Chapter 4.7

Object Detection and Probabilistic Object Representation for Grasping with Two Arms

P. Ramon Soria, B.C. Arrue

Abstract This chapter explains a technique for detection and grasping of objects using two arms of an aerial robot. The method robustly obtains the 6D pose of the robot to grasp it regardless the environment. The method is based on modeling the objects surfaces under the probabilistic framework of Gaussian Processes. A probabilistic framework has been proposed to tackle the problem of shape uncertainty when the robot has partial information about the object to be manipulated. This uncertainty is modeled using GPIS and evaluated using the quality metric: probability of force closure.

1 Introduction

Object detection and surface modeling are crucial for any manipulation task. In this chapter, a method for modeling objects' surface for grasping using a probabilistic representation is presented. Nowadays, 3D sensors are increasingly being used for robotics applications as they provide grateful information about the environment in the form of 3D point clouds. These data can be used with RGB information from monocular cameras to detect objects in the scene.

However, these sensors may see their information distorted while working outdoors. For this reason, is it essential to smooth and model object's surface in a way it can be used for grasping them. At first instance, an algorithm is introduced that uses 3D information from depth sensors to create a probabilistic model of object's surface using Gaussian Process Implicit Surfaces (GPIS). The probabilistic information is used later for generating bi-manual grasps.

Manipulation with robots has been studied extensively for a long time. Mathematical and mechanical models of robot hands and their interaction with the object are a fundamental aspect in the analysis of robotic manipulation. The huge amount of combinations of objects and hands configurations make this area of research challenges. Miller et al. [187] propose the use of primitive shapes that approximate the object, facilitating the generation of grasps. More recent research uses machine

Table 28 Acronyms and symbols.

Definition	Symbol
Mean function of GP	$m(\cdot)$
Kernel function of GP	$K(\cdot, \cdot)$
Regressed values of a GP given observations	f^*
Regressed covariance of GP given observation	Σ^*
Mean values of observations a GP	μ_f^*
Mean values of GP on set points	μ_f
Values of observations used for regression	Y
Covariance of observations of a GP	K^{**}
Partial Covariance of observations and regression points	K^*
Covariance of regression points	K

learning methods to generate better grasps, where the system is trained using synthetic datasets [278, 266] and reinforcement learning [17]. Additionally to the grasp generation, it is necessary to quantify their quality. Several quality metrics are studied in [243, 161]. In this work, an approach for grasping objects by an aerial manipulator has been studied under the probabilistic framework of Gaussian Processes.

In order to be able to manipulate, it is necessary to establish the modeling of the objects. Typically, this is tackled using explicitly defined shapes, typically meshes. In [186] authors used predefined geometrical shapes as primitives. Williams and Fitzgibbon in [310] propose an algorithm that uses Gaussian Processes (GPs) for reconstruction of 3D objects in a probabilistic framework, denoted Gaussian Process Implicit Surfaces (GPIS). In this approach, the surface of the object is assigned to a level set of the regressed function from the data. This algorithm gives a probabilistic representation of the surface of the object in contrast with previous algorithms that use deterministic shapes. GPIS are also used in the literature for smoothing input data [96] or fusing data from different sensors [71, 30]. In [71, 30, 139] authors extended GPIS by integrating surface normals to improve the reconstruction of the object.

Table 28 summarizes all the symbols that will be used above.

2 Gaussian processes implicit surfaces

A Gaussian Process is a probabilistic model where each of an arbitrarily large set of variables is assumed to be jointly normally distributed. This work follows the notation given in [178]. The GP models the joint distribution of all these variables and is defined as

$$f \sim GP(m(\cdot), K(\cdot, \cdot)) \quad (213)$$

where f is the vector of random variables, $m(\cdot)$ denotes the mean function and $K(\cdot, \cdot)$ the kernel or covariance function, respectively.

Given a set of data points in 3D $\mathcal{D} = \{(x_i, y_i), \forall i = 1, \dots, N, x_i \in \mathbf{R}^3 \text{ and } y_i \in \mathbf{R}\}$, it is possible to perform regression, and for any point $x^* \in \mathbf{R}^3$,

$$f^* = \mu_f^* + K^* \cdot K^{-1} \cdot (Y - \mu_f) \quad (214)$$

$$\Sigma^* = K^{**} - K^{*T} \cdot K^{-1} \cdot K^*, \quad (215)$$

The kernel function $K(\cdot, \cdot)$ defines the correlation between data and any query points. In this work, the exponential kernel has been used,

$$K(x, y) = \exp\left(-\frac{(x - y)' * (x - y)}{2 * l^2}\right) \quad (216)$$

Authors in [310, 71, 30], propose the use of GPs to reconstruct the surface of 3D objects. The surface of the object is modeled as a level-set of f function, usually the zero-level, so that:

$$f = \begin{cases} Interior, & \text{if } f < 0 \\ Surface, & \text{if } f = 0 \\ Exterior, & \text{if } f > 0. \end{cases} \quad (217)$$

According to the previous statement, all observed data points $x_i \in \mathcal{D}$ are hence set to $y_i = 0$ as they are assumed to belong to the surface.

Additionally, to the information that the observed points lie on the surface, it is possible to acquire surface normals on the observed data-points. The camera can capture this information or computed from the observed data points. This extra information can feed the GP regression by extending the mean and covariance functions. Derivatives of a GP are also Gaussian, so the covariance functions between data points and derivatives can also be computed [238, 5]. Hence, it can also be inferred the normals on new points.

$$\begin{cases} K\left(\frac{\partial f(\mathbf{x})}{\partial x_i}, f(\mathbf{x}')\right) &= \frac{\partial k(\mathbf{x}, \mathbf{x}')}{\partial x_i} \\ K\left(\frac{\partial f(\mathbf{x})}{\partial x_i}, \frac{\partial f(\mathbf{x}')}{\partial x'_j}\right) &= \frac{\partial^2 k(\mathbf{x}, \mathbf{x}')}{\partial x_i \partial x'_j}. \end{cases} \quad (218)$$

As a result, the dataset $\mathcal{D}$ is extended as $\mathcal{D}^+ = \{(x_i, y_i), \forall i = 1 \dots N, x_i \in \mathbf{R}^3 \text{ and } y_i \in \mathbf{R}^4\}$ with y_i being the concatenation of the function value at x_i (usually zero for GPIS) and the components of the normal $[\nabla x, \nabla y, \nabla z]$. Equations 214 and 215 are modified using $Y \in \mathbf{R}^{4N}$, $K \in \mathbf{R}^{(4N) \times (4N)}$, $K^* \in \mathbf{R}^{(4N) \times 1}$, $K^{**} \in \mathbf{R}^{4 \times 4}$, $\mu_f \in \mathbf{R}^{4N}$ and $\mu_f^* \in \mathbf{R}^4$.

Up to this point, a probabilistic surface model has been presented. Grasp analysis considers the determination of whether a grasp is stable and has desirable closure properties given an object and a set of contact points. A contact point of the manipulator is represented as a point in the space and a coordinate system whose z axis

has the same direction as the surface normal but pointing inward. All information about the grasp is gathered by the Grasp Matrix ($\mathbf{G}$) [161]. Given a set of contact points $\{\vec{c}_i, \forall i = 1, \dots, n_c\}$.

The grasp matrix is essential for determining the properties and quality of a grasp. However, if the shape of the object is not certain, the measured contact points and normals can vary significantly from the real ones. The previously introduced GPIS provide a probabilistic representation of the surface and return for each grasping point the parameters of the distribution of their values.

In order to evaluate grasps there exist several quality metrics [243]. The most common one is the so-called force closure. For a given grasp, it is possible to define the Grasp Wrench Space (GWS), which is the set of wrenches generated on its contact points. A wrench (in 3D) is a 6D vector composed of the three forces and the three moments generated on a cp. It is said that the grasp has force closure if the convex hull composed by the 6D wrenches contains the origin [36].

The probability of force closure is defined as the probability of a grasp with uncertain contact points and normals of having force closure. Its analytical solution is complicated, so usually, this metric is obtained by Monte Carlo sampling. Each contact point is sampled, and then it is computed if it has force closure. The probability can then be approximated by evaluating N samples.

Authors usually set a constant mean level function for the reconstruction of the objects: $m(\cdot) = c$ [310, 71, 30], where c is a constant. However, in some situations, it is beneficial to model the mean function, and authors in [178] propose the use of non-constant mean functions. Such prior functions can be set to model geometrical shapes (such as spheres, ellipsoids, among other simple shapes.), improving the reconstruction of the expected volume of the object according to the prior. The use of prior knowledge in the shape of the object is undoubtedly helpful for grasping, as standard constant-mean functions result in holes in the volume of the objects which is undesirable for the grasping process.

In this work, it is proposed a method for adding prior information in the form of extra data instead of modifying the mean function. The data is carried with varying covariance terms associated to the errors from sensors and the detection algorithms.

For simplicity, a 1D example is considered first. Figure 173(a) shows the standard regression using a constant mean function and fixed observation noise is set for each point. Figure 173(b) shows an example of regression using a model aligned with the data. However, this does not take into account that points that lie far away from the observations cannot have high confidence. For that reason, the varying sigma is introduced depending on the distance to the data points as shown in Figure 173(c). This allows the use of custom complex mean shapes for the GPIS regression and also to increase the variance far away from the observation points. Comparing Sub-figures 173(a) and 173(c) it can be seen how both have the desired mean shape, but our model produces larger sigmas for non-confident points that are further away from the observations.

The same process can be applied to 3D shapes. Firstly, the object needs to be detected and segmented. Let D be a set of points obtained from an input sensor, which are assumed to be part of object's surface. Then, let S be a set of points that

(a) No model and fixed sigma noise

(b) Model and fixed sigma noise

(c) Model and varying sigma

Fig. 173 1D example of varying sigma for prior model for GPIS. The green circles correspond to real observations, the blue line the mean function and the pair of black lines represent the variances on the value of the function. Red asterisks are the data points from the aligned model.

models the shape of the targeted object. The algorithm aligns the stored model with the observed point cloud using RANSAC [28], which has been proved to be a robust algorithm for fitting models using 3D descriptors without initial guesses.

However, the alignment is prone to have deviations. Hence, a variance is introduced on the points of the aligned model to be used in the GPIS process. We infer a variance term for each point in the model data set according to $\sigma(x_i^{model}) = \sigma_{alignment} * e^{min_dist(x_i^{model}, observations)}$, being $\sigma_{alignment}$ the variance of the correspondences used for the alignment of the model and $min_dist(x_i^{model}, observations)$ is the minimum distance between the point and the observed points.

The overall process is shown in Figure 174. It can be seen that the mean shape of the GPIS bends towards the model points reconstructing the shape of the object correctly. It would not be possible with only a constant mean function and the ob-

servations. Other authors use the thin plate covariance kernel which minimizes the changes in the gradients, but it still fails when the gaps in the observed surfaces are large [178].

Fig. 174 Sequence for 3d object. Firstly input cloud is received, then the model is aligned and voxelized to reduce the number of points used for the GPIS regression. Finally, the regression is performed obtaining the shape of the object.

The computation time of GP regressions scales poorly with the size of the input data. For this reason, the input point cloud of the model is downsampled using any filtering. We chose a super-voxelization algorithm presented in [223], but any other simple voxel filtering or downsampling algorithm could also be used.

The grasping process is outlined in algorithm 1. Initially, a random point away from the observations is chosen. Next, using the GPIS information and a gradient descent algorithm, it is conducted towards the object's surface. Then, a point in the opposite direction of the normal of the first point is taken and converged to the surface. Finally, it is checked if the grasp is feasible for inverse kinematics (IK) of the robot. Figure 175 shows examples of grasps generated.

Algorithm 1 Grasp Generation using GPIS for parallel gripper

```

1: Choose random init point.
2: while is point on the surface? do
3:    $gradient \leftarrow GPIS(point)$ 
4:    $movepoint$ 
5: end while
6: Choose a point in the opposite direction of the normal of the first point.
7: while is point on surface? do
8:    $gradient \leftarrow GPIS(point)$ 
9:    $movepoint$ 
10: end while
11: if Is feasible for the IK of the arm? then
12:   Return grasp.
13: else
14:   Resample grasp.
15: end if

```

Fig. 175 Examples of grasps generated on a cylinder. The reconstructed surface is represented by a gray mesh. Red lines belong to the convex cone of forces generated in the grasping points.

Once grasps are generated, it can be performed a grasp planning in order to obtain the best possible grasp taking into account the constraints of the hand and the quality of the grasps. This work focuses on selecting the largest probability of force closure, but different quality metrics can also be used, such as the large minimum resisted wrench [243] or the isotropy index [243]. It can also be applied more constraints as movement constraints on the robot; for example, collision avoidance or inverse kinematics.

The whole grasping process is summarized in Algorithm 2

Algorithm 2 Grasp planning algorithm with improved shape reconstruction using GPIS

```

1: model  $\leftarrow$  load_model()
2: voxels  $\leftarrow$  super_voxels(model)
3: cloud  $\leftarrow$  sensor()
4: T  $\leftarrow$  align(cloud, model)
5: alignedVoxels  $\leftarrow$  align(voxels, T)
6: for Number grasps do
7:   Compute grasp
8:   Compute  $P_{Q_f}$ 
9: end for
10: Select best grasp

```

3 Simulations

In order to test the algorithm, partial observations (in the form of point clouds) are generated using the BlenSor [107]. BlenSor is a modified version of Blender 3D tool that can simulate different depth sensors.

Figure 176 shows the results of the surface reconstruction using our algorithm in comparison with the result without using the prior shape model and a constant mean level function.

Fig. 176 Simulation environment of crawler on a tube. The algorithm processes the input cloud, align the model using RANSAC, and performs the grasp generation.

Note that the complete surface reconstruction is not needed and it is stored implicitly in the GPIS model. It is computationally expensive. For this reason, a coarse representation is computed (Subfigure 176(d)) for demonstration purposes using the algorithm described in [178]. Actually, in Subfigure 176(f), it can be seen a grasp is generated holding the antenna.

4 Real experiments for probabilistic object's surface reconstruction for grasping

In this section, real experiments are shown in which a UAV detects a crawler robot and the probabilistic model is used to generate a meshed model which contains the uncertainty on the shape to be used for grasping. The experiments have been carried out in a hexacopter platform equipped with an Intel RealSense D435 camera, as shown in Figure 177.

Figure 178 show captures of a real experiment. At first instance, the object has been detected and its pose estimated using the 3D model and the point cloud ac-

Fig. 177 Aerial platform used for the experiments.

quired by the camera, as introduced in Sec. 2. Then, an approximated reconstruction of the object's surface has been computed using the GPIS model.

Fig. 178 Snapshots of outdoor experiments. Reconstruction of the complete surface with colored covariance from the GPIS

Furthermore, the GPIS model can calculate a covariance term for each of the points on the approximated surface of the object which is used to compute the probability of force closure for the grasp planning process. Figure 178 shows the experiment in three different instants. From left to right columns; the crawler is completely observed, for this reason, the uncertainty is uniform and low close to the observation points. In the second column, the crawler started leaving the frustum of the camera, the alignment algorithm still have enough information for acquiring an approximate pose of the target object, but as part of the object is not observed, the uncertainty increased on the backside. In the last column, the top-right part of the crawler is out of camera's sight. Thus the uncertainty increased in the right side too.

The computation of this covariance terms in the surfaces allows taking into account both the shape of the object and which part is better to be used for grasping in situations where the object cannot be completely observed.

5 Conclusions

In this chapter, we have presented a method to robustly obtain the 6D pose of the robot to grasp it regardless of the environment. The modeling of the object's surface is important to be able to plan suitable grasp for the dual manipulators described in Section 2. A probabilistic framework has been proposed to tackle the problem of shape uncertainty when the robot has partial information of the target object. This uncertainty is modeled using GPIS and evaluated using the quality metric: the probability of force closure. A probabilistic framework has been developed and presented in both simulated and real environments.