



HAL
open science

In-hand manipulation planning using human motion dictionary

Ali Hammoud, Valerio Belcamino, Alessandro Carfi, Veronique Perdereau,
Fulvio Mastrogiovanni

► **To cite this version:**

Ali Hammoud, Valerio Belcamino, Alessandro Carfi, Veronique Perdereau, Fulvio Mastrogiovanni. In-hand manipulation planning using human motion dictionary. 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2022), Aug 2022, Naples & virtual conference, Italy. pp.927-933, 10.1109/RO-MAN53752.2022.9900858 . hal-04609538

HAL Id: hal-04609538

<https://hal.science/hal-04609538v1>

Submitted on 12 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

In-hand manipulation planning using human motion dictionary

Ali Hammoud¹, Valerio Belcamino², Alessandro Carfi², Veronique Perdereau¹ and Fulvio Mastrogiovanni²

Abstract—Dexterous in-hand manipulation is a peculiar and useful human skill. This ability requires the coordination of many senses and hand motion to adhere to many constraints. These constraints vary and can be influenced by the object characteristics or the specific application. One of the key elements for a robotic platform to implement reliable in-hand manipulation skills is to be able to integrate those constraints in their motion generations. These constraints can be implicitly modelled, learned through experience or human demonstrations. We propose a method based on motion primitives dictionaries to learn and reproduce in-hand manipulation skills. In particular, we focused on fingertip motions during the manipulation, and we defined an optimization process to combine motion primitives to reach specific fingertip configurations. The results of this work show that the proposed approach can generate manipulation motion coherent with the human one and that manipulation constraints are inherited even without an explicit formalization.

I. INTRODUCTION

Humans’ daily activities often require interacting with objects and dexterously manipulating them in hand. These abilities result from a continuous learning process, during human lives, based on the observation of other humans’ actions and personal attempts and errors. For robots to successfully integrate and operate in the human environment, they should interact with the environment as humans do [1]. Therefore, robots should be able to manipulate unknown objects dexterously, adapting their previous experience to new scenarios. Furthermore, robots should be able to learn new manipulation skills by observing other “agents” actions. A robotic platform can achieve this by integrating advanced perception tools and flexible learning methods to represent and plan new manipulation actions.

Given a predefined manipulation goal, planning a dexterous manipulation consists in determining the necessary finger trajectories to reach it. Following the literature, we can divide approaches for planning robotic manipulations into two categories: data-driven and classical [2]–[10]. In data-driven approaches, dexterous manipulation models are trained either by robot trial and error or by observing human demonstrations [2]–[5]. Instead, classical approaches

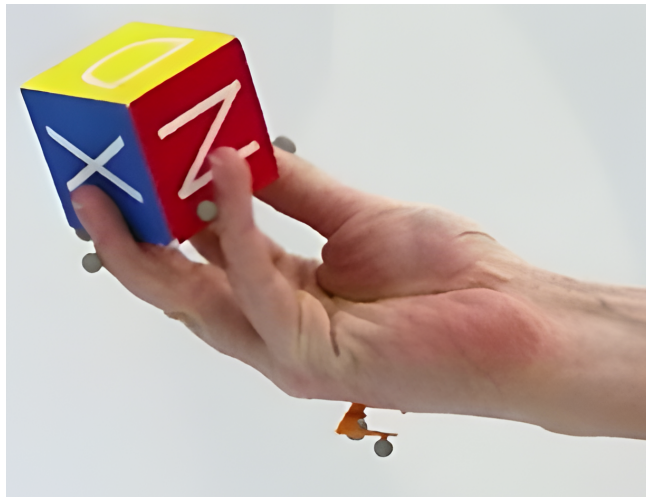


Fig. 1: Example of an in-hand manipulation supported by the fingertips.

are based on robotics principles, dividing complex tasks into sets of elementary actions [6]–[9].

In the context of in-hand manipulation path planning, most data-driven approaches rely on dynamic movement primitives (DMP) [2]. This solution is elegant and generates smooth trajectories while keeping a small number of parameters. DMP is made up of a set of generalized dynamics system equations that flexibly express movements of any shape by altering a simple linear dynamical system with a non-linear component [3]. Non-linear components of DMP systems can be determined using data either from human demonstrations or robot ones. Alternative systems use Markov Decision Processes (MDP). With MDP, a Dynamical Bayesian Network models the evolution of an agent’s state according to its actions and the environment dynamics [4]. Finding an in-hand manipulation grasp sequence to pass from an initial hand position to a final one is what the MDP idea entails [5]. However, this does not include movement between grasps. In general, in-hand manipulation based on training methods necessitates a large amount of training data and a significant processing time.

On the other side, in-hand manipulation planning in classical robotics is solved by modelling the environment and the robotic hand. In this context, robot hand actions are decomposed into atomic sequences such as the in-hand regrasping and finger reallocation. In-hand regrasping consists in repositioning the object to the desired pose within the robotic hand. Sundaralingam and Hermans 2017 [6] proposed an

*This work was supported by the European project Index

¹A. Hammoud and V. Perdereau are with Sorbonne Universite, Institut des Systemes Intelligents et de Robotique, ISIR, F-75005 Paris, France ali.hammoud, veronique.perdereau@sorbonne-universite.fr

²V. Belcamino, A. Carfi and F. Mastrogiovanni are with TheEngineRoom, Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa, Via Opera Pia 13, 16145, Genoa, Italy valerio.belcamino@edu.unige.it, alessandro.carfi@dibris.unige.it, fulvio.mastrogiovanni@unige.it.

efficient solution to this problem with a purely kinematic trajectory optimization. Various approaches based on rolling and sliding motions have also been explored [7]. Instead, in finger relocation, sometimes known as finger gaiting, the robot shifts contact points between the object and the fingertip by moving a single finger at a time. Solutions to finger gaiting based on geometric approaches were presented by Sundaralingam and Hermans, 2018 [8] and Fan et al., 2017 [9]. Classical approaches have a predictable planning outcome since the designer sets all the constraints. However, this category of solutions can only approach atomic actions and not the whole in-hand manipulation due to the increasing complexity in modelling the system constraints.

Both classical and data-driven approaches have their limitations. With the former, one should manage large training datasets and long training times while, with the latter, one encounters difficulties in modelling complex constraints. We aim to overcome these limitations by generating human-like gaiting and in-hand regrasping with a hybrid approach that creates an in-hand manipulation primitives dictionary from human demonstrations [10]. With this approach, combining primitives, we create robotic in-hand manipulation trajectories that respect the constraints imposed by the environment. This solution is possible since the dictionary is built from human demonstrations, and therefore, the dictionary elements respect all the complex constraints without the need for their formal definition. Finally, the possibility to build the primitives dictionary from small data samples resolves the computational problem often associated with data-driven approaches.

The remainder of this paper is organized as follows. The concept of a primitive dictionary and its application into in-hand manipulation are covered in the first section. In the second section, path planning by integrating an in-hand manipulation dictionary is explained. The third part presents the implemented steps to integrate an in-hand manipulation dictionary into the path planning problem passing through data acquisition and model training and testing. In the fourth part, we show the obtained testing results for simulated trajectory .

II. PRIMITIVES DICTIONARY

Primitives dictionaries have been extensively used in fields such as computer vision [11], clustering of documents [12], astronomy [13] and generating human motion trajectories [14]–[16]. Given the desired behaviour to represent (V) it is possible to reconstruct it with a dictionary of primitives (W) combining them with a set of weights (h):

$$V = Wh \quad (1)$$

We should point out that W has dimensionality $|V| \times I$, where $|V|$ is the cardinality of the "behaviour to represent" and I is the number of primitives. The weight vector h has dimensionality I . The primitives dictionary can be learned from data using different approaches. Most common solutions for primitive extractions involve the usage of non-negative matrix factorization method (NMF) [17] or principal

component analysis (PCA) [18]. The selection process of the weights allows generating various behaviours starting from the same dictionary. For this reason, weights are typically selected through an optimization process considering constraints that can vary according to the scenario.

A. In-Hand Manipulation Primitives

In this work, we start from the assumption that it is possible to model the in-hand manipulation of an object (see Figure 1), from an initial to the desired pose, as the evolution of the 3D position of the fingertips over time. The problem is therefore characterized by two main components: the fingers ($F(t)$) and object ($O(t)$) trajectories in space. Starting from this assumption, we can approach the problem by describing the fingertips as 3D points and their trajectories as follows:

$$F(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_5(t) \end{bmatrix} \\ \text{s.t. } t \in [1, \dots, N]$$

where

$$f_i(t) = \begin{bmatrix} x_i(t) \\ y_i(t) \\ z_i(t) \end{bmatrix}$$

At the same time, we can describe the object pose evolution considering pose its 3D position and its orientation, expressed through Euler angles:

$$O(t) = [p_o(t), r_o(t)] \\ \text{s.t. } t \in [1, \dots, N]$$

where

$$p_o(t) = \{x_o(t), y_o(t), z_o(t)\} \\ r_o(t) = \{\psi_o(t), \theta_o(t), \phi_o(t)\}$$

Given these definitions the overall problem of the in-hand manipulation can be summarized as finding the intermediate points of the fingertips trajectories to carry the object from an initial pose ($O(1)$) to a final desired pose ($\hat{O}(N)$) and reach a desired final fingertips configuration ($\hat{F}(N)$).

If we consider the problem one finger at a time and we recall Eq. 1 we can express the problem as follows:

$$f_i(t) = W_i(t)h \quad (2)$$

Where the primitives dictionary for the i -th finger is influenced by the time evolution. Given this formalization we can describe the overall problem as:

$$F(t) = \mathbb{W}(t)h \quad (3)$$

where

$$\mathbb{W}(t) = \begin{bmatrix} W_1(t) \\ \vdots \\ W_5(t) \end{bmatrix}$$

Notice that with this description, the primitives are not described individually for each finger. Therefore, the weight

h is shared over all the finger primitives, and we describe the in-hand manipulation as the interplay of all the fingers. With this formalization of the in-hand manipulation problem, it is possible to learn the primitives dictionary \mathbb{W} using the non-negative matrix factorization method (NMF), as described by Hammoud et al., 2021 [10].

III. GENERATION OF MANIPULATIONS

Once the primitives dictionary is available, we can generate in-hand manipulation trajectories by infusing the dictionary with the weights. However, the generated manipulation should follow some specific constraints to be useful for robotic applications. Before introducing these constraints and explaining the procedure to generate new manipulation skills from our dictionary, we have to point out that the proposed approach is limited by a few assumptions:

- The robot and gravity are the only factors that can affect the object's pose (i.e there are no external forces acting on our system).
- The robotic hand and the object are rigid.
- The initial object grasp is stable.
- The desired object pose is in the reachable zone of the fingertips.
- Contacts between the hand and the object are only made at fingertips.
- For fine in-hand manipulations, the final object pose is fully determined by the fingertips motions and the initial object pose.

Furthermore, we should point out that our formalization describes the in-hand manipulation as the trajectories in the Cartesian space of five fingertips, ignoring their orientations. This simplification is reasonable considering that, for in-hand manipulations, the fingertips' orientations, expressed with respect to the hand back, are strongly influenced by their 3D position. Although these assumptions limit the range of in-hand manipulation our method can achieve, we should point out that our method is able to produce a wide range of manipulations and that it could be extended to relax some of these assumptions. Given the problem formalization and the initial assumption, we can now introduce the procedure to compute the appropriate weights to solve a specific manipulation. This is done through an optimization process that has an objective function J described as:

$$J(F, O) = J_F + \alpha J_O \quad (4)$$

where J_F represents the distance between the final fingertips configuration $F(N)$ and the desired final fingertips configuration $\hat{F}(N)$. J_F is defined as follows:

$$J_F = \sum_{i=1}^5 \|f_i(N) - \hat{f}_i(N)\|$$

J_O is the distance between the final object pose $O(N)$ and the desired final pose ($\hat{O}(N)$). Since the object pose is composed of both 3D position and orientation, the distances for

the two components are computed separately and the overall distance J_O is defined as the sum of the two distances:

$$J_O = \|p_o(N) - \hat{p}_o(N)\| + \|r_o(N) - \hat{r}_o(N)\|$$

Finally, alpha is a scalar weight fine-tuning the trade-off between the two cost components. While solving the in-hand manipulation problem for a robotic hand, other constraints linked to the robotic hand kinematics should be addressed. The i -th fingertip should be in the reachable workspace (\mathbb{F}_i)

$$f_i(t) \in \mathbb{F}_i, \forall i \in [1, 5], \forall t \in [1, N] \quad (5)$$

and at the same time, each fingertip should respect kinematic constraints on their instantaneous velocity

$$\dot{f}_{i,min} \leq \dot{f}_i(t) \leq \dot{f}_{i,max}, \forall i \in [1, 5], \forall t \in [1, N] \quad (6)$$

While performing the in-hand manipulation, other constraints apply. Fingertips should not collide with each other and at each instant, during the manipulation, the object should be in contact with at least three fingertips to ensure a stable grasp [19]. We can express the first constraint by checking the Euclidean distances between fingertips for all the manipulation time:

$$\min_{\forall j \in (i, 5)} \|f_i(t) - f_j(t)\| \neq 0, \forall i \in [1, 4], \forall t \in [0, N] \quad (7)$$

For the second constraint, we have instead to introduce the distance between the fingertip ($f_i(t)$) and the object surface, represented as a point-cloud ($M(t)$) of 3D points ($m_k(t)$, with $k \in [1, |M|]$). Given this representation, the distance between the i -th fingertip and the object ($d_i(t)$) is defined as the distance between the fingertip and the closest point of the object point-cloud:

$$d_i(t) = \min_{\forall k \in [1, N]} \|f_i(t) - m_k(t)\|$$

and the subset of fingertips in contact with the object ($Fc(t)$)

$$Fc(t) = \{f_j(t) \in F(t) \mid f_j(t) \leq \tau\} \quad (8)$$

where τ is a threshold value defining the maximum distance between the fingertip and the object when they are in contact. Given this definition, the cardinality of the subset of fingertips in contact should be always at least 3 to maintain a stable grasp during the manipulation

$$|Fc(t)| \geq 3, \forall t \in [1, N] \quad (9)$$

A. Relaxed Optimization Problem

Human manipulation typically follows the constraints previously described and our approach uses human data to derive the manipulation primitives. Therefore, it is reasonable to assume that at least some of these constraints are going to be transposed to the new generated trajectories without including them in the optimization problem. This consideration should apply for most of the constraints except for the fingertips velocity constraint because of the differences between the human and robot's velocity ranges. Furthermore, since each dictionary is trained for a specific object and the pose of the object is strictly related to the fingertips

motions, because of the first assumption, we can simplify the optimization function considering the fingertips component only. Therefore, Eq. 4 can be rewritten as:

$$J(F) = J_F = \sum_{i=1}^5 \|f_i(N) - \hat{f}_i(N)\|_2^2$$

that using Eq. 2 can be written as:

$$J(F) = \sum_{i=1}^5 \|W_i(N)h - \hat{f}_i(N)\|_2^2 \quad (10)$$

The optimization process should find the weights h that minimizes Eq. 10 while considering the fingertips velocity constraint expressed in Eq. 6.

All the other constraints not directly considered in the optimization phase (Eq. 5,7,9) are going to be evaluated during the testing phase to determine if they are embedded into dictionaries learned from human demonstrations.

IV. IMPLEMENTATION

Given the problem statement introduced in Section II and the optimization procedure detailed in Section III, human demonstrations of in-hand manipulations are needed to implement and evaluate the presented approach. These demonstrations should consist of the 3D positions of the fingertips during manipulation. In fact, this information is enough to both train the dictionary and generate new manipulations. However, to test the adherence to the constraints introduced in Section III, the pose of the object is also necessary.

Many solutions exist to track the position and the orientation of objects. Information from RGB cameras can be processed to extract the position of the fingertips [10] or datagloves can be used to track human motions [20]. At the same time, the pose of an object can be tracked using embedded inertial measurement units [21] or through image processing [22]. Since the main objective of this work is to validate the proposed approach to generate in-hand manipulations, we used a motion capture (MoCap) system to gather precise and reliable data.

A. Data Acquisition

For the data acquisition we used a MoCap system composed of six OptiTrack Flex-3 cameras¹ (see Figure 3). This MoCap system can track the 3D position of reflective markers (see Figure 2a and 2b), and position and orientation of sets of markers mounted on a rigid object, i.e., rigid body. The MoCap acquires all the data at the constant frequency of 100Hz.

Data about the hand motion during the manipulation are collected using five markers on the fingertips and one rigid body on the hand back (see Figure 2a). More precisely, the reflective markers are placed between the nail and the distal-interphalangeal joint to minimize collisions and occlusions. In this way, we maximize the accuracy of the tracking system. The objects poses are tracked with a single rigid body. As can be seen in Figure 2b, the experimental scenario

involves two different objects: i) a cube with an edge of a 5 cm; ii) a cylinder with 5 cm diameter and 5 cm height.

For both objects, we have collected two datasets, one for training and another one for testing. For what regards the training dataset only the hand was tracked by the MoCap system and the experiments consisted of rotations and translations of the object on multiple axes while its weight is supported by the fingers. The palm of the hand keeps facing upwards during these actions for a stronger grip and to guarantee that the object is exclusively rotated by the movement of the fingers. Figure 1 shows a snapshot of the experiment.

To record the data for testing the proposed method, we adopted a slightly different approach since also tracking the object pose was necessary. As shown in Figure 2b, a set of markers were attached to the objects to track their pose. The tracking of the object with the MoCap system inserts a trade off between the freedom of movement and the tracking precision. We opted to place the markers on a single face in a way that their centroid coincides with the axis of symmetry of the object. This choice allows precise tracking of the rigid body pose and easy computation of the absolute object position from the measurements with a simple translation in the local system of reference. However, this marker disposal limits the possible manipulation, allowing complete rotation along the axis orthogonal to the markers only.

Both training and testing data have been recorded thanks to a single subject performing for approximately 2 minutes an in-hand manipulation: five times for the training and one for the test. This procedure has been repeated for both objects.

B. Model Training and Testing

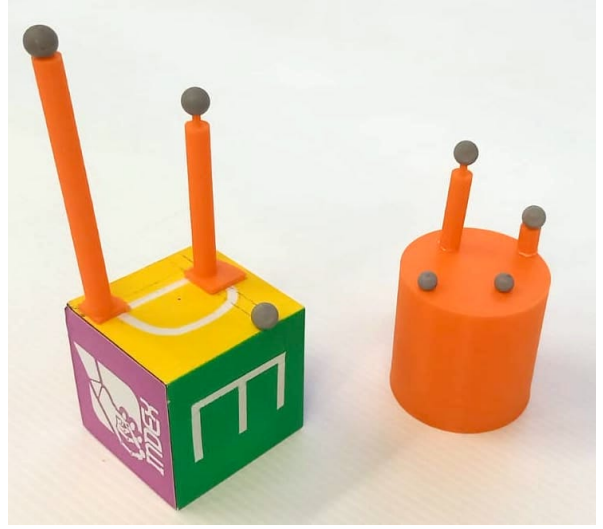
Before the training process can take place the collected data should be processed. The processing is divided into three phases. In the first phase, the position of the fingertips and the full pose of the object are transformed according to the hand back pose. Following the first assumption presented in Section III, we are going to study the components of the manipulation related to the fingertips only. Therefore, by performing the described transformation we can work only on the fingertip and object poses assuming the hand is static. In the second phase, all the processed data are filtered using a median filter of window size 50 samples. In the third and last phase, given the object model and its pose, it is generated a pointcloud ($M(T)$) of the object surface that will be used in the testing phase to compute the distance between the fingertips and the object. The final cube pointcloud has a dimension of around fifty thousand points while the cylinder one has a dimension of seventy thousand points.

After their processing the training data are used to derive two dictionaries, one for the cube and one for the cylinder, using the approach proposed by Hammoud et al., 2021 [10]. Both dictionaries have been trained on 1 second segments of the full in-hand manipulation. This allows to simplify the training and to model intermediate configurations of the in-hand manipulation. However, this implies that these dictionaries could be used to generate in-hand manipulation

¹<https://optitrack.com/cameras/flex-3/>



(a) Hand Marker Placement.



(b) Objects Marker Placement.

Fig. 2: On the left the placement of the MoCap markers on the hand, on the right the placement of markers on the two different objects.



Fig. 3: The motion capture system.

of 1 second only. The final dictionaries have both 200 primitives (i.e., $I = 200$).

Finally, to test the method, the collected data are divided into segments of 1 second. For each segment, the last fingertip configuration is used as target ($\hat{f}_i(N)$) for the optimization process described by Eq. 10. To solve the optimization problem and find the weights (h) we used the MATLAB optimization toolbox with a quadratic programming feature. The resulting weights are then used in Eq. 3 to generate

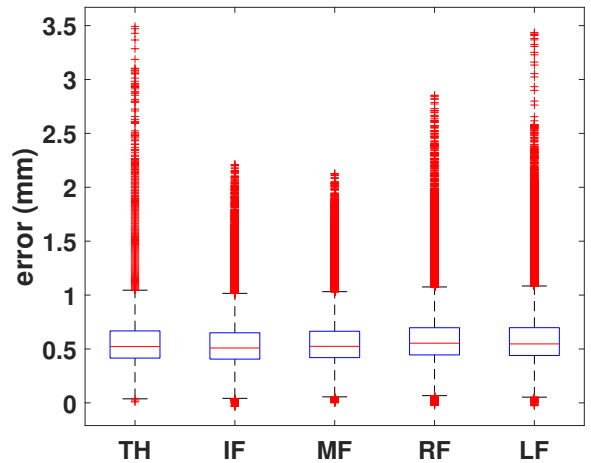


Fig. 4: The box plot of the Euclidean distance between recorded and generated fingertips positions for the cube manipulation. TH - Thumb; IF - Index; MF - Middle; RF - Ring; LF - Little.

all the entire fingertips trajectory $F(t)$. Notice that since the trajectory length is 1 second and the data frequency is 100Hz, the trajectory is composed of 100 samples ($t \in [1, 100]$). From the single sequence recorded for the testing, we were able to extract and generate 100 sequences of 1 second for each object.

V. RESULTS

To evaluate the performance of our approach to generate new manipulation trajectories, we carried out an in-depth

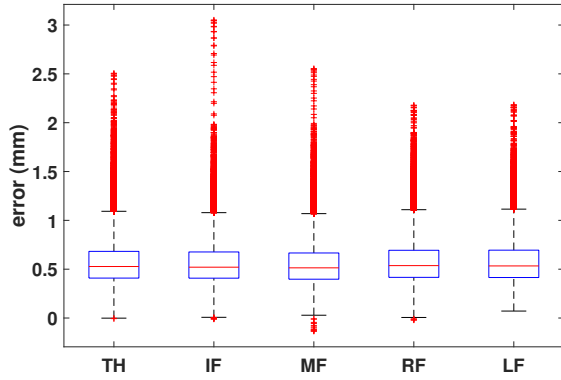


Fig. 5: The box plot of the Euclidean distances between recorded and generated fingertips positions for the cylinder manipulation.

analysis. At first, we analyzed the error between recorded and generated manipulations. We computed this error at each instant as the Euclidean distance between recorded and generated points. These results are summarized by the box plot in Figure 4 for the cube and Figure 5 for the cylinder. These two figures show the errors for each finger over approximately ten thousand time instants. The results present many outliers, up to $3.5mm$ of error. However, for all the fingers, in 75% of the time instants, the error is below $1.2mm$.

To check the capability of the system to generate new manipulation trajectories while preserving the characteristics of the human manipulation we tested the adherence of the new trajectories to the three constraints introduced in Section III: i) reachability, ii) finger collision, and iii) minimum contact points.

To test the reachability constraint we defined the reachable workspace \mathbb{F}_i , for each finger, as the set containing all the fingertip positions observed in the training set. Then we checked if the new generated fingertip positions are inside the reachable workspace as described in Eq. 5. All the points generated with our approach satisfied the reachability constraint.

For the finger collision, we computed the minimum finger to finger distance as defined in Eq. 7. Over the one-hundred trajectories generated for each object, we detected 4 time instants with a collision for the cube and only one for the cylinder. Therefore, the generated trajectories also satisfied this constraint.

Finally, for what regards the minimum contact points between the fingertips and the object we decided to divide its evaluation into three steps: i) compute the finger to object distance both for recorded and generated data; ii) compute the distance difference between the recorded and generated data to check for similarities in the behaviours; and iii) given the finger to object distance for the generated data determine at each instant how many contacts are present. The Euclidean distance between the fingertips (recorded and

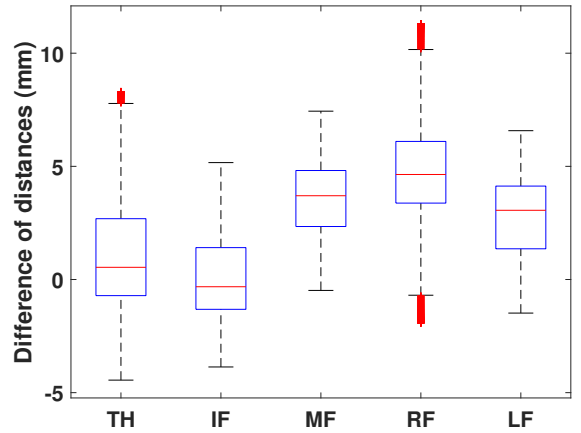


Fig. 6: Difference of distances between human and object and created path and object for the testing cube data.

generated) and the object has been computed as described in Section III using the "object pointcloud" generated during the data preprocessing. For each time instant of the one-hundred sequences, the difference between these two fingers to object distances is presented in the box plots of Figure 6 (for the cube) and 7 (for the cylinder). From the box plots we can see that, contrary to what happens for the cylinder, the difference varies a lot over the different fingers for the cube. This behaviour could be linked to the different shapes but we do not have enough details to conclude on it. Overall the median value for each finger over all the objects is between 0 and $5mm$ and maximum differences never overcome $15mm$. According to this, we can conclude that the relation between the object and fingertips can be partially different when considering human and generated trajectories. This aspect should be the subject of future studies but does not imply that the minimum contact points constraints are not satisfied. To determine when a fingertip is in contact with the object we refer to Eq. 8 setting $\tau = 5mm$. This value has been chosen since according to Delhay et al., 2021 [23] the fingertip deformation can go up to $5mm$. According to this parameter, the generated trajectories were able to maintain three points of contact for 97% and 93% of the time respectively for the cube and the cylinder.

VI. CONCLUSIONS

This paper introduced a method based on motion primitives dictionaries to generate human-like fingertip trajectories for in-hand manipulation. The approach leverages human demonstrations to learn a manipulation dictionary that implicitly includes in-hand manipulation constraints.

The proposed approach, tested for the in-hand manipulation of two different objects (i.e., a cube and a cylinder), could generate new manipulation trajectories coherent with the human demonstrations. This coherence is also evident in the adherence of the manipulations with three constraints that characterizes the in-hand manipulation (i.e., reachability, finger collision, and minimum contact points). A relevant

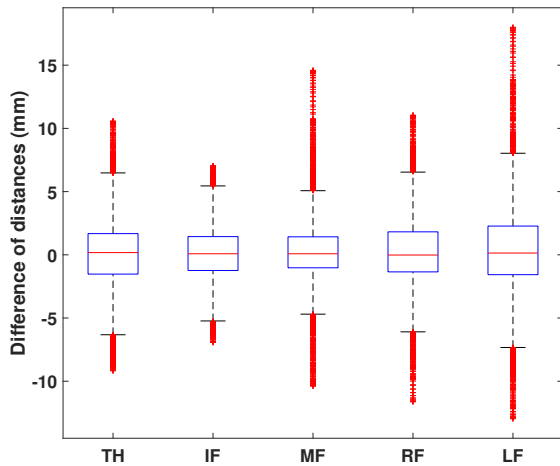


Fig. 7: Difference of distances between human and object and created path and object for the testing cylinder data.

result is that the proposed approach inherits these constraints from the human demonstrations without the need for a formal representation.

Given the observed results, trajectories generated with this approach could drive robotic hands in dexterous manipulations. However, future work in this field should focus on relaxing the adopted assumptions to target more complex in-hand manipulations, for example, by extending the effect on the manipulation from the fingertips to other hand parts.

ACKNOWLEDGMENT

This work is supported by the CHIST-ERA (2014-2020) project InDex and received funding from Agence Nationale de la Recherche (ANR) under grant agreement No. ANR-18-CHR3-0004 and the Italian Ministry of Education and Research (MIUR).

REFERENCES

- [1] A. Carfi, T. Patten, Y. Kuang, A. Hammoud, M. Alameh, E. Maiettini, A. I. Weinberg, D. Faria, F. Mastrogiovanni, G. Alenyà, *et al.*, “Hand-object interaction: From human demonstrations to robot manipulation,” *Frontiers in Robotics and AI*, vol. 8, p. 316, 2021.
- [2] E. Theodorou, J. Buchli, and S. Schaal, “A generalized path integral control approach to reinforcement learning,” *The Journal of Machine Learning Research*, vol. 11, pp. 3137–3181, 2010.
- [3] A. J. Ijspeert, J. Nakanishi, and S. Schaal, “Learning attractor landscapes for learning motor primitives,” in *15th International Conference on Neural Information Processing Systems (NIPS)*, (Vancouver, British Columbia, Canada), January 2002.
- [4] G. E. Monahan, “State of the art—a survey of partially observable markov decision processes: theory, models, and algorithms,” *Management science*, vol. 28, no. 1, pp. 1–16, 1982.
- [5] U. Prieur, V. Perdereau, and A. Bernardino, “Modeling and planning high-level in-hand manipulation actions from human knowledge and active learning from demonstration,” in *IEEE/RSJ International Conference on intelligent Robots and Systems*, (Vilamoura-Algarve, Portugal), December 2012.
- [6] B. Sundaralingam and T. Hermans, “Relaxed-rigidity constraints: kinematic trajectory optimization and collision avoidance for in-grasp manipulation,” *Autonomous Robots*, vol. 43, no. 2, pp. 469–483, 2019.
- [7] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 1994.

- [8] B. Sundaralingam and T. Hermans, “Geometric in-hand regrasp planning: Alternating optimization of finger gaits and in-grasp manipulation,” in *IEEE International Conference on Robotics and Automation (ICRA)*, (Brisbane Convention & Exhibition Centre, Brisbane, Australia), May 2018.
- [9] Y. Fan, W. Gao, W. Chen, and M. Tomizuka, “Real-time finger gaits planning for dexterous manipulation,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 12765–12772, 2017.
- [10] A. Hammoud, A. Diouf, and V. Perdereau, “A robotic in-hand manipulation dictionary based on human data,” in *20th International Conference on Advanced Robotics (ICAR)*, (Ljubljana, Slovenia), December 2021.
- [11] E. Cao, K. Cao, K. Feng, and J. Wang, “Nmf based image sequence analysis and its application in gait recognition,” *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 2, pp. 86–96, 2020.
- [12] Z. Akata, C. Thureau, and C. Bauckhage, “Non-negative matrix factorization in multimodality data for segmentation and label prediction,” in *16th Computer vision winter workshop*, (Mitterberg, Austria), February 2011.
- [13] O. Berne, C. Joblin, Y. Deville, J. Smith, M. Rapacioli, J. Bernard, J. Thomas, W. Reach, and A. Abergel, “Analysis of the emission of very small dust particles from spitzer spectro-imagery data using blind signal separation methods,” *Astronomy & Astrophysics*, vol. 469, no. 2, pp. 575–586, 2007.
- [14] C. Vollmer, S. Hellbach, J. Eggert, and H.-M. Gross, “Sparse coding of human motion trajectories with non-negative matrix factorization,” *Neurocomputing*, vol. 124, pp. 22–32, 2014.
- [15] A. J. Ijspeert, J. Nakanishi, and S. Schaal, “Movement imitation with nonlinear dynamical systems in humanoid robots,” in *Proceedings of the IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, (Washington, DC, USA), May 2002.
- [16] M. S. Lewicki and T. J. Sejnowski, “Learning overcomplete representations,” *Neural computation*, vol. 12, no. 2, pp. 337–365, 2000.
- [17] S. Sra and I. Dhillon, “Generalized nonnegative matrix approximations with bregman divergences,” *Advances in neural information processing systems*, vol. 18, p. 283–290, 2005.
- [18] H. Hotelling, “Analysis of a complex of statistical variables into principal components.,” *Journal of educational psychology*, vol. 24, no. 6, p. 417, 1933.
- [19] D. Prattichizzo and J. C. Trinkle, *Grasping*, pp. 671–700, 2008.
- [20] T. Funatomi, T. Yamane, H. Ouchida, M. Iiyama, and M. Minoh, “Deriving motion constraints in finger joints of individualized hand model for manipulation by data glove,” in *International Conference on 3D Vision (3DV)*, (Seattle, WA, USA), June 2013.
- [21] R. Niewiadomski and A. Sciutti, “Multimodal emotion recognition of hand-object interaction,” in *26th International Conference on Intelligent User Interfaces*, (College Station, TX, USA), April 2021.
- [22] K. Park, T. Patten, and M. Vincze, “Pix2pose: Pixel-wise coordinate regression of objects for 6d pose estimation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (Seattle, WA, USA), 2019.
- [23] B. P. Delhayé, E. Jarocka, A. Barrea, J.-L. Thonnard, B. Edin, and P. Lefèvre, “High-resolution imaging of skin deformation shows that afferents from human fingertips signal slip onset,” *Elife*, vol. 10, p. e64679, 2021.