

MIT Open Access Articles

Embedded Object Detection and Mapping in Soft Materials Using Optical Tactile Sensing

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: SN Computer Science. 2024 Mar 29;5(4):372

As Published: 10.1007/s42979-024-02731-6

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/154045>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution





Embedded Object Detection and Mapping in Soft Materials Using Optical Tactile Sensing

Jose A. Solano-Castellanos¹ · Won Kyung Do² · Monroe D. Kennedy III²

Received: 15 December 2023 / Accepted: 17 February 2024
© The Author(s) 2024

Abstract

In this paper, we present a methodology that uses an optical tactile sensor for efficient tactile exploration of embedded objects within soft materials. The methodology consists of an exploration phase, where a probabilistic estimate of the location of the embedded objects is built using a Bayesian approach. The exploration phase is then followed by a mapping phase which exploits the probabilistic map to reconstruct the underlying topography of the workspace by sampling in more detail regions where there are expected to be embedded objects. To demonstrate the effectiveness of the method, we tested our approach on an experimental setup that consists of a series of quartz beads located underneath a polyethylene foam that prevents direct observation of the configuration and requires the use of tactile exploration to recover the location of the beads. We show the performance of our methodology using ten different configurations of the beads where the proposed approach is able to approximate the underlying configuration. We benchmark our results against a random sampling policy. Our empirical results show that our method outperforms the fully random policy in both the exploration and mapping phases. The exploration phase produces a better probabilistic map with fewer samples which enables an earlier transition to the mapping phase to reconstruct the underlying shape. On both the exploration and mapping phases, our proposed method presents a better consistency as compared to the random policy, with smaller standard deviation across the ten different bead configurations.

Keywords Force and tactile sensing · Sensor-based control · Probabilistic inference · Active perception · Shape reconstruction

Introduction

The sense of touch is fundamental for humans to safely and robustly interact with their environment, it is arguably the most informative sense, in some instances more so than the sense of sight. With a single touch humans can detect temperature, surface smoothness, object shape and orientation, and object hardness. With minimal effort, people can

leverage their sense of touch in occluded spaces like pockets or drawers and accurately assess what is being touched.

While robotics have seen a significant increase in capabilities, dexterity that is comparable to humans has not yet been achieved. In the last decade, a great effort has been dedicated to providing robots the ability to sense the environment through touch, specifically with optical tactile sensors. Examples of such sensors are GelSight [1], which is able to provide high-resolution relief maps of object surfaces; Soft-Bubble [2] which has shown to be able to classify, estimate the pose and tracking objects; and DenseTact 2.0 [3], capable of not only reconstruct the shape of the object but also provide a 6-axis wrench estimation. In this paper, we have decided to use DenseTact 2.0, which has a force range of -11 to 3 N in the vertical direction. Furthermore, DenseTact 2.0 has been shown to provide an accuracy of 0.3633 mm per pixel for shape reconstruction, 0.410 N for forces and 0.387 N mm for torques [3]. Successful applications of DenseTact 2.0 in robotics tasks include manipulation of small objects for in-hand reorientation and identification for

✉ Jose A. Solano-Castellanos
jsolanoc@mit.edu

Won Kyung Do
wkdo@stanford.edu

Monroe D. Kennedy III
monroek@stanford.edu

¹ Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA

² Department of Mechanical Engineering, Stanford University, 438 Panama St., Stanford, CA 94305, USA

effective classification [4], as well as the ability to pick up small objects from flat surfaces, where the relative geometry of the sensor to the object poses a challenge [5]. For more details on the manufacturing and training, please consult [6].

Tactile information for robotic manipulation is particularly appealing when the sense of sight cannot provide reliable information to complete the task. A guiding example is a person forming an inventory of objects within an occluded pocket without taking the objects out for visual inspection. This task requires the classification of soft versus hard objects and the person may explore the environment through tactile motion primitives such as pinching, squeezing, or pushing to identify the contents.

Robots with this capability will likely have a significant societal impact. For industrial applications, robots would be able to classify packages at distribution centers without having to open them; for in-home care, robots that are equipped with this capacity could assist humans in retrieving objects hidden within soft objects such as blankets or within clothes in a more efficient and natural way. In medical applications, tactile sensation could aid physicians in assessing superficial anatomical structures providing reliable relief maps for further evaluation. Even for bionic prosthesis, forming a relief map and reproducing the haptic sensation on another part of the user's body could provide the user with the sensation of touch that was lost (Fig. 1).

Extracting Local Information Through Touch

Optical tactile sensors to estimate mechanical properties and differentiate heterogeneous objects have proven to be an effective sensing method in previous work. Yuan et al. [7, 8] used the GelSight sensor to estimate the hardness of everyday objects with different shapes, improving the object's recognition that could provide information for an appropriate

grasping strategy. Yuan et al. [9] also demonstrated the use of optical tactile sensors to actively recognize properties in clothing including characteristics such as thickness, fuzziness, softness, and durability and even estimate characteristics such as preferred washing methods or the season it is meant to be worn.

In the medical setting, Gwilliam et al. [10] proposed the use of a capacitive tactile sensor for the detection of lumps in soft tissues of variable size and depth. The sensor was compared with the human finger outperforming it in the detection of lumps, requiring lower indentation depths and pressures. Jia et al. [11] later replicated the work using GelSight instead of the capacitive sensor, also showing better performance than the human finger.

Active Exploration for Shape Reconstruction

The above-mentioned works show the efficacy of the use of optical tactile sensors to infer local properties of the objects. However, rich interactions with the environment require not only the capacity to detect local properties such as hardness, but also to gain global information of the object such as shape. One common choice for shape representation using haptic sensory information is through the use of implicit surfaces via Gaussian Processes (GP). Dragiev et al. [12] show the performance of implicit surfaces GP shape estimation using tactile information from a seven-degrees-of-freedom hand.

Efficient shape reconstruction requires active and purposeful exploration of objects' surfaces. Martins & Ferreira [13] implemented a haptic exploration in simulation that allowed a humanoid robot to reconstruct the boundary between two different materials using a Bayesian model. Jamali et al. [14] used a GP to probabilistically reconstruct the shape of the object driving the exploration based on the estimated boundaries of the object. A similar approach was shown to work by Yi et al. [15]. A more recent work involving the use of GelSight was developed by Wang et al. [16] integrating the use of visual and tactile exploration with learned shape priors for 3D shape reconstruction. Liu et al. [17] demonstrated the application of optical tactile sensing along the entire surface of a finger for a three-finger robotic hand which enabled the recognition of a set of rigid objects with a single grasp.

Merging both local and global information for shape reconstruction was recently explored by Zhao et al. [18] with FingerSLAM. This method uses both visual and tactile information to reconstruct the relief map by loop closure of the two sensing modalities. This method requires direct visualization of the surface and has only been tested on rigid objects.

As shown before, shape reconstruction has been extensively studied for rigid and homogeneous objects using

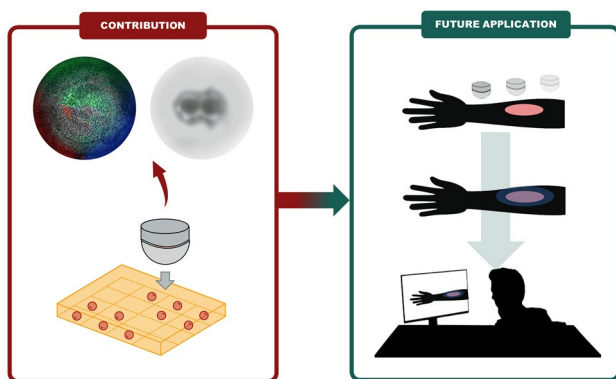


Fig. 1 Embedded object detection. In this work, we demonstrate the ability to detect and map objects embedded in a soft medium. This work has future applications which include medical automated assistance for patient palpation

tactile information. When the object is heterogeneous and/or deformable, previous work has been focused on gathering local information to obtain mechanical properties or estimate the presence of anomalies such as lumps in medical applications. To the best of our knowledge, simultaneous detection and mapping for heterogeneous and deformable objects has not been addressed in the past. This paper makes a significant contribution toward addressing this goal.

The main contributions of this paper are: first, we generate a probabilistic map of embedded objects locations using a Bayesian approach that samples a large region, improving on previous methods that sampled only local regions for object presence estimation. Second, we use this probabilistic map to drive the exploration of the workspace and approximate the shape of the embedded objects (heterogeneous medium), different from other papers where such shape reconstruction is done on rigid objects (homogeneous medium). Finally, our method is able to detect and approximate the shapes of closely embedded and discontinuous objects within the soft material, independent of their continuity, relying solely on their spatial grouping within the workspace.

The paper is divided in the following sections: Sect. 2 describes the proposed framework which is divided into an exploration and a mapping phase. Each phase is driven, respectively, by a reduction of uncertainty and the maximization of the probability of the presence of embedded objects. Section 3 describes the experimental setup that was used to show the efficacy of our method using ten different configurations of the embedded objects. Section 4 presents the empirical results obtained across the ten different configurations where we show that our framework outperforms a fully random policy on both the exploration and mapping phases.

The project website which contains results shown in figures and videos can be found here: sites.google.com/stanford.edu/embeddedobjectdetection/

Exploration and Mapping

We propose a framework for efficient tactile exploration and mapping for embedded rigid objects within matrices of soft materials using an optical tactile sensor. The method first generates a map that indicates the presence of hard objects below the surface (exploration). In this phase, a probabilistic map is iteratively built by concurrently minimizing the expected uncertainty (as quantified by a Gaussian Process) and maximizing the spread of the samples (as measured by the space-filling metric of the discrepancy) for efficient exploration. The exploration is followed by a more thorough interaction in the areas of interest for an approximate reconstruction of the underlying topography (mapping). This is achieved by maximizing both the probability of sampling

an embedded object as well as the spread of the samples. The method is tested in an experimental setup that involves a series of bead clusters in a planar array which is then covered by a polyethylene foam. Figure 2 graphically represents the proposed method.

It is important to highlight that our proposed method follows the Sense-Plan-Act autonomy methodology of robotics as depicted in Fig. 3.

Exploration

The exploration phase of the method, which provides a probabilistic map of the location of the hard objects below the soft surface starts by sampling the workspace (*E.1* in Fig. 2). Assuming that initially there is no prior information of the location of the embedded objects, the sampling can be performed using any desired strategy (e.g. randomized sampling). It is desired to get a comprehensive understanding of the workspace avoiding oversampling some regions of the workspace or leaving some regions unexplored. A sampling plan that encourages a diversity of samples and maximizes coverage of the workspace is called a space-filling sampling plan. There exist many sampling plans that are considered space-filling, from which we have tested the Latin-hypercube sampling, Halton Sequence, and Sobol Sequences. Among these sampling strategies, including a randomized sampling plan, we found empirically that Sobol Sequences [19] yields the best initialization for our experimental setup with eight initial samples to form the prior. Naturally, in an instance where a prior is available, the use of a sampling strategy can be omitted.

Using DenseTact 2.0, each sample produces a $640 \times 640 \times 1$ depth image (*E.2* in Fig. 2). The image is subsequently processed to check for the presence or absence of hard objects below the surface. To this end, the depth image obtained from the sample is compared against the image of the undeformed sensor, and a threshold ϵ is applied: if the deformation (in absolute value) $|\delta|$ associated with a given pixel is larger than ϵ , then the depth information is preserved, otherwise the deformation is set to zero. The value of ϵ is closely related to the error of the sensor: if there is a deformation smaller than the error of the sensor, it is likely that it is noise from the sensor and not an actual deformation on the sensor. However, the presence of the soft material may affect the value of the threshold whose influence is difficult to quantify and deserves an analysis of its own. From our experiments, it was observed that the value of ϵ needed to be set to 1.55 mm to produce a good performance of our method. The depth image is saved as a point cloud for later use in the mapping phase of the algorithm.

Once the threshold is applied and deformation is detected in the depth image, the next step is to find the centroid of the indentations (*E.3* in Fig. 2). The depth

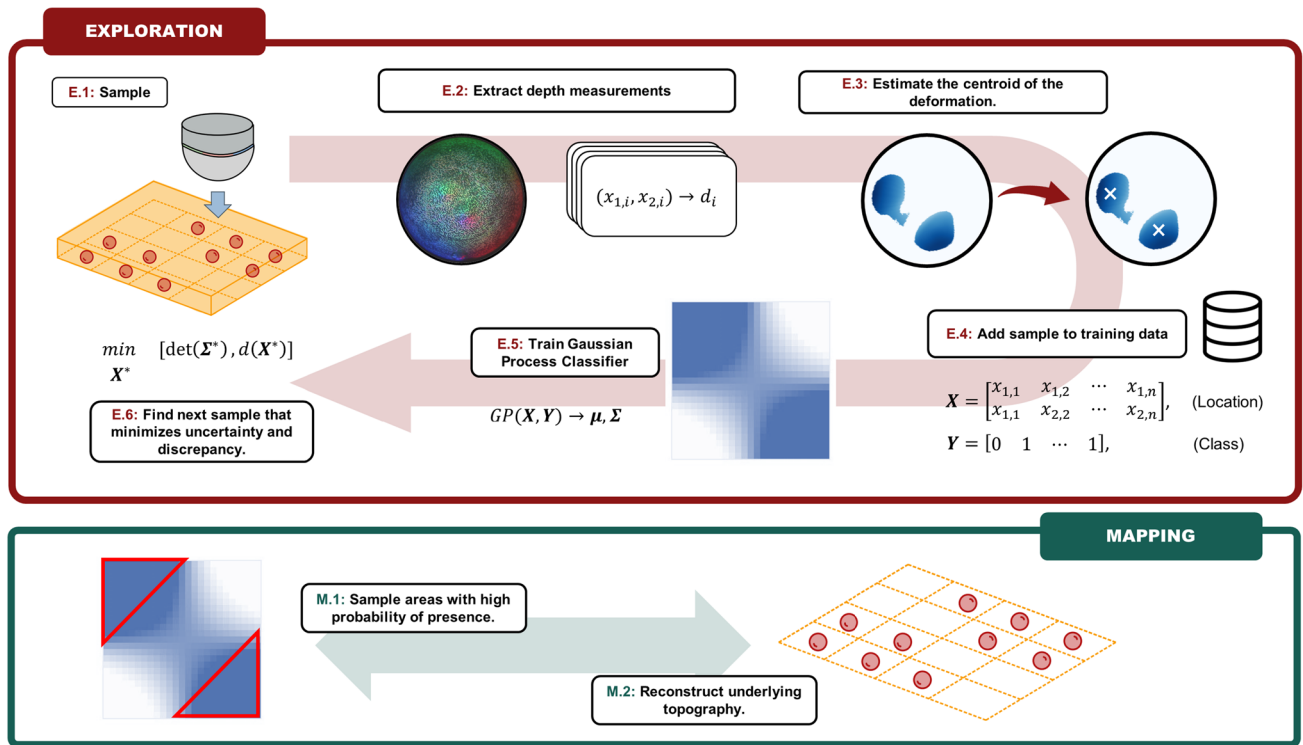


Fig. 2 Graphical representation of the proposed method. In the exploration phase (E.1 through E.6) the probable areas of hard embedded objects below the soft surface are estimated. In the mapping phase

(M.1 and M.2) a more thorough interaction of such areas is conducted to approximate the underlying topology

Fig. 3 Relation of the proposed method with the Sense-Plan-Act cycle of robotics

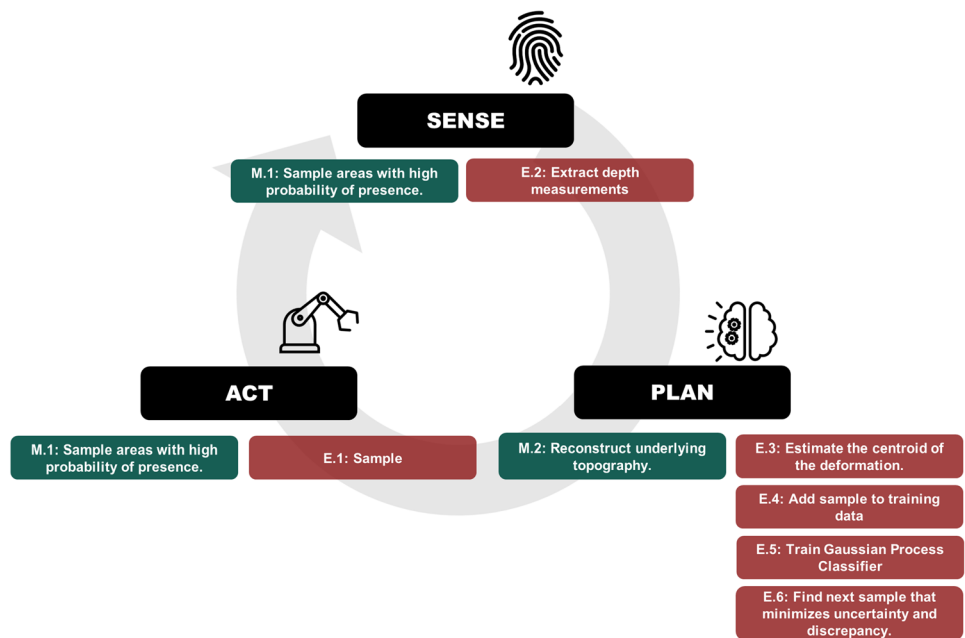


image is run through a Canny Edge Detector [20] to find the contours of the clusters observed in the depth image. This allows us to form the convex hull of each cluster defined by the vertex set $\{v_{i,1}, \dots, v_{i,n}\}$ and easily obtain

the centroid $\bar{c}_i = (x_i, y_i)$ of such set (1) which serves as an approximation of the centroid of the convex hull. In (1) x_i and y_i are the horizontal and vertical coordinates of the i -th centroid respectively.

$$\bar{c}_i = \frac{1}{n} \sum_{j=i}^n v_{i,j} \tag{1}$$

This entire process is performed using the libraries `cv.Canny`, `cv.findContours`, `cv.convexHull` and `cv.moments` from OpenCV [21].

If deformation is present on the depth image, the location of the centroid of each cluster is added to the training data with label 1; if no deformation is present on the depth image, the location of the sample (center of the depth image) is added to the training data with label 0 (E.4 in Fig. 2). We have opted for the strategy involving the inclusion of a single training data point per cluster, as opposed to incorporating each individual pixel exhibiting deformation. This choice is driven by the need to address scenarios involving disjoint embedded objects and to facilitate predictions over larger areas of presence. By doing so, we prevent the Gaussian Process from optimizing for local information and encourage it to capture a more holistic understanding.

To generate the probabilistic map of the location of the embedded objects, a Gaussian process for classification is used (E.5 in Fig. 2 and outlined in Algorithm 1). We have opted for the utilization of a Gaussian Process based on the premise that hard objects tend to form clusters and/or the sensor’s size is comparatively small in relation to them, mirroring scenarios outlined in the introductory applications of our manuscript. Given this assumption, the identification of a hard object, whether partially or entirely, has a significant influence on the likelihood of detecting additional objects in its proximity. This rationale justifies the adoption of a Gaussian Process in our approach.

A Gaussian process $\mathcal{GP}(\mu(\cdot), \kappa(\cdot, \cdot))$ is a prior over functions defined by a mean function $\mu(\cdot)$ and a kernel function $\kappa(\cdot, \cdot)$. Different from a Gaussian process for regression, where we want to predict a continuous function value, in a Gaussian process for binary classification we care about a discrete variable $y^* \in \{1, 0\}$ at location x^* given the observed classes \mathbf{Y} at locations \mathbf{X} using a predictive distribution $p(y^* = 1 | x^*, \mathbf{X}, \mathbf{Y})$ [22, 23].

To compute such distribution, a latent function $\mathbf{f}(x)$ is introduced over which we perform a Gaussian process regression and then its output is passed through a logistic function (2), i.e. $p(y = 1 | x) = \sigma(\mathbf{f}(x))$. Please keep in mind that the latent function \mathbf{f} serves as an intermediary variable within our method and does not represent a physically interpretable quantity.

$$\sigma(z) = [1 + \exp(-z)]^{-1} \tag{2}$$

The probabilistic prediction is then defined as:

$$p(y^* = 1 | x^*, \mathbf{X}, \mathbf{Y}) = \int \sigma(\mathbf{f}^*) p(\mathbf{f}^* | x^*, \mathbf{X}, \mathbf{Y}) d\mathbf{f}^* \tag{3}$$

This integral is analytically intractable and is solved by first approximating the logit function with a probit function $\Phi(f^*)$ and the second term by using a Laplace approximation [22]. With these approximations, the integral is Gaussian and can be computed, therefore:

$$p(\mathbf{f}^* | x^*, \mathbf{X}, \mathbf{Y}) \approx \mathcal{N}(\mathbf{f}^* | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{4a}$$

$$\boldsymbol{\mu} = \mathbf{k}_*^T (\mathbf{Y} - \sigma(\hat{\mathbf{f}})) \tag{4b}$$

$$\boldsymbol{\Sigma} = \mathbf{k}_{**} - \mathbf{k}_*^T (\mathbf{W}^{-1} + \mathbf{K}_f)^{-1} \mathbf{k}_* \tag{4c}$$

where $\mathbf{k}_* = \kappa(\mathbf{X}, \mathbf{X}^*)$, $\mathbf{k}_{**} = \kappa(\mathbf{X}^*, \mathbf{X}^*)$, \mathbf{W} is a diagonal matrix whose entries are $W_{ii} = \sigma(f_i)(1 - \sigma(f_i))$, $\mathbf{K}_f = \kappa(\mathbf{X}, \mathbf{X}) + \nu \mathbf{I}$ and ν is a small value that provides numerical stability (1×10^{-5} in our case). The mean $\hat{\mathbf{f}}$ can be obtained iteratively, i.e. $\hat{\mathbf{f}} \leftarrow \mathbf{f}^{new}$ using (5).

$$\mathbf{f}^{new} = \mathbf{K}_f (\mathbf{I} + \mathbf{W} \mathbf{K}_f)^{-1} (\mathbf{Y} - \sigma(\mathbf{f}) + \mathbf{W} \mathbf{f}) \tag{5}$$

The predictive distribution can be approximated as follows [23]:

$$p(y^* = 1 | x^*, \mathbf{X}, \mathbf{Y}) \approx \sigma(\boldsymbol{\mu}(1 + \pi \boldsymbol{\Sigma}/8)^{-1/2}) \tag{6}$$

We use an (isotropic) square exponential kernel (7) with parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$ which are optimized automatically by maximizing the log marginal likelihood $\log p(\mathbf{Y} | \boldsymbol{\theta})$. In the experimental validation of our method Sect. 4 the values of $\boldsymbol{\theta}$ converged to [30, 15].

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \theta_2^2 \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{2\theta_1^2}\right) \tag{7}$$

Algorithm 1 Gaussian process classifier (E.5 in Fig. 2)

Input: \mathbf{X} (Seen Locations), x^* (Unseen Location), \mathbf{Y} (Class)

Output: $p(y^* = 1 | \mathbf{X}^*, \mathbf{Y})$ (Presence probability)

- 1: Initialization: $\nu = 1 \times 10^{-5}$, $\mathbf{f} = \mathbf{0}$ Compute: \mathbf{K}_f
- 2: **for** j in $\text{max_iter} = 20$ **do**
- 3: Compute: \mathbf{W} , $\mathbf{f}^{new} = \text{Eqn. (5)}$
- 4: **if** $|\mathbf{f}^{new} - \mathbf{f}| < 1 \times 10^{-9}$ **then**
- 5: **break**
- 6: **end if**
- 7: $\mathbf{f} \leftarrow \mathbf{f}^{new}$
- 8: **end for**
- 9: $\hat{\mathbf{f}} \leftarrow \mathbf{f}^{new}$
- 10: Compute: \mathbf{k}_* , \mathbf{k}_{**} and \mathbf{W}
- 11: Compute: $\boldsymbol{\mu} = \text{Eqn. (4b)}$, $\boldsymbol{\Sigma} = \text{Eqn. (4c)}$
- 12: **return** $p(y^* = 1 | \mathbf{X}^*, \mathbf{Y}) = \text{Eqn. (6)}$

With every sample, we aim to reduce the uncertainty in our probabilistic map. Since a Gaussian process also

returns the confidence of the predicted label y , a naive approach would be to choose a sample x^* that has the highest uncertainty as measured by the covariance matrix Σ . However, in the tests conducted on real data, we found that this policy tends to oversample a region of the workspace before moving to others, which significantly delays the mapping phase of our algorithm. For this reason, we incorporate the discrepancy of the set of samples $\mathbf{X}^* = \{\mathbf{X}, x^*\}$ to decide the next sample. The discrepancy $d(\mathbf{X})$ is a criterion used to quantify the space-filling characteristics of a set of samples normalized over a unit hypercube and can be defined as follows [24]

$$d(\mathbf{X}) = \sup_{\mathcal{H}} \left| \frac{\#(\mathbf{X} \cap \mathcal{H})}{\#\mathbf{X}} - \lambda(\mathcal{H}) \right| \quad (8)$$

\mathcal{H} is the hyper-rectangular subset of the hypercube, $\#\mathbf{X}$ is the number of samples, $\#(\mathbf{X} \cap \mathcal{H})$ is the number of samples that lie in \mathcal{H} , and $\lambda(\mathcal{H})$ is the volume of \mathcal{H} . In general, the discrepancy is difficult to compute exactly and there exist approximate methods to do so. In our implementation, we compute the discrepancy of the samples using the Centered Discrepancy (CD) method [25]. With this, we can find the next sample using both the uncertainty of the Gaussian process and the discrepancy of the samples (E.6 in Fig. 2).

$$\text{minimize}_{\mathbf{X}^*} [\det \Sigma(\mathbf{X}^*), d(\mathbf{X}^*)] \quad (9)$$

The selection of the next sample could be benefited by the presence of a human in the loop, which could be valuable in medical applications or in prostheses as outlined in the introduction. The entire loop (steps E.1 through E.6) is repeated to choose the next sample.

Mapping

The mapping phase of the algorithm has a similar procedure to the exploration phase, however, the emphasis of this

phase is to reconstruct the shape of the embedded objects rather than forming the probabilistic map of their location. Because of this, we sample the areas with a high probability of presence according to the Gaussian process (M.1 in Fig. 2) following the same loop as in the exploration phase. The difference now is that we aim to sample areas with a high probability of presence rather than reduce the uncertainty in the workspace. The sampling strategy is then modified in step E.6 as:

$$\text{minimize}_{\mathbf{X}^*} [-p(y^* = 1 | \mathbf{X}^*, \mathbf{Y}), d(\mathbf{X}^*)] \quad (10)$$

With each additional sample, we continue training the Gaussian process as well as reconstructing the underlying topography of the workspace (M.2 in Fig. 2) which can then be visualized as a point cloud and further processing can be done if needed.

Experimental Setup

The experimental setup is presented in Fig. 4 where DenseTact 2.0 is attached to a desktop computer numerical control (CNC) machine for data collection. A perforated acrylic sheet with 5.5-millimeter holes in a 10×34 array with a center distance of 8.5 mms both vertically and horizontally, provides an easy to reconfigure test bed Fig. 4a. The perforated acrylic sheet allows us to locate quartz beads of 6 mms of diameter in the desired configurations for our experimental validation of the framework described in section 2. This array provides an effective area of 90.4×300 mms.

Once the beads have been placed in the desired locations, the array is covered with a sheet of polyethylene foam with a thickness of half an inch (12.7 mm). The foam is secured in place with two longitudinal acrylic strips that are clamped to the CNC bed as shown in Fig. 4b.

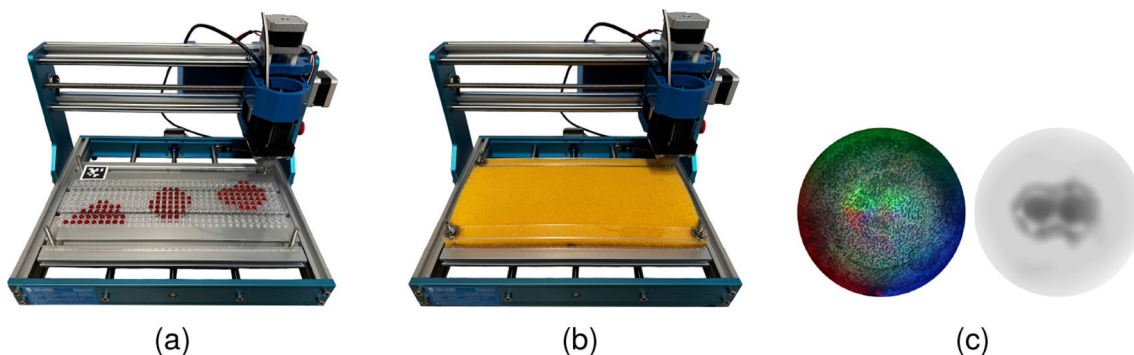


Fig. 4 Experimental configuration for the detection and mapping of embedded rigid objects in soft material matrices. **a** A perforated acrylic sheet allows to accommodate quartz bead clusters in any desired configuration. **b** The beads are covered with a layer of poly-

ethylene foam and secured in place with two longitudinal acrylic strips. **c** Side by side of the raw image from camera and the resulting depth image from DenseTact 2.0

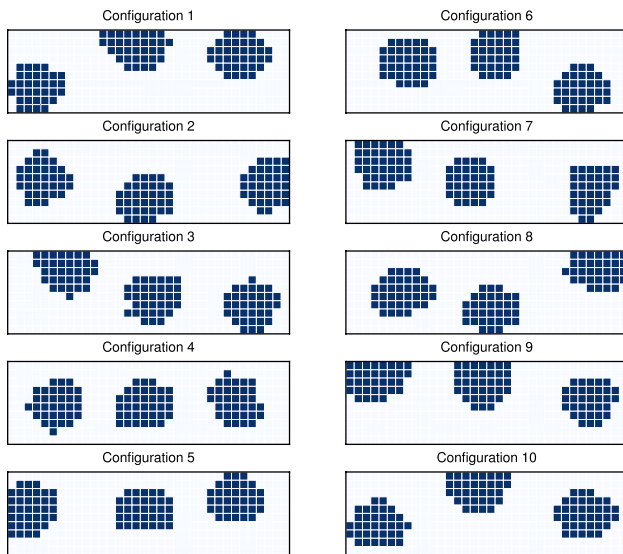


Fig. 5 Top-down view of the configurations of the quartz beads below the foam for the evaluation of the performance of the proposed method

For the evaluation of the performance of the framework, we evaluated ten different configurations of the beads on the acrylic sheet (Fig. 5). Due to the dimensions of the workspace, each configuration was comprised of three bead clusters, each cluster with thirty-four beads on average. All the data was collected at the same depth of 12 mms below the surface of the foam. The beads are dyed red to be identified with a computer vision program, that together with an AprilTag [26] allows us to easily reconstruct the configuration for evaluation of the performance of our framework.

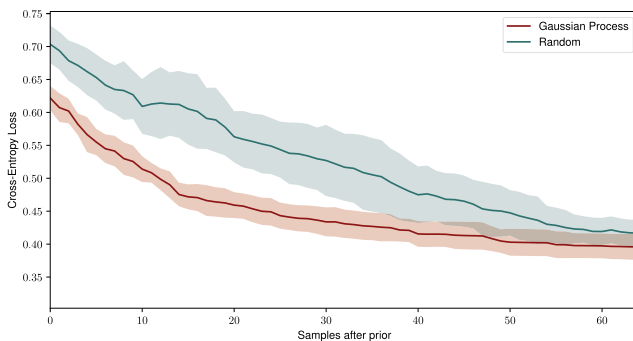
We are interested in reconstructing the underlying shape of the embedded objects. To accomplish this, it is necessary for the soft material to have relatively low

stiffness against the sensor. In our case, DenseTact 2.0 has a Shore 00 hardness of 49.7, while the polyethylene foam has a Shore 00 hardness of 24.7 when pressed seven millimeters. For this reason, we limited our experiments to a single type of soft material. However, it is important to highlight that with a stiffer material, it is still possible to apply the exploration phase of our method as long as the sensor is able to detect the presence of the embedded materials as in [10] and [11].

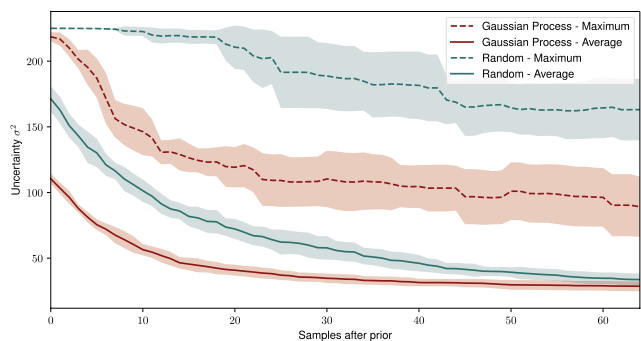
Results and Discussion

Figure 6 shows the performance of the exploration phase with up to sixty-four samples after the prior. The results are benchmarked against a fully random policy where the prior is also formed with eight randomly chosen samples. Regarding this random policy, we implemented a discretization of the workspace at 2.5 mm intervals in both vertical and horizontal directions. This discretized grid serves as the basis for randomly selecting samples from across the workspace. We only benchmark our approach against the aforementioned random policy, as the existing literature has not addressed the problem of embedded object localization and mapping. While no prior work has directly tackled this specific problem, the closest method we found is the work conducted by Jamali et al. [14]. This previous work relies on the continuity of the mapped object, which in our case such continuity is not necessary.

Because the main result of the exploration phase is the probabilistic map of the presence or absence of embedded objects (categorical variable), we evaluate the performance using the Cross-Entropy loss (11) with a density of 0.5 mms both horizontally and vertically:



(a)



(b)

Fig. 6 Performance of the exploration phase showing the average over the ten configurations shown in Fig. 5 with a solid/dashed lines and the shaded region indicating the corresponding standard deviation. **a** Shows the Cross-Entropy loss of our proposed method in red

and a fully random policy in green. **b** Shows the uncertainty, both maximum (dashed line) and average (solid line), for our proposed method and a fully random policy

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N [y_i \log p(y_i = 1 | \mathbf{X}, \mathbf{Y}) + (1 - y_i) \log p(y_i = 0 | \mathbf{X}, \mathbf{Y})] \tag{11}$$

As can be seen in Fig. 6a the proposed method has a better performance as compared with the fully random policy. Our proposed method has two distinct behaviors: before 14 to 16 samples after the prior, the loss decreases rapidly; after 14 to 16 samples after the prior, the loss continues decreasing but with a slower rate. As shown in Fig. 7, with sixteen samples after the prior, the probabilistic map has capture the main characteristic of the true distribution of the embedded objects. Different from our proposed method, the random policy more or less decreases at the same rate up to sixty additional samples after the prior. Eventually, with enough samples, both strategies reach a similar loss. However, it is important to highlight that our proposed method has a better consistency as reflected in a tighter standard deviation.

In terms of uncertainty, Fig. 6b, there is a similar behavior to the loss: before 14 to 16 samples there is a rapid decrease of uncertainty, both the maximum and average value, followed by a slower rate of decrease. As shown in the figure, both

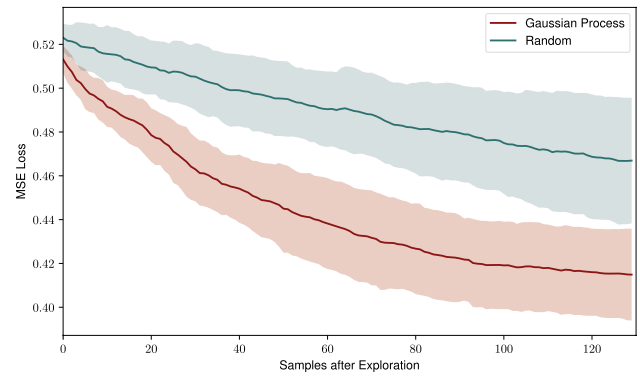


Fig. 8 Performance of the mapping phase showing the average over the ten configurations shown in Fig. 5 with a solid line and the shaded region indicating the corresponding standard deviation

maximum and average uncertainty are lower in our method compared to the random strategy.

Since we are interested in forming a quick idea of the location of the embedded objects to sample the areas with a high probability for the mapping phase, it is valuable that our method can capture the main features with approximately 14 to 16 samples, with a lower uncertainty than a fully random

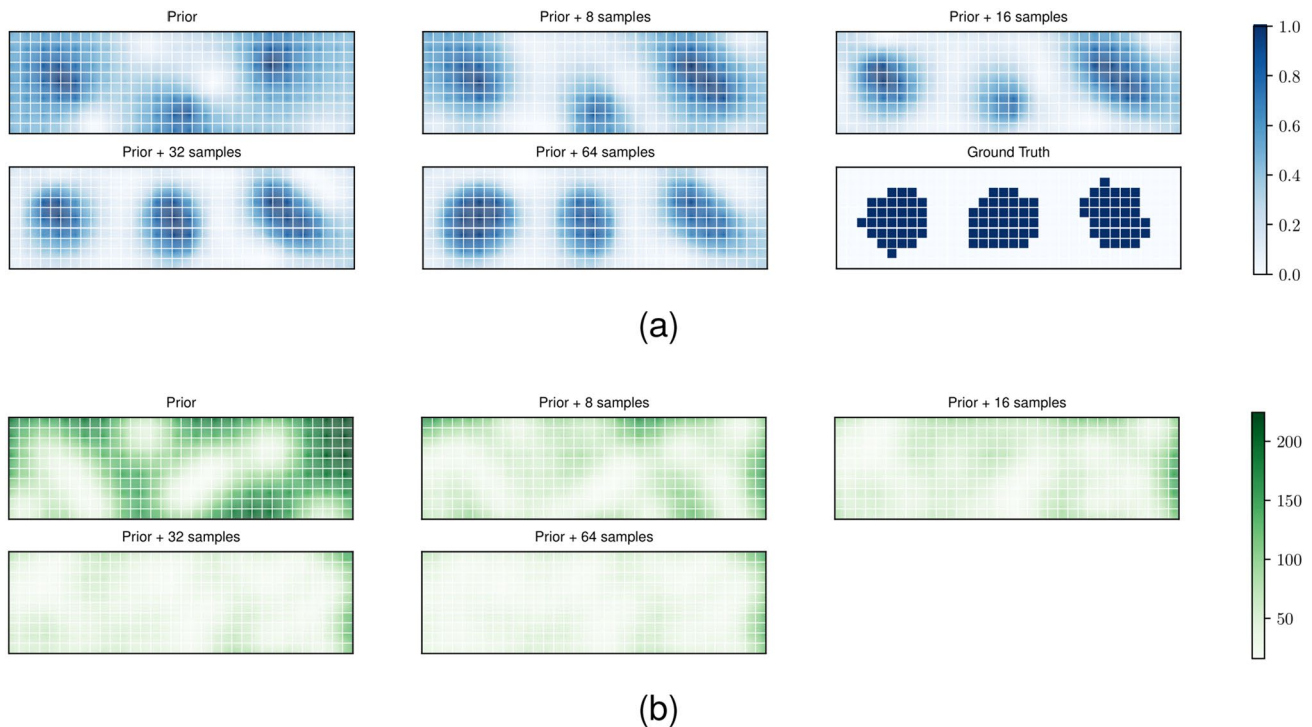


Fig. 7 Example of the progression of the exploration phase of the proposed method (E.4 in Fig. 2) as a top-down view of one configuration. **a** Shows the probabilistic map, given by equation (6), obtained after the prior and additional eight, sixteen, thirty-two, and sixty-four additional samples as well as the ground truth. Darker regions

denote a high probability of presence, while lighter regions denote a low probability of presence. **b** Shows the associated variance of the plots shown in (a) given by equation (4c). Darker regions denote high uncertainty, while lighter regions denote low uncertainty

sampling strategy. Given the observed trend in Fig. 6, we choose to have an exploration phase of 16 samples (or equivalently a Cross-Entropy loss of around 0.45 for the given size of the workspace) after the prior before proceeding to the mapping phase of our method.

For the mapping phase, the evaluation of the performance is accomplished by using the Mean Squared Error (MSE) loss (12) between the ground truth z and the height reported by the sensor \tilde{z} , again with a density of 0.5 mms both horizontally and vertically:

$$\mathcal{L}_{MSE} = \frac{1}{N} \sum_{i=1}^N (z_i - \tilde{z}_i)^2 \tag{12}$$

As seen in Fig. 8 the proposed method has a better performance as compared with the fully random policy: throughout the one hundred and twenty-eight samples the MSE error is consistently lower for our proposed method. This is expected since using a random policy it is possible that regions that do not have any embedded objects are sampled, especially in this case where the majority of the workspace does not have embedded objects. Our method in contrast does focus its sampling efforts on the areas that the exploration phase estimates are embedded objects. Finally, although not as noticeable as in the exploration phase, our method is more consistent than the fully random policy as depicted by the tighter standard deviation in Fig. 8. Figure 9 provides an example of the progression of the shape reconstruction of the embedded objects below the polyethylene foam.

We choose these loss functions as they are standard for binary classification (Cross-Entropy loss) and regression (MSE). Furthermore, the choice of logarithm in cross-entropy loss heavily penalizes misclassifications, better illustrating the need for an accurate probabilistic map before going to the mapping phase of our algorithm.

The performance of our method could be improved by incorporating local information about stress and force. Although DenseTact 2.0 provides a 6-axis wrench estimation over the entire sensor, when we incorporated this

information in our method, due to some inconsistency and high variance of the measurements (either from issues in the calibration or undesired interactions between the soft material and the sensor) our method had a tendency to substantially favor depth information and for this reason, we dispensed with force and moment information.

Conclusion

In this paper, we have proposed a method for the exploration and mapping of heterogeneous materials, where hard embedded objects are present within a matrix of a soft and deformable material, a situation in which sight cannot provide reliable or any kind of information and tactile sensing becomes the main mechanism of interaction. Our method has been motivated in applications such as package sorting, medical diagnosis, and restoration of haptic sensation in individuals with prostheses.

The proposed framework is divided into two main phases. In the exploration phase, where a probabilistic map is generated by sampling the workspace and using a Gaussian process for classification we estimate the probability of the presence or absence of the hard embedded objects. The mapping phase exploits the probabilistic map generated by the exploration phase by sampling the areas of high probability of presence to reconstruct the underlying shape of the hard objects. We also present a strategy to obtain a prior of the location of the embedded objects using Sobol sequences when such a prior is not available.

We validate our approach using an experimental setup that located a series of quartz beads in clusters underneath a polyethylene foam, and by using the optical tactile sensor DenseTact 2.0 together with the aid of a computer numerical control (CNC) machine, we sample the workspace.

Our empirical results show that our method outperforms a fully random policy in both the exploration and mapping phases. In the exploration phase, our method presents two distinct behaviors, before 14 to 16 samples after the prior

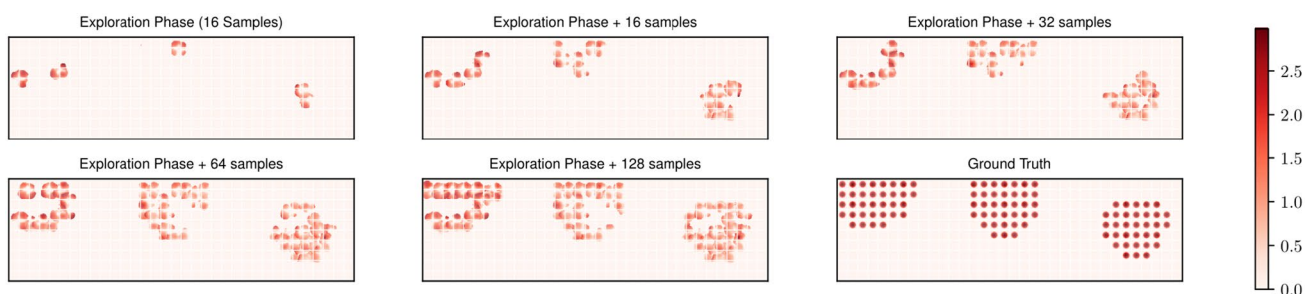


Fig. 9 Example of the progression of the mapping phase of the proposed method ($M.2$ in Fig. 2) as a top-down view of one configuration. The image shows the approximate topology obtained after the

exploration phase and additional sixteen, thirty-two, sixty-four, and one hundred and twenty-eight additional samples as well as the ground truth. The scale on the right is given in millimeters

there is a rapid decrease in both uncertainty and Cross-Entropy loss, followed by a continuous decrease of both the uncertainty and loss with a slower rate. It is valuable to see the rapid decrease in both loss and uncertainty with few samples to be able to transition to the mapping phase and focus the attention of the samples on the shape reconstruction aspect of the method. For the given size of the workspace, 16 samples (or equivalently a Cross-Entropy loss of around 0.45) were chosen to be the number of samples before switching to the mapping phase where we also demonstrate that our method outperforms the fully random policy. On both the exploration and mapping phases, our proposed method presents a better consistency as compared to the random policy, shown by the smaller standard deviation across the ten different bead configurations.

Immediate extensions include the post-processing of the reconstructed shape depending on the application: for item recognition in packages and home assistance, the reconstructed shape can be passed through a convolutional neural network to identify the different objects; for medical applications, the reconstructed shape can be displayed to the physician using virtual reality for assessment; and for haptic sensation, the reconstructed shape can be reproduced on another part of the prosthesis owner's body. Future work could include the sampling and reconstruction of three-dimensional and/or curved surfaces and incorporate stress and force measurements into the method to obtain additional information on the embedded objects.

Funding Open Access funding provided by the MIT Libraries. This study was funded by the National Science Foundation (grant 2142773).

Data availability The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Li R, Platt R, Yuan W, Pas A, Roscup N, Srinivasan MA, Adelson E. Localization and manipulation of small parts using GelSight tactile sensing. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014; pp. 3988–3993. <https://doi.org/10.1109/IROS.2014.6943123>.
- Alspach A, Hashimoto K, Kuppuswamy N, Tedrake R. Soft-bubble: A highly compliant dense geometry tactile sensor for robot manipulation. In: 2019 2nd IEEE International Conference on Soft Robotics (RoboSoft), 2019; pp. 597–604. <https://doi.org/10.1109/ROBOSOFT.2019.8722713>.
- Do WK, Jurewicz B, Kennedy M. DenseTact 2.0: Optical Tactile Sensor for Shape and Force Reconstruction. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023; pp. 12549–12555. <https://doi.org/10.1109/ICRA48891.2023.10161150>.
- Do WK, Aumann B, Chungyoung C, Kennedy M. Inter-finger small object manipulation with DenseTact optical tactile sensor. *IEEE Robot Autom Lett.* 2024;9(1):515–22. <https://doi.org/10.1109/LRA.2023.3333735>.
- Do WK, Dhawan AK, Kitzmann M, Kennedy III M. DenseTact-Mini: An optical tactile sensor for grasping multi-scale objects from flat surfaces. 2023. arXiv preprint [arXiv:2309.08860](https://arxiv.org/abs/2309.08860).
- Do WK, Kennedy M. DenseTact: Optical Tactile Sensor for Dense Shape Reconstruction. In: 2022 International Conference on Robotics and Automation (ICRA), 2022; pp. 6188–6194. <https://doi.org/10.1109/ICRA46639.2022.9811966>.
- Yuan W, Srinivasan MA, Adelson EH. Estimating object hardness with a GelSight touch sensor. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016; pp. 208–215. <https://doi.org/10.1109/IROS.2016.7759057>.
- Yuan W, Zhu C, Owens A, Srinivasan MA, Adelson EH. Shape-independent hardness estimation using deep learning and a GelSight tactile sensor. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017; pp. 951–958. <https://doi.org/10.1109/ICRA.2017.7989116>.
- Yuan W, Mo Y, Wang S, Adelson EH. Active Clothing Material Perception Using Tactile Sensing and Deep Learning. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018; pp. 4842–4849. <https://doi.org/10.1109/ICRA.2018.8461164>.
- Gwilliam JC, Pezzementi Z, Jantho E, Okamura AM, Hsiao S. Human vs. robotic tactile sensing: Detecting lumps in soft tissue. In: 2010 IEEE Haptics Symposium, 2010; pp. 21–28. <https://doi.org/10.1109/HAPTIC.2010.5444685>.
- Jia X, Li R, Srinivasan MA, Adelson EH. Lump detection with a gelsight sensor. In: 2013 World Haptics Conference (WHC), 2013; pp. 175–179. <https://doi.org/10.1109/WHC.2013.6548404>.
- Dragiev S, Toussaint M, Gienger M. Gaussian process implicit surfaces for shape estimation and grasping. In: 2011 IEEE International Conference on Robotics and Automation, 2011; pp. 2845–2850. <https://doi.org/10.1109/ICRA.2011.5980395>.
- Martins R, Ferreira JF, Dias J. Touch attention Bayesian models for robotic active haptic exploration of heterogeneous surfaces. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014; pp. 1208–1215. <https://doi.org/10.1109/IROS.2014.6942711>.

14. Jamali N, Ciliberto C, Rosasco L, Natale L. Active perception: Building objects' models using tactile exploration. In: 2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids), 2016; pp. 179–185. <https://doi.org/10.1109/HUMANOIDS.2016.7803275>.
15. Yi Z, Calandra R, Veiga F, Hoof H, Hermans T, Zhang Y, Peters J. Active tactile object exploration with Gaussian processes. In: 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2016; pp. 4925–4930. <https://doi.org/10.1109/IROS.2016.7759723>.
16. Wang S, Wu J, Sun X, Yuan W, Freeman WT, Tenenbaum JB, Adelson EH. 3D Shape Perception from Monocular Vision, Touch, and Shape Priors. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018; pp. 1606–1613. <https://doi.org/10.1109/IROS.2018.8593430>.
17. Liu SQ, Yañez LZ, Adelson EH. GelSight EndoFlex: A Soft Endoskeleton Hand with Continuous High-Resolution Tactile Sensing. In: 2023 IEEE International Conference on Soft Robotics (RoboSoft), 2023; pp. 1–6. <https://doi.org/10.1109/RoboSoft55895.2023.10122053>.
18. Zhao J, Bauza M, Adelson EH. FingerSLAM: Closed-loop Unknown Object Localization and Reconstruction from Visuo-tactile Feedback. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), 2023; pp. 8033–8039. <https://doi.org/10.1109/ICRA48891.2023.10161489>.
19. Sobol' IM. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Mathe Math Phys.* 1967;7(4):86–112. [https://doi.org/10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9).
20. Canny J. A Computational Approach to Edge Detection. *IEEE Trans Patt Anal Mach Intell PAMI.* 1986;8(6):679–98. <https://doi.org/10.1109/TPAMI.1986.4767851>.
21. Bradski G. The OpenCV library. *Dr. Dobb's J Softw Tools.* 2000.
22. Rasmussen CE, Williams CKI. Gaussian processes for machine learning. Adaptive computation and machine learning. Cambridge: MIT Press; 2006. p. 2–248.
23. Bishop CM. Pattern recognition and machine learning (information science and statistics). Berlin, Heidelberg: Springer; 2006.
24. Kochenderfer MJ, Wheeler TA. Algorithms for optimization. Cambridge: The MIT Press; 2019.
25. Zhou YD, Fang KT, Ning JH. Mixture discrepancy for quasi-random point sets. *J Complex.* 2013;29(3–4):283–301. <https://doi.org/10.1016/J.JCO.2012.11.006>.
26. Olson E. AprilTag: A robust and flexible visual fiducial system. In: 2011 IEEE International Conference on Robotics and Automation, 2011; pp. 3400–3407. <https://doi.org/10.1109/ICRA.2011.5979561>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.