2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2025)
October 19-25, 2025. Hangzhou, China

2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 979-8-3315-4393-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/IROS60139.2025.11246764

Multimodal Deformation Estimation of Soft Pneumatic Gripper During Operation

Changheng Cai[1], Fei Xiao[1,2], Marcellus Vanza[1,2], Taoyang Wang[1]
Fangbing Zhou[1], Xuanyang Xu[1,2], Jian Zhu[1,2†] and Yuan Gao[1,2†]

***Abstract*—Soft pneumatic robots are gaining significant attention due to their compliance and adaptability in unstructured environments. While emerging dual-chamber soft pneumatic robots can achieve complex 3D deformations beyond conventional single-axis bending, real-time proprioception remains challenging due to the high degrees of freedom and the complex interaction between chambers. To address this issue, we propose a multimodal learning-based sensing method that combines camera and inertial measurement unit (IMU) and then extracts full-body shape information using deep learning algorithms. Our method enhances proprioception by effectively processing high-dimensional sensor data, providing real-time feedback on the gripper shape. The average error of key points was found to be 3.67mm (Var 8.39) for our method, while the error was 4.36mm (Var 10.47) when a camera was used alone, or 9.32mm (Var 21.29) when an IMU was used alone. Our multimodal learning-based shape estimation and reconstruction empower soft pneumatic grippers to be seamlessly integrated into the embodied AI framework, significantly improving their reliability and thus paving the way for applications in service robotics, ehabilitation robotics, and human-robot collaborations.**

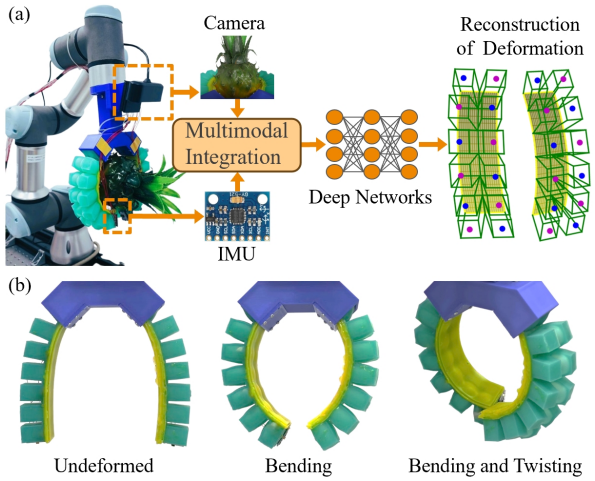

Fig. 1. (a) To effectively predict the deformation of soft grippers, the sensor signals collected from camera and IMU are fused and processed by a learning-based method. (b) Dual-chamber soft pneumatic grippers can perform bending and twisting.

I. Introduction

SOFT pneumatic robots offer unique advantages in complex manipulation tasks [1], [2]. Their high degrees of freedom and inherent material compliance enhance safety and reliability, particularly in delicate operations. [3], [4]. These features also enable soft grippers to achieve more stable and diverse grasps, increasing the contact surface area [5]. Although emerging dual-chamber soft pneumatic robots can perform complex 3D motions [6]–[8] beyond conventional single-axis bending [9], their deployment remains hindered by the challenge of real-time full-body shape sensing, a critical requirement for precise closed-loop control in dynamic tasks [10]. Without reliable shape estimation or feedback, control strategies are less effective, which limits the practical use of soft robots in real-world applications.

However, compared to other robotic systems, observing the shape of soft robots is more costly. Many previous studies have utilized exterior cameras to capture the complex deformation behaviors of soft robots for manipulation [11]. Exterior cameras are prone to occlusion, and the gripper's workspace must be restricted to prevent obstruction in the camera's field of view. Only using cameras mounted at the bottom or inside of the gripper without external ones, occlusion of the gripper's end by the grasped object during the grasping state still remains a challenge [12]. Therefore, systems that rely solely on vision for shape capture either provide incomplete information or are limited by the observational posture, which restricts the execution space. Another approach is to use embedded self-sensing sensors, which has achieved some success [13], [14]. But sensor deployment must consider how to effectively monitor deformation without compromising the gripper's functionality, softness, and overall structure. [15]. Moreover, the deformation of soft pneumatic grippers is highly flexible and complex, making it difficult to describe the deformation process using traditional rigid models [16]. Self-sensing systems based on this approach often struggle with data sparsity and fail to accurately capture subtle 3D deformations. In response, researchers have started to leverage advances in deep learning to capture the complex and often nonlinear relationships between sensor observations and the soft robot's state, achieving some success [17].

Building on this, we propose a multimodal fusion method that combines visual sensors and inertial measurement units (IMU) for 3D shape modeling of soft pneumatic robots, utilizes attention mechanisms and LSTM. This multimodal fusion approach effectively addresses the issue of incomplete

*This work is supported by the Start-up Funding from the Chinese University of Hong Kong, Shenzhen (UDF01001987); the Funding of Shenzhen Institute of Artificial Intelligence and Robotics for Society (AC01202201007-02); the Funding of Guangdong Province (2021CX02Z251); the 2023 SZSTI stable support scheme; Guangdong Basic and Applied Basic Research Foundation (2022A1515110787 and 2024A1515012065); and the Shenzhen Science and Technology Program (JSGGKQTD20221101115656029).

[1]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen.
[2]Shenzhen Institute of Artificial Intelligence and Robotics for Society.
†Corresponding authors: Jian Zhu (zhujian@cuhk.edu.cn) and Yuan Gao (gaoyuan@cuhk.edu.cn).

information in a single vision system and the data sparsity problem in a self - sensing system. It showcases enhanced robustness and flexibility when dealing with complex scenarios. Our method generates real-time digital models for a soft pneumatic gripper handling objects of various shapes. The average error of key points was found to be 3.67mm (Var 8.39) when comparing our method's outputs with an external motion capture system, with the average error for using the camera alone being 4.36mm (Var 10.47) and for using the IMU alone being 9.32mm (Var 21.29). This work not only extends the functional boundaries of soft pneumatic grippers but also offers new design insights for low-cost, high-performance robotic manipulation systems.

II. Related Work

This section examines the research background and recent advancements in soft pneumatic grippers. It first outlines types of end effectors and their selection criteria, then reviews vision-based deformation capture methods, and finally analyzes proprioceptive sensing-based approaches.

A. Design of Dual-chamber Soft Pneumatic robots

Soft pneumatic robots have gained attention for their cost-effectiveness, reversibility, controllability, and rapid response. Various designs with specific deformations (e.g., elongation, bending, twisting) have been proposed [25]–[27]. The soft pneumatic two-finger gripper is widely used due to its simple structure and ease of control [9]. However, its single motion mode limits adaptability in diverse tasks. To address this limitation, researchers have introduced parallel chamber designs to enable spatial motion [6]–[8]. By specially designing the shape and layout of air chambers and constraint layers, soft robots can achieve a variety of motion modes [28], [29]. Among these, dual-chamber actuators offer significant advantages in adaptability and manufacturing cost, making them well-suited for precise manipulation and integration with sensing and control functions. In our design, we propose a dual-chamber soft pneumatic actuator with a narrow limiting layer, which can not only bend but also twist (Fig. 1).

B. Deformation Estimation based on Vision

Vision-based deformation capture methods are widely used for high precision and resolution in dynamic motion capture [30]. However, they often require complex hardware and precise initialization, making them costly and less practical for general applications. To reduce these barriers, recent studies have explored monocular RGB or RGB-D cameras [31]–[33].Despite these advancements, vision systems remain susceptible to occlusion, lighting variations, and visual interference—critical challenges in highly interactive scenarios [34]. Some studies have mounted visual sensors directly within pneumatic actuators to address occlusion, but these designs still face challenges with complex modeling and significant bending angles [35]. This study uses a wrist-mounted camera on a robotic arm to capture visual signals, a common setup in robotic manipulation platforms, for estimating object deformation near the gripper.

C. Deformation Estimation based on Self-sensing Sensors

Proprioceptive methods for non-visual deformation capture rely on contact-based sensors to measure soft structure deformation directly, without external equipment [36]. These methods are summarized in Table I. Most proprioceptive technologies capture only 1D or 2D spatial information (e.g., resistive [13], [18], capacitive [14], [15], piezoelectric [19], [20], and optical fiber sensors [21], [22]). Machine learning is often employed to map low-dimensional sensor data to 3D deformations [13], [18], but creating large datasets and deploying sensors without affecting soft actuators are challenging. In contrast, IMU sensors can directly acquire 3D orientation data with minimal interference [23], [24]. This study uses an IMU-based method for high-precision endpoint pose capture, laying a solid foundation for subsequent task control.

III. Methodology

This section describes the methodology for dynamic deformation estimation and reconstruction of a pneumatic soft gripper mounted on a UR3e robotic arm. The system integrates a monocular RGB camera and two IMU sensors to capture deformation data, with an external motion capture system providing ground truth. A novel kinematic chain model characterizes the gripper's deformation, while a multimodal perception system fuses visual and IMU data. A deep learning-based approach processes this data to enable accurate tracking and modeling of the gripper's behavior, ensuring precise real-time monitoring of its shape and pose in complex, dynamic environments.

A. Hardware Setup

We employ a UR3e robotic arm as the carrier for a pneumatic soft gripper. The most important hardware component

TABLE I
Summary of Widely Used Self-sensing Sensors for Measuring Deformation

Sensor Type	Spatial Dimension	Sensing Principle	Key Characteristics
Resistive [13], [18]	1D, 2D	Converts shape or structural changes (e.g., stretch, bend, press) into resistance variations.	Low cost and widely used; prone to significant drift.
Capacitive [14], [15]	1D, 2D	Geometric changes affect the overlap area of capacitive units, causing capacitance variations.	Flexible and fast response; commonly used in wearable electronics.
Piezoelectric [19], [20]	1D, 2D	Deformation of piezoelectric elements alters surface charge density, generating polarization signals.	No external power required; sensitive to dynamic deformation.
Optical Fiber [21], [22]	1D, 2D	Stress alters curvature radius and bending direction, leading to changes optical characteristics of transmitted light.	Compact and precise, but requires expensive signal processing equipment.
IMU [23], [24]	3D	Uses accelerometers, gyroscopes, magnetometers, or their combinations to detect 3D orientation information.	Data drift issues but lightweight and compact.

is the pneumatic finger, which is manufactured as shown in Fig. 2. The three views and dimensions can be found in Fig. 3. We use electromagnetic proportional valves to control the input air pressure.

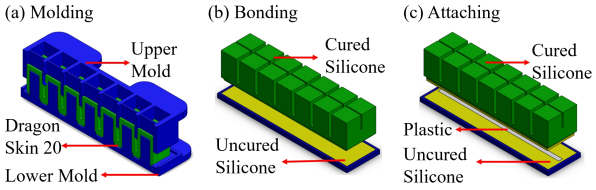

Fig. 2. Fabrication process. (a) Molding: The upper and lower molds are assembled, and silicone (Dragon Skin 20) is poured into the cavity. (b) Bonding: Cured silicone is bonded with uncured silicone to cover the lower part of the finger. (c) Attaching: The cured silicone is attached to a plastic layer, which serves as an expansion limit, with an additional layer of uncured silicone.

A monocular RGB camera is mounted near the base of the robotic arm to capture visual information around the root of the gripper. This configuration is commonly used in grasping systems, eliminating the need for additional cameras and simplifying system setup. Additionally, we use an external stereo motion capture camera to track key points as the ground truth of the gripper shape, which we compare with our estimate. An Inertial Measurement Unit (IMU) sensor is integrated into the gripper's tip, enabling the detection of dynamic motion and 3D pose at the endpoint. Its lightweight design ensures that it does not interfere with the gripper's deformation.

To mathematically represent the complex deformation of the pneumatic gripper, we propose a new kinematic chain model based on physical reality to represent the minimum deformation unit and achieve relative position estimation, as shown in Fig 3.

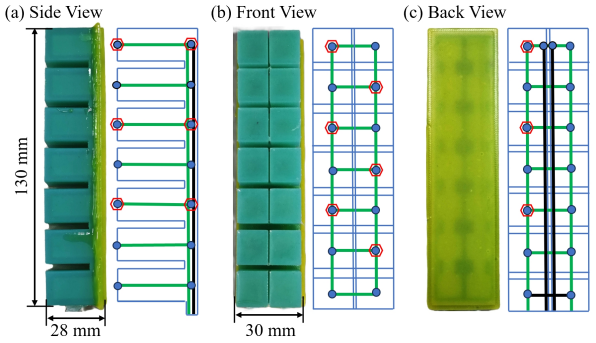

Fig. 3. Geometric prior knowledge. The lateral (a) and dorsal (b) sections of the gripper (denoted by green lines) experience significant deformation due to pneumatic expansion, serving as the primary source for extracting deformation data. The palmar center section (c) (outlined by black lines) is made from a non-extensible material. The deformation of the base can be effectively represented by a set of feature points (marked by red circles), which are fewer compared to all key points (marked by blue points).

B. Perception System Architecture

1. Visual Perception Module

We have developed a dynamic monitoring framework based on a visual system. To accurately extract the gripper's contours and dynamic features from the captured images, we utilize the state-of-the-art SAM2 segmentation and detection model to extract the flexible gripper, which is the target of our processing, from the video stream. Following this, we designed a specialized feature extractor based on Hiera to extract shape-related features from the segmented gripper images. These features include, but are not limited to, the gripper's contour shape, geometric dimensions, and the degree of local deformation. Through this feature extractor, we can transform visual information into quantifiable data thus supporting the subsequent dynamic deformation analysis. This approach allows us to accurately capture the intricate details and variations in the gripper's shape, providing valuable data for further analysis and decision-making processes.

2. IMU Perception Module

The IMU (Inertial Measurement Unit) Perception Module is crucial for accurate pose estimation in dynamic environments and can provide a powerful supplement when the visual image is obscured. The processing of IMU data is mainly to eliminate drift. Startup drift depends on factors such as environmental conditions at startup and the randomness of electrical parameters. Once the startup is completed, this drift will remain at a fixed value, but this fixed value is a random variable, so this component can be described by a random constant (1). Fast drift, on the other hand, manifests as chaotic high-frequency fluctuations superimposed on the aforementioned components, which exhibit minimal or virtually no correlation at two closely spaced time points. It can be abstracted as a white noise process w_{gi}, that is (2) where $\delta(t-\tau)$ is the Dirac function. In summary, the drift can be modeled as (3). We employed initial zero-bias calibration to correct (1), and utilized the Kalman filter to reduce bias (2).

$$\dot{\varepsilon}_{bi} = 0 \quad i = x, y, z. \tag{1}$$

$$\mathbb{E}\left[w_{g_i}(t) w_{g_i}(\tau)\right] = q_{k_i}\delta(t-\tau) \quad i = x, y, z, \tag{2}$$

$$\varepsilon_i(t) = \varepsilon_{bi}(t) + w_{ki}(t) \quad i = x, y, z. \tag{3}$$

3. True shape annotation:

This study uses an external binocular motion capture system (OptiTrack) to provide high-precision annotation of the gripper's real shape during operation. Active marker tracking is employed by attaching reflective markers (infrared spheres) to the gripper and target object. The system calculates the 3D coordinates of these markers in real-time with sub-millimeter accuracy via multi-view triangulation. The captured shape data is stored in structured formats and synchronized with visual features and IMU sequences through timestamps. Considering factors such as feature point distinctiveness (with clear geometric characteristics), stability (maintaining relatively stable spatial relationships in task), even independent distribution (each feature point should provide independent information) , and an appropriate number (not too many to tracking burden, while not too few to accurately reflect deformation), a set of feature points are selected and arranged as red circles shown in Fig. 3.

C. Learning-Based Modeling

In this study, we propose a deep learning-based multi-modal perception modeling approach, with the following

framework in Fig. 4. Our architecture combines attention layer and LSTM. The processed visual features are first passed through a self-attention layer, which is responsible for managing the spatial relationships between camera features and IMU features, dynamically selects salient features while disregarding irrelevant information, and aiding the model in better fusing visual and IMU information. Then the LSTM captures the temporal dependencies of the two perception, utilizes historical information to predict the position of key-points, and maintains the smoothness and continuity of the predicted trajectory. This design offers complementarity, with the Attention aiding in the selection of important features and the LSTM maintaining temporal consistency. The synergy between the two is evident in multi-level feature extraction. Additionally, this design is adaptable, with the Attention able to adjust feature weights according to different scenarios, and the LSTM capable of adapting to various motion patterns.

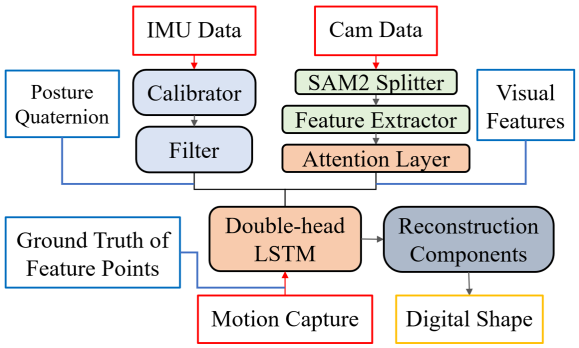

Fig. 4. Utilizing the SAM2 segmenter and feature extractor to process camera data and obtain visual features. Then the processed visual features are transmitted through a self-attention layer. Meanwhile, the IMU signals are processed by a calibrator and a filter. Subsequently, the two types of features are fused and processed by a dual-headed LSTM. Finally, the resultant values are used to reconstruct the shape of the gripper.

For learning rate scheduling, we use ReduceLROnPlateau method to reduce the learning rate when the loss stops decreasing. The loss function we use is the MSE of the feature point locations. An early stopping mechanism is implemented to stop the training process if the validation loss does not improve for 10 consecutive epochs. The optimizer we use is Adam optimizer. In the training loop, all the data batches are iterated every epoch. The best performing model is saved throughout the training process.

D. Deformation Reconstruction and Visualization

We can inferred feature points that effectively characterize the dynamics deformation of a single gripper. Leveraging these feature points and the geometric prior knowledge3, we can deduce the critical positional feature points essential for accurate shape reconstruction as show in Fig. 5.

We define the 12 dorsal structural points as $S_{1.i}$ and the 12 palmar structural points as $S_{2.j}$, where $i, j \in \{1, 2, \ldots, 12\}$; ω is the angle between the vector $\overrightarrow{S_{1.1}S_{1.5}}$ and the vector $\overrightarrow{S_{1.5}S_{1.9}}$; θ is half of ω. R is the rotation matrix for rotating the vector $\overrightarrow{S_{1.5}S_{1.1}}$ by θ. Part (b) shows the solution process of equation 4; oc_1 is the midpoint between $S_{1.1}$ and $S_{1.4}$; ic_1 is the medial counterpart of oc_1, and d is the distance between the dorsal and medial sides. The positions of the palmar structural points to palmar central points show a similar relationship to those of the dorsal structural points to dorsal central points. Part (d) shows the solution process of equation 5.

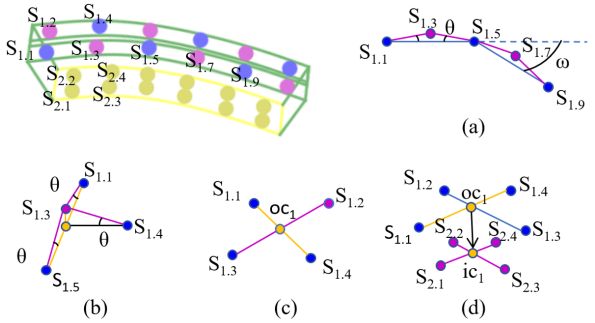

Fig. 5. The blue points in the figure are known points, the pink points are the predicted points based on calculation, and the orange points are the midpoints used in the solving process.

$$S_{1.3} = S_{1.4} + \frac{R\left(\frac{S_{1.1}+S_{1.5}}{2} - S_{1.4}\right)}{\cos\theta}, \tag{4}$$

$$ic_1 = \frac{S_{1.1} + S_{1.4}}{2} + \frac{\overrightarrow{S_{1.3}S_{1.2}} \times \overrightarrow{S_{1.4}S_{1.1}}}{\|\overrightarrow{S_{1.3}S_{1.2}} \times \overrightarrow{S_{1.4}S_{1.1}}\|} \cdot d. \tag{5}$$

The mathematical framework achieves precise reconstruction of the gripper shape by systematically analyzing the deformation patterns captured by feature points. Based on this, we propose a simplified point-line geometric model for visual modeling. This model is highly computationally efficient, making it suitable for real-time visualization and interaction with intelligent agents. Additionally, the positions of the feature points can be imported into finite element modeling software to achieve high-fidelity digital reconstruction, thereby meeting the demands of applications requiring high precision.

IV. Experiments and Results

A. Experimental Design

To comprehensively assess the accuracy of our shape estimation system, we designed 5 different experimental scenarios, as shown in Fig. 6: (1) Unobstructed motion, (2) Rigid object manipulation of a cube, (3) Rigid object manipulation of irregular shapes, (4) Soft object manipulation, and (5) Manipulation with human disturbed situations (shaking, touching, bending). These scenarios were carefully selected to evaluate the system's performance under various operational conditions and external influences. In addition, to assess the contribution of multimodal data integration, we conducted comparative experiments using models trained with a single perception modality under the same hardware configuration. This comparative analysis not only highlights the performance advantages of our multimodal fusion framework but also provides valuable insights into the individual contributions of different perception modalities to soft gripper deformation estimation. We analyzed the temporal consistency of the shape estimation results to evaluate the

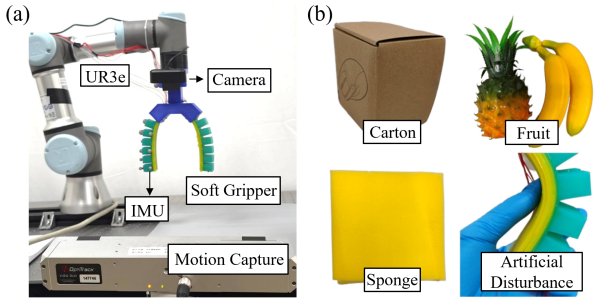

Fig. 6. Experimental setup. (a) A motion capture system collects test data during the gripper operation. (b) Five manipulation scenarios are tested: cubic rigid objects, irregular objects, soft objects, artificially disturbed motion, and unobstructed motion.

system's stability over time. In these experimental scenarios, we systematically compared the inferred shape feature points with the ground truth values collected by a motion capture system. To further quantify the system's performance, we employed a variety of evaluation metrics, including mean average distance, chamfer distance, max distance and variance of these distances, to assess both the average and time-varying performance across different tasks. These metrics were calculated for each scenario to provide a comprehensive understanding of the system's accuracy and robustness under diverse conditions.

B. Results

We tested and compared the keypoint coordinates obtained through model inference with those from an external high-precision motion capture system to measure the model's accuracy. The metrics evaluated included the average distance of keypoints, the Chamfer distance of point sets, the maximum distance of keypoint errors, and the variances of these measures. These results are presented in Fig. 7, Table III and II.

The IMU provides only the end-effector orientation, limiting accurate estimation of the manipulator's overall shape. Although using visual signals alone slightly improves accuracy, robustness is insufficient in scenarios with occlusions or lighting changes. The fusion of these two sensing modalities

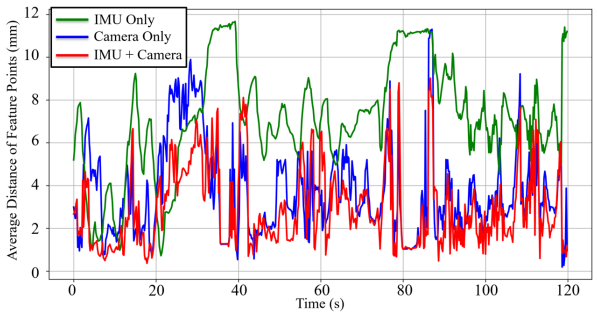

Fig. 7. This figure illustrates the temporal variation of prediction error, measured as the distance between keypoints and ground truth, in a cube manipulation task across three modalities: IMU Only, Camera Only, and IMU + Camera. The results indicate that the combined modality (IMU + Camera) achieves the highest accuracy, followed by the Camera Only modality, and finally the IMU Only modality.

effectively mitigates performance degradation caused by data distortion or noise. As illustrated in the Table III, both the error values and their variances are reduced across various metrics. As shown in the Fig. 7, in a typical cube manipulation task, the prediction accuracy using the IMU modality alone is lower than that using the camera alone, which in turn is lower than the accuracy achieved by combining the IMU and camera modalities during the majority of the time period.

Through a comparison with ground truth, we demonstrate the accuracy of the model in shape estimation across various environments. In all cases, the average distance between the ground truth and the estimated shape keypoints, as well as the chamfer distance, are close, indicating that our feature points are reasonably set with minimal confusion. In simple environments, such as tasks without obstructions, the model shows high accuracy with an average distance of $2.27mm$. In more complex environments, including irregular objects, soft materials, and unobstructed motion, the model's errors stay at low range. When subjected to external disturbances, such as high-frequency touches applied to the actuators, the model's accuracy declines, resulting in an error of approximately 6.45 mm. Despite this, the overall performance in all conditions still outperforms traditional methods. The error when manipulating soft objects is lower than when handling irregular objects, which, in turn, is better than with cubic objects, suggesting that the more compliant and rounded the object, the smaller the estimation error by the model.

We visualized the reconstruction of feature points to intuitively assess their accuracy. As shown in Fig. 8, we analyzed the deformations of a pneumatic actuator under various air pressure inputs and compared the actual photographic projections with model-predicted projections. The results demonstrate that our reconstruction model accurately reproduced the deformations for both single-channel and dual-channel inflation.

V. Conclusion and Future Works

In this paper, we propose a novel multimodal framework for the real-time estimation and modeling of deformations in soft pneumatic actuators. Our method employs a deeplearning-based pipeline that integrates data from cameras and IMUs, leveraging multiple signal processing techniques and models to fuse multimodal inputs into the pose estimation of geometric feature points. This effectively establishes a mapping from sensor signals to the shape of the soft manipulator. We validate the effectiveness of our method through extensive physical experiments, achieving precise proprioception across a wide range of payload types and task scenarios. The average prediction error on feature point set across various scenarios is 3.67mm (Var 8.39). Our approach provides a promising framework for perceiving soft manipulators with complex deformations.

Although this study focuses exclusively on estimating the deformation of pneumatic soft grippers using multimodal deep networks, its potential applications extend far beyond this specific domain. In the future, this work can bring numerous operational benefits to various soft actuator tasks. For example, it can detect whether an object being grasped has a tendency to slip or if its orientation is skewed, thereby

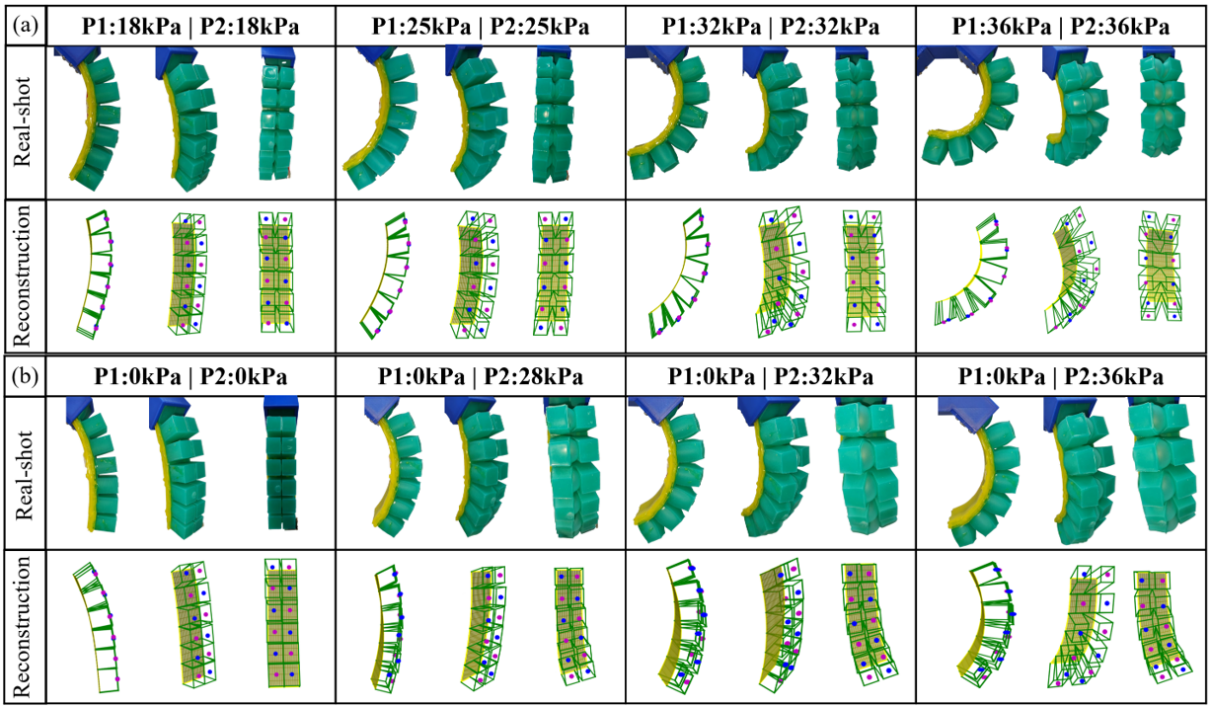

Fig. 8. The comparison between the real-shot and reconstructed images is shown from multiple perspectives. The pressures of the dual-chamber actuators indicated as P1 and P2. The reconstructed images are obtained from the pipeline shown in Fig. 4.

TABLE II
PERFORMANCE METRICS FOR DIFFERENT SCENARIOS

Scenarios	Test volume	Average distance (Var)	Chamfer distance (Var)	Max distance (Var)
No obstruction	162	2.27mm (4.25)	2.26mm (4.18)	25.24mm (24.08)
Cubic rigid object	216	3.41mm (7.82)	3.41mm (7.82)	18.86mm (21.77)
Irregular objects	215	2.78mm (3.27)	2.78mm (3.27)	19.45mm (18.11)
Soft objects	200	2.52mm (2.73)	2.52mm (2.73)	20.50mm (11.93)
Artificial disturbed	253	6.45mm (20.34)	6.33mm (17.27)	61.63mm (107.73)
Averages	1046	3.67mm (8.39)	3.64mm (7.63)	30.63mm (40.29)

TABLE III
PERFORMANCE METRICS FOR DIFFERENT PERCEPTIONS

Perception	Test volume	Average distance (Var)	Chamfer distance (Var)	Max distance (Var)
Camera Only	1046	4.36mm (10.47)	4.28mm (8.29)	70.84mm (147.83)
IMU Only	1046	9.32mm (21.29)	8.79mm (12.89)	97.58mm (637.26)
Camera + IMU	1046	3.67mm (8.39)	3.64mm (7.63)	30.63mm (40.29)

enabling adjustments to the object's posture. These capabilities will significantly enhance downstream robotic tasks. Moreover, the realization of deformation perception in soft manipulators paves the way for the development of sensor-based closed-loop control algorithms. It also facilitates the integration of soft actuators into contemporary intelligent agent frameworks through precise digital modeling.

REFERENCES

[1] N. Obayashi, D. Howard, K. L. Walker, J. Jørgensen, M. Gepner, D. Sameoto, A. Stokes, F. Iida, and J. Hughes, "A democratized bimodal model of research for soft robotics: Integrating slow and fast science," *Science Robotics*, vol. 10, no. 99, p. eadr2708, 2025.

[2] A. Billard and D. Kragic, "Trends and challenges in robot manipulation," *Science*, vol. 364, no. 6446, p. eaat8414, 2019.

[3] G. Gu, N. Zhang, H. Xu, S. Lin, Y. Yu, G. Chai, L. Ge, H. Yang, Q. Shao, X. Sheng *et al.*, "A soft neuroprosthetic hand providing simultaneous myoelectric control and tactile feedback," *Nature biomedical engineering*, vol. 7, no. 4, pp. 589–598, 2023.

[4] D. Rus and M. T. Tolley, "Design, fabrication and control of soft robots," *Nature*, vol. 521, no. 7553, pp. 467–475, 2015.

[5] M. T. Mason, "Toward robotic manipulation," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 1, no. 1, pp. 1–28, 2018.

[6] S. Abondance, C. B. Teeple, and R. J. Wood, "A dexterous soft robotic hand for delicate in-hand manipulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5502–5509, 2020.

[7] X. Yang, N. Zhang, X. Huang, R. Bian, M. Feng, X. Zhu, and G. Gu, "Multidirectional bending soft pneumatic actuator with fishbone-like strain-limiting layer for dexterous manipulation," *IEEE Robotics and Automation Letters*, vol. 9, no. 4, pp. 3815–3822, 2024.

[8] J. Yoon, J. Yang, and D. Yun, "A two-chamber soft actuator with an expansion limit line for force enhancement," *IEEE Robotics and*

Automation Letters, vol. 9, no. 5, pp. 4567–4574, 2024.
[9] B. Mosadegh, P. Polygerinos, C. Keplinger, S. Wennstedt, R. F. Shepherd, U. Gupta, J. Shim, K. Bertoldi, C. J. Walsh, and G. M. Whitesides, "Pneumatic networks for soft robotics that actuate rapidly," *Advanced functional materials*, vol. 24, no. 15, pp. 2163–2170, 2014.
[10] O. Yasa, Y. Toshimitsu, M. Y. Michelis, L. S. Jones, M. Filippi, T. Buchner, and R. K. Katzschmann, "An overview of soft robotics," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 6, no. 1, pp. 1–29, 2023.
[11] Z. Zhang, A. Petit, J. Dequidt, and C. Duriez, "Calibration and external force sensing for soft robots using an rgb-d camera," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2356–2363, 2019.
[12] H. Bezawada, C. Woods, and V. Vikas, "Shape reconstruction of soft manipulators using vision and imu feedback," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9589–9596, 2022.
[13] T. G. Thuruthel, B. Shih, C. Laschi, and M. T. Tolley, "Soft robot perception using embedded soft sensors and recurrent neural networks," *Science Robotics*, vol. 4, no. 26, p. eaav1488, 2019.
[14] Z. Zhou, R. Zuo, B. Ying, J. Zhu, Y. Wang, X. Wang, and X. Liu, "A sensory soft robotic gripper capable of learning-based object recognition and force-controlled grasping," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 1, pp. 844–854, 2022.
[15] F. Xiao, Z. Wei, Z. Xu, H. Wang, J. Li, and J. Zhu, "Fully 3d-printed soft capacitive sensor of high toughness and large measurement range," *Advanced Science*, vol. 12, no. 8, p. 2410284, 2025.
[16] Z. Chen, F. Renda, A. Le Gall, L. Mocellin, M. Bernabei, T. Dangel, G. Ciuti, M. Cianchetti, and C. Stefanini, "Data-driven methods applied to soft robot modeling and control: A review," *IEEE Transactions on Automation Science and Engineering*, 2024.
[17] K. Chin, T. Hellebrekers, and C. Majidi, "Machine learning for soft robotic sensing and control," *Advanced Intelligent Systems*, vol. 2, no. 6, p. 1900171, 2020.
[18] R. L. Truby, C. Della Santina, and D. Rus, "Distributed proprioception of 3d configuration in soft, sensorized robots via deep learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3299–3306, 2020.
[19] S. Yoo, H. I. Park, J. Kang, and Y. Cha, "Soft origami module with piezoelectric sensors for multi-directional bending," *IEEE Robotics and Automation Letters*, vol. 10, no. 1, pp. 17–24, 2025.
[20] Y. H. Jung, S. K. Hong, H. S. Wang, J. H. Han, T. X. Pham, H. Park, J. Kim, S. Kang, C. D. Yoo, and K. J. Lee, "Flexible piezoelectric acoustic sensors and machine learning for speech processing," *Advanced Materials*, vol. 32, no. 35, p. 1904020, 2020.
[21] H. Zhao, K. O'brien, S. Li, and R. F. Shepherd, "Optoelectronically innervated soft prosthetic hand via stretchable optical waveguides," *Science robotics*, vol. 1, no. 1, p. eaai7529, 2016.
[22] J. L. Molnar, C.-A. Cheng, L. O. Tiziani, B. Boots, and F. L. Hammond, "Optical sensing and control methods for soft pneumatically actuated robotic manipulators," in *2018 IEEE international conference on robotics and automation (ICRA)*. IEEE, 2018, pp. 3355–3362.
[23] F. Stella, C. Della Santina, and J. Hughes, "Soft robot shape estimation with imus leveraging pcc kinematics for drift filtering," *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1945–1952, 2023.
[24] Y. Meng, G. Fang, J. Yang, Y. Guo, and C. C. L. Wang, "Spring-IMU fusion-based proprioception for feedback control of soft manipulators," *IEEE/ASME Transactions on Mechatronics*, vol. 29, no. 2, pp. 832–842, 2024.
[25] F. Connolly, C. J. Walsh, and K. Bertoldi, "Automatic design of fiber-reinforced soft actuators for trajectory matching," *Proceedings of the National Academy of Sciences*, vol. 114, no. 1, pp. 51–56, 2017.
[26] G. Singh and G. Krishnan, "Designing fiber-reinforced soft actuators for planar curvilinear shape matching," *Soft robotics*, vol. 7, no. 1, pp. 109–121, 2020.
[27] D. Drotman, M. Ishida, S. Jadhav, and M. T. Tolley, "Application-driven design of soft, 3-d printed, pneumatic actuators with bellows," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 1, pp. 78–87, 2018.
[28] S. Y. Kim, R. Baines, J. Booth, N. Vasios, K. Bertoldi, and R. Kramer-Bottiglio, "Reconfigurable soft body trajectories using unidirectionally stretchable composite laminae," *Nature communications*, vol. 10, no. 1, p. 3464, 2019.
[29] P. Jiang, J. Luo, J. Li, M. Z. Chen, Y. Chen, Y. Yang, and R. Chen, "A novel scaffold-reinforced actuator with tunable attitude ability for grasping," *IEEE Transactions on Robotics*, vol. 39, no. 2, pp. 1164–1177, 2022.
[30] B. Guo, S. Duan, P. Wang, H. Lei, Z. Zhao, and D. Fang, "A review on the deformation tracking methods in vision-based tactile sensing technology," *Acta Mechanica Sinica*, vol. 41, no. 10, p. 424436, 2025.
[31] A. Rosinol, J. J. Leonard, and L. Carlone, "Nerf-slam: Real-time dense monocular slam with neural radiance fields," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 3437–3444.
[32] T. Bhattacharjee, G. Lee, H. Song, and S. S. Srinivasa, "Towards robotic feeding: Role of haptics in fork-based food manipulation," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1485–1492, 2019.
[33] P. Sundaresan, S. Belkhale, and D. Sadigh, "Learning visuo-haptic skewering strategies for robot-assisted feeding," in *6th Annual Conference on Robot Learning*, 2022.
[34] J. Chen, X. Wei, X. Liang, H. Xu, L. Zhou, W. He, Y. Ma, and Y. Yin, "High precision 3d reconstruction and target location based on the fusion of visual features and point cloud registration," *Measurement*, vol. 243, p. 116455, 2025.
[35] U. Yoo, H. Zhao, A. Altamirano, W. Yuan, and C. Feng, "Toward zero-shot sim-to-real transfer learning for pneumatic soft robot 3d proprioceptive sensing," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 544–551.
[36] H. Wang, M. Totaro, and L. Beccai, "Toward perceptive soft robots: Progress and challenges," *Advanced Science*, vol. 5, no. 9, p. 1800541, 2018.