A Latent and Semantic Framework for Deformable Object Representation

Peng Zhou[1], *Student Member, IEEE*, Jihong Zhu[2], *Member, IEEE*, Shengzeng Huo[1], *Student Member, IEEE*, and David Navarro-Alarcon[1], *Senior Member, IEEE*

***Abstract*—Soft object manipulation has recently gained popularity within the robotics community due to its potential applications in many economically important areas. Although great progress has been recently achieved in these types of tasks, most state-of-the-art methods are case-specific; They can only be used to perform a single deformation task (e.g. bending), as their shape representation algorithms typically rely on "hard-coded" features. In this paper, we present a new latent and semantic framework for soft object representation. Our new method introduces internal latent representation layers between low-level geometric feature extraction and high-level semantic shape analysis. The proposed latent framework allows to explore the semantic deformation process in the built latent space and generate deformation planning with a geodesic path-based algorithm for deformable objects. To validate this new framework, we report a corresponding experimental study.**

***Index Terms*—Representation Learning; Shape Deformation Planning; Latent Space;**

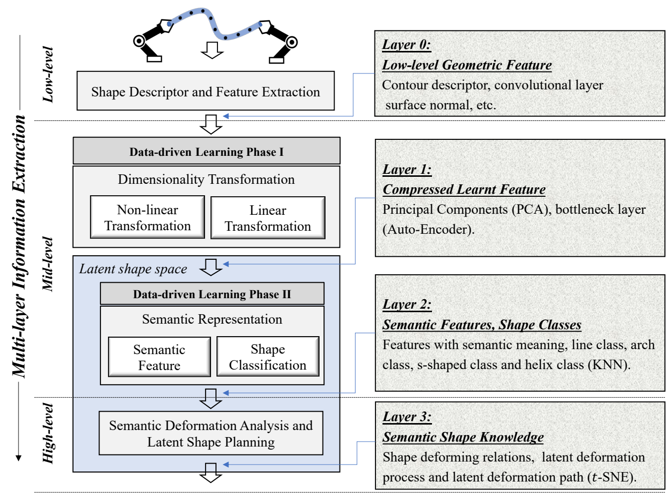

Fig. 1. Conceptual representation of the proposed framework that fully describes and represents the soft objects from four layers, namely, the low-level geometric feature layer, compressed learnt feature layer, semantic features and shape classes layer, and semantic shape knowledge layer.

I. Introduction

RECENT studies have shown that the manipulation of soft objects is crucial and indispensable to achieve high autonomy in robots [1]. Although great progress has been recently achieved, the *feedback* manipulation of soft objects is still a challenging research question. The implementation of these types of advanced manipulation capabilities is complicated by various issues. Amongst the most important is the difficulty in characterizing the feedback shape of a soft object. Our aim in this work is to develop new data-driven methods that can quantitatively describe deformable shapes.

Classical methods are based on geometric features e.g. angles, curvatures, catenaries [2]–[5]; Its disadvantage is that they are case-specific, thus, can only be used to perform a single shaping action. Some works have addressed this issue by developing generic representations that only require sensory data. For example, [6], [7], and [8] characterize shapes using Fourier series and feature histograms; These methods, however, create very large feature vectors, which may not be the most efficient feedback metric. A more effective solution is to automatically compute generic feedback features (e.g. as in direct visual servoing [9], [10]) and combine them with dimension reduction techniques, as in e.g. [11], [12]. Data-driven based shape analyses [13], [14] have gained in popularity as it offers a useful alternative to model-based approaches. An increasing amount of research have focus on different-level segmentation and shape classifications (see [15], [16], and [17]). However, these methods purely depend on the designed end-to-end pipeline which ignores the semantic meaning of internal features and thus failing to interpret the entire analytical process. Therefore, latest applications started to examine attribute-based approaches, such as binary attributes [18], relative attributes [19], and semantic image color palette editing [20]. Several works [21], [22] further combine shape analysis and semantic attributes for a in-depth deformation analysis.

Latent space approaches have recently achieved many successful results in robotic manipulations (e.g. [23]), due to its capability to encode high-dimensional data into a meaningful internal representation. By using concise low-dimensional latent variables and highly flexible generators, a latent space allows us to generate new data samples on data space. In this manner, a deformation planning problem of soft objects can be solved in a novel way by constructing a feasible sequence of deformable shapes in latent space. However, many works [24] have adopted a linear interpolation in remapping the latent variables back to data space, which could cause serious distortions on the generated samples for a shape planning scenario. For example, consider a generator g and a latent

[1] are with Robotics and Machine Intelligence Laboratory, The Hong Kong Polytechnic University (e-mail: jeffery.zhou@connect.polyu.hk; kyle-sz.huo@connect.polyu.hk; david.navarro-alarcon@polyu.edu.hk)

[2] is with Delft University of Technology, Mekelweg 2, 2628CD, The Netherlands and Honda Research Institute Europe, Offenbach, Germany. (e-mail: j.zhu-3@tudelft.nl)

variable $\mathbf{z}$ with two infinitesimal shifts δ_1 and δ_2, then the distance with Taylor's expansion [25] is formulated by:

$$\|g(\mathbf{z_0}+\delta_1)-g(\mathbf{z_0}+\delta_2)\|^2 = (\Delta_{12})^\top \left(\mathbf{J}_{\mathbf{z_0}}^\top \mathbf{J}_{\mathbf{z_0}}\right)(\Delta_{12}) \quad (1)$$

for $\mathbf{J}_{\mathbf{z_0}} = \left.\frac{\partial g}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{z_0}}$ and $\Delta_{12} = \delta_1 - \delta_2$, which indicates that the normal distance in $\mathcal{Z}$ space changes locally as it is determined by the local Jacobian. Consequently, seeking the shortest curve along a curved surface, a manifold, manifold is a more reasonable way to compute the interpolation and generate undistorted samples.

As a feasible solution to these problems, we present a general data-driven representation framework for soft object representation depicted in Fig. 1, which is composed of three layers: A low-level soft object geometric shape processing, a mid-level data-driven representation learning, and a high-level semantic shape analysis. This abstract's main contributions are summarized as follows:

- An effective representation framework for soft object analysis during manipulation tasks.
- A novel semantic approach for deformation process in a latent space.
- A solution for shape planning with a geodesic path-based interpolation algorithm in the latent space.

II. Methods

With dimensionality transformations, we embed the low-level features of the collected shapes in a low-dimensional latent shape space as shown in Fig. 2. The learnt latent space is an immersed manifold (see [25]) and can avoid collapses during training with some tricks [26].

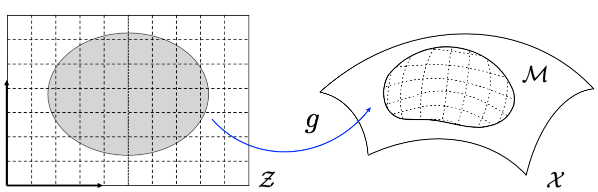

Fig. 2. Conceptual representation of a generator g as a mapping from low-dimensional latent space $\mathcal{Z}$ into a manifold in input data space $\mathcal{X}$.

A. Semantic Analysis

semantic deformation analysis is introduced to establish a mapping from soft object deformations to latent variables in latent shape space. Intuitively, if the dimensionality reduction technique is invertible, then we can explore deformation rules between different shape classes by observing the latent shape space. With performing classification on the latent variables encoded from collected shapes, this path will travel through different spaces enclosed by pre-defined shape classes, thus revealing some rules of shape deformations in real-world applications.

B. Geodesic Path on Manifolds

Through the mapping g, for each point $z \in \mathcal{Z}$, the Riemannian metric is defined as below:

$$G(z) = J_g(z)^T J_g(z) \quad (2)$$

Therefore, the inner product of two tangent vectors $u, v \in T_z Z$ is $\langle u, v\rangle = u^T G(z) v$. Consider a smooth curve in the latent space $\gamma_t : [a, b] \to \mathcal{Z}$, then it has length $\int_a^b \|\dot{\gamma}_t\| \,\mathrm{d}t$, where $\dot{\gamma}_t = \mathrm{d}\gamma_t/\mathrm{d}t$ denotes the velocity of the curve. The length of this curve L lying on the manifold ($g \circ \gamma(t) \in M$) is computed as:

$$L[g(\gamma_t)] = \int_a^b \|\dot{g}(\gamma_t)\| \,\mathrm{d}t = \int_a^b \|\mathbf{J}_{\gamma_t}\dot{\gamma}_t\| \,\mathrm{d}t \quad (3)$$

where $\mathbf{J}_{\gamma_t} = \left.\frac{\partial g}{\partial \mathbf{z}}\right|_{\mathbf{z}=\gamma_t}$ and the last step follows from Taylor's Theorem, which implies the length of a curve γ_t along the surface can be computed directly in the latent space using below defined norm:

$$\|\mathbf{J}_\gamma \dot{\gamma}\| = \sqrt{\dot{\gamma}^\top \left(\mathbf{J}_\gamma^\top \mathbf{J}_\gamma\right)\dot{\gamma}} = \sqrt{\dot{\gamma}^\top \mathbf{M}_\gamma \dot{\gamma}} \quad (4)$$

Here, $\mathbf{M}_\gamma = \mathbf{J}_\gamma^\top \mathbf{J}_\gamma$ and it is a symmetric and positive definite matrix, that gives rise to the definition of a Riemannian metric for each point z in the latent space $\mathcal{Z}$. The arc length with metric $\mathbf{M}_\gamma$ can be re-expressed as:

$$L(\gamma) = \int_a^b \sqrt{\dot{\gamma}_t^\top \mathbf{M}_{\gamma_t}\dot{\gamma}_t}\mathrm{d}t \quad (5)$$

To obtain a geodesic curve, the curve length $L(\gamma)$ is locally minimized through an energy functional $E(\gamma)$ defined as:

$$E(\gamma) = \frac{1}{2}\int_a^b \dot{\gamma}(t)^T G_{\gamma(t)} \dot{\gamma}(t) dt \quad (6)$$

In Riemannian geometry, taking a variation of the geodesic energy function can lead to the Euler-Lagrange equation calculated as:

$$\frac{d^2\gamma^\mu}{dt^2} = -\Gamma^\mu_{\alpha\beta}\frac{d\gamma^\alpha}{dt}\frac{d\gamma^\beta}{dt} \quad (7)$$

where $\Gamma^\mu_{\alpha\beta}$ is the Christoffel symbol of the metric G, which is defined as:

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2}G^{v\mu}\left(\frac{\partial G_{v\beta}}{\partial \gamma^\alpha} + \frac{\partial G_{v\alpha}}{\partial \gamma^\beta} - \frac{\partial G_{\alpha\beta}}{\partial x^\mu}\right) \quad (8)$$

where $G^{v\mu}$ is the inverse of $G_{v\mu}$. To avoid expensive calculations of the Christoffel symbols, instead of getting the entire geodesic path, we only calculate out few discrete points along on the geodesic path with discrete geodesic energy (6).

Formally, consider a discretized curve $\gamma : [0, 1] \to \mathcal{Z}$ denoted by a series of coordinates $z_0, z_1, \ldots, z_N \in \mathcal{Z}$. With T time steps, a sequence of discrete time intervals, $\delta t = 1/N$, is generated, which matches a discretized points on the manifold $\mathcal{M}$, $g(z_i)$. With a small shift, the velocity of $g(z_i)$ can be formulated by $v_i = (g(z_{i+1}) - g(z_i))/\delta t$. Similarly, the energy of this curve can be given:

$$E_{z_i} = \frac{1}{2}\sum_{i=0}^{N}\frac{1}{\delta t}\|g(z_{i+1}) - g(z_i)\|^2 \quad (9)$$

Fixing the first and last points, z_0 and z_N, as the beginning and ending points of the geodesic curve, minimizing this energy function would result in an approximated geodesic path, which can be obtained by performing a gradient descent algorithm for

$z_1, \ldots, z_{N-1}$, along this curve. The gradient at z_i is computed as:

$$\nabla_{z_i} E = -\frac{1}{\delta t} J_g^T (z_i) \left(g(z_{i+1}) - 2g(z_i) + g(z_{i-1})\right) \quad (10)$$

Based on this geodesic path, we can perform a shape planning with Algorithm 1.

Algorithm 1: Latent Shape Planning

Input: Current shape $\boldsymbol{x}_0$, target shape $\boldsymbol{x}_*$, iteration N, encoder h, decoder g

Output: Planned deformation trace $\mathcal{D}_p$

Compute the coordinates using $(\boldsymbol{z}_0, \boldsymbol{z}_*) = h(\boldsymbol{x}_0, \boldsymbol{x}_*)$

$\mathcal{S}_{low} = \{\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_*\} = \text{ShortestPath}(\boldsymbol{z}_0, \boldsymbol{z}_*)$

$\mathcal{S}_{high} = g(\mathcal{S}_{low})$

$\mathcal{G}_{low} = \{\boldsymbol{z}_0, \boldsymbol{z}'_1, \ldots, \boldsymbol{z}_*\} = \text{Interpolation}(\boldsymbol{z}_0, \boldsymbol{z}_*, N)$

if g is not linear **then**

  Update $\mathcal{G}_{low}$ with the geodesic path.

end

$\mathcal{G}_{high} = g(\mathcal{G}_{low})$

$\mathcal{D}_p = \{\text{Visualizer}(\mathcal{S}_{high}), \text{Visualizer}(\mathcal{G}_{high})\}$

return $\mathcal{D}_p$

III. Results

As shown in Fig. 3, a foam bar with markers were used to collect deformed shapes. During the collection, the Prime 13 motion tracking system was used to track the position of each marker mounted on the its surface in 30 FPS.

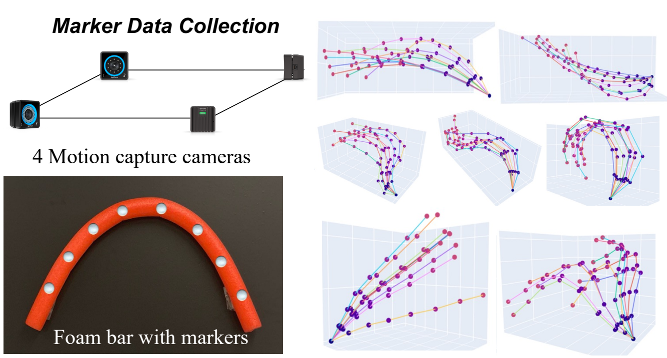

Fig. 3. Left: Experimental setups of the shape data collection; Right: the collected shape categories.

1) Semantic Deformation: As Fig. 4 (a) shows, all the shapes collected from the foam bar (dataset #1) are encoded into a 3D latent shape space with t-SNE built from AE. In this space, the deformation path generated from gesture controls is represented as a red curve and different shape categories of dataset #1 were organized with *mesh3D* from *Plotly* and rendered with different colors according to the prediction of kNN. The beginning shape located at the position of the triangle marker, and then the foam bar started from the line category area denoted by a blue color. As the shape deformed, the current point moved continuously toward the positive arch category denoted by the yellow color in area #1, and then moved to the negative S-shaped category denoted by the cyan color in area #2. Subsequently, the foam bar went back to the positive arch shape from area #2 which form a identical but inverse path. And so forth, the deformed foam bar ended up with its original shape state. Therefore, the entire trace semantically reflects the entire process of shape deformation in a latent space when manipulating a soft object.

2) Latent Shape Planning: With dimensionality transformations, we embed the low-level features of the collected shapes in a low-dimensional latent shape space. We use Algorithm 1 to perform a shape planning through a generator $(g : \mathcal{Z} \to \mathcal{X})$ to map paths calculated in the latent space into shapes on the generated manifold ($\mathcal{M}$). Fig. 4(b) shows a beginning line and target S-shape of a foam bar. Figs. 4 (c) and (d) show the resulting deformation processes from a geodesic interpolation and shortest path, respectively. We can clearly observed that the geodesic path-based interpolation deformation process is smoother compared with the process with a shortest path. For more results, please refer to [27].

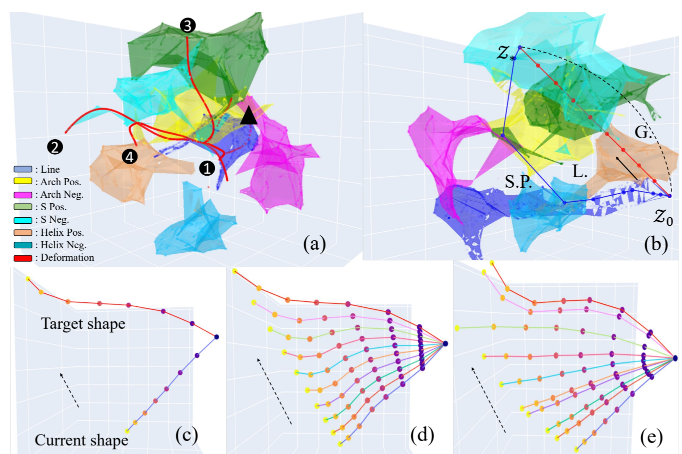

Fig. 4. Visualization of the process of latent shape planning for the foam bar. (a) Deformation trace of the manipulation task with Leap motion in latent shape space; (c) shows the beginning shape and the target shape; figures (d) and (e) present the planned shape deformations; (b) presents their corresponding deformation paths with shape planning algorithm.

IV. Conclusions

In this paper, we present a generic latent representation framework for semantic soft object manipulation tasks. With dimensionality transformations, we embed the shapes of soft objects from the originally high-dimensional shape space into a semantically low-dimensional latent shape space and solve the shape planning with designed geodesic path-based algorithms on the data manifold.

References

[1] H. B. Amor, A. Saxena, *et al.*, "Special issue on autonomous grasping and manipulation," *Auton. Robots*, vol. 36, no. 1-2, pp. 1–3, 2014.

[2] D. Navarro-Alarcon, Y.-H. Liu, *et al.*, "Model-free visually servoed deformation control of elastic objects by robot manipulators," *IEEE Trans. Robot.*, vol. 29, no. 6, pp. 1457–1468, 2013.

[3] D. Navarro-Alarcon, Y.-h. Liu, *et al.*, "On the visual deformation servoing of compliant objects: Uncalibrated control methods and experiments," *Int. J. Robot. Res.*, vol. 33, no. 11, pp. 1462–1480, 2014.

[4] D. Navarro-Alarcon, H. M. Yip, *et al.*, "Automatic 3-d manipulation of soft objects by robotic arms with an adaptive deformation model," *IEEE Trans. Robot.*, vol. 32, no. 2, pp. 429–441, 2016.

[5] M. Laranjeira, C. Dune, and V. Hugel, “Catenary-based visual servoing for tether shape control between underwater vehicles,” *Ocean Engineering*, vol. 200, pp. 1–19, 2020.

[6] D. Navarro-Alarcon *et al.*, “Fourier-based shape servoing: A new feedback method to actively deform soft objects into desired 2D image shapes,” *IEEE Trans. Robot.*, vol. 34, no. 1, pp. 272–1279, 2018.

[7] J. Zhu, B. Navarro, P. Fraisse, A. Crosnier, and A. Cherubini, “Dual-arm robotic manipulation of flexible cables,” in *IEEE/RSJ Int. Conf. on Robots and Intelligent Systems*, 2018, pp. 479–484.

[8] Z. Hu, P. Sun, and J. Pan, “Three-dimensional deformable object manipulation using fast online gaussian process regression,” *IEEE Robot. Autom. Lett.*, vol. 3, no. 2, pp. 979–986, 2018.

[9] C. Collewet and E. Marchand, “Photometric visual servoing,” *IEEE Trans. Robot.*, vol. 27, no. 4, pp. 828–834, 2011.

[10] E. Marchand, “Subspace-based direct visual servoing,” *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2699–2706, 2019.

[11] K. M. Digumarti, B. Trimmer, A. T. Conn, and J. Rossiter, “Quantifying dynamic shapes in soft morphologies,” *Soft Robot.*, 2019.

[12] J. Zhu, D. Navarro-Alarcon, R. Passama, and A. Cherubini, “Vision-based manipulation of deformable and rigid objects using subspace projections of 2d contours,” *arXiv preprint arXiv:2006.09023*, 2020.

[13] K. Xu, V. G. Kim, *et al.*, “Data-driven shape analysis and processing,” in *SIGGRAPH ASIA 2016 Courses*, 2016, pp. 1–38.

[14] H. Zhang, A. Sheffer, D. Cohen-Or, Q. Zhou, O. Van Kaick, and A. Tagliasacchi, “Deformation-driven shape correspondence,” in *Computer Graphics Forum*, vol. 27, no. 5, 2008, pp. 1431–1439.

[15] A. Golovinskiy and T. Funkhouser, “Consistent segmentation of 3d models,” *Computers Graphics*, vol. 33, no. 3, pp. 262–269, 2009.

[16] O. Sidi, O. van Kaick, Y. Kleiman, H. Zhang, and D. Cohen-Or, “Unsupervised co-segmentation of a set of shapes via descriptor-space spectral clustering,” in *Proc. SIGGRAPH Asia Conf.*, 2011, pp. 1–10.

[17] O. Van Kaick, K. Xu, H. Zhang, Y. Wang, S. Sun, A. Shamir, and D. Cohen-Or, “Co-hierarchical analysis of shape structures,” *ACM Trans. Graphic.*, vol. 32, no. 4, pp. 1–10, 2013.

[18] L. Tao, L. Yuan, and J. Sun, “Skyfinder: attribute-based sky image search,” *ACM Trans. Graphic.*, vol. 28, no. 3, pp. 1–5, 2009.

[19] D. Parikh and K. Grauman, “Relative attributes,” in *Int. Conf. Comput. Vis.*, 2011, pp. 503–510.

[20] P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays, “Transient attributes for high-level understanding and editing of outdoor scenes,” *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–11, 2014.

[21] G. Leifman, R. Meir, and A. Tal, “Semantic-oriented 3d shape retrieval using relevance feedback,” *Visual Comput.*, vol. 21, no. 8-10, pp. 865–875, 2005.

[22] M. Attene *et al.*, “Characterization of 3d shape parts for semantic annotation,” *Comput. Aided Des.*, vol. 41, no. 10, pp. 756–763, 2009.

[23] M. Lippi, P. Poklukar, M. C. Welle, A. Varava, H. Yin, A. Marino, and D. Kragic, “Latent space roadmap for visual action planning of deformable and rigid object manipulation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2020.

[24] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas, “Learning representations and generative models for 3d point clouds,” in *Int. conf. Machine Learning*. PMLR, 2018, pp. 40–49.

[25] G. Arvanitidis *et al.*, “Latent space oddity: On the curvature of deep generative models,” in *Int. Conf. Learn Represent.*, 2018.

[26] H. Shao, A. Kumar, and P. Thomas Fletcher, “The riemannian geometry of deep generative models,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 315–323.

[27] P. Zhou, J. Zhu, S. Huo, and D. Navarro-Alarcon, “Lasesom: A latent and semantic representation framework for soft object manipulation,” *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5381–5388, 2021.