Toward Real-Time 3D Shape Tracking of Deformable Objects for Robotic Manipulation and Shape Control

Angel J. Valencia, Félix Nadon and Pierre Payeur

School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada
{avale073, fnado079, ppayeur}@uottawa.ca

***Abstract*—Dexterous robotic manipulation of complex non-rigid objects is a requirement for automating many delicate tasks. This work explores various sensing and modeling strategies for tracking the 3D shape of objects during manipulation, allowing for more accurate and responsive planning and control of robotic manipulators. These approaches are evaluated in terms of their ability to correctly capture the behaviour of non-rigid objects as well as their computational complexity, which is strongly constrained by the need for real-time control on robotic platforms with limited computing power.**

***Keywords*—RGB-D imaging, 3D sensing and modeling, non-rigid deformable objects, robotic manipulation, force closure, diminishing rigidity.**

I. Introduction

Many delicate or labour-intensive tasks that have yet to be automated in the contexts of food processing, surgery, industrial assembly and household environments involve the manipulation of objects which change their shape as a result of the manipulation. In order for robots and automated systems to perform these tasks, it is necessary that they become able to handle such non-rigid objects in a safe, dexterous manner. Much attention has been given so far to 1D (e.g. rope, knotting) and 2D (e.g. clothes folding) non-rigid objects in the scope of robotic manipulation. However, shape measurement and deformation sensing has mostly been restricted to 2D imaging [1], [2]. These approaches yielded interesting results for the classification of objects and materials according to their intrinsic elasticity, but revealed problematic for more accurate dynamic shape tracking. Here, the sensing approach is extended and aims to tackle a more comprehensive set of measurements, targeting the full 3D reshaping of objects when submitted to external forces. The use of 3D shape sensing will lead to more accurate planning and control and allow a robotic system to handle complex everyday objects which cannot be correctly described with a 2D representation only. The research is performed in the context of safe manipulation and object reshaping with multi-finger robotic hands. Different 3D representation techniques are investigated to experimentally determine the most robust, while computationally tractable, approaches for dynamic 3D shape representation from raw RGB-D sensors. The context of development is illustrated with experimental tests on deformable objects with various initial shapes and elasticity properties.

II. State of the Art

Sensing strategies used to capture the deformation of 3D non-rigid objects [3] mainly comprise depth sensors installed either as standalone or as part of a multi-modal system, with the aim of obtaining a 3D pointcloud which captures sufficient information about the object. Similarly, the computational representation of these 3D non-rigid objects is mainly obtained by two classes of methods: those that use a formal model and those that do not require a particular model and instead perform online estimation techniques. Within model-based methods, the most common approach is to use a physics-based representation. However, these models require the specification of physical properties of the object, which are often unknown or difficult to estimate when dealing with complex structures. On the other hand, model-free or learning approaches are interesting alternatives because they provide better generalization and scaling capabilities to handle more complex structures without the problem of defining physical properties beforehand.

As for robotic manipulation of non-rigid objects, most recent approaches highlighted in [3], [4] focus on linear (rope) and planar (cloth-like) objects. Some work has also been performed regarding manipulations in a 2D environment, either in simulation or with relatively flat real scenes observed through a simple 2D vision system. Even though research efforts are slowly turning to 3D objects, they have not yet received much attention. Capturing the behaviour of deformable 3D objects requires more complex models than for their 1D and 2D counterparts, which in turn require more elaborate sensing setups and increased computing power, all of which are challenging issues for robotic manipulation outside of simulation environments.

III. Deformable Object Representation and Use for Robotic Manipulation

In order to capture the deformation of complex objects from real world scenes, a method to structure the sensor data is needed. Typically, raw data taken from sensors (usually a 3D pointcloud) is not suitable for representing objects because these basic structures are formed as a set of unrelated points non-uniformly distributed. Instead, objects should be represented as a collection of points in which there exists some relation between points around a neighborhood. In this way, there is a mechanism to capture the deformation. Likewise, this representation must also be active, in the sense that it should be able to evaluate functions in such a way that it can predict how the structure would change given certain input variables. Given these constraints, two models that may prove suitable for this problem are discussed. Growing Neural Gas (GNG) and Particle Graph Networks (PGN) are architectures based on graphs which demonstrate an object-relation (as nodes-edges) interaction in addition to being defined as adaptable models. In the following, both models are briefly described, emphasizing their features and limitations as well as how they can be applied to represent deformable objects.

A. Growing Neural Gas

Growing Neural Gas (GNG), originally presented in [5], is a type of unsupervised learning algorithm derived as a vector quantization technique, but that has also shown interesting capabilities for structure representation. In essence, GNG learns a topological representation of some data distribution as a dynamic undirected graph $G = \langle A, N \rangle$, where $A = \{a_i\}$ is the set of nodes (or neurons) and $N = \{n_k\}$ is the set of edges. Also, each node $a_i = \{w_i\}$ has an associated weight vector $w_i \in \mathbf{R}^n$ of the same dimension as the input space n. The network is seen as dynamic since the nodes A are not fixed but rather added incrementally and edges N are also changing over time according to some criterion (e.g., threshold distance). This adaptability coupled with intrinsic signal filtering properties, such as denoising and downsampling [6], makes GNG particularly suitable to represent complex changing structures in an efficient manner while maintaining robustness against noisy data coming from sensors.

B. Particle Graph Network

Particle Graph Network (PGN) [7], [8] is a novel representation that takes inspiration from physics simulation models used in computer graphics, in the way of how particles can be associated to represent more complex entities. However, unlike the graphics model, PGN represents each entity as a graph neural network which has the ability to predict the dynamics of the system by learning parameters from data. Many architectures have been proposed for this representation, but in essence PGN is defined as a directed graph $G = \langle O, R \rangle$ where $O = \{o_i\}$ is the set of nodes (or objects) and $R = \{r_k\}$ is the set of edges (or relations). Each node is described as $o_i = \{x_i, a_i^o\}$, where $x_i = \{q_i, \dot{q}_i\}$ and a_i^o corresponds to the state (e.g, position, velocity) and features (e.g, stiffness, radius) of node i, respectively. Moreover, each relation is described as $r_k = \langle u_k, v_k, a_k^r \rangle$, where u_k is the node receiver, v_k the node sender, whereas a_k^r corresponds to the features (e.g, connections) of relation k. Afterwards, the updated graph G at time $t+1$ is obtained by predicting the next state of each object $o_{i,t+1} = f_O\left(o_{i,t}, e_{k,t}\right)$ which evaluates a function over the current object state and its relation effect. In turn, the next relation effect $e_{k,t+1} = f_R\left(o_{u_k,t}, o_{v_k,t}, a_k^r\right)$ evaluates a function over each relation and its corresponding nodes. Functions f_O and f_R are normally approximated using artificial neural networks.

This model, as opposed to GNG, does not include the unsupervised construction of a graph structure from raw data. However, PGN is able to predict the dynamics of a system, which is something that GNG cannot achieve. Therefore, the direction for this work focuses on analyzing GNG initially as a structure representation method and then include PGN as a structure predictor method. Thus, a complete forward simulation of the deformation can be obtained which is trained end-to-end from real world observations.

C. Robotic Manipulation Guided by Shape Tracking

Robotic manipulation presents many issues which are inherently dependent on the quality of the representation of the object to manipulate. In the case of non-rigid objects, the situation is complicated by the need to track the shape of the object in real-time to ensure the safety and accuracy of the manipulation. In the initial planning stages, it is useful to have a representation of the object which is as complete as possible in order to select the contact points and manipulator paths that will optimize the manipulation task while ensuring the object's stability [2]. The main criteria for stability is the ability to compensate all internal and external forces, a situation known as *force closure* [9], [10]. In the case of non-rigid objects, this condition must be verified throughout the manipulation to account for changes in the shape of the object. Simulation-based planning approaches such as [11], [12] provide good results, but they are computationally expensive, which makes it difficult to adapt to unexpected changes in the object's shape or behaviour. Other approaches rely on heuristics such as *diminishing rigidity* [13] to estimate the behaviour of an unknown object, allowing for real-time control without requiring expensive simulations.

IV. Feasibility Analysis

In addition to the accuracy of the sensing setup and the ability of the model to capture the behavior of the object, robotic manipulation of non-rigid objects imposes constraints on the computational complexity of the selected representation, especially if real-time control is desired. This requirement is even greater if the system is to be integrated into a mobile robotic platform with limited computing power and battery life. Based on our recent experiments in [2], it was found that a processing speed of about two frames per second was sufficient to match the speed at which the Barrett hand [14] (considered in this research) moves, which is thus used as the target for "real-time" processing.

A. Deformable Object Representation

GNG, as described in section III is analyzed in the context of structure representation. Within this scope, it is observed

TABLE I: GNG processing time for various inputs.

Input	Data Points	Compression Ratio	CPU Time(s)
RGB 480x640	370,200	92:1	168.89
RGB 320x240	76,800	20:1	33.22
RGB 128x128	16384	8:1	6.35
RGB 64x64	4096	3:1	1.45
Depth frame 1	3769	125:1	27.10
Depth frame 2	3766	125:1	1.27
Depth frame 10	3609	120:1	1.13
Depth frame 30	3279	109:1	0.94

that GNG can properly describe unstructured noisy data such as the pointclouds obtained from a commercial Kinect sensor. This idea is supported by study cases such as object segmentation in complex environments [15] and their subsequent shape tracking [16]. While it may be said that GNG is a model that can represent changing structures, our experiments have shown that is not well optimized to perform real-time processing. Table I shows the time to reach a quantization error of 2 pixel units for RGB images (such as Fig. 1) and 5mm for depth filtered pointclouds (as in Fig. 2). In order to respect time constraints, still images require increasing the number of GNG nodes, and therefore reducing the compression ratio, to achieve a similar performance as if the model was reused between frames [16]. In both cases, reaching an average processing speed of two frames per second would require a very small number of input points, thus significantly degrading the accuracy of 3D shape estimation. Such large requirements in terms of computing time and power deter from using GNG as an online model for real-time control. However, there is evidence [6] suggesting that an optimized version of GNG can handle sequences of pointclouds in real time.

Early results with PGNs demonstrate how this model is able to predict the physics of complex scenes in which liquids, deformable objects and collisions are present [8]. However, this model is trained mostly using simulation engines, so its ability to work based on real sensor data is not yet fully explored. Therefore, it is justified to consider that the robust structure representation obtained from GNG could be beneficial to clean and compress the raw data that would then be use by PGN. This combination of approaches allows to build a system with the ability to predict the physics of deformable objects from real world observations.

B. Robotic Manipulation Guided by Shape Tracking

The safe manipulation and reshaping of deformable objects with initially unknown shape and elasticity must rely on some form of real-time tracking of the objects shape in order to ensure the success of the operation and to react to changes in the object behaviour. In the absence of depth information, the contour of the object can be tracked through color-based approaches, such as in [1], [2], [17]. While this leads to fast and efficient shape tracking, it constrains the manipulation to objects with a high contrast and relatively even color. Moreover, recent experiments based on the work performed in [2] have shown that while taking a 2D view is suitable for many objects, it may lead to failure in cases where the manipulated object does not have a straight profile in the direction normal to the 2D plane (Fig. 1). The use of 3D sensing and modelling allows for more accurate planning of the vertical position of the fingers and enables the tracking of objects with more varied colors.

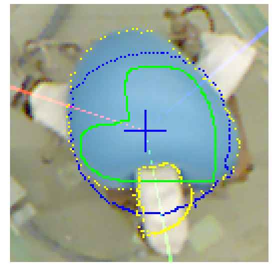

Fig. 1: A 2D view is not suitable for handling spherical non-rigid objects, as the fingers can easily slip in the vertical direction. The manipulation shown is an attempt to deform a balloon from its initial shape (blue contour) to a target shape (green contour) by applying forces along the radial lines. The current detected contour is in yellow.

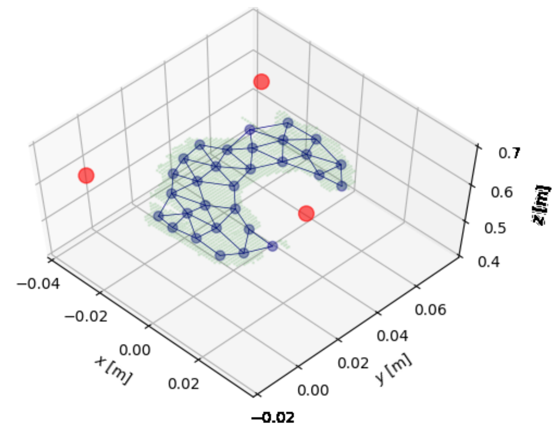

Fig. 2: Deformable rectangular surface (green) manipulated by a three fingered robotic hand (red) and its corresponding graph representation (blue) obtained using GNG.

V. Conclusion

This work explores the ability of GNG to represent complex changing structures in the context of non-rigid object manipulation. It was found that its most important limitations for this task were in terms of the computing power required to perform real-time tracking with this architecture. Moreover, it highlights the need for highly accurate sensing and modelling techniques with low computational cost to support the real-time tracking of non-rigid objects during robotic manipulation. A promising direction for overcoming this issue is to optimize GNG so that it can be used in conjunction with PGN to predict the shape of deformable objects from real world observations. Future works expanding on this topic will experiment with higher accuracy sensors with real-time acquisition capability.

References

[1] F. Hui, P. Payeur, and A.-M. Cretu, "Visual Tracking of Deformation and Classification of Non-Rigid Objects with Robot Hand Probing," *Robotics*, vol. 6, no. 1, p. 5, Mar. 2017. [Online]. Available: http://www.mdpi.com/2218-6581/6/1/5

[2] F. Nadon and P. Payeur, "Automatic Selection of Grasping Points for Shape Control of Non-Rigid Objects," in *Proceedings of the IEEE International Symposium on Robotic and Sensors Environments (ROSE 2019)*, Ottawa, ON, June 2019, pp. 170–176.

[3] F. Nadon, A. 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. [Online]. Available: http://www.mdpi.com/2218-6581/7/4/74

[4] 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, June 2018. [Online]. Available: https://doi.org/10.1177/0278364918779698

[5] B. Fritzke, "A Growing Neural Gas Network Learns Topologies," in *Advances in Neural Information Processing Systems 7*. MIT Press, 1995, pp. 625–632.

[6] S. Orts-Escolano, J. Garcia-Rodriguez, V. Morell, M. Cazorla, M. Saval, and J. Azorin, "Processing point cloud sequences with Growing Neural Gas," in *2015 International Joint Conference on Neural Networks (IJCNN)*, July 2015, pp. 1–8.

[7] D. Mrowca, C. Zhuang, E. Wang, N. Haber, L. F. Fei-Fei, J. Tenenbaum, and D. L. Yamins, "Flexible neural representation for physics prediction," in *Advances in Neural Information Processing Systems 31*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds. Montreal, Canada: Curran Associates, Inc., 2018, pp. 8799–8810.

[8] Y. Li, J. Wu, R. Tedrake, J. B. Tenenbaum, and A. Torralba, "Learning Particle Dynamics for Manipulating Rigid Bodies, Deformable Objects, and Fluids," in *International Conference on Learning Representations*, 2019.

[9] A. Morales, G. Recatalá, P. J. Sanz, and A. P. Del Pobil, "Heuristic vision-based computation of planar antipodal grasps on unknown objects," in *Proceedings - IEEE International Conference on Robotics and Automation*, vol. 1, 2001, pp. 583–588.

[10] K. Gopalakrishnan and K. Goldberg, "D-space and Deform Closure Grasps of Deformable Parts," *The International Journal of Robotics Research*, vol. 24, no. 11, pp. 899–910, 2005. [Online]. Available: http://journals.scholarsportal.info/detailsundefined

[11] S. Patil and R. Alterovitz, "Toward automated tissue retraction in robot-assisted surgery," in *IEEE International Conference on Robotics and Automation*, May 2010, pp. 2088–2094.

[12] L. Zaidi, J. A. Corrales, B. C. Bouzgarrou, Y. Mezouar, and L. Sabourin, "Model-based strategy for grasping 3D deformable objects using a multi-fingered robotic hand," *Robotics and Autonomous Systems*, vol. 95, pp. 196–206, Sept. 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0921889016308089

[13] D. Berenson, "Manipulation of deformable objects without modeling and simulating deformation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nov. 2013, pp. 4525–4532.

[14] Barrett Advanced Robotics, "Barrett hand," 2019. [Online]. Available: https://advanced.barrett.com/barretthand/

[15] G. Vachkov and H. Ishihara, "Data Partitioning and Image Segmentation by Use of Information Compression and Graph Structures," in *2009 11th International Conference on Computer Modelling and Simulation*, Mar. 2009, pp. 65–70.

[16] J. Garcıa-Rodrıguez, E. Domınguez, A. Angelopoulou, A. Psarrou, F. J. Mora-Gimeno, S. Orts, and J. M. Garcıa-Chamizo, "Video and Image Processing with Self-Organizing Neural Networks," in *Advances in Computational Intelligence*, ser. Lecture Notes in Computer Science, J. Cabestany, I. Rojas, and G. Joya, Eds. Springer Berlin Heidelberg, 2011, pp. 98–104.

[17] A. Cretu, E. M. Petriu, P. Payeur, and F. F. Khalil, "Deformable Object Segmentation and Contour Tracking in Image Sequences Using Unsupervised Networks," in *Canadian Conference on Computer and Robot Vision*, May 2010, pp. 277–284.