

# The Sum of Its Parts: Visual Part Segmentation for Inertial Parameter Identification of Manipulated Objects

Philippe Nadeau, Matthew Giamou, and Jonathan Kelly<sup>‡</sup>

**Abstract**—To operate safely and efficiently alongside human workers, collaborative robots (cobots) require the ability to quickly understand the dynamics of manipulated objects. However, traditional methods for estimating the full set of inertial parameters rely on motions that are necessarily fast and unsafe (to achieve a sufficient signal-to-noise ratio). In this work, we take an alternative approach: by combining visual and force-torque measurements, we develop an inertial parameter identification algorithm that requires slow or “stop-and-go” motions only, and hence is ideally tailored for use around humans. Our technique, called Homogeneous Part Segmentation (HPS), leverages the observation that man-made objects are often composed of distinct, homogeneous parts. We combine a surface-based point clustering method with a volumetric shape segmentation algorithm to quickly produce a part-level segmentation of a manipulated object; the segmented representation is then used by HPS to accurately estimate the object’s inertial parameters. To benchmark our algorithm, we create and utilize a novel dataset consisting of realistic meshes, segmented point clouds, and inertial parameters for 20 common workshop tools. Finally, we demonstrate the real-world performance and accuracy of HPS by performing an intricate ‘hammer balancing act’ autonomously and online with a low-cost collaborative robotic arm. Our code and dataset are open source and freely available.

## I. INTRODUCTION

In order to operate effectively and safely, robots require an understanding of the dynamics of their interactions with the environment. For collaborative robots (cobots) that are designed to work alongside people, the ability to infer the inertial parameters of manipulated objects is particularly important. An object’s inertial parameters include its mass, centre of mass (COM), and moments of inertia, all of which are used to perform dynamic manipulation [1]. Force-torque sensors (FT) located in the joints or at the wrist of a robot arm can be used to determine the inertial parameters of a manipulated object. However, sensor noise limits parameter identification accuracy, especially for cobots that must (should) move within the safe velocity limits defined by ISO standards 10218 [2] and 15066 [3]. These slow motions lead to lower signal-to-noise ratios in force-torque data, prohibiting the use of inertial parameter estimation techniques appropriate for large, fast-moving industrial robots [4]. In this paper, we propose an alternative identification method that relaxes the need for accurate kinematics estimates but

All authors are with the STARS Laboratory at the University of Toronto Institute for Aerospace Studies, Toronto, Ontario, Canada. <firstname>.<lastname>@robotics.utoronto.ca

<sup>‡</sup>Jonathan Kelly is a Vector Institute Faculty Affiliate. This research was supported in part by the Canada Research Chairs program.

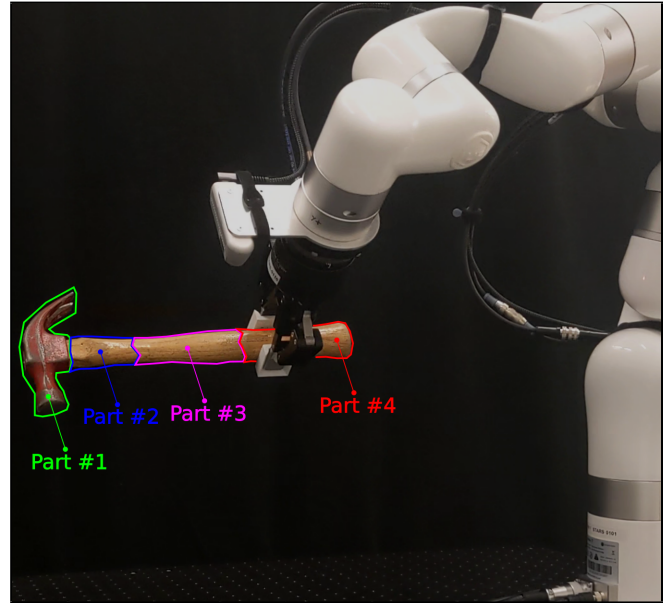


Fig. 1: The shape of a manipulated object is segmented into parts that are assumed to be homogeneous. By identifying the mass of each part, the full set of inertial parameters can be estimated from safer “stop-and-go” motions.

can nonetheless determine the complete set of inertial parameters, under some mild assumptions.

Humans are able to easily infer inertial properties through *dynamic touch* [5] and cobots would benefit from a similar capability [6], [7]. Additionally, humans combine proprioception with vision, suggesting that complementary perceptual modalities can speed up the identification process. Our technique uses information obtained from an RGB-D camera to improve the speed and accuracy of inertial parameter identification. We make the key observation that man-made objects are usually built from a few parts having approximately homogeneous densities.

**Assumption 1** (Homogeneous Density of Parts). *An object is made of  $N$  parts  $\{r_1, \dots, r_N\}$ , each with a shape  $V_j \subset \mathbb{R}^3$ . The mass density is such that  $\forall \mathbf{p} \in V_j, \rho(\mathbf{p}) = \rho_j$  where  $\rho_j > 0$  is part  $r_j$ ’s constant density.*

Critically, Assumption 1 can be used to simplify the identification problem by exploiting measurements of the shape of a manipulated object. This is accomplished by noting that, once a cobot has determined the shape of an object, the object’s inertial parameters depend solely on the mass of each of its homogeneous parts. We refer to our overall

approach as *Homogeneous Part Segmentation* (HPS).<sup>1</sup> Our main contributions are:

- a formulation of inertial parameter identification that incorporates Assumption 1;
- a method combining the algorithms proposed in [8] and [9] to improve part segmentation speed;
- a dataset of models of 20 common workshop tools with 3D meshes, point clouds, ground truth inertial parameters, and ground truth part-level segmentation for each object; and
- experiments highlighting the benefits of HPS when compared to two benchmark algorithms.

We undertake a series of simulation studies with our novel dataset and carry out a real-world experiment to assess the performance of the proposed algorithm on a real cobot. We show that, under noisy conditions, our approach produces more accurate inertial parameter estimates than competing algorithms that do not utilize shape information.

## II. RELATED WORK

This section provides a brief overview of work related to the two main algorithmic components of our system: inertial parameter or load identification and part-level object segmentation. A more thorough survey of inertial parameter identification for manipulation is provided by Golluccio et al. in [10].

### A. Load Identification

In [11], a least squares method is applied to determine the inertial parameters of a load manipulated by a robot arm. The authors of [11] underline that poor performance can be caused by the low signal-to-noise ratios in the data used for the regression. A recursive total least squares (RTLS) approach is proposed in [12] to account for noise in the regressor matrix. However, an experimental evaluation in [13] of RTLS on multiple datasets reports large estimation errors, demonstrating that RTLS has limited utility with noisy sensors. In this work, we also minimize a least squares cost, but we employ a unique formulation that is well-suited to data gathered by a slower-moving cobot, for example.

Without appropriate constraints on regression variables, many identification algorithms find unphysical solutions [14], [15]. Sufficient conditions for the physical consistency of the identified parameters are stated in [15], where a constrained optimization problem is formulated that guarantees the validity of the solution. Similarly, physical consistency is enforced through linear matrix inequalities as part of the method proposed in [16]. Geodesic distance approximations from a prior solution are used in [17] to regularize the optimization problem without introducing very long runtimes. In this work, we *implicitly* enforce the physical consistency of estimated inertial parameters by discretizing the manipulated

object with point masses [18], [4]. Fixed-sized voxels can also be used to achieve the same effect [19].

Finally, the authors of [20] augment the traditional force-torque sensing system with tactile sensors to estimate the inertial parameters and the friction coefficient of a manipulated object. Our contribution similarly makes use of an additional sensing modality in the form of a camera.

### B. Part-Level Object Segmentation

Although there is no formal definition of a “part” of a manipulated item, humans tend to follow the so-called *minima rule*: the decomposition of objects into approximately convex contiguous parts bounded by negative curvature minima [21]. This rule has provided inspiration for many part segmentation methods reviewed in [22], which benchmarks several techniques on the Princeton dataset [23] and divides approaches into surface-based, volume-based, and skeleton-based algorithms. Surface-based and volume-based mesh segmentation algorithms are reviewed in [24], highlighting a tradeoff between segmentation quality and the number of parts produced by the algorithm. Skeleton-based segmentation methods for 3D shapes, which capture structural information as a lower-dimensional graph (i.e., the shape’s *skeleton*), are reviewed in [25]. The approach in [26], which is based on the medial axis transform [27], exploits prior knowledge of an object’s skeleton to produce a *top-down* segmentation about 10 times faster than *bottom-up* methods that use local information only.

The method of Kaick et al. [28] also segments incomplete point clouds into approximately convex shapes, but is prohibitively slow for manipulation tasks, with an average computation time of 127 seconds on the Princeton benchmark [23] as reported in [26]. In contrast, learning-based methods for part segmentation can struggle to generalize to out-of-distribution shapes but are usually faster than geometric techniques [29], [30], [22]. Our approach to segmentation does not apply any learned components but is fast enough for real-time use by a collaborative robot.

A hierarchical volume-based segmentation technique (HTC) proposed in [8], enabled by fast tetrahedralization [31] and quick convex hull computation [32], can perform well if a watertight mesh can be reconstructed from the input point cloud. This technique, which we describe in detail in Section IV, is an important element of our part segmentation pipeline. Another key component of our approach is the surface-based algorithm proposed in [9], which uses a dissimilarity metric to iteratively merge nearby mesh patches, taking special care to preserve part boundaries.

## III. INERTIAL PARAMETERS OF HOMOGENEOUS PARTS

In this section, we describe our inertial parameter identification technique, which assumes that an object has been segmented into its constituent parts. By determining the mass of each segment, we are able to identify the full set of inertial parameters, or to provide an approximate solution when Assumption 1 is not respected.

<sup>1</sup>We provide code, our complete simulation dataset, and a video showcasing our algorithm at <https://papers.starslab.ca/part-segmentation-for-inertial-identification/>.

### A. Notation

Reference frames  $\mathcal{F}_b$  and  $\mathcal{F}_s$  are attached to the object and to the force-torque (FT) sensor, respectively. The reference frame  $\mathcal{F}_w$  is fixed to the base of the robot and is assumed to be an inertial frame, such that the gravity vector expressed in the sensor frame is given by  $\mathbf{g}_s = {}^s\mathbf{R}^w[0, 0, -9.81]^\top$ . The orientation of  $\mathcal{F}_b$  relative to  $\mathcal{F}_s$  is given by  ${}^s\mathbf{R}^b$  and the origin of  $\mathcal{F}_b$  relative to  $\mathcal{F}_s$ , expressed in  $\mathcal{F}_w$ , is given by  ${}^s\mathbf{t}_w^b$ . The skew-symmetric operator  $[\cdot]_\times$  transforms a vector  $\mathbf{u} \in \mathbb{R}^3$  into a  $\mathbb{R}^{3 \times 3}$  matrix such that  $[\mathbf{u}]_\times \mathbf{v} = \mathbf{u} \times \mathbf{v}$ .

### B. Formulation of the Optimization Problem

For a part  $r_j$ , under Assumption 1, the  $k$ -th moment of a mass distribution discretized into  $n$  point masses is given by

$$\int_{V_j} \mathbf{p}^k \rho_j(\mathbf{p}) dV_j \approx \frac{m}{n} \sum_i^n (\mathbf{p}_i)^k, \quad (1)$$

where the position of the  $i$ -th point mass relative to  $\mathcal{F}_b$  is given by  $\mathbf{p}_i$ . For a homogeneous mass density, the centre of mass corresponds to the centroid

$${}^b\mathbf{t}_b^{r_j} = \frac{1}{n} \sum_i^n \mathbf{p}_i. \quad (2)$$

The inertia tensor of the  $i$ -th point mass relative to  $\mathcal{F}_b$  is

$$\mathbf{J}(\mathbf{p}_i) = -m_i [\mathbf{p}_i]_\times [\mathbf{p}_i]_\times \quad (3)$$

$$= m_i \begin{bmatrix} y^2 + z^2 & -xy & -xz \\ -yx & x^2 + z^2 & -yz \\ -zx & -zy & x^2 + y^2 \end{bmatrix}, \quad (4)$$

where  $\mathbf{p}_i^\top = [x, y, z]$  and  $m_i$  is the mass of the point.

The part's inertial parameters with respect to the sensor frame  $\mathcal{F}_s$  are

$${}^s\phi^{r_j} = [m, {}^s\mathbf{c}^{r_j}, {}^s\mathbf{J}^{r_j}]^\top = m \begin{bmatrix} 1 \\ {}^s\mathbf{t}_s^{r_j} \\ \text{vech}\left({}^s\mathbf{R}^b \frac{1}{n} \sum_i^n (-[\mathbf{p}_i]_\times [\mathbf{p}_i]_\times) {}^b\mathbf{R}^s - [{}^s\mathbf{t}_s^{r_j}]_\times [{}^s\mathbf{t}_s^{r_j}]_\times\right) \end{bmatrix}, \quad (5)$$

where  ${}^s\mathbf{R}^b$  is the rotation that aligns  $\mathcal{F}_b$  to  $\mathcal{F}_s$ ,  ${}^s\mathbf{t}_s^{r_j}$  is the translation that brings the centroid of  $r_j$  to  $\mathcal{F}_s$ , and  $\text{vech}(\cdot)$  is the *vector-half* operator defined in [33] that extracts the elements on and above the main diagonal.

By keeping  $\mathbf{p}_i$  fixed and by measuring  ${}^s\mathbf{t}_s^b$  and  ${}^s\mathbf{R}^b$  with the robot's perception system, it becomes clear that only the part's mass  $m$  needs to be inferred in (5) as  ${}^s\mathbf{t}_s^{r_j} = {}^s\mathbf{R}^b {}^b\mathbf{t}_b^{r_j} + {}^s\mathbf{t}_s^b$ . Hence, assuming that the robot's perception system can provide  $\mathbf{p}_i$  and that  ${}^s\mathbf{t}_s^b$  and  ${}^s\mathbf{R}^b$  are either known or measured, the inertial parameters of a homogeneous part depend solely on its mass. Similarly, the inertial parameters of a rigid object can be expressed as a function of the masses of its constituent parts.

For "stop-and-go" motions, where force measurements are taken while the robot is immobile, only the mass and COM

are identifiable [4]. Nonetheless, a stop-and-go trajectory greatly reduces noise in the data matrix, because accurate estimates of the end-effector kinematics are not needed [4]. Assuming that the manipulated object is built from up to four homogeneous parts<sup>2</sup>, finding the mass of each part enables the identification of the complete set of inertial parameters even when stop-and-go trajectories are performed. This identification requires measuring the wrench  $\mathbf{b}_j$  at time step  $j$  and relating the wrench to the masses  $\mathbf{m}$  via

$$\underbrace{\begin{bmatrix} \mathbf{f}_s \\ \boldsymbol{\tau}_s \end{bmatrix}}_{\mathbf{b}_j} = \underbrace{\begin{bmatrix} \mathbf{g}_s [1 & 1 & 1 & 1] \\ -[\mathbf{g}_s]_\times [{}^s\mathbf{t}_s^{r_1} & {}^s\mathbf{t}_s^{r_2} & {}^s\mathbf{t}_s^{r_3} & {}^s\mathbf{t}_s^{r_4}] \end{bmatrix}}_{\tilde{\mathbf{A}}_j} \underbrace{\begin{bmatrix} m_{r_1} \\ m_{r_2} \\ m_{r_3} \\ m_{r_4} \end{bmatrix}}_{\mathbf{m}}, \quad (6)$$

stacking  $K$  matrices such that  $\mathbf{A} = [\tilde{\mathbf{A}}_1^\top, \dots, \tilde{\mathbf{A}}_K^\top]^\top$  and  $\mathbf{b} = [\mathbf{b}_1^\top, \dots, \mathbf{b}_K^\top]^\top$ . Minimizing the Euclidean norm leads to the convex optimization problem

$$\begin{aligned} \min_{\mathbf{m} \in \mathbb{R}^4} \quad & \|\mathbf{A}\mathbf{m} - \mathbf{b}\|_2 \\ \text{s.t.} \quad & m_{r_j} \geq 0 \quad \forall j \in \{1, \dots, 4\}, \end{aligned} \quad (7)$$

which can be efficiently solved with standard methods.

## IV. VISUAL PART SEGMENTATION

In this section, we combine a part segmentation method that uses local information (e.g., surface normals) with a second method that relies on structural information (e.g., shape convexity). The Python implementation of our part segmentation algorithm, which is described by Algorithm 1, includes an open-source version of [8] as well as our variant of [9] that makes use of colour information.

Defining the shape of an object from a point cloud involves reconstructing the object surface (i.e., shape reconstruction). In this work, the ball-pivoting algorithm [34] is used owing to its speed and relative effectiveness on point clouds of low density. Shape reconstruction can be challenging for objects with very thin parts (e.g., a saw) and a thickening of the object through voxelization is performed beforehand.

As stated by (1), the moments of a mass distribution are computed by integrating over the shape of the distribution. Hence, volumetric part segmentation is sufficient for identification of the true inertial parameters of an object. To obtain such a representation of the object from its surface mesh, tetrahedralization is performed via TetGen [31] and the resulting tetrahedral mesh is supplied as an input to the part segmentation algorithm.

Our method makes use of the Hierarchical Tetrahedra Clustering (HTC) algorithm [8], which iteratively merges clusters of tetrahedra such that the result is as convex as possible while also prioritizing smaller clusters. HTC maintains a graph with nodes representing clusters and edges representing adjacency, and with an edge cost based on the

<sup>2</sup>For stop-and-go trajectories, the rank of the data matrix is four when using non-degenerate poses as described in the appendix of [4]. Dynamic trajectories can increase the rank of  $\mathbf{A}$  to 10, enabling mass identification of up to 10 unique homogeneous parts.

concavity and size of the connected nodes. The concavity of a cluster is computed by subtracting its volume from its convex hull and an edge cost is defined by

$$c_{ij} = \text{CVXHULL}(C_i \cup C_j) - \text{VOL}(C_i \cup C_j)$$

$$\text{Cost}(i, j) = \begin{cases} c_{ij} + 1, & \text{if } c_{ij} > 0 \\ \frac{|C_i|^2 + |C_j|^2}{N^2}, & \text{otherwise} \end{cases}, \quad (8)$$

where  $|C_i|$  is the number of elements in cluster  $C_i$  and  $N$  is the total number of elements. The edge associated with the lowest cost is selected iteratively and the connected nodes are merged into a single cluster, resulting in the hierarchical segmentation.

To make part segmentation faster, we perform an initial clustering such that HTC requires fewer convex hull computations, which is by far the most expensive operation. The initial clustering is provided through a bottom-up point cloud segmentation algorithm [9] that clusters points together based on heuristic features and chooses a representative point for each cluster. Two adjacent clusters are merged if the dissimilarity between their representative points is lower than some threshold  $\beta$  (n.b. the  $\lambda$  symbol is used in [9]). The value of  $\beta$  increases with each iteration; the process halts when the desired number of clusters is obtained. In our implementation, the dissimilarity metric is defined as:

$$D(C_i, C_j) = \lambda_p \|\mathbf{p}_i - \mathbf{p}_j\| + \lambda_t \|\mathbf{l}_i - \mathbf{l}_j\| + \lambda_n (1 - |\mathbf{n}_i \cdot \mathbf{n}_j|) \quad (9)$$

where each  $\lambda$  is a tunable weight,  $\mathbf{p}_i$  is the position of the representative point of cluster  $C_i$ ,  $\mathbf{l}_i \in \mathbb{N}^3$  is the RGB representation of its colour, and  $\mathbf{n}_i$  is its local surface normal.

To enable accurate part segmentation, the initial clustering should not cross the boundaries of the parts that will subsequently be defined by the HTC algorithm. Therefore, initial clustering is stopped when the number of clusters (e.g., 50 in our case) is much larger than the desired number of parts. The desired number of clusters does not need to be tuned on a per-object basis as long as it is large enough.

## V. EXPERIMENTS

In this section, each component of the proposed method is evaluated on 20 objects with standard metrics, and our entire identification pipeline is benchmarked in 80 scenarios. The practicality of the proposed approach is also demonstrated through a real ‘hammer balancing act’ experiment using relatively inexpensive robot hardware (see Fig. 4).

To conduct simulation experiments with realistic objects and evaluate the performance of part segmentation and identification, a dataset with ground truth inertial parameters and segments is needed. To the best of the authors’ knowledge, no such dataset is freely available. From CAD files contributed by the community, we built a dataset containing 20 commonly-used workshop tools. For each object, our dataset contains a watertight mesh, a coloured surface point cloud with labelled parts, the object’s inertial parameters, and a reference frame specifying where the object is usually

---

### Algorithm 1: HTC (2) with initial clustering (1)

---

**input** : A point cloud  
**output**: A segmented tetrahedral mesh

- 1 Initialize similarity threshold  $\beta$  as done in [9]  
Initialize each point as a cluster  
**while** *number of clusters* > *desired number* **do**
  - foreach** *existing cluster*  $C_i$  **do**
    - foreach** *neighboring cluster*  $C_j$  **do**
      - if**  $D(C_i, C_j) < \beta$  **then**
        - Merge  $C_j$  into  $C_i$
  - $\beta \leftarrow 2\beta$
- foreach** *point*  $\mathbf{p}$  *at the border of two clusters* **do**
  - Merge  $\mathbf{p}$  into the most similar cluster

- 2 Perform surface reconstruction and tetrahedralization  
Associate each tet. to its nearest cluster  
Initialize a node in the graph for each cluster  
**foreach** *node* **do**
  - if** *two tet. from two nodes share a face* **then**
    - Create an edge between the nodes and
    - compute edge-cost with (8)
- while** *number of edges* > 0 **do**
  - Find edge with lowest cost and merge its nodes
  - Create a parent node that contains merged nodes

Label each tet. with its associated cluster number

---

grasped. This dataset of realistic objects enables the evaluation of shape reconstruction, part segmentation, and inertial parameter identification.

#### A. Experiments on Objects from Our Dataset

The quality of a shape reconstructed from point cloud data is evaluated by computing the Hausdorff distance between the ground truth mesh and the reconstructed mesh, both scaled such that the diagonal of their respective bounding box is one metre in length. Part segmentation performed with our proposed variation of HTC is evaluated via the undersegmentation error (USE) [35], and with the global consistency error (GCE) [36], [23], with results summarized in Table II. The USE measures the proportion of points that are crossing segmentation boundaries, or *bleeding out* from one segment to another. The GCE measures the discrepancy between two segmentations, taking into account the fact that one segmentation can be more refined than the other. Both evaluation metrics represent dimensionless error ratios and are insensitive to over-segmentation, as discussed in Section VI.

Table I summarizes the runtime performance obtained by augmenting HTC with the initial clustering in Algorithm 1. While the average segmentation error is almost identical in both cases, Algorithm 1 executes in about a third of the time, owing to the smaller number of convex hull computations performed. The reconstruction and part segmentation can also be qualitatively evaluated on point clouds obtained from RGB-D images taken by a Realsense D435 camera at the wrist of the manipulator. To complete a shape that is partly

TABLE I: Average computation time and segmentation error per object from our dataset with standard deviations in parentheses. Initial clustering significantly reduces the runtime with little impact on the part segmentation.

Algorithm	USE	GCE	Time (s)
HTC	0.1 (0.13)	0.05 (0.10)	9.73 (5.56)
HTC with Initial Clustering	0.1 (0.11)	0.07 (0.11)	3.48 (1.00)

TABLE II: Evaluation of part segmentation (USE and GCE) and shape reconstruction (Hausdorff). As expected, objects with a single part do not have GCE and USE errors while objects with many parts have larger segmentation errors.

Object	USE	GCE	Hausdorff (mm)	Mass (g)	#Parts
Allen Key	0	0	8.2	128	1
Box Wrench	0	0	10.9	206	1
Measuring Tape	0	0	11.4	136	1
Ruler	0	0	14.8	9	1
Screwdriver	0.013	0.001	12.1	30	2
Nut Screwdriver	0.002	0.002	16	81	2
Rubber Mallet	0.011	0.009	15.8	237	2
Bent Jaw Pliers	0.176	0.01	16.5	255	3
Machinist Hammer	0.018	0.012	13.6	133	2
Pliers	0.123	0.041	10.7	633	3
C Clamp	0.103	0.077	9.6	598	5
Adjustable Wrench	0.117	0.098	14.2	719	4
Hammer	0.08	0.098	14.7	690	3
File	0.057	0.102	17.9	20	3
Socket Wrench	0.141	0.123	7.4	356	5
Hacksaw	0.129	0.128	11.1	658	7
Clamp	0.259	0.181	17.6	340	7
Vise Grip	0.158	0.287	17.2	387	8
Electronic Caliper	0.42	0.316	9.1	174	14
Vise Clamp	0.296	0.373	15.1	225	9

occluded by the support table, the point cloud is projected onto the table plane, producing a flat bottom that might not correspond to the true shape of the object. For instance, the left side of the rotated screwdriver point cloud in Fig. 2 is the result of such a projection.

The proposed identification algorithm is tested in simulation under four noise scenarios where zero-mean Gaussian noise is added to the signals. The standard deviations of the noise values are based on the specifications of the Robotiq FT-300 sensor as described in Table III. The identification trajectory used to generate the regressor matrix  $\mathbf{A}$  in (7) has the robot lift the object and successively orient it such that gravity is aligned with each axis of  $\mathcal{F}_b$ , stopping for a moment at each pose and moving relatively quickly between poses, as shown in Fig. 3. The average condition

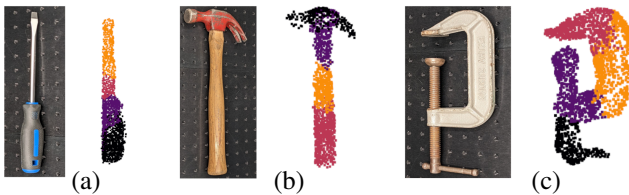


Fig. 2: Pictures of objects scanned by the RGB-D camera on the manipulator, next to their segmented meshes; points are located at the tetrahedra's centroids (screwdriver is rotated).

TABLE III: Standard deviations of the zero-mean Gaussian noise added to accelerations and force signals in simulation.

Scenario	Ang. Acc.	Lin. Acc.	Force	Torque
Low Noise	0.25	0.025	0.05	0.0025
Moderate Noise	0.5	0.05	0.1	0.005
High Noise	1	0.1	0.33	0.0067

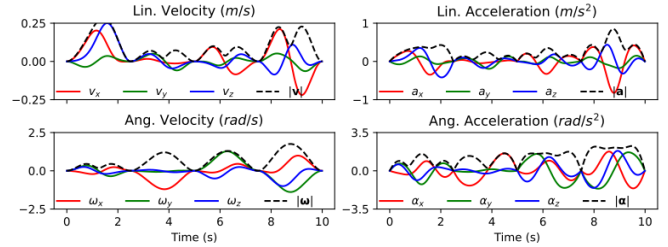


Fig. 3: Kinematics of the identification trajectory performed with the Rubber Mallet. The norm of the velocity (black) goes slightly above the standard maximum speed for a cobot.

number [37] of the scaled regressor matrix for all simulated trajectories is 74, 92, and 99 for the low, moderate, and high noise scenarios, respectively. These relatively low condition numbers confirm that our identification trajectory is non-degenerate.

The performance of our proposed algorithm (HPS) is compared against the classical algorithm proposed in [11] (OLS) and to the more modern algorithm proposed in [17] (GEO). The latter was provided with a prior solution consisting of the true mass of the object, and the COM and inertia tensor resulting from a homogeneous mass distribution for the object ( $\alpha = 10^{-5}$  was used). HPS only uses measurement data when both the linear and angular accelerations are below 1 unit/ $s^2$ , corresponding to the *stalled* timesteps of the stop-and-go motion, whereas OLS and GEO use all data available.

The accuracy of inertial parameter identification is measured via the Riemannian geodesic distance  $e_{\text{Rie}} = \sqrt{(\frac{1}{2} \sum \lambda)}$  [38], where  $\lambda$  is the vector of eigenvalues for  $P^{(s}\phi^1)^{-1}P^{(s}\phi^2)}$  and  $P(\phi)$  is the *pseudoinertia* matrix that is required to be symmetric positive-definite (SPD) [16]. The metric  $e_{\text{Rie}}$  is the distance between the estimated and ground truth inertial parameters in the space of SPD matrices. To assist interpretation of the identification error, we also compute the size-based error metrics proposed in [4], which use the bounding box and mass of the object to produce average *percentage* errors  $\bar{e}_m$ ,  $\bar{e}_C$ , and  $\bar{e}_J$  associated, respectively, with the mass, centre of mass, and inertia tensor estimates. Table IV reports the mean of the entries of the error vector  $\bar{e}$  for each quantity of interest.

### B. Demonstration in Real Settings

To test our proposed method in a realistic setting, we used a uFactory xArm 7 manipulator equipped with a RealSense D435 camera and a Robotiq FT-300 force-torque sensor, as shown in Fig. 1. First, the hammer in Fig. 2 was scanned with the camera, producing 127 RGB-D images of the scene in about 30 seconds. The object was then picked up at a predetermined grasping pose where the Robotiq 2F-85

TABLE IV: Comparison between HPS (ours), OLS, and GEO with various levels of noise, indicating the percentage of solutions that were physically consistent (Cons).

Noise	Algo.	Cons. (%)	$\bar{e}_m(\%)$	$\bar{e}_C(\%)$	$\bar{e}_J(\%)$	$e_{\text{Rie}}$
No	OLS	100	<0.1	<0.1	<b>0.09</b>	<b>0.03</b>
	GEO	100	<0.1	1.35	53.22	1.14
	HPS	100	0.27	0.1	10.28	0.72
Low	OLS	14	0.19	2.13	>500	N/A
	GEO	100	<b>0.18</b>	1.34	52.79	1.14
	HPS	100	0.40	<b>0.32</b>	<b>11.12</b>	<b>0.74</b>
Mod.	OLS	4.5	0.36	5.33	>500	N/A
	GEO	100	<b>0.35</b>	1.37	51.73	1.14
	HPS	100	0.74	<b>0.48</b>	<b>11.81</b>	<b>0.77</b>
High	OLS	2	0.69	10.11	>500	N/A
	GEO	100	<b>0.64</b>	1.58	48.49	1.13
	HPS	100	2.79	<b>1.07</b>	<b>15.00</b>	<b>0.87</b>

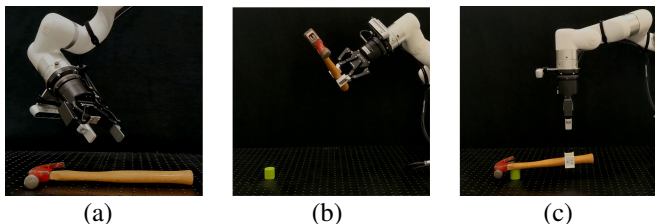


Fig. 4: The hammer is (a) scanned, (b) picked up for inertial identification, and (c) balanced onto the small green target.

gripper could fit into plastic holders attached to the object (enabling a stable grasp of the handle). A short trajectory that took about 10 seconds to execute was performed while the dynamics of the object were recorded at approximately 100 Hz. Point cloud stitching, mesh reconstruction, and part segmentation can be performed concurrently while the robot executes the trajectory, since these operations take 2.87, 0.24, and 2.82 seconds, respectively. Finally, our proposed method identified the inertial parameters in about 0.5 seconds with MOSEK [39]. Using only its estimate of the COM, the robot autonomously balanced the hammer on a cylindrical target with a radius of 17.5 mm. The entire process is shown in our accompanying video, with summary snapshots provided in Fig. 4. In contrast, both OLS and GEO returned inaccurate parameter estimates, causing hammer balancing to fail.

## VI. DISCUSSION

This work exploits object shape information to produce fast and accurate inertial parameter estimates. The object reconstructions used by our method to determine object shape are often already computed as components of planning and manipulation algorithms, limiting the computational burden introduced by our approach. The decomposition of objects into parts also enables fast inertial parameter identification of bodies that have joints (e.g., hedge scissors), since the parameters can be trivially updated following a change in object configuration.

Wrong or inaccurate part segmentation may lead to erroneous parameter estimates. However, if Assumption 1 holds, *over-segmenting* an object does not affect the result of the

identification as for a given part, (1) can be decomposed into

$$\int_{V_1} \mathbf{p}^k \rho_1(\mathbf{p}) dV_1 + \int_{V_2} \mathbf{p}^k \rho_2(\mathbf{p}) dV_2 = \int_V \mathbf{p}^k \rho(\mathbf{p}) dV, \quad (10)$$

which is true since  $V = V_1 \cup V_2$  and  $\rho_1(\mathbf{p}) = \rho_2(\mathbf{p})$ . Similarly, if two conceptually distinct parts with the *same* mass density are erroneously combined into one, the result of the identification will remain unaffected. However, if parts with *different* mass densities are considered to be a single part by the algorithm, the identification will fail since (10) does not hold when  $\rho_1(\mathbf{p}) \neq \rho_2(\mathbf{p})$ .

The comparison in Table IV suggests that OLS outperforms other algorithms in the noiseless scenario, which is expected since it is not biased by any prior information. However, OLS nearly always converges to physically inconsistent solutions and becomes inaccurate in the presence of even a small amount of sensor noise. The GEO algorithm performs similarly across noise levels, possibly due to the very good prior solution (i.e., correct mass, homogeneous density) provided, corresponding to the exact solution for objects that have a single part. On average, HPS outperforms OLS and GEO for the identification of COM and J when the signals are noisy. The slightly higher  $\bar{e}_m$  for HPS may be caused by the approximation made when using stalling motions [4].

Experiments with objects from our dataset do not reveal any obvious trends relating the quality of the shape reconstruction (measured via the Hausdorff distance), the quality of the part segmentation (measured via USE and GCE), and the quality of the inertial parameter identification (measured via  $e_{\text{Rie}}$ ). This can be explained by the fact that an object’s mass and shape largely determine the signal-to-noise ratios that any identification algorithm has to deal with.

As demonstrated by experiments with objects that are mostly symmetrical (e.g., the screwdriver), if the shape of the object is such that the parts’ centroids are coplanar, the optimizer will ‘lazily’ zero out the mass of some parts since they are not required to minimize (7). An improved version of HPS could use the hierarchy from HTC to intelligently define segments whose centroids are not coplanar.

## VII. CONCLUSION

In this paper, we leveraged the observation that man-made objects are often built from a few parts with homogeneous densities. We proposed a method to quickly perform part segmentation and inertial parameter identification. We ran 80 simulations in which our approach outperformed two benchmark methods, and we demonstrated real-world applicability by autonomously balancing a hammer on a small target. On average, our proposed algorithm performs well in noisy conditions and can estimate the full set of inertial parameters from ‘stop-and-go’ trajectories that can be safely executed by collaborative robots. Promising lines of future work include formulating the optimization problem as a mixed-integer program where mass densities are chosen from a list of known materials, and improving the segmentation algorithm such that parts’ centroids are never coplanar.

## REFERENCES

- [1] M. T. Mason and K. M. Lynch, "Dynamic manipulation," in *1993 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 1, pp. 152–159, 1993.
- [2] "Robots and robotic devices - Safety requirements for industrial robots - Part 1: Robots," Tech. Rep. ISO 10218-1:2011, ISO, Geneva (2011).
- [3] "Robots and robotic devices - Collaborative robots," Tech. Rep. ISO/TS 15066:2016, ISO, Geneva (2016).
- [4] P. Nadeau, M. Giamou, and J. Kelly, "Fast Object Inertial Parameter Identification for Collaborative Robots," in *IEEE International Conference on Robotics and Automation*, May 2022.
- [5] C. C. Pagano and M. Turvey, "Eigenvectors of the inertia tensor and perceiving the orientation of a hand-held object by dynamic touch," *Perception & psychophysics*, vol. 52, no. 6, pp. 617–624, 1992.
- [6] J. Lukos, C. Ansuini, and M. Santello, "Choice of contact points during multidigit grasping: Effect of predictability of object center of mass location," *Journal of Neuroscience*, vol. 27, no. 14, pp. 3894–3903, 2007.
- [7] J. B. Hamrick, P. W. Battaglia, T. L. Griffiths, and J. B. Tenenbaum, "Inferring mass in complex scenes by mental simulation," *Cognition*, vol. 157, pp. 61–76, 2016.
- [8] M. Attene, M. Mortara, M. Spagnuolo, and B. Falcidieno, "Hierarchical convex approximation of 3D shapes for fast region selection," in *Computer Graphics Forum*, vol. 27, pp. 1323–1332, 2008.
- [9] Y. Lin, C. Wang, D. Zhai, W. Li, and J. Li, "Toward better boundary preserved supervoxel segmentation for 3D point clouds," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 143, pp. 39–47, 2018.
- [10] G. Golluccio, G. Gillini, A. Marino, and G. Antonelli, "Robot dynamics identification: a reproducible comparison with experiments on the kinova jaco2," *IEEE Robotics and Automation Magazine*, 2020.
- [11] C. G. Atkeson, C. H. An, and J. M. Hollerbach, "Estimation of inertial parameters of manipulator loads and links," *The International Journal of Robotics Research*, vol. 5, no. 3, pp. 101–119, 1986.
- [12] D. Kubus, T. Kroger, and F. M. Wahl, "On-line estimation of inertial parameters using a recursive total least-squares approach," in *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3845–3852, 2008.
- [13] S. Farsoni, C. T. Landi, F. Ferraguti, C. Secchi, and M. Bonfè, "Real-time identification of robot payload using a multirate quaternion-based kalman filter and recursive total least-squares," in *2018 IEEE International Conference on Robotics and Automation*, pp. 2103–2109, 2018.
- [14] C. D. Sousa and R. Cortesao, "Physical feasibility of robot base inertial parameter identification: A linear matrix inequality approach," *The International Journal of Robotics Research*, vol. 33, no. 6, pp. 931–944, 2014.
- [15] S. Traversaro, S. Brossette, A. Escande, and F. Nori, "Identification of fully physical consistent inertial parameters using optimization on manifolds," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5446–5451, 2016.
- [16] P. M. Wensing, S. Kim, and J.-J. E. Slotine, "Linear matrix inequalities for physically consistent inertial parameter identification: A statistical perspective on the mass distribution," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 60–67, 2017.
- [17] T. Lee, P. M. Wensing, and F. C. Park, "Geometric robot dynamic identification: A convex programming approach," *IEEE Transactions on Robotics*, vol. 36, no. 2, pp. 348–365, 2019.
- [18] K. Ayusawa and Y. Nakamura, "Identification of standard inertial parameters for large-DoF robots considering physical consistency," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 6194–6201, 2010.
- [19] C. Song and A. Boularias, "A probabilistic model for planar sliding of objects with unknown material properties: Identification and robust planning," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5311–5318, October 2020.
- [20] B. Sundaralingam and T. Hermans, "In-Hand Object-Dynamics Inference Using Tactile Fingertips," *IEEE Transactions on Robotics*, vol. 37, no. 4, pp. 1115–1126, 2021.
- [21] D. D. Hoffman and W. A. Richards, "Parts of recognition," *Cognition*, vol. 18, no. 1-3, pp. 65–96, 1984.
- [22] R. S. Rodrigues, J. F. Morgado, and A. J. Gomes, "Part-based mesh segmentation: a survey," in *Computer Graphics Forum*, vol. 37, pp. 235–274, 2018.
- [23] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 1–12, 2009.
- [24] A. Shamir, "A survey on mesh segmentation techniques," in *Computer Graphics Forum*, vol. 27, pp. 1539–1556, 2008.
- [25] A. Tagliasacchi, T. Delame, M. Spagnuolo, N. Amenta, and A. Telea, "3d skeletons: A state-of-the-art report," in *Computer Graphics Forum*, vol. 35, pp. 573–597, 2016.
- [26] C. Lin, L. Liu, C. Li, L. Kobbelt, B. Wang, S. Xin, and W. Wang, "Seg-mat: 3d shape segmentation using medial axis transform," *IEEE Transactions on Visualization and Computer Graphics*, 2020.
- [27] P. Li, B. Wang, F. Sun, X. Guo, C. Zhang, and W. Wang, "Q-mat: Computing medial axis transform by quadratic error minimization," *ACM Transactions on Graphics*, vol. 35, no. 1, pp. 1–16, 2015.
- [28] O. V. Kaick, N. Fish, Y. Kleiman, S. Asafi, and D. Cohen-Or, "Shape segmentation by approximate convexity analysis," *ACM Transactions on Graphics*, vol. 34, no. 1, pp. 1–11, 2014.
- [29] Z. Dou, C. Lin, R. Xu, L. Yang, S. Xin, T. Komura, and W. Wang, "Coverage axis: Inner point selection for 3d shape skeletonization," vol. 41, no. 2, pp. 419–432, 2022.
- [30] C. Lin, C. Li, Y. Liu, N. Chen, Y.-K. Choi, and W. Wang, "Point2Skeleton: Learning Skeletal Representations from Point Clouds," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4277–4286, 2021.
- [31] H. Si, "TetGen, a Delaunay-based quality tetrahedral mesh generator," *ACM Transactions on Mathematical Software*, vol. 41, no. 2, pp. 1–36, 2015.
- [32] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, 1996.
- [33] H. V. Henderson and S. Searle, "Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics," *Canadian Journal of Statistics*, vol. 7, no. 1, pp. 65–81, 1979.
- [34] F. Bernardini, J. Mittleman, H. Rushmeier, C. Silva, and G. Taubin, "The ball-pivoting algorithm for surface reconstruction," *IEEE Transactions on Visualization and Computer Graphics*, vol. 5, no. 4, pp. 349–359, 1999.
- [35] A. Levinshstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2290–2297, 2009.
- [36] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *2001 IEEE International Conference on Computer Vision*, vol. 2, pp. 416–423, 2001.
- [37] M. Gautier and W. Khalil, "Exciting trajectories for the identification of base inertial parameters of robots," *The International Journal of Robotics Research*, vol. 11, no. 4, pp. 362–375, 1992.
- [38] S. Lang, *Fundamentals of Differential Geometry*, vol. 1. Springer Science and Media, 1999.
- [39] E. D. Andersen and K. D. Andersen, "The mosek interior point optimizer for linear programming: An implementation of the homogeneous algorithm," in *High performance optimization*, pp. 197–232, 2000.