Toward a General Framework for 3D Deformable Object Grasping and Manipulation

Félix Nadon[1], Angel J. Valencia[1], Narmadha Sambandam[1], Stephen Rowlands[1], James Dickens[1], Pierre Payeur[1]

***Abstract*— The intelligent handling of deformable objects by robotic manipulators has enabled the automation of various delicate or labour-intensive tasks, but still faces many challenges when it comes to more complex situations. In particular, the lack of a general framework to manage deformation makes it difficult to manipulate unknown objects, especially in the 3D case. This work aims to provide a path forward to solve some of the challenges pertaining to the robotic manipulation of deformable objects and integrate these solutions in a flexible system able to manipulate various deformable objects in a task-independent way.**

I. INTRODUCTION

Deformable objects are commonplace in everyday life, and their intelligent manipulation is a requirement for automating many delicate or labour-intensive tasks in e.g., agriculture, food processing, or healthcare. While multiple robotic platforms and algorithms have been proposed to perform specific tasks with such objects [1], [2], they still lack the human-like skills and flexibility that would allow a general-purpose robot to perform varied tasks with different deformable objects while retaining the ability to handle the ever-present rigid objects. Much work has been done on manipulating linear and planar objects such as rope and cloth, but the lack of a task-independent framework to handle deformation limits applications when objects are not fully known a priori. This issue is accentuated when dealing with more complex 3D objects, when non-rigid and rigid objects are mixed up in an application, or when dealing with objects with dissimilar material characteristics that make them partially rigid and partially deformable.

This paper draws upon our work on the intelligent robotic manipulation of non-rigid objects to present a generalizable framework for manipulating deformable objects and highlights some key observations and challenges to overcome.

II. OBSERVATIONS FROM PREVIOUS WORK AND CURRENT CHALLENGES

A. Deformable Objects Sensing, Detection and Understanding

Interactions with deformable objects are inherently complex since it can be challenging to predict their behaviour. When probed, the object undergoes a change in shape and in turn the reaction forces experienced can be nonlinear and vary based on the inhomogeneous nature of the object. The quality of data acquired plays a key role in obtaining accurate and complete models, especially when a scene contains complex or multiple occluded objects. Models built from single-view information have achieved great results for bounding box detection but do not support full 3D representation due to limitations in deducing the shape and characteristics of objects from a single viewpoint. Symmetry based shape completion techniques work well for objects that are made up of simple shapes, while object detection techniques work well when the 3D objects in the scene are known. Learning based solutions have shown very high degrees of success when objects are known but are not as successful when dealing with novel objects.

As a result, the object should be observed from different perspectives to reach a more comprehensive description of the object shape and characterize its material [3]. While creating such models is rather complex, dynamic 3D representations of deformable objects can be built by leveraging the data from an optimized set of viewpoints provided by a network of RGB-D sensors in different locations. This includes multi-sensor systems, mobile sensors that can be moved to obtain numerous views of the scene, as well as systems where the manipulator provides multiple views to a single sensor while controlling the object.

Mainly relying on single view data and in order to validate and compare various object detection and semantic segmentation algorithms, computer vision researchers have typically conducted experiments through the use of publicly available datasets such as ImageNet [4] or Microsoft COCO [5] using standard metrics. Notably, the authors of these datasets do not typically distinguish between rigid and deformable objects in their descriptions, but rather implicitly note the difficulty involved with detecting and segmenting both types of objects in the context of different viewpoints, different object poses, and as parts of a hierarchy.

Because current large-scale datasets for object recognition do not explicitly consider the objects' deformability, but rather imply it through multiple poses and moveable parts, it is challenging to apply machine learning algorithms to identify and track non-rigid objects. This increases the dependency of the initialization sequence on human interaction

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada. {fnado079, avale073, nsamb018, srowl017, jdick088, ppayeur}@uottawa.ca

Workshop on Robotic Manipulation of Deformable Objects (ROMADO)
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2020

to identify objects and validate the detected shape before performing the manipulation. Researchers would benefit from large-scale RGB/RGB-D datasets explicitly used for training models on deformable objects for standardization and comparison. It is worth distinguishing two kinds of deformable objects that could form different datasets in this context. First, there are objects that can be broken down into a discrete set of parts which may be annotated, such as the Pascal VOC *Person Layout* dataset [6]. The second kind of objects are those that are not discretizable in a reasonable sense, e.g., plasticine and sponges. A greater variety of objects is required in datasets of both categories.

Prior to the ubiquitous use of end-to-end training on large datasets with convolutional neural networks, a popular approach in object detection to deal with deformations in objects makes use of deformable part models (DPM), notably implemented by Felzenszwalb et al. [7]. In their work, a set of learned filters conditioned on histogram of oriented gradients (HOG) features are trained on an image pyramid, applied at different scales at a fixed location. Displacement of each filter is based on a set of chosen anchored part locations and spatial priors, relative to a root for a fixed number of parts. The model computes a displacement of the parts relative to the anchors for each part and uses a latent SVM formulation for classification. Notably, DPM explicitly takes into account the alignment of the parts of the model to form a "hypothesis of objectness", and is agnostic to whether the object is rigid or not. Deformable part-based models do not require a-priori knowledge of the configuration of the object and its deformability, and so can be used in the context of detection of both rigid and non-rigid objects.

B. Modelling and Behaviour Prediction from Partial Observations

Modelling objects, either deformable or rigid, from a single point of view provides a limited 2.5D representation. This leads to significant challenges in the grasping and control stages due to the requirement of estimating 3D objects from 2.5D data. An estimate can be made using symmetry, object recognition or learning based techniques. Each method provides a unique approach and may be suitable depending on the problem. The difficulty of estimating 3D shape from 2.5D information is exacerbated once an object begins to deform or is occluded by the manipulator. These challenges demand novel solutions that can be applied generally across a range of deformable and rigid objects.

The major limitation of current research is the lack of a general-purpose deformation modelling methodology. Physics-based models are very accurate, but not very flexible since these are tied to a particular configuration. Strong assumptions are necessary to support the selection of the type of model, e.g., homogeneous composition or isotropic materials. However, determining these conditions in advance is very difficult in robotic environments since objects and their properties are often unknown. In contrast, learning-based models are capable of extracting latent properties from data and support generalization across objects.

A particular challenge of learning-based models is the prediction of long-term deformation behaviours. The collection of a large amount of data to achieve better performance is an option, though an expensive and dangerous one when using robots. An interesting approach to more efficiently solve this problem is to exploit the structure of the data directly in the models. The use of deep learning techniques that explicitly integrate relational constraints such as graph neural networks demonstrates capabilities to learn long-term dynamics of physical systems [8].

A more general problem for any modelling approach is handling partial observations. In some cases, specific sensing techniques can be used to augment the shape data, e.g., multi-view sensor fusion or occlusion removal algorithms. The physical data can be augmented by e.g., contact detection by force and tactile sensors. But even in a sensor-rich environment, certain properties are very difficult to capture directly from the data. A promising approach to managing complex deformation behaviours is the operation of the models in latent space representations [9].

C. Grasp Optimization and Control

In our previous work on grasp selection and in-hand shape control of deformable objects [10], we found that it was possible to use a simplified model of the hand motion and various heuristics to quickly identify a stable, task-optimized grasp based solely on a visual description of the non-rigid object to manipulate. This is especially useful for automatically selecting the initial grasp on an object for which the rigidity and other material properties are unknown. We also observed that shape control algorithms and principles that are effective for reshaping linear and planar deformable objects may not be directly applicable to volumetric objects, as they tend to show an increase in rigidity as they are compressed and a more well-defined shape when at rest, two properties that are not as present in e.g., ropes and cloth.

Our work in [10] also highlighted some significant challenges for the control of reshaping tasks. First and foremost, we note that the selected grasp and performance of the control scheme are heavily dependant on the quality of the object description, both initially and throughout the manipulation. If the initial shape of the object is not correctly captured, for instance due to occlusions or insufficient contrast, the selected grasp may fail to remain stable or to optimize the task. In terms of control, inaccuracies in the object's contour make it difficult to automatically quantify the quality of the end result and to define the control error in the manipulation.

As highlighted by the significant differences between linear, planar, and volumetric types, deformable objects have a wide range of possible behaviours depending on their dimensionality, rigidity, and whether they are elastic, plastic, or somewhere in between [3]. This variety of possible behaviours presents additional challenges when considering objects composed of heterogeneous materials, or when a mix of non-rigid and rigid objects are present in the environment.

In complex situations involving the manipulation of non-rigid objects, it is not feasible to rely on offline simulations to derive a motion plan that guarantees object integrity, slippage avoidance, or even the success of the manipulation, as this would require perfect and complete a priori knowledge of the objects, robot, and environment. Given the potentially unpredictable behaviour of non-rigid objects, it is necessary to rely on a real-time quality assurance scheme to ensure that the above constraints are respected.

Another significant challenge in developing this module is the availability of easy-to-integrate tactile sensors that are accurate enough to perform contact detection and slippage avoidance on soft non-rigid objects as well as the supporting algorithms. Moreover, the real-time adaptation of the motion plan to ensure the quality of the manipulation requires fast and accurate models and algorithms for the tracking, behaviour prediction, and control of deformable objects.

III. FRAMEWORK DEFINITION AND DEVELOPMENT

A. Proposed Framework

Derived from observations on the state-of-the-art and our previous work on the topic, Fig. 1 proposes a general framework for the robotic manipulation of deformable objects based on visual and tactile sensing. This framework is divided in three main stages, namely the initialization, planning, and execution phases. In the initialization phase, the system primarily uses visual information, from one or more sensors, to identify the workspace and relative positions of the robot, objects to manipulate, and any obstacles that should be avoided. This stage also includes the task definition by identifying the object to manipulate and the task to perform along with desired locations, shapes or paths. In the planning phase, initial knowledge of the object, task, and robotic system is applied to derive grasping points and trajectories for the robot and manipulator that are expected to complete the defined task. Finally, the execution phase is when the initial contact with the object is made and the manipulation task is performed using real-time visual and tactile tracking of the object. Running in parallel with the manipulation task, we define a task-independent “quality assurance” step to handle issues such as slippage avoidance and maintaining the structural integrity of the object throughout the manipulation. This part of the system is also responsible for tracking the true behaviour of the object under manipulation and updating the knowledge encoded in the object model and behaviour prediction module, adapting and changing the initial grasp and robot trajectory as necessary. This adaptation behaviour would provide the system with the flexibility to handle objects with initially uncertain characteristics, including objects made of non-homogeneous materials, or if a variety of rigid and non-rigid objects are present.

B. From Simulation to Physical Manipulation

While much of the research on deformable object manipulation has so far relied on simulated environments, transferring these skills to the physical manipulation of 3D non-rigid objects presents substantial integration challenges in all phases of the manipulation sequence.

In the initialization phase, multiview and multimodal sensing capabilities are needed to obtain a complete description of the objects and environment due to occlusions and limited prior knowledge. However, implementing these sensing capabilities in the physical world requires careful extrinsic calibration of all sensors as well as fusion of data from multiple sources. Given the variety of sensor types and modalities that may be involved, these tasks must be carefully handled to generalize to different setups. Object identification schemes used in initialization must also deal with complex backgrounds and measurement noise that may not be as significant in simulations.

In the planning phase, we note that modelling the physics of deformable objects tends to be computationally expensive. As such, when dealing with robotic applications, object representations must be selected with care and optimized to support real-time operation. Moreover, the physical characteristics and mechanical accuracy of the robotic hands and manipulators are taken into account: the selected grasp points and motions are limited to those reachable by the robot, and the motion planning aims to minimize the impact of inaccuracies, for instance by avoiding grasps on sharp corners.

When executing the manipulation task, it is important to note that the modelling of deformable objects is more error-prone than that of rigid objects as their shape and behaviour can change during the manipulation. This calls for live grasp stability verification and quality assurance throughout the manipulation, which may also require updating the object model and manipulation plan in real-time.

C. Sample Results

As a sample of our results obtained in [10] and [11], Fig. 2a shows a spherical deformable object modelled using an optimized neural gas network. Fig. 2b shows the grasp and finger paths selected for deforming the initially circular object into a rectangular shape (shown in gold) with the 3-finger Barrett hand.

IV. CONCLUSION

When attempting to manipulate deformable objects with a robot, it is crucial for the object to be correctly and completely identified, or for the manipulation system to be highly adaptable such that it may detect failures and adapt to new information as the object is handled. Given the complexity and variety of possible behaviours of non-rigid

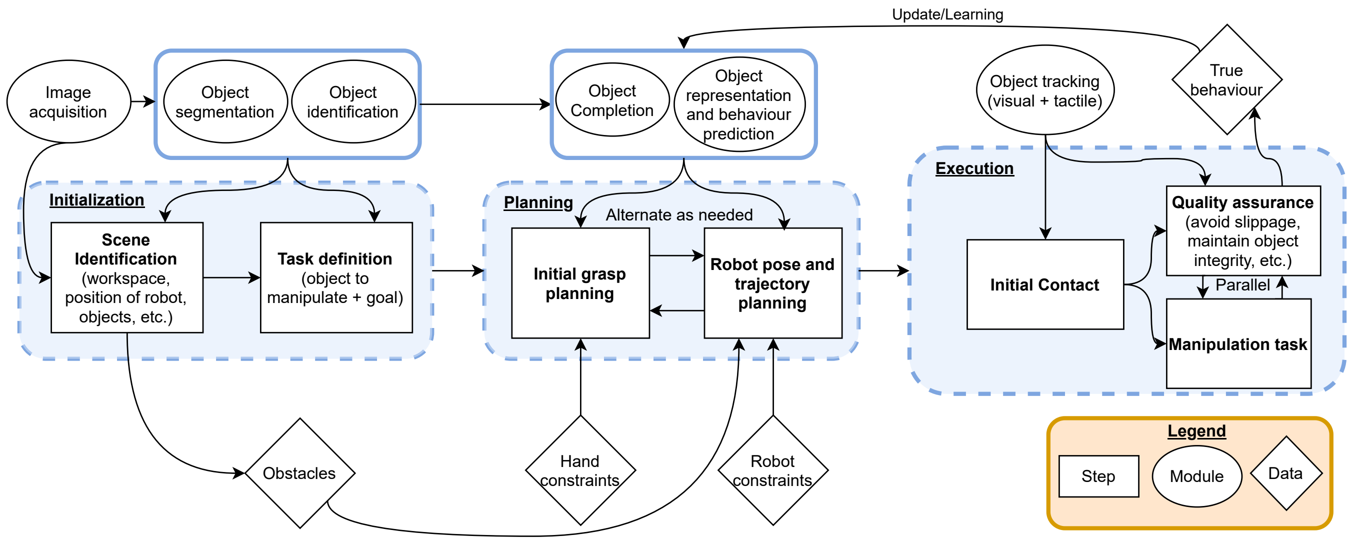
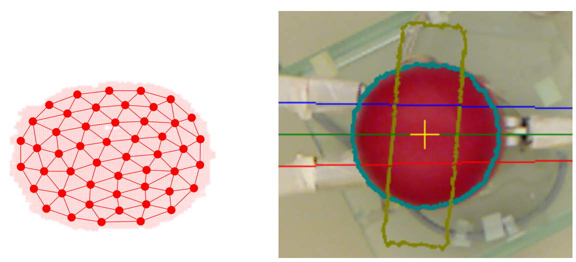


Fig. 1. Proposed framework for manipulation of non-rigid objects



(a) (b)

Fig. 2. (a) Optimized neural gas representation of a spherical object and (b) selected grasp for object reshaping based on a 2D view

objects, it is not practical to rely on offline simulation for planning robotic actions. This is especially true in cases when the properties of the objects to handle are not fully known beforehand, including when considering objects composed of heterogeneous materials, or when a mix of different non-rigid and rigid objects are present in the environment.

Based on these observations and the challenges highlighted, this paper proposes a general framework to handle a mixture of potentially unknown, deformable objects in a task-independent way. Many key challenges to overcome in developing this system involve multi-view and multi-modal sensing capabilities. These include the accurate detection and tracking of deformable objects, strategies to deal with occlusions, and reliable contact detection and slippage avoidance systems for non-rigid objects. Moreover, it is essential to develop fast and adaptable models that can be used to predict the behaviour of deformable 3D objects in real time.

ACKNOWLEDGMENTS

The authors wish to acknowledge support to this research from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ontario Graduate Scholarship (OGS) program, and from the Ontario Queen Elizabeth II Graduate Scholarship in Science and Technology program.

REFERENCES

[1] J. Sanchez, J.-A. Corrales, B.-C. Bouzgarrou, and Y. Mezouar, ‘Robotic Manipulation and Sensing of Deformable Objects in Domestic and Industrial Applications: a Survey’, The International Journal of Robotics Research, vol. 37, no. 7, pp. 688–716, Jun. 2018, doi: 10.1177/0278364918779698.

[2] F. Nadon, A. J. Valencia, and P. Payeur, ‘Multi-Modal Sensing and Robotic Manipulation of Non-Rigid Objects: A Survey’, Robotics, vol. 7, no. 4, p. 74, Nov. 2018, doi: 10.3390/robotics7040074.

[3] F. Hui, P. Payeur, and A.-M. Cretu, ‘Visual Tracking of Deformation and Classification of Non-Rigid Objects with Robot Hand Probing’, Robotics, 6(1)-5, 2017, doi: 10/gfs798.

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, and L. Fei-Fei, ‘ImageNet Large Scale Visual Recognition Challenge’. International Journal of Computer Vision, 2015.

[5] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C.L. Zitnick, ‘Microsoft Coco: Common Objects in Context’, Proceedings of the European Conference on Computer Vision (ECCV), 2014.

[6] M. Everingham and J. Winn, ‘The Pascal Visual Object Classes Challenge 2012 (VOC2012) Development Kit’, Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep. 8, 2011.

[7] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, ‘Object Detection with Discriminatively Trained Part-based Models’, IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9):1627–1645, 2010.

[8] P. Battaglia, R. Pascanu, M. Lai, and D.J. Rezende, ‘Interaction Networks for Learning About Objects, Relations and Physics’, Advances in Neural Information Processing Systems, pp. 4502-4510, 2016.

[9] D. Rodriguez, C. Cogswell, S. Koo, and S. Behnke, ‘Transferring Grasping Skills to Novel Instances by Latent Space Non-Rigid Registration,’ IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, 2018, pp. 4229-4236, 2018, doi: 10.1109/ICRA.2018.8461169.

[10] F. Nadon and P. Payeur, ‘Grasp Selection for In-Hand Robotic Manipulation of Non-Rigid Objects with Shape Control’, in Annual IEEE International Systems Conference (SysCon), Montréal, QC, Aug. 2020.

[11] A. J. Valencia, F. Nadon, and P. Payeur, ‘Toward Real-Time 3D Shape Tracking of Deformable Objects for Robotic Manipulation and Shape Control,’ 2019 IEEE SENSORS, Montreal, QC, Canada, 2019, pp. 1-4, doi: 10.1109/SENSORS43011.2019.8956623.