

Design of an Object Scanning System and a Calibration Method for a Fingertip-Mounted Dual-Modal and Dual Sensing Mechanisms (DMDSM)-based Pretouch Sensor for Grasping

Di Wang, Fengzhi Guo, Cheng Fang, Jun Zou, and Dezhen Song

Abstract—To enable robots to grasp unknown objects, we have developed a new type of fingertip-mounted sensor that can detect distance, material type, and interior structure without making contact with the object to be grasped. Due to its working principle, the sensor is named as Dual-Modal and Dual Sensing Mechanisms (DMDSM) pretouch sensor. To enable the wide deployment of the DMDSM sensor, we need to scan a large number of common household items using the sensor to establish an object/material database. Here we report our progress in designing an automatic object scanning system and the sensor calibration algorithm with the new sensor. The object scanning system is constructed by a refitted 3D printer with a motorized turntable mounted on its printing stage. The extruder of the 3D printer is replaced by the sensor to perform 3D translation. The turntable rotates the object of interest to allow a full-body scan. A prototype of the scanning system has been built, and a new calibration algorithm has been developed to estimate the parameters of both the sensor and the scanning system. The system design and ranging accuracy have been verified by physical experiments, and the collected data from seven types of common household objects have shown promising prospects of using DMDSM sensors in grasping.

I. INTRODUCTION

To enable robots to grasp unknown objects is of great importance in the field of robotics [1], [2]. It is considered a grand challenge because of unknown factors such as object shape, surface friction, material type, impact/contact characteristics, etc. For unknown objects, detecting object shape/interior structure prior to grasping contact is necessary for devising a successful grasping plan. In our previous works [3]–[6], we have developed a new generation of fingertip-mounted dual-modal and dual sensing mechanisms (DMDSM)-based pretouch sensor for object ranging and material/structure sensing based on pulse-echo ultrasound (US) and optoacoustics (OA) (Fig. 1(a)). The new sensor can provide information about shape, material type, and interior structure about the unknown object of interest.

To enable the successful deployment of this new sensor, we need to scan object signatures for common household items. The scanning will form a database that can be used to assist real-time perception and recognition. The collected

D. Wang, F. Guo, and D. Song are with CSE Department, Texas A&M University, College Station, TX 77843, USA, Email: dzsong@cs.tamu.edu. F. Guo and D. Wang are co-first authors of this paper.

C. Fang and J. Zou are with ECE Department, Texas A&M University, College Station, TX 77843, USA, Email: junzou@tamu.edu.

This work was supported in part by National Science Foundation under IIS-2119549 and NRI-1925037, by Amazon Research Award, and by GM/SAE AutoDrive Challenge.

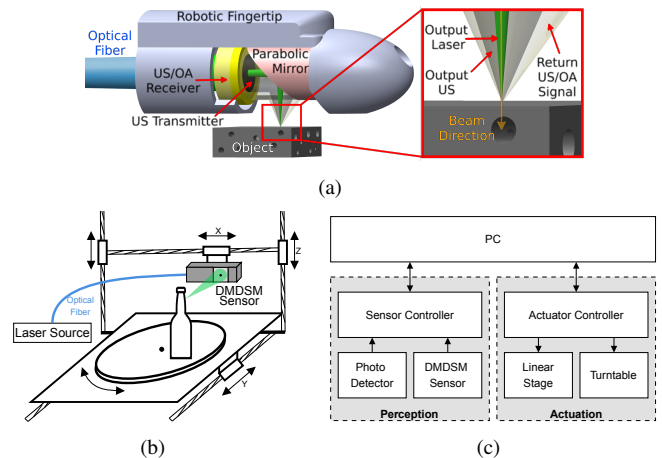


Fig. 1. Schematics of (a) the DMDSM sensor design (US: pulse-echo ultrasound; OA: optoacoustics). The hardware design (b) and system diagram (c) of the object scanning system.

database can provide both training and testing datasets. Building such a database is a large-scale data collection process which demands automatic object scanning. Therefore, we design an object scanning system, which is the main focus of this paper.

We build our scanning system by refitting a 3-dimension (3D) printer. We mount a motorized turntable on the 3D stages of the 3D printer. We present the system hardware design (Fig. 1(b)), electronic design (Fig. 1(c)). We develop a new calibration algorithm that considers characteristics of the new sensor to enable the automatic object scanning system and future sensor deployment. We have constructed a prototype system and the experimental results show that our system and algorithm design is successful. We are able to automatically scan and analyze the collected data for common household items with a size profile of less than $27 \times 27 \times 43$ cm³. The experimental results show the calibrated object scanning system achieves a contour reconstruction accuracy of 0.06 mm with a standard deviation of 0.06 mm when using the OA modality. The material recognition classifies 7 common household objects with over 98.5% accuracy for objects with artificial materials and 79.0% accuracy for objects with organic material.

II. RELATED WORK

The main purpose of our work is to design an object scanning system and a calibration algorithm for creating a

material database to facilitate perception algorithm development and the deployment of the new DMDSM sensor [3]–[6]. The related works include existing efforts in grasping databases and calibration algorithms.

The idea of building a material database for object grasping has been inspired by many existing efforts on grasping datasets in vision and tactile sensing domains. Image-based material datasets include [7]–[9], where Sharan et al.’s dataset [7] has 10 categories and contains 100 images of each category. Bell et al.’s database [9] has 23 material categories and over 20,000 images. Texture and reflectance datasets like [10], [11] have been built by taking photos of specific samples. Dana et al.’s dataset [10] contains images of 61 material samples under over 200 different lighting and viewing conditions. These image-based datasets that have been built are designed to use existing sensors such as regular or RGB-D cameras. While the calibration and scanning are well understood for cameras, we need to develop new algorithms and systems to enable an object scanning system with the new DMDSM sensor.

Beyond camera images, existing datasets also include depth images [12], [13] or 4-Dimension (4D) Light-Fields [14] to help material classification. Tactile material classification datasets vary because of the difference in tactile sensors. Murali et al.’s dataset [15] is a grasping force dataset with 7800 samples of 52 objects. Besides contact force, Erickson et al.’s dataset [16] also contains temperature measurements. Culbertson et al.’s dataset [17] and Strese et al.’s dataset [18] both contain sample texture images as well as accelerometer and contact force sensor readings. In [19], texture images, sliding acceleration, contact sound, and force are recorded on 69 surfaces. Erickson et al.’s datasets [20], [21] are the only non-contact material classification datasets which [20] contain 5 categories. Each category has 10 different materials and readings from 2 spectroscopes with different wavelengths recorded, [21] has both texture images and spectroscope readings of 144 objects from 8 material categories. The YCB set [22] is a set of household objects for the grasping evaluation purpose. Due to a different sensor design, our scanning system will generate different data types. However, the existing efforts in grasping datasets shed light on how to build such databases.

One significant component of the object scanning system is calibration, by considering the characteristics of the new DMDSM sensor. Calibration is a fundamental task in developing and maintaining a robotic system [23], [24], which has two common types: sensor calibration and robot mechanism calibration. Sensor calibration focuses on estimating the parameters of a sensing model of a sensor such as RADAR [25], [26], camera [27], IMU [28] and LIDAR [29], [30]. Robot mechanism calibration focuses on estimating robot kinematic or inertial parameters from actuator inputs and sensor measurements [31], [32]. Hand-eye calibration is a subtopic of robot mechanism calibration that aims at estimating the transformation between a robot end-effector and a mounted-on camera [33], [34]. The calibration of our scanning system is similar to a hand-eye calibration

but includes different sensor measurement characteristics. We focus on estimating the sensor parameters between our DMDSM sensor and the actuator frame by designing a new calibration rig, a calibration procedure, and an algorithm.

III. SYSTEM DESIGN

We review our DMDSM sensor before introducing the overall scanning system design.

1) *DMDSM Sensor Review:* As shown in Fig. 1(a), the new dual-modal and dual sensing mechanisms (DMDSM)-based pretouch sensor [6] utilizes both US and OA modalities to detect object distance, material type, and interior structure. The sensor transmits co-centered and co-registered ultrasound and laser pulses from its side window onto the target, and detects the reflected US and induced OA signals.

OA effect refers to the phenomenon that when a short laser pulse is incident on the target surface, part of its energy is absorbed and converted into a heat pulse to create a fast transient temperature rising, thermal expansion, and contraction, which induces the ultrasound generation. It is worth noting that the OA effect is used in both sensing modalities, and one laser pulse triggers the collection of both US and OA signals. For the US modality, the pulsed laser beams drive the ultrasound transmitter inside the sensor to send both low- and high-frequency (wideband) ultrasound pulses to the target, and the ultrasound echoes are received by the ring-shaped transducer. For OA modality, the laser pulses directly pass through the center hole of the ultrasound transmitter and are directly incident on the target surface, and the induced wideband OA signals are also collected by the same ring-shaped transducer. In both modalities, the signal temporal delay is used to determine the target-sensor distance for ranging, while their frequency spectra are used to extract distinctive features for classifying the target materials or interior structures. Details of the DMDSM sensor design can be found in our previous paper [6].

Using the OA effect in both modalities allows generating acoustic waves with wide bandwidth, which is important for material type and interior structure recognition. One can view the two modalities as indirect OA vs. direct OA. Direct OA tends to better contain the target information, however, for the transparent or highly-reflective targets, direct OA can fail. That is the reason why indirect OA is also included in the sensor design.

2) *Scanning System Design:* Now let us introduce the hardware and electronic design of our object scanning system.

The hardware design of the object scanning system is shown in Fig. 1(b), it has three main mechanical components: a DMDSM sensor, a motorized turntable (TBVECHI™ HT03RA100), and a motorized three-axis linear stage which is from an existing 3D printer (Anycubic™ Chiron). The sensor is mounted on the X -axis linear rail of the linear stage, and the turntable is mounted on the Y axis linear rail. The relative pose between the sensor and target object placed on the turntable is adjusted by actuating the linear stages

and turntable, which enables the sensor to scan the vertical surface of the object.

The electronic design of the object scanning system is shown in Fig. 1(c), it has two main parts controlled by a PC through the serial port. The perception part is driven by a sensor controller (STM32™ NUCLEO-H743ZI) with internal ADC and triggered by the photo detector upon the firing of a laser pulse to sample the pre-amplified signal from the sensor. The actuation part is driven by an actuator controller (Atmel™ ATmega2560), which interprets control commands from PC to driving signals of the linear stage and turntable.

IV. SENSOR AND SCANNING SYSTEM CALIBRATION

Just like any other sensors, deploying a DMDSM sensor requires a calibration process. This step is not only for building an object scanning system but also necessary for future deployment of the sensor to other robotic grasping applications.

A. Calibration Parameters

To mount DMDSM in our object scanning system or any other robotic end-effector for grasping, we need to calibrate its installation parameter. Ultimately, our DMDSM sensor is a 1-dimensional (1D) scanning sensor. The key sensor parameter is the scanning beam vector in the frame of reference. In our object scanning system, it is the coordinate of our 3D linear stages. In a general grasping application, the frame of reference can be its world frame. The scanning beam direction vector $\mathbf{v}$ is the direction of output laser and ultrasound beams (marked with yellow arrows in Fig. 2(a)). Here we assume the output laser beam and ultrasound signal are co-directional and co-axial due to our sensor design. The two modalities share the same scanning beam vector. Therefore, we only need to estimate the laser beam direction due to the fact that it is better focused than that of the ultrasonic beam. In fact, each beam is a focused beam reflected by the same parabolic mirror as shown in Fig. 1(a).

The rest of the parameters pertain to the scanning system itself, which is the pose of the turntable in the 3D coordinate system. The pose is uniquely defined by the rotation axis vector (i.e. normal of turntable plane surface) and its center location which is the yellow vector $\mathbf{n}$ and its position X_R as shown in Fig. 2(b). Before detailing calibration rig design and measurements, we introduce common notations as follows.

B. Nomenclature

$\mathbb{S}^2$ is the unit 2-sphere in 3D Euclidean coordinate system, $T_q\mathbb{S}^2$ is the tangent space at point $\mathbf{q} \in \mathbb{S}^2$.

$\{0\}$ represents sensor initial frame, it is a right-handed 3D Euclidean system defined by sensor initial position. Its origin is at the intersection point of output laser beam and outer surface of the robotic fingertip. Its X -, Y - and Z -axis are parallel to X , Y and Z direction of the 3D linear stages, respectively. All variables are default in $\{0\}$.

$\mathbf{n}$ is the turntable surface normal vector, $\mathbf{n} \in \mathbb{S}^2$.

$\mathbf{v}$ is the sensor output laser beam direction vector, $\mathbf{v} \in \mathbb{S}^2$.

S is a sensor position reading from linear stage, $S \in \mathbb{R}^3$.

X is a point in 3D Euclidean space, $X \in \mathbb{R}^3$.

E is an edge in 3D Euclidean space. $E = [[\mathbf{q}]_{\times} \mathbf{m}]$, where $\mathbf{q} \in \mathbb{S}^2$ is its unit length direction vector and $\mathbf{m} \in T_q\mathbb{S}^2$ is its moment vector in Plücker coordinate [1]. $[\cdot]_{\times}$ denotes the skew-symmetric matrix.

C. Calibration Rig Design and Measurements

The calibration rig is designed to provide precise measurements for calibration. The calibration rig includes a thin straight rigid graphite filament and a 3D printed base frame. The dark color of graphite generates strong OA signal responses, and the thin filament structure provides a precise point and edge measurement during scanning. The base frame is used to hold the filament either horizontally or vertically for scanning (Fig. 2).

A raw point is perceived when the DMDSM sensor's laser beam hits the object surface, and the response signal is received by the sensor. The 3D position of a raw point is recovered from sensor depth measurements d , sensor position readings S from the linear stages, and the sensor parameter $\mathbf{v}$. The filament is either horizontally pointing toward the sensor scanning window (Fig. 2(a)) or vertically (Figs. 2(b) and 2(c)). Points from the former are called tip points while points from the latter are called edge points.

Multiple raw points are aggregated into a 2D frame scan by applying an X - Z planar linear stage motion to the sensor during scanning, where this motion is called the in-frame motion (green dash arrows in Figs. 2(a) and 2(b)). The bottom-right black boxes in Figs. 2(a) and 2(b) show the response signal amplitude heatmap of the perceived points in a frame when the filament is held horizontally or vertically, respectively.

With a frame scan of the filament, the center point(s) in a frame is determined based on the thresholded response signal of the raw points. The center point is the rudimentary element used in our calibration. Subscript i is used to note frame index, and j is used to note center point index, point index is omitted if there is only one center point in a frame. For example, center point X_{ij} denotes the j -th center point scanned in the i -th frame. For a tip scan, the tip center point (red dot in the bottom-right black box of Fig. 2(a)) is the average of all thresholded tip points scanned with the X - Z planar motion. For an edge scan, the edge center points (red dots in the bottom-right black box of Fig. 2(b)) is the average of thresholded edge points scanned with X -axis motion. For a center point X_{ij} calculated from n raw points, by the central limit theorem, the noise in its corresponding sensor position reading S_{ij} from the linear stage follows a Gaussian distribution $\mathcal{N}(0, \Sigma_{ij})$, where $\Sigma_{ij} = \frac{1}{n} \Sigma_{Pij}$ is the covariance matrix of the noise, and Σ_{Pij} is the covariance matrix of the raw points' sensor position readings. The noise in the sensor depth measurement d_{ij} of a center point follows a similar derivation and has a variance $\sigma_{ij} = \frac{1}{n} \sigma_{Pij}$ where σ_{Pij} is the variance of the raw points' sensor depth readings.

Multiple frames are generated by moving the sensor and rig with linear stage and turntable to obtain frame scans of the rig from different depth settings or perspectives, where this motion is called the between-frame motion (black solid arrows in Fig. 2). In the between-frame motion, the rotation motion applied between the i -th and k -th frame scan is determined by turntable rotation angle readings θ_{ik} , the turntable normal vector $\mathbf{n}$ and center X_R . The noise in turntable rotation angle readings θ_{ik} follows a Gaussian distribution $\mathcal{N}(0, \sigma_{ik})$, where σ_{ik} is the turntable precision.

D. Calibration Procedure

The calibration procedure begins with estimating the sensor parameter $\mathbf{v}$ first and then calibrating scanning system parameter $\mathbf{n}$ and X_R . As shown in Fig. 2, both steps employ in-frame motions and between-frame motions. The calibration rig is mounted on the turntable rigidly, and the turntable base is fixed on the 3D linear stages. The relative motion is generated by the 1D rotation or the 3D linear motions.

1) *Sensor Parameter Calibration Procedure:* As shown in Fig. 2(a), for sensor parameter calibration, the graphite filament is held horizontally to point its tip toward the laser beam from the sensor. Firstly, in-frame motion is applied on the sensor to obtain one frame scan of the filament tip. The tip center point X_c in the c -th frame is determined following the process mentioned in Sec. IV-C, and its corresponding depth measurements and position readings (d_c, S_c) are recorded. An example of the obtained frame is shown in the bottom-right black box of Fig. 2(a), and the tip center point is marked with a red dot. After obtaining one frame, between-frame translational motions along Y axis is applied on the sensor, followed by another round of in-frame motion to get a frame scan of the tip from a different depth. The beam direction vector $\mathbf{v}$ is estimated from at least two tip frame scans with the raw measurements of one tip center point in each frame.

2) *Scanning System Parameters Calibration Procedure:* As shown in Fig. 2(b), for scanning system parameters calibration, the graphite filament is held vertically to allow for frame scans to capture edge points along the filament. In-frame motion is applied to obtain one frame scan of the filament edge. Similar to the previous step, edge center points X_{ij} are determined from a frame and their corresponding raw measurements (d_{ij}, S_{ij}) are recorded. An example of the obtained frame scan is shown in the bottom-right black box of Fig. 2(b), and the edge center points are marked with red dots. Between-frame motions in this step include sensor translational motions and rig rotation motion with angle θ_{ik} , it is followed by another round of in-frame motion to get a frame scan of the edge from different perspectives. Raw measurements (d_{kj}, S_{kj}) of the new edge center points from the frame are recorded to pair with their pre-rotation counterparts. The turntable surface normal vector $\mathbf{n}$ and center X_R are estimated from at least two edge frame scans with the raw measurements of no less than two edge center points in each frame.

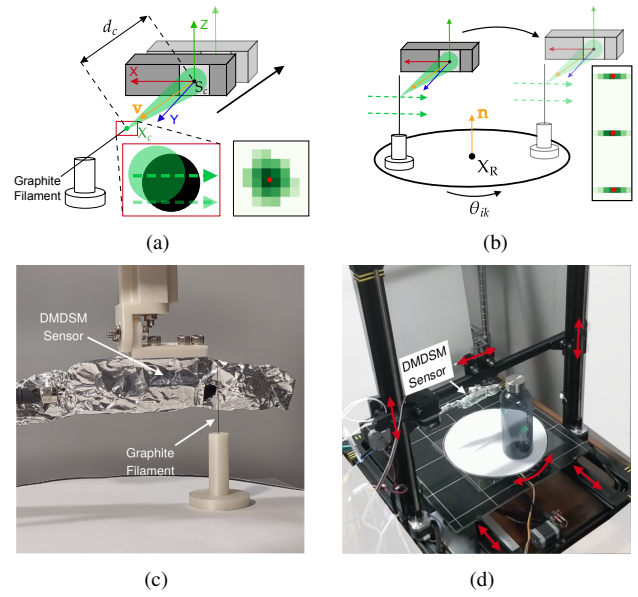


Fig. 2. Illustration of calibration procedure and resulting data (best viewed in color), in-frame motions and between-frame motions are marked with green dash arrows and black solid arrows, respectively. In (a) and (b), the gray 3D rectangle is DMDSM sensor with scanning window shooting out green focusing signal beams. Typical signal readings in one frame scan is shown in the bottom-right black box. (a) Beam direction estimation procedure and typical tip point readings in a frame. (b) Turntable center and normal estimation procedure and sample edge point readings. (c) A photo of actual scanning set up. The DMDSM sensor is wrapped in aluminum foil for electromagnetic shielding. The filament is mounted vertically. (d) The overall view of the scanning system with a water bottle on the turntable.

E. Problem Definition

The calibration of our object scanning system is a two-step process. The first step is a 2 degrees-of-freedom (DoF) sensor parameter calibration problem, followed by a 4 DoF scanning system parameters calibration problem.

Definition 1 (Sensor Parameter Calibration): Given tip center points' sensor depth measurements and position readings (d_c, S_c), estimate sensor parameter $\mathbf{v}$.

Definition 2 (Scanning System Parameters Calibration): Given sensor parameter $\mathbf{v}$, turntable rotation angle readings θ_{ik} between a pair of edge E_i and E_k , and their edge center points' sensor depth measurements and position readings (d_{ij}, S_{ij}) and (d_{kj}, S_{kj}), estimate scanning system parameters $\mathbf{n}$ and X_R .

F. Calibration Algorithm

Now the calibration algorithm is presented for our problem, it starts with estimating sensor parameter $\mathbf{v}$ and scanning system parameter $\mathbf{n}$ and X_R , then jointly optimize the estimations with a Maximum Likelihood Estimation (MLE) method.

1) *Estimate Sensor Parameter $\mathbf{v}$:* Before elaborating the beam direction estimation scheme, we first show how a center point is recovered from sensor depth measurements d_c and sensor position S_c from linear stage readings.

As shown in Fig. 2(a), during the c -th frame scan, sensor output laser beam starts from sensor position S_c and travel

distance d_c along direction vector $\mathbf{v}$ until it hits point X_c on the graphite filament. Therefore, point X_c satisfies

$$X_c = S_c + d_c \mathbf{v} \quad (1)$$

In sensor parameter calibration, laser beam direction $\mathbf{v}$ is estimated from the tip center points in multiple frame scans (red dot in the bottom-right black box of Fig. 2(a)). Denote the set of frame indices collected in this step as $\mathcal{I}_1$. Because all frames capture the same point X_C on the filament tip from different depth as tip center points, ideally we should have $X_c = S_c + d_c \mathbf{v} = X_C$ for all frame index $c \in \mathcal{I}_1$. That means when scanning the tip, all sensor position S_c are collinear and parallel to $\mathbf{v}$. Therefore

$$(S_c - \bar{S}_c) \times \mathbf{v} = \mathbf{0}, \quad (2)$$

where $\bar{S}_c = \frac{1}{|\mathcal{I}_1|} \sum_{c \in \mathcal{I}_1} S_c$ is the averaged tip center points' corresponding sensor position, and $\times$ is the cross product. By stacking (2) for all the points S_c with $i \in \mathcal{I}_1$, the least-squares estimation of beam direction $\mathbf{v}$ is obtained using singular value decomposition (SVD).

2) Estimate Scanning System Parameters $\mathbf{n}$ and X_R :

In scanning system parameters calibration, $\mathbf{n}$ and X_R are estimated from the rotation motion using edge center points in multiple frame scans (red dots in the bottom-right black box of Fig. 2(b)). We start with estimating edges from edge center points, then elaborate the rotation motion and show how to utilize the edges to obtain $\mathbf{n}$ and X_R from the motion.

An edge is estimated from at least two center points lying on it, and edge center points are determined from the estimated $\mathbf{v}$ and raw measurements following the derivation of (2). Denote the filament scanned in the i -th frame as edge $\mathbf{E}_i$, X_{ij} is the j -th edge center point on $\mathbf{E}_i$, and the set of point indices collected in this step is denoted as $\mathcal{I}_2$. The direction vector $\mathbf{q}_i$ of $\mathbf{E}_i$ is parallel to the vector between any two points on the edge

$$(X_{ij} - \bar{X}_i) \times \mathbf{q}_i = \mathbf{0}, \quad (3)$$

where $\bar{X}_i = \frac{1}{|\mathcal{I}_2|} \sum_{j \in \mathcal{I}_2} X_{ij}$ is the averaged edge center points on $\mathbf{E}_i$. The moment vector $\mathbf{m}_i$ of $\mathbf{E}_i$ is calculated by

$$\mathbf{m}_i = \bar{X}_i \times \mathbf{q}_i \quad (4)$$

following the conventions in [1]. By stacking (3) for all the edge points X_{ij} on $\mathbf{E}_i$, the least-squares estimation of the direction vector $\mathbf{q}_i$ is obtained using singular value decomposition (SVD). The moment vector $\mathbf{m}_i$ of edge $\mathbf{E}_i$ is solved from (4), which completes the estimation of $\mathbf{E}_i = [[\mathbf{q}_i]_{\times} \mathbf{m}_i]$. Edge $\mathbf{E}_k$ is estimated in the same manner from its raw measurements (d_{kj}, S_{kj}) .

The rotation motion $\mathbf{T}_R$ between two consecutive frame scans can be decomposed into three steps. First, we translate the origin of $\{0\}$ to the turntable center. Then we apply the rotating motion. Finally, we translate the origin back. The

rotation motion $\mathbf{T}_R$ can be written as

$$\begin{aligned} \mathbf{T}_R &= \begin{bmatrix} \mathbf{I} & X_R \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_n(\theta) & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I} & -X_R \\ \mathbf{0} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}_n(\theta) & (\mathbf{I}_{3 \times 3} - \mathbf{R}_n(\theta))X_R \\ \mathbf{0} & 1 \end{bmatrix}, \end{aligned} \quad (5)$$

where $\mathbf{R}_n(\theta) = \mathbf{I}_{3 \times 3} + \sin \theta [\mathbf{n}]_{\times} + (1 - \cos \theta) [\mathbf{n}]_{\times}^2$ is the Rodrigues' formula for axis-angle rotation, $\mathbf{n}$ is the normal vector of turntable surface, and θ is the rotation angle of the motion. X_R is the center of turntable.

Using at least two edges and their underlying rotation motion, the scanning system parameters $\mathbf{n}$ and X_R can be estimated. Let $\mathbf{E}_k$ be the counterpart of edge $\mathbf{E}_i$ after applying the rotation motion $\mathbf{T}_R$. The direction vector $\mathbf{q}_i$ of edge $\mathbf{E}_i$ is parallel to its counterpart $\mathbf{q}_k$ after rotation

$$\mathbf{R}_n(\theta) \mathbf{q}_i \times \mathbf{q}_k = \mathbf{0}. \quad (6)$$

Besides, a point X_{ij} on edge $\mathbf{E}_i$ lies on the transformed edge $\mathbf{E}_k$ after rotation

$$\mathbf{E}_k(\mathbf{T}_R \mathbf{X}_i) = \mathbf{0}, \quad (7)$$

where $\mathbf{X}_i = \left[\dots, \begin{bmatrix} X_{ij} \\ 1 \end{bmatrix}, \dots \right]$ are all the edge points on $\mathbf{E}_i$. Combining (6) and (7), the turntable normal $\mathbf{n}$ and center X_R can be solved using SVD.

3) *MLE Estimation:* Given the least-squares estimations of $\mathbf{v}$, $\mathbf{n}$ and X_R obtained from previous steps, a Maximum Likelihood Estimation (MLE) problem that jointly estimates the parameters is formulated to obtain the optimal estimations under the Gaussian noise assumption.

Suppose that $\mathbf{E}_k$ is the counterpart of edge $\mathbf{E}_i$ after applying the rotation motion $\mathbf{T}_R$. The collection of raw measurements from the two frames is denoted as $\mathcal{X}_{ik} = [\theta_{ik}, \mathbf{d}_i^T, \mathbf{S}_i^T, \mathbf{d}_k^T, \mathbf{S}_k^T]^T$. Here θ_{ik} is the turntable angle reading of rotation motion between the i -th and k -th frame. $\mathbf{d}_i = [\dots, d_{ij}, \dots]^T$ and $\mathbf{S}_i = [\dots, S_{ij}, \dots]^T$ are depth and sensor position readings of edge center points on edge $\mathbf{E}_i$, $\mathbf{d}_k$ and $\mathbf{S}_k$ are those for edge $\mathbf{E}_k$.

The Maximum Likelihood Estimation (MLE) optimization problem is formulated as follows

$$\begin{aligned} \min_{\hat{\mathcal{X}}_{ik}} \sum_{i \neq k} \left\| \mathcal{X}_{ik} - \hat{\mathcal{X}}_{ik} \right\|_{\Sigma_{ik}}^2 \\ \text{s.t.} \quad \mathcal{C}(\hat{\mathcal{X}}_{ik}) = \mathbf{0}, \end{aligned} \quad (8)$$

where $\|\cdot\|_{\Sigma_{ik}}^2$ is Mahalanobis distance with covariance matrix Σ_{ik} of $\mathcal{X}_{ik}$, and $\Sigma_{ik} = \text{Diag}(\sigma_{ik}, \Sigma_{Si}, \Sigma_{di}, \Sigma_{Sk}, \Sigma_{dk})$. Covariance matrices $\Sigma_{Si} = \text{Diag}(\dots, \Sigma_{ij}, \dots)$ and $\Sigma_{di} = \text{Diag}(\dots, \sigma_{ij}, \dots)$ are obtained from Sec. IV-C, and the same apply for σ_{ik} , Σ_{Sk} and Σ_{dk} . Constraint function $\mathcal{C}(\hat{\mathcal{X}}_{ik})$ is a concatenation of (3), (6) and (7). The MLE problem is solved using the Levenberg-Marquardt algorithm.

V. EXPERIMENTS AND RESULTS

The proposed object scanning system and calibration algorithm have been validated with physical experiments. The experiment setup is shown in Fig. 2(c) and Fig. 2(d). We

first show the calibration experiment results, then present the household object scanning experiment results.

A. Calibration Experiment

1) *Calibration Experiment Data and Results:* The calibration dataset is collected by following the procedure described in Sec.IV-D. In the dataset, 4 tip center points are extracted from 70 tip point measurements, and 14 edge center points are extracted from 88 edge point measurements in two edge frame scans, between which the rotation angle is $\theta_{01} = \pi$. Using the collected dataset, calibration parameters are estimated with the algorithm mentioned in Sec.IV-F, and their values are shown in Table. I.

TABLE I
ESTIMATED CALIBRATION PARAMETERS

$\mathbf{v}$	[	0.0656;	0.9955;	-0.0678]
$\mathbf{n}$	[	-0.0007;	0.0022;	0.9999]
X_R (mm)	[	235.21;	288.17;	0.00]

2) *Calibration Result Validation:* The estimated calibration parameters and co-directional alignment of the OA and US beams are validated by reconstructing an aluminum block with a known shape. The reconstructed points are compared with the ground-truth measurements from the vernier caliper. The error metric $e = d(X_{ij}, \mathbf{E}_i)$ is defined based on the Euclidean distance function d between a reconstructed point X_{ij} and its corresponding ground-truth contour line $\mathbf{E}_i$.

The reconstruction error and standard deviation of OA and US are (0.06 ± 0.06) and (0.15 ± 0.11) mm, respectively. The reconstruction result using OA and US signals is visualized in Fig. 3.

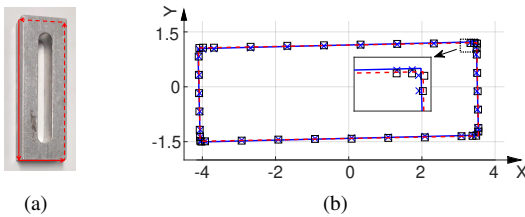


Fig. 3. Aluminum block contour reconstruction for validating the system parameters. (a) Aluminum block with scanning path. (b) Reconstructed Contour. The unit of X-Y axes is centimeter. Symbols '□' and '×' represent OA and US scanned points, respectively. The red and blue boxes represent the OA and US fitted contours, respectively.

The results show both OA and US reconstruction achieving a reconstruction accuracy level on par with the DMDSM sensor. For the depth measurements, the sensor has lower than 0.10 mm average error and deviation for the OA modality, and around 0.15 mm average error and 0.11 mm deviations for the US modality, respectively. These results verify the calibration result and indicate that the two co-directional signal beams are aligned well.

B. Household Object Scanning Experiment Results

After validating the calibration experiment result, household object scanning experiments are conducted on seven common household objects to demonstrate the shape and

material sensing capability of our scanning system. The scanning experiment includes a contour reconstruction test and a material classification test.

1) *Contour Reconstruction Test:* The contour reconstruction test aims at showing the shape sensing capability of the system. The six scanned items and their responses to OA and US signals are listed in Fig. 4.

As shown in Fig. 4, it is not uncommon that one of the OA and US signals is not perceivable on some objects. For transparent or reflective objects, the OA signal will not be presented due to lacks of light absorptive material. For porous objects, the US signal will not be presented due to sound absorptive material [6]. Nonetheless, our scanning system is still able to reconstruct the contour of the scanned items by utilizing our DMDSM sensor.

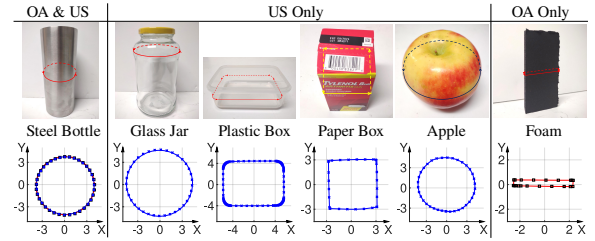


Fig. 4. Object Contour Reconstruction Results. The first row shows working modality on each object. The second row shows object photos with scanning paths. The third row illustrates their reconstructed contours. The unit of X-Y axes is the centimeter. The symbols '□' and '×' represent the OA and US scanned points, respectively. The red and blue lines represent the OA and US connected lines between adjacent points, respectively.

2) *Material Differentiation Test:* The material differentiation test aims at showing the material sensing capability of the system. Similar to our previous works [3]–[6], the BOSS classifier is used to distinguish the materials of the seven objects mentioned above using 50 random trials and a 3:1 ratio of training and testing data. Fig. 5 shows the confusion matrix of the material differentiation test, and it shows that we are able to achieve over 98.5% accuracy for objects with artificial materials and 79.5% for organic object (apple). The lower accuracy of the apple may result from the non-uniformity of acoustic impedance in organic materials and the signal discrepancy caused by the varied incidence angles of the sensor signal beams. More organic objects and objects with complex shapes will be scanned in the future to optimize the training and classification algorithm.

VI. CONCLUSION AND FUTURE WORK

In this paper, the design of an object scanning system and a calibration algorithm with the new DMDSM sensor has been demonstrated. The collected data from seven types of common household objects have established an object/material database to assist real-time perception and recognition, which shows promising prospects of using the DMDSM sensor in robotic grasping. In the future, we plan to scan more objects to optimize the training and classification algorithm, improve the scanning flexibility and speed, and establish the 3D contour/material database. After that, we will integrate

the DMDSM sensor onto a robot hand to enable real-time close-loop grasping.

True Material \ Predicted Material	Al	Apple	Foam	Glass	Paper	Plastic	Steel
Al	100.0	0.0	0.0	0.0	0.0	0.0	0.0
Apple	0.0	79.0	0.0	0.0	0.0	21.0	0.0
Foam	0.0	0.0	100.0	0.0	0.0	0.0	0.0
Glass	0.0	0.0	0.0	100.0	0.0	0.0	0.0
Paper	0.0	0.0	0.0	0.0	100.0	0.0	0.0
Plastic	0.0	1.5	0.0	0.0	0.0	98.5	0.0
Steel	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Fig. 5. The classification confusion matrix of the seven materials.

ACKNOWLEDGMENT

We are grateful to K. Goldberg, X. Duan, S. Ye, S. Xie, A. Kingery, A. Angert, C. Qian and Y. Jiang for their inputs and contributions.

REFERENCES

- [1] M. T. Mason, *Mechanics of robotic manipulation*. MIT press, 2001.
- [2] M. Ciocarlie, K. Hsiao, E. G. Jones, S. Chitta, R. B. Rusu, and I. A. Sucan, "Towards reliable grasping and manipulation in household environments," in *Experimental Robotics*. Springer, 2014, pp. 241–252.
- [3] C. Fang, D. Wang, D. Song, and J. Zou, "Toward fingertip non-contact material recognition and near-distance ranging for robotic grasping," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 4967–4974.
- [4] —, "Fingertip non-contact optoacoustic sensor for near-distance ranging and thickness differentiation for robotic grasping," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 10 894–10 899.
- [5] —, "Fingertip pulse-echo ultrasound and optoacoustic dual-modal and dual sensing mechanisms near-distance sensor for ranging and material sensing in robotic grasping," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 14 105–14 111.
- [6] —, "The second generation (g2) fingertip sensor for near-distance ranging and material sensing in robotic grasping," in *2022 IEEE International Conference on Robotics and Automation (ICRA)(Accepted)*. IEEE, 2022.
- [7] L. Sharan, R. Rosenholtz, and E. Adelson, "Material perception: What can you see in a brief glance?" *Journal of Vision - J VISION*, vol. 9, pp. 784–784, 8 2010.
- [8] D. Hu, L. Bo, and X. Ren, "Toward robust material recognition for everyday objects," in *British Machine Vision Conference, BMVC 2011, Dundee, UK, August 29 - September 2, 2011. Proceedings*, 2011, pp. 1–11.
- [9] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 3479–3487.
- [10] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, 1999.
- [11] E. Hayman, B. Caputo, M. Fritz, and J. Eklundh, "On the significance of real-world conditions for material classification," in *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV*, 2004, pp. 253–266.
- [12] J. DeGol, M. G. Fard, and D. Hoiem, "Geometry-informed material recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 1554–1562.
- [13] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. B. Hullin, and W. Heidrich, "Material classification using raw time-of-flight measurements," in *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2016, pp. 3503–3511.
- [14] T. Wang, J. Zhu, H. Ebi, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4d light-field dataset and CNN architectures for material recognition," in *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III*, 2016, pp. 121–138.
- [15] A. Murali, Y. Li, D. Gandhi, and A. Gupta, "Learning to grasp without seeing," *CoRR*, vol. abs/1805.04201, 2018.
- [16] Z. M. Erickson, S. Chernova, and C. C. Kemp, "Semi-supervised haptic material recognition for robots using generative adversarial networks," in *1st Annual Conference on Robot Learning, CoRL 2017, Mountain View, California, USA, November 13-15, 2017, Proceedings*, 2017, pp. 157–166.
- [17] H. Culbertson, J. J. L. Delgado, and K. J. Kuchenbecker, "One hundred data-driven haptic texture models and open-source methods for rendering on 3d objects," in *IEEE Haptics Symposium, HAPTICS 2014, Houston, TX, USA, February 23-26, 2014*, 2014, pp. 319–325.
- [18] M. Strese, J. Lee, C. Schuwerk, Q. Han, H. Kim, and E. G. Steinbach, "A haptic texture database for tool-mediated texture recognition and classification," in *2014 IEEE International Symposium on Haptic, Audio and Visual Environments and Games, HAVE 2014, Richardson, TX, USA, October 10-11, 2014*, 2014, pp. 118–123.
- [19] M. Strese, C. Schuwerk, A. Iepure, and E. G. Steinbach, "Multimodal feature-based surface material classification," *IEEE Trans. Haptics*, vol. 10, no. 2, pp. 226–239, 2017.
- [20] Z. Erickson, N. Luskey, S. Chernova, and C. C. Kemp, "Classification of household materials via spectroscopy," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 700–707, 2019.
- [21] Z. Erickson, E. Xing, B. Srirangam, S. Chernova, and C. C. Kemp, "Multimodal material classification for robots using spectroscopy and high resolution texture imaging," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 10 452–10 459.
- [22] B. Çalli, A. Singh, A. Walsman, S. S. Srinivasa, P. Abbeel, and A. M. Dollar, "The YCB object and model set: Towards common benchmarks for manipulation research," in *International Conference on Advanced Robotics, ICAR 2015, Istanbul, Turkey, July 27-31, 2015*, 2015, pp. 510–517.
- [23] Z. Roth, B. Mooring, and B. Ravani, "An overview of robot calibration," *IEEE Journal on Robotics and Automation*, vol. 3, no. 5, pp. 377–385, 1987.
- [24] T. Verma and N. R. Chauhan, "A critical review on calibration of robots," in *Advances in Interdisciplinary Engineering*, M. Kumar, R. K. Pandey, and V. Kumar, Eds. Singapore: Springer Singapore, 2019, pp. 677–683.
- [25] A. Freeman, "Sar calibration: an overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 30, no. 6, pp. 1107–1121, 1992.
- [26] C. Chou, H. Li, and D. Song, "Encoder-camera-ground penetrating radar sensor fusion: Bimodal calibration and subsurface mapping," *IEEE Transactions on Robotics*, vol. 37, no. 1, pp. 67–81, 2021.
- [27] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [28] Y. Yang, P. Geneva, X. Zuo, and G. Huang, "Online imu intrinsic calibration: Is it necessary?" *Robotics: Science and Systems XVI*, vol. 16, 2020.
- [29] C. L. Glennie and D. D. Lichti, "Static calibration and analysis of the velodyne hdl-64e s2 for high accuracy mobile scanning," *Remote. Sens.*, vol. 2, no. 6, pp. 1610–1624, 2010.
- [30] J. K. Huang, C. Feng, M. Achar, M. G. Jadidi, and J. W. Grizzle, "Global unifying intrinsic calibration for spinning and solid-state lidars," *CoRR*, vol. abs/2012.03321, 2020.
- [31] B. Mooring, M. Driels, and Z. Roth, *Fundamentals of Manipulator Calibration*. John Wiley & Sons, Inc., 1991.
- [32] J. Hollerbach, W. Khalil, and M. Gautier, *Model Identification*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 321–344.
- [33] R. Horaud and F. Dornaika, "Hand-eye calibration," *The International Journal of Robotics Research*, vol. 14, no. 3, pp. 195–210, 1995.
- [34] R. Tsai and R. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, 1989.