Synergies of Motion Planning and Visual Perception by Reasoning about Possible Outcomes

ERF 2014, Rovereto, 13.03.2014

Zoltan-Csaba Marton

German Aerospace Center (DLR) Institute of Robotics and Mechatronics

Outline

1. Introduction

2. Object Categorization in Clutter
Unsupervised Part Learning
Scene Subgraphs and Additive Features

3. Exploiting Embodiment
Scenes From Multiple Views
Robot Interacting with the Scenes

4. Merging Pose Estimates
Pose Distances and Means
Discrete Bayes Filtering

5. Summary

Outline

1. Introduction

2. Object Categorization in Clutter
Unsupervised Part Learning
Scene Subgraphs and Additive Features

3. Exploiting Embodiment
Scenes From Multiple Views
Robot Interacting with the Scenes

4. Merging Pose Estimates
Pose Distances and Means
Discrete Bayes Filtering

5. Summary

Introduction

- For successful grasp planning, an accurate object segmentation and pose estimation is necessary.
- Current perception methods for such tasks, however, have different strengths and weaknesses, and are influenced by occlusions in real-world applications.
- On the other hand, robots have the ability to move around to acquire more information, and even to interact with the scene in order to ease perceptual interpretation.
- Thus perception and action can mutually benefit from each other, given the ability to reason about possible gains of certain actions or perception methods.
- I will present past and ongoing efforts in these directions, aimed to enable robots to interpret complex scenes.

Introduction

- For successful grasp planning, an accurate object segmentation and pose estimation is necessary.
- Current perception methods for such tasks, however, have different strengths and weaknesses, and are influenced by occlusions in real-world applications.
- On the other hand, robots have the ability to move around to acquire more information, and even to interact with the scene in order to ease perceptual interpretation.
- Thus perception and action can mutually benefit from each other, given the ability to reason about possible gains of certain actions or perception methods.
- I will present past and ongoing efforts in these directions, aimed to enable robots to interpret complex scenes.

Introduction

- For successful grasp planning, an accurate object segmentation and pose estimation is necessary.
- Current perception methods for such tasks, however, have different strengths and weaknesses, and are influenced by occlusions in real-world applications.
- On the other hand, robots have the ability to move around to acquire more information, and even to interact with the scene in order to ease perceptual interpretation.
- Thus perception and action can mutually benefit from each other, given the ability to reason about possible gains of certain actions or perception methods.
- I will present past and ongoing efforts in these directions, aimed to enable robots to interpret complex scenes.

Introduction

- For successful grasp planning, an accurate object segmentation and pose estimation is necessary.
- Current perception methods for such tasks, however, have different strengths and weaknesses, and are influenced by occlusions in real-world applications.
- On the other hand, robots have the ability to move around to acquire more information, and even to interact with the scene in order to ease perceptual interpretation.
- Thus perception and action can mutually benefit from each other, given the ability to reason about possible gains of certain actions or perception methods.
- I will present past and ongoing efforts in these directions, aimed to enable robots to interpret complex scenes.

Introduction

- For successful grasp planning, an accurate object segmentation and pose estimation is necessary.
- Current perception methods for such tasks, however, have different strengths and weaknesses, and are influenced by occlusions in real-world applications.
- On the other hand, robots have the ability to move around to acquire more information, and even to interact with the scene in order to ease perceptual interpretation.
- Thus perception and action can mutually benefit from each other, given the ability to reason about possible gains of certain actions or perception methods.
- I will present past and ongoing efforts in these directions, aimed to enable robots to interpret complex scenes.

Outline

1. Introduction

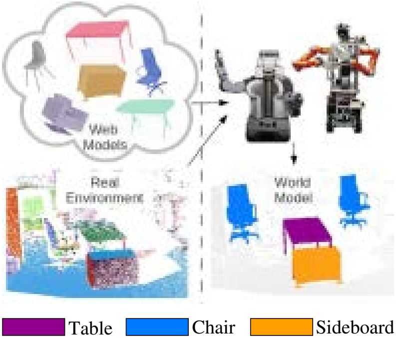
2. Object Categorization in Clutter
Unsupervised Part Learning
Scene Subgraphs and Additive Features

3. Exploiting Embodiment
Scenes From Multiple Views
Robot Interacting with the Scenes

4. Merging Pose Estimates
Pose Distances and Means
Discrete Bayes Filtering

5. Summary

Unsupervised Part Learning



[Mozos et al. RAM'11]

PCL-based implementation:

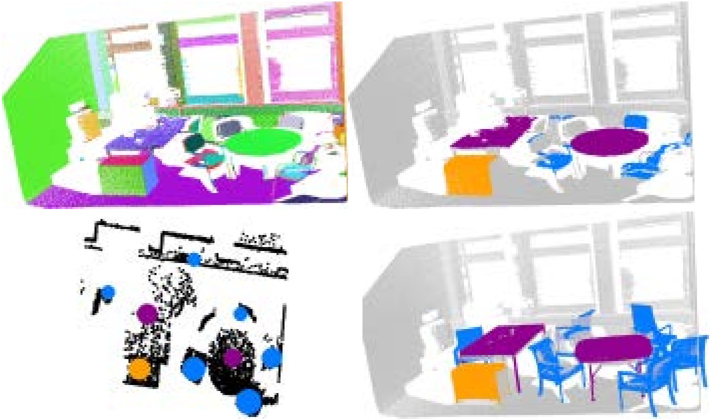
`http://www.ros.org/wiki/furniture_classification`

[by Vlad Usenko]

- Identification of furniture pieces for which similar CAD models are available from online stores.
- Common parts are grouped into a codebook based on simple statistics and they cast votes for object hypotheses
- Pose estimation by model matching and geometric verification.
- Can increase robustness by using multiple segmentations or views.

Unsupervised Part Learning

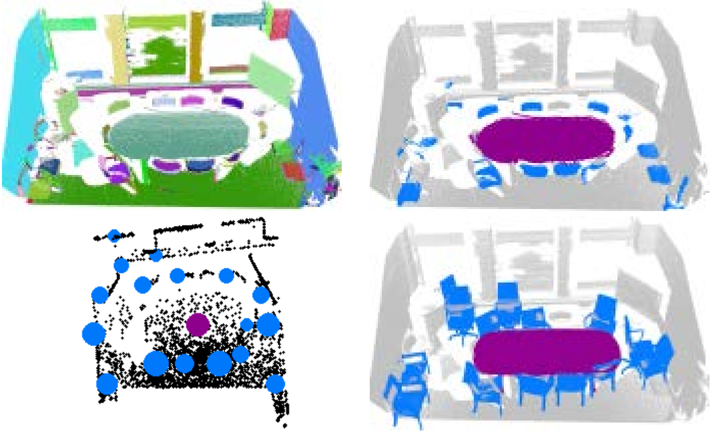
Results for Office



- remaining false matches due to high occlusions

Unsupervised Part Learning

Results for Seminar Room



- remaining false matches due to high occlusions

Scene Subgraphs

Considering all possible part groupings

- Object categorization in cluttered scenes where accurate segmentation can be difficult.
- Over-segmentation and multiple hypotheses better than relying on a single, possibly bad segmentation.
- Approach based on scene- or part-graphs, using additive RGBD feature descriptors.

[Marton et al. SC'12, RSS-WS'13, extended version in JINT'14]

Scene Subgraphs

Segmentation and classification on cluttered table scenes

Object parts are segmented and categorized as *spherical*, *box*, *flat* and *cylindrical* (training and large-scale testing done using the RGB-D Object Dataset *[Lai et al. ICRA'11]*). Geometric modeling performed using the priors from the categorization.

Outline

1. Introduction

2. Object Categorization in Clutter
Unsupervised Part Learning
Scene Subgraphs and Additive Features

3. Exploiting Embodiment
Scenes From Multiple Views
Robot Interacting with the Scenes

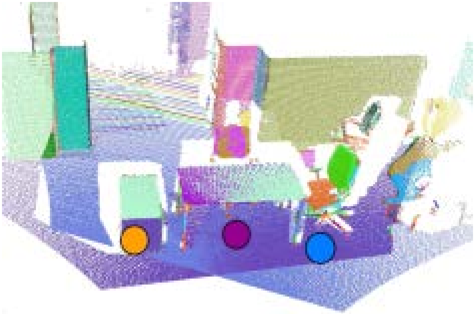
4. Merging Pose Estimates
Pose Distances and Means
Discrete Bayes Filtering

5. Summary

Exploiting Embodiment

Scenes From Multiple Views

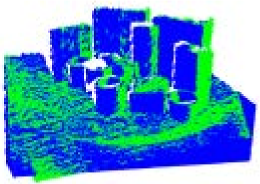
- Decreasing occlusion, increasing number of parts

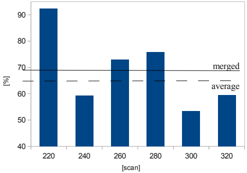


- The model fitting and verification steps also do not assume one viewpoint or one segmentation per scan

Exploiting Embodiment

Scenes From Multiple Views

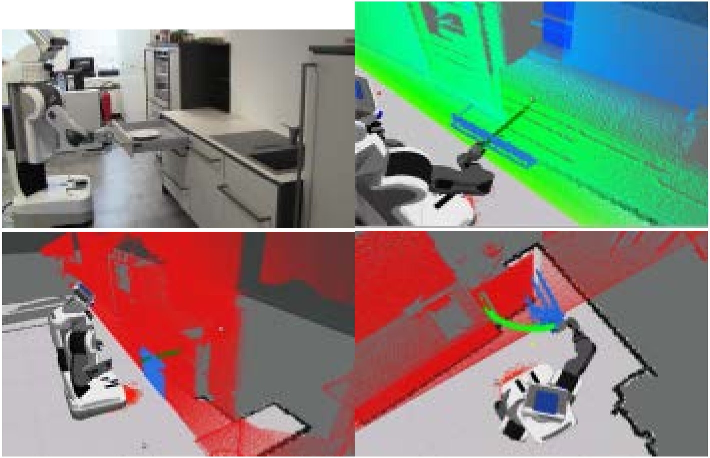




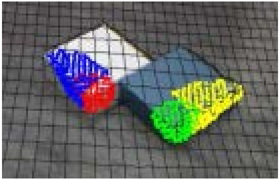
As the camera is moved (left), multiple frames can be captured that cover different parts of the objects in the scene (right).

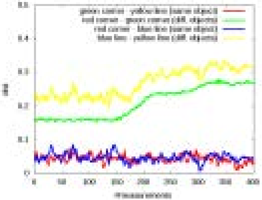
Exploiting Embodiment

Door and Drawer Hypothesis Validation through Interaction



Exploiting Embodiment





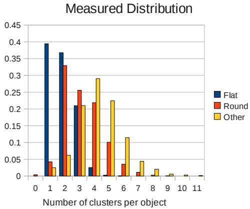
Detecting when and where to push and tracking 3D features in order to segment objects (using *openni_tracking* from U-Tokyo):

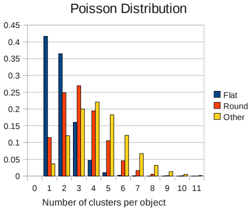
`http://www.ros.org/wiki/interactive_segmentation_textureless`

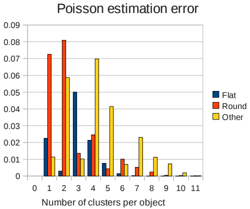
[Bersch et al. RSS'12/WS, Hausmann et al. ICRA'13]

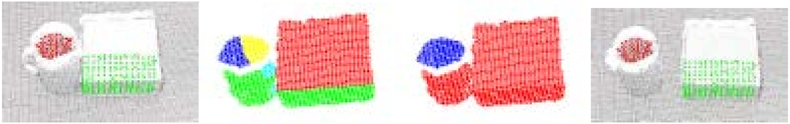
Exploiting Embodiement

Robot Interacting with the Scenes









Thanks to Karol Hausmann, Ferenc Balint-Benczedi, Dejan Pangercic and Ryohei Ueda (Univ. Tokyo / Prof. Kei Okada)!

Outline

1. Introduction

2. Object Categorization in Clutter
Unsupervised Part Learning
Scene Subgraphs and Additive Features

3. Exploiting Embodiment
Scenes From Multiple Views
Robot Interacting with the Scenes

4. Merging Pose Estimates
Pose Distances and Means
Discrete Bayes Filtering

5. Summary

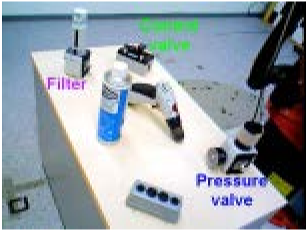
Merging Pose Estimates

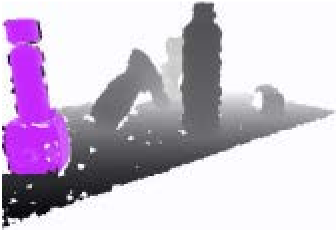
Example application scenario

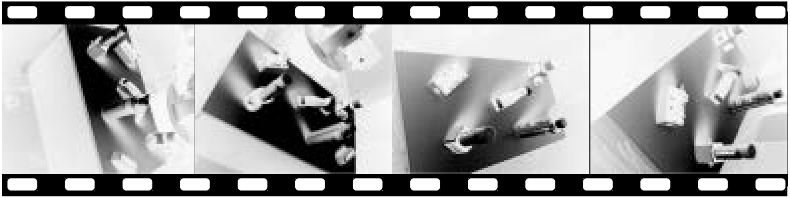
Estimating the relative transformation between 3D model and the robot (its sensor), based on 3D images. Different sensors, object types, methods – develop a “high-level” solution.

Merging Pose Estimates

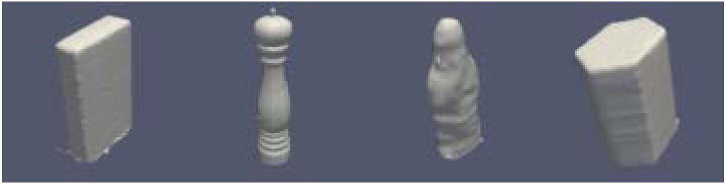
Basic idea of multi-view recognition

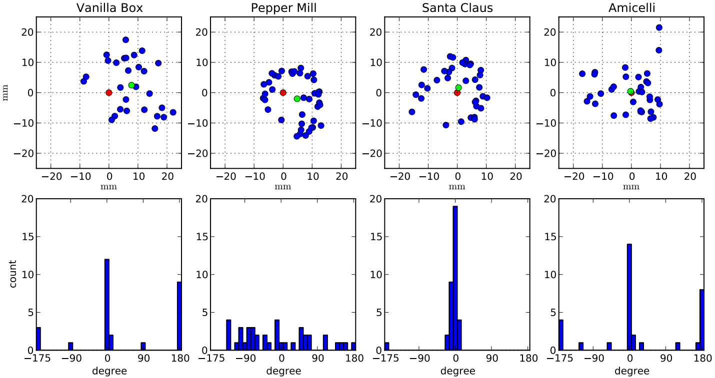






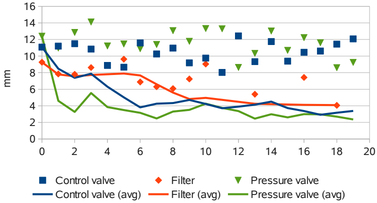
Pose Distances and Means

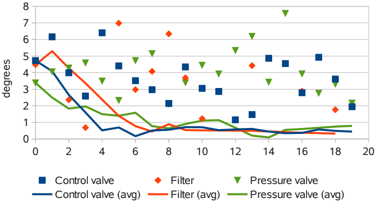




Pose Distances and Clustering

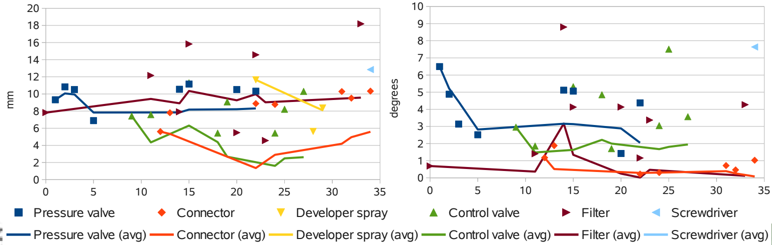
Multi-view pose estimation of single objects





[Work done together with Simon Kriegel and Manuel Brucker]

Pose Distances and Clustering



Discrete Bayes Filtering

Storing histograms

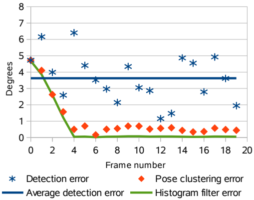
Representation for histogram filter:

- represent the space of rotation in a grid
- evenly divided the space of quaternions, and selected those cells that contain unit quaternions
- not uniform sampling, but the area of SO(3) that falls into each cell ca be estimated using random sampling
- 64x64x64x32 division of the 4D quaternions with dimensions between -1 and 1 ($w \geq 0$)
- accuracy in the range of 1-2 degrees, roughly 500,000 cells covering SO(3)
- for initial tests half or quarter of this resolution used

Discrete Bayes Filtering

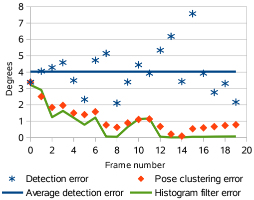
Evaluation and comparison

Rotational error for the control valve object

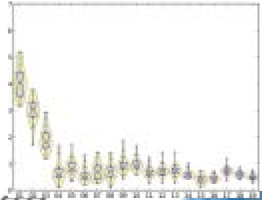


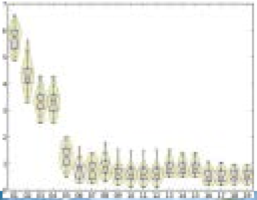
Rotational error for the filter object

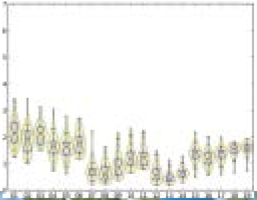
Rotational error for the pressure valve object



Histogram filter improves over the simple pose clustering. More robust than particle filters (evaluated over 20 runs):



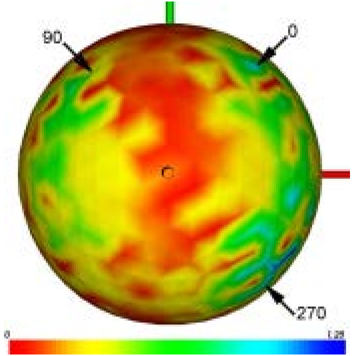




DLR

Discrete Bayes Filtering

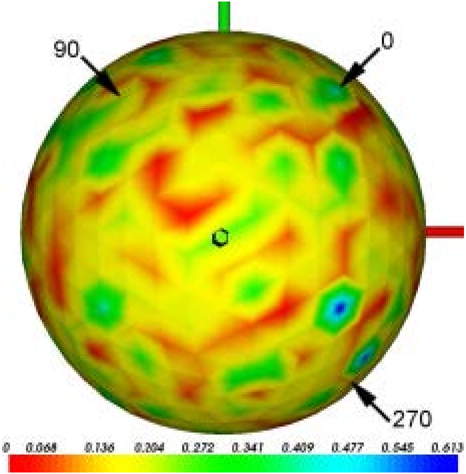
Error model evaluation



Using a low-resolution histogram, and a fast (but inaccurate) pose estimation method (feature-RANSAC with few iterations, detections shown right). Mistakes of up to 6 in 95% of the cases for the error model's histogram cells that were not hit during evaluation.

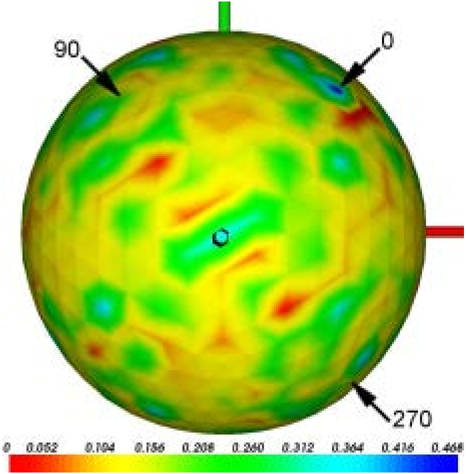
Discrete Bayes Filtering

Error model based evaluation: step 1 (correct pose at 0)



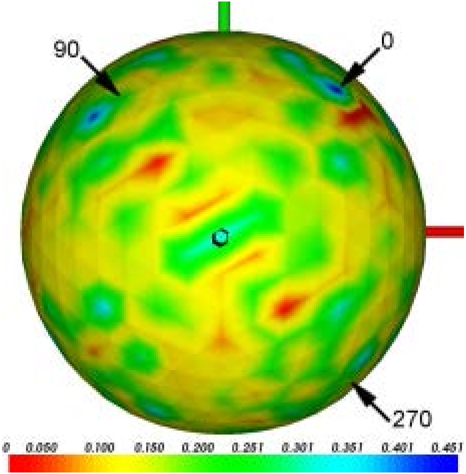
Discrete Bayes Filtering

Error model based evaluation: step 13 (correct pose at 0)



Discrete Bayes Filtering

Error model based evaluation: step 24 (correct pose at 0)



Outline

Summary

- Segmentation and pose estimation in real-world scenes is challenging, but embodiment can help.
- Need a good way of reasoning about possible outcomes, and checking the results of actions.
- A good modeling of expected results is difficult, and might not be accurate enough.
- Preliminary results show the feasibility of such approaches and their use for view and interaction planning.

Summary

- Segmentation and pose estimation in real-world scenes is challenging, but embodiment can help.
- Need a good way of reasoning about possible outcomes, and checking the results of actions.
- A good modeling of expected results is difficult, and might not be accurate enough.
- Preliminary results show the feasibility of such approaches and their use for view and interaction planning.

Summary

- Segmentation and pose estimation in real-world scenes is challenging, but embodiment can help.
- Need a good way of reasoning about possible outcomes, and checking the results of actions.
- A good modeling of expected results is difficult, and might not be accurate enough.
- Preliminary results show the feasibility of such approaches and their use for view and interaction planning.

Summary

- Segmentation and pose estimation in real-world scenes is challenging, but embodiment can help.
- Need a good way of reasoning about possible outcomes, and checking the results of actions.
- A good modeling of expected results is difficult, and might not be accurate enough.
- Preliminary results show the feasibility of such approaches and their use for view and interaction planning.

Summary

- Segmentation and pose estimation in real-world scenes is challenging, but embodiment can help.
- Need a good way of reasoning about possible outcomes, and checking the results of actions.
- A good modeling of expected results is difficult, and might not be accurate enough.
- Preliminary results show the feasibility of such approaches and their use for view and interaction planning.

Questions?

Contact: `zoltan.marton@dlr.de`