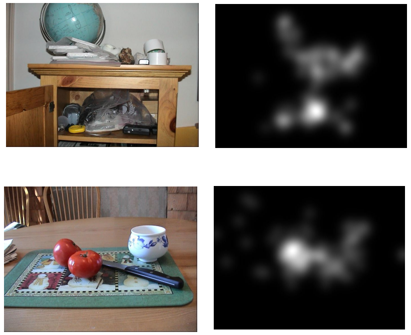
feature change

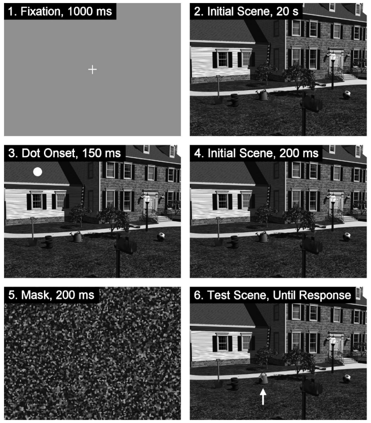
location change

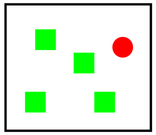
binding change

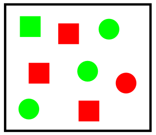
- human eye movement is not random (saliency)
- visual search highlights target objects
- details of a scene are kept in memory
- no full representation (change blindness)

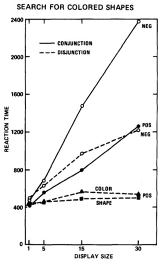
Scene Representation in Humans

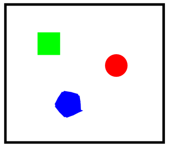


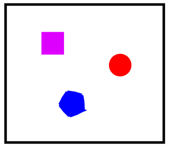


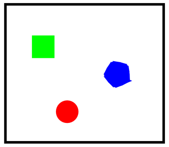

feature search

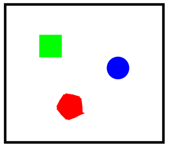

conjunctive search




initial scene


feature change

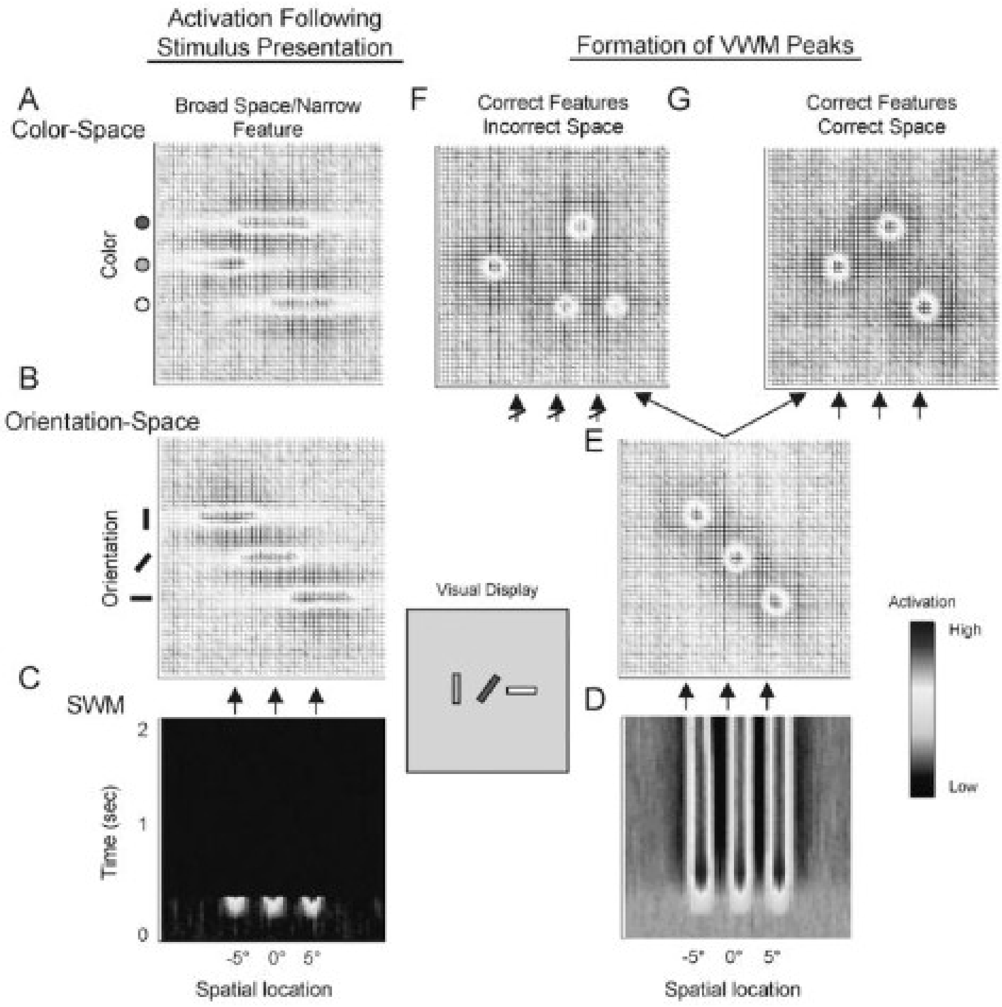

location change


binding change

- human eye movement is not random (saliency)
- visual search highlights target objects
- details of a scene are kept in memory
- no full representation (change blindness)

attention is a key theme in visual processing

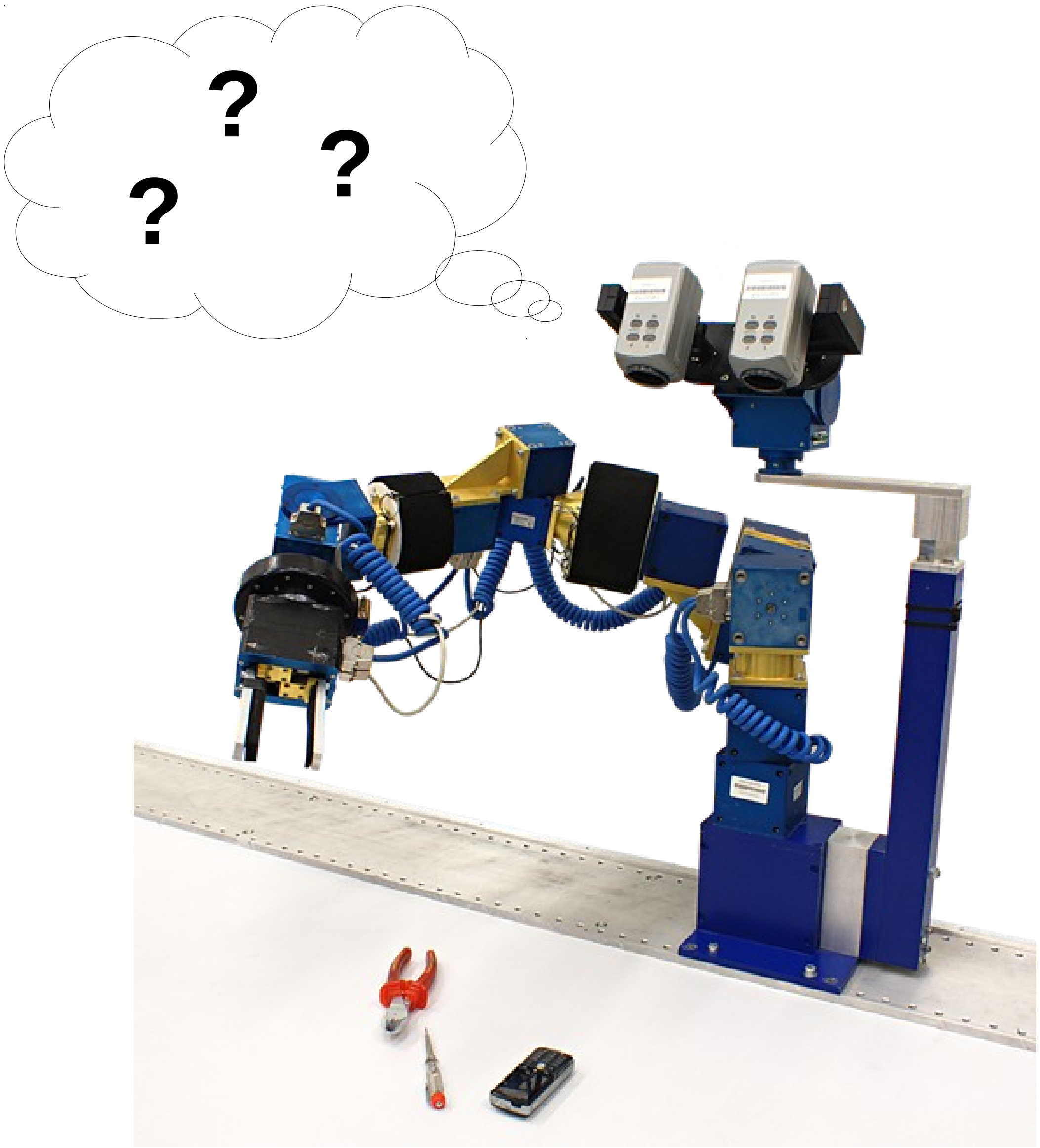
DFT Model



Johnson, Spencer, Schöner (2008)

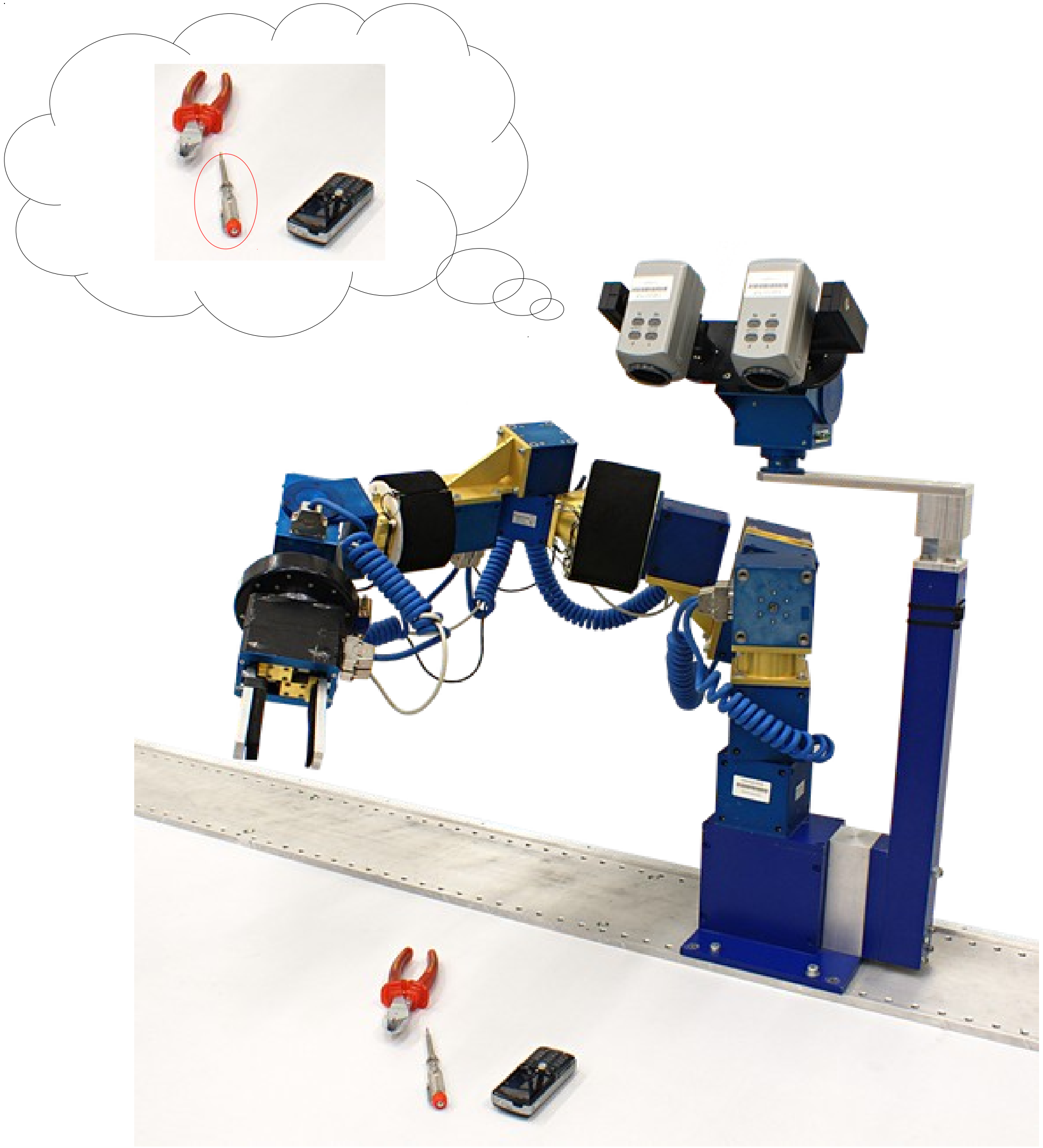
- model of human visual working memory based on dynamic neural fields
- representation of low-level features
- decomposition of features, binding through shared space

Robotic Scenario



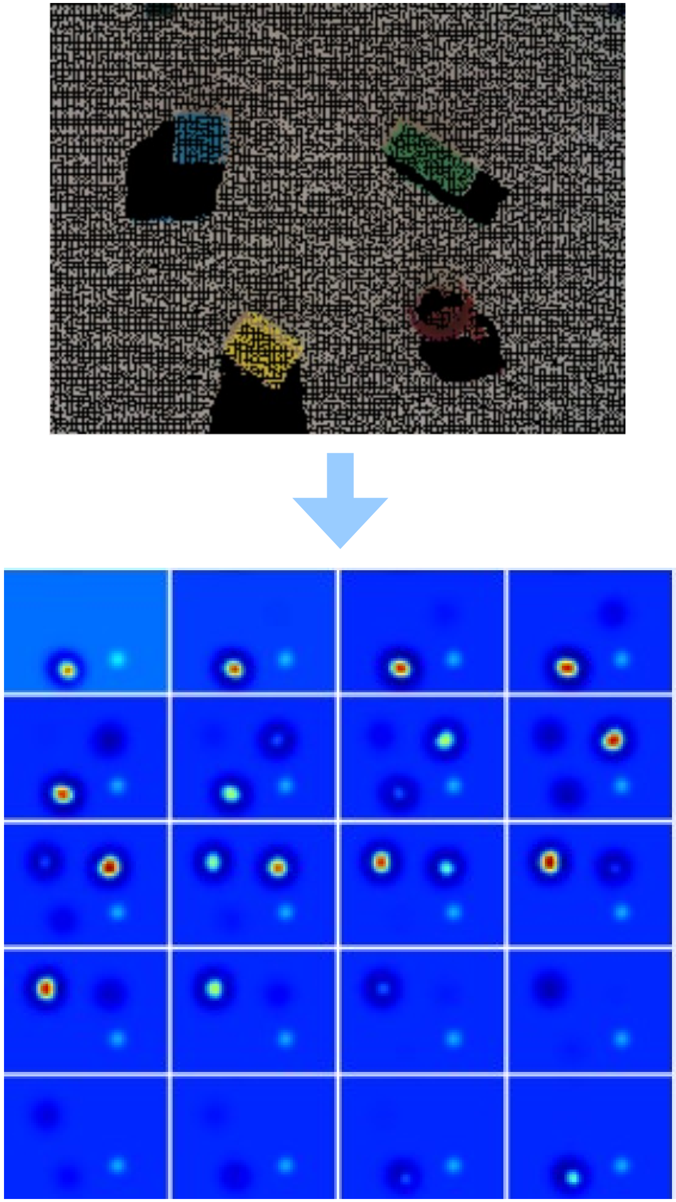
- apply to table-top scenario and human interaction
- use the internal representation for behavior generation (e.g., grasping)
- interact with humans (“hand me the red screwdriver”, “what's to the left of the pliers?”)

Robotic Scenario



- explore the environment and store objects and their features internally
- maintain the internal representation
- query the represen- tation to create autonomous action based on the representation

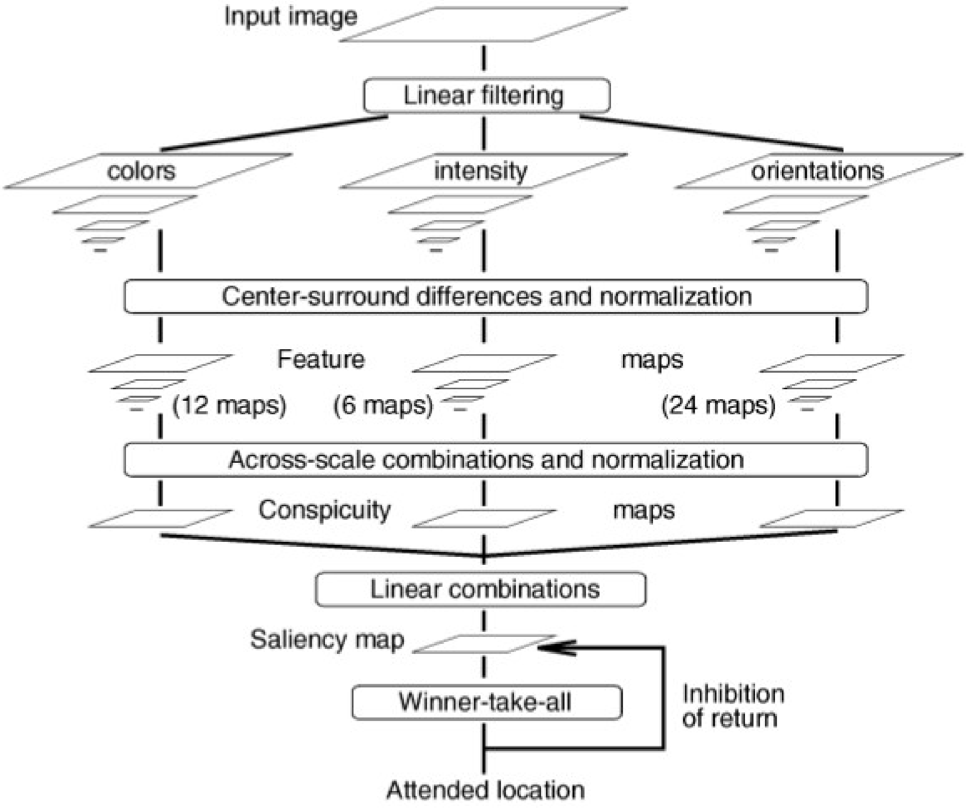
Robotic Scenario: Challenges

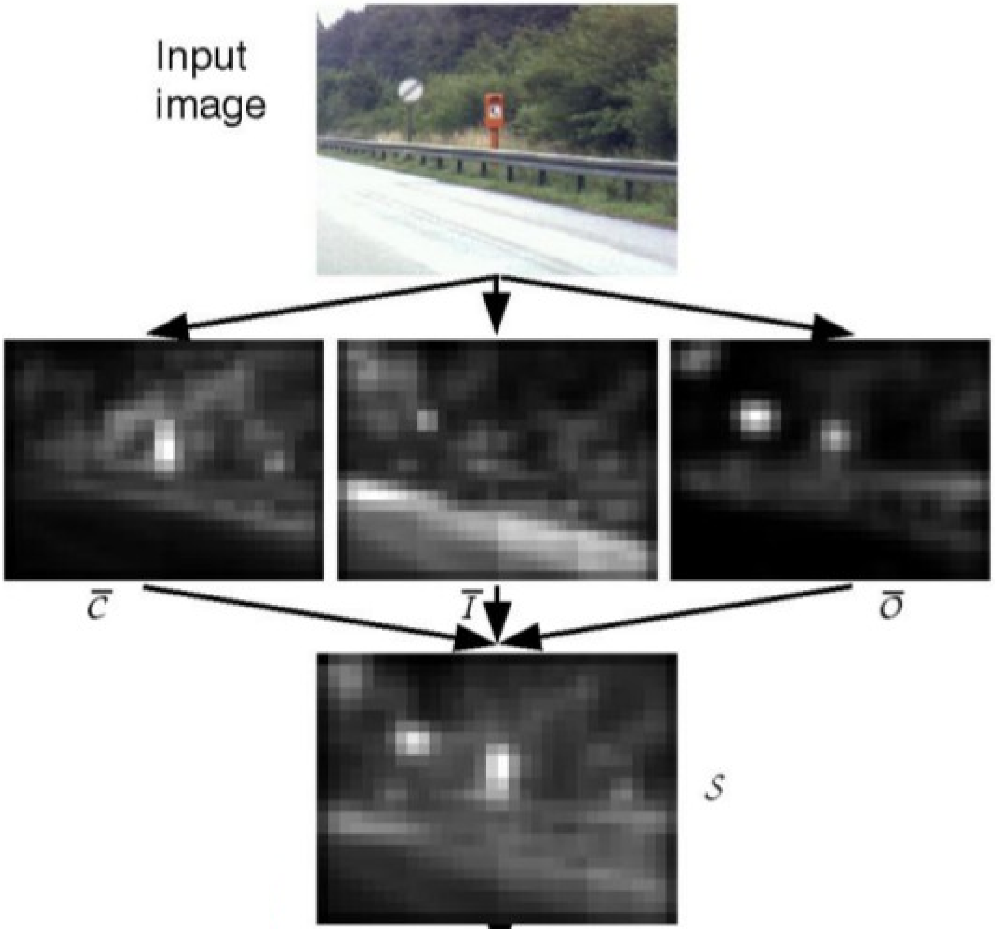


- real sensory input
- moving sensors
- limited field of view
- 3D space
- dynamic scenes
- multiple behaviors
- computational constraints
- and many more ...

Building Blocks

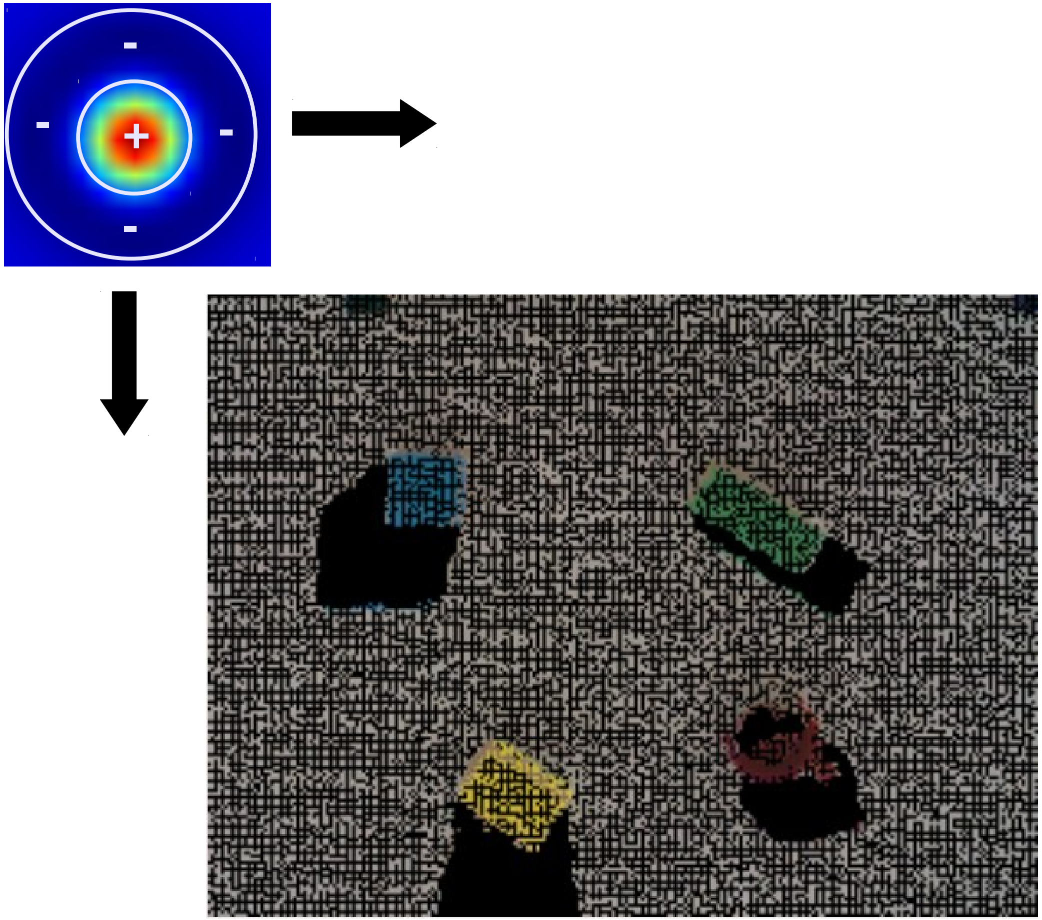
Saliency





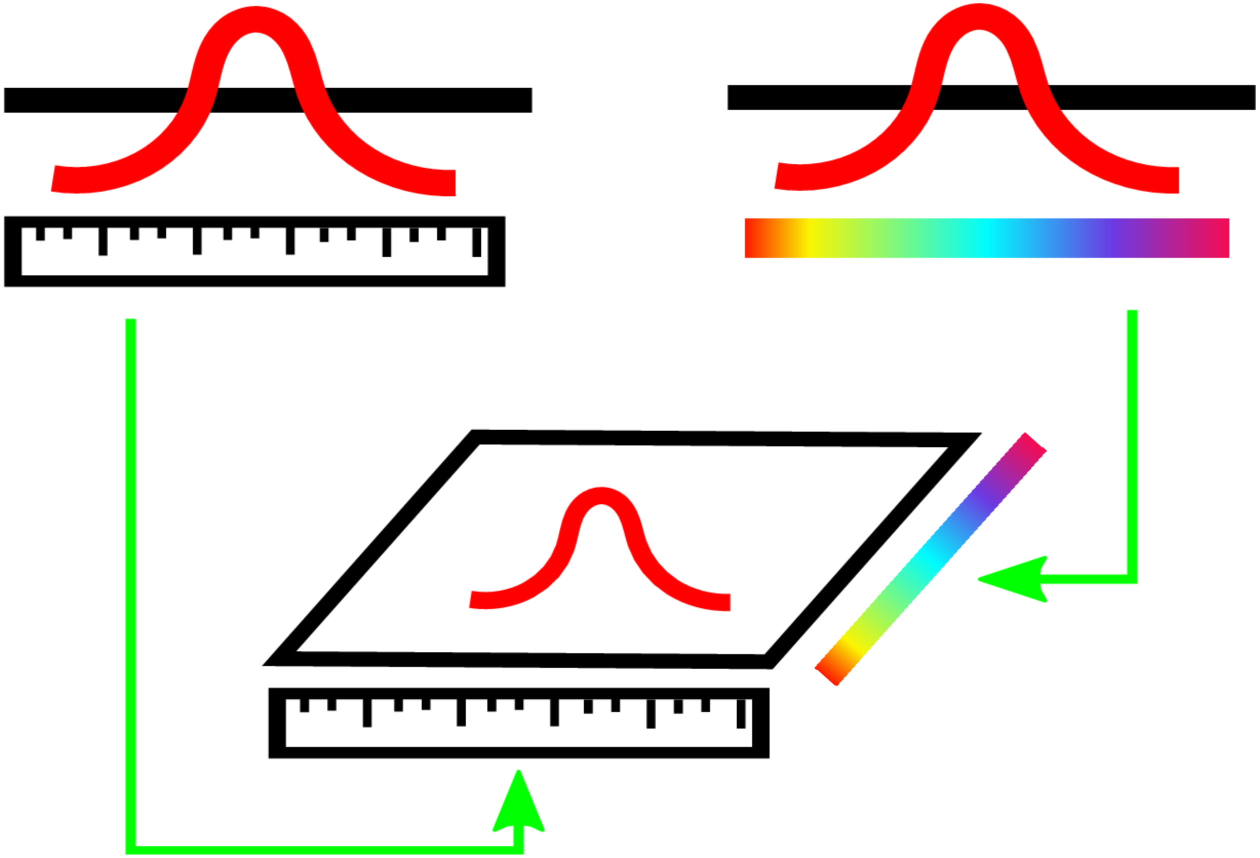
Itti, Koch, Niebur (1998)

Saliency



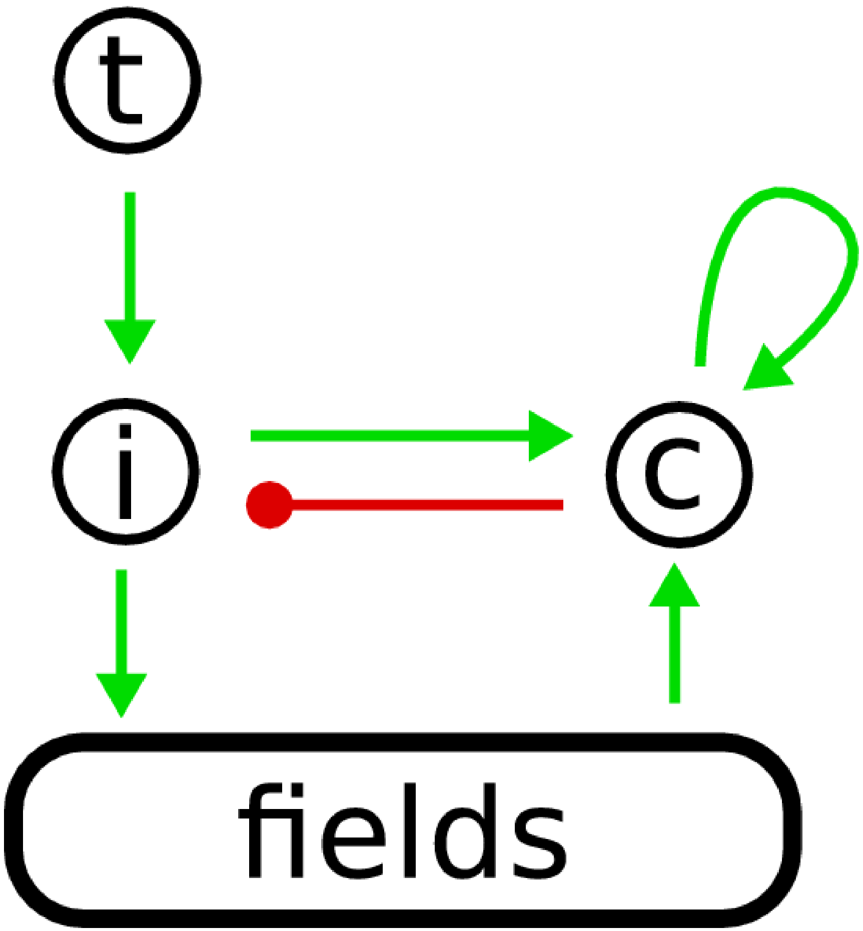
- on-/off-center responses
- uniform regions result in zero responses
- objects fitting into on-center region produce non-zero responses
- these lead to detection decision in fields

Space-Feature Links



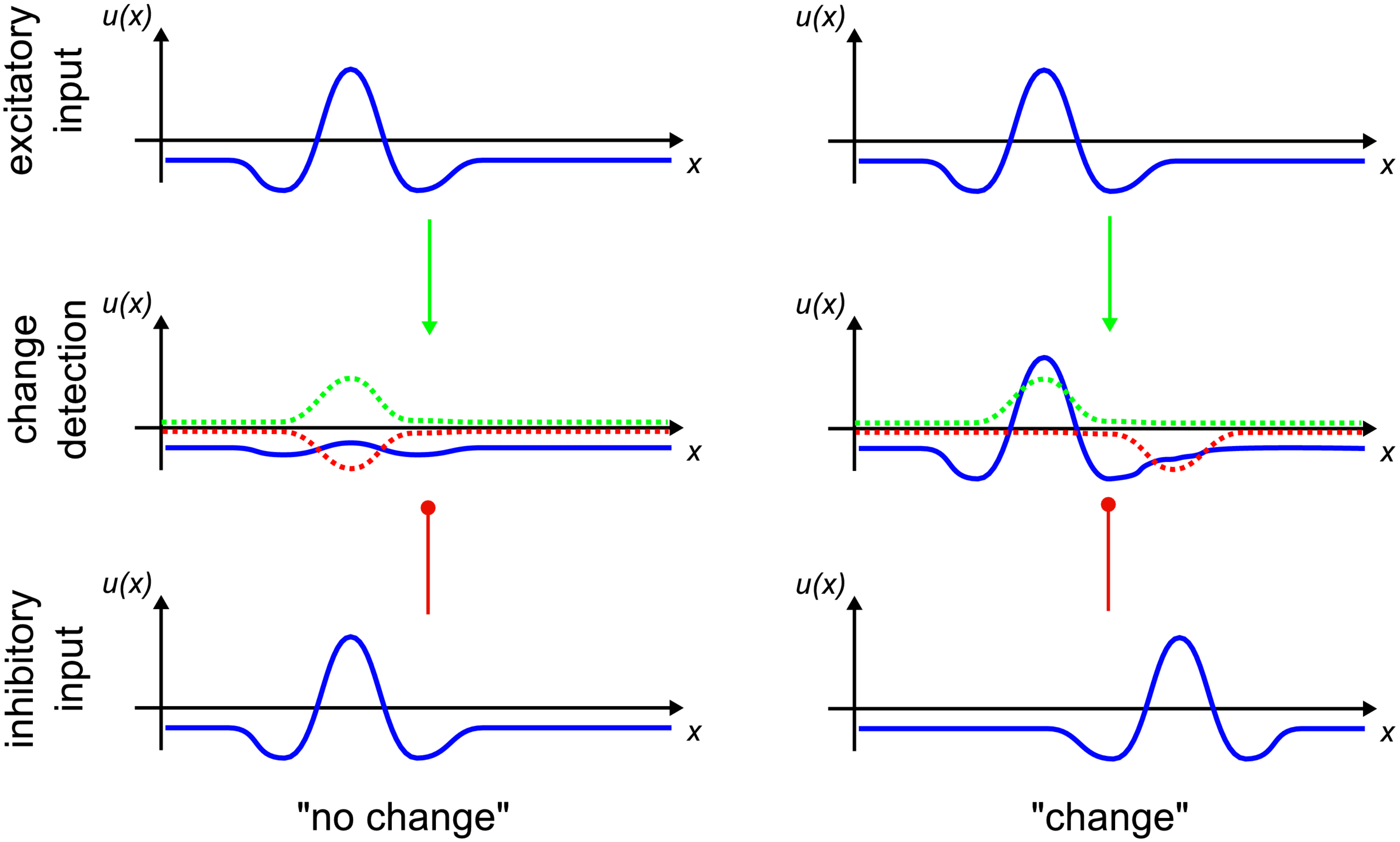
Feature estimates are linked to spatial positions.

Autonomy of Behaviors



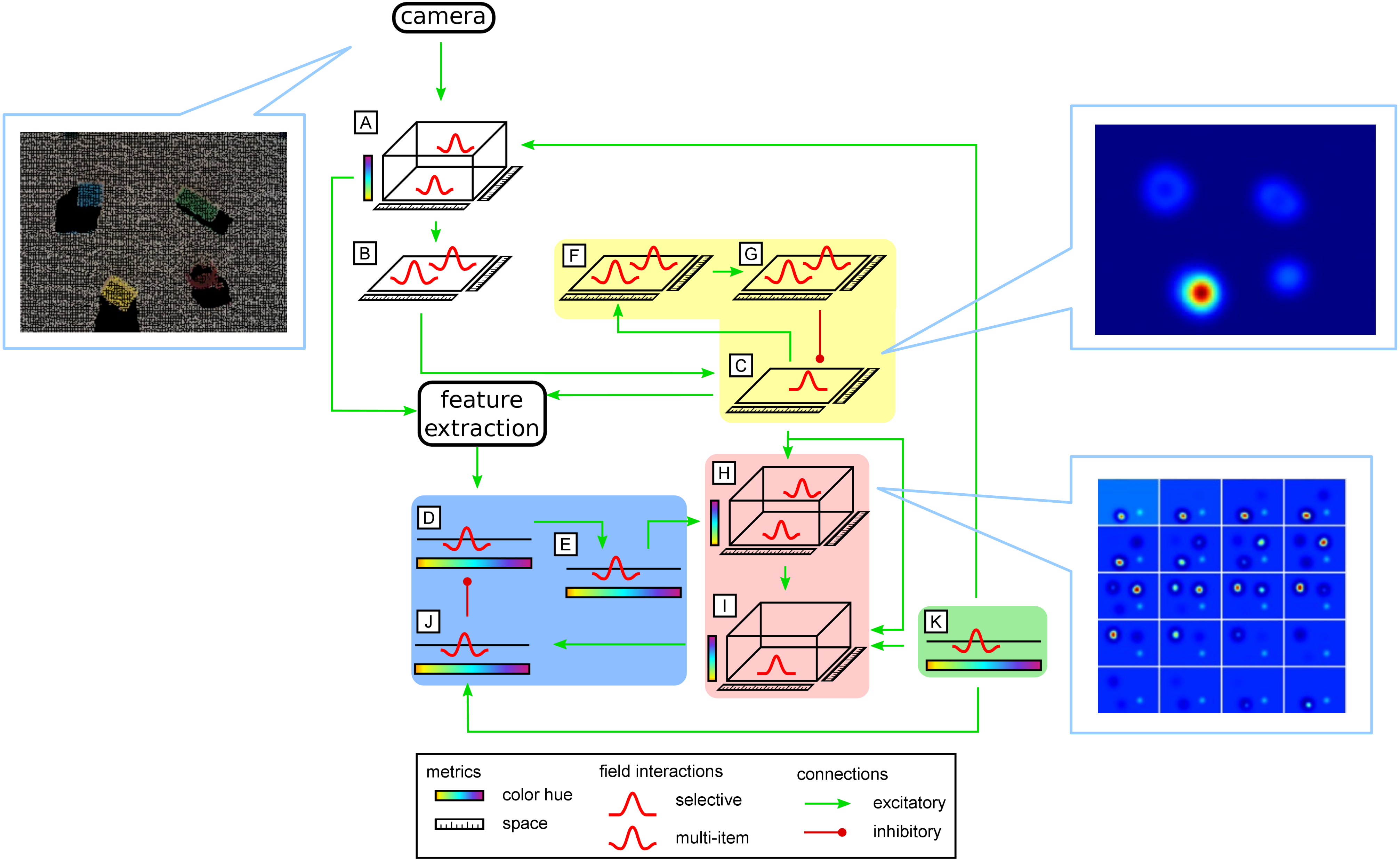
- elementary cognitive units (ECU)
- intention node boosts fields
- CoS node detects completion of behavior in fields
- sequences or exclusions through precondition and suppression nodes

Change Detection with Fields



Elementary Behaviors of Scene Representation

Architecture Overview



camera

A

B

F

G

C

feature extraction

D

E

J

H

I

K

metrics

color hue

space

field interactions

selective

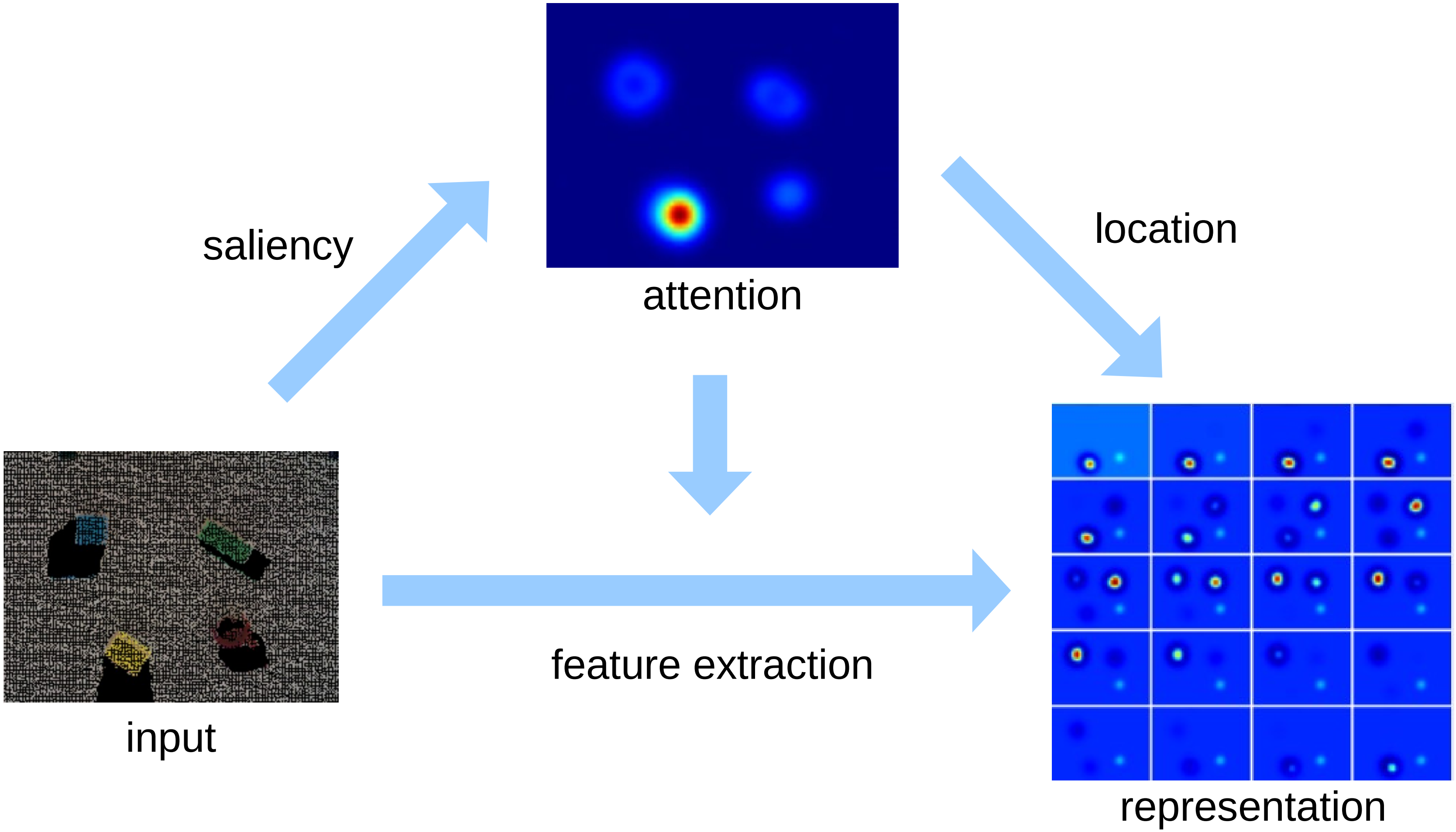
multi-item

connections

excitatory

inhibitory

Exploration Behavior



saliency

attention

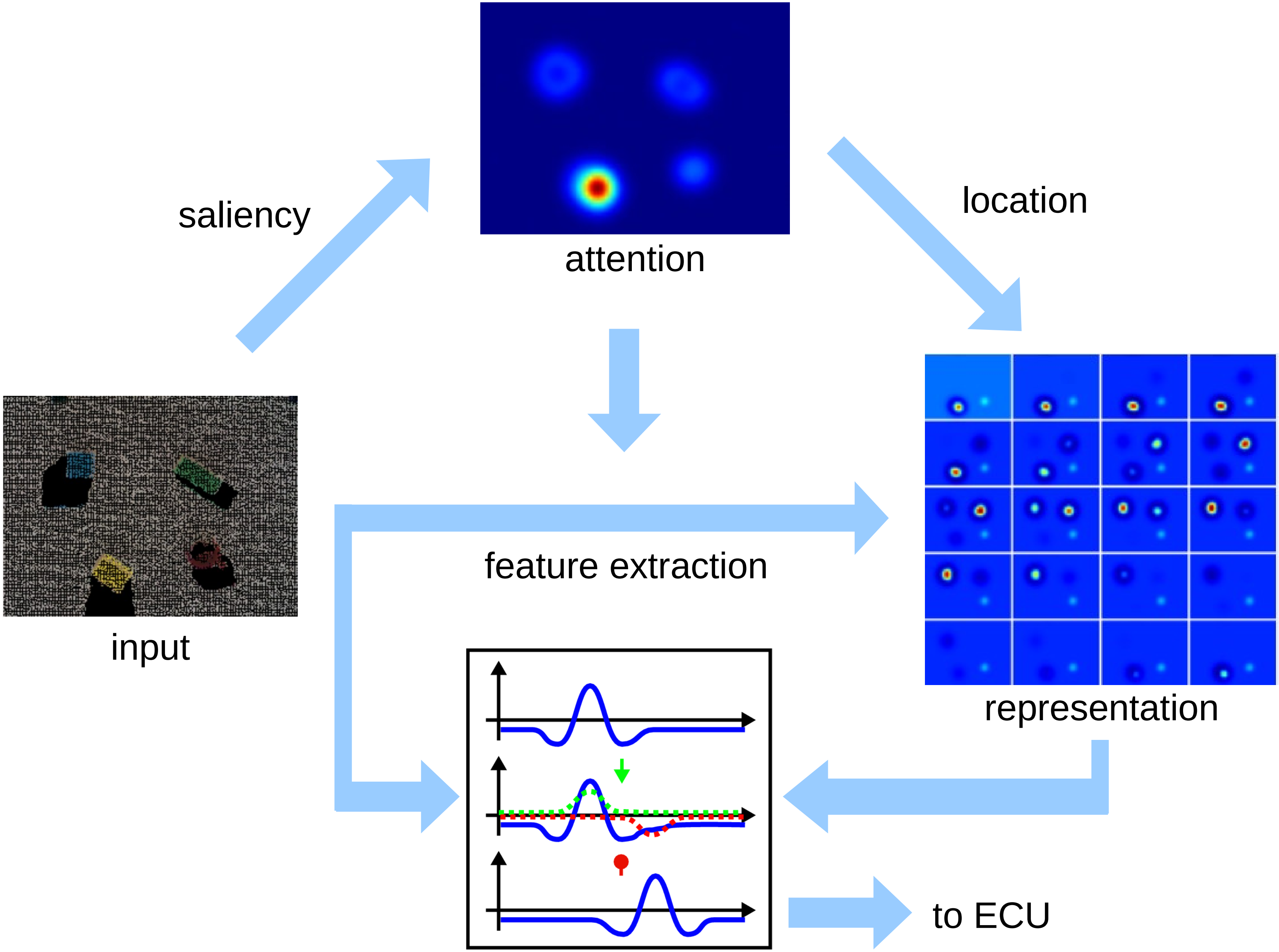
location

input

feature extraction

representation

Autonomy of Exploration



saliency

attention

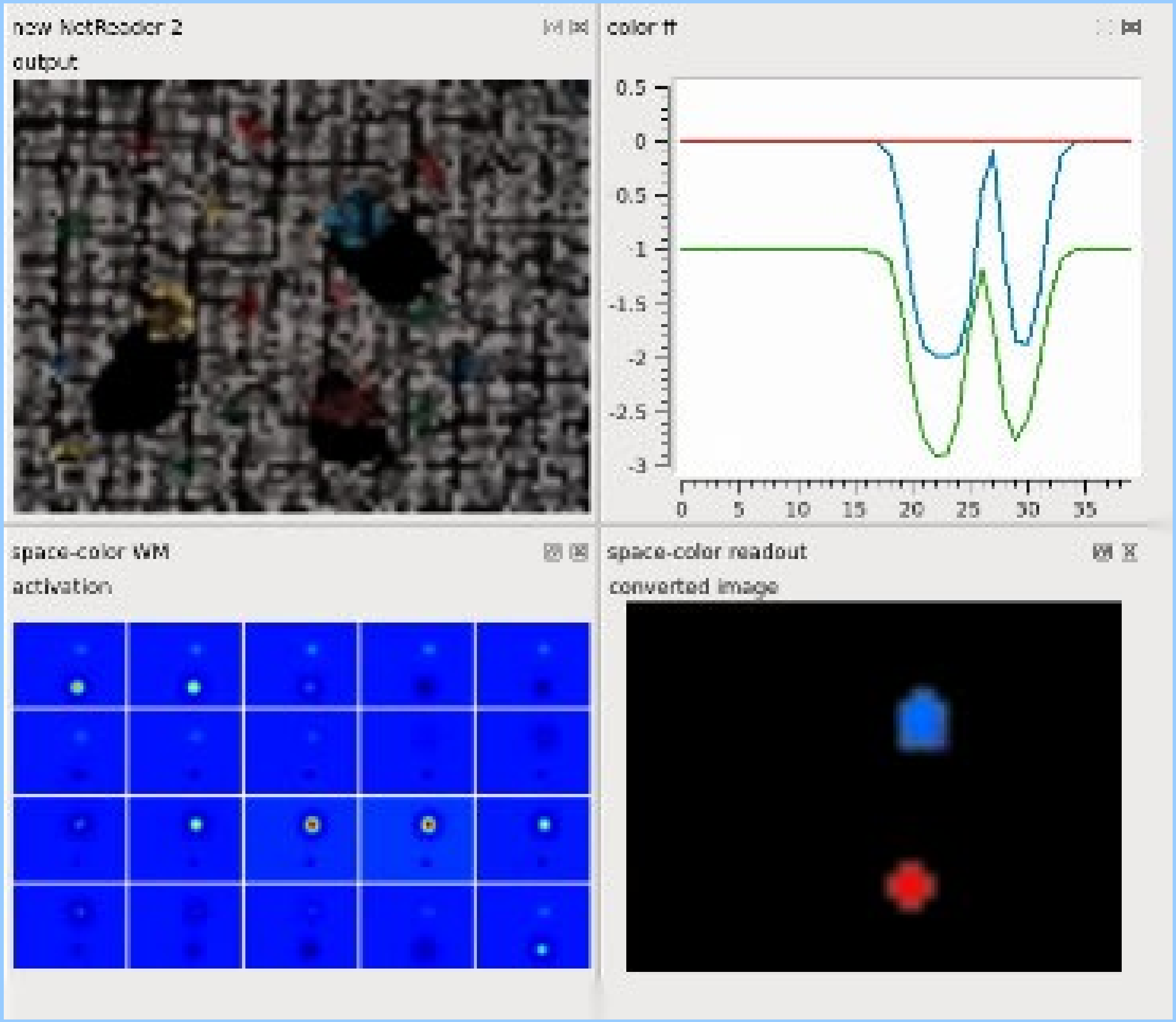
location

input

feature extraction

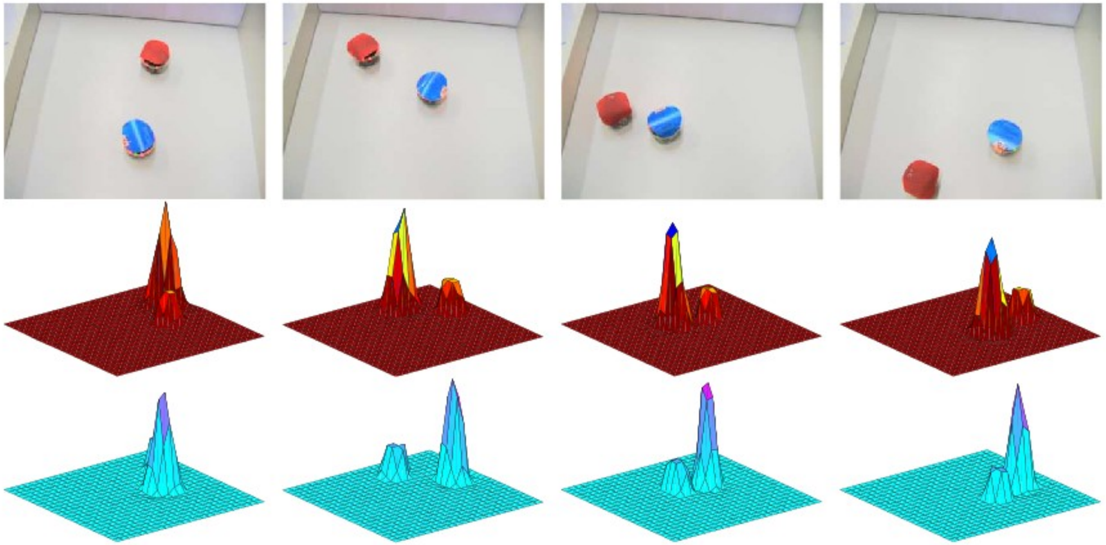
representation

to ECU


new NetReader 2
output
color ff
0.5
0
0.5
-1
-1.5
-2
-2.5
-3
0
5
10
15
20
25
30
35
space-color WM
activation
space-color readout
converted image

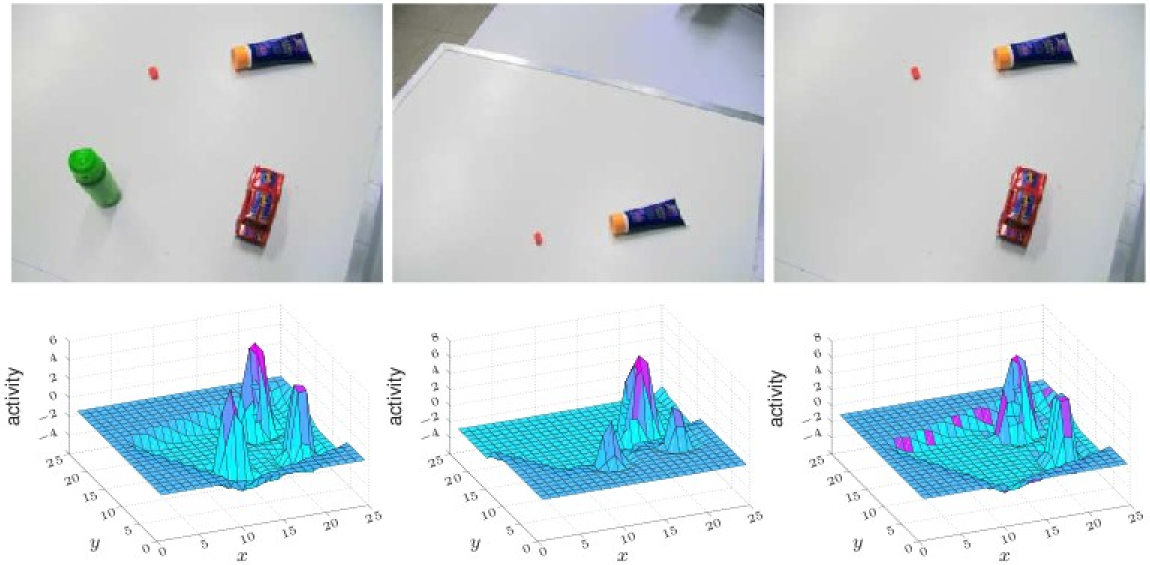
Video

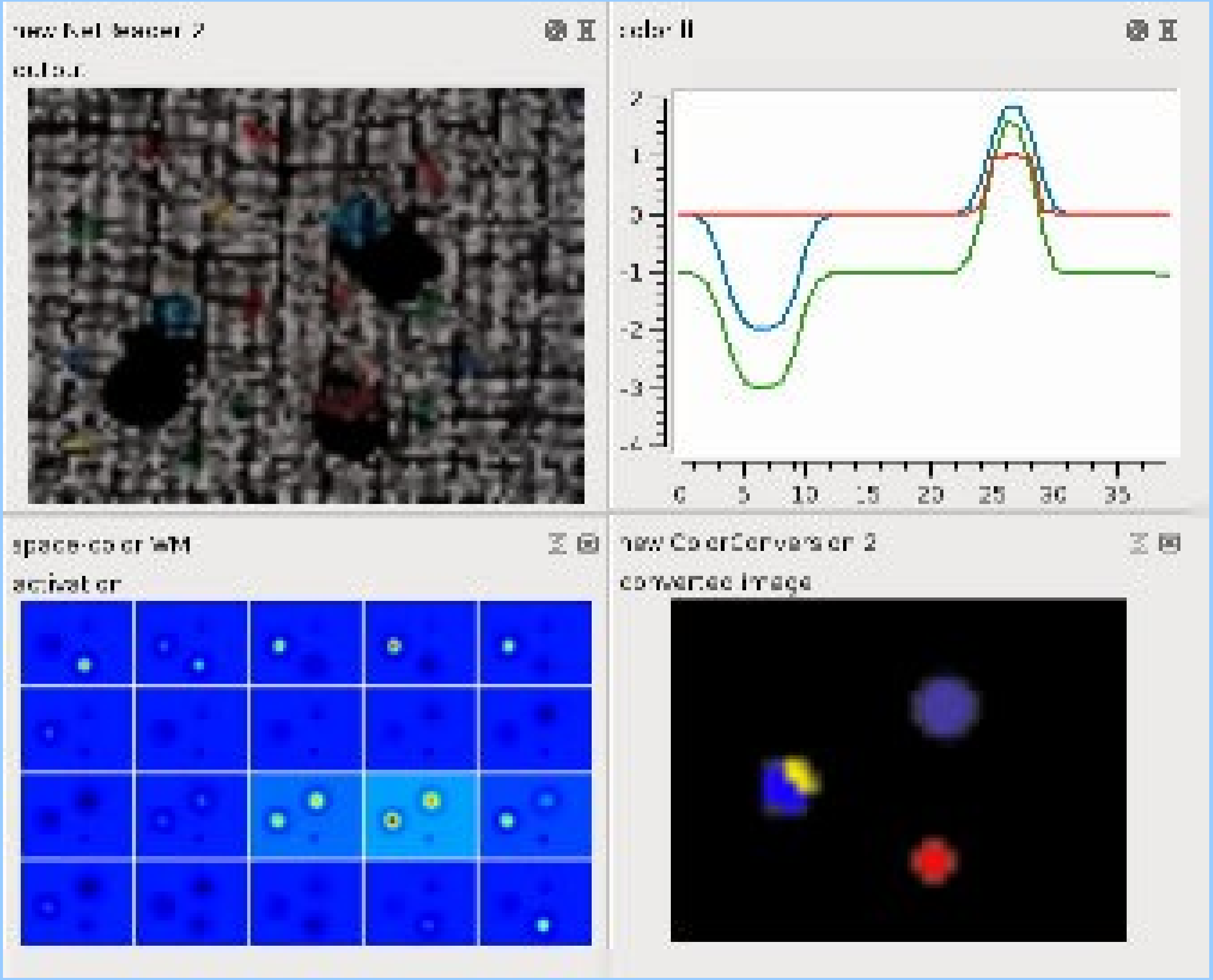
Maintenance Behavior



tracking

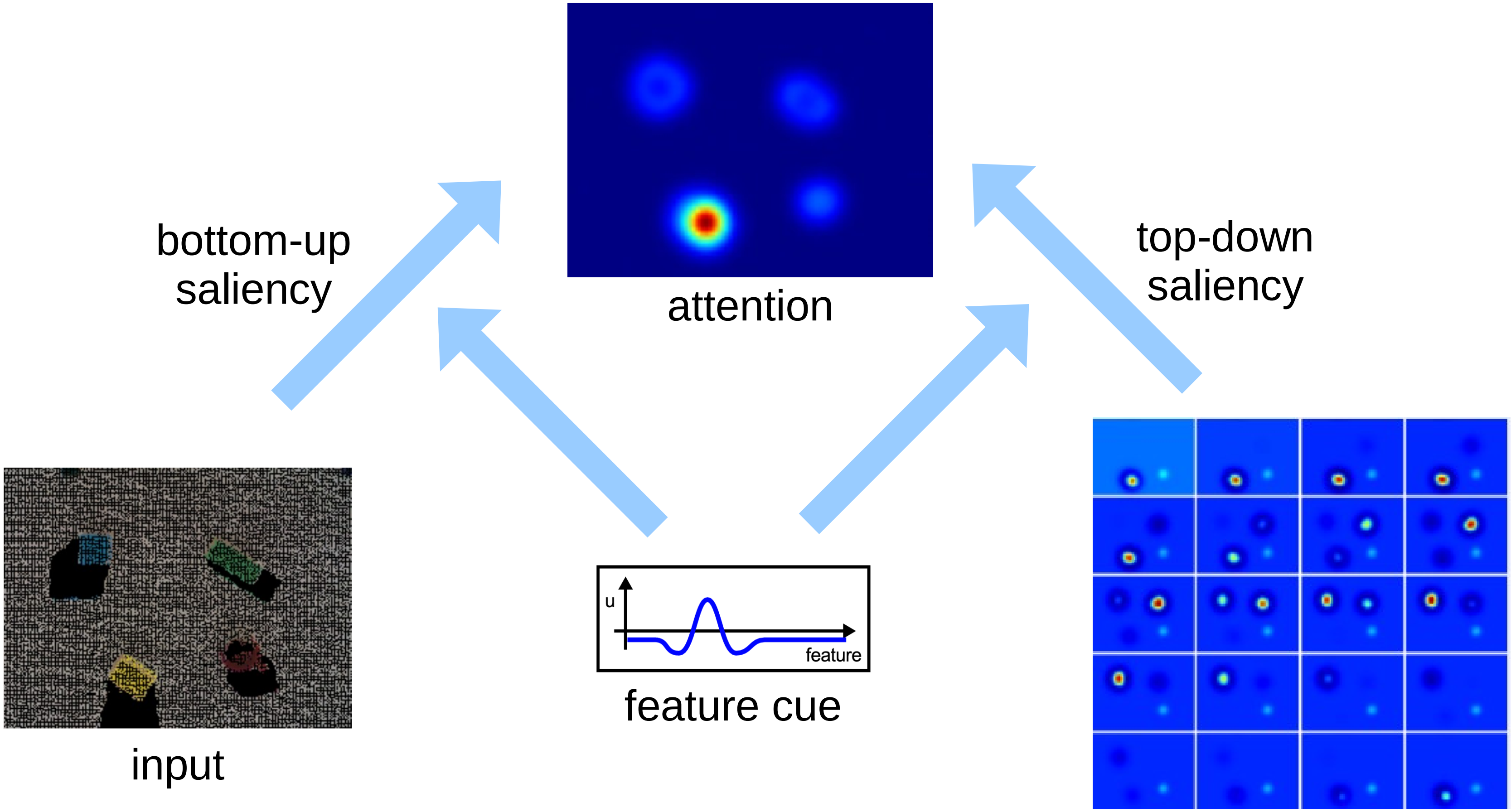
working memory
vs.
updating




new NetReader 2
output
color II
2
1
0
-1
-2
-3
-4
0
5
10
15
20
25
30
35
space-color WM
activation
new ColorConversion 2
converted image

Video

Query Behavior



attention

bottom-up saliency

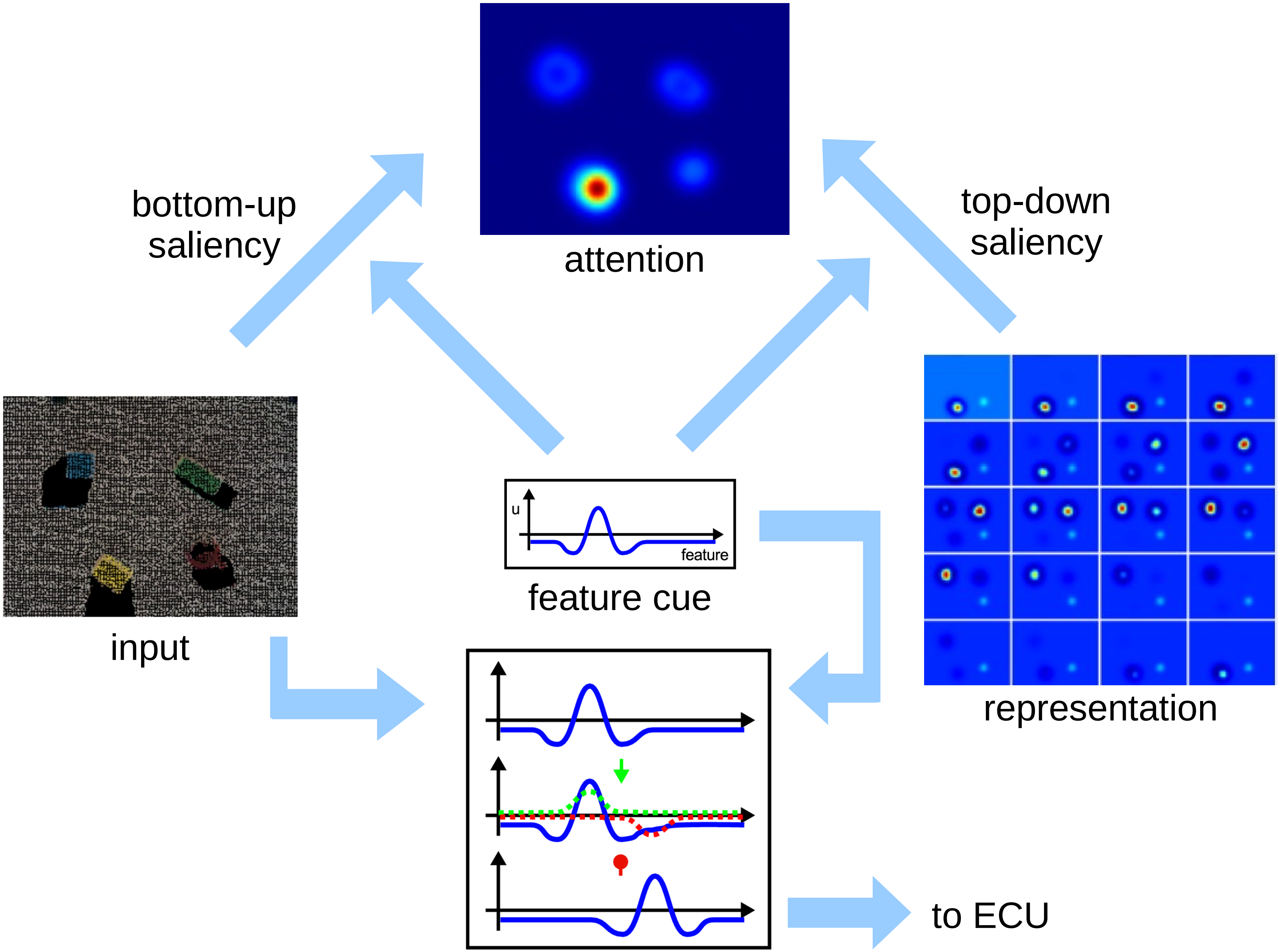
top-down saliency

input

feature cue

representation

Autonomy of Query



bottom-up saliency

attention

top-down saliency

input

feature cue

representation

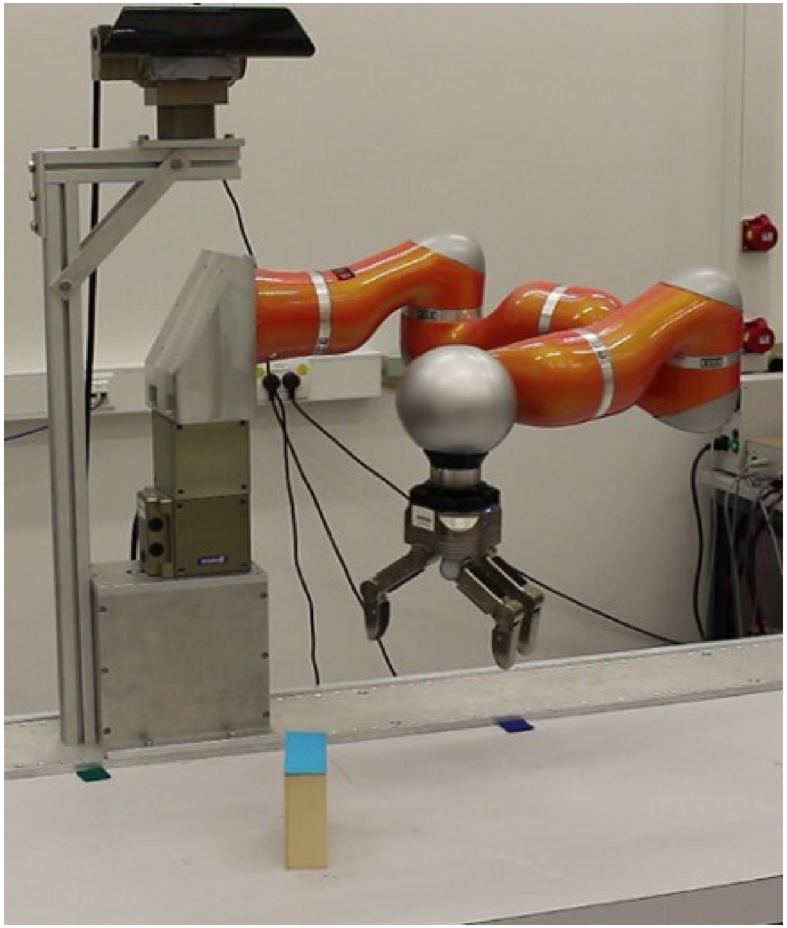
to ECU

Reaching and Grasping

Challenges

focus attention

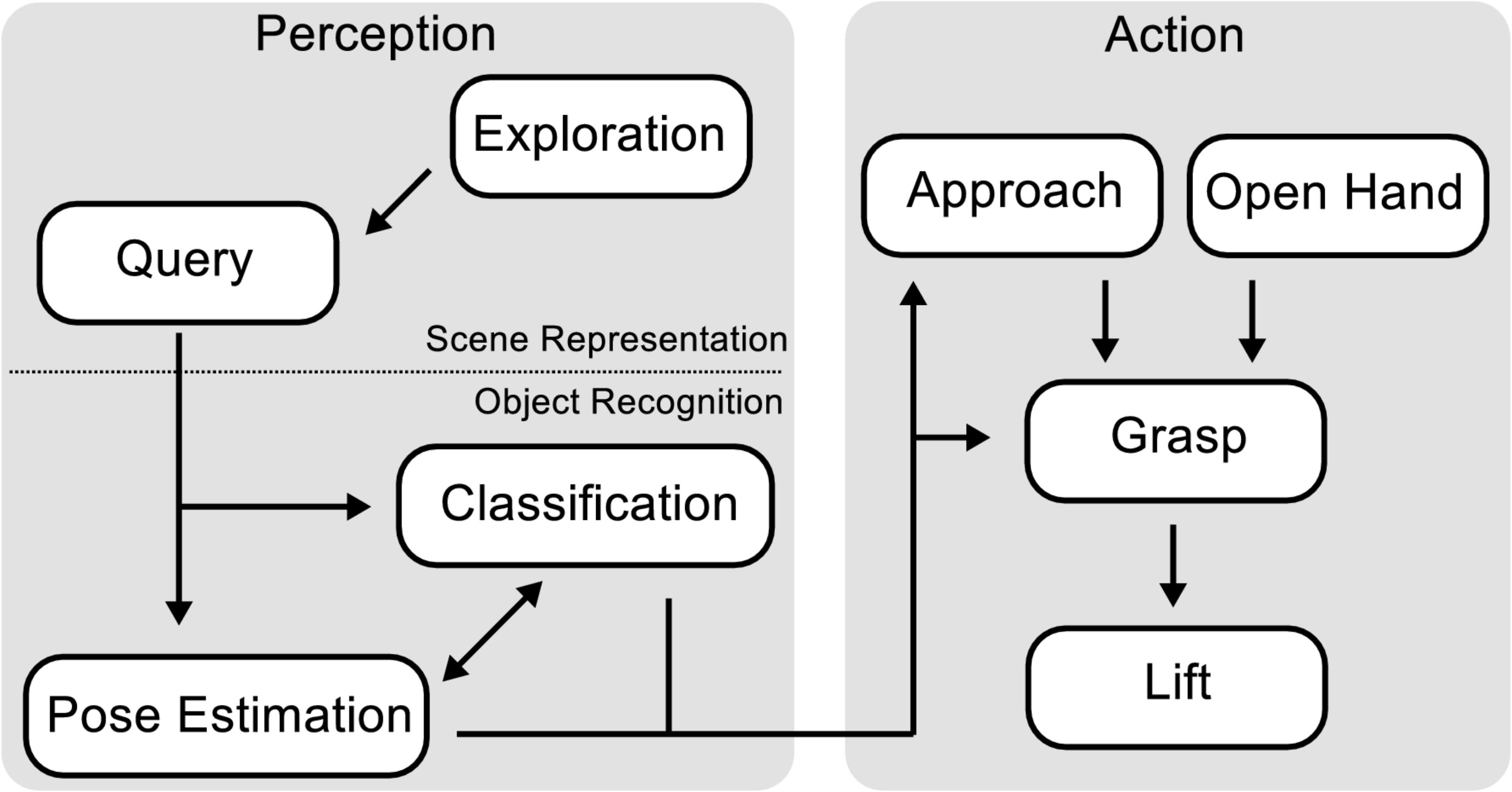
estimate shape



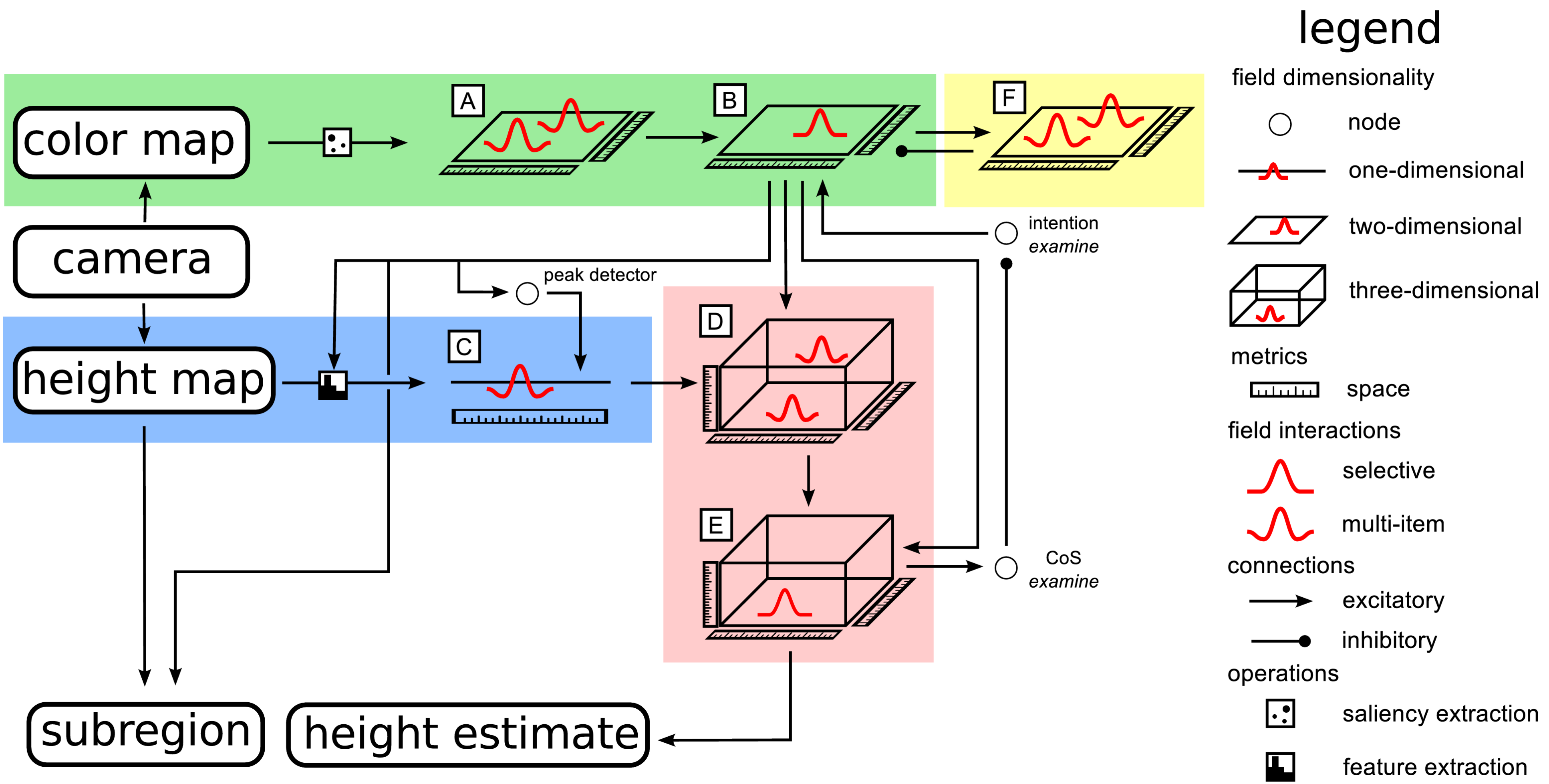
multiple behaviors

arm movement

Behaviors Involved in Grasping



From Camera to a Height Map



color map

camera

height map

subregion

height estimate

peak detector

intention
examine

CoS
examine

legend

field dimensionality

○ node

one-dimensional

two-dimensional

three-dimensional

metrics

space

field interactions

selective

multi-item

connections

excitatory

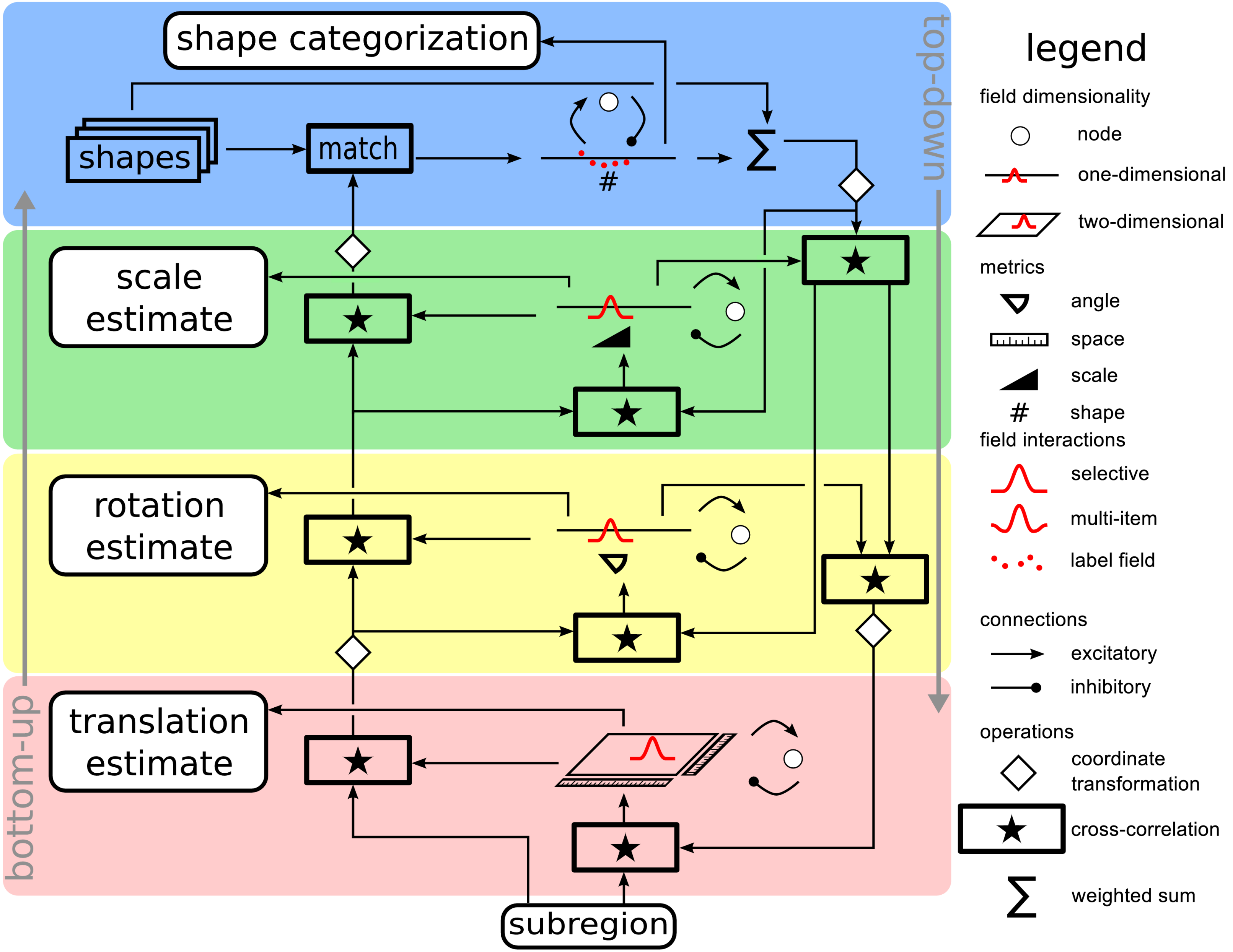
inhibitory

operations

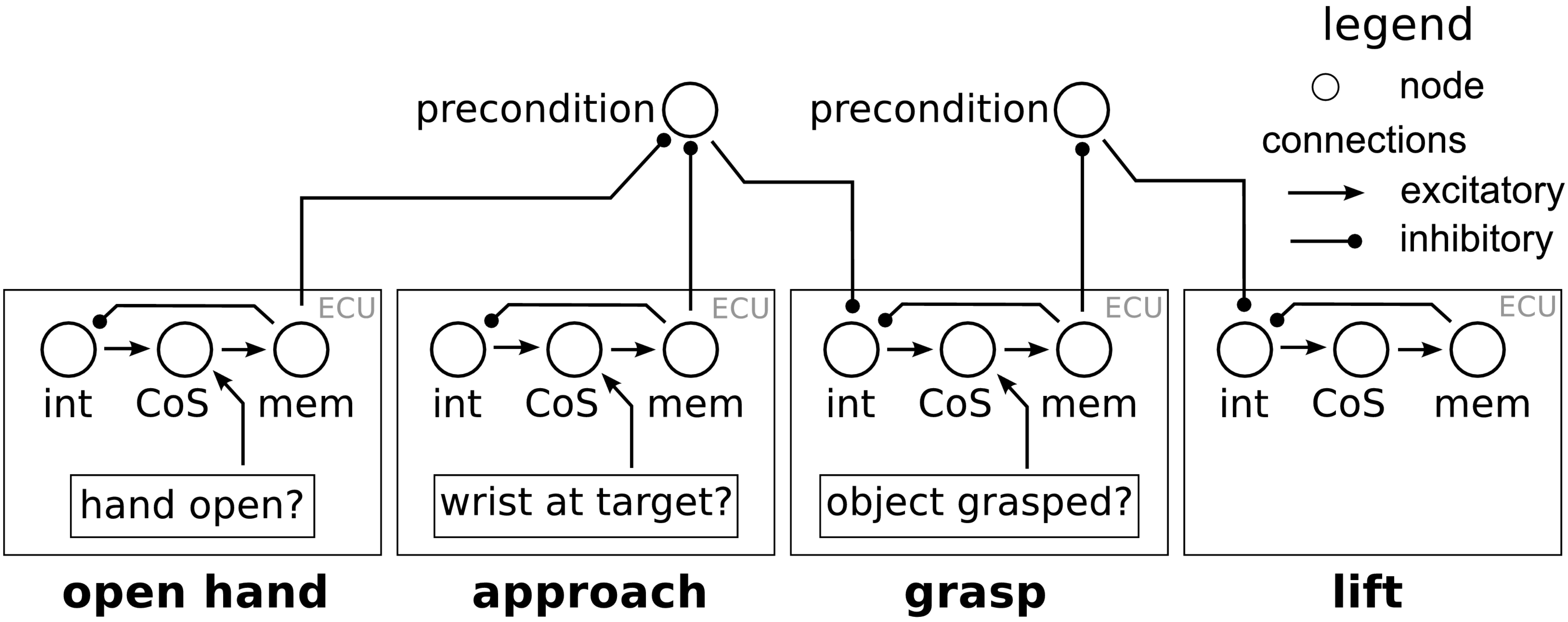
saliency extraction

feature extraction

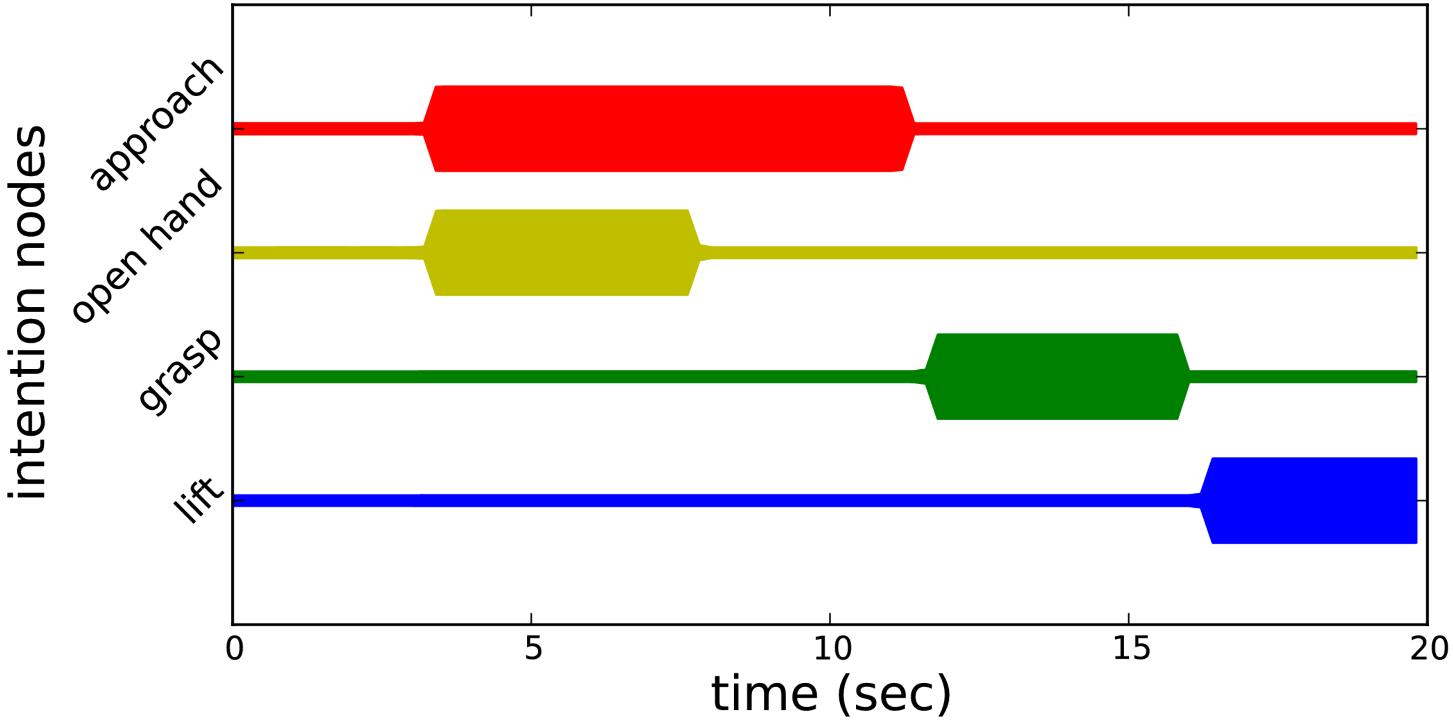
From Height Map to Grasp Parameters

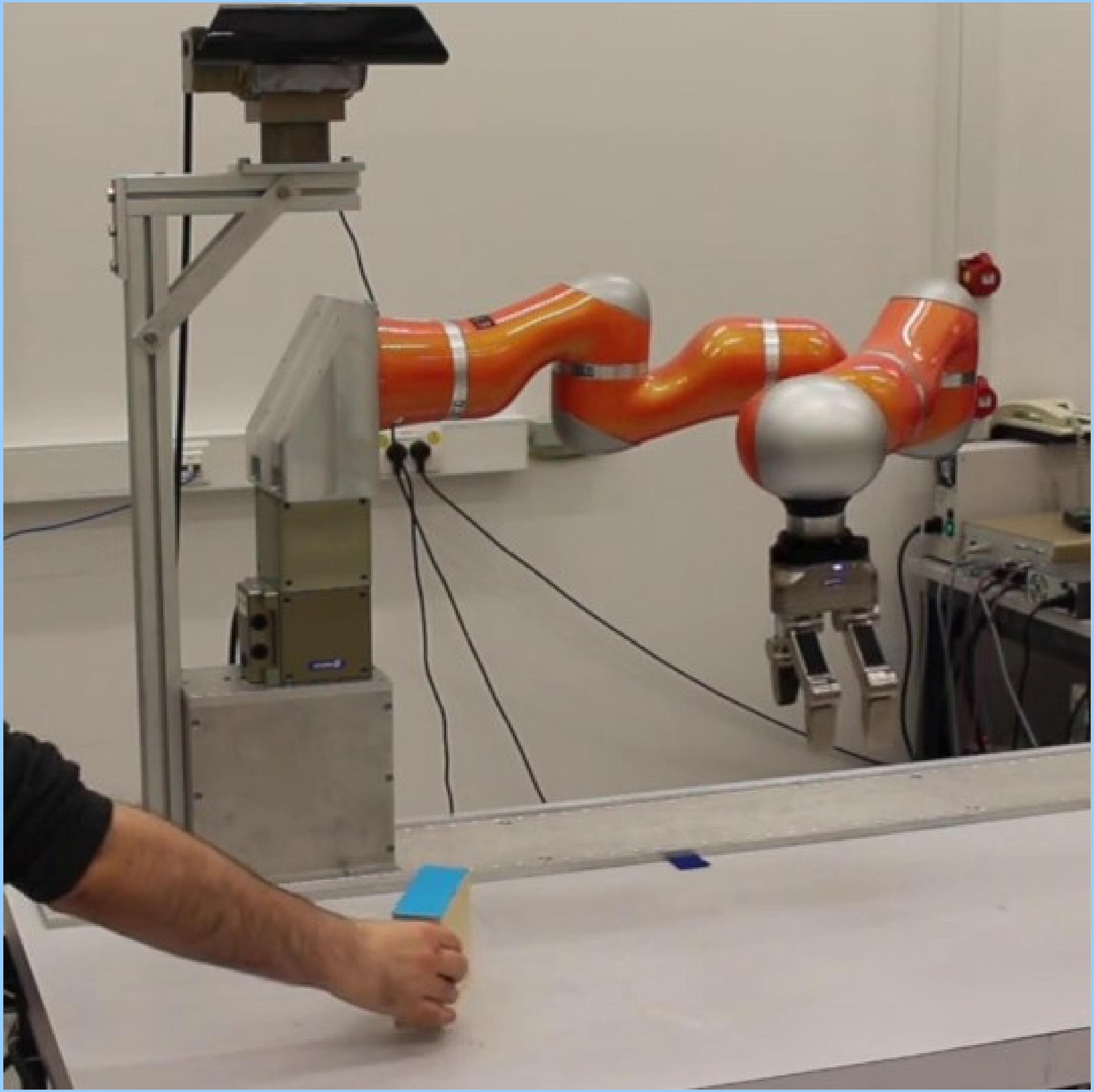


Grasp Execution



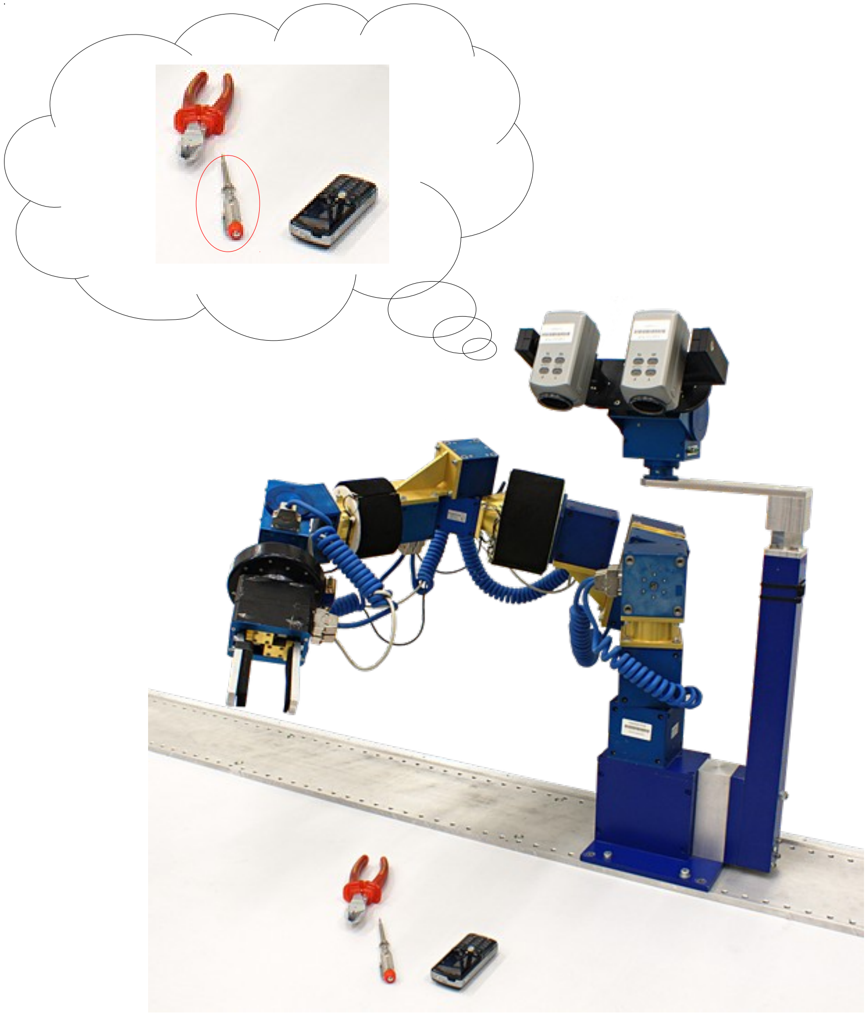
Grasp Execution





Video

Take-home Message



- exploration, maintenance and query are the core behaviors of scene representation
- change detection is a driving force for autonomy
- integration with other DFT architectures yields complex behaviors such as grasping
- integration is facilitated by DFT framework

Thanks for your attention!