

# Probabilistic Surface Friction Estimation Based on Visual and Haptic Measurements

Tran Nguyen Le , Francesco Verdoja , Fares J. Abu-Dakka , and Ville Kyrki 

**Abstract**—Accurately modeling local surface properties of objects is crucial to many robotic applications, from grasping to material recognition. Surface properties like friction are however difficult to estimate, as visual observation of the object does not convey enough information over these properties. In contrast, haptic exploration is time consuming as it only provides information relevant to the explored parts of the object. In this letter, we propose a joint visuo-haptic object model that enables the estimation of surface friction coefficient over an entire object by exploiting the correlation of visual and haptic information, together with a limited haptic exploration by a robotic arm. We demonstrate the validity of the proposed method by showing its ability to estimate varying friction coefficients on a range of real multi-material objects. Furthermore, we illustrate how the estimated friction coefficients can improve grasping success rate by guiding a grasp planner toward high friction areas.

**Index Terms**—Perception for grasping and manipulation, probabilistic inference, sensor fusion.

## I. INTRODUCTION

**N**OWADAYS, robots are used extensively to perform various tasks from simple pick-and-place to sophisticated object manipulation in complex environments from factory floors to hospitals. For such tasks, robots are required to interact with, and adapt to, unknown environments and objects. In order to successfully accomplish these tasks, robots need to identify various properties of the objects to be handled. For these reasons, identifying object models that can represent the properties of objects has become a crucial issue in robotics.

Many object-modelling approaches have focused on object shape and geometry by utilizing vision [1]–[3]. However, other physical properties also play an important role in characterizing object behavior during interaction and handling. In particular, surface properties such as surface friction, texture, and roughness are vital for manipulation planning.

Various methods have been proposed to learn object surface properties from vision [4]–[6] or haptic feedback [7]–[11]. Also the combination of both vision and haptic cues [12], [13] has

Manuscript received October 15, 2020; accepted February 12, 2021. Date of publication February 26, 2021; date of current version March 17, 2021. This letter was recommended for publication by Associate Editor F. Meier and Editor M. Vincze upon evaluation of the reviewers' comments. This work was supported in part by the Academy of Finland Strategic Research Council under Grant 314180 and in part by CHIST-ERA Project IPALM (326304). (Corresponding author: Tran Nguyen Le.)

The authors are with Intelligent Robotics Group, Department of Electrical Engineering and Automation, School of Electrical Engineering, Aalto University, 02150 Espoo, Finland (e-mail: tran.nguyenle@aalto.fi; francesco.verdoja@aalto.fi; fares.abudakka@gmail.com; ville.kyrki@aalto.fi).

Digital Object Identifier 10.1109/LRA.2021.3062585

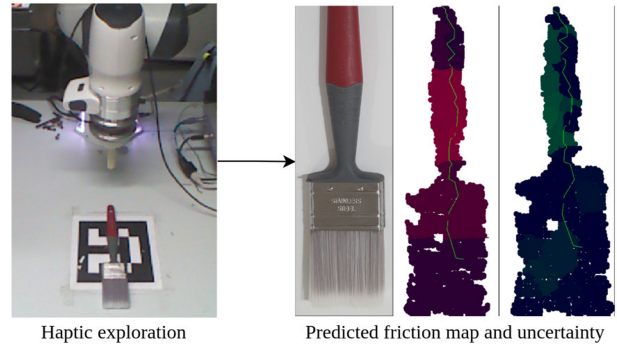


Fig. 1. After the shape of an object has been captured from a camera, surface friction on a small part of the object is estimated using haptic exploration. The friction over the entire object is then predicted by coupling visual information with the local haptic measurements.

been proposed, similar to human perception [14]. However, most published works assume the surface properties to be identical across the whole surface of the object. This assumption does not hold for many real objects, since objects often consist of multiple materials.

To address this, we propose a method to estimate the surface properties of a multi-material object by combining a visual prior with haptic feedback obtained while performing haptic exploration on the object. We focus on one property, the surface friction coefficient, but the proposed method could be applied to other properties such as texture and roughness. The approach is based on the assumption that visual similarity implies similarity of surface properties. By measuring a property directly using haptic exploration over a small part of the object, the joint distribution of visual and haptic features can be constructed. Using the joint distribution, the measurement can then be generalized over all parts of the object that are visible. The inference allows recovering both the expected value of friction for each part as well as a respective measure of prediction confidence.

The main contributions of this letter are:

- a probabilistic method to estimate the object friction coefficient based on visual prior and haptic feedback without being restricted by the assumption that objects have uniform homogeneous friction everywhere;
- a set of experiments on a physical robot showing the proposed method working on a wide range of objects, including multi-material objects;
- a case study, demonstrating the ability of the estimated friction coefficients to guide grasp planning towards areas of high friction, improving grasping success rate.

## II. RELATED WORK

### A. Friction Estimation From Vision

In the context of friction estimation from vision, most recent improvements come from the adoption of deep learning. Majority of the works focuses on recognizing the material using images then assigning the friction coefficient based on a material dataset. For instance, Zhang *et al.* [4] presented a material recognition method of deep reflectance code that encodes information about the physical properties of a surface from reflectance measurements. The predicted friction coefficient of a sample is then assigned as the average of friction coefficients of corresponding samples in the dataset. Another approach for material recognition was proposed by Xue *et al.* [5], where they developed a deep neural network architecture using angular variation features. Brandão *et al.* [6] proposed a solution to this problem by combining a state-of-the-art deep Convolutional Neural Network (CNN) architecture to predict broad material classes from images with known distributions of material friction. The predicted friction coefficient is then used to plan the locomotion of biped robot on different terrains.

One downside of the aforementioned approaches is that wrong material recognition will lead to wrong friction coefficient estimation. As the mentioned works utilize different visual features to recognize the materials, the recognition result depends heavily on the quality of visual input. However, vision is usually impaired by occlusions, lighting condition, and location of the sensor. Furthermore, even assuming the visual input to be perfectly gathered, these approaches may also fail in cases such as

- 1) different visual features having the same friction coefficients, or
- 2) same visual features having different friction coefficients.

In this letter, we overcome these limitations by combining the visual input with haptic feedback to directly estimate friction coefficients of an object. The addition of haptic feedback offers the possibility of accessing information that is hardly perceptible visually. In addition, the aforementioned works only attempt to estimate surface properties for outdoor scenes and applied to mobile robotics domain while in this work we target household objects.

### B. Friction Estimation From Haptic Feedback

The idea of using exploratory actions to estimate physical properties of objects has been carried out in many works [15]. One of the earliest work on friction estimation from haptic feedback was proposed by Yoshikawa *et al.* [7], where the authors described a method to estimate the friction distribution and the centre of friction of an object by pushing it with a mobile manipulator. Similar works on estimating friction coefficient were carried out with different exploration actions such as pushing [16], pressing [17], or lateral sliding [8]–[11]. However, all these works are only valid under the assumption that the surface properties are identical across the whole object surface. This limitation makes it difficult to apply the methods on a wider range of objects including multi-material objects,

something that we address in this work. Rosales *et al.* [12] attempted to lift this limitation by proposing a representation that consists of both shape and friction that were gathered through an exploration strategy. They then used a Gaussian Process to approximate the distribution of the friction coefficient over the surface. However, the results presented in that work show that the friction coefficient is only estimated for the regions that are explored by the robot. Unexplored areas on the object surface are then assigned a non-sense value. Our method, on the other hand, estimates friction coefficient also for unexplored areas based on information gathered from explored areas. Additionally, [12] only considers single-material objects such as a paperboard box, and a metallic can for experimental evaluation. In this work, we experimentally evaluated our method with different objects including both single-material and multi-material objects.

## III. PROBLEM FORMULATION

The problem of identifying the friction coefficients of an unknown object is difficult. A typical approach for this problem is to utilize haptic sensory feedback. However, haptic feedback is usually high cost, noisy, and unreliable due to the fact that haptic sensing introduces many issues such as low durability, low performance, high cost, and less compatibility with other sensors [18]. Another way of gathering data that can be processed for friction estimation is visual sensing. Although visual feedback is cheap and intuitive, it does not always provide enough information to identify friction coefficient of an unknown object. In this work, we address the problem of estimating friction coefficient of unknown objects lying on a supporting plane by combining both visual and haptic sensing. The goal of this letter is to understand the correlation between the known visual and haptic feature of the target object and, based on that correlation, to then extrapolate the unknown haptic features from the visual features.

Visual information about the scene is obtained by a RGB-D camera whose pose is known relative to the robot. Haptic feedback is gathered while performing exploratory actions on the target object using a robotic arm equipped with a plastic finger and force/torque sensor. To make it easier for latter representation, the regions that are touched by the robot during the haptic exploration are called *explored regions*, and the one that are not touched *unexplored regions*. Let  $\mathbf{V}$  and  $\mathbf{H}$  denote visual and haptic features, respectively. For the *explored regions*, both  $\mathbf{V}$  and  $\mathbf{H}$  are known while for *unexplored regions* only  $\mathbf{V}$  is known. The goal is to infer haptic feature  $\mathbf{H}$  from the visual feature  $\mathbf{V}$  of *unexplored regions* based on the joint  $(\mathbf{V}, \mathbf{H})$  model of the *explored regions*. In other words, the objective is to find the conditional probability  $\mathcal{P}(\mathbf{H} | \mathbf{V})$  for *unexplored regions*.

To this end, we propose to model the joint distribution  $\mathcal{P}(\mathbf{V}, \mathbf{H})$  by fitting a Gaussian Mixture Model (GMM) over all the visuo-haptic features of all points in the *explored regions*. More precisely, for an object with  $n$  materials, let us build the  $C$ -component GMM that better fit the visuo-haptic data from the *explored regions*, with one component for each of the  $n$  materials, plus one background component describing an uninformative visuo-haptic prior, i.e.,  $C = n + 1$ . Formally, we build

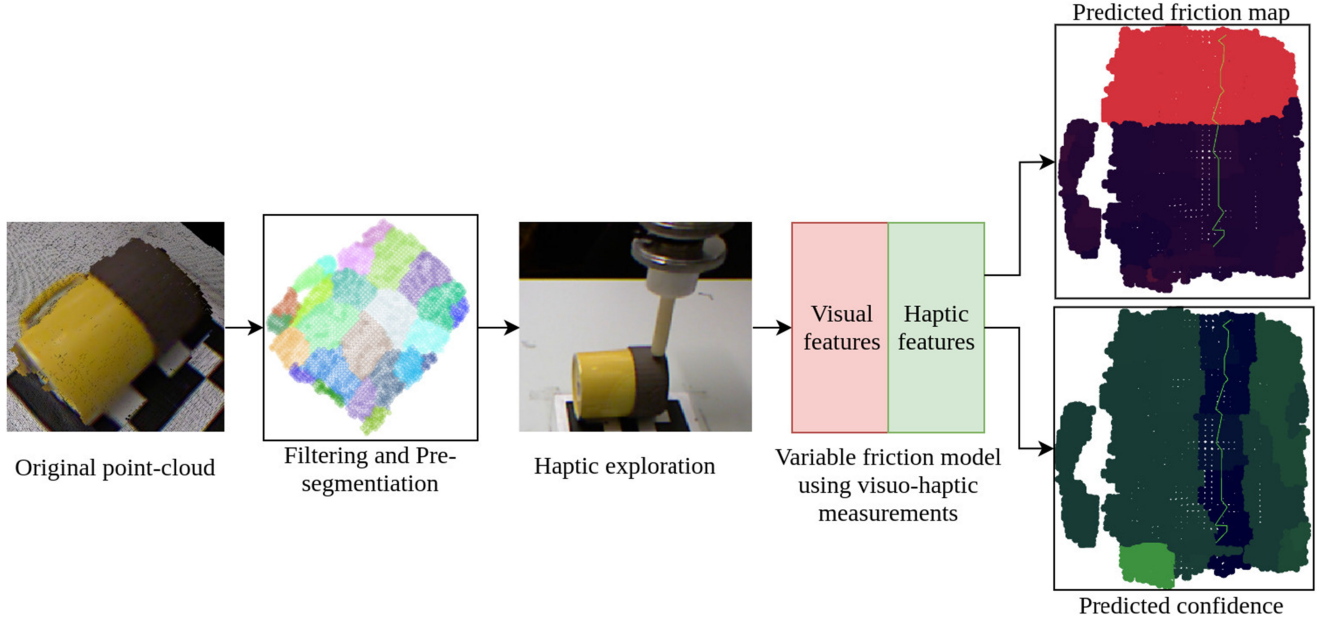


Fig. 2. The proposed pipeline: the object’s visual properties are first acquired as point-cloud, which is then filtered and pre-segmented into regions. The robot then performs a haptic exploration over some of the regions. The proposed model then is used to estimate the friction coefficient over the whole object, together with the corresponding confidence.

a multivariate GMM model  $M$  to estimate the joint probability distribution  $\mathcal{P}(\mathbf{V}, \mathbf{H})$  from gathered data in the *explored regions*, i.e.,

$$\mathcal{P}(\mathbf{V}, \mathbf{H}) \sim \sum_{c=1}^C \pi_c \mathcal{N}(\mu_c, \Sigma_c), \quad (1)$$

where  $\pi_c$ ,  $\mu_c$ , and  $\Sigma_c$  denote the prior probability, mean, and covariance of the  $c$ -th Gaussian component respectively. Formally, let us decompose the GMM parameters  $\mu_c$  and  $\Sigma_c$  as

$$\mu_c = \begin{bmatrix} \mu_c^{\mathbf{V}} \\ \mu_c^{\mathbf{H}} \end{bmatrix}, \quad \Sigma_c = \begin{bmatrix} \Sigma_c^{\mathbf{V}} & \Sigma_c^{\mathbf{VH}} \\ \Sigma_c^{\mathbf{HV}} & \Sigma_c^{\mathbf{H}} \end{bmatrix}. \quad (2)$$

After the model  $M$  is fitted to a given dataset, by using Gaussian Mixture Regression (GMR) [19], we can estimate the haptic feature  $\mathbf{H}_i$  at each data point  $i = 1, \dots, N$  in the *unexplored regions* given input  $\mathbf{V}_i$  by means of the conditional probability  $\mathcal{P}(\mathbf{H} | \mathbf{V})$  expressed as

$$\mathcal{P}(\mathbf{H}_i | \mathbf{V}_i) \sim \sum_{c=1}^C h_c(\mathbf{V}_i) \mathcal{N}(\hat{\mu}_c^{\mathbf{H}}(\mathbf{V}_i), \hat{\Sigma}_c^{\mathbf{H}}). \quad (3)$$

Then, given an input  $\mathbf{V}_i$ , the mean  $\hat{\mu}_i^{\mathbf{H}}$  and covariance  $\hat{\Sigma}_i^{\mathbf{H}}$  of its corresponding output  $\mathbf{H}_i$  are computed by

$$\begin{aligned} \hat{\mu}_i^{\mathbf{H}} &= \mathbb{E}(\mathbf{H}_i | \mathbf{V}_i) = \sum_{c=1}^C h_c(\mathbf{V}_i) \hat{\mu}_c^{\mathbf{H}}(\mathbf{V}_i), \text{ and} \\ \hat{\Sigma}_i^{\mathbf{H}} &= \sum_{c=1}^C h_c(\mathbf{V}_i) (\hat{\Sigma}_c^{\mathbf{H}} + \hat{\mu}_c^{\mathbf{H}}(\mathbf{V}_i) \hat{\mu}_c^{\mathbf{H}}(\mathbf{V}_i)^\top \\ &\quad - \hat{\mu}_i^{\mathbf{H}} (\hat{\mu}_i^{\mathbf{H}})^\top), \end{aligned} \quad (4)$$

$$\begin{aligned} \hat{\mu}_c^{\mathbf{H}}(\mathbf{V}_i) &= \mu_c^{\mathbf{H}} + \Sigma_c^{\mathbf{HV}} (\Sigma_c^{\mathbf{V}})^{-1} (\mathbf{V}_i - \mu_c^{\mathbf{V}}), \\ \hat{\Sigma}_c^{\mathbf{H}} &= \Sigma_c^{\mathbf{H}} - \Sigma_c^{\mathbf{HV}} (\Sigma_c^{\mathbf{V}})^{-1} \Sigma_c^{\mathbf{VH}}, \\ h_c(\mathbf{V}_i) &= \frac{\pi_c \mathcal{N}(\mathbf{V}_i | \mu_c^{\mathbf{V}}, \Sigma_c^{\mathbf{V}})}{\sum_{k=1}^C \pi_k \mathcal{N}(\mathbf{V}_i | \mu_k^{\mathbf{V}}, \Sigma_k^{\mathbf{V}})}. \end{aligned} \quad (5)$$

#### IV. IMPLEMENTATION

The system pipeline shown in Fig. 2 consists of:

- i) filtering and pre-segmenting the real objects,
- ii) conducting an haptic exploration process to gather and couple haptic data with visual data,
- iii) modeling the variable friction model using visuo-haptic data, and
- iv) inferring the friction coefficients and inference confidence of *unexplored regions* from the modelled distribution.

##### A. Visual Filtering and Pre-Segmentation

The scene in the original point-cloud  $Y$  contains a target object lying on a table. As the pose of the camera with respect to the robot is known, we can first remove points that are part of the supporting surface, and points that belong to the background indicated by their distance exceeding a certain threshold. We then obtain a filtered point-cloud  $\bar{Y}$  containing only the view of the object to be pre-segmented.

Formally, let  $y = (\mathbf{x}_p, \mathbf{V})$  denote a point in the filter point-cloud  $\bar{Y}$  ( $y \in \bar{Y}$ ),  $\mathbf{x}_p \in \mathbb{R}^3$  is the position of the point with respect to the camera frame, and  $\mathbf{V} \in \mathbb{R}^3$  is the RGB component vector representing its visual feature. It should be noted that for

objects with textured appearance, visual texture features could be used instead of color.

As discussed in the previous section, the main goal of this work is to estimate haptic features for *unexplored regions* based on the gathered visual and haptic features of *explored regions*. Thus, we consider a region-based representation of the target object. The idea is to divide the target object into a large number  $N$  of connected regions, wherein each region  $i = 1, \dots, N$  has its own visual feature and haptic feature. We achieved this using a state-of-the-art supervoxel segmentation named VCCS [20]. Given the filtered point-cloud  $\bar{Y}$ , we defined an  $N$ -region segmentation as a partition  $R = \{r_i\}_{i=1}^N$  of the points of  $\bar{Y}$ . More precisely, the regions must satisfy the following constrains:

$$\begin{aligned} \forall y \in \bar{Y} \quad (\exists r \in R \mid y \in r); \\ \forall r \in R \quad (\nexists r' \in R \mid r \cap r' \neq \emptyset). \end{aligned} \quad (6)$$

These constrains guarantees that all points in the filtered point-cloud  $\bar{Y}$  belong to a region, and no point can belong to two regions. The second step in Fig. 2 shows the result after the filtering and pre-segmentation step. The next step is to gather and couple haptic data with visual data through haptic exploration.

### B. Haptic Exploration and Tracking

In this work, haptic data is gathered using a force/torque (F/T) sensor attached to the wrist of a robot arm. The haptic data consists of the contact point on the object surface  $\mathbf{x}_c \in \mathbb{R}^3$  with respect to the robot base frame calculated using forward kinematics, and the contact force between the finger and the object  $\mathbf{f} \in \mathbb{R}^3$  expressed in contact frame. The dynamic friction coefficient, which represents haptic feature  $\mathbf{H}$ , can be then estimated using Coulomb friction model as

$$\mathbf{H} = c.o.f. = \frac{|\mathbf{f}_t|}{|\mathbf{f}_z|}, \quad (7)$$

where  $\mathbf{f}_t$  and  $\mathbf{f}_z$  are the tangent and normal forces at the contact point, respectively. The reason we chose Coulomb friction model is that it produces less noisy point-wise estimation and requires less computational time compared to others friction model such as LuGre model [21]. This claim is made based on the result of conducted experiments where we used both Coloumb and LuGre friction models to estimate the friction coefficient of a plastic plate (Fig. 3) and compared the results with the ground-truth presented in [22]. The calculated c.o.f. ranges from 0.2 to 0.35, while the ground-truth ranges from 0.2 to 0.4. In this work, we consider dynamic friction because it is easier and more efficient to measure compared to static friction, since static friction measurement requires starting and stopping motion for each measurement point. Furthermore, as static friction is often higher than dynamic friction, the latter will act as the safe lower bound in grasping scenarios.

To obtain haptic data of an object, we perform an exploratory action called lateral sliding on the object where we use the robot arm equipped with an F/T sensor to slide on the object surface along a linear path. Hybrid force and position control is used during the exploration to guarantee the contact with the object surface. The exploration depends on the object remaining immobile. In our case, the object is manually held to prevent the object from moving during the exploration. While not optimal,

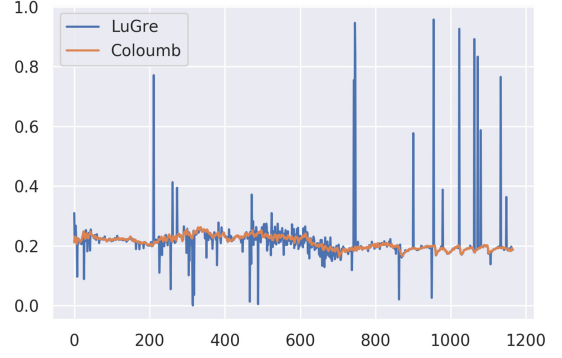


Fig. 3. A qualitative comparison between friction models. The orange line denotes the friction coefficient estimated using Coloumb friction model, while the blue line denotes the one using LuGre friction model (best viewed in color).

this is a typical way to stabilize objects under exploration procedures [23]. The problem could also be solved by using a dual arm setup to remove the need for human intervention.

During the exploration, we also need to map the gathered haptic data to the corresponding position in the pre-segmented visual data obtained from the previous step. This is done by tracking the position of the contact point  $\mathbf{x}_c \in \mathbb{R}^3$  through forward kinematics, and finding the point  $y$  in the object point cloud that has the smaller Euclidean distance to the current contact point ( $\mathbf{x}_p \approx \mathbf{x}_c$ ). We then assign to  $y$  the friction coefficient computed by (7) at  $\mathbf{x}_c$ . The benefit of using this approach is that some uncertainties caused by calibration procedure can be reduced. After the exploration and tracking, the explored points include both visual  $\mathbf{V}$  and haptic data  $\mathbf{H}$ , which can be represented as  $y = (\mathbf{x}_p, \mathbf{V}, \mathbf{H})$ . The next step is to learn a model from the *explored regions* and use it to infer haptic feature for *unexplored regions*.

### C. Variable Friction Model Using Visuo-Haptic Measurements

We consider a visuo-haptic dataset  $\xi = \{\xi_j\}_{j=1}^N$  defined by  $N$  observations  $\xi_j \in \mathbb{R}^D$ . Each datapoint  $\xi_j$  is represented with input/output components indexed by  $\mathbf{V}$  and  $\mathbf{H}$ , so that  $\xi_j = [\xi_j^{\mathbf{V}}; \xi_j^{\mathbf{H}}]$  and  $D = D^{\mathbf{V}} + D^{\mathbf{H}}$ . In this work,  $\xi_j$  is a concatenation of visual features, (input component) and haptic features (output component).

As the haptic exploration is conducted only once on a linear path along the object, the tip of the robot only touches a few points in each *explored region*. Thus, the number of points that has been assigned a friction coefficient value is always smaller than the total number of points of each *explored region*. Therefore, the covariance between RGB color components is computed using all of the points in the region while the covariance between friction coefficient and color components are computed using only the points that are touched by the robot. Then, given the visuo-haptic dataset  $\xi$ , we use a GMM with  $C$ -components optimized through Expectation Maximization (EM) to encode the joint probability distribution  $\mathcal{P}(\xi^{\mathbf{I}}, \xi^{\mathbf{O}})$ . After a GMM is fitted to the dataset, GMR can subsequently be used to estimate haptic features  $\xi_*^{\mathbf{O}}$  for visual features  $\xi_*^{\mathbf{I}} \in \mathbb{R}^D$  of *unexplored regions* as mentioned in Section III.



Fig. 4. The experimental setup.

Additionally, as discussed in Section III we include a background component in the GMM model to reflect the estimation uncertainty. The background component is constructed using measurements from the entire scene, thus representing the entire variability. If an input, i.e., visual feature, is not close to any main components, the corresponding output will depend primarily on the background component. This allows the model to capture the estimation uncertainty such that a high variance is predicted when a particular region is not visually similar to any of the regions with haptic measurements. This uncertainty measurement can be used to actively make requests for new haptic explorations in uncertain regions.

## V. EXPERIMENTS AND RESULTS

To demonstrate the ability of the proposed approach to estimate surface friction without being restricted by the assumption that objects have uniform homogeneous friction everywhere, we first evaluate the capabilities of the method by testing it with different objects, including multi-material objects. Furthermore, we report on the repeatability of the results. Afterwards, we present a grasping case study in order to demonstrate the benefits of accurate friction estimation in practical robotics applications.

### A. Experimental Setup

The experiments are performed using the *Franka Emika Panda* robot and a Kinect 360° camera to capture the input point-clouds as shown in Fig. 4. We used an Aruco marker for the extrinsic calibration of the camera. Once the input point-cloud was captured, it was filtered and pre-segmented as explained in Section IV. To perform the haptic exploration, we used a six-axis force-torque sensor (ATI Mini45) attached between the robot's wrist and the gripper. The haptic exploration is performed once for each object.

### B. Model Representation With Real Robot and Objects

To study the capability of the proposed method, we ran the experiment on fifteen different objects, shown in Fig. 5. Of these

objects, the book (Fig. 5(a)), the boardgame (Fig. 5(b)), the cereal box (Fig. 5(c)) and the toy blocks (Fig. 5(l)) represent single-material objects, where friction coefficient is identical across the whole surface, while the rest of the objects are composed of multiple materials (ranging from two to four).

Fig. 5 shows qualitative results obtained from the proposed method applied on the target objects. We show in the result the original object, the estimated friction, and the estimated uncertainty of each target object respectively. The haptic exploration is visualized as the green path. For the estimated friction, regions are colored according to their friction coefficient value, i.e., more red corresponds to higher friction. For the estimated uncertainty, the level of uncertainty is represented by green color, such that the greener the higher uncertainty. The results show that the proposed method has successfully estimated the friction coefficient for all the target objects even in presence of multiple materials. Specifically, the method is able to not only produce similar friction coefficients across the object surface in the case of single-material object (i.e., book and cereal box), but also to provide different friction coefficients for different parts of objects in the case of multi-material objects. For example, in the case of the yellow mug (Fig. 5(d)) the method estimated higher friction coefficient value for the rubber band, and provided an accurate boundary between the rubber part and the ceramic part of the object. Furthermore, the proposed method is shown to work well even for objects that are made from more than two materials such as the hammer (Fig. 5(o)) and the brush (Fig. 5(j)). In the case of the brush, the method was able to segment the object into four parts representing four different materials. This predicted friction map is accurate compared to the real object where the handle is made from both plastic and rubber, and the head is made from steel and fabric.

Another interesting point from the result is that our method works even in cases where visual features vary a lot but haptic features are identical across the surface, like the toy blocks and the board game. As this scenario would likely cause problems to methods estimating friction only from vision as discussed in Section II, this experiment shows the benefit of combining both visual and haptic feedback for friction estimation. Additionally, in the case of the butcher knife (Fig. 5(i)) with rubber handle, without encapsulating haptic feedback, one would hardly know if the handle has lower (made of plastic) or higher (made of rubber) friction coefficient compared to the steel blade. Since we conducted haptic exploration across the object, our method is capable of estimating higher friction coefficient for the handle, which is consistent with the ground truth object. However, objects having similar visual features for surfaces having different friction has proven challenging for our approach. This is the case for the modified cork trivet (Fig. 5(m)), where we covered half of the surface with a wood-textured wallpaper. The trivet is now composed of two materials with similar visual features. The result shows that our method provides a uniform friction map over the whole object surface. However, due to the variations in the haptic features along the explored path, the estimated uncertainty is shown to be high in the lower area of the trivet. One of the potential solution to overcome this problem is to conduct new haptic explorations in uncertain regions as discussed in Section IV.

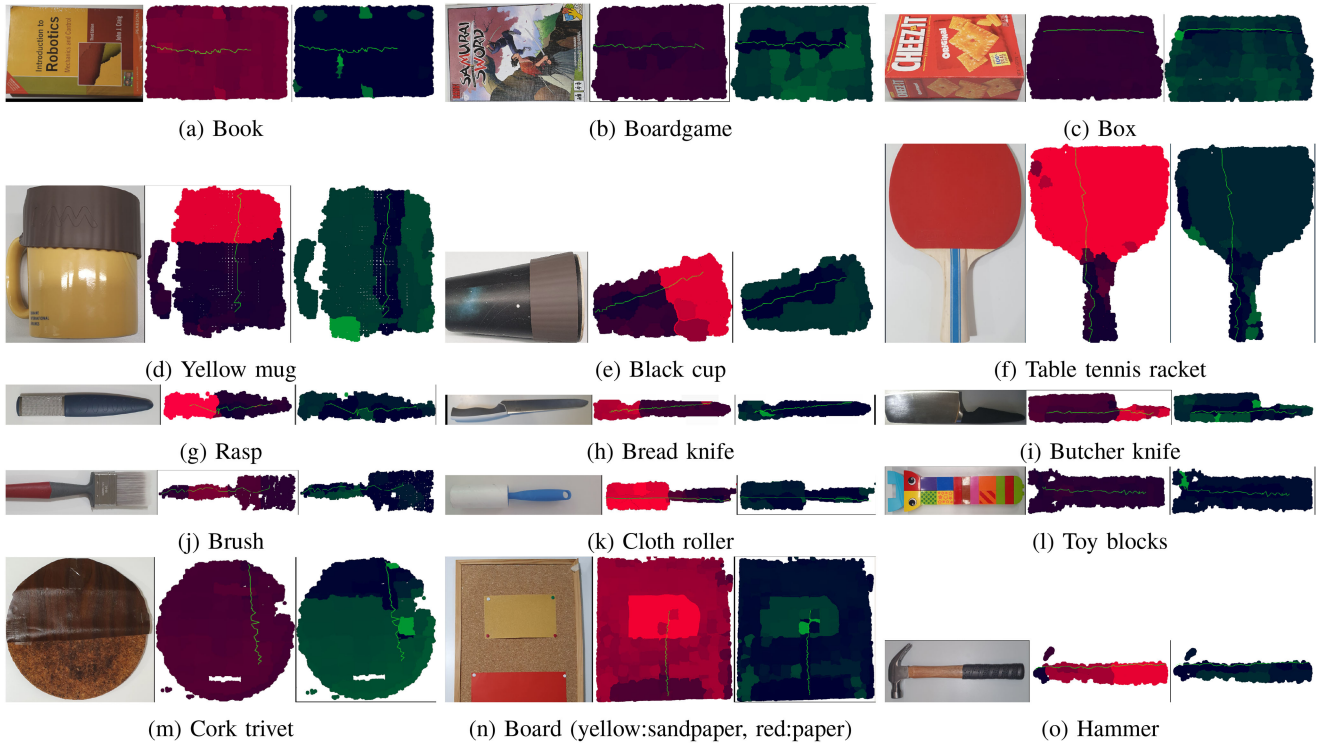


Fig. 5. Real objects, along with estimated friction coefficient and uncertainty returned by the proposed method. Red indicates higher friction coefficient value, while green denotes higher uncertainty. The haptic exploration is showed by the path in green (best viewed in color).

Furthermore, the experimental results also show the uncertainty map of the proposed method. In the case of the book and the cereal box, the uncertainty is extremely low because there are no abnormal visual features. In other words, all visual features of unexplored regions are almost as varied as the ones of explored regions. On the other hand, for the yellow mug case, the uncertainty is high at the left edge of the mug where the blue pattern is located. This result is suspected to be because the haptic exploration is carried out only in the middle of the mug, where the color is always yellow or brown. As the blue color is far from the explored visual feature, that region is classified by the background class, which in turn produces a high uncertainty for the prediction. A similar result is shown in the case of the board game. However, in all these cases, even though the uncertainty is high, the estimated friction is still similar to that of the rest of the regions of the same material.

*C. Repeatability of the Results*

Next we evaluated the repeatability of the proposed method by running the experiment five times on each object. The estimated friction coefficient is recorded and plotted as density plots for analysis as shown in Fig. 6, where, for each friction coefficient, we show the number of regions found having that value. For brevity, only the results of the cereal box representing single-material objects and the yellow mug representing multi-material objects are plotted for discussion. The results show that the proposed method produces reasonable result in term of repeatability. Note that despite small variation between repetitions, we

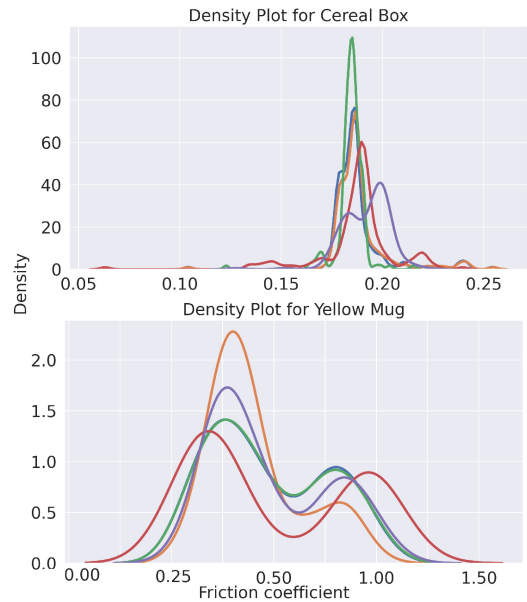


Fig. 6. The density plots of the cereal box and the yellow mug, showing consistent friction coefficient estimation over five repetitions of the experiment.

still can clearly see that both density plots represent accurately the number of the material in each case. For example, in the case of the cereal box the density plot only has one peak which denotes a single-material object while the density plot of the yellow mug contains two peaks representing a multi-material object.

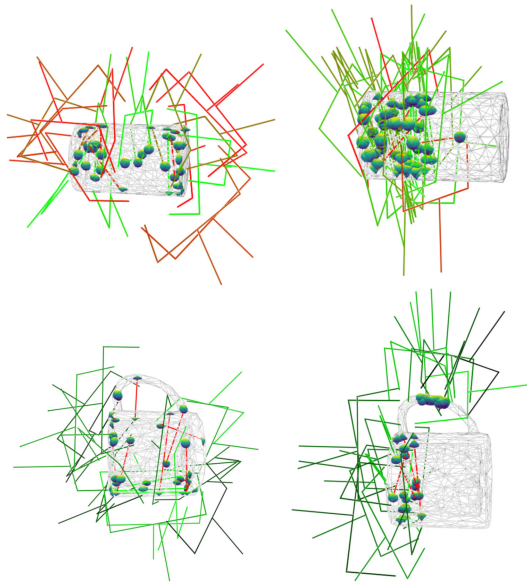
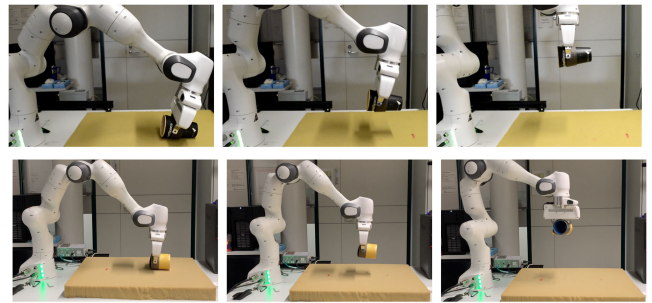


Fig. 7. Grasp sampling results for two objects, black cup (top) and yellow mug (bottom), in two cases: using uniform friction (left), and non-uniform friction model (right).

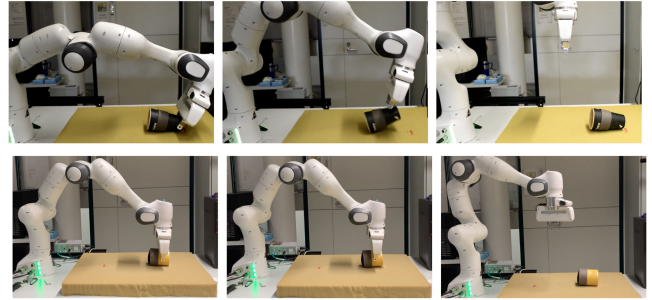
#### D. Grasping Case Study

In robotic grasping, grasp selection is usually complex as its stability takes into account different factors such as gripper geometry, object shape, mass, and surface friction. Although methods have been proposed to grasp without needing the estimation of physical properties [24], most common approaches for grasping rely on metrics based on physical properties to evaluate grasp quality. In particular, a set of best grasps are sampled and evaluated from a grasp sampling and evaluation algorithm. Typically, parameters like surface friction are kept constant during the process; however, the selected grasps that are executed in the real worlds may fail as well due to the contact with low friction surface. In order to demonstrate the usefulness of our method, we conducted a grasping case study where the estimated friction information is used to sample and evaluate the grasps.

In this demonstration, we first capture the target object from different viewpoints and merge them together in order to obtain a multi-view point-cloud of the object. The proposed method is then applied to the given point-cloud to produce the surface friction estimation. As the input of the grasp sampler is typically a mesh, we converted the estimated point-cloud to a mesh using Meshlab. Since the mesh does not contain any information about the estimated friction, we calculate the center of each face of the mesh, find its closet point and assign the friction coefficient value of the point to the corresponded face. Next, the mesh with assigned friction coefficient is fed to a grasp sampler to generate grasp candidates. Grasp candidates are sampled using antipodal grasp sampling method. Specifically, we randomly select a point on the mesh and assume this point as the first contact point. At this first contact point, a direction ray that lies inside of the friction cone is generated. If the ray intersects with a certain face, the intersect point will be the second contact point, and



(a) Successful grasps executed on a region of high friction



(b) Unsuccessful grasps executed on a region of low friction

Fig. 8. The robot executing different grasps on an object. The grasps are proposed by a sampler employing Fig. 8(a) the proposed friction model, or Fig. 8(b) a uniform friction model.

the closing vector is assured. Next, we randomly generate an approach vector along the closing vector. The generated grasps will then be checked for different type of collisions. The grasps that collide with the mesh are filtered out and the grasps that are not in collision are then evaluated using Ferrari & Canny L1 quality metric [25].

In this study, we sample 1000 grasps each on two object models: one with uniform friction coefficient and one with non-uniform friction coefficient computed using the proposed method. The coefficient of friction in the case of the uniform friction was chosen similar to that of the high friction area computed by our method. The object used in this study are the black cup and the yellow mug as shown in Fig. 5. Good grasp candidates together with friction cones at contact points of both cases are presented in Fig. 7. These results show that the sampler behaves as expected: In the case of the model with a uniform friction coefficient (Fig. 7 left figures), the grasp candidates are distributed across the entire object, while the grasp candidates on the non-uniform friction coefficient model (Fig. 7 right figures) only appear around the left side of the object, where the friction coefficient is high.

To study the effect of the proposed method on grasp performance in the real world, grasp candidates were generated using the antipodal grasp sampling method. Five grasps with the highest quality according to the Ferrari-Canny minimum quality metric were chosen in each case and executed with the real robot. For the uniform friction coefficient case, we chose the five best grasps that are on the right side of the object. By doing this, we can see how the grasps actually behave if they make contact with low friction area of the object. The grasps in

both cases were executed with the same grasping force of 20 N. To evaluate if a grasp was successful, the robot moved to the planned grasp pose, closed its fingers, and moved the arm back to the starting position. Once there, the gripper was rotated around the last joint (Fig. 8). A grasp was considered successful if the object was grasped in a stable manner for the whole procedure and unsuccessful if the object was dropped.<sup>1</sup>

All evaluated grasp candidates generated with the sampler that utilized the proposed method were successful as they grasped the object at the high friction area. The chosen grasp candidates generated in the uniform friction case failed to grasp the object since they aimed for the low friction area of the object, which in turn, produced slippage during grasping. The results show that even when the grasp sampler generates grasps with high quality under the uniform friction case, the grasps still may fail when executed in the real world due to incorrect friction assumption.

## VI. CONCLUSION

We presented an approach that enables the estimation of local object physical properties, like the surface friction coefficient, from visual and haptic cues, which goes beyond the state-of-the-art by lifting the assumption that the target object has uniform friction across its surface. The key component in this work is the use of a probabilistic model to estimate the surface friction coefficient of the unexplored areas from visuo-haptic data gathered by haptic exploration. Furthermore, we also presented an approach to represent a level of uncertainty of the estimate. This could be useful in future work to actively make requests for new haptic explorations in the regions with high uncertainty. We demonstrated the capability and the repeatability of the approach through experiments on a wide range of objects including single-material and multi-material objects. The results show that the proposed approach is capable of providing object representations with varying surface friction coefficient. Moreover, the friction coefficients can be used to guide grasp planning towards areas of high friction, improving robotic grasping success rate. Despite the good results, there is still room for improvements. For example, using more robust and complex visual features, like the one that could be obtained by modern deep learning approaches, could improve the quality and the robustness of the estimated results against, *e.g.*, reflection and variations of lighting.

## ACKNOWLEDGMENT

The authors would like to thank Jens Lundell for the invaluable help in developing the grasping experiments.

## REFERENCES

- [1] T. Fulhammer *et al.*, "Autonomous learning of object models on a mobile robot," *IEEE Robot. Automat. Lett.*, vol. 2, no. 1, pp. 26–33, Jan. 2017.

<sup>1</sup>A video showcasing the grasping case study can be [Online]. Available: <https://irobotics.aalto.fi/video-friction/>

- [2] J. Prankl, A. Aldoma, A. Svejda, and M. Vincze, "Rgb-d object modelling for object recognition and tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 96–103.
- [3] T. Weise, T. Wismer, B. Leibe, and L. Van Gool, "In-hand scanning with online loop closure," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2009, pp. 1630–1637.
- [4] H. Zhang, K. Dana, and K. Nishino, "Friction from reflectance: Deep reflectance codes for predicting physical surface properties from one-shot in-field reflectance," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 808–824.
- [5] J. Xue, H. Zhang, K. Dana, and K. Nishino, "Differential angular imaging for material recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6940–6949.
- [6] M. Brando, Y. M. Shiguematsu, K. Hashimoto, and A. Takanishi, "Material recognition cnns and hierarchical planning for biped robot locomotion on slippery terrain," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, 2016, pp. 81–88.
- [7] T. Yoshikawa and M. Kurisu, "Identification of the center of friction from pushing an object by a mobile robot," in *Proc. IEEE/RSJ Int. Workshop Intell. Robots. Syst.*, 1991, pp. 449–454.
- [8] M. R. Tremblay and M. R. Cutkosky, "Estimating friction using incipient slip sensing during a manipulation task," in *Proc. IEEE Int. Conf. Robot. Automat.*, 1993, pp. 429–434.
- [9] H. Liu, X. Song, J. Bimbo, L. Seneviratne, and K. Althoefer, "Surface material recognition through haptic exploration using an intelligent contact sensing finger," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 52–57.
- [10] G. De Maria, C. Natale, and S. Pirozzi, "Slipping control through tactile sensing feedback," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2013, pp. 3523–3528.
- [11] H. C. Lin and M. Mistry, "Contact surface estimation via haptic perception," in *Proc. IEEE Int. Conf. Robot. Automat.*, 2020, pp. 5087–5093.
- [12] C. Rosales, A. Ajoudani, M. Gabbicini, and A. Bicchi, "Active gathering of frictional properties from objects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 3982–3987.
- [13] T. Bhattacharjee, A. A. Shenoi, D. Park, J. M. Rehg, and C. C. Kemp, "Combining tactile sensing and vision for rapid haptic mapping," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2015, pp. 1200–1207.
- [14] T. A. Whitaker, C. Simoes-Franklin, and F. N. Newell, "Vision and touch: Independent or integrated systems for the perception of texture?," *Brain Res.*, vol. 1242, pp. 59–72, 2008.
- [15] S. J. Lederman and R. L. Klatzky, "Extracting object properties through haptic exploration," *Acta Psychologica*, vol. 84, no. 1, pp. 29–40, 1993.
- [16] K. M. Lynch, "Estimating the friction parameters of pushed objects," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 1993, pp. 186–193.
- [17] T. Maeno, T. Kawamura, and Sen-Chieh Cheng, "Friction estimation by pressing an elastic finger-shaped sensor against a surface," *IEEE Trans. Robot. Automat.*, vol. 20, no. 2, pp. 222–228, Apr. 2004.
- [18] A. Yamaguchi and C. G. Atkeson, "Recent progress in tactile sensing and sensors for robotic manipulation: Can we turn tactile sensing into vision?," *Adv. Robot.*, vol. 33, no. 14, pp. 661–673, 2019.
- [19] S. Calinon, "A tutorial on task-parameterized movement learning and retrieval," *Intell. Serv. Robot.*, vol. 9, no. 1, pp. 1–29, 2016.
- [20] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel cloud connectivity segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2027–2034.
- [21] C. Canudas de Wit, H. Olsson, K. J. Astrom, and P. Lischinsky, "A new model for control of systems with friction," *IEEE Trans. Autom. Control*, vol. 40, no. 3, pp. 419–425, Mar. 1995.
- [22] "Coefficient of friction, rolling resistance and aerodynamics," Accessed: Jan. 14, 2021. [Online]. Available: <https://www.tribology-abc.com/abc/cof.htm>
- [23] R. S. Jamisola, P. Kormushev, A. Bicchi, and D. G. Caldwell, "Haptic exploration of unknown surfaces with discontinuities," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2014, pp. 1255–1260.
- [24] M. Adjigble, N. Marturi, V. Ortenzi, V. Rajasekaran, P. Corke, and R. Stolkin, "Model-free and learning-free grasping by local contact moment matching," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2018, pp. 2933–2940.
- [25] C. Ferrari and J. Canny, "Planning optimal grasps," in *Proc. IEEE Int. Conf. Robot. Automat.*, 1992, pp. 2290–2295.