Object Shape Estimation and Modeling Combining Visual Data and Tactile Exploration

Yasemin Bekiroglu

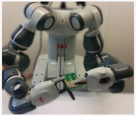
University College London

Chalmers University of Technology

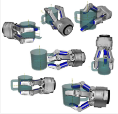
July 8, 2020

Shape Modeling

- Shape is an important parameter for various tasks.



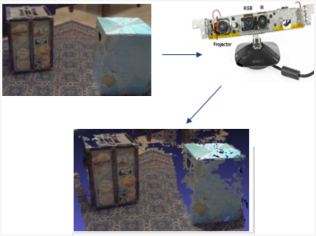
Assembly



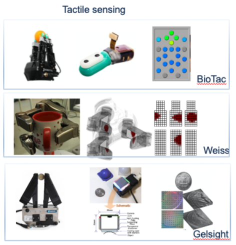
Grasp planning

Pose estimation[1]

- Sensory data is noisy and incomplete.



Visual data

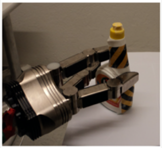


Tactile data

[1] Simtrack: A simulation-based framework for scalable real-time object pose detection and tracking, IROS 2015.

Shape Modeling[2]

- Enhance shape perception by complementing visual data with actively acquired tactile measurements.
- Find low-dimensional representation of shape data that captures manipulation affordances.

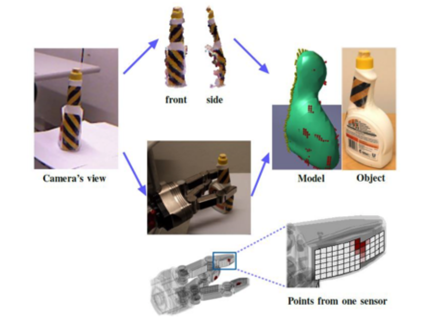


Incomplete visual data can be complemented by touch sensing.

[2] Enhancing Visual Perception of Shape through Tactile Glances, IROS 2013.

Shape Modeling[3]

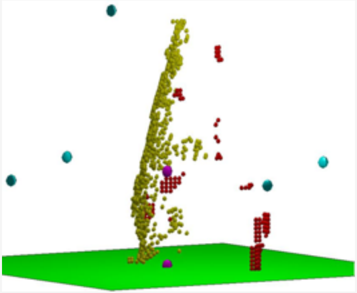
- Enhance shape perception by complementing visual data with actively acquired tactile measurements.
  - Raw sensory data → point cloud → shape



[3] Enhancing Visual Perception of Shape through Tactile Glances, IROS 2013.

Regression for Shape Modeling: Data and Representation

- Measurements given as pairs (x_i, y_i), $x_i \in \mathbb{R}^3$, $y_i \in \mathbb{R}$
  - On surface (visual and tactile data points): $y_i = 0$
  - Outside (e.g. on borders of bounding cube): $y_i = 1$
  - Inside (1 cm behind visual point cloud centroid): $y_i = -1$



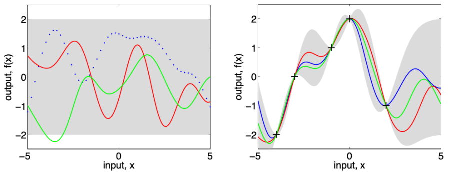
Vision points in yellow and tactile points in red.

Regression for Shape Modeling: Implicit Surfaces

- Implicit surface representation to describe shape of a given object $f(\mathbf{x}) = 0, x \in \mathbb{R}^3$
- Modeled by Gaussian Processes using a Thin Plate prior[4]: $k(\mathbf{x}, \mathbf{x}') = 2r^3 - 3Rr^2 + R^3, r = |\mathbf{x} - \mathbf{x}'|$
- Each observation subjected to Gaussian noise, $y_i = f(\mathbf{x}_i) + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$
- A GP can be interpreted as a distribution: $f(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}'))$
- For test points $\mathbf{X}^* \in \mathbb{R}^{n^*,3}$, $p(\mathbf{f}^* | \mathbf{y}, \mathbf{X}, \mathbf{X}^*) \sim \mathcal{N}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$,

$$\boldsymbol{\mu}^* = K(\mathbf{X}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1}\mathbf{y}$$

$$\boldsymbol{\Sigma}^* = K(\mathbf{X}^*, \mathbf{X}^*) - K(\mathbf{X}^*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I]^{-1}K(\mathbf{X}, \mathbf{X}^*)$$



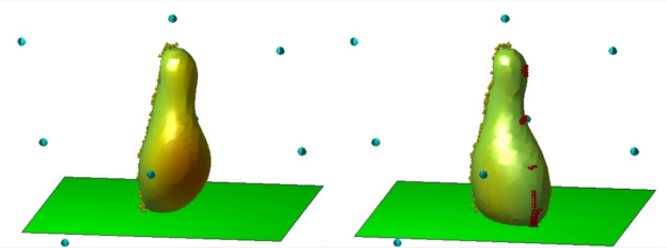
(a), prior (b), posterior

Rasmussen and Williams, Gaussian Processes for Machine Learning, 2006.

[4] R is a maximum possible value of r.

Touching actions guided by uncertainty

- Maximize information gain by touching most uncertain regions.
- Uncertainty given by variance from GP estimates.
  - Color coded surface: low uncertainty, high uncertainty

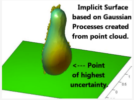


Initial surface with no touches   After 4 touches

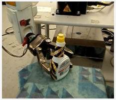
Convergence

vision only | 1 t. | 3 t. | 6 t. | 12 t. ✓ | 54 t.

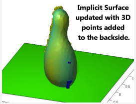
Experiments with a spray bottle



v. only

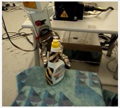


1st touch

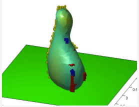


updated GPIS

all objects



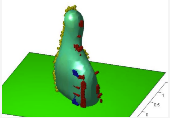
4th touch



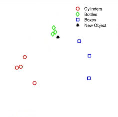
GPIS

all objects

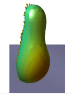
9th touch



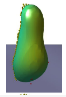
GPIS



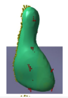
all objects



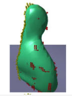
v. only



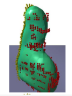
1 t.



4 t.



12 t.



54 t.

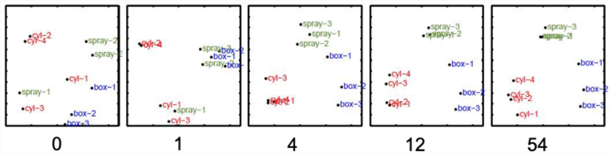
Enhancing Visual Perception of Shape through Tactile Glances, IROS 2013 CoTeSys Cognitive Robotics Best Paper Award Finalist

Shape Representation: Evaluations

- Distributions of principal curvatures from resulting surfaces.
  - Comparison by kernel based two sample test.

Similarity matrix

- Spectral clustering from similarity matrix.



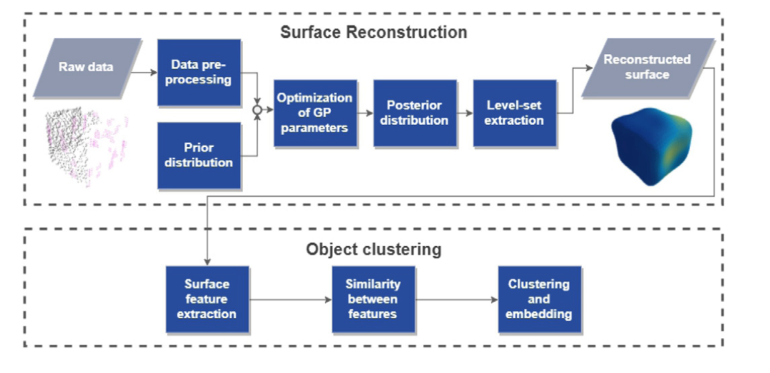
Sparse GPs

- Limitation of GPs: complexity is $\mathcal{O}(n^3)$.
- Sparse approximate method: $\mathcal{O}(n(n^u)^2)$, for $n^u \ll n$.
- n^u pairs of auxiliary input-output variables, inducing variables, $\mathbf{x}_i^u \in \mathbb{R}^3$ and $f_i^u \in \mathbb{R}$, for $i = 1, \ldots, n^u$.
- Variational approximation of posterior[5]

$$
\begin{aligned}
q(\mathbf{f}^*|\mathbf{X}^*) &\sim \mathcal{N}(\boldsymbol{\mu}^{q(\mathbf{f}^*|\mathbf{X}^*)}, \boldsymbol{\Sigma}^{q(\mathbf{f}^*|\mathbf{X}^*)}), \\
\boldsymbol{\mu}^{q(\mathbf{f}^*|\mathbf{X}^*)} &= K(\mathbf{X}^*, \mathbf{X}^u)K(\mathbf{X}^u, \mathbf{X}^u)^{-1}\boldsymbol{\mu}^{q(\mathbf{f}^u)} \\
\boldsymbol{\Sigma}^{q(\mathbf{f}^*|\mathbf{X}^*)} &= K(\mathbf{X}^*, \mathbf{X}^*) - K(\mathbf{X}^*, \mathbf{X}^u)K(\mathbf{X}^u, \mathbf{X}^u)^{-1}K(\mathbf{X}^u, \mathbf{X}^*) + \\
&\quad K(\mathbf{X}^*, \mathbf{X}^u)K(\mathbf{X}^u, \mathbf{X}^u)^{-1}\boldsymbol{\Sigma}^{q(\mathbf{f}^u)}K(\mathbf{X}^u, \mathbf{X}^u)^{-1}K(\mathbf{X}^u, \mathbf{X}^*)
\end{aligned}
$$

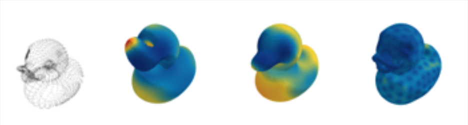
[5] Variational learning of inducing variables in sparse Gaussian processes, AISTATS 2009.

System Outline

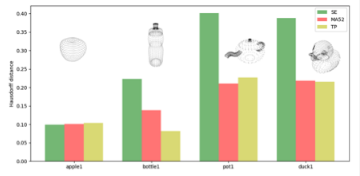


Results using Sparse GPs for Shape Modeling[6]

- Kernel selection



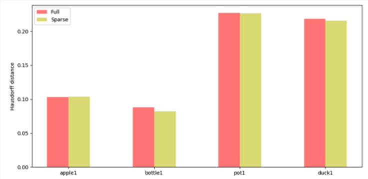
Raw data, SE, MA, TP



Distance to ground truth

- Full vs Approximation

Raw data, Full, Sparse

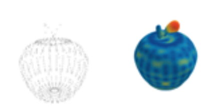


Distance to ground truth for full and Sparse GPs

[6] Object shape estimation and modeling, based on sparse Gaussian process implicit surfaces, combining visual data and tactile exploration, Robotics and Autonomous Systems, 2020.

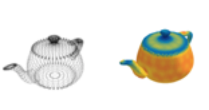
Results using Sparse GPs for Shape Modeling[7]

- More reconstructions from synthetic and real data:

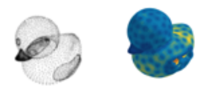


Raw data, GPIS

Raw data, GPIS

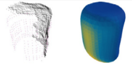


Raw data, GPIS



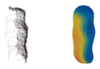
Raw data, GPIS

G. truth



Raw data, GPIS

G. truth

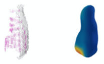


Raw data, GPIS

G. truth

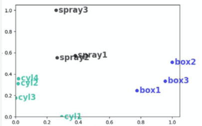
Raw data, GPIS

G. truth

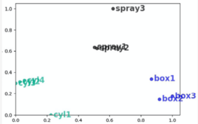


Raw data, GPIS

- Clusters from Kuka and Schunk robot data:



After a few touches

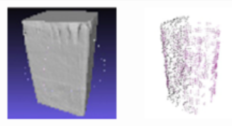


Exhaustive exploration

7 Object shape estimation and modeling, based on sparse Gaussian process implicit surfaces, combining visual data and tactile exploration, RAS, 2020.

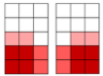
Dataset available[8]

- Visual and tactile point cloud data from two robots for shape modeling and completion.
  - 20 objects, 4 categories, ground truth scans.
  - https://data.mendeley.com/datasets/ztkctgvgw6/1



Point cloud

PR2

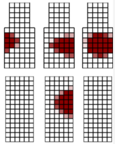


Tactile data

Name	Object	# of sensory points Visual	# of sensory points Tactile	# of vertices in scans	Name	Object	# of sensory points Visual	# of sensory points Tactile	# of vertices in scans
spray1		4252	1214	94658	bottle1		3759	478	82667
spray2		4084	1508	92439	bottle2		3044	327	83357
spray3		2937	1166	63231	bottle3		2915	321	63770

Data description

Kuka, Schunk, Kinect

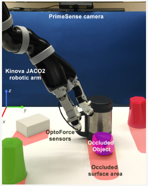


Tactile data

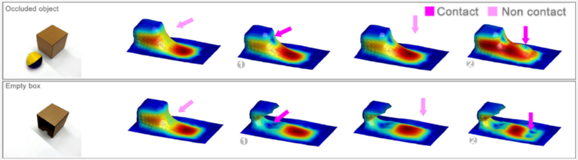
[8] Visual and Tactile 3D Point Cloud Data from Real Robots for Shape Modeling and Completion, Data in Brief, 2020.

Extensions

- Exploring surfaces and building 3D representations of the environment.[9]



Setup

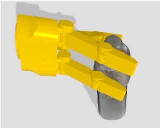


Exploration

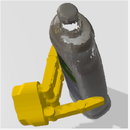
- Grasp planning using GPIS representation

Given object

Grasp configurations



Example grasp



Example grasp

9 Active Exploration Using Gaussian Random Fields and Gaussian Process Implicit Surfaces, IROS 2016

Conclusions

- Probabilistic shape models based on Implicit Surfaces using Gaussian Processes.
- Incremental refinement of shape for visually observed objects through uncertainty guided touches.
  - Resulting models similar to real object shapes.
- Some future directions: improve accuracy in reconstructions, exploration strategy (bimanual); explore geometric priors, kernel choice, category level modeling.