

# Enhanced GPIS learning based on local and global focus areas

Zuka Murvanidze<sup>1</sup>, Marc Peter Deisenroth<sup>1</sup>, and Yasemin Bekiroglu<sup>1,2</sup>

**Abstract**—Implicit surface learning is one of the most widely used methods for 3D surface reconstruction from raw point cloud data. Current approaches employ deep neural networks or Gaussian process models with the trade-offs across computational performance, object fidelity, and generalization capabilities. We propose a novel method based on Gaussian process regression to build implicit surfaces for 3D surface reconstruction (GPIS), which leads to better accuracy in comparison to the standard GPIS formulation. Our approach encodes local and global shape information from the data to maintain the correct topology of the underlying shape. The proposed pipeline works on dense, sparse, and noisy raw point clouds and can be parallelized to improve computational efficiency. We evaluate our approach on synthetic and real point cloud datasets obtained from laser scans, synthetic CAD objects and robot visual and tactical sensors. Results show that our approach leads to high accuracy compared to baselines.

**Index Terms**—Perception for Grasping and Manipulation, Visual Learning

## I. INTRODUCTION

**S**URFACE reconstruction is an important step for modelling the environment in many applications varying from computer vision to robotics domain. There are various approaches proposed for reconstruction of object surfaces and geometry, following analytic [1], or data-driven [2], [3] solutions. A good surface reconstruction algorithm should handle inputs with varying sampling densities and at the same time generalize well across different object topologies. Despite the drastic proliferation of 3D scanning devices, efficient and accurate reconstruction remains an active research topic [4].

Gaussian Process Implicit Surface (GPIS) representation has been proposed [5] as a probabilistic approach to surface reconstruction. In particular, it has been shown to be efficient for different tasks in robotics: learning continuous sliding paths [6], single-finger tactile exploration to guide a sensor to high uncertainty regions [7], grasp planning in 2D using only visual data [8], blind grasping by following surface contours for shape estimation [9], and grasping based on pre-trained systems



Fig. 1: Example ground truth models (the first row) and reconstructions based on our approach (the second row).

using wrist poses [10], grasping in 3D combining visual and tactile data [11], exploring surfaces and building compact 3D representations of the environment [12]. Recent research also proved its effectiveness in real-time applications, such as robot vision, using a continuous mapping from sparse measurements [13].

Standard GPIS approaches lead to approximate reconstructions capturing general shape leaving out details. We address this issue and focus on improving reconstruction accuracy capturing more details about the underlying shape. Our approach provides an efficient, generic solution that generalizes well across the different types of inputs and fits the needs of various applications without an explicit parametrization leading to high reconstruction accuracy of implicit surface learning [14] (example reconstructions can be seen in Figure 1). We compose a complex training dataset from sparse and dense observations of synthetic CAD models and real-world scans from robot visual and tactile sensors. The goal is to highlight commonly encountered problems when dealing with the different types of raw point clouds and outline the trade-offs between reconstruction accuracy and computational speed.

We present a generic pipeline for 3D surface reconstruction extending standard GPIS approach, that can work on dense sparse, and noisy point clouds. This enables reconstructions that yield higher accuracy in comparison to the standard GPIS based approaches [15]. To achieve this, we first introduce an automated data augmentation process for a raw point cloud. We then structure the input based on local and global views of the data using an Octree-based partitioning method, where we fit a global shape function based on sampling from entire object and local shape functions from partitioned segments. During inference, we compare uncertainty values between local and global models per query point and use the prediction values from the model with higher confidence. This allows us to maintain the correct topology of the shape and retain

Manuscript received: March, 25, 2022; Revised June, 18, 2022; Accepted July, 11, 2022.

This paper was recommended for publication by Editor Markus Vincze upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by Chalmers AI Research Center (CHAIR), Chalmers Gender Initiative for Excellence (Genie), and the project AIMCoR - AI-enhanced Mobile Manipulation Robot for Core Industrial Applications.

<sup>1</sup>All the authors are with the Department of Computer Science, University College London, London WC1E 6BT, U.K. <sup>2</sup>Y. Bekiroglu is also with the Department of Electrical Engineering, Chalmers University of Technology, Göteborg SE-41296, Sweden. (Corresponding author: Zuka Murvanidze, email: zuka.murvanidze.17@alumni.ucl.ac.uk; m.deisenroth@ucl.ac.uk; y.bekiroglu@ucl.ac.uk; yaseminb@chalmers.se)

Digital Object Identifier (DOI): see top of this page.

accurate information about the global shape. Fidelity of the reconstruction is boosted by the information regressed by the local models, which enable reconstruction of the intricate details. We present results using both synthetic and real datasets.

The remainder of the paper is organized as follows. Section I-A discusses related works. Section II introduces our surface reconstruction approach from local and global approximations, detailing the data augmentation, and partitioning steps. Section III presents evaluations using different datasets and baseline approaches, and Section IV provides concluding remarks and suggestions for future work.

### A. Related Work

A wide variety of methods are available for surface reconstruction. Some of the earliest research includes analytical approaches such as Alpha-Shape [16], Ball Pivoting [17], and Poisson Reconstruction [18] that convert point clouds into meshes. One of the most frequently used classic method is the Poisson Reconstruction, which works well with dense (roughly 1,000,000 points) and noise-free data, but its performance degenerates with sparse, non-uniformly sampled, or noisy observations. Implicit surface representation is another widely used approach [19], where a regression step is applied to fit an implicit function to an unorganized set of points. Various implicit function choices have been proposed, such as wavelets [20], Fourier coefficients [21] radial-basis functions [22], and multi-scale approaches [23], [24]. More recent works extend this formulation based on various data-driven approaches, e.g. deep neural networks [2], Gaussian processes [5], [15], [25].

An emerging trend in the application of deep neural networks is to learn latent feature space representations of object classes from large datasets. Scan2Mesh [26] is an example of early work in this direction. Even though this network is robust in completing the missing parts of the input, it produces non-smooth, coarse outputs and does not generalize to unseen shapes. A slightly different approach in AtlasNet [27] represents a 3D shape as a collection of parametric surface elements and, in contrast to prior work, directly infers an object's surface. Other approaches present trained models that learn the direct mapping from latent vector encodings to implicit surface functions [2], [28]–[31]. Such implicit representations are more suitable for modeling objects with complex topologies and produce visually appealing smooth reconstructions. However, they can encode only a limited number of classes of objects and reconstruct the shapes from those categories. Points2Surf [3] encodes a feature representation of local patches of geometry and relies on the assumption that local regions of objects have somewhat similar properties. These models generalize well across the unseen shapes and produce high fidelity reconstructions. However, due to its patch-based nature, it is prone to introducing topological noise. Furthermore, for each query point, the model evaluates the entire local patch coupled with the global encoding of the object. It has a high computational cost, which causes long training and inference times that make it impractical for real-time applications.

A good representation of object shape should ideally allow for the following requirements: i) encode uncertainty about the

shape, with uncertainty varying over different surface regions; ii) optimally combining different sources of information, e.g. tactile and visual. A probabilistic shape representation, Gaussian Process Implicit Surfaces, GPIS, [5], [32] is a good candidate to address these requirements. They have been shown to produce good surface reconstructions to identify or categorize objects [15], [33], [34]. After its introduction, GPIS gained popularity in the field, and its applications have quickly emerged.

Due to the flexible nature of the Gaussian process (GP), further extensions have been proposed that enhance its capabilities to fulfill even more demands across the applications. One such extension is the incorporation of geometric priors [25] to model expectations about the object's shape. The method enables the reconstruction of the missing parts from the partial object observations. However, the primary challenge is to find an appropriate geometric prior for novel data, which is a non-trivial task limiting the generalization capabilities. Priors also force reconstructed meshes to align with specific requirements resulting in overly smooth and vague outputs.

Another line of research tries to increase the accuracy of reconstruction by sacrificing the computational demand and run-time of the algorithm. GPs suffer from the cubic training complexity  $\mathcal{O}(n^3)$  with respect to the size  $n$  of the training dataset. This limits their scalability to large training sets that we often encounter in synthetic objects or high-resolution 3D scans. Many approaches [35] are being proposed by the GP community to offset this limitation using various methods that fall into two main categories; making global approximations to distillate the entire data using sparse Gaussian approximations [15], [36]–[38], and fitting local approximations involving the division of the data into learning subspace [39]–[45]. For instance, GP-MPU [39], applies multi-level-partitioning of unity and replaces polynomial functions used in the original approach [42] with Gaussian processes. Other work uses overlapping local GPs [46] with the underlying hypothesis that having the subset of points captured by the neighboring patches will improve the results. However, using the same data across multiple models diminishes the computational performance and seems redundant. The authors suggest the observation of empty spaces in reconstructed objects, where training input is sparse or non-existent. Further iterations on these models produced hierarchical [41] and skeleton-based [40] approaches, where the latter achieve comparable results to the state-of-the-art but relies on the assumption that input point cloud is dense.

The primary focus of our work is to increase the accuracy of reconstructed GPIS, given sparse, dense, and noisy point clouds, while maintaining adequate run-time and computational efficiency. We implement a segmentation algorithm to extract a critical training set from the input and employ powerful data augmentation techniques to maximize reconstruction fidelity. Simple parameterization allows our approach to adapt and fit a wide range of applications such as robot grasping, navigation, and its interaction with the environment. At the same time, it can satisfy the demands of graphics-intensive applications that require visually appealing objects with high resolution. This is achieved by implicit selection of augmentation technique, based

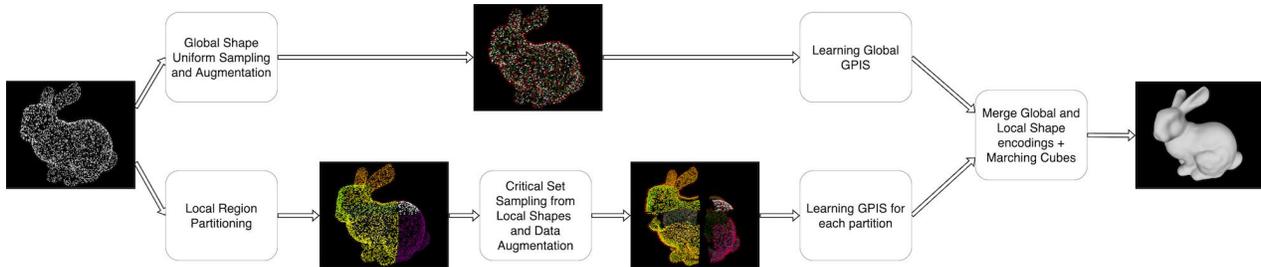


Fig. 2: System outline starting from raw point cloud input to surface reconstruction based on GP regression.

on the available training data (unless explicitly parameterized), and variable learning parameters that adjust noise and length-scale depending on individual input shape. Furthermore, unlike the approach proposed in [25], we omit any prior assumptions about the shape. Similar to [39], [40] and [41] we do partition the data into multiple local regions. However, our partitioning relies on an Octree data-structure, that allows recursive subdivision of the shape. Parent nodes in the tree encode global shape while children contain detailed information about the local shapes, such representation also allows tight control over the recursion depth, which directly correlates to reconstruction speed/fidelity trade-off. This coupled with proposed data pre-processing methods compose our approach to directly obtain implicit surface function of the 3D shape from unordered point clouds.

## II. SURFACE RECONSTRUCTION

Our method combines analytical pre-processing of a given point cloud and probabilistic learning of an implicit surface function that captures the underlying shape using Gaussian process regression. In this work, we focus on learning a model for a given object. Figure 2 illustrates multiple stages of our method, which begins with data standardization and augmentation. In the first step, we augment the data and generate additional training points that lie inside and outside of the observed point cloud. We identify the areas with more details and sample them more densely than slowly-varying regions. As a result, we obtain a critical training set, which is used in Gaussian process regression step to learn the implicit surface function. We evaluate the learned model using a cubic grid of size  $\mathbf{n}$  placed around the input object. Finally, we extract the output mesh using the Marching Cubes algorithm [47].

### A. Surface Reconstruction using Gaussian Processes

We construct an implicit surface representation based on GP regression [5] to create a model of an unknown surface. The implicit surface is defined as  $f(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$ , in which  $f(\mathbf{x})$  is the piece-wise function where  $\mathbf{x} \in \mathbb{R}^3$  is the observed point:

$$f(\mathbf{x}) = \begin{cases} -1, & \text{if } \mathbf{x} \text{ is below the surface} \\ 0, & \text{if } \mathbf{x} \text{ is on the surface} \\ 1, & \text{if } \mathbf{x} \text{ is above the surface.} \end{cases} \quad (1)$$

A GP is a stochastic process, so that any finite collection of random variables has a multivariate normal distribution. A

GP is the joint distribution of all (infinitely many) random variables and therefore it can be interpreted as a distribution over the functions  $f(\cdot)$ , and we write  $f(\cdot) \sim \mathcal{GP}(0, k(\cdot, \cdot))$ , where  $k$  is the covariance function of the GP.

Given a training set  $T$  with inputs  $\mathbf{x}_i \in \mathbb{R}^3$  and corresponding noisy observations  $y_i \in \mathbb{R}$ , the relation between the function values and corresponding observations defined as  $y_i = f(\mathbf{x}_i) + \epsilon_i$ , where  $\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)$  are independent and identically distributed noise variables, that follow a zero-mean Gaussian distribution with variance  $\sigma_n^2$ . Given a finite set  $\mathbf{X}' \in \mathbb{R}^{n'}$ ,<sup>3</sup> that comprises  $n'$  novel input points  $\mathbf{x}'$ , i.e the points outside the training set  $T$ , the distribution of  $\mathbf{f}' = [f(\mathbf{x}'_1), \dots, f(\mathbf{x}'_{n'})]^T$  is a multivariate Gaussian  $p(\mathbf{f}' | \mathbf{y}, \mathbf{X}, \mathbf{X}') = \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$ , whose mean and covariance are given by

$$\begin{aligned} \boldsymbol{\mu}' &= k(\mathbf{X}', \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y}, \\ \boldsymbol{\Sigma}' &= k(\mathbf{X}', \mathbf{X}') - k(\mathbf{X}', \mathbf{X})[k(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} k(\mathbf{X}, \mathbf{X}') \end{aligned} \quad (2)$$

respectively. Here,  $k(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{n, n}$ ,  $k(\mathbf{X}', \mathbf{X}') \in \mathbb{R}^{n', n'}$  and  $k(\mathbf{X}, \mathbf{X}') \in \mathbb{R}^{n, n'}$  comprise of elements  $k(\mathbf{x}_i, \mathbf{x}_j)$ ,  $k(\mathbf{x}'_i, \mathbf{x}'_j)$  and  $k(\mathbf{x}_i, \mathbf{x}'_j)$  accordingly, for  $i$  and  $j$  denoting a row and column indices. Thus  $k(\mathbf{X}', \mathbf{X}) = k(\mathbf{X}, \mathbf{X}')^T$ . The choice of the covariance function plays a crucial role in solving problems using Gaussian processes, because they discriminate between the functions to be selected for the inference, based on their characteristics, for instance stationarity and smoothness [48]. In our implementation, we use the radial basis function covariance function (also called squared-exponential, eq. (3)), which is characterized by the two hyper-parameters the (function) variance  $\sigma^2$  and the length scale  $l$ :

$$k^{SE}(\mathbf{x}, \mathbf{x}' | \boldsymbol{\theta}^{SE} = \{\sigma^2, l\}) = \sigma^2 \exp \left[ -\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2l^2} \right]. \quad (3)$$

Hyper-parameters are optimized using the standard training method for Gaussian processes, i.e., maximizing the marginal likelihood [48]. Empirical evaluations show that the length-scale hyper-parameter affects generalization performance across various shapes considerably, i.e. low values help capturing more details about the underlying surface features while higher values lead to smoother surfaces. The optimization procedure leads to results where we can both capture details and smooth out areas with missing points satisfactorily.

### B. Data Augmentation and Processing

In addition to the points from the raw point cloud input, auxiliary training points are generated in this step to increase the

accuracy of the resulting reconstructions. Ideally, the training set should contain points both inside and outside the object's surface. In order to generate these points we propose to use two approaches. The first method, radial augmentation, illustrated in Figure 3 on the left generates spheres inside and outside the object surface to sample auxiliary points. This data augmentation method computes two distances, between the centre of the object and its furthest and closest points. These distances are then offset by a fixed constant  $\epsilon = 0.035$  and used as radii to generate spheres inside and outside of the object surface, from which augmented points are sampled. A limitation of this method is that it makes a prior assumption about the modeled shape that the geometric center lies inside the surface. Consequently, they perform poorly when applied to the inputs that violate this premise such as torus.

The second technique, producing visually appealing and high-fidelity results, is per-point normal augmentation. The key idea is to generate two novel points for each vertex, projected in the positive and negative directions along the normal, where the projection magnitude is multiplied by a small offset value  $\epsilon$ . We set this offset empirically to be  $\epsilon = 0.035$  on normalized data (mean-centered, unit variance). Even though this augmentation method is very effective in terms of improving reconstruction accuracy, it triples the size of the initial training data. We often need to down-sample such datasets, so that non-parametric models, such as Gaussian processes, can efficiently handle them. We randomly down-sample the resulting augmented data preserving a set of points inside and outside the surface of the object.

If normal information is available in the data set, we apply one more step before down-sampling that is critical point selection. This selection step is performed by computing the dot product between the normals of the points in the same local neighbourhood grouped together based on their locations. The points that are selected as critical have normals that are sufficiently dissimilar to the normals of the neighbouring points. After this step, both critical and non-critical training points are sampled with the corresponding ratio of 4:1. This way we do not lose information about the less detailed areas of the object while including more points from complex areas of geometry.

### C. Dataset Partitioning and Local Approximations

We apply a dataset partitioning step to create local regions where we can fit local GP models for capturing more details in our reconstructions. We use Octree data structure for subdividing the space occupied by the input point cloud into local regions, which is commonly used for compact representation of 3D volume. It also enables nearest neighbor search in logarithmic time making it further useful for segmentation and processing tasks. Figure 4 shows different coloring of each such sub-region, where we fit the local GP models following the same pipeline as described in Figure 2. We set the depth of the octree to 1, producing 8 equal-sized partitions. The octree depth should be increased for high accuracy reconstructions, which in turn produces more local GPs requiring additional time and computational resources for training, unless parallelized.

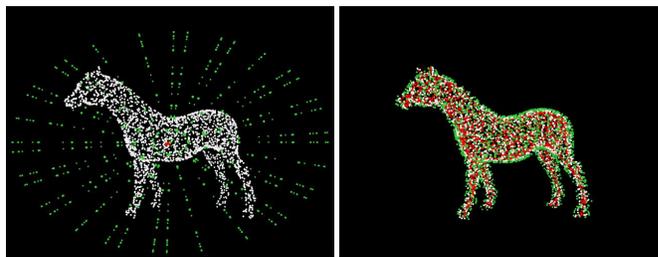


Fig. 3: Data augmentation: Radial (left) and Normal (right) based. Points inside the surface, on the surface, and outside the surface are colored in red, white, and green respectively.

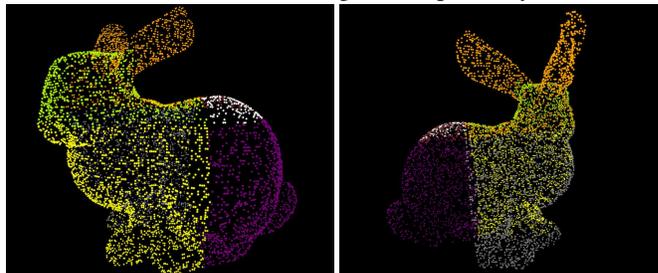


Fig. 4: Training input partitioning into local regions

### D. Merging Local and Global Approximations

The final step in the proposed pipeline is to combine global and local model predictions. Initially our distance voxel only consists of global model predictions. For each partition, we pick corresponding local model and infer distance value and uncertainty value (variance) for every query point in that partition. As a result, for each point of interest we have two predictions, one from global GPIS and one from its corresponding local GPIS. We compare the uncertainty values of these predictions and select the one with the lowest variance, which improves final surfaces in comparison to only using the local predictions. We correct the global model predictions in the distance voxel where local model obtains distance measure with lower uncertainty value.

Let  $G_0$  be a global GPIS, and  $G_1, \dots, G_n$  be local GPISs. And  $\mu_i$  and  $\sigma_i$  their corresponding means and variances. Let  $j$  denote the  $j^{th}$  query point in test sample, then for each point we select the mean value which corresponds to smaller variance:

$$\mu_{0,j} = \begin{cases} \mu_{i,j} & \text{if } \sigma_{i,j} < \sigma_{0,j} \\ \mu_{0,j} & \text{otherwise} \end{cases} \quad (4)$$

This way we merge the overall knowledge about the shape into single voxel  $\mu_0$ . This approach implicitly deals with potential discontinuities that may arise when combining different local regions together. This is because GPIS yields much higher variances at the edges of the local shapes where non-smooth (discontinued) transitions are present. In this case, global GPIS, which captures the entire geometry of the object, yields significantly lower variance, therefore the predictions from the global model are selected. This way, when two non-overlapping neighboring regions are merged, better accuracy is achieved. This volumetric grid then is fed into marching cubes algorithm to reconstruct continuous smooth surfaces with high fidelity.

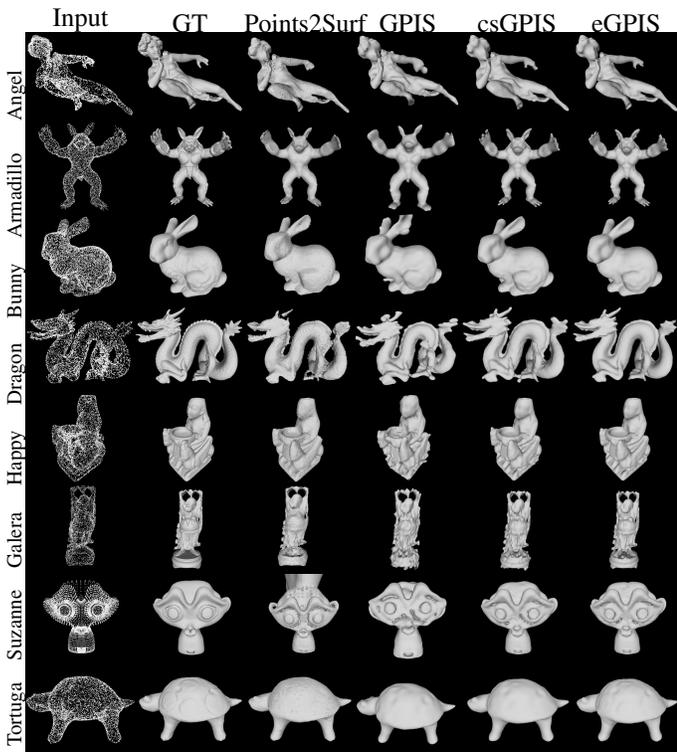


Fig. 5: Surface reconstruction results. The first column: input point cloud. GT denotes ground truth, Points2Surf is the deep learning based approach [3]. GPIS denotes the standard GP based reconstruction approach [5], [15], [32], csGPIS denotes our approach without local enhancements, and eGPIS denotes the proposed approach with all features including augmentation and local enhancements.

### III. RESULTS

In this section we present results and analysis of applying our approach for a set of reconstruction tasks, and also provide comparisons to baseline approaches. We aim to answer:

- 1) Does the data augmentation procedure help to improve GPIS reconstructions?
- 2) Do local GP estimations improve accuracy and quality of the reconstructions?
- 3) How robust is the approach to noise in the data?

We use synthetic and real datasets to evaluate our approach, eGPIS, and compare it with three baseline methods, standard GPIS approach [5], [15], [32] without the proposed data augmentation procedure, GPIS approach (denoted by csGPIS in the tables) without local enhancements using the same data augmentation introduced in the previous section, and a state of the art deep learning approach, Points2Surf [3] (P2S).

We present quantitative results in Tables I, II, and III, which detail the number of points used in the reconstruction and the distances to the ground truth using three types of distance measures: mean-square, Hausdorff, Chamfer. Table I shows the reconstruction results in terms of the distances to the ground truth for the standard GPIS approach without the new augmentation technique, and csGPIS, using the proposed data augmentation approach. We can see that the proposed augmentation approach leads the better accuracy in comparison

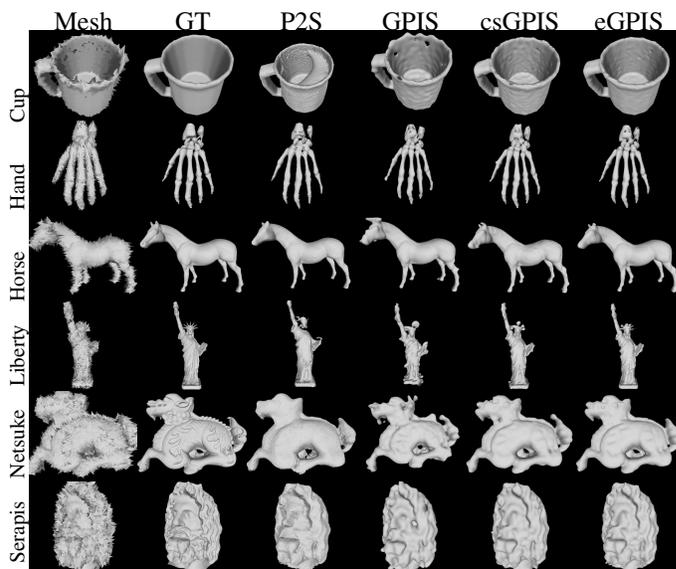


Fig. 6: Reconstructions from noisy point clouds. First Column: mesh with added Gaussian noise. GT denotes ground truth, Points2Surf is the baseline approach [3] and eGPIS is the proposed approach.

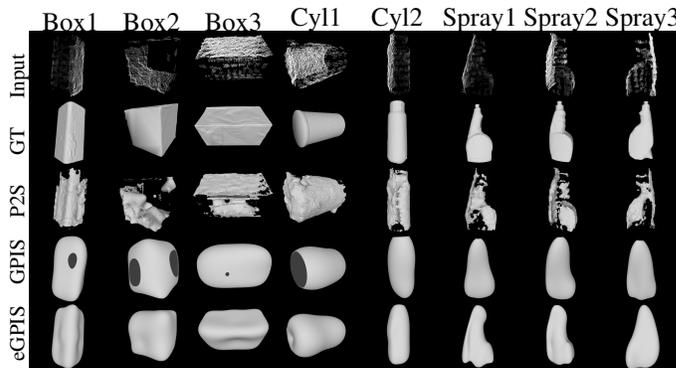


Fig. 7: Reconstructions from real robot data [49].

to the baseline GPIS approach with and without noise in the data. We also observe that denser point sampling leads to higher accuracy in reconstructions at the cost of more computational time, and that given similar number of points, eGPIS leads to higher accuracy in comparison to the baseline approaches as seen from Table I and III <sup>1</sup>.

Table II summarizes the performance differences of a deep learning approach, Points2Surf (P2S), standard GPIS approach, GPIS without local enhancements (csGPIS) and the proposed approach (eGPIS). The proposed approach eGPIS in comparison to the baselines leads to better accuracy in general for the test objects as shown by the resulting distances to the ground truth, also supported by the averaged results in the table. Resulting GPIS models demonstrate less topological noise or errors compared to P2S. In general GPIS based results are

<sup>1</sup>Data preprocessing time including critical set segmentation and augmentation steps for csGPIS and eGPIS is around 2 to 20 seconds depending on the input point cloud. P2S preprocessing time can take minutes on a high spec GPU machine.

Points	GPIS (Original)				csGPIS (Original)				GPIS (Noisy)				csGPIS (Noisy)			
	CD	HD	MD	Time	CD	HD	MD	Time	CD	HD	MD	Time	CD	HD	MD	Time
2000	3.526	31.270	10.324	0.80	<b>2.377</b>	<b>29.069</b>	<b>7.031</b>	<b>0.18</b>	36.978	31.887	11.065	0.78	<b>2.497</b>	<b>30.069</b>	<b>8.011</b>	<b>0.18</b>
3000	2.932	27.371	8.958	1.59	<b>0.840</b>	<b>18.761</b>	<b>4.198</b>	<b>0.41</b>	30.288	28.316	9.836	1.76	<b>0.637</b>	<b>20.616</b>	<b>4.537</b>	<b>0.70</b>
4000	2.833	25.963	9.027	2.80	<b>0.456</b>	<b>14.906</b>	<b>3.145</b>	<b>0.62</b>	2.818	28.286	9.154	3.38	<b>0.584</b>	<b>16.336</b>	<b>3.908</b>	<b>1.19</b>
5000	2.236	21.567	7.284	6.98	<b>0.414</b>	<b>13.627</b>	<b>3.035</b>	<b>0.90</b>	2.489	24.889	7.318	8.43	<b>0.452</b>	<b>15.862</b>	<b>3.423</b>	<b>1.48</b>
6000	2.110	21.035	6.898	9.22	<b>0.389</b>	<b>13.502</b>	<b>2.983</b>	<b>1.15</b>	2.194	21.027	7.333	10.00	<b>0.421</b>	<b>15.546</b>	<b>3.134</b>	<b>2.38</b>
9000	2.035	20.996	3.910	22.12	<b>0.338</b>	<b>12.160</b>	<b>2.625</b>	<b>2.51</b>	1.783	20.165	6.913	34.23	<b>0.392</b>	<b>13.160</b>	<b>2.625</b>	<b>3.67</b>
10000	1.856	18.003	3.852	23.48	<b>0.302</b>	<b>12.033</b>	<b>2.596</b>	<b>3.47</b>	1.649	17.117	4.156	35.17	<b>0.345</b>	<b>12.809</b>	<b>2.596</b>	<b>5.26</b>

TABLE I: Results averaged across all objects, GPIS vs csGPIS comparison for original and noisy inputs, distances are based on measurements in mm and time in minutes. Minimum values per distance and elapsed time for each approach and experiment (with and without noise) are given in bold.

Name	P2S			GPIS			csGPIS			eGPIS		
	CD	HD	MD	CD	HD	MD	CD	HD	MD	CD	HD	MD
angel	90.607	17.226	5.445	0.049	6.758	0.963	0.032	6.577	0.925	<b>0.029</b>	<b>6.133</b>	<b>0.847</b>
armadillo	2.920	13.377	4.427	0.094	10.291	1.748	0.077	8.961	1.460	<b>0.074</b>	<b>8.399</b>	<b>1.439</b>
bunny	81.713	44.280	98.086	3.361	61.306	<b>9.532</b>	2.594	39.887	9.867	<b>2.288</b>	<b>38.212</b>	<b>9.739</b>
cup	88.272	48.203	8.406	1.944	30.050	10.123	0.136	15.116	1.938	<b>0.129</b>	<b>14.168</b>	<b>1.878</b>
dragon	13.973	9.260	0.477	0.043	<b>8.956</b>	1.149	0.044	9.325	1.149	<b>0.040</b>	<b>9.007</b>	<b>1.058</b>
galera	14.752	42.649	<b>0.452</b>	5.179	34.944	17.033	0.166	10.774	<b>2.408</b>	<b>0.160</b>	<b>10.662</b>	<b>2.475</b>
hand	15.588	62.683	<b>0.699</b>	10.033	40.317	21.887	0.017	2.761	0.786	<b>0.011</b>	<b>2.419</b>	<b>0.760</b>
happy	49.724	<b>6.630</b>	2.386	6.056	41.859	19.176	0.055	<b>8.310</b>	1.282	<b>0.051</b>	<b>8.476</b>	<b>1.232</b>
horse	86.058	151.203	53.938	0.003	1.796	0.317	<b>0.001</b>	<b>1.039</b>	<b>0.168</b>	<b>0.003</b>	<b>1.718</b>	<b>0.313</b>
liberty	99.856	16.237	5.703	0.037	7.138	0.923	0.031	6.962	0.909	<b>0.030</b>	<b>6.911</b>	<b>0.858</b>
netsuke	16.580	<b>7.742</b>	<b>0.452</b>	0.119	10.334	<b>2.071</b>	<b>0.116</b>	9.323	<b>2.119</b>	<b>0.116</b>	<b>9.033</b>	2.121
serapis	16.452	<b>3.225</b>	<b>0.547</b>	0.290	11.757	3.599	<b>0.273</b>	11.333	<b>3.539</b>	<b>0.274</b>	<b>11.219</b>	<b>3.558</b>
suzanne	44.418	24.986	49.538	1.010	24.862	<b>7.173</b>	0.932	24.752	<b>7.191</b>	<b>0.925</b>	21.293	<b>7.348</b>
tortuga	12.228	<b>2.059</b>	<b>0.360</b>	0.274	<b>3.572</b>	<b>1.090</b>	<b>0.254</b>	15.126	3.007	0.266	<b>13.329</b>	<b>2.986</b>
Average	59.510	32.126	16.494	2.035	20.996	6.913	0.338	12.160	2.625	<b>0.314</b>	<b>11.499</b>	<b>2.615</b>

TABLE II: Results per-object: Baseline approaches P2S and GPIS vs the proposed csGPIS and eGPIS. Distances are based on measurements in mm. Minimum values per distance for each approach are given in bold, the second best result in red and the third best result in cyan.

Global	Local	Original				Noisy			
		CD	HD	MD	Time	CD	HD	MD	Time
5000	3000	<b>0.345</b>	<b>13.188</b>	<b>2.708</b>	<b>1.47</b>	0.362	14.094	2.814	2.74
8000	5000	<b>0.330</b>	<b>12.612</b>	<b>2.619</b>	<b>6.16</b>	0.358	13.230	2.735	7.06
9000	5000	<b>0.286</b>	<b>11.858</b>	<b>2.522</b>	<b>15.97</b>	0.316	12.249	2.624	16.51
10000	7000	<b>0.274</b>	<b>11.596</b>	<b>2.561</b>	<b>18.10</b>	0.291	12.239	2.601	19.83
11000	7000	<b>0.206</b>	<b>11.118</b>	<b>2.531</b>	<b>21.54</b>	0.284	12.192	2.590	22.94

TABLE III: Results across all objects using eGPIS. Distances are based on measurements in mm and time in minutes.

Name	GPIS	P2S	eGPIS
box1	2.8995	12124.6591	<b>2.1812</b>
box2	2.4051	7708.5759	<b>2.1053</b>
box3	2.3028	18928.7009	<b>1.9836</b>
cyl1	3.2799	9470.8379	<b>2.081</b>
cyl2	2.9721	8729.1513	<b>2.1625</b>
spray1	2.4376	11309.7142	<b>2.2386</b>
spray2	2.5245	12296.1702	<b>2.105</b>
spray3	2.7742	11945.5462	<b>2.0406</b>
Average	2.6994	11564.1694	<b>2.1122</b>

TABLE IV: Results from real robot dataset comparing Chamfer distances to the ground truth.

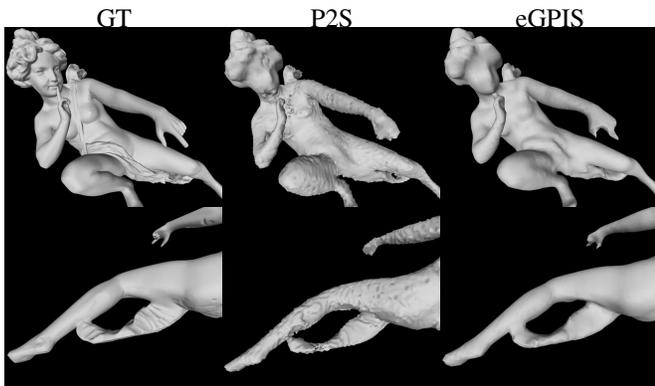


Fig. 8: Close inspection of the reconstructions by P2S and eGPIS. GT denotes ground truth reconstructions.

better than the deep learning approach, consistently based on the Chamfer distance. As for the other two distance measures P2S leads to better accuracy for a few objects such as *tortuga*,

*serapis*, *netsuke* where the GPIS based approaches tend to smooth out very fine and sharp details.

We visualize the reconstruction results of input point clouds for experiment objects in Figure 5 for synthetic objects, Figure 6 with noisy inputs for the synthetic objects, and Figure 7 from real data collected by a robot [49], using the baseline approaches and the proposed approach. Results from Figure 5 demonstrate the eGPIS can both capture fine details of surfaces while at the same time providing smoothness in comparison to the baselines. Figure 8 provides a closer look at reconstructed surfaces from an object model that has fine details. As seen, the baseline approach P2S tend to generate details that are not available on the original surface also missing existing details such as the fingers, while eGPIS produces smooth surfaces and captures those missing details better. Figure 6 presents the experiment results where the original input point clouds are corrupted by noise. Proposed csGPIS and eGPIS lead to better reconstructions in comparison to the other baselines and eGPIS in general produces the best results in terms of

reconstruction accuracy, completeness and smoothness. These characteristics also appear in the experiments where we use real data generated using tactile and visual measurements in Figure 7. eGPIS based reconstructions lead to complete, smooth, most accurate surfaces in comparison to the baseline approaches P2S and GPIS, yielding the smallest distances to the ground truth data as shown in TableIV.

As for computational requirements, the deep learning approach, Points2Surf, was trained for approximately 5 days on a specialized hardware. In comparison, our approach regresses independent models per object without parallelizing or GPU acceleration. P2S does heavy pre-processing using Blender software to sample and convert point clouds into appropriate format, while our approach has more rapid and efficient pre-processing. Inference time in P2S, heavily depends on the hardware and GPU specifications, however, single CPU suffices efficient inference in our method and the partitioning together with the local GP calculations can also be easily parallelized to improve efficiency.

#### IV. CONCLUSION

The paper presents a 3D reconstruction approach to build implicit surface representation of a given object point cloud observation via Gaussian process regression. The Gaussian process formulation for surface reconstruction has been shown to model unknown object shape taking into account the uncertainties in observations successfully. In this paper, we further extend this approach and demonstrate that GPIS formulation can yield high quality reconstructions by exploiting local and global models. Results on synthetic and real-world datasets demonstrate that Gaussian processes outperform baseline approaches both in reconstruction accuracy, completeness and smoothness of the resulting surfaces.<sup>2</sup> We plan to further improve the quality of the reconstructed surfaces where fine details need to be captured. Another future work direction is to learn generic models based on training data including various objects rather than learning models per a given object.

#### REFERENCES

- [1] A. Pentland and S. Sclaroff, "Closed-form solutions for physically based shape modeling and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, pp. 715–729, 1991.
- [2] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning continuous signed distance functions for shape representation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [3] P. Erler, P. Guerrero, S. Ohrhallinger, M. Wimmer, and N. J. Mitra, "Points2surf: Learning implicit surfaces from point cloud patches," *arXiv:2007.10453*, 2020.
- [4] M. Berger, A. Tagliasacchi, L. M. Seversky, P. Alliez, J. A. Levine, A. Sharf, and C. T. Silva, "State of the art in surface reconstruction from point clouds," in *Eurographics 2014—State of the Art Reports*. The Eurographics Association, 2014.
- [5] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," in *Gaussian Processes in Practice*, 2007.
- [6] D. Driess, P. Englert, and M. Toussaint, "Active learning with query paths for tactile object shape exploration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2017.
- [7] Z. Yi, R. Calandra, F. Veiga, H. van Hoof, T. Hermans, Y. Zhang, and J. Peters, "Active tactile object exploration with Gaussian processes," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.
- [8] J. Mahler, S. Patil, B. Kehoe, J. van den Berg, M. Ciocarlie, P. Abbeel, and K. Goldberg, "GP-GPIS-OPT: Grasp planning with shape uncertainty using Gaussian process implicit surfaces and sequential convex programming," in *IEEE International Conference on Robotics and Automation*, 2015.
- [9] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *IEEE International Conference on Robotics and Automation*, 2011.
- [10] S. Ottenhaus, D. Renninghoff, R. Grimm, F. Ferreira, and T. Asfour, "Visuo-haptic grasping of unknown objects based on Gaussian process implicit surfaces and deep learning," in *IEEE International Conference on Humanoid Robots*, 2019.
- [11] C. de Farias, N. Marturi, R. Stolkin, and Y. Bekiroglu, "Simultaneous tactile exploration and grasp refinement for unknown objects," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, 2021.
- [12] S. Caccamo, Y. Bekiroglu, C. H. Ek, and D. Kragic, "Active exploration using Gaussian random fields and Gaussian process implicit surfaces," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2016.
- [13] B. Lee, C. Zhang, Z. Huang, and D. D. Lee, "Online continuous mapping using Gaussian process implicit surfaces," in *IEEE International Conference on Robotics and Automation*, 2019.
- [14] G. Turk and J. F. O'Brien, "Variational implicit surfaces," 1999.
- [15] G. Z. Gandler, C. H. Ek, M. Björkman, R. Stolkin, and Y. Bekiroglu, "Object shape estimation and modeling, based on sparse Gaussian process implicit surfaces, combining visual data and tactile exploration," *Robotics and Autonomous Systems*, vol. 126, 2020.
- [16] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, "On the shape of a set of points in the plane," *IEEE Transactions on Information Theory*, vol. 29, no. 4, pp. 551–559, 1983.
- [17] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 5, no. 4, pp. 349–359, 1999.
- [18] M. B. Kazhdan and H. Hoppe, "Poisson surface reconstruction," *Eurographics*, 2006.
- [19] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, "Surface reconstruction from unorganized points," *SIGGRAPH Comput. Graph.*, vol. 26, no. 2, p. 71–78, 1992.
- [20] J. Manson, G. Petrova, and S. Schaefer, "Streaming surface reconstruction using wavelets," *Computer Graphics Forum*, vol. 27, pp. 1411–1420, 2008.
- [21] M. Kazhdan, "Reconstruction of solid models from oriented point sets," in *Eurographics Symposium on Geometry Processing*, 2005.
- [22] Y. Ohtake, A. Belyaev, and H.-P. Seidel, "3D scattered data interpolation and approximation with multilevel compactly supported RBFs," *Graphical Models*, vol. 67(3), pp. 150–165, 05 2005.
- [23] Y. Ohtake, A. Belyaev, and H. P. Seidel, "A multi-scale approach to 3D scattered data interpolation with compactly supported basis functions," in *Shape Modeling International.*, 2003, pp. 153–161.
- [24] Y. Nagai, Y. Ohtake, and H. Suzuki, "Smoothing of partition of unity implicit surfaces for noise robust surface reconstruction," in *Symposium on Geometry Processing*, 2009.
- [25] W. Martens, Y. Poffet, P. Soria, Ramón, R. Fitch, and S. Sukkarieh, "Geometric priors for Gaussian process implicit surfaces," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 373–380, 2017.
- [26] A. Dai and M. Niessner, "Scan2Mesh: from unstructured range scans to 3D meshes," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019.
- [27] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "AtlasNet: A papier-mâché approach to learning 3D surface generation," *arXiv:1802.05384*, 2018.
- [28] L. Mescheder, M. Oechsle, M. Niemeyer, S. Nowozin, and A. Geiger, "Occupancy networks: Learning 3D reconstruction in function space," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [29] J. Mu, W. Qiu, A. Kortylewski, A. Yuille, N. Vasconcelos, and X. Wang, "A-SDF: learning disentangled signed distance functions for articulated shape representation," in *International Conference on Computer Vision*, 2021.
- [30] V. Sitzmann, E. Chan, R. Tucker, N. Snavely, and G. Wetzstein, "MetaSDF: meta-learning signed distance functions," in *Advances in Neural Information Processing Systems*, 2020.
- [31] K. Gupta, S. Jabbireddy, K. Shah, A. Shrivastava, and M. Zwicker, "Improved modeling of 3D shapes with multi-view depth maps," *arXiv:2009.03298*, vol. abs/2009.03298, 2020.

<sup>2</sup>The source code can be accessed via <https://github.com/Zuka98/eGPIS>.

- [32] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," 2011.
- [33] M. Björkman, Y. Bekiroglu, V. Högman, and D. Kragic, "Enhancing visual perception of shape through tactile glances," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013, pp. 3180–3186.
- [34] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Active learning with Gaussian processes for object categorization," in *IEEE International Conference on Computer Vision*, 2007.
- [35] H. Liu, Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian process meets big data: A review of scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, pp. 4405–4423, 2020.
- [36] M. K. Titsias, "Variational learning of inducing variables in sparse Gaussian processes," in *Artificial Intelligence and Statistics*, 2009, pp. 567–574.
- [37] J. Quiñero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *Journal of Machine Learning Research*, vol. 6, no. 12, pp. 1939–1959, 2005.
- [38] J. Hensman, A. Matthews, and Z. Ghahramani, "Scalable variational Gaussian process classification," in *Artificial Statistics and Machine Learning*, 2015.
- [39] M. G. López, B. Mederos, and O. Dalmau, "GP-MPU method for implicit surface reconstruction," in *Human-Inspired Computing and Its Applications*, 2014, pp. 269–280.
- [40] L. Wu, R. Falque, V. Perez-Puchalt, L. Liu, N. Pietron, and T. Vidal-Calleja, "Skeleton-based conditionally independent Gaussian process implicit surfaces for fusion in sparse to dense 3D reconstruction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1532–1539, 2020.
- [41] S. Kim and J. Kim, "Hierarchical Gaussian processes for robust and accurate map building," in *IEEE International Conference on Robotics and Automation*, 2015.
- [42] Y. Ohtake, A. Belyaev, M. Alexa, G. Turk, and H.-P. Seidel, "Multi-level partition of unity implicits," *ACM Transactions on Graphics*, vol. 22, no. 3, pp. 463–470, 2003.
- [43] V. Tresp, "A Bayesian committee machine," *Neural Computation*, 2000.
- [44] Y. Cao and D. J. Fleet, "Transductive log opinion pool of Gaussian process experts," *arXiv:1511.07551*, 2015.
- [45] S. Cohen, R. Mubvha, T. Marwala, and M. P. Deisenroth, "Healing products of Gaussian process experts," in *International Conference on Machine Learning*, 2020.
- [46] S. Kim and J. Kim, "Continuous occupancy maps using overlapping local Gaussian processes," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.
- [47] T. Lewiner, H. Lopes, A. W. Vieira, and G. Tavares, "Efficient implementation of marching cubes' cases with topological guarantees," *Journal of Graphics Tools*, vol. 8, no. 2, pp. 1–15, 2003.
- [48] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- [49] Y. Bekiroglu, M. Bjorkman, G. Z. Gandler, J. Exner, C. H. Ek, and D. Kragic, "Visual and tactile 3D point cloud data from real robots for shape modeling and completion," *Data in Brief*, vol. 30, 2020.