Visual 3D Perception for Interactive Robotic Tactile Data Acquisition

Jain, Siddarth; Corcodel, Radu; van Baar, Jeroen

TR2021-092 August 24, 2021

Abstract

In this paper, we present a novel approach for tactile saliency computation on 3D point clouds of unseen object instances, where we define salient points as those that provide informative tactile sensory information with robotic interaction. Our intuition is that the local 3D surface geometries of objects contain characteristic information both in terms of texture and shape which can provide important discriminating information for tactile interactions. We solve the problem by taking as input a 3D point cloud of an object and develop a geometric approach which computes the tactile saliency map for the object without requiring pre-training. We furthermore develop a formulation to compute grasps using the tactile saliency for prehensile probing manipulation. We demonstrate our framework with evaluation on a variety of household objects in real-world experiments. Since it is difficult to manually define a ground truth tactile saliency measure, we evaluate our approach by having a human subject provide saliency information as baseline in pilot experiments. Results show good performance of our algorithm both in terms of the computation of tactile saliency and its usefulness to acquire informative tactile sensory data with a real-world robot.

IEEE International Conference on Automation Science and Engineering (CASE)

Mitsubishi Electric Research Laboratories, Inc.
201 Broadway, Cambridge, Massachusetts 02139

2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)
August 23-27, 2021. Lyon, France

Visual 3D Perception for Interactive Robotic Tactile Data Acquisition

Siddarth Jain, Radu Corcodel, and Jeroen van Baar

Abstract— **In this paper, we present a novel approach for tactile saliency computation on 3D point clouds of unseen object instances, where we define salient points as those that provide informative tactile sensory information with robotic interaction. Our intuition is that the local 3D surface geometries of objects contain characteristic information both in terms of texture and shape which can provide important discriminating information for tactile interactions. We solve the problem by taking as input a 3D point cloud of an object and develop a geometric approach which computes the tactile saliency map for the object without requiring pre-training. We furthermore develop a formulation to compute grasps using the tactile saliency for prehensile probing manipulation. We demonstrate our framework with evaluation on a variety of household objects in real-world experiments. Since it is difficult to manually define a ground truth tactile saliency measure, we evaluate our approach by having a human subject provide saliency information as baseline in pilot experiments. Results show good performance of our algorithm both in terms of the computation of tactile saliency and its usefulness to acquire informative tactile sensory data with a real-world robot.**

I. INTRODUCTION

For performing everyday tasks, humans utilize multiple sensory modalities to infer object properties. Although humans can distinguish many object and surface properties by touch alone, visual information has been shown to influence tactile judgments of object identities, spatial features, texture, and even heaviness [1]. For robotic manipulation, tactile information is of key importance to enable close physical interactions in applications such as autonomous manufacturing. Tactile sensory information can help enable perception of intrinsic object properties such as stiffness and hardness, along with object identities and their pose information for robotic manipulation. Although intrinsic object properties cannot be effectively perceived by visual perception, vision is a fast and global modality which can complement and guide tactile sensory data acquisition for robotic manipulation.

Learning to make inferences from tactile sensory data is a challenging task due to the sparse and local nature of touch. Furthermore, if we consider the problem while performing the task with a high degree-of-freedom (DOF) robot, the space of possible actions to acquire tactile sensory data becomes continuous and high-dimensional (6D). Thus, collecting informative representations of data samples for a variety of object classes and geometries becomes a difficult task. One could randomly and/or uniformly sample objects to acquire tactile sensory data [2], but this would be extremely time-consuming and performance might not generalize across object types. Thus, one needs to (i) know how to efficiently explore a variety of objects types and geometries and (ii) be able to discriminate between the object types based on the acquired tactile information. These two elements correspond to estimating which aspect of the object are salient to touch (and which are not), and utilizing that information to intelligently acquire tactile sensory data.

In this paper, we explore determination of tactile saliency for informative tactile data acquisition with robotic interactions. We aim to determine tactile saliency from visual perception, using 3D point cloud depth data of objects. We refer to this problem as *tactile saliency* determination and we define tactile saliency map points on an object's representation as those which are more likely to provide useful tactile information with robotic interactions. Our algorithm performs the determination of tactile saliency from point cloud depth data of a scene containing unknown objects by leveraging their geometric information. We believe that certain geometric characteristics and regularities in objects can be exploited for tactile saliency determination. Our intuition is that the 3D local surface geometries of objects contain characteristic information both in terms of texture and shape (as curvature) and that geometrical analysis of object's point cloud local neighborhood can provide important discriminating information. We further explore the utilization of tactile saliency for computation of 6D grasp poses and demonstrate informative acquisition of tactile sensory data from objects by robotic prehensile probing manipulation.

In summary, our work makes the following contributions:

- We present an algorithm which leverages geometric information for determination of tactile saliency from point cloud depth data, with the aim to guide acquisition of useful tactile data with robotic interactions. Our approach does not require pre-training and work with novel objects.
- We develop computation of 6D grasp poses based on tactile saliency map information and demonstrate its utilization to acquire tactile sensory data with prehensile probing manipulation by a robotic gripper.
- We validate our contributions with comparative experiments on a variety of household objects in real-world experiments on a robotic arm.

The remainder of the article is organized as follows. Related work is discussed in Section II. Section III presents the problem formulation and and details of our algorithm. Section IV presents the experimental details with results. In Section V, we present the discussion and then conclude with directions for future research in Section VI

Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA sjain@merl.com, corcodel@merl.com, jeroen@merl.com

II. Related Work

Visual saliency is a well-studied topic which has been explored in multiple disciplines of computer vision [3], where highlighting visually salient regions or objects in an image has drawn extensive attention. Previous works compute image saliency maps [4] and identify salient objects and regions in 2D images and videos [5]. Recently, deep learning based models using convolutional neural networks (CNNs) have been explored for image saliency determination [6]. 3D meshes and point cloud data contain depth information (in addition to RGB color) and usually consist of larger data size compared to RGB images and videos, making saliency determination more challenging. Recently, saliency determination for meshes has been explored, for example mesh saliency computation with Gaussian-weighted mean curvatures for selection of visually interesting regions [7].

In the field of robotics, there exists literature on affordance detection which aims to localize object parts and identify their functions [8], and grasp detection using RGB images and point cloud data [9], [10] for manipulation. With the recent advancements in tactile sensing, the state-of-the-art in robotics is moving toward closed loop grasping and advanced in-hand manipulation. Recent work focuses on tactile recognition problems including recognition of object instances [2], [11], [12], surface texture information [13], shape information [14], and stiffness properties [15]. Tactile sensing has also been utilized for improving robotic manipulation and grasping [16]. While existing research focuses on inferring contact and grasping using tactile sensing, in this paper we focus on the determination of tactile saliency which aims at inferring from 3D visual perception the tactile salient aspects of objects for robotic interaction.

There has been recent work in computer graphics targeted towards tactile saliency detection with human interactions. Tactile saliency determination is performed for the purpose of determining salient points for human interactions with objects on 3D virtual meshes [17] and 2D sketches [18] using deep learning. In contrast, we study a different aspect of tactile saliency with the aim to visually determine salient points which are most useful for a robot to acquire tactile information from objects. Our problem setting is challenging as we aim at computing a tactile saliency map on novel objects to enable an autonomous robot to gather informative tactile sensory data. Unlike previous works in computer graphics, our approach does not require a training phase, and provide tactile saliency determination along with associated grasp poses for robotic probing of objects.

III. Problem Formulation & Framework

Acquisition of tactile sensory observations from an object are restricted to the local regions of object's surface felt by robot fingers. Thus, obtaining more complete tactile information would require a multitude of systematic touches on the object, which deems expensive in practice. We propose to make the acquisition of the tactile modality more efficient by determining tactile saliency map of objects. We autonomously select a small number of touches on objects, using the computed tactile saliency maps. The steps in our proposed framework involves: point cloud preprocessing and segmentation; tactile saliency map computation; and 6D touch poses generation for probing with robotic manipulation for tactile data acquisition.

A. Preprocessing and Segmentation

The visual sensory input to our framework is 3D point cloud data of the scene from a RGB-D sensor, where each point is represented by a six-tuple (x, y, z, r, g, b), i.e., its 3D position in camera coordinates and RGB color information. The sensory input is transformed into the base frame of the robot Rb_f, in which the Z-axis is perpendicular to the ground (Z_R), the Y-axis is front-to-back (Y_R) and the X-axis is left-to-right (X_R). We first preprocess the input data to filter out sparse points which are considered to be noise, and then segment the scene into individual objects. We find the individual objects by performing Euclidean clustering such that the resultant is the set of individual object's point cloud. For simplicity, in further discussion we consider the point cloud of an object of interest denoted as P.

B. Tactile Saliency Map Determination

In this section, we describe our algorithm for tactile saliency map determination. Our algorithm performs the determination of tactile saliency from point cloud depth data of a scene containing unknown objects by leveraging their geometric information. Our intuition is that the 3D local surface geometries of objects contain characteristic information both in terms of texture and shape, i.e, curvature, and that geometrical analysis of the local neighborhood for the object's point cloud can provide important discriminating information for the determination of tactile saliency.

Given the point cloud of an object we first find clusters which satisfy smoothness constraints based on surface normals and curvature. For the identified clusters, if their surface area is greater than a threshold value determined by the tactile sensing patch used to probe the objects, we perform voxelization of the points and identify supervoxels in each cluster which conform to geometric relationships and object boundaries. We then compute tactile saliency based on the interconnections of the computed clusters and all individual supervoxels, with the intuition that the local neighborhood at these interconnections contains characteristic information both in terms of change in texture and shape, which can provide important discriminating tactile information.

We next describe our algorithm for tactile saliency map computation in detail. For a point cloud of an object represented by P, the representative tactile saliency map is comprised of a saliency value for all points in P, where the saliency value is based on the likelihood to provide informative tactile information. For recognizing an object by touch, its shape is an important cue that associates several characteristic features, such as edges, curvature, and surface area. We form this intuition as our basis and we focus on geometric computation of tactile saliency.

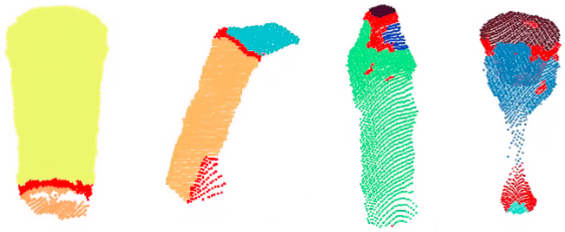

Fig. 1. Examples of the clustering approach where the set of points are considered to be a part of the same smooth surfaces are clustered together shown for a sample of objects (including a paper cup, box, water bottle, and wine glass), where different colors indicate the identified clusters.

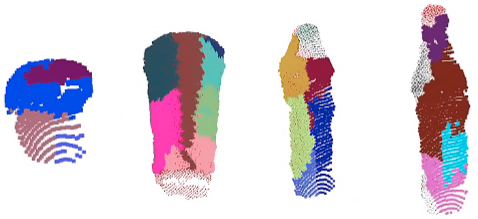

Fig. 2. Supervoxles computed on identified clusters in point clouds of sample objects (including an apple, paper cup, water bottle, and oil bottle) where different colors indicate supervoxels in the cluster. Uncolored or gray regions represent the small sized clusters which are considered to be a part of the same smooth surfaces and were not processed for the supervoxel computation.

The first step of our tactile saliency determination algorithm involves computation of the normal and curvature value of each point in P. Next, the points are clustered according to the normal and curvature values. We implement an approach for clustering similar to [19] [1]. Figure 1 shows an example of the clusters found using the above approach on several point clouds of objects, where different colors represent different identified clusters.

Once we have identified the clusters $\mathbf{C}$ in P which are close enough in terms of the smoothness constraints, for every cluster C_k in $\mathbf{C}$ we check if its surface area is greater than a threshold value θ_{surf} (set as the surface area of the tactile sensing patch). For every such cluster, we perform voxelization of the points in C_k based on the distance and density of the points in the cluster with the aim to identify supervoxels in each cluster, which conform to geometric relationships and object boundaries. We generate supervoxels similar to the VCCS algorithm [20] with an adjacency graph. Initially, voxels are generated by an octree algorithm on points in C_k, and next an adjacency graph is constructed for identifying the spatial relationship between voxels using a 26-adjacency neighborhood connectivity. This is accomplished by searching the voxel Kd-tree, and for a given voxel, the centers of all 26-adjacent voxels are contained within $\sqrt{3}$ $\times$ V_{res}, where V_{res} represents the voxel resolution used for segmentation, set to $0.005m$ in our experiments. We utilize the adjacency graph for clustering to generate supervoxels and finding neighborhoods of the supervoxels. We initialize seeds for initial supervoxel centers, by first dividing the space into a voxelized grid with seed resolution higher than the resolution with which the input cluster point cloud is quantified. We then compute supervoxel feature vectors by finding the center of a seed voxel and its connected neighbors within 2 voxels, where each seed is described by 39 dimensional features that describe spatial coordinates, colors and local surface model properties computed by FPFH pose-invariant features [21], given as,

$$\mathbf{F} = [x, y, z, L, a, b, FPFH_{1\ldots33}]. \tag{1}$$

Next, the voxels in C_k are clustered to form supervoxels based on their spectral and geometrical relationship in 3D space, where a supervoxel is a group of voxels that share similar characteristics based on feature $\mathbf{F}$. Clustering is performed iteratively by means of the distance metric, the adjacency graph, and the search volume of the supervoxel, where the normalized distance D is given as,

$$D = \sqrt{\frac{\lambda D_c^2}{m^2} + \frac{\mu D_s^2}{3S_{dist}^2} + \varepsilon D_H^2} \tag{2}$$

where D_c is the color distance in euclidean space normalized by a constant m, D_s is the spatial distance, D_H is the distance in FPFH space. λ, μ and ϵ control the influence of color, spatial and geometric relationship between voxels respectively. We use $\lambda = 0.2, \mu = 0.8, \epsilon = 0.8$ resp. S_{dist} determines the distance between supervoxels, we set S_{dist} to $0.01m$. The influence parameters for supervoxels computation were set according to importance to normal and spatial geometric relationship for identification of tactile saliency. Figure 2 shows example results of the supervoxel clusters found using the above approach on the clusters of several point clouds of objects, where different colors represent the identified supervoxels in the cluster. We note that the size of the voxel and the resolution of seeds can affect the performance and we set these factors empirically according to the object densities and the varying range from the sensor to the objects.

Once we have identified the clusters and supervoxles in P, we compute a tactile saliency map which comprises of points in the point cloud P which are at the interconnections of the computed clusters and all individual supervoxels in P. These points contain characteristic information in the local neighborhood, both in terms of change in texture and change in shape, and thus provides important discriminating tactile information.

C. Generation of Tactile Grasp Poses

Given the computed tactile saliency map for an object, we then compute 6D tactile grasps. Such grasp poses are aimed to be feasible for prehensile probing on the object in order to efficiently acquire informative tactile sensory data. Our hypothesis is that such a grasp pose when used for probing

[1] For threshold values, we empirically set the angle threshold to 0.06 rad and curvature threshold to 1.0

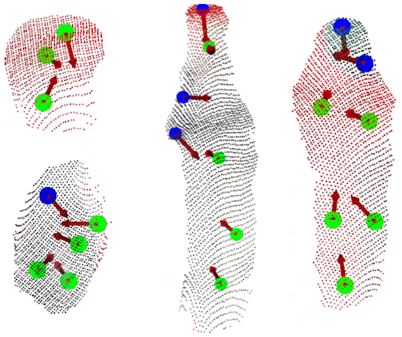

Fig. 3. Figure shows the results of tactile grasp poses computed by our approach on a sample of test object point clouds (an apple, football, water bottle, and oil bottle). Note the green and blue color represents whether the tactile grasp pose was generated from an initially segmented small size cluster or a supervoxel region respectively.

an object would make contact with the tactile salient aspects of the object to acquire informative tactile sensory data.

For each of the identified clusters and supervoxels in P, we compute their surface normal from the viewpoint of the object's 3D centroid. Next, we compute the cross product of the surface normal vector and the up vector along the Z-axis (Z_R) of the referential (base frame) to get the axis of rotation. We then take the dot product of the surface normal vector and the up vector to compute the angle of rotation, and compute the quaternion orientation. With the 3D centroid of the respective cluster or supervoxel, the quaternion orientation, and considering the surface area of the tactile patch and gripper geometry, we compute grasp poses (referred to as tactile grasp poses). Figure 3 shows some example poses for test objects.

We utilize the computed tactile grasp poses to determine viable and safe grasp candidates that can be executed by a robot. We use the grasp pose detector (GPD) [10] and generate a set of all feasible 6D grasps (GPD grasp poses) on the object's point cloud where the grasps are filtered to avoid collisions, and also satisfy the gripper and kinematics constraints. Next, we perform a radius based nearest neighbor classifier training on the GPD grasp poses and locate the GPD grasp poses that are within a given fixed radius of each of the tactile grasp pose computed by our algorithm. The fixed radius is based on the associated cluster or supervoxel size that comprises the tactile pose. Next, we compute the orientation similarity between each of the nearest neighbors in the GPD grasp set and the associated tactile poses. We compute the orientation similarity based on the Euler angles and denote (α_1, β_1 γ_1) and (α_2, β_2 γ_2) as the two sets of Euler angles, then

$$\begin{aligned} &\phi : E \times E \rightarrow \mathbb{R}^+, \\ &\phi((\alpha_1, \beta_1, \gamma_1), (\alpha_2, \beta_2 \gamma_2)) \\ &= \sqrt{d(\alpha_1, \alpha_2)^2 + d(\beta_1, \beta_2)^2 + d(\gamma_1, \gamma_2)^2} \end{aligned} \tag{3}$$

where $d(x, y)$ = min { | x - y |, 2π - | x - y | } denotes the normalized difference between the two angles such that $0 \leq d(x, y) \leq \pi$, and α, $\gamma \in$ [-π, π); $\beta \in$ [-π/2, π/2). For each tactile pose, we then select the first n nearest neighbors GPD poses, with $n = 3$, based on the orientation similarity to the corresponding tactile pose, in order to perform prehensile probing on the object at the detected poses by GPD.

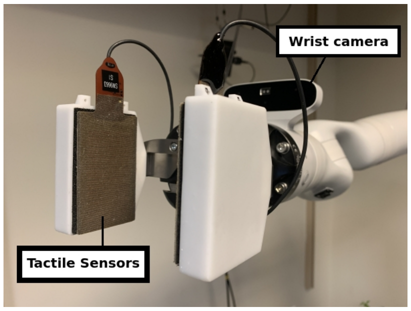

Fig. 4. Figure shows the robot hardware used in our experiments with the instrumented tactile sensor and the wrist camera. The flexible tactile sensor is affixed to a foam backing to allow compliance.

IV. EXPERIMENTS & RESULTS

Our experimental work aims to evaluate the performance of our approach both for tactile saliency determination and its utilization to intelligently acquire tactile sensory data with 6D grasp pose generation. We used a test set of $n_o = 11$ household objects in our experiments, namely: apple (ap), box (bx), camera lens (ln), koala bear plush toy (kb), glass oil bottle (ob), paper cup (pc), foam football (fb), soda can (sc), tennis ball (tn), plastic water bottle (wb), and a wine glass (wg). The test set of objects is shown in Fig. 5.

Robot Hardware & Tactile Sensors: Our research platform for the experiments described in this section is the Kinova Gen3 robot arm (Kinova Robotics, Canada) [22], a 7-DoF manipulator with a Robotiq Hand-E parallel jaw gripper. We instrumented both gripper finger pads with pressure based tactile sensor arrays [23] as shown in Figure 4. Each planar tactile array contain 10 × 18 calibrated capacitive taxels (tactile sensing pixels) which measure contact pressure in the range 0-5PSI with a 14-bit resolution. The active area of a tactile array is 55mm × 30mm. The calibrated pressure data is converted into 3D localized tactile data, similar to [2], constructed as a NURBS polynomial surface patch [24]. The visual input to our algorithm is from the wrist mounted RGB-D camera of the robotic arm.

Experiment Protocol: We aim to evaluate the similarity between the algorithm computed tactile saliency maps to those generated by a human as baseline on the test set objects, and furthermore we also conduct an evaluation of the similarity between the acquired tactile sensory data with prehensile probing manipulation across the two modalities. In experiments, both the algorithm and a human subject were presented with a large set of GPD generated 6D grasp poses (>200) for each object in the test set ($n_o = 11$), and they were tasked to select grasp poses which would result in better tactile sensory information when used to perform prehensile probing on the object. The grasp pose detector (GPD) ranks the quality of generated proposals by first pruning proposals that are infeasible based on the robot's gripper

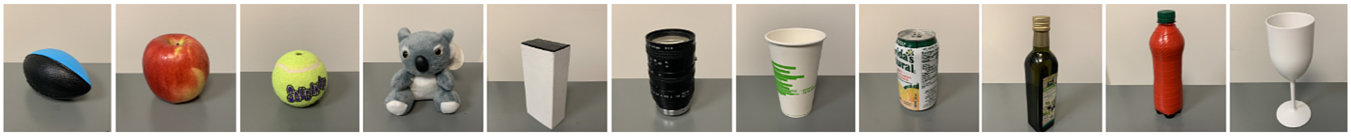

Fig. 5. Test set of objects used in our experiment spanning a wider range of material properties and geometry.

TABLE I

Average distance between grasp poses determined by our algorithm and the nearest neighbor human-directed grasp pose. The first two rows are Euclidean positional error (m), and the last two rows are orientation error according to Eq. 4. Our algorithm is able to propose grasps poses in close proximity of the human baseline to acquire tactile sensory data.

	Human baseline vs.	ap	box	kb	le	ob	pc	fb	tb	wg	wb	sc
Position error	Ours (trial 1)	0.012	0.018	0.0004	0.021	0.015	0.009	0.034	0.030	0.017	0.10	0.013
	Ours (trial 2)	0.019	0.018	0.011	0.013	0.018	0.029	0.017	0.027	0.040	0.011	0.020
Orientation error	Ours (trial 1)	0.405	0.166	0.241	0.173	0.146	0.143	0.342	0.420	0.203	0.219	0.195
	Ours (trial 2)	0.433	0.057	0.341	0.229	0.225	0.147	0.388	0.290	0.203	0.205	0.176

and kinematics constraints, followed by pruning based on a cost function [10]. The cost function takes into account the approach angles and the distances in the configuration space. In contrast, our algorithm computes a set of tactile grasps, with the aim to acquire informative tactile sensory information.

The test objects were presented one by one in a random pose configuration, on a table in front of the robot two times (trial 1 and trial 2), for a total of 22 trials and selection was the same for both the human and algorithm. The human subject was able to view the experimental setup, as well as the 3D point cloud with color information of the test object on a computer monitor screen, where the selectable GPD generated 6D grasp poses on the test objects were also displayed with an overlay on the point cloud. The visualization could be rotated by the human in 6D using a computer mouse to change viewing angles (see Figure 6). The human subject was tasked to select the best set of grasp poses from the visualization of the object which when executed in order to perform prehensile probing on the object would result in informative tactile sensory data. Our algorithm was provided as input with the same point cloud and the available pool of GPD generated grasp poses, which then were processed autonomously to determine a tactile saliency map on the object (Section IV-B) and a set of tactile 6D grasps based on the saliency information (Section IV-C). Only the number of grasps computed by the algorithm (n_a) were communicated to the human subject. The subject then select the same number of grasps on the object from the pool of GPD grasps which according to them would result in salient tactile information when used on the object.

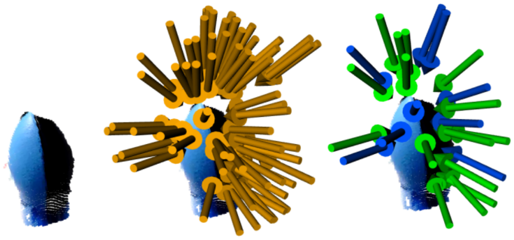

Fig. 6. Figure shows visualization of the RGB-D point cloud of football, the initial grasp pool set computed by GPD, the human baseline for grasps shown in green and our algorithm determined tactile grasps using the tactile saliency are shown in blue. Some grasps choices overlap between human baseline and algorithm selection.

Human baseline vs. algorithm generated tactile saliency: We first perform an evaluation of the tactile saliency by computing the difference between the 6D grasp poses selected by the human subject and those by our algorithm. Our intuition is that the grasp selected by the human subject are targeted to tactile salient areas on the object and thus proximity to those grasp locations on the object would provide information about tactile saliency computation for the object. Thus, our hypothesis is that the grasps selected according to the determination of saliency maps should better correspond with those selected by the human subject (in terms of 3D location and orientation).

We first perform a nearest neighbor classifier training on the human selected grasp poses, and then for each grasp computed by our algorithm on the object, we determine its nearest neighbor based on Euclidean distance ranking and we determine both a positional error (distance in 3D Cartesian space between grasp locations) and an orientation error. The orientation error is computed according to Equation 3. Table I summarizes the result over the test set of objects and results show that our approach generated grasps based on tactile saliency determination are targeted on similar areas of the objects as selected by the human subject and are closer to human selections of grasps both in terms of position and orientation.

Human baseline vs. algorithm directed acquisition of tactile sensory information: In addition to performing an evaluation of spatial similarity, we also perform an evaluation of the tactile sensory information acquired by using human subject as a baseline vs. tactile saliency determination. The selected grasp poses (human vs. algorithm) were executed by the 7-DoF Kinova robotic arm to perform prehensile probing manipulation on the test objects. Each object was palpated by the robot gripper equipped with a tactile sensor

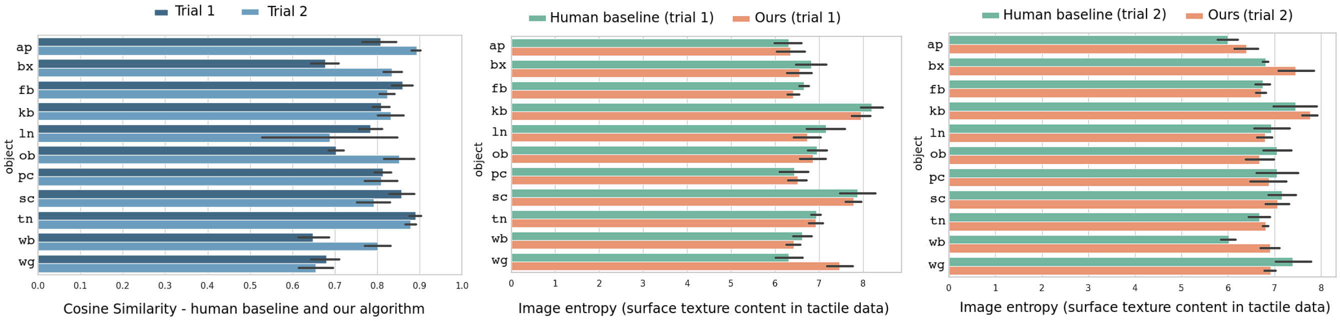

Fig. 7. *Left:* Figure shows results (mean ± S.E.) for the cosine similarity between the human baseline and algorithm for the acquired tactile sensory data in the two trials of the experiment. *Middle and Right:* Figure shows results (mean ± S.E.) for surface texture content in terms of image entropy between the human baseline and algorithm for the acquired tactile sensory data in trail 1 (middle) and trial 2 (right) of the experiment.

on the selected grasp locations to acquire tactile sensory information about the object. The calibrated pressure maps and the NURBS polynomial surface data for the sensors are recorded to the disk for each palpation. Our hypothesis is that the tactile sensory data acquired by the grasps generated by our algorithm based on tactile saliency determination would show similarity to the human baseline data. In order to process and examine the informative content and similarity of the acquired data, we perform two set of evaluations.

We first encode all the NURBS polynomial surface data as a Viewpoint Feature Histogram(VFH) [25] which is a 308-dimensional descriptor for 3D data. We compute one for each finger of the gripper, and store combination of two VFHs, as a 616-dimensional feature vector. We use the nearest neighbor classifier training on the human selected grasp poses, and for each grasp pose computed by our algorithm on the object, we determine the nearest neighbor human directed grasp based on Euclidean distance ranking, and orientation computed according to Equation 3. Next, we compute the similarity between the feature descriptors computed by the tactile data acquired by the human baseline to that acquired by our algorithm, according to the cosine similarity. The cosine similarity similarity between the tactile data feature histograms $\mathbf{f_1}$, $\mathbf{f_2} \in \mathbb{R}^{616}$ is computed as,

$$c(\mathbf{f_1}, \mathbf{f_2}) = cos(\theta) = \frac{\langle \mathbf{f_2} \cdot \mathbf{f_2} \rangle}{\|\mathbf{f_1}\| \|\mathbf{f_2}\|} \tag{4}$$

where $\langle \cdot \rangle$ defines the inner product of f_1, f_2, θ denotes the angle they form and $\| \cdot \|$ denotes the Euclidean norm. Note that the cosine similarity takes values in the range $[-1, 1]$. Figure 7 (*left*) shows the cosine similarity for the two sets of test objects. Results shows high values (> 0.5) of cosine similarity for our algorithm directed tactile sensory data on the test objects (trial 1 mean: 0.76 ± 0.011, trial 2 mean: 0.81 ± 0.012). Cosine similarity between human directed and algorithm directed tactile sensory data acquisition indicates that the computed tactile saliency maps on the objects resulted in informative tactile sensory data.

Since the recorded tactile imprints from the tactile gripper can be represented as images, we propose an additional comparison between algorithm selected grasps and human selected grasps. Our hypothesis is that for tactile salient grasps, the tactile imprints should record surface details. Although we do not have access to real ground truth data, we rely on the assumption that human subjects will pick grasps which are most likely to provide informative tactile data. We thus aim to verify whether the grasps selected by our proposed algorithm capture the same amount of surface detail. There are different ways to analyze the image content in terms of the amount of variation, or texture, that is present. We choose to compute image entropy from grey level co-occurrences matrices (GLCM) [26] determined for the image representing the tactile imprint for texture information. We simply convert the acquired tactile pressure values to a grey-scale image and GLCM (CM) is then computed according to:

$$CM_{\Delta x, \Delta y}(i, j) = \sum_{x=1}^{n} \sum_{y=1}^{m} \begin{cases} 1, & \text{if } I(x, y) = i, \text{ and } I(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Here, i and j represent the possible image intensity values for the $n \times m$ image I. In our case the offsets are determined from a distance d and angle θ. The entropy can then be calculated as:

$$E = -\sum_{i=1}^{c} \sum_{j=1}^{c} CM(i, j) \log_2 CM(i, j). \tag{6}$$

We have tried different distances and angles, and they gave mostly similar results. We report the case for $d = 1$ and $\theta = 0$. Fig. 7 *middle* and Fig. 7 *right* show the results. We can see that the entropies for the algorithm directed grasps are similar to the entropies for the human directed grasps for most objects in the test set for both trials, indicating that a similar amount of detail is captured by our algorithm selected grasps based on the computation of tactile saliency.

V. DISCUSSION AND FUTURE WORK

We have introduced tactile saliency computation with visual perception on novel objects for robots to intelligently acquire informative tactile sensory data. One assumption we have made is that the local neighborhood of the object's point cloud can provide important discriminating information for tactile saliency. Our intuition was to first segment and label the local geometry of the object, which then would provide an understanding for computing tactile saliency. An advantage of our method is that it works without any existing training data on real-world depth data for many types of shapes as shown in the experiments.

In the absence of ground truth or a comparative approach, tactile saliency selection by a human subject provides a baseline in our pilot experiments. We demonstrate the effectiveness of our approach with an evaluation of proximity to human-selected grasp poses for probing salient areas of objects and we also analysed and evaluated the acquired tactile sensory data. Results indicate that our algorithm successfully computed and utilized the tactile saliency maps on a variety of objects, and performance is comparable to the human baseline. We observed that performance may vary based on the pose in which the object is presented as shown in the two trials. Although we performed prehensile probing, the computed tactile saliency information can also be useful for other types of object interactions.

Although we achieve promising results, we would like to further explore a learning based framework and experiment with large network architectures for further improving our approach. In future work, we will recruit a higher number of human subjects to provide the baseline for experiments and explore the possibility of manually annotated 3D point clouds for tactile saliency ground truth information. We will also explore ways to indicate confidence on the computed saliency maps and handle uncertainty when applied in the real world. Finally, the computed saliency maps can also provide an understanding of functional grasping for the objects, which would be an interesting future direction.

VI. CONCLUSION

We have presented an algorithm for the computation of tactile saliency from visual 3D perception with the aim to guide the acquisition of informative tactile sensory data by robotic manipulation. Our method has several advantages as it does not require known models of the objects and is independent of any object recognition or training phase. Results from our experiments have shown that the approach is capable of computing tactile saliency that is similar to human-selection as baseline for a wide range of unseen household objects, and the computed saliency was also used to perform autonomous prehensile probing with 6D grasps by a robotic arm to intelligently acquire informative tactile sensory data.

REFERENCES

[1] Y. Xu, S. O'Keefe, S. Suzuki, and S. L. Franconeri, "Visual influence on haptic torque perception," *Perception*, vol. 41, no. 7, pp. 862–870, 2012.

[2] R. Corcodel, S. Jain, and J. van Baar, "Interactive tactile perception for classification of novel object instances," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020.

[3] R. Cong, J. Lei, H. Fu, M.-M. Cheng, W. Lin, and Q. Huang, "Review of visual saliency detection with comprehensive information," *IEEE Transactions on circuits and Systems for Video Technology*, vol. 29, no. 10, pp. 2941–2959, 2018.

[4] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley, "Image saliency: From intrinsic to extrinsic context," in *CVPR 2011*, 2011, pp. 417–424.

[5] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 10, pp. 1915–1926, 2011.

[6] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1265–1274.

[7] C. H. Lee, A. Varshney, and D. W. Jacobs, "Mesh saliency," in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 659–666.

[8] K. Mo, L. Guibas, M. Mukadam, A. Gupta, and S. Tulsiani, "Where2act: From pixels to actions for articulated 3d objects," *arXiv preprint arXiv:2101.02692*, 2021.

[9] S. Jain and B. Argall, "Grasp detection for assistive robotic manipulation," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 2015–2021.

[10] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt, "Grasp pose detection in point clouds," *The International Journal of Robotics Research*, vol. 36, no. 13-14, pp. 1455–1473, 2017.

[11] G. Rouhafzay and A.-M. Cretu, "Object recognition from haptic glance at visually salient locations," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 672–682, 2019.

[12] J. Lin, R. Calandra, and S. Levine, "Learning to identify object instances by touch: Tactile recognition via multimodal matching," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019.

[13] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in neurorobotics*, vol. 6, p. 4, 2012.

[14] 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*, 2013, pp. 3180–3186.

[15] W. Yuan, C. Zhu, A. Owens, M. A. Srinivasan, and E. H. Adelson, "Shape-independent hardness estimation using deep learning and a gelsight tactile sensor," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[16] R. Calandra, A. Owens, M. Upadhyaya, W. Yuan, J. Lin, E. H. Adelson, and S. Levine, "The feeling of success: Does touch sensing help predict grasp outcomes?" *arXiv preprint arXiv:1710.05512*, 2017.

[17] M. Lau, K. Dev, W. Shi, J. Dorsey, and H. Rushmeier, "Tactile mesh saliency," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–11, 2016.

[18] J. Jiao, Y. Cao, M. Lau, and R. Lau, "Tactile sketch saliency," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3072–3080.

[19] T. Rabbani, F. Van Den Heuvel, and G. Vosselmann, "Segmentation of point clouds using smoothness constraint," *International archives of photogrammetry, remote sensing and spatial information sciences*, vol. 36, no. 5, pp. 248–253, 2006.

[20] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel cloud connectivity segmentation-supervoxels for point clouds," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2027–2034.

[21] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (fpfh) for 3d registration," in *Proceedings of the IEEE international conference on robotics and automation (ICRA)*, 2009, pp. 3212–3217.

[22] Kinova, "Gen3 Lightweight 7-DOF robot arm," https://www.kinovarobotics.com/en/products/gen3-robot, 2021, [Online; accessed 15-Mar-2021].

[23] PPS, "Pressure Profile Systems," https://pressureprofile.com/, 2021, [Online; accessed 15-Mar-2021].

[24] L. Piegl and W. Tiller, *The NURBS book*. Springer Science & Business Media, 1996.

[25] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *Proceedings of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, October 2010.

[26] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.