Task-Specific Representations of Dynamic Environments

Darius Burschka

Machine Vision and Perception Group
Department of Computer Science

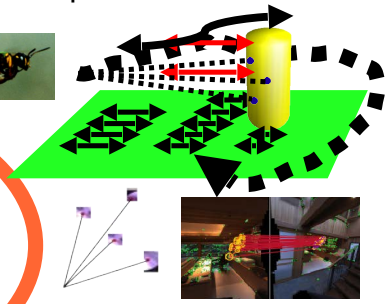
and
Scientific Board Member
Munich School of Robotics and Machine Intelligence (MSRM)
Technische Universität München

Research of the MVP Group

Perception for manipulation

Visual navigation

Biologically motivated perception



The *Machine Vision and Perception Group* @TUM works on the aspects of **visual perception** and **control** in medical, mobile, and HCI applications

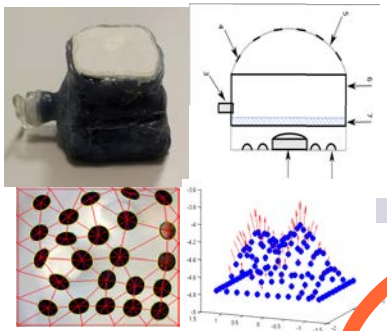
Rigid and Deformable Registration

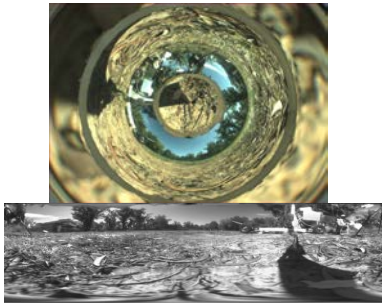
Photogrammetric monocular reconstruction

Visual Action Analysis

Research of the MVP Group

Sensor substitution

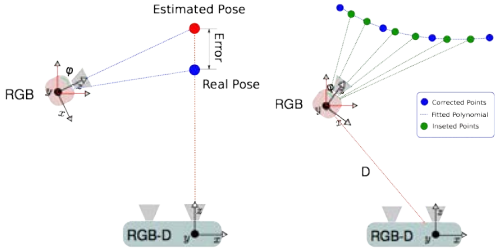




Development of new Optical Sensors

The *Machine Vision and Perception Group* @TUM works on the aspects of **visual perception** and **control** in medical, mobile, and HCI applications

Exploration of physical object properties



Multimodal Sensor Fusion

Applications

Why do we need Perception in Robotics?

"known" clutter

"unknown" clutter

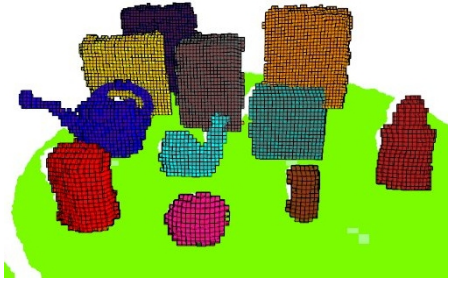
Current State of Task Definition In Manipulation...

Big Bang Theory

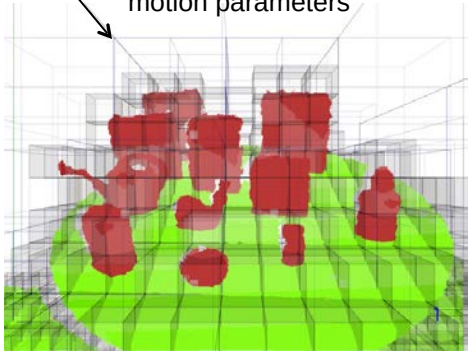
What do we need to define a task?

- Acquisition of the scene
- Clustering of data
- Labeling of data
- Identification of independent motion components
- Estimation of action
- Understanding actions

labeling



motion parameters



Sensors in Robotics

Flood of sensors required on modern service robots to cover varying measurement ranges and operating spaces

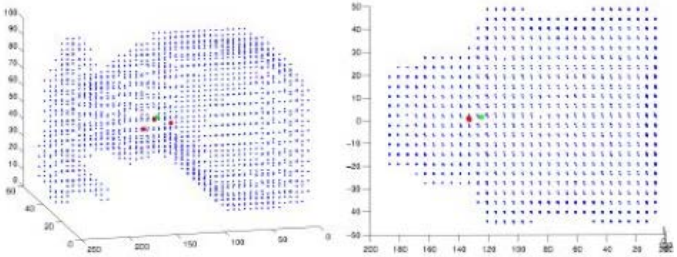
- Video Cameras
- Laser-Range Finders
- Accelerometers
- Force Sensors
- Artificial Skin
- Chemical Sensors

…..

The sensor configuration does change also with the operating conditions, e.g., baseline dependency of a binocular stereo setup

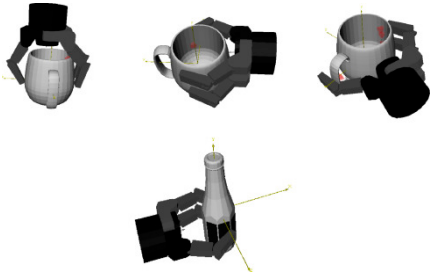
➔Implementation of missing sensing modality from a basic set of physical sensors

Stable Grasp Requires Knowledge of Physical Object Properties (weight, CoM?)



(a) Spray bottle CoM (b) Juice bottle CoM

(d) Spray bottle mass distribution (e) Juice bottle mass distribution



Material (weight): Glass or Plastic?

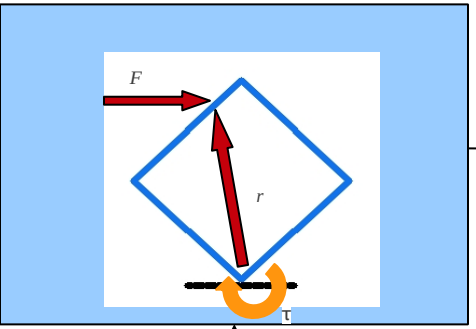
Center of Mass not equal to center of the geometry

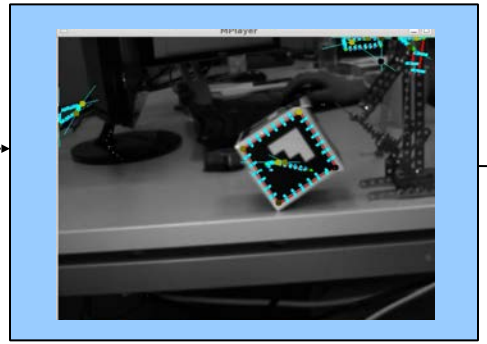
Active Exploration

predict

act

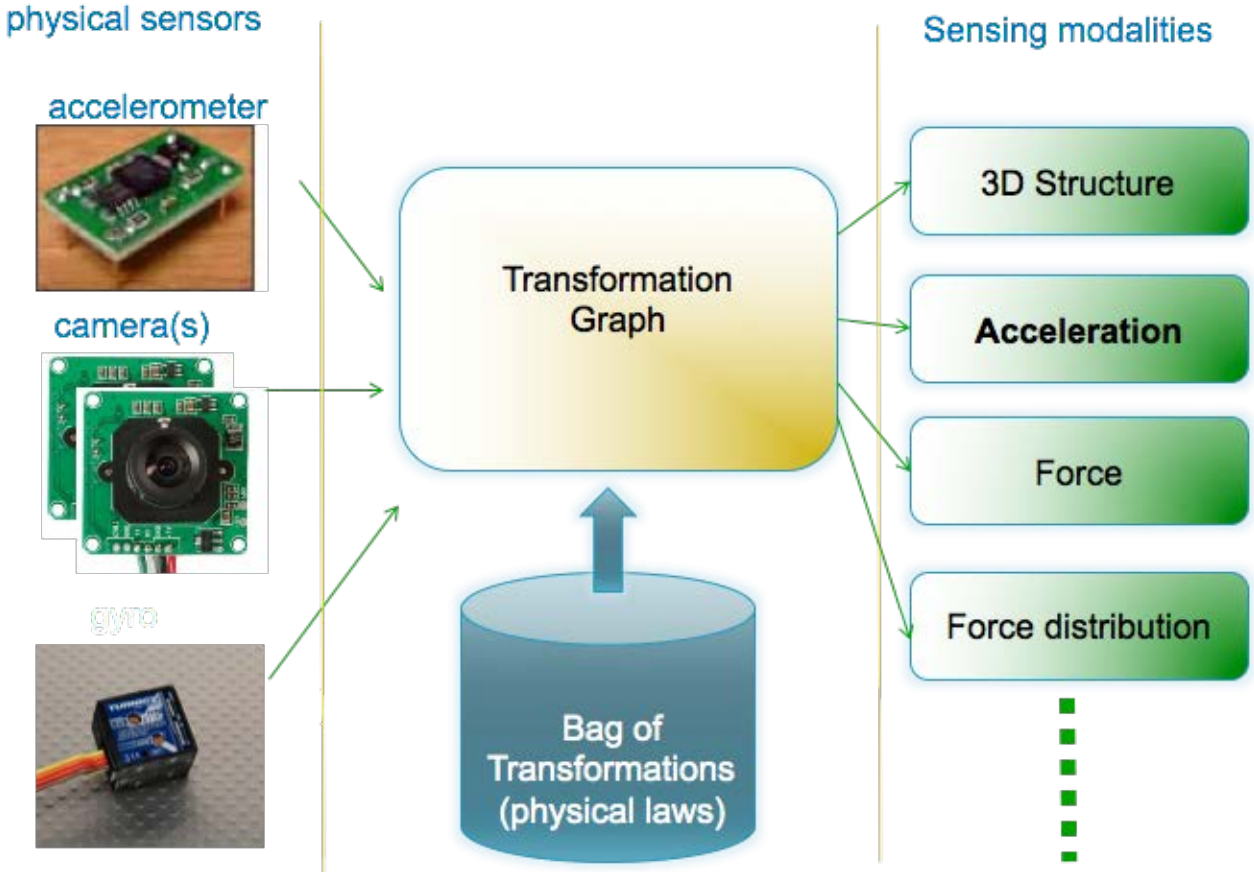
perceive



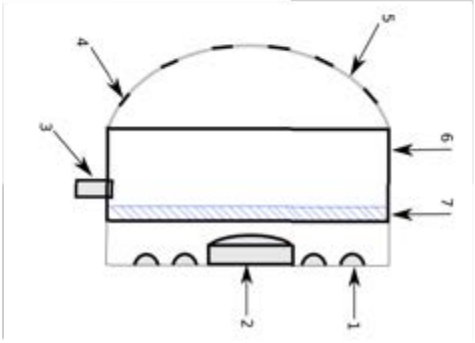


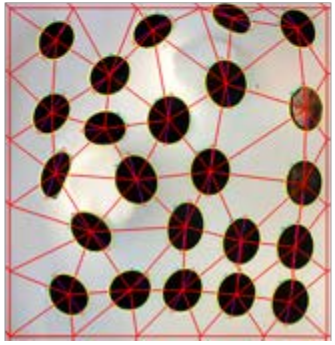
simulator

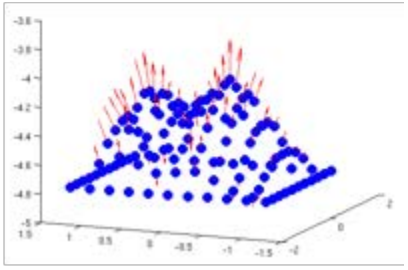
Substitution Framework



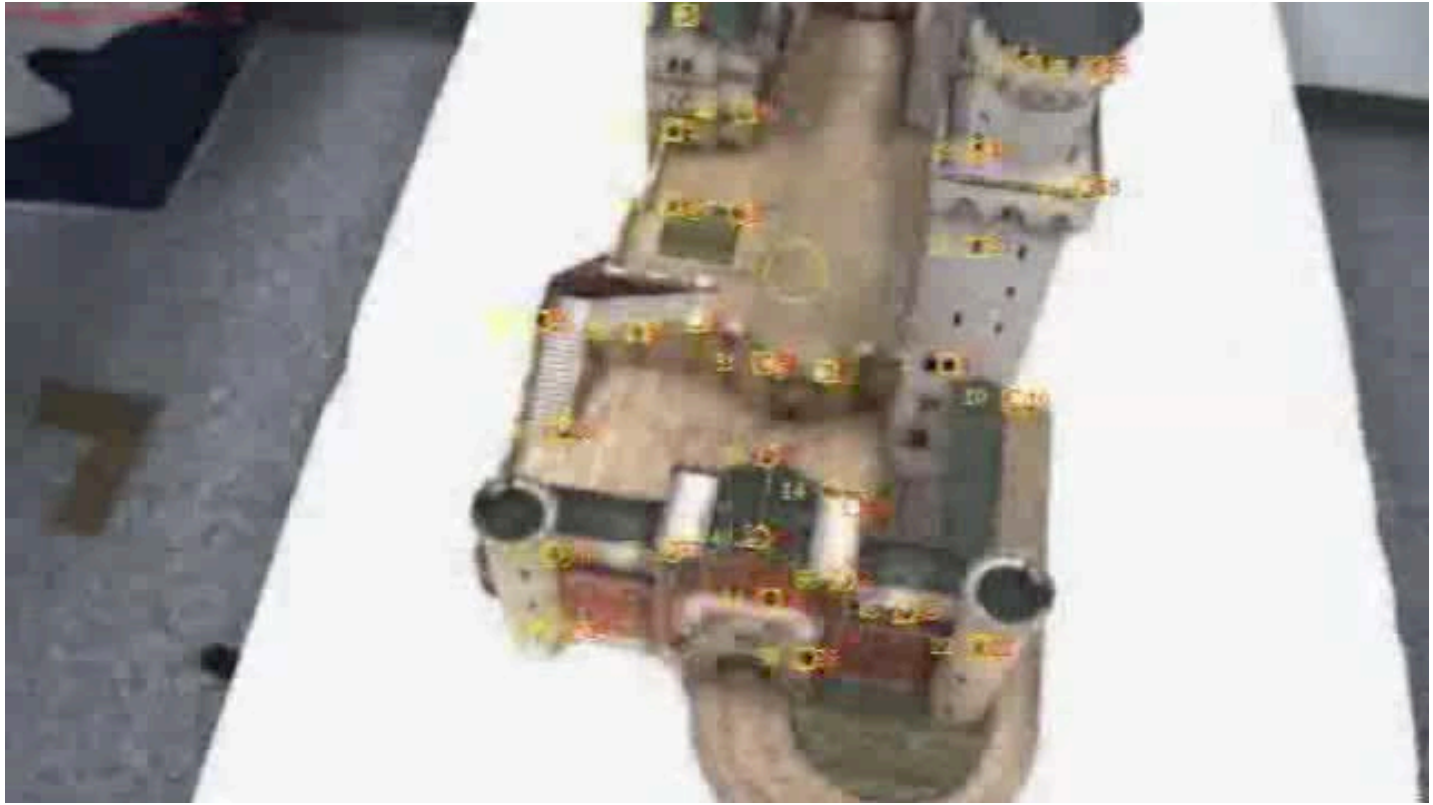
Force Sensing without Force Sensor?



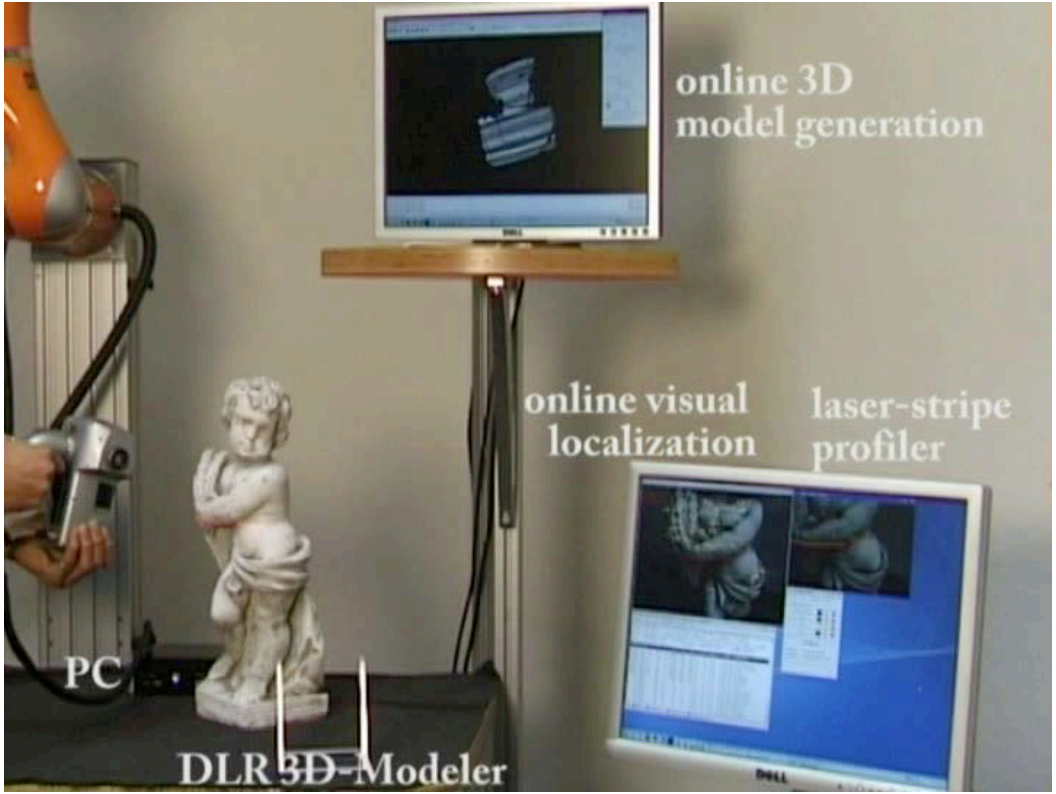




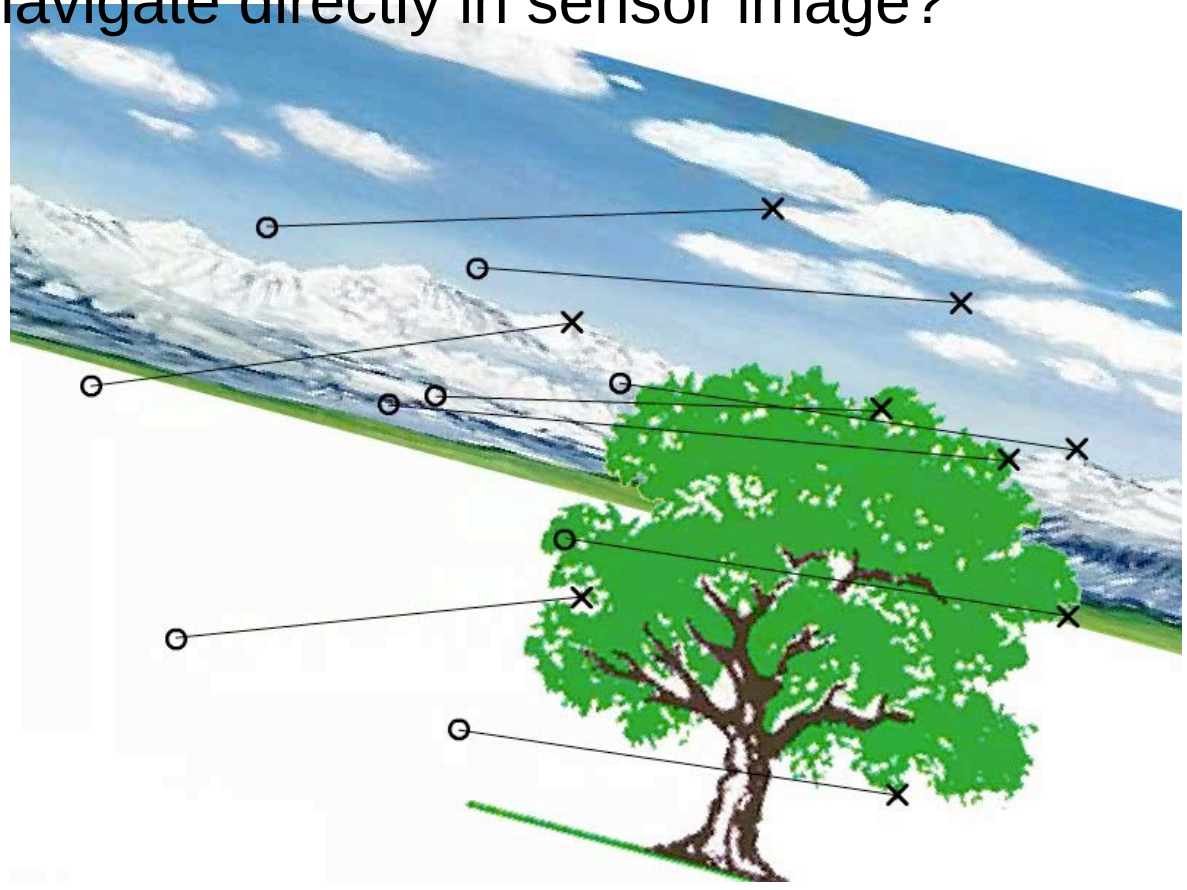
Early Monocular Navigation Approaches VGPS (IROS 2003)



Fusion of sensor readings - Construction of 3D models

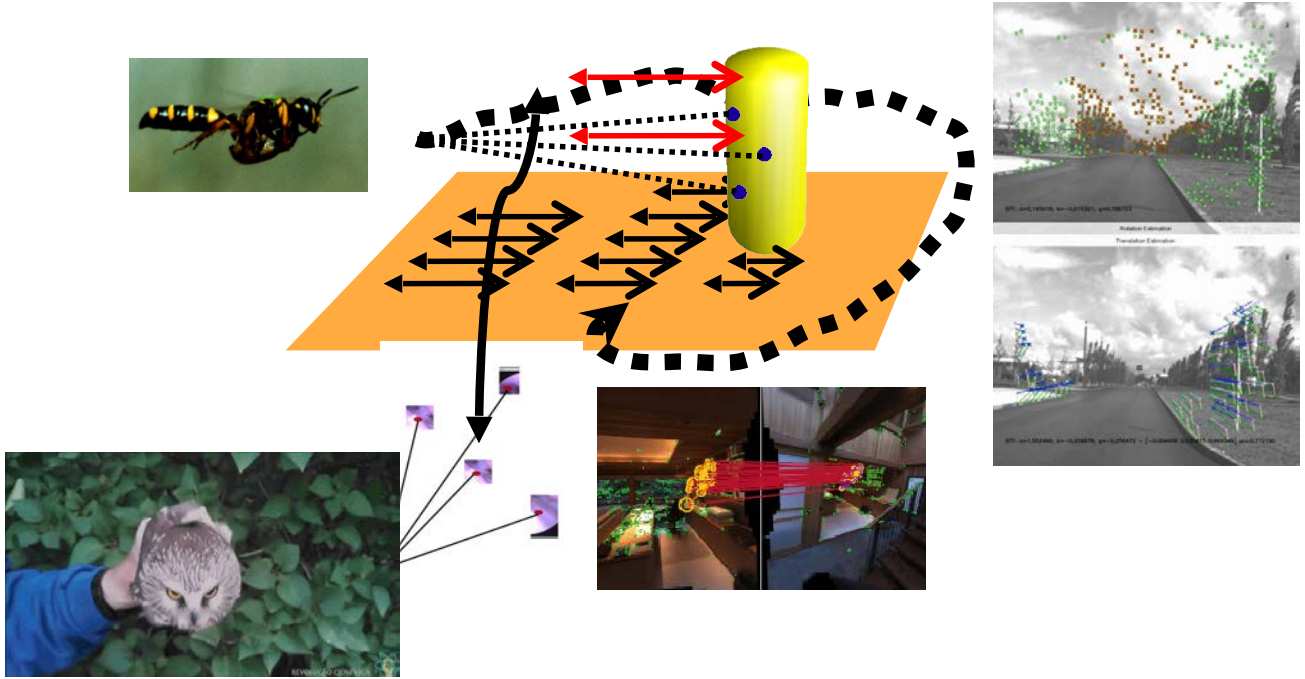


Can we navigate directly in sensor image?



Z_∞ – Algorithm at Work

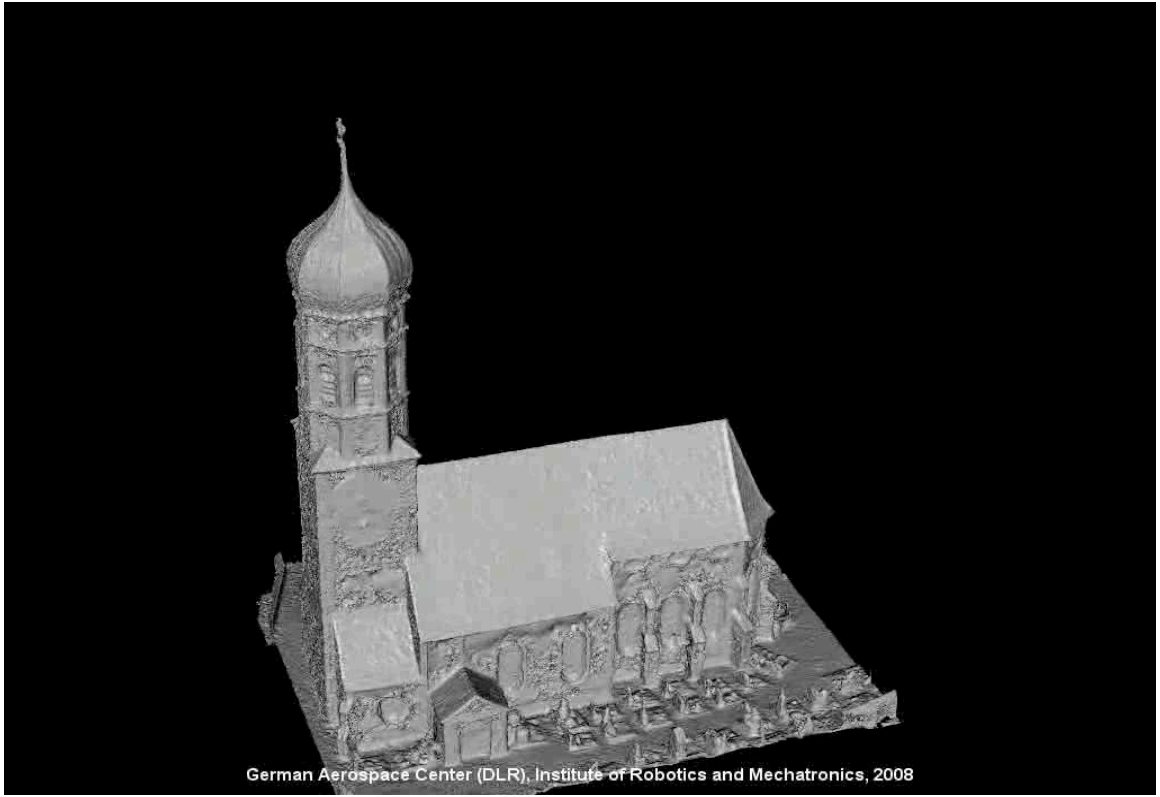
Mair, Burschka
Mobile Robots Navigation, book chapter, In-Tech, 2010



Real-Time Navigation Data from an Image Sequence

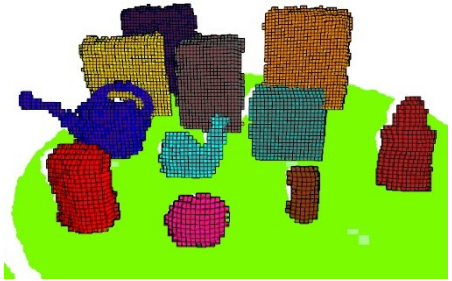
Estimation of the 6 Degrees of Freedom

We used to do this in 2007… now again?

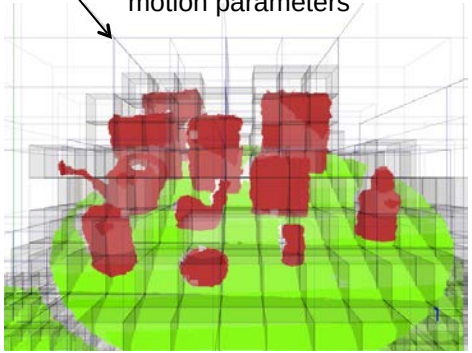


What can we do with the 3D PointClouds?

labeling



motion parameters



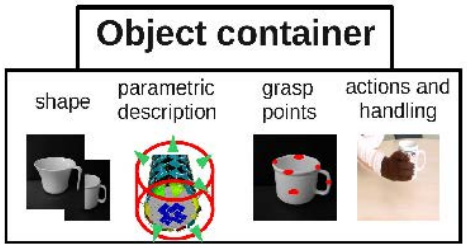
What is in the scene? (labeling)

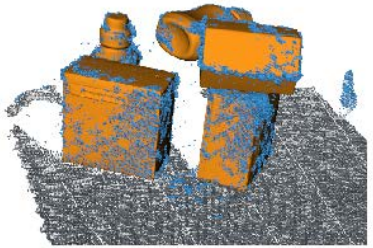
Indexing of the Atlas information from 3D perception

Real-world scenario

scene setup

input point cloud

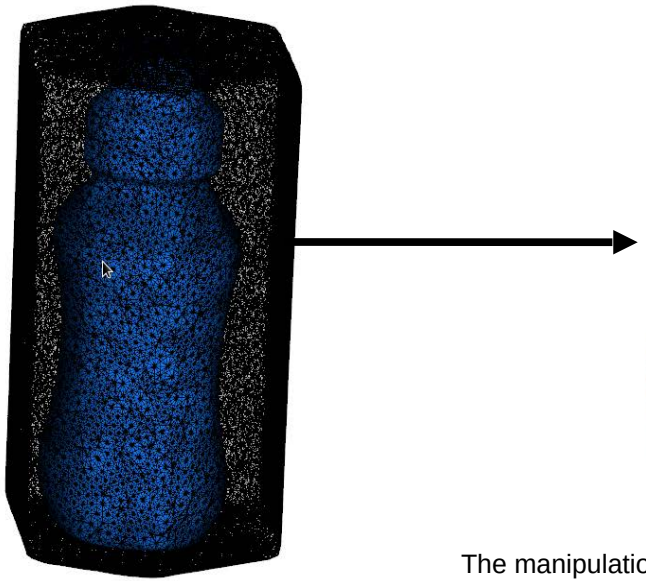




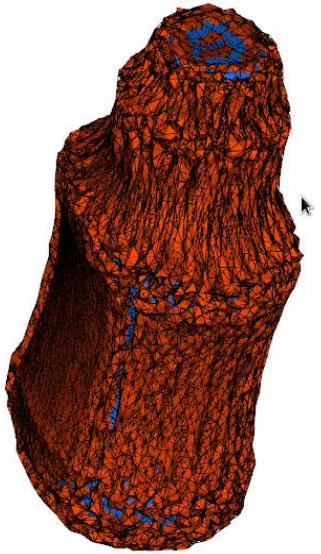
recognized models


Copyright MVP
Brandt

Deformable Registration from generic models (special issue SGP'11 Papazov et al.)

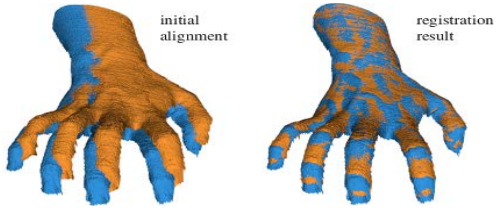


Matching of a detailed shape to a primitive prior



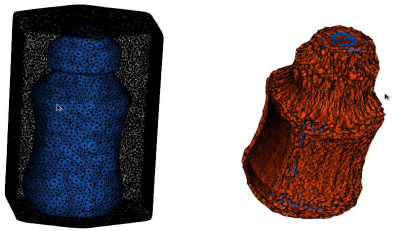
The manipulation “heat map” from the generic model gets propagated

A p p l i c a t i c r s c f D e

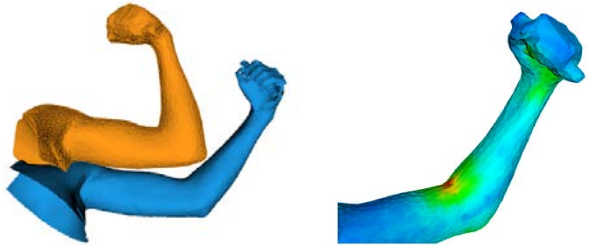


Alignment of non-rigid shapes for inspection and localization

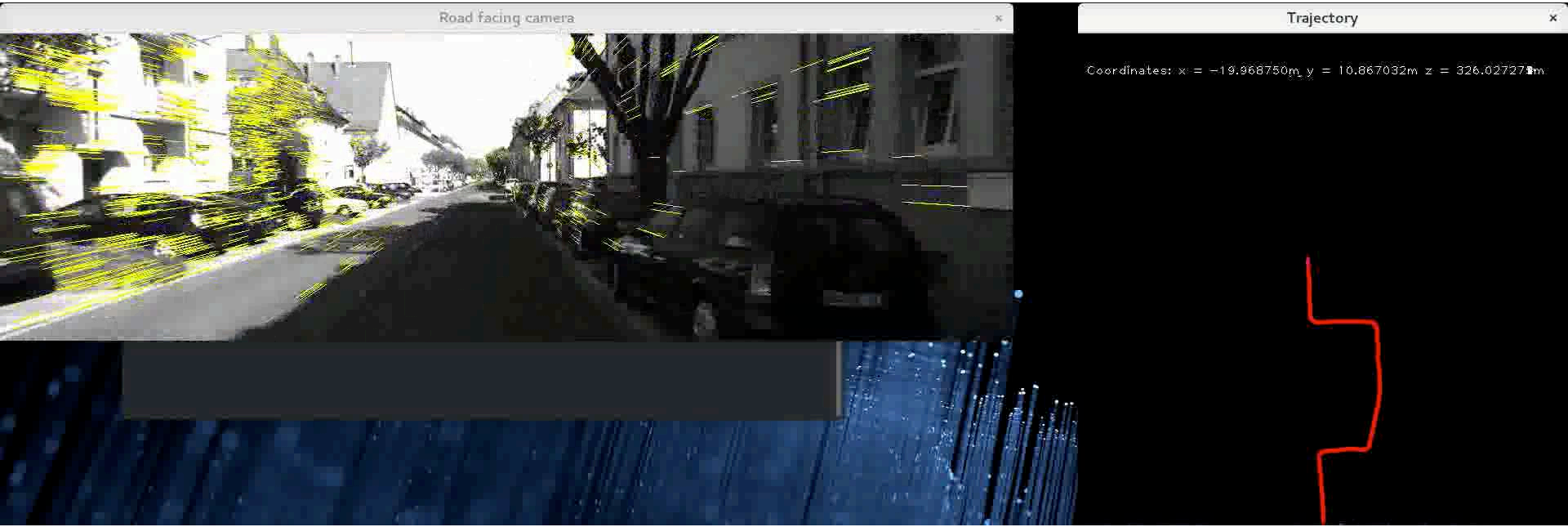
Similarity estimates for categorization of objects



Visual identification of kinematic chains

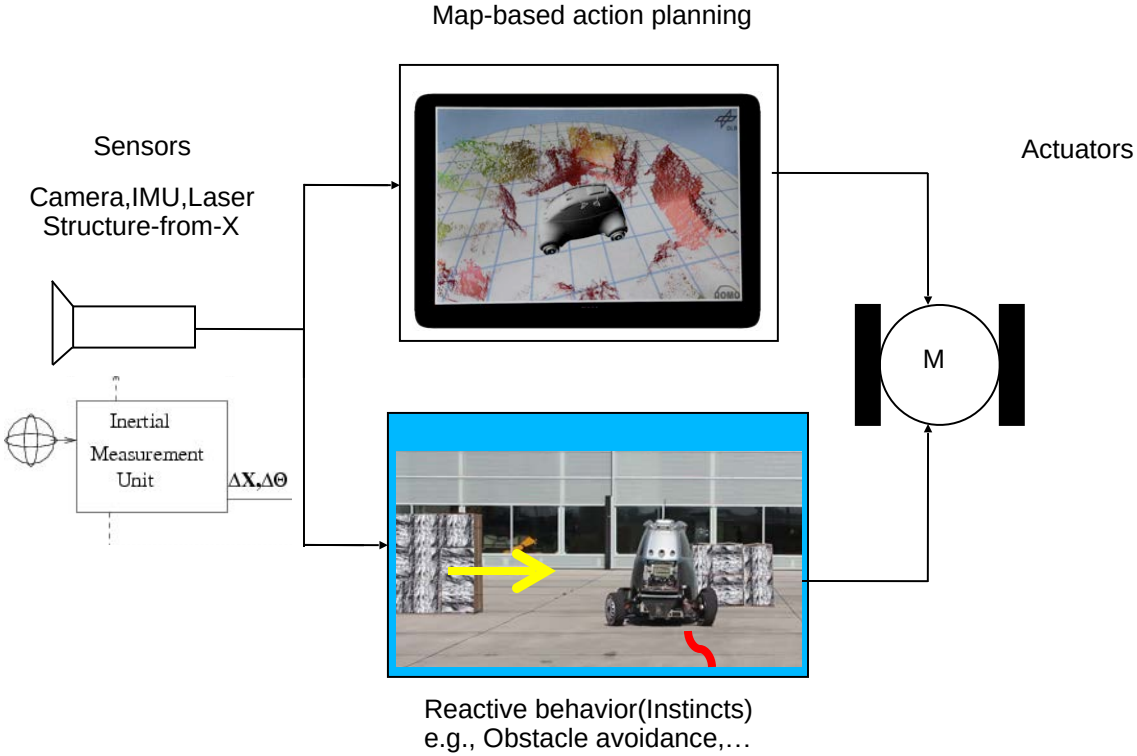


120fps Monocular Navigation from Flow

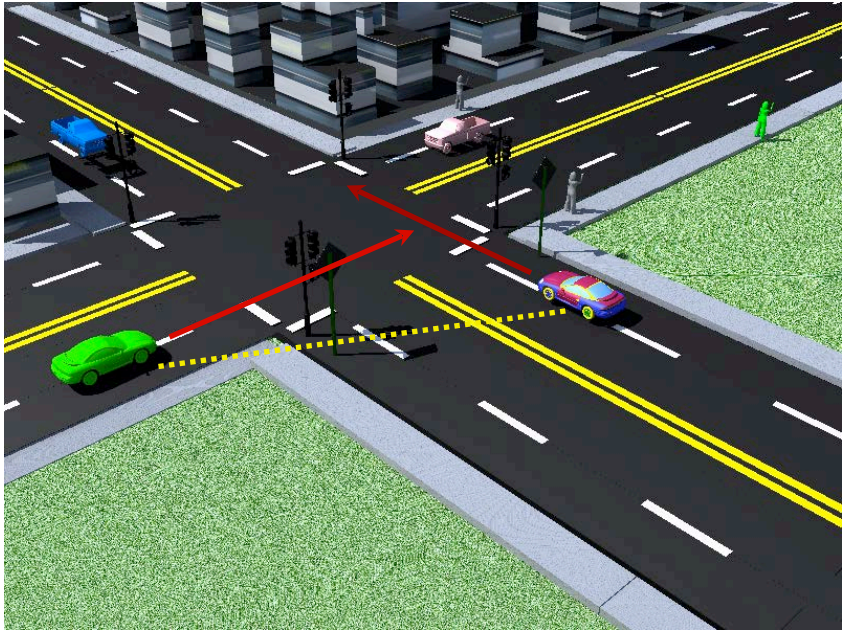


Reconstruction Task seemed to be solved? (2010)

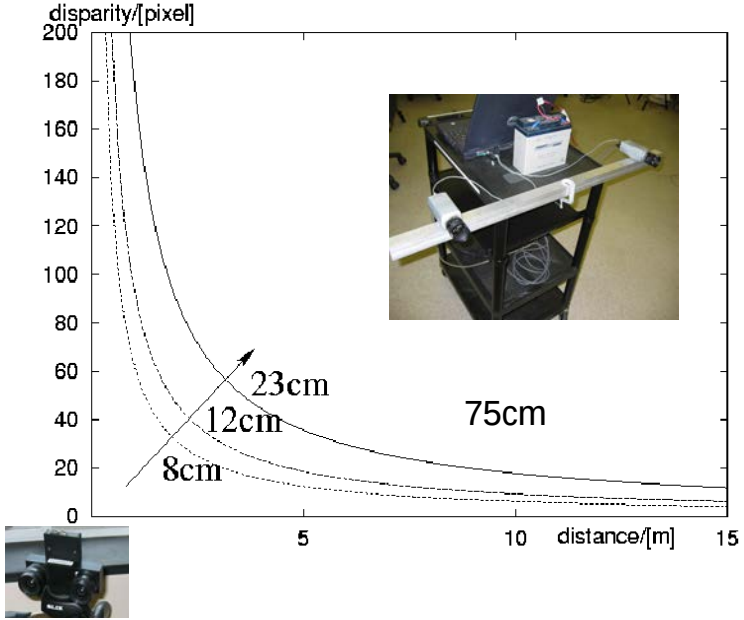
Coupling Alternatives for Perception Modules



Capturing Motion Properties of Large Dynamic Scenes



Are lab approaches transferrable to automobile and avionic applications?

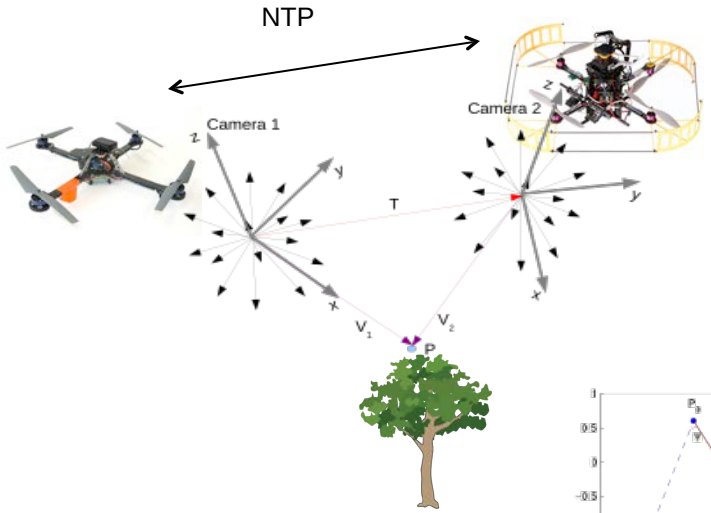


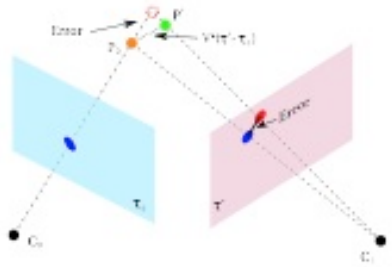
$$d_p = \frac{B \cdot f}{p_x} \cdot \frac{1}{z} [pixel]$$

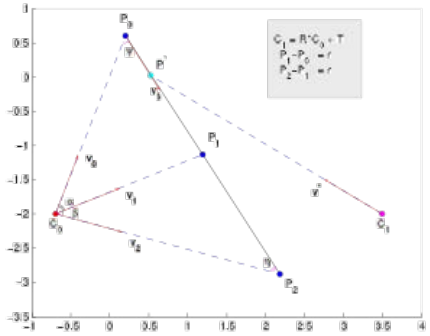
Sensitivity increase:

- Larger baseline (B)
- Longer focal length (f) ® field of view
- Smaller pixelsize (p_x) ® “pixel explosion”

Asynchronous stereo for dynamic scenes

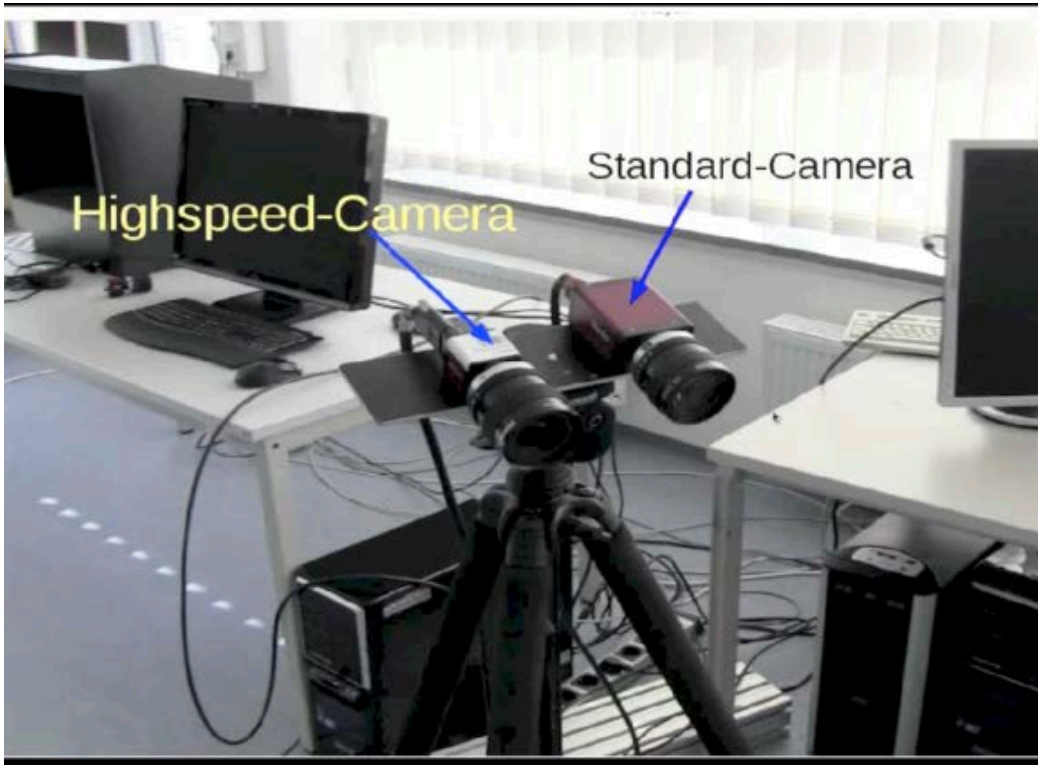




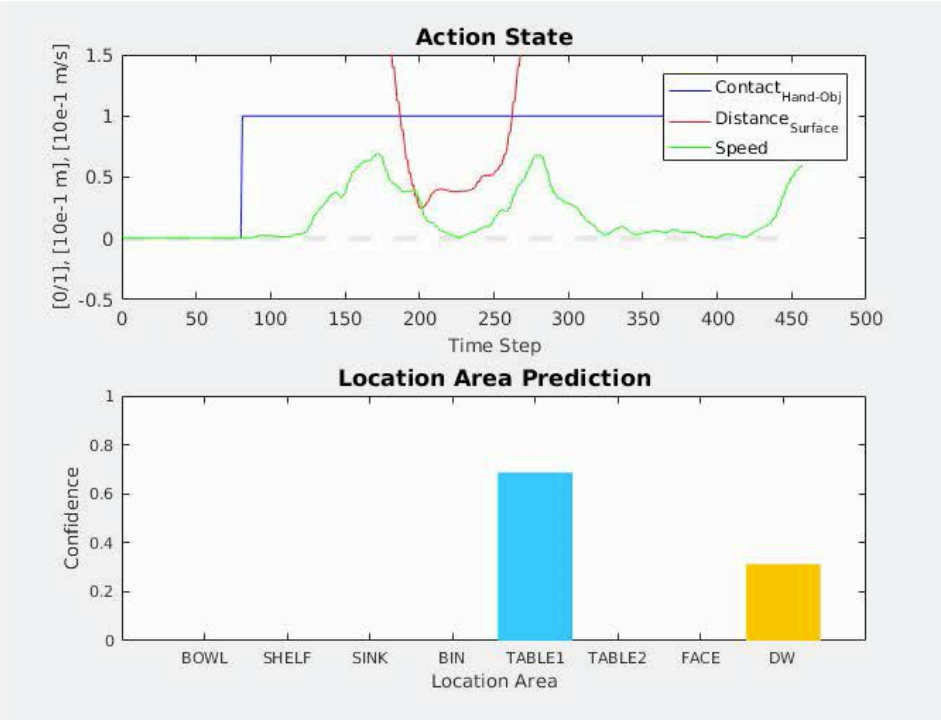


Mkhitaryan, Burschka VISAPP 2014

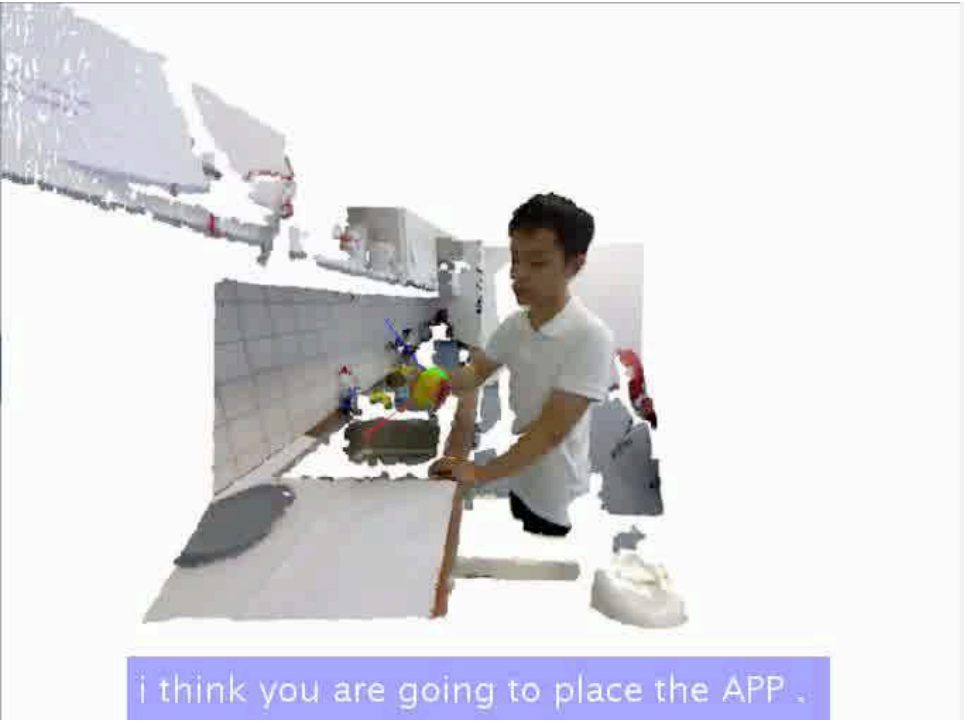
Hybrid High-Speed Stereo System

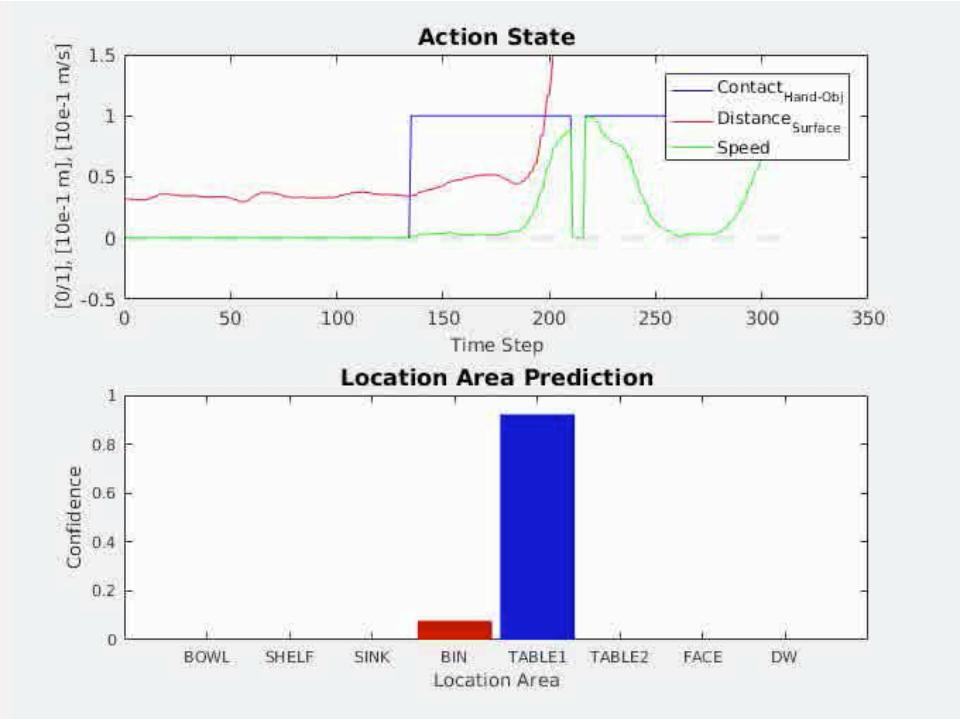


Understanding Tasks to Predict and to Understand the Physical Object State



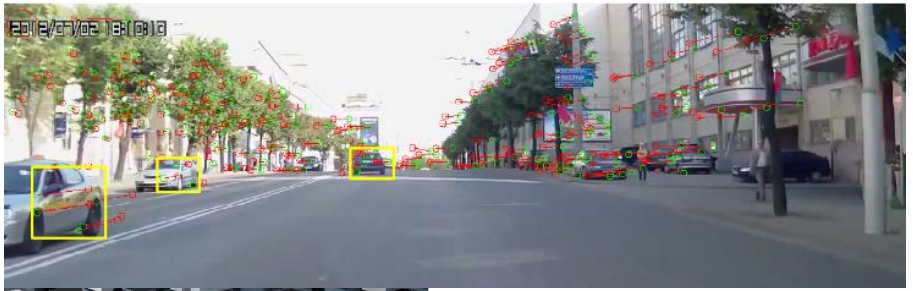
…similar observation, different action label

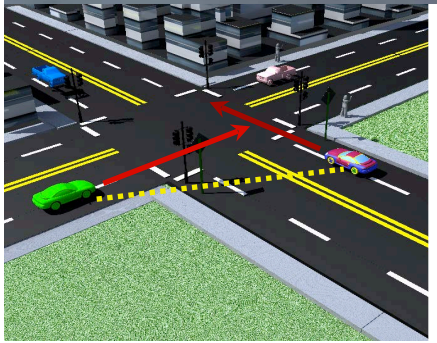


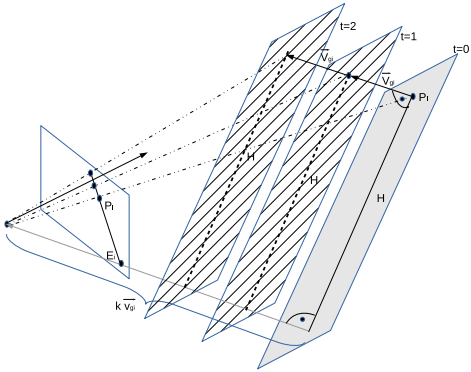


Detection of Independent Motion Groups from Optical Flow

- Our goal is a robust detection of **motion direction** and **collision times** from a **monocular uncalibrated** camera sequence.
- Representation of the **dynamic scene ordered by collision times** instead of Cartesian coordinates aenables monocular processing (**no scale necessary**) and better priorisation of collision candidates than in conventional methods
- Independent estimation of motion direction and collision time allows collision categorization in large distances from the camera





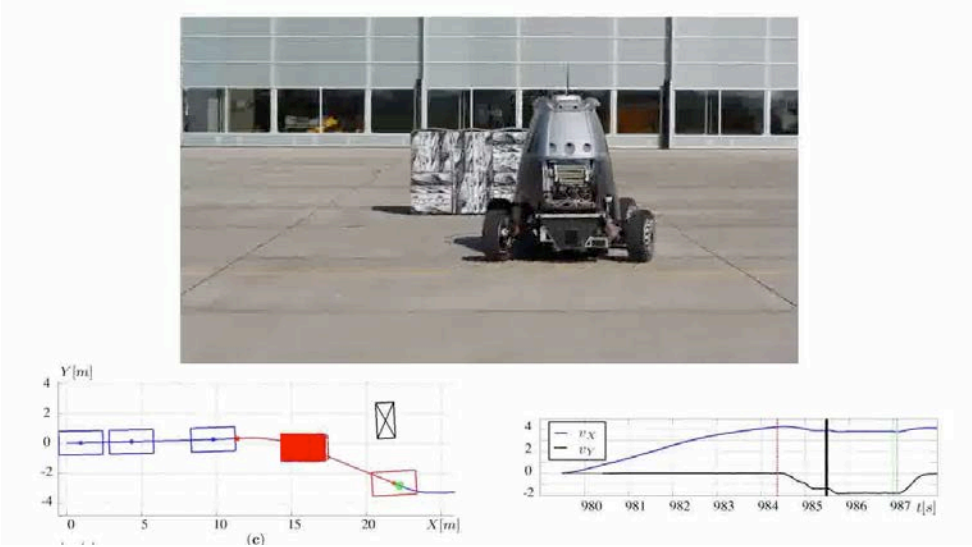


Schaub et al., Journal ITSC
Burschka, BMVC 2017

Navigation based on Pixel-Information from Monocular View

Concept: Shifting the optical flow epipole out of the object's boundaries → no collision

Schaub, Burschka ITSCC 2015



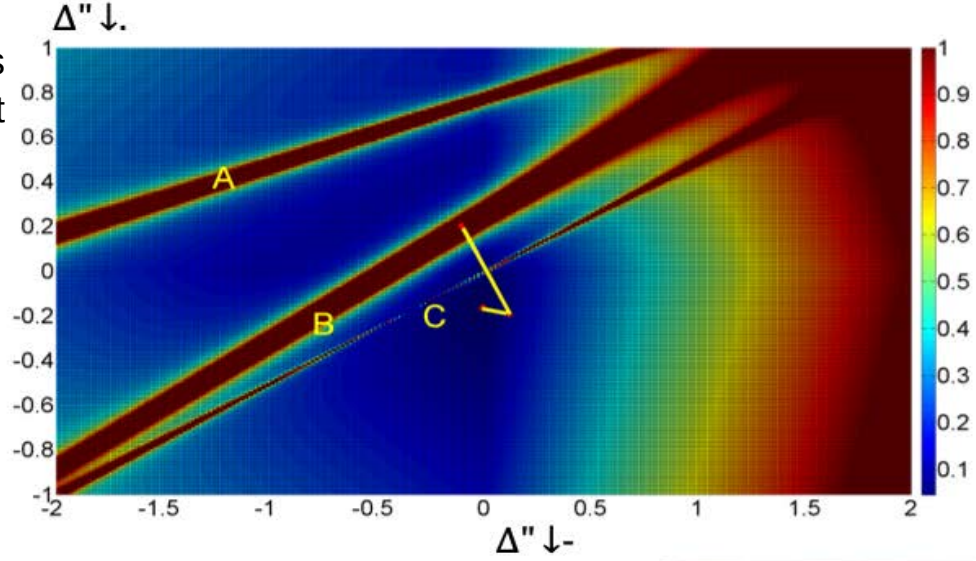
Planar motion of the objects (

Effect of th $E_x = c_x + f s_x \frac{v_Y}{v_X}$ e
Epipole's 2 $E_y = c_y$ ind: :

$$\Delta E_x = E_x(v + \Delta v) - E_x(v) = f_x \frac{v_Y \Delta v_X - v_X \Delta v_Y}{(v_X + \Delta v_X) v_X}$$

Novel Control Design for non-metric Control Input

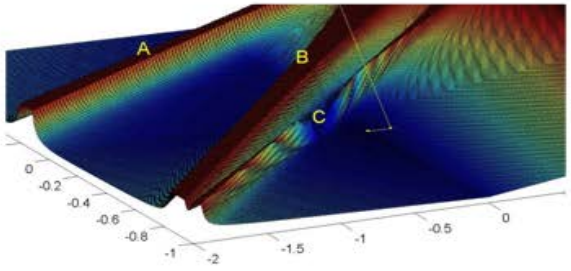
- Planning space represented as collision times for different velocities



- Non-Linear Gradient-Descent with an Adaptive Lagrange Interpolation Search (ALIS)
- Weights: $J_d > J_{a_X} > J_{a_Y}$
- Good performance: 2 Steps to reach the optimum

$$J = 0.0349 \qquad J_{fmincon} = 0.3792$$

- Realtime implementation with 25 Hz



Contributions

- Active Exploration: Estimation of additional physical properties of objects that cannot be observed by passive sensing to robustify grasping of objects
- Semantic Labeling: Semantic abstraction of sensory observations for easier knowledge exchange with the human operator - MMI
- Interaction with Dynamic Environments: Generalization of the observed tasks to support motion planner in execution of the shown task