Acquiring Semantic Object Maps for Household Tasks

Dejan Pangeric
Intelligent Autonomous Systems Group
Technische Universität München

Munich, November 2011
1. Semantic Object Maps

2. Texture Blending

3. Perception of Handles

4. Cabinet Opening

5. Conclusions
Structures that enable the robot to answer the following categories of queries:

- “What do parts of the kitchen look like?”;
- “How can a container be opened and closed?”;
- “What is inside of cupboards/drawers?”;
- “Where do objects of daily use belong?”.

Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks
System Overview

- End-to-end system for generation of SOM+s
- Low-cost low-quality sensor
- Logic-based formal language and background knowledge for the representation of SOM+s maps
- Detection of the handles with surfaces with specular reflectivity

[ICRA 2012, RAS 2011]
1. Semantic Object Maps

2. Texture Blending

3. Perception of Handles

4. Cabinet Opening

5. Conclusions
Goal 1

⇒ Photo-realistic visualization

Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks

Dejan Panceric
Goal 2
Goal 3

Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks

Dejan Panceric
Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks

Dejan Pangeric
Texture Map defined as $\tau \subset \mathbb{R}^2$ and mapped onto the 2-dimensional space.

Texture mapping means the mapping of a texture $\tau$ onto a surface $S \subseteq M$, where $M$ - triangle mesh.

Mapping function: $f : S \mapsto \tau$.

Topology-preserving bijective mappings preferred.
Surface Segmentation  Surface Unfolding  Color Reconstruction  Color Blending

- Assuming Lambertian Bidirectional Reflectance Distribution Function.
Surface Partitioning

- Region-growing based segmentation (with normal feature).
- Region Adjacency Graphs to correct over-segmentation.
Surface Unfolding

- Using conformal mapping which preserves angles.
- Transformation of an elementary circle in the texture domain into an elementary circle on the surface.
- Minimization of discrete conformal energy that corresponds to the “non-conformality” in a least-square sense.
Minimization of discrete conformal energy that corresponds to the “non-conformality” in a least-square sense:

\[
\min \sum_{\{i,j,k\} \in F} \left\| M_T \begin{pmatrix} v_i \\ v_j \\ v_k \end{pmatrix} - \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} M_T \begin{pmatrix} u_i \\ u_j \\ u_k \end{pmatrix} \right\|^2
\]
Estimation of the correct camera image for each point (excluding occlusions) based on the neighborhood of points.

Poor resolution.
Mesh Re-parametrization

- Re-parametrization of the mesh to obtain densely sampled surface.
- Barycentric coordinates to interpolate the mapping between the vertices.
Minimization of the visibility of discontinuity artifacts.

Guidance vector field $V$ composed of color gradients and boundary constraints.
Membrane interpolant problem:

\[
\hat{f}_{1:n} = \min_{f_{1:n}} \sum_i \iint_{\Omega_i} |\nabla f_i - \mathbf{V}|^2 \]

- First boundary constraint - smooth transition.
- Second boundary constraint - invariance of gradient operator for multiplicative factors.
Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks

Dejan Pangeric
Munich, November 2011
Acquiring Semantic Object Maps for Household Tasks
Dejan Pangerlec
1. Semantic Object Maps

2. Texture Blending

3. Perception of Handles

4. Cabinet Opening

5. Conclusions
Non-reflective Handles

- min distance $h_d$
- According to “ADA Standards for Accessible Design”
- Rusu et al., ICAR2009
Specular Handles

- Kinect sensor
- Distorted projected pattern leads to erroneous matching and consequentially to invalid measurement
- **IDEA**: Lack of missing data as indicator of the presence of handles
Detect ROI in 3D using RANSAC

Create binary masks from detected ROI and invalid measurements
Handle Candidates

- Series of dilation and erosion operations
- Perform bit-wise conjunction between both masks
- Euclidean clustering using a region growing approach
- Keep clusters $C$ that correspond to the expected size and height (according to ADA)

Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks

Dejan Panceric
Pose Estimation

- Grasp point $P_{xyzrpy}$: compute convex hull $H$ around every cluster $c_i$, hull’s centroid as $YZ$ in robot’s base
- $X$ component as $h_d$ (ADA) or contact point between the handle and fingertip
- Hough line fitting for RPY orientation
2 handles, 9 different poses:

<table>
<thead>
<tr>
<th></th>
<th>positive</th>
<th>negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td>negative</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
Outline

1. Semantic Object Maps
2. Texture Blending
3. Perception of Handles
4. Cabinet Opening
5. Conclusions
Using impedance controller from Willow Garage (pr2_cockpit stack)

Initialization: gently pull the handle backwards by moving the Cartesian equilibrium point towards the robot

Record the trajectory of robot’s gripper $y_{1:n}$ with $y_i \in \text{SE}(3)$
Learning of Articulation Models - Model Fitting

Iteratively:

- (Re-)estimate the kinematic model $\mathcal{M} \in \{\text{rigid, prismatic, rotational}\}$
- Estimate model-specific parameter vector $\theta \in \mathbb{R}^d$ (encoding radius, rotation axis) of the articulated object: $\hat{\mathcal{M}}, \hat{\theta} = \arg \max_{\mathcal{M}, \theta} p(\mathcal{M}, \theta \mid y_{1:n})$
- Fit the parameter vector of all model candidates using an MLESAC estimator
Learning of Articulation Models - Model Selection

- Select the best model according the Bayesian information criterion (BIC)
- Use the model to predict the continuation of the trajectory and to generate the next Cartesian equilibrium point $x_{CEP}^{n+1}$.
- Finally, determine opening angle / opening distance
- Output: Kinematic model

VIDEO: *
Sample a noise-free trajectory $\mathbf{x}_{1:n}$ from the model over the configuration range.

Store and retrieve the trajectory from KnowRob DB (Tenorth, IROS 2009):

$$
?- \text{rdf\_triple}(\text{knowrob:'in-ContGeneric'}, \text{knowrob:'Cup67'}, B), \\
\text{rdf\_has}(B, \text{knowrob:openingTrajectory}, \text{Traj}), \\
\text{findall}(P, \text{rdf\_has}(\text{Traj}, \text{knowrob:pointOnTrajectory}, P), \\
\text{Points}).
$$
29 cabinets, 5 kitchens of Bosch and TUM, PR2 Robot, 104 trials
Evaluation of the opening distances and estimated opening radii:

<table>
<thead>
<tr>
<th>cabinet type</th>
<th>trials</th>
<th>translational err.</th>
<th>rotational err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>drawers</td>
<td>18</td>
<td>0.664cm</td>
<td>4.96°</td>
</tr>
<tr>
<td>doors</td>
<td>35</td>
<td>0.516cm</td>
<td>36.2°</td>
</tr>
<tr>
<td>average</td>
<td>53</td>
<td>0.567cm</td>
<td>25.6°</td>
</tr>
</tbody>
</table>
## Results - Whole System

<table>
<thead>
<tr>
<th>activity</th>
<th>trials</th>
<th>successes</th>
<th>success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>handle detection</td>
<td>104</td>
<td>97</td>
<td>93.3 %</td>
</tr>
<tr>
<td>operate cabinet and learn model</td>
<td>97</td>
<td>54</td>
<td>56.2 %</td>
</tr>
<tr>
<td>re-execute learned model</td>
<td>54</td>
<td>54</td>
<td>100 %</td>
</tr>
<tr>
<td>overall performance</td>
<td>104</td>
<td>54</td>
<td>51.9 %</td>
</tr>
</tbody>
</table>
Results - Whole System - Failure Causes

<table>
<thead>
<tr>
<th>type of failure</th>
<th>trials</th>
<th>failed trials</th>
<th>rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) handle detection</td>
<td>104</td>
<td>7</td>
<td>6.7%</td>
</tr>
<tr>
<td>(b) slippage of gripper</td>
<td>97</td>
<td>6</td>
<td>6.2%</td>
</tr>
<tr>
<td>(c) robot too weak at start</td>
<td>97</td>
<td>6</td>
<td>6.2%</td>
</tr>
<tr>
<td>(d) robot too weak during motion</td>
<td>97</td>
<td>23</td>
<td>23.7%</td>
</tr>
<tr>
<td>(f) overall failures in learning phase</td>
<td>97</td>
<td>43</td>
<td>44.3%</td>
</tr>
<tr>
<td>(g) sum (a)+(f)</td>
<td>104</td>
<td>50</td>
<td>48.1%</td>
</tr>
</tbody>
</table>
Outline

1. Semantic Object Maps
2. Texture Blending
3. Perception of Handles
4. Cabinet Opening
5. Conclusions
Conclusions

- End-to-end system that covers all steps required to automatically reconstruct textured SOM$^+$ models of kitchens using low-cost Kinect sensor.
- Texture reconstruction approach that consists of the following steps: surface partitioning (segmentation and slicing), surface unfolding, color reconstruction, and color blending.
- Open source
Collaborators: Ben Pitzer, Jürgen Sturm, Thomas Rühr, Martin Schuster, Moritz Tenorth, Zoltan-Csaba Marton, Michael Beetz

Contact:
{pangercic}@cs.tum.edu

Munich, November 2011

Acquiring Semantic Object Maps for Household Tasks