From Computer to Robotic Vision: just a step *or* a leap forward?

Lazaros Nalpantidis

Computer Vision and Active Perception Lab.
Centre for Autonomous Systems

KTH - Royal Institute of Technology
Stockholm, **Sweden**

Centre for Autonomous Systems

KTH VETENSKAP OCH KONST ROYAL INSTITUTE OF TECHNOLOGY

30 May 2012 - Thessaloniki, Greece
Information Technologies Institute - Centre of Research and Technology Hellas

OUTLINE

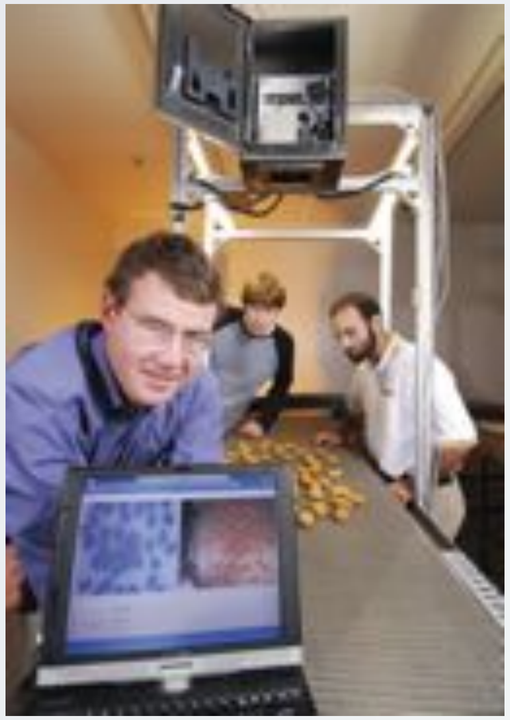
- what’s the difference?
- from computers to Robots
    - goals / priorities / challenges / adaptation
- new possibilities
    - active perception
    - cognitive Vision/Robotics
- some thoughts

what's the difference?

robot vision as the
"daughter"
of computer vision

- younger
- “simple” methods
- “simple” problems
- happy with “small” achievements
- still not sure how to use its capacities

CV | RV



RV

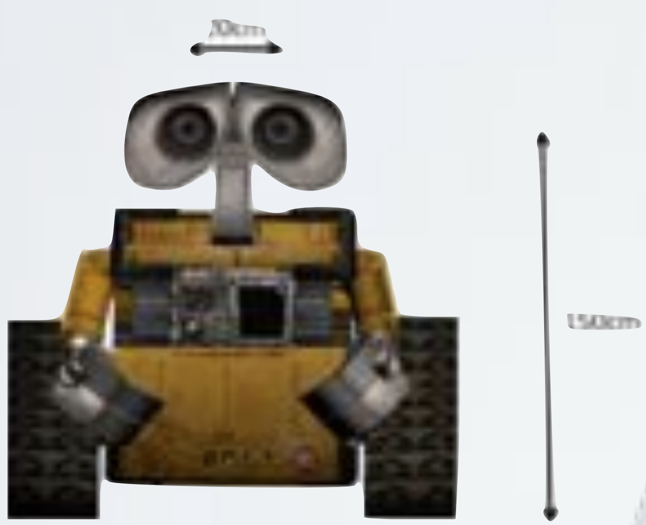
from computers to Robots

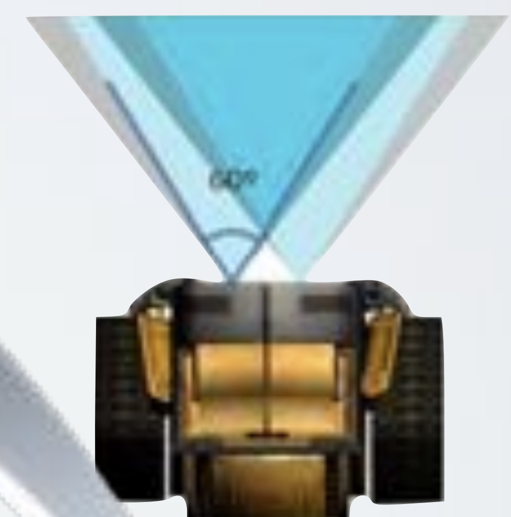
+

robot + vision

robot + vision

+

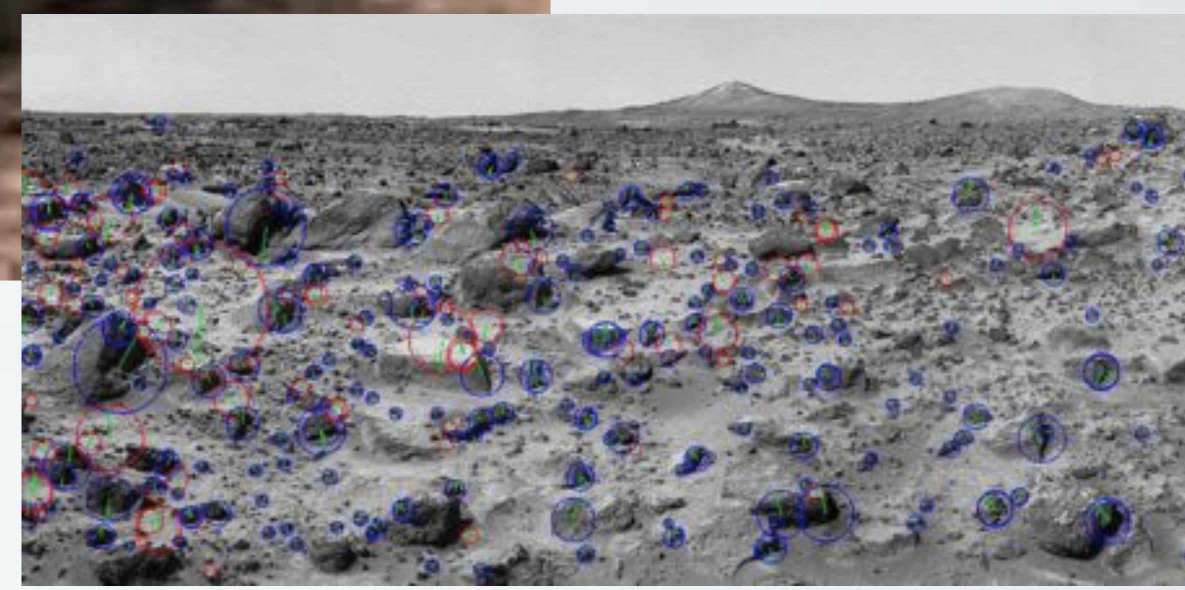




ro on

GOALS

- categorization
- recognition
- **navigation**
- **manipulation**

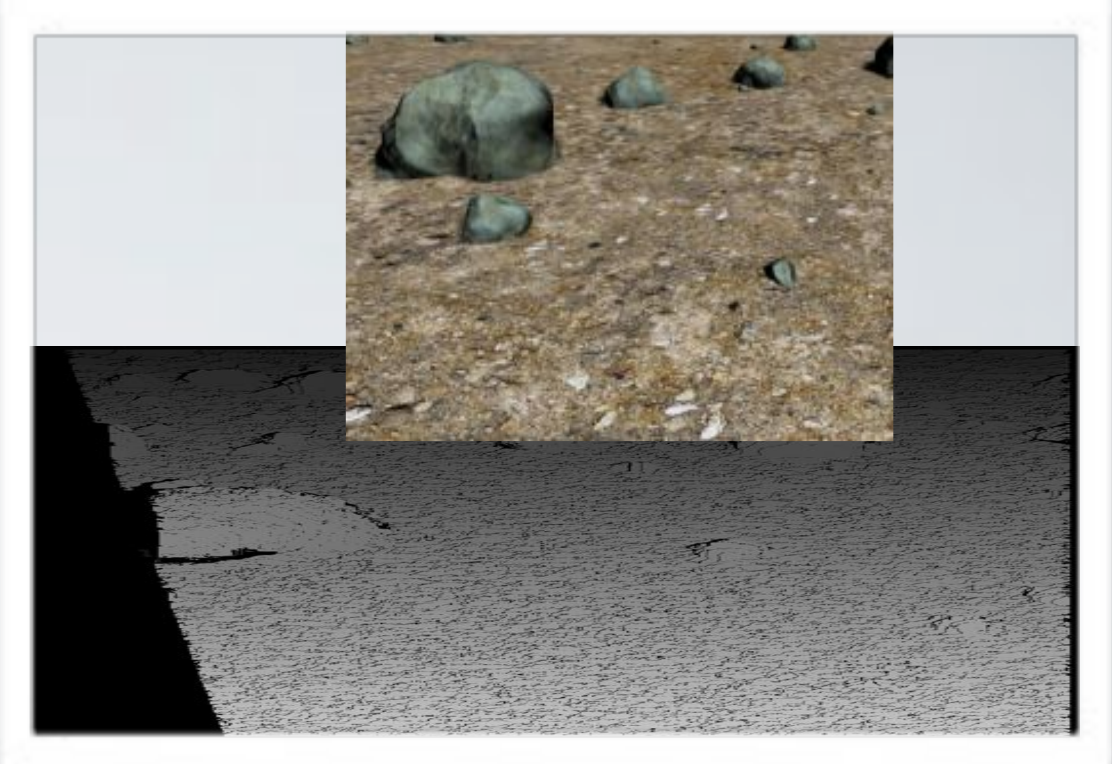


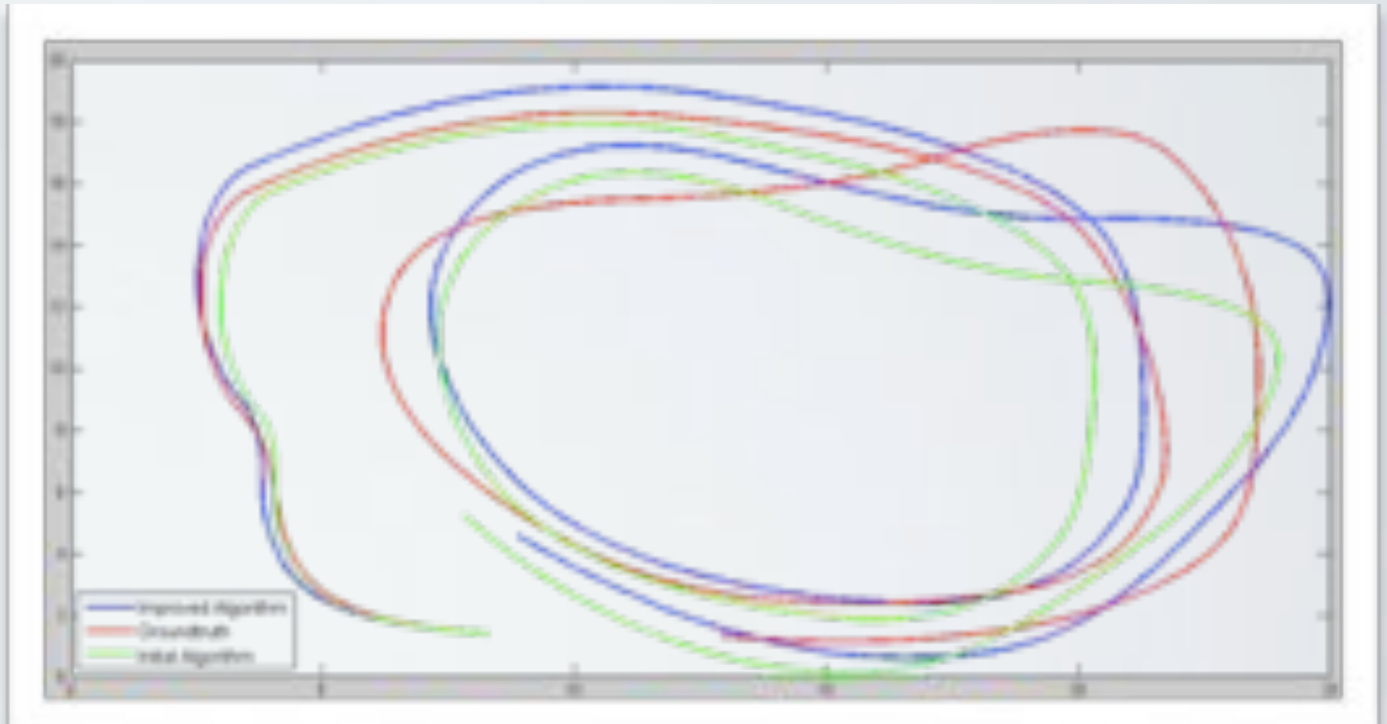
SPARTAN

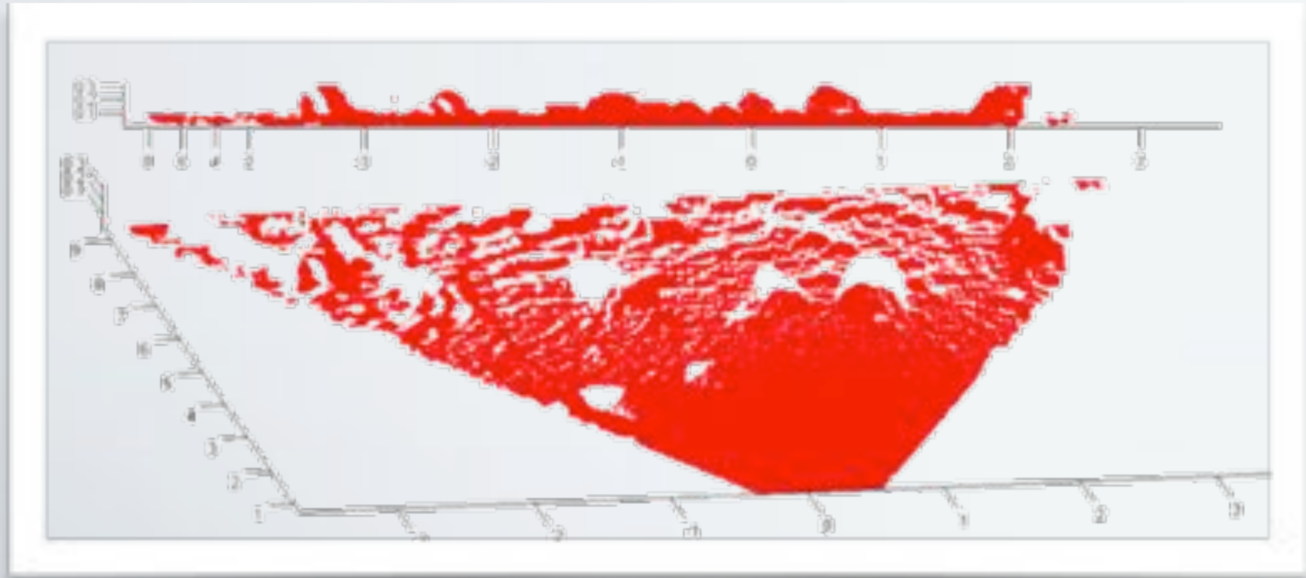
SPAring Robotics Technologies for Autonomous Navigation

SPACE EXPLORATION

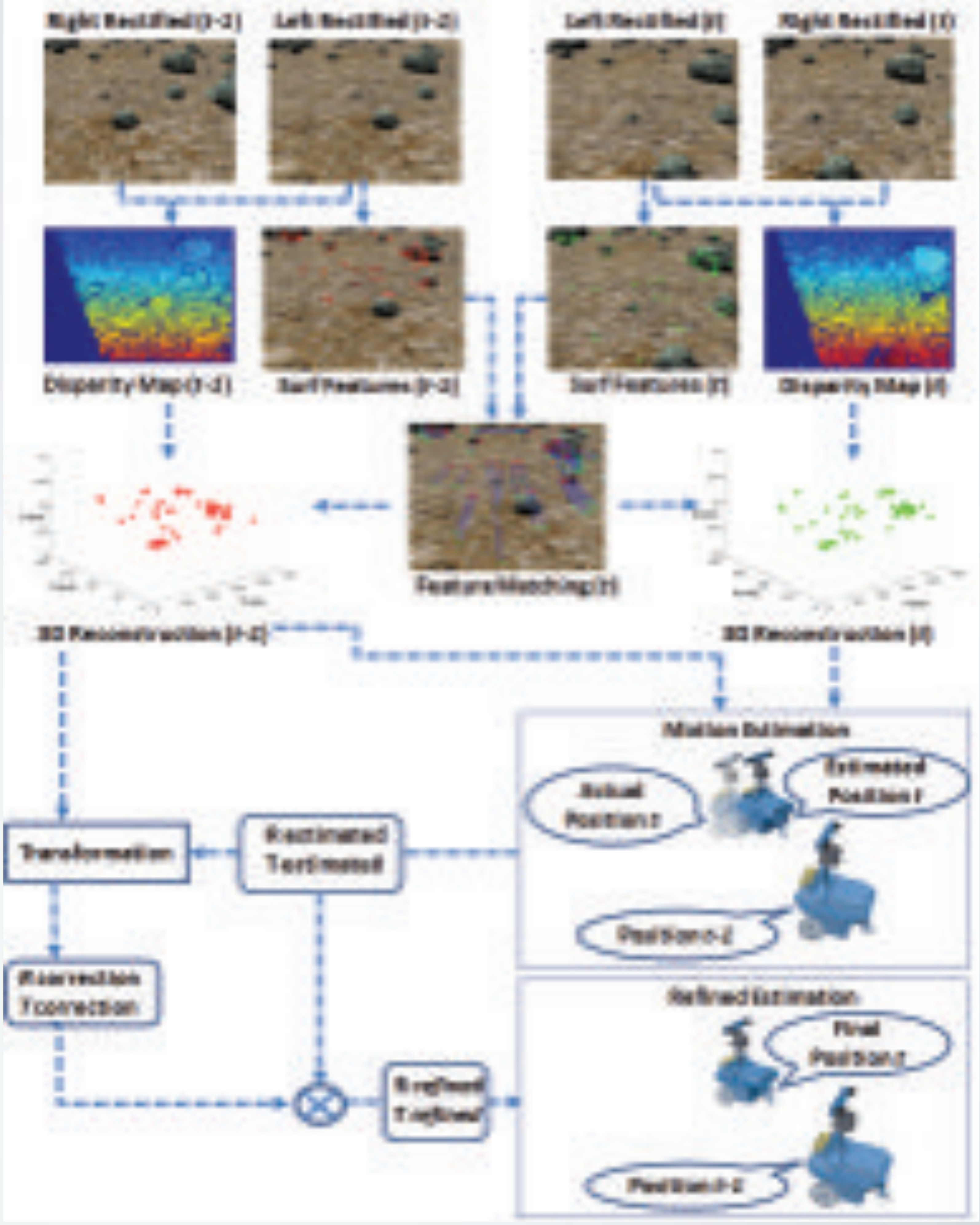
- Stereo Vision
- Visual Odometry
- 3D reconstruction







- Visual Odometry



PRIORITIES

- real-time
- robustness
- reliability

- employing bio-inspired & perceptual organization rules in stereo depth estimation

- Circular support region

- Adaptipe support weights aggregation (**Gestalt** laws of proximity, similarity, and continuity)

$$distance(x', y')|_{x,y} = \sqrt{(x - x')^2 - (y - y')^2}$$

$$dissimilarity(x', y')|_{x,y} = \frac{1}{3} \sum_{c \in R,G,B} |I_c(x, y) - I_c(x', y')|$$

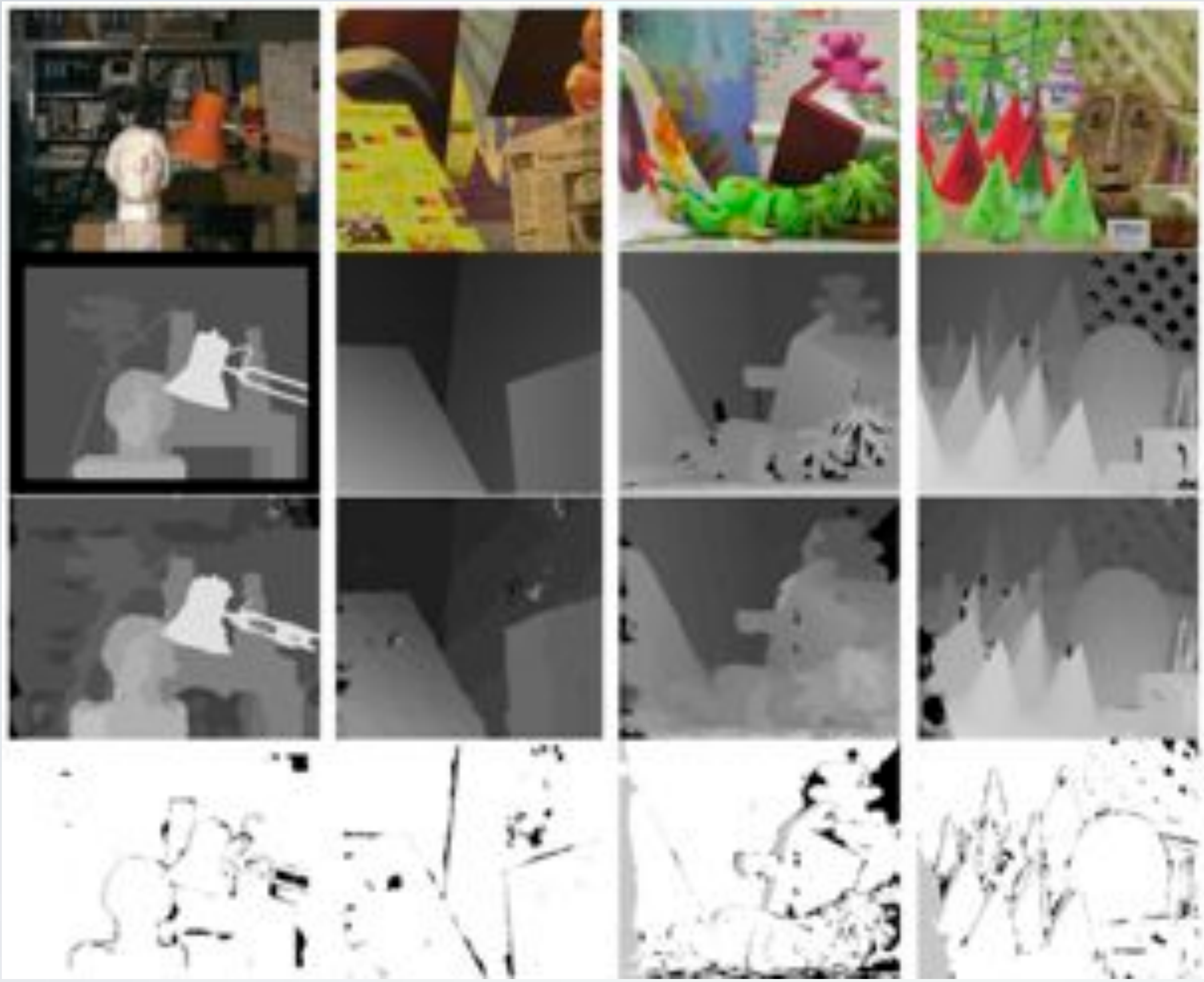
$$discontinuity(x', y', d)|_{x,y,d} = \frac{AD(x', y', d)}{\max(AD)}$$

- Logarithmic response to stimuli (Weber-Fechner law)

$$w_{dist}(x', y', d)|_{x,y,d} = -k_1 \ln\left(distance(x', y', d)|_{x,y,d}\right)$$

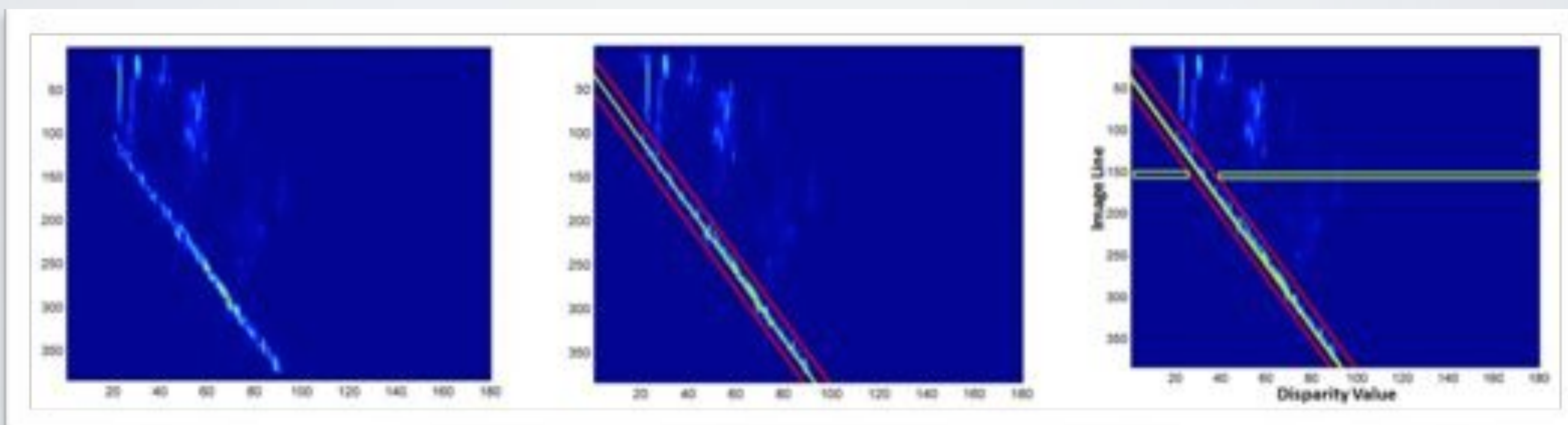
$$w_{dissim}(x', y', d)|_{x,y,d} = -k_2 \ln\left(dissimilarity(x', y', d)|_{x,y,d}\right)$$

$$w_{discon}(x', y', d)|_{x,y,d} = -k_3 \ln\left(discontinuity(x', y', d)|_{x,y,d}\right)$$



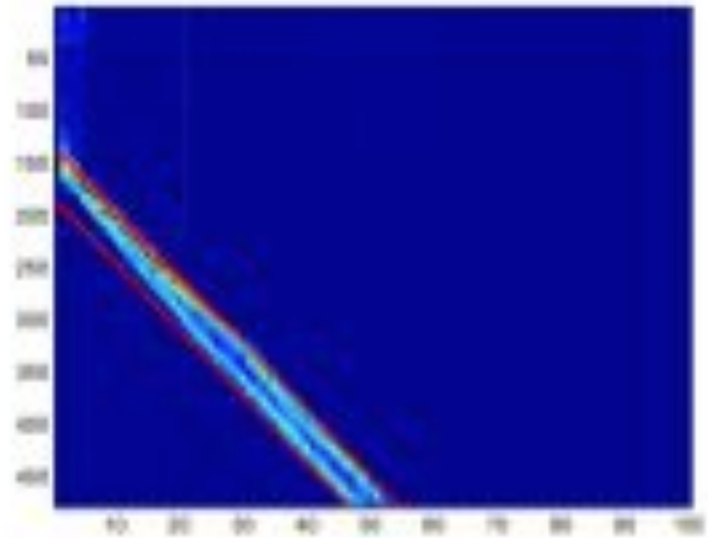
	nonocc		all		disc	
	error	variation	error	variation	error	variation
proposed	**3.62**		**5.52**		**14.6**	
no continuity	5.19	+43.37%	7.17	+29.89%	21.7	+48.63%
no log. response	8.89	+145.58%	10.5	+90.22%	36.1	+147.26%
no circ. window	3.79	+4.70%	5.62	+1.81%	15.8	+8.22%

- using 2D histograms of depth images to detect “dominant” planes and “obstacles”

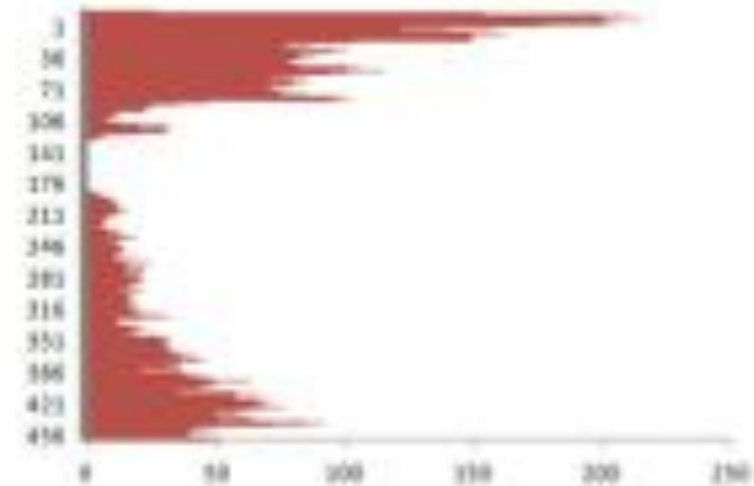


50
100
150
200
250
300
350
20 40 60 80 100 120 140 160 180
Image Line
Disparity Value

(a) Traversable image

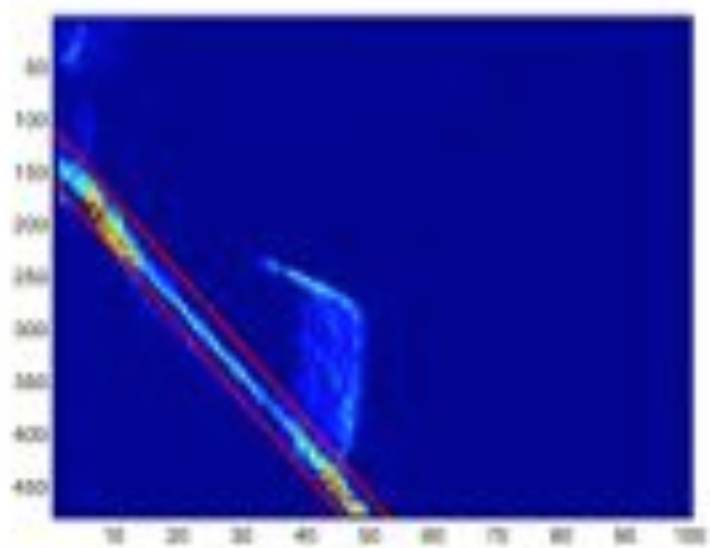


(b) V-disparity image

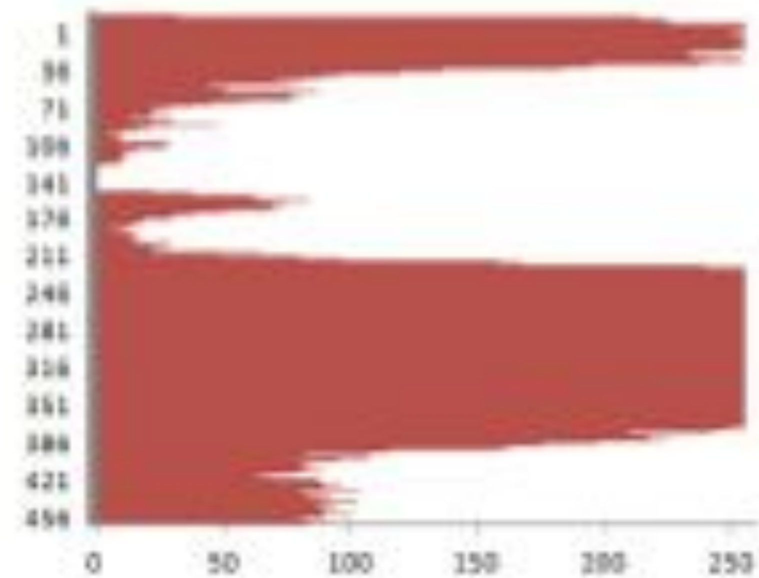


(c) Histogram of features

(d) Non-Traversable image



(e) V-disparity image



(f) Histogram of features

CHALLENGES

- simplicity of calculations
- non-ideal lighting
- de-calibrated input

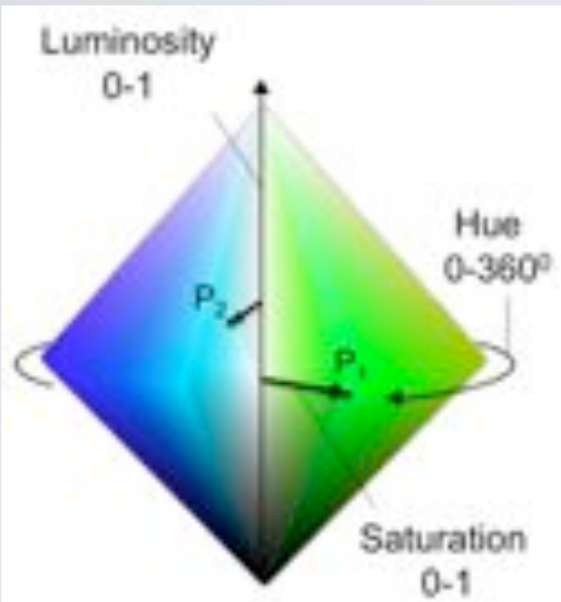
View Finder

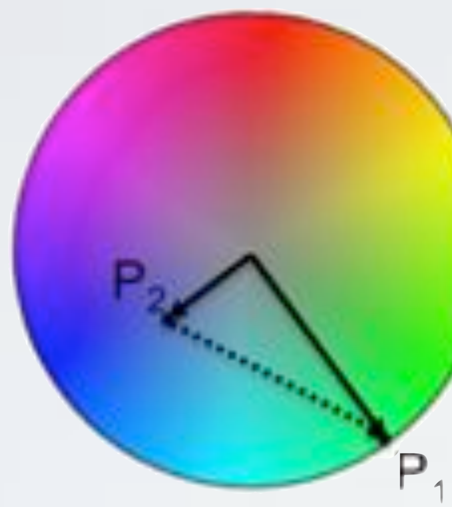
Information Society Technologies

Dealing with non-Ideal Lighting

Dealing with non-Ideal Lighting

Luminosity-Compensating Dissimilarity Measure



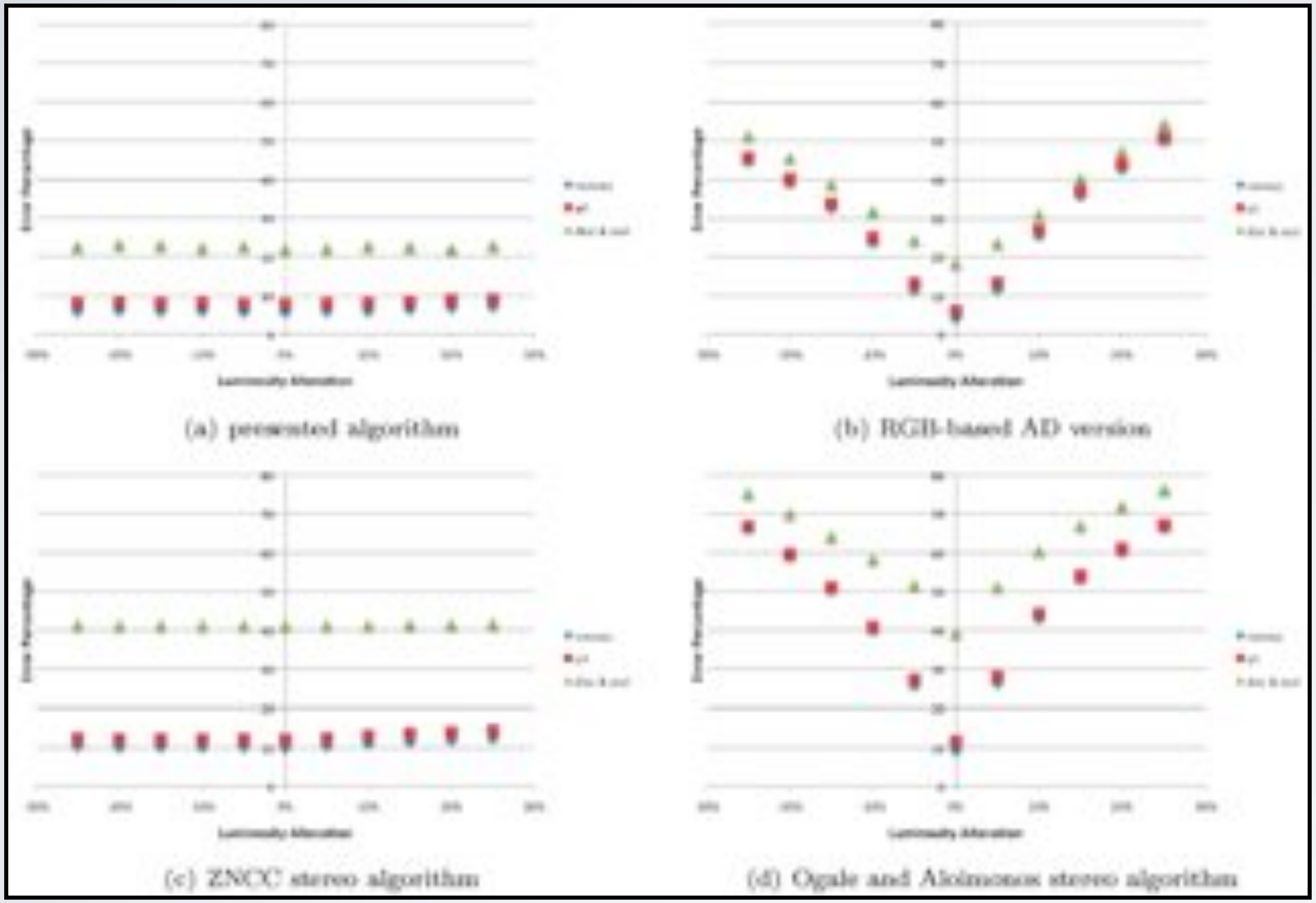


Definition of LCDM
in the
HSL color space

$$\mathbf{P}_k = S_k e^{iH_k}$$

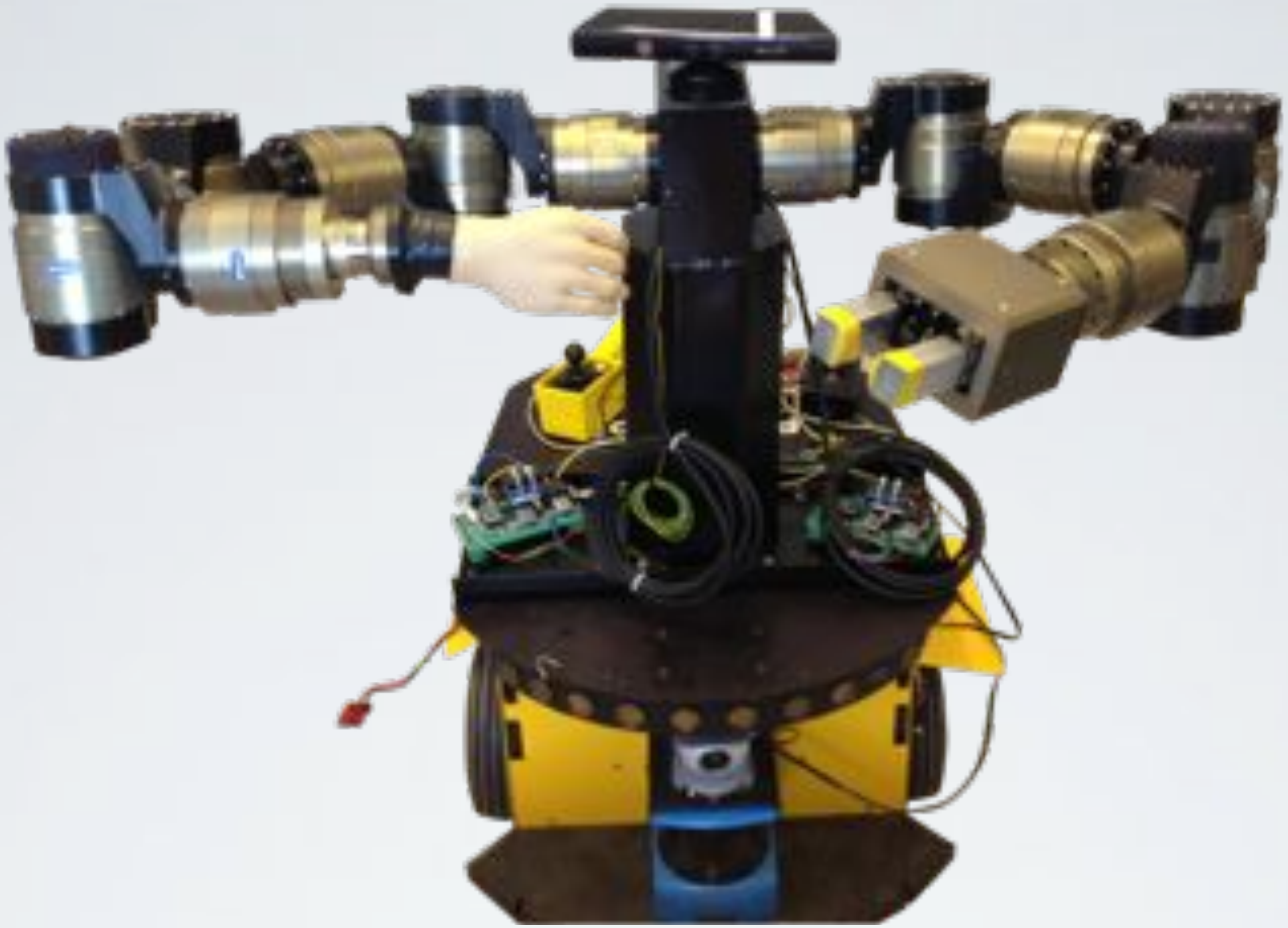
$$\begin{aligned} LCDM_{P_1,P_2} &= |\mathbf{P}_1 - \mathbf{P}_2| \\ &= \left| S_1 e^{iH_1} - S_2 e^{iH_2} \right| \\ &= \sqrt{{S_1}^2 + {S_2}^2 - 2S_1 S_2 \cos(H_1 - H_2)} \end{aligned}$$

Dealing with non-Ideal Lighting
LCDM-based Algorithm



Swedish
Research Council

Vetenskapsrådet



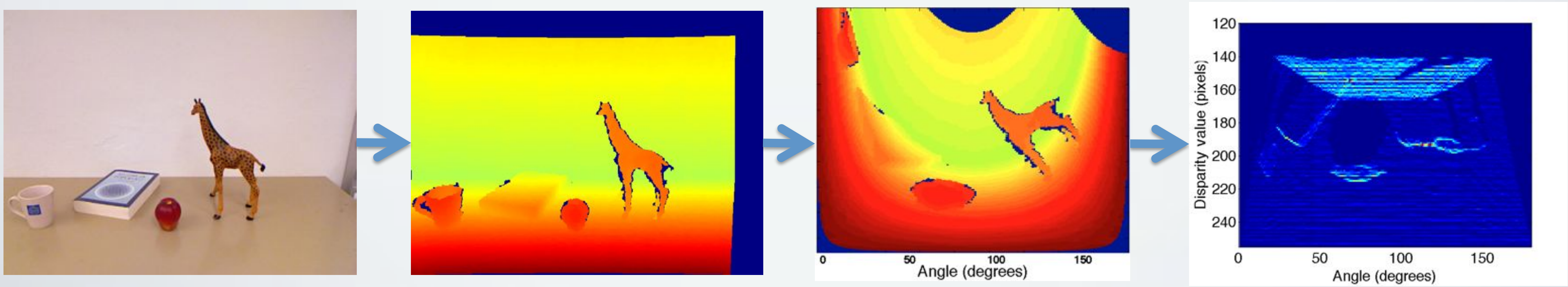
our DAM robot: moves & manipulates

Theta-disparity

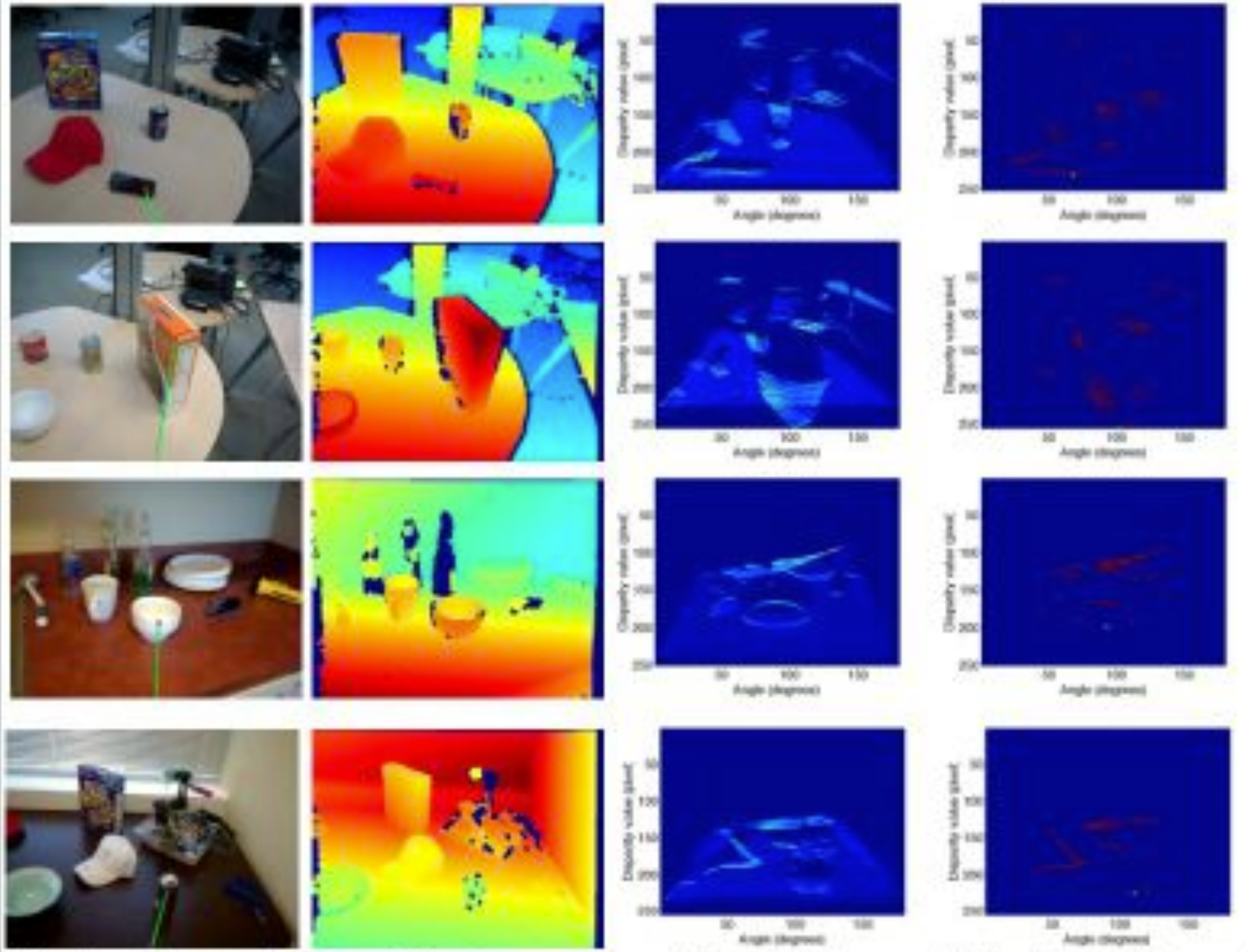
- radial arrangement of objects
- basic attention mechanism
- common treatment of
  - object detection
  - obstacle avoidance
  - ...

(a) A synthetic scene of the considered scenari

(b) View of the Kinect sensor



(a) (b) (c) (d)

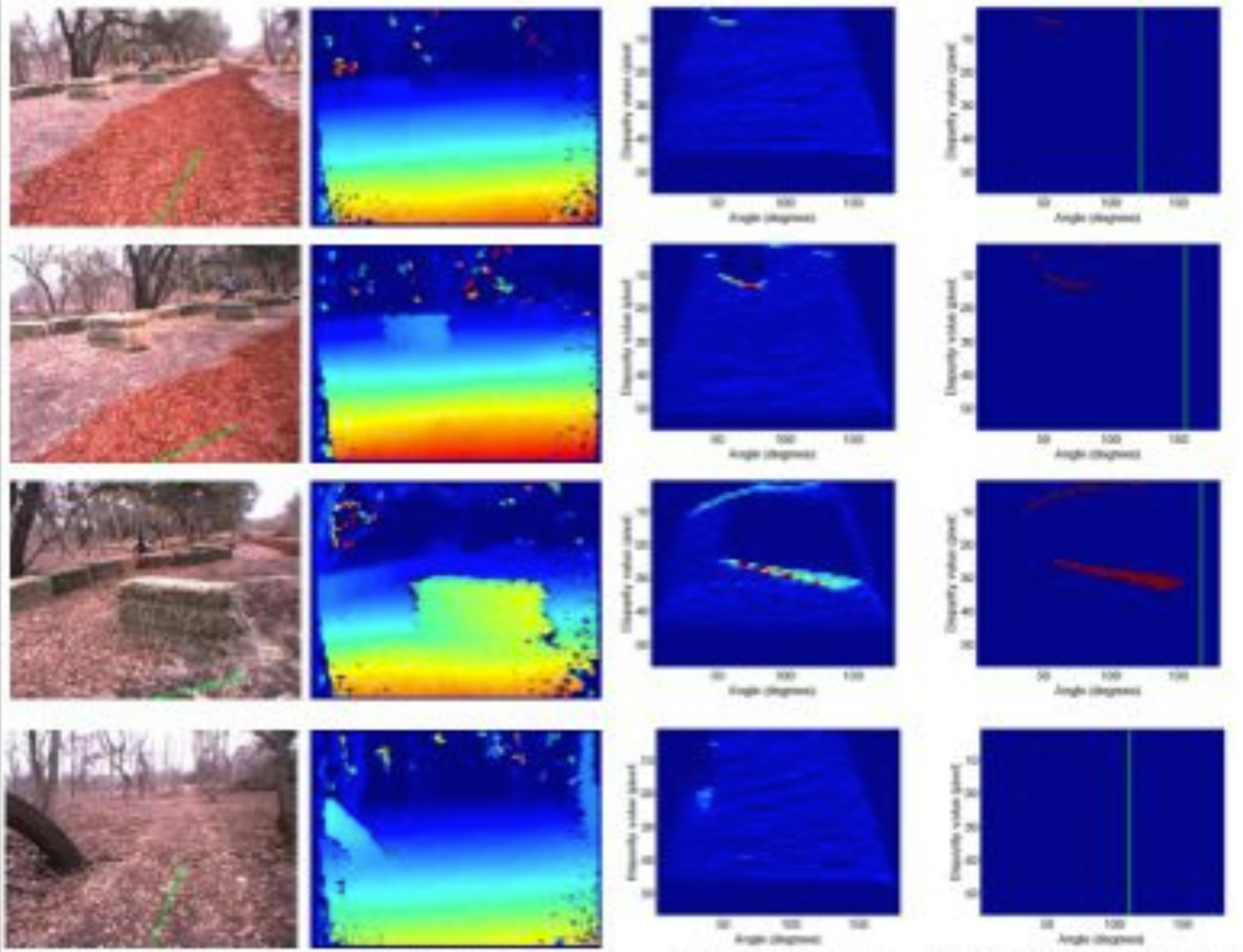


(a) Input image with suggested direction

(b) Disparity map

(c) Theta-disparity image

(d) Binary theta-disparity image with the closest trace highlighted



(a) Input image with suggested direction

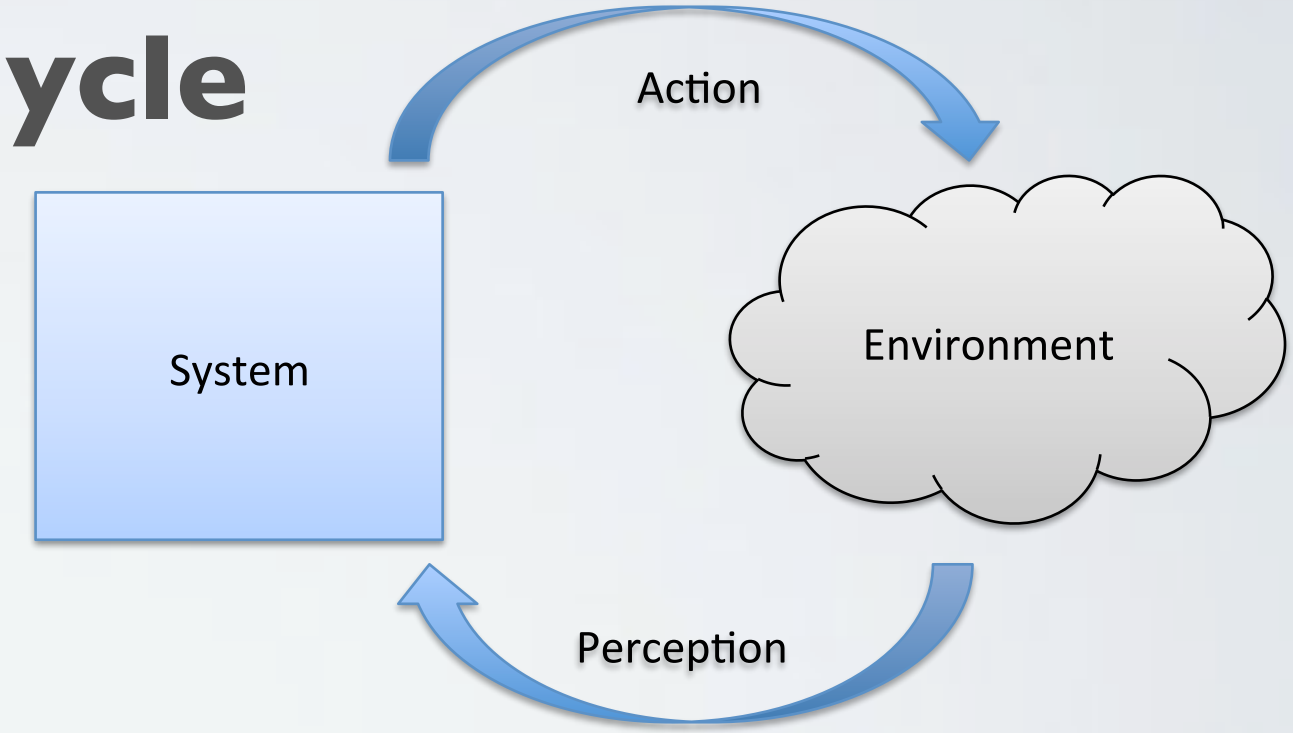
(b) Disparity map

(c) Theta-disparity image

(d) Binary theta-disparity image with suggested direction new possibilities

- a seeing system/robot uses **Vision** to control its **Motions** **(Perception** precedes **Action)**

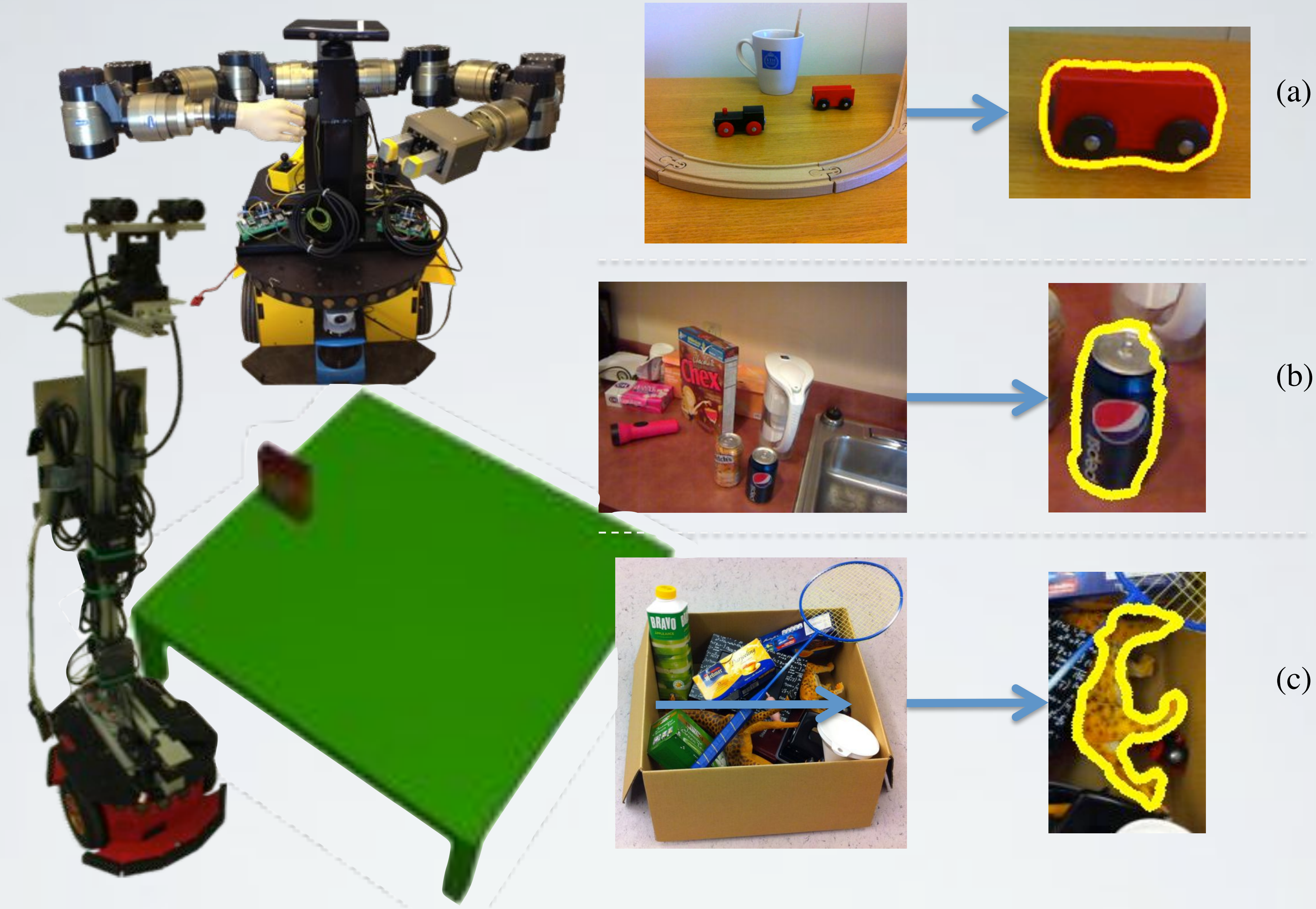
- **Perception-Action Cycle**



- a system/robot might use **Motion** to control its **Visual** input **(Action** precedes **Perception)** ...

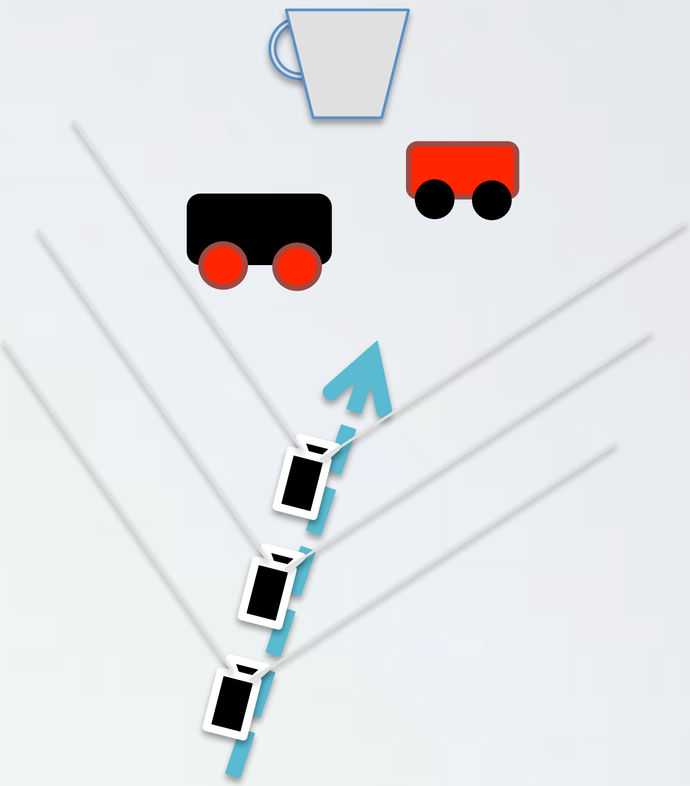
KTH head "Charlie"
in 90's

ACTIVE VISION

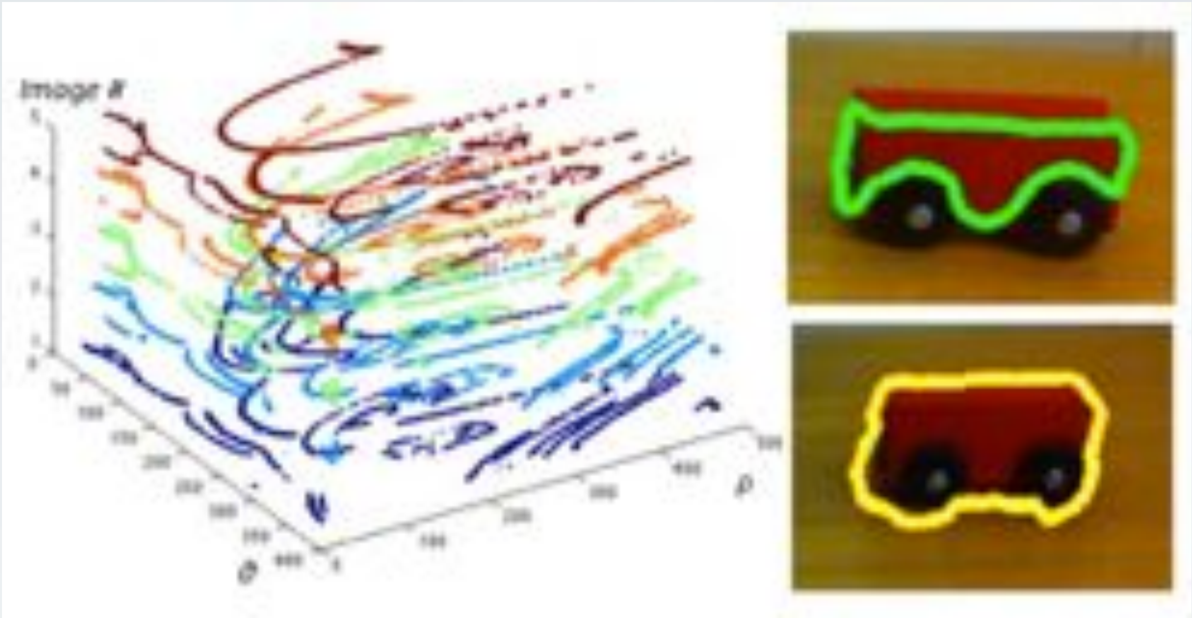


Object Segmentation

- Taking **time** into account
- Agent moving in the scene
- Multiple observations improve/ simplify the segmentation
- Object segmentation by active perception



Object Segmentation



Object Segmentation

eSMCs | Extending Sensorimotor Contingencies to Cognition

SEVENTH FRAMEWORK PROGRAMME

Sensorimotor Contingency theory

O'Regan & Noe, 2001

- Philosophers
- Psychologists
- Neurophysiologists
- Pathophysiologists
- Neuroscientists
- Roboticists

Beyond **perception**, complex aspects of **cognition** are grounded in sensorimotor interactions

“Red” is knowing the structure of the changes that “red” causes.

e.g.
a system can only truly understand what a sponge is if it can experience the sponginess by squeezing the object and observing the sensory consequences. Having this understanding allows the system to grasp and use the sponge correctly

O'Regan and Noë, “A sensorimotor account of vision and visual consciousness”, Behavioral and Brain Sciences, vol. 24, no. 5, pp. 939-1031, 2001.

- constant self-calibration
- tolerance to changes/damages/failures
- adaption to a dynamic & ambiguous environments

some thoughts

- the transition from Computer to Robot Vision most importantly involves a body - **embodiment**

- AI seems to need both **mind & body** (just like humans!!)

- **abstraction** is required for "Scene Understanding" (pixels, features, patches, histograms...)

- robots (seeing ones!!) can help understand how **cognition** emerges**!!!**

- autonomous robots are not mature enough
- learning methods difficult to scale-up to real-world
- “*Give a man a fish and you feed him for a day; teach a man to fish and you feed him for a lifetime.*”
  > a cognitive system requires rules rather than facts

Thank you very much!!

Lazaros Nalpantidis
web: www.csc.kth.se/~lanalpa
email: lanalpa@kth.se