**IEEE Robotics and Automation Letters (RAL) paper presented at the**
**2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**
**October 25-29, 2020, Las Vegas, NV, USA (Virtual)**

# Exploiting Visual-outer Shape for Tactile-inner Shape Estimation of Objects Covered with Soft Materials

Tomoya Miyamoto[1*], Hikaru Sasaki[1*], and Takamitsu Matsubara[1]

***Abstract*— In this paper, we consider the problem of *inner-shape* estimation of objects covered with soft materials, e.g., pastries wrapped in paper or vinyl, water bottles covered with shock-absorbing fabrics, or human bodies dressed in clothes. Due to the softness of the covered materials, tactile information obtained through physical touches can be useful to estimate such inner shape; however, using only tactile information is inefficient since it can collect local information at around the touchpoint. Another approach would be taking visual information obtained by cameras into account; however, it is not straightforward since the visual information only captures the *outer shape* of the covered materials, and it is unknown how much such visual-outer shape is similar/dissimilar to the tactile-inner shape. We propose an active tactile exploration framework that can utilize the visual-outer shape to efficiently estimate the inner shape of objects covered with soft materials. To this end, we propose the Gaussian Process Inner-Outer Implicit Surface model (GPIOIS) that jointly models the implicit surfaces of inner-outer shapes with their similarity by Gaussian processes. Simulation and real-robot experimental results demonstrated the effectiveness of our method.**

## I. Introduction

In our living environment, there are many objects covered with soft materials. Examples include pastries wrapped in paper or vinyl, water bottles covered with shock-absorbing fabrics, and human bodies dressed in clothes. When a robot grasps and manipulates such an object, it needs the shape information of the inner object, not the outer shape covered with a soft material.

Tactile information can be useful to estimate such inner shape assuming the outer material is soft enough, and the robot can contact the inner object. In particular, tactile exploration for object shape estimation of unknown objects (not covered with soft materials) has been studied intensively [1]–[12]. However, this approach is essentially inefficient since only the local information of the touched part is obtained from tactile sensors.

To improve the efficiency of tactile exploration, exploiting visual information obtained through cameras has been explored since visual information can be in a wide range without physical touch. For example, both visual and tactile information is simply used in a mixed way [13], [14], or utilizing uncertainty of shapes constructed from visual information for active tactile shape estimation [15], [16]. However, it is not straightforward for objects covered with soft materials, since the visual information captures the *outer shape* of the covered materials, and it is unknown how much such visual-outer shape is similar/dissimilar to the tactile-inner shape.

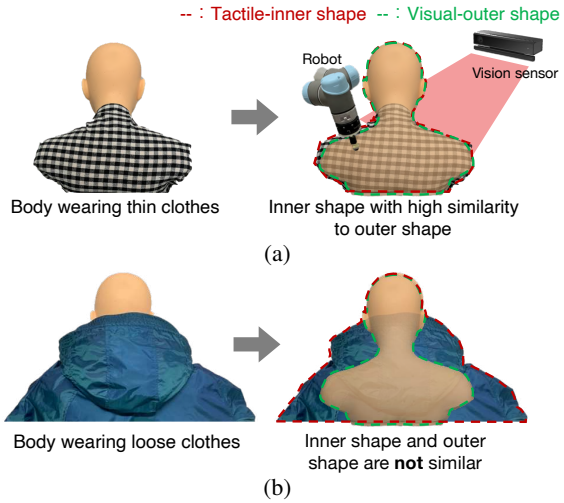



Fig. 1: Mannequin wearing clothing as an example of objects covered with soft materials. (a) The mannequin fits a cloth tightly, the visual-outer shape is similar to tactile-inner shape. (b) The mannequin fits a cloth loosely, the visual-outer shape is dissimilar to the tactile-inner shape.

Fig. 1 shows a mannequin wearing clothing as an example. If the mannequin fits a cloth tightly, we can use visual information directly in combination with the tactile information for inner shape estimation, since the visual-outer shape is similar to tactile-inner shape. If the mannequin fits a cloth loosely, it may not be beneficial for inner shape estimation since the visual-outer shape is dissimilar to the tactile-inner shape. In other words, for effectively utilizing the visual information in tactile-inner shape estimation of objects covered with soft materials, it is of primal importance to evaluate the similarity between the visual-outer shape and the tactile-inner shape.

With the above in mind, in this paper, we propose an active tactile exploration framework by extending an active tactile exploration strategy proposed in [7]. Our framework can utilize the outer shape estimated from visual information to efficiently estimate the inner shape of objects covered with soft materials. To this end, we propose the Gaussian Process Inner-Outer Implicit Surface model (GPIOIS) that jointly models the implicit surfaces of inner-outer shapes with their similarity by Gaussian processes. The GPIOIS can be interpreted as a novel combination of Gaussian process

[1]T. Miyamoto, H. Sasaki, and T. Matsubara are with the Division of Information Science, Graduate School of Science and Technology, Nara Institute of Science and Technology (NAIST), Japan.
*Both authors contributed equally
This research was supported by MEXT/JSPS KAKENHI Grant Number 19H01124

implicit surface (GPIS) [17] with Gaussian process-based multi-task learning [18], [19]. One of the objectives of multi-task learning [20], [21] is, in general, to estimate the similarities among multiple tasks from observations. Our method considers the estimation problems of visual-outer and tactile-inner shapes as different tasks. Then the similarity estimation problem between these tasks is tackled by a multi-task learning scheme, and the estimated similarity is utilized for exploiting visual-outer shape for the active tactile-inner shape estimation of the covered objects.

## II. Related Work

Accurate shape representation of 3D objects is crucial in many situations. The implicit surface [22] has been widely used for expressing shape information because it is easy to obtain such geometric information as normal vectors and curvature from it. A practical extension of the implicit surface is the Gaussian process implicit surface (GPIS) [17], which can express the uncertainty of the shape estimation with the implicit surface using a Gaussian process [23]. With the uncertainty of shape estimation, active tactile exploration can be formulated to actively collect useful information on unknown object shapes by intelligent touching behaviors [1]–[6]. Martens et al. proposed geometric object priors and introduced them into GPIS by modifying the mean function to encode shape primitives in the prior distribution for performance improvement [24]. Sommer et al. proposed a strategy of bimanual compliant tactile exploration to guide two arms and an object for identification [25]. By extending work by Matsubara et al. that minimizes shape uncertainty and travel cost to the next touchpoint [6], Ottenhaus et al. presented a strategy for object shape estimation and the evaluation of many 3D object models in simulations [26]. Yang presented an alternative strategy in which the next touchpoint is selected by the uncertainty reduction and the local curvature for fidelity enhancement [27]. Driess et al. explored another interesting attempt that incorporated sliding motions rather than discrete touch behaviors in the tactile shape estimation of objects [7], [12].

Integrating visual and tactile information for object shape estimation [13]–[16] as well as object recognition [28] has been studied. However, these studies commonly assumed that the visual and tactile shapes are very similar to the actual object shape. Therefore, both visual and tactile information can be processed together in the same methodologies. Such methods cannot be directly applied to objects covered with soft materials like those we tackled in this paper since visual and tactile shapes can be different.

In addition to shape estimation, some studies proposed data-driven exploration using Gaussian processes. For example, Saal et al. proposed the GP-tactile-based active estimation of object dynamics parameters with tactile sensors. Tanaka et al. extended it a GP-based tactile exploration strategy, which focuses more on object identification [29]. Poon et al. extended a framework [30] for object search in clutter. Kaboli et al. proposed a tactile-based GP approach to execute the robot to discriminate objects in an unknown workspace, object characteristic identification and object recognition [31].

## III. Tactile Exploration with Sliding Motions

This section briefly summarizes the framework for sliding-based exploration from [7] as preparation for our framework in the next section.

### A. Gaussian Process Implicit Surface

The implicit surface is a method that represents shape $\mathcal{S}$ of an object in $d$ dimensions space from point cloud-based data using implicit function $F$: $\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^d \mid F(\mathbf{x}) = 0\}$. GPIS estimates an object's shape by regressing implicit function $F$ by GP using explored data $\mathscr{D} = \{(\mathbf{x}_i, h_i)\}_{i=1}^N$ where $N$ is data size. $\mathbf{x}_i$ is indicating a position. $h_i$ is a surface label indicating the position of $\mathbf{x}_i$ with respect to the object. $h_i$ takes 0 or 1, respectively indicating the outer and surface of an object.

GP infers a predictive distribution of implicit function $F$ using exploration data $\mathscr{D}$ as a Gaussian distribution:

$$p(F(\mathbf{x}) \mid \mathbf{x}, \mathscr{D}) = \mathcal{N}(F(\mathbf{x}) \mid \mu(\mathbf{x}), V(\mathbf{x})), \tag{1}$$

$$\mu(\mathbf{x}) = m + \mathbf{k}_*^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{y} - m\mathbf{1}), \tag{2}$$

$$V(\mathbf{x}) = k_{**} - \mathbf{k}_*^T(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\mathbf{k}_*, \tag{3}$$

where $\mathbf{k}_*$ is a kernel vector as $[\mathbf{k}_*]_i = \mathrm{k}_x(\mathbf{x}, \mathbf{x}_i)$, $\mathbf{K}$ is a kernel gram matrix as $[\mathbf{K}]_{ij} = \mathrm{k}_x(\mathbf{x}_i, \mathbf{x}_j)$, kernel value $k_{**} = \mathrm{k}_x(\mathbf{x}, \mathbf{x})$, $\mathbf{y} = [h_1, \ldots, h_N]^T$, $\mathrm{k}_x(\cdot, \cdot)$ is a kernel function, and $\mathbf{I}$ is an identity matrix. $m$ and $\sigma^2$ are the mean and variance of a prior distribution. We set $m = 1$ by assuming that almost all the space is outside of the object [2], [7], [12]. Variance function $V$ indicates the uncertainty of the inference. The value of variance function $V(\mathbf{x})$ increases in areas where data is scarce. GPIS estimates shape $\hat{\mathcal{S}}$ using mean function $\mu$:

$$\hat{\mathcal{S}} = \{\mathbf{x} \in \mathbb{R}^d \mid \mu(\mathbf{x}) = 0\}. \tag{4}$$

In this paper, we employ an inverse-multiquadric kernel that was previously suggested [7] as a suitable choice for implicit surfaces:

$$\mathrm{k}_x(\mathbf{x}_i, \mathbf{x}_j) = (\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 + \theta_x^2)^{-1/2}, \tag{5}$$

where $\theta_x$ is a parameter of the kernel function.

Although the original GPIS [17] contains three potentials, 0, -1, 1, by assuming that the objects are closed, the robot can acquire only two pieces of information, outside and on the surface. Thus, -1 is omitted.

### B. Sliding-Motion Planning for Exploration with GPIS

A sliding motion is defined as moving a robot in the direction that maximizes the uncertainty of estimated shape $\hat{\mathcal{S}}$ on the tangent plane. The direction is found by solving the following maximization:

$$\mathbf{d}^* = \arg\max_{\mathbf{d}} \nabla L(\mathbf{x})^T\mathbf{d}, \tag{6}$$

$$\text{s.t.} \quad \mathbf{n}_\mu(\mathbf{x})^T\mathbf{d} = 0, \quad \|\mathbf{d}\|_2 = \alpha, \tag{7}$$

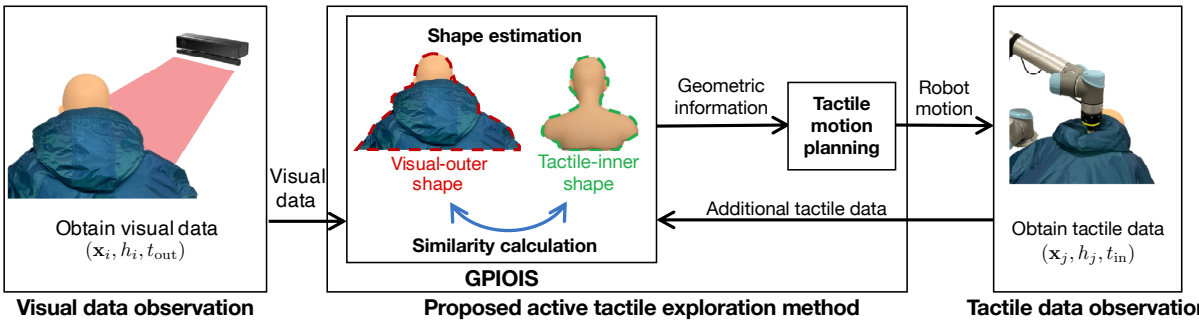



Fig. 2: Overview of our proposed framework for active tactile exploration of an object's inner shape covered with soft material. It is composed of a Gaussian process inner-outer implicit surface model (GPIOIS) and sliding-motion planning for exploration with GPIOIS. Inner shape estimation and tactile motion planning (and execution) are alternatively executed until inner shape estimation converges.

where $\mathbf{x}$ is the current robot position, $\mathbf{n}_\mu(\mathbf{x})$ indicates the normal vector of the estimated shape, $\alpha$ is the moving distance in one step, $L(\mathbf{x}) = V(\mathbf{x}) + C(\mathbf{x})$ is the objective function that is a sum of variance function $V$ and penalty function $C$. $C(\mathbf{x}) < 0$ indicates the areas that a robot cannot reach. Optimal direction $\mathbf{d}^*$ is found by solving the maximization with a Lagrange multiplier:

$$\mathbf{d}^* = \alpha \frac{\mathbf{P}_\mu(\mathbf{x}) \nabla L(\mathbf{x})}{\|\mathbf{P}_\mu(\mathbf{x}) \nabla L(\mathbf{x})\|_2}, \tag{8}$$

where $\mathbf{P}_\mu(\mathbf{x})$ is the tangent plane projector of estimated shape on current robot position $\mathbf{x}$. The robot moves to next position $\mathbf{x}' = \mathbf{x} + \mathbf{d}^*$.

## IV. Proposed Method

Figure 2 overviews our proposed framework, which is composed of a) a Gaussian process inner-outer implicit surface model (GPIOIS) and b) sliding-motion planning for exploration with GPIOIS.

### A. Gaussian Process Inner-Outer Implicit Surface

Here we introduce our Gaussian process inner-outer implicit surface (GPIOIS) model, which jointly models the object's inner and outer shapes by GPIS and can be interpreted as a special case of a multi-task Gaussian process [18], [19] combined with GPIS.

Our GPIOIS learns two implicit functions $F_{\text{in}}$ for the inner shape of an object covered by soft materials and $F_{\text{out}}$ for its outer shape jointly using tactile and visual-based data $\mathscr{D}_{\text{IO}} = \{(\mathbf{x}_i, h_i, t_i)\}_{i=1}^N$. $h_i$ is a surface label indicating the position $\mathbf{x}_i$ with respect to the object. $h_i$ takes 0 or 1, respectively indicating the outer and surface of an object. $t_i \in \{t_{\text{in}}, t_{\text{out}}\}$ is an inner-outer (IO) label. We assign inner label $t_{\text{in}}$ to tactile-based data and outer label $t_{\text{out}}$ to visual-based data without specifying their values; they are treated as a latent variable.

GPIOIS employs kernel function $\mathrm{k}_{\text{IO}}(\cdot,\cdot)$ to treat data label $t_i$ as input:

$$\mathrm{k}_{\text{IO}}((\mathbf{x}_i, t_i), (\mathbf{x}_j, t_j)) = \mathrm{k}_x(\mathbf{x}_i, \mathbf{x}_j)\ \mathrm{k}_t(t_i, t_j), \tag{9}$$

where $\mathrm{k}_t(\cdot,\cdot)$ is a kernel function that calculates the similarity between labels. It is vital to note that the similarity is defined in between the estimated shapes associated with the labels. Since the data label has only two patterns, kernel $\mathrm{k}_t(\cdot,\cdot)$ is defined using matrix $\mathbf{K}_t$:

$$\mathbf{K}_t = \begin{bmatrix} \mathrm{k}_t(t_{\text{in}}, t_{\text{in}}) & \mathrm{k}_t(t_{\text{in}}, t_{\text{out}}) \\ \mathrm{k}_t(t_{\text{out}}, t_{\text{in}}) & \mathrm{k}_t(t_{\text{out}}, t_{\text{out}}) \end{bmatrix}. \tag{10}$$

$\mathbf{K}_t$ must be a positive semi-definite matrix since $\mathrm{k}_t(\cdot,\cdot)$ is a kernel function. To meet this requirement, by decomposing $\mathbf{K}_t$ into $\mathbf{K}_t = \mathbf{L}\mathbf{L}^T$ by lower triangular matrix $\mathbf{L}$ using Cholesky decomposition, kernel $\mathrm{k}_t(\cdot,\cdot)$ that indicates the similarity can be parametrized by $\theta_t = \mathbf{L}$.

Note that we employ an approach for learning the similarity in a free form [18], [19]. We don't specifically design either the values of $t$ or the type of kernel $\mathrm{k}_t(\cdot,\cdot)$; instead we directly optimize $\mathbf{K}_t$ through $\theta_t$. See [18], [19] for more details.

GPIOIS infers the predictive distribution of implicit function $F_{\text{in}}$ for tactile-inner shape using data $\mathscr{D}_{\text{IO}}$ as a Gaussian distribution:

$$p(F_{\text{in}}(\mathbf{x}) \mid \mathbf{x}, t_{\text{in}}, \mathscr{D}_{\text{IO}}) = \mathcal{N}(F_{\text{in}}(\mathbf{x}) \mid \mu_{\text{in}}(\mathbf{x}, t_{\text{in}}), V_{\text{in}}(\mathbf{x}, t_{\text{in}})), \tag{11}$$

$$\mu_{\text{in}}(\mathbf{x}, t_{\text{in}}) = m + \mathbf{k}_{\text{IO}}^T (\mathbf{K}_{\text{IO}} + \sigma^2 \mathbf{I})^{-1} (\mathbf{y}_{\text{IO}} - m\mathbf{1}), \tag{12}$$

$$V_{\text{in}}(\mathbf{x}, t_{\text{in}}) = k_{\text{IO}} - \mathbf{k}_{\text{IO}}^T (\mathbf{K}_{\text{IO}} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_{\text{IO}}, \tag{13}$$

where kernel vector $\mathbf{k}_{\text{IO}}$ as $[\mathbf{k}_{\text{IO}}]_i = \mathrm{k}_{\text{IO}}((\mathbf{x}_i, t_i), (\mathbf{x}, t_{\text{in}}))$, kernel gram matrix $\mathbf{K}_{\text{IO}}$ as $[\mathbf{K}_{\text{IO}}]_{ij} = \mathrm{k}_{\text{IO}}((\mathbf{x}_i, t_i), (\mathbf{x}_j, t_j))$, and kernel value $k_{\text{IO}} = k_{\text{IO}}((\mathbf{x}, t_{\text{in}}), (\mathbf{x}, t_{\text{in}}))$, $\mathbf{y}_{\text{IO}} = [h_1, \ldots, h_N]^T$.

By using GPIOIS, tactile-inner shape $\hat{\mathcal{S}}_{\text{in}}$ is estimated using mean function $\mu_{\text{in}}$:

$$\hat{\mathcal{S}}_{\text{in}} = \{\mathbf{x} \in \mathbb{R}^d \mid \mu_{\text{in}}(\mathbf{x}, t_{\text{in}}) = 0\}. \tag{14}$$

Parameter $\theta_t$ is optimized by a standard marginal likelihood maximization approach [18], [19]:

$$\theta_t^* = \arg\max_{\theta_t} \log p(\mathbf{y}_{\text{IO}} \mid \mathscr{D}_{\text{IO}}, \theta_t), \tag{15}$$

$$\log p(\mathbf{y}_{\text{IO}} \mid \mathscr{D}_{\text{IO}}, \theta_t) = -\frac{1}{2} \log |\mathbf{K}_{\text{IO}} + \sigma^2 \mathbf{I}| - \frac{1}{2} (\mathbf{y}_{\text{IO}} - m\mathbf{1})^T (\mathbf{K}_{\text{IO}} + \sigma^2 \mathbf{I})^{-1} (\mathbf{y}_{\text{IO}} - m\mathbf{1}) + C. \tag{16}$$

The other kernel parameters are fixed in the exploration like the previous method [7].

### B. Sliding-Motion Planning for Exploration with GPIOIS

Inspired by the strategy with GPIS explained in Section III our proposed framework plans a sliding motion by moving the robot in the direction that maximizes the uncertainty of tactile-inner shape estimation $\hat{\mathcal{S}}_{\text{in}}$ on the tangent plane. The direction is found by solving the following maximization:

$$\mathbf{d}_{\text{in}}^{*} = \arg\max_{\mathbf{d}_{\text{in}}} \nabla L_{\text{in}}(\mathbf{x}, t_{\text{in}})^{T} \mathbf{d}_{\text{in}}, \tag{17}$$

$$\text{s.t.} \quad \mathbf{n}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})^{T} \mathbf{d}_{\text{in}} = 0, \quad \|\mathbf{d}_{\text{in}}\|_2 = \alpha, \tag{18}$$

where $\mathbf{n}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})$ indicates the normal vector of the estimated tactile-inner shape and $L_{\text{in}}(\mathbf{x}, t_{\text{in}}) = V_{\text{in}}(\mathbf{x}, t_{\text{in}}) + C(\mathbf{x})$ is the objective function. Optimal direction $\mathbf{d}_{\text{in}}^{*}$ is found by solving the maximization with a Lagrange multiplier:

$$\mathbf{d}_{\text{in}}^{*} = \alpha \frac{\mathbf{P}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}}) \nabla L_{\text{in}}(\mathbf{x}, t_{\text{in}})}{\|\mathbf{P}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}}) \nabla L_{\text{in}}(\mathbf{x}, t_{\text{in}})\|_2}, \tag{19}$$

where $\mathbf{P}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})$ is a tangent plane projector of estimated shape $\hat{\mathcal{S}}$ on current robot position $\mathbf{x}$. The robot moves to next position $\mathbf{x}' = \mathbf{x} + \mathbf{d}_{\text{in}}^{*}$.

Normal vector $\mathbf{n}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})$ and tangent plane projector $\mathbf{P}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})$ of estimated tactile-inner shape $\hat{\mathcal{S}}_{\text{in}}$ are calculated using mean function $\mu_{\text{in}}(\mathbf{x}, t_{\text{in}})$:

$$\mathbf{g}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}}) = \frac{\partial \mu_{\text{in}}(\mathbf{x}, t_{\text{in}})}{\partial \mathbf{x}}^{T}, \tag{20}$$

$$\mathbf{n}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}}) = \frac{\mathbf{g}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})}{\|\mathbf{g}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})\|_2}, \tag{21}$$

$$\mathbf{P}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}}) = \mathbf{I} - \mathbf{n}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}}) \mathbf{n}_{\mu_{\text{in}}}(\mathbf{x}, t_{\text{in}})^{T}. \tag{22}$$

The computational complexity for the optimal direction in one step is $O(N^3)$, which is the same as the exploration method proposed in the previous method [7].

## V. Simulation

To investigate our method's effectiveness, we conducted simulations with four objects with different inner and outer shapes.

### A. Experimental Settings

*1) Environment:* The environment of each object in the simulation is shown in Fig. 3. We assume that the robot is a point mass with a tactile sensor and can move in any direction in 2D space. Therefore, we set $C(\mathbf{x}) = 0, \forall x$.

We conducted 2D simulations with four shapes. Regarding the relationship between the internal and external shapes, we first considered three cases: high similarity (object A), low similarity (object D), and local mixing (object C). We prepared the shapes of these three patterns by simple rectangles in the 2D simulation. We made object B's shape more complicated to investigate the applicability to non-convex shapes. Based on these considerations, we designed the four shapes for the simulations. Each shape was validated only once since there are no effects of noise and randomness in this environment; no statistical evaluations are required in this case.

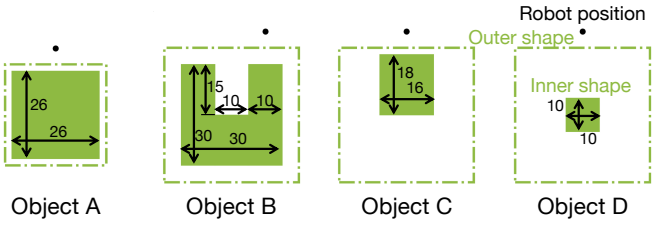



Fig. 3: Four objects with different inner and outer shapes. Filled green shapes indicate inner shape and dash-dotted lines indicate outer shape of objects. Black dots are starting points of exploration of each object. Unit of measurement is centimeters.

TABLE I: Hausdorff distance between inner and outer shapes

| | Object A | Object B | Object C | Object D |
|---|---|---|---|---|
| Hausdorff (cm) | 6.4 | 12.0 | 26.9 | 37.5 |

*2) Exploration Setting:* We compared our visual-tactile exploration method with a previous active tactile exploration method [7]. Our method obtains visual information once before exploration. We initially gave each method a direction to an object. The exploration starts when the robot touches the object the first time and terminates when it goes around the inner shape. The exploration parameters were set as follows: $\alpha = 1.0$ cm, $\sigma^2 = 0.006$, and, $\theta_x = 0.1$.

*3) Evaluation Method:* We evaluated each exploration method by the following three criteria.

**Uncertainty measure:** We used the variance of the predictive distribution on estimated shape $\hat{\mathcal{S}}$ as the uncertainty of the estimated shape by following Dragiev et al. [3]:

$$\frac{1}{N_{\hat{\mathcal{S}}}} \sum_{\mathbf{x} \in \hat{\mathcal{S}}} V(\mathbf{x}), \tag{23}$$

where $N_{\hat{\mathcal{S}}}$ is the number of points on the surface of estimated shape $\hat{\mathcal{S}}$. The uncertainty measure is normalized to make it comparable among objects of different shapes and sizes.

**Hausdorff distance:** We employed the Hausdorff distance [32] as a measure of the distance between shapes $\mathcal{S}_a$ and $\mathcal{S}_b$:

$$\mathrm{d}_H(\mathcal{S}_a, \mathcal{S}_b) = \max \left[ \max_{\mathbf{x} \in S_a} \mathrm{d}(\mathbf{x}, \mathcal{S}_b), \max_{\mathbf{x} \in \mathcal{S}_b} \mathrm{d}(\mathbf{x}, \mathcal{S}_a) \right], \tag{24}$$

where $\mathrm{d}(\mathbf{x}, \mathcal{S}) = \min_{\mathbf{s} \in \mathcal{S}} ||\mathbf{x} - \mathbf{s}||_2$ is a minimum distance between points $\mathbf{x}$ and $\mathbf{s}$ on the surface.

**Similarity:** We defined the similarity between our method's visual-inner shape and tactile-outer shape as follows: $\mathrm{k}_t(t_{\text{in}}, t_{\text{out}}) / \mathrm{k}_t(t_{\text{in}}, t_{\text{in}})$.

### B. Result

Figure 4 shows the result of the our visual-tactile-based exploration, the previous tactile-based exploration, the transition of the uncertainty measure, and the Hausdorff distance. Fig. 5 summarizes the estimated similarities with the four objects in our method.

In the our method for object A, which has similar inner and outer shapes, the estimated shape closely resembled the actual shape even from the beginning of the exploration. Both the low uncertainty measure and the short Hausdorff distance were obtained continuously. Therefore, an efficient

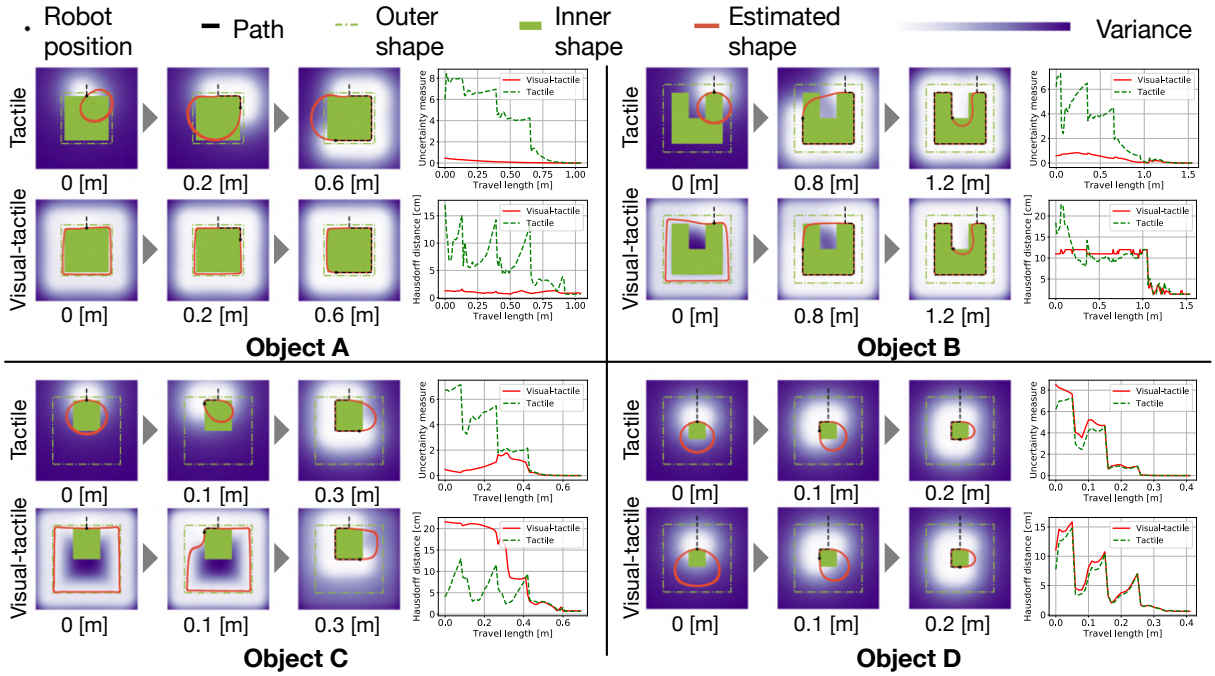



Fig. 4: Exploration results of our visual-tactile-based exploration method and previous tactile-based exploration method for objects A to D. Black dashed lines indicate robot's exploration path. Red lines indicate estimated shape. Purple indicates uncertainty of shape estimation in each position. Right figures are transitions of uncertainty measure and Hausdorff distance between estimated and true shapes.

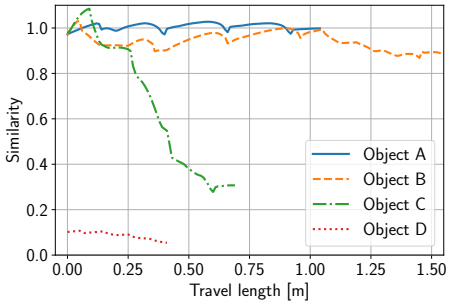



Fig. 5: Transitions of similarity between visual-outer shape and tactile-inner shape of our visual-tactile exploration method in exploration for objects A to D.

inner-shape estimation was performed by estimating and exploiting the similarity between the visual-outer and tactile-inner shapes. For object B, our method obtained low uncertainty and short Hausdorff distance with less exploration than tactile exploration until it reached the hole. Both exploration methods obtained almost the same transition of the uncertainty measure and the Hausdorff distance in the exploration of the hole. Regarding object C, the estimated shape by our method was affected by the outer shape in the early stage of the exploration. However, in the later stage, the estimated similarity decreased quickly and resembled the tactile exploration behaviors without exploiting the visual-outer shape. Finally, for object D during the exploration, the estimated shapes by our method and the tactile exploration method were very similar, and the uncertainty measure and area error were almost the same. Due to the low similarity between the inner and outer shapes, our method did not exploit the visual-outer shape for inner shape estimation.

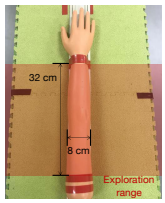
In summary, all the simulation results suggest that our method can estimate and exploit the similarity between the visual-outer and tactile-inner shapes for the efficient inner shape estimation of the objects covered by soft materials. Even with low similarity, since the visual-outer shape is not exploited by the inner shape estimation, there are almost no adverse effects on estimation accuracy.

## VI. Actual Robot Experiment

Finally, we investigate the effectiveness of our method in an actual robot experiment. The role of 2D simulations in the previous section validated the effectiveness of our visual-tactile-based method compared with tactile-based exploration. Since the same path can be followed for both cases, we can focus on validating the mechanism's shape estimation performance with similarity inference between the visual-outer and tactile-inner shapes and its effect on the inner-shape estimation in detail. Here, we validate our method's effectiveness with 3D objects in real robot experiments where our visual-tactile-based method and the tactile-based method comparison follow different paths in exploration.

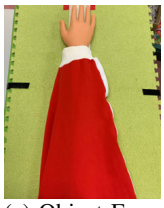
### A. Experimental Settings

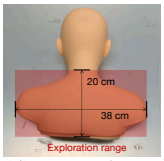
*1) Environment:* We used arm and body mannequins covered by different clothes (Fig. 6) for an exploration experiment by an actual robot. Fig. 7 shows its environment. We used Universal Robots' UR5 with six degrees of freedom as an actual robot, Kinect v2 as a visual sensor, and a 3-axis OptoForce OMD-20-SE-40N as a tactile sensor. To restrict



(a) Inner shape (true)


(b) Object E (fitted)


(c) Object F (loosely fitted)

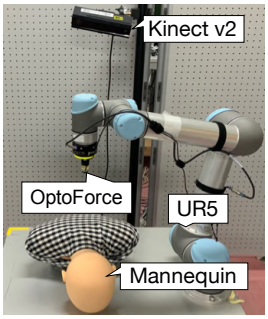


(d) Inner shape (true)

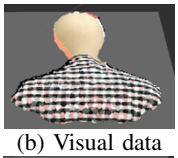

(e) Object G (fitted)


(f) Object H (loosely fitted)

Fig. 6: Objects for exploration experiment by actual robot



(a) Sensor and robot


(b) Visual data

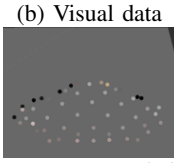

(c) Downsampled visual data

Fig. 7: Environment for actual robot experiment

the exploration range, we designed penalty function $C(\mathbf{x})$:

$$C(\mathbf{x}) = \begin{cases} -\beta \ \mathrm{dist}(\mathbf{x})^2 & \text{(out of range)} \\ 0 & \text{(otherwise)}, \end{cases} \tag{25}$$

where $\beta = 100$, $\mathrm{dist}(\mathbf{x})$ is the distance between robot position $\mathbf{x}$ and the exploration range's boundary shown in Figs. 6a and 6d. To improve the calculation efficiency, we downsampled vision-based point cloud by averaging points divided in a 5 cm voxel (Fig. 7c). We evaluated each method using the uncertainty measure, the Hausdorff distance, and the similarity described in previous section.

*2) Exploration Setting:* Similar to the simulation, we compared our method with a previously proposal [7]. Our method obtains visual information once before exploration. We initially gave each method a direction to an object. The exploration starts when the robot touches the object for the first time and terminates when the travel length reaches 1.5 m. The exploration parameters are set as $\alpha = 1.0$ cm, $\sigma^2 = 0.08$, and, $\theta_x = 0.15$.

*3) Robot Control Framework:* The robot movements can be divided into the tangential direction and the normal direction to the object. Since we placed a 3-axis OptoForce OMD-20-SE-40N in the robot's fingertip in this experiment, the square root of the sum of the squares of these three forces is set on the robot fingertip as the contact force, and contact is detected if the magnitude of the contact force exceeds 1.0 N. Because the robot searches for unknown objects, unexpected force may be exerted on it in tangential and normal movement directions. To prevent this, when the contact force exceeds 5.0 N, it moves in a normal direction opposite the object until the contact force becomes 5.0 N or less. These constraints allow the robot to search safely.

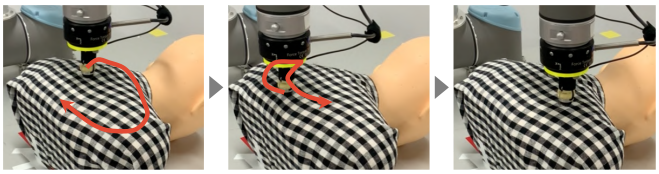


Fig. 8: Exploration by our method with object G

### B. Result

Figure 8 shows the exploration by the actual robot using our method with object G. Fig. 9 shows the result of our visual-tactile-based exploration and the previous tactile-based exploration in actual robot environments, with the results of the uncertainty measure and the Hausdorff distance. For objects E and G that wear thin clothes, the Hausdorff distance indicates that our method had high estimation accuracy even from the beginning of the exploration. Since both the uncertainty measure and the Hausdorff distance were continuously low, we did an efficient inner-shape estimation by estimating and exploiting the similarity between the visual-outer and tactile-inner shapes. For object F that wears loose clothes but the arm's shape is almost recognizable, the estimated shape was initially affected by the outer shape, and the estimated shape shrank to the accurate inner shape as the exploration progressed. For object H that wears loose clothes, our method and the previous method had similar estimated shapes, and the uncertainty measure and the Hausdorff distance were almost the same. Thus, our method did not exploit the visual-outer shapes since inner and outer shapes were dissimilar. Visual-tactile based and tactile based exploration methods took $20.38 \pm 3.14$ min and $16.75 \pm 2.38$ min for 1.5 m exploration, including 150 times' motion planning and execution. Each motion planning took $8.15 \pm 1.26$ s and $6.70 \pm 0.95$ s, respectively.

Figure 10 shows the transitions of the estimated similarity for the four objects. Objects E and G obtained continuously high similarity during the exploration. Object F obtained slightly less similarity as the exploration progressed, although it remained relatively high. The transition of object H's similarity is lower than the other objects. Our method can estimate the reasonable similarity between inner and outer shapes.

Figure 11 shows the initial exploration direction on each object with our method and the previous tactile exploration method. By using visual information to a tactile exploration, the initial exploration distances were changed.

In summary, in actual robot experiments, our method achieved efficient exploration based on the estimated similarity between visual-outer and tactile-inner shapes. Even with low similarity, since the visual-outer shape is not exploited in the inner shape estimation, we found almost no adverse effects on the estimation accuracy.

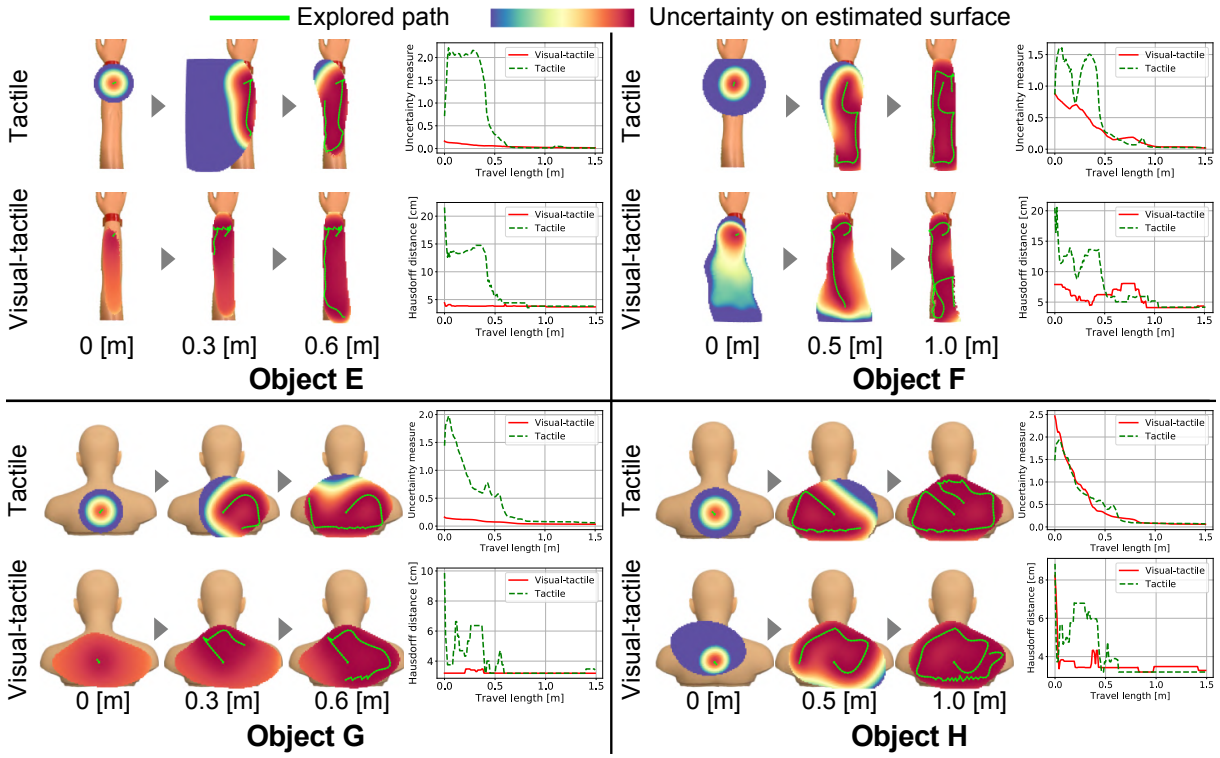



Fig. 9: Exploration results of our visual-tactile-based exploration method and previous tactile-based exploration method for objects E to G in actual robot setting. Green lines indicate exploration path of actual robot. Color maps on mannequin indicate estimated shape and uncertainty on its surface. Right figures are transitions of uncertainty measure and Hausdorff distance between estimated and true shapes.

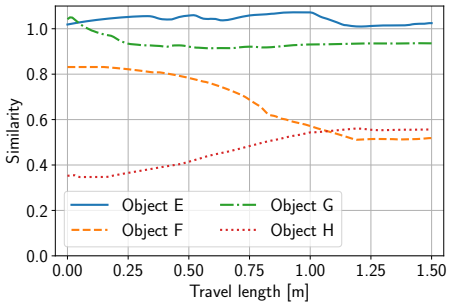



Fig. 10: Transitions of similarity between visual-outer and tactile-inner shapes of our visual-tactile exploration method in exploration for objects E to H.

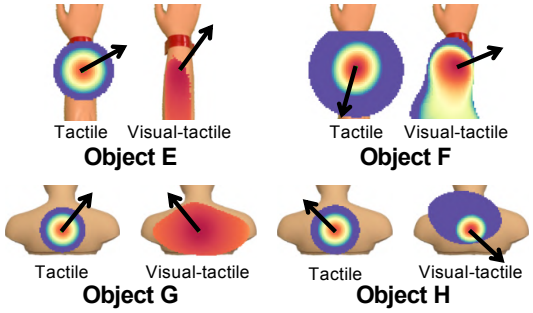



Fig. 11: Initial exploration direction on objects E to H with tactile and visual-tactile information

## VII. Discussion

Our primary motivation for this study is partner robots that talk, watch, and care for people in such private spaces as homes and care facilities. To improve the affinity between humans and robots, touch care is useful because it provides not only visual and auditory stimulation through dialogue but also tactile stimulation at the same time. In such situations, people will probably touch the outside of their clothing more often than they will directly touch another person's body.

Many other cases exist where our approach becomes important, including when pastries are wrapped in paper or plastic and water bottles are covered with shock-absorbing fabrics. Generally, if the inner object is fragile or deformable and if the outer surface is loose, handling with tremendous force/torque is dangerous. Another future direction of our work is investigating how to adjust the compliance of the robot control with the looseness/tightness of the outer surface.

Our current framework has several limitations. First, our framework implicitly assumes that just one object is covered with soft materials due to utilized tactile motion planning for maintaining contact with the object that is currently being touched. Second, it assumes that the internal object is not moving over the shape estimation process. Future work will manage this limitation by introducing another latent variable in the model to represent such moving position and orientation.

Another limitation of our framework is its estimation of the global similarity between visual-outer and tactile-inner shapes. Thus, our method may fail to capture the local similarities of shapes. Our framework can be extended with

a local similarity measure that would extend latent task label $t_{\mathrm{in/out}}$ from a scalar to a set where each element is assigned to a local area of the target space. This framework could be implemented by expanding the size of task matrix $\mathbf{K}_t$ in (10) so that each element of the gram matrix captures the local similarities between the inner and outer shapes.

Intriguing future work will extend our framework so that it can explore not only the inner shape of rigid objects covered with soft materials but also such other characteristics of deformable-inner objects as stiffness, inertia, surface texture, center of mass, and shape [31].

## VIII. Conclusion

In this paper, we considered the problem of *inner-shape* estimation of objects covered with soft materials. We proposed an active tactile exploration framework that can utilize the visual-outer shape to efficiently estimate the inner shape of objects covered with soft materials. We confirmed through actual robot experiments that the proposed framework achieved an efficient exploration based on the estimated similarity between visual-outer and tactile-inner shapes. Even with low similarity, since the visual-outer shape is not exploited into the inner shape estimation, there were almost no adverse effects on the estimation accuracy.

## References


[1] A. Bierbaum, M. Rambow, T. Asfour, and R. Dillmann, "A potential field approach to dexterous tactile exploration of unknown objects," in *2008 IEEE-RAS 8th International Conference on Humanoid Robots (Humanoids)*, 2008, pp. 360–366.

[2] S. Dragiev, M. Toussaint, and M. Gienger, "Gaussian process implicit surfaces for shape estimation and grasping," in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 2845–2850.

[3] S. Dragiev, M. Toussaint, and M. Gienger, "Uncertainty aware grasping and tactile exploration," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*, 2013, pp. 113–119.

[4] N. Jamali, C. Ciliberto, L. Rosasco, and L. Natale, "Active perception: Building objects' models using tactile exploration," in *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, 2016, pp. 179–185.

[5] Z. Yi, R. Calandra, F. Veiga, H. van Hoof, T. Hermans, Y. Zhang, and J. Peters, "Active tactile object exploration with gaussian processes," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 4925–4930.

[6] T. Matsubara and K. Shibata, "Active tactile exploration with uncertainty and travel cost for fast shape estimation of unknown objects," *Robotics and Autonomous Systems*, vol. 91, pp. 314–326, 2017.

[7] D. Driess, P. Englert, and M. Toussaint, "Active learning with query paths for tactile object shape exploration," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 65–72.

[8] K. Yao, M. Kaboli, and G. Cheng, "Tactile-based object center of mass exploration and discrimination," in *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, 2017, pp. 876–881.

[9] M. Kaboli, D. Feng, K. Yao, P. Lanillos, and G. Cheng, "A tactile-based framework for active object learning and discrimination using multimodal robotic skin," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2143–2150, 2017.

[10] M. Kaboli, D. Feng, and G. Cheng, "Active tactile transfer learning for object discrimination in an unstructured environment using multimodal robotic skin," *International Journal of Humanoid Robotics*, vol. 15, no. 01, p. 1850001, 2017.

[11] D. Feng, M. Kaboli, and G. Cheng, "Active prior tactile knowledge transfer for learning tactual properties of new objects," *Sensors*, vol. 18, no. 2, p. 634, 2018.

[12] D. Driess, D. Hennes, and M. Toussaint, "Active multi-contact continuous tactile exploration with gaussian process differential entropy," in *2019 IEEE International Conference on Robotics and Automation (ICRA)*, 2019, pp. 7844–7850.

[13] J. Ilonen, J. Bohg, and V. Kyrki, "Fusing visual and tactile sensing for 3-d object reconstruction while grasping," in *2013 IEEE International Conference on Robotics and Automation (ICRA)*, May 2013, pp. 3547–3554.

[14] J. Bimbo, L. D. Seneviratne, K. Althoefer, and H. Liu, "Combining touch and vision for the estimation of an object's pose during manipulation," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 4021–4026.

[15] M. Björkman, Y. Bekiroglu, V. Högman, and D. Kragic, "Enhancing visual perception of shape through tactile glances," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2013, pp. 3180–3186.

[16] S. Wang, J. Wu, X. Sun, W. Yuan, W. T. Freeman, J. B. Tenenbaum, and E. H. Adelson, "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.

[17] O. Williams and A. Fitzgibbon, "Gaussian process implicit surfaces," in *Gaussian Processes In Practice*, 2006.

[18] E. V. Bonilla, F. V. Agakov, and C. K. I. Williams, "Kernel multi-task learning using task-specific features," in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2007, pp. 43–50.

[19] E. V. Bonilla, K. M. Chai, and C. Williams, "Multi-task gaussian process prediction," in *Advances in Neural Information Processing Systems 20 (NIPS)*, 2008, pp. 153–160.

[20] H. Kashima, T. Kato, Y. Yamanishi, M. Sugiyama, and K. Tsuda, "Link propagation: A fast semi-supervised learning algorithm for link prediction," in *Proceedings of the 2009 SIAM international conference on data mining (SDM)*, 2009, pp. 1100–1111.

[21] K. M. Chai, *Multi-task learning with gaussian processes*. The University of Edinburgh, 2010.

[22] A. Hilton, A. Stoddart, J. Illingworth, and T. Windeatt, "Implicit surface-based geometric fusion," *Comput. Vis. Image Underst.*, vol. 69, no. 3, pp. 273–291, 1998.

[23] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[24] W. Martens, Y. Poffet, P. R. Soria, R. Fitch, and S. Sukkarieh, "Geometric priors for gaussian process implicit surfaces," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 373–380, 2016.

[25] N. Sommer, M. Li, and A. Billard, "Bimanual compliant tactile exploration for grasping unknown objects," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 6400–6407.

[26] S. Ottenhaus, L. Kaul, N. Vahrenkamp, and T. Asfour, "Active tactile exploration based on cost-aware information gain maximization," *International Journal of Humanoid Robotics*, vol. 15, no. 01, p. 1850015, 2018.

[27] S. Yang, "Enhanced learning strategies for tactile shape estimation and grasp planning of unknown objects," 2019.

[28] P. Falco, S. Lu, C. Natale, S. Pirozzi, and D. Lee, "A transfer learning approach to cross-modal object recognition: From visual observation to robotic haptic exploration," *IEEE Transactions on Robotics*, vol. 35, no. 4, pp. 987–998, 2019.

[29] D. Tanaka, T. Matsubara, K. Ichien, and K. Sugimoto, "Object manifold learning with action features for active tactile object recognition," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014, pp. 608–614.

[30] J. Poon, Y. Cui, J. Ooga, A. Ogawa, and T. Matsubara, "Probabilistic active filtering for object search in clutter," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 7256–7261.

[31] M. Kaboli, K. Yao, D. Feng, and G. Cheng, "Tactile-based active object discrimination and target object search in an unknown workspace," *Autonomous Robots*, vol. 43, no. 1, pp. 123–152, 2019.

[32] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.