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/IROS60139.2025.11247770

AirTouch: A Low-Cost Versatile Visuotactile Feedback System for Enhanced Robotic Teleoperation

Shoujie Li[1*], Xingting Li[2*], Yan Huang[1*], Ken Jiankun Zheng[3], Ran Yu[1], Xueqian Wang[1], Wenbo Ding[1†]

***Abstract*— Vision-based teleoperation systems are widely used due to their cost-effectiveness and intuitive operation. However, these systems often suffer from challenges such as hand occlusions, environmental variability, and the lack of tactile feedback, limiting their precision and applicability in complex tasks. To address these limitations, we present AirTouch, a novel, low-cost visuotactile teleoperation system that integrates air pressure-based tactile feedback with lightweight hand pose estimation. AirTouch features an inflatable tactile bubble that provides adjustable feedback through closed-loop pneumatic control, enhancing the operator's sense of interaction with remote environments. The system's robust hand-tracking algorithm ensures accurate control even under dynamic and occlusion-prone conditions, while its hardware design eliminates the need for wearable devices, enabling intuitive operation. AirTouch supports a wide range of robotic end-effectors, including dexterous hands, parallel grippers, and suction cups, demonstrating versatility across multiple platforms. Extensive experiments validate AirTouch's performance, achieving high precision in hand pose estimation and a 91% success rate in complex teleoperation tasks, all with a hardware cost as low as $39. These results highlight AirTouch as a scalable and practical solution for enhancing robotic teleoperation across industrial, medical, and hazardous scenarios.**

I. INTRODUCTION

Teleoperation systems are pivotal for enabling humans to remotely control robots in diverse applications, ranging from industrial automation to medical surgery and operations in hazardous environments. Among these, vision-based teleoperation has gained significant traction due to its cost-effectiveness, intuitive operation, and ease of deployment [1], [2]. By leveraging minimal hardware, such systems allow operators to control robotic end-effectors with high scalability, making them suitable for real-world applications [3].

Fig. 1. Overview of AirTouch system, which integrates a flexible inflatable bubble, a camera, and 3D-printed components. The three-mode platform enables versatile teleoperation across various robotic end effectors.

Despite their advantages, vision-based teleoperation systems face critical limitations. Hand occlusions, environmental variability, and the lack of tactile feedback hinder their precision and applicability in complex tasks [4], [5]. Existing solutions incorporating haptic feedback, such as commercial gloves or bilateral control systems, often involve high costs, bulky hardware, and extensive calibration requirements, limiting their accessibility and scalability [6], [7].

To address these challenges, we propose **AirTouch**, a novel, low-cost ($39) visuotactile teleoperation system that combines air pressure-based tactile feedback with lightweight hand pose estimation, as shown in Fig. 1. AirTouch introduces an inflatable bubble as its core component, which provides adjustable tactile feedback through closed-loop pneumatic control. Unlike conventional controllers that rely on wearable devices or complex hardware setups, AirTouch features an intuitive and portable design that supports a wide range of robotic end-effectors, including dexterous hands, parallel grippers, suction cups, and jamming grippers.

The key contributions of this work are as follows:

- **Visuotactile feedback mechanism**: AirTouch employs a pneumatic feedback system to provide operators with a sense of contact and interaction with remote environments, enhancing task precision and reducing object drop rates.
- **Lightweight hand pose estimation algorithm**: The system combines RGB and depth-based sensing with advanced feature fusion to ensure robust performance even in dynamic and occlusion-prone environments.

∗These authors contributed equally to this work.

[1]Shenzhen Ubiquitous Data Enabling Key Lab, Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China.
[2]School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China.
[3]UC Berkeley Electrical Engineering and Computer Sciences (EECS), University of California, Berkeley, Berkeley, CA 94720, United States.

This work was supported by National Key R&D Program of China (No.2024YFB3816000), Shenzhen Key Laboratory of Ubiquitous Data Enabling (No. ZDSYS20220527171406015), Guangdong Innovative and Entrepreneurial Research Team Program (2021ZT09L197), Meituan Academy of Robotics Shenzhen, and Tsinghua Shenzhen International Graduate School-Shenzhen Pengrui Young Faculty Program of Shenzhen Pengrui Foundation (No. SZPR2023005).

This paper has supplementary material available at https://sites.google.com/view/airtouch-bubble.

†Corresponding author: ding.wenbo@sz.tsinghua.edu.cn.

- **Versatile operational modes**: AirTouch supports three distinct configurations, the handheld controller, the ALOHA end effector, and the desktop controller, demonstrating flexibility on various robotic platforms.
- **Comprehensive experimental validation**: Extensive experiments demonstrate AirTouch's ability to achieve an average keypoint error of 0.001949m in hand pose estimation and an 91% task success rate across various robotic end-effectors.

With its modular design, low cost, and high adaptability, AirTouch represents a scalable and practical solution for enhancing robotic teleoperation in industrial, medical, and hazardous environments.

II. RELATED WORK

Teleoperation systems have a well-established history, with a wide range of strategies deployed to facilitate human-robot interaction, as shown in Table. I. Common input devices for these systems include hand Motion Capture (MoCap) setups [8], [9], [10], [11], spacemice [12], Virtual Reality (VR) systems [1], [13], [14], [15], [16], and others [17], [18].

A. Vision-based Teleoperation

Some vision-based robotic teleoperation systems offer cost-effective, adaptable alternatives to solutions that feature sophisticated devices like VR, haptic gloves, and specialized cameras [1], [6], [19]. These often incur steep costs, limiting their accessibility and restricting efforts at large-scale collection of data. On top of requiring minimal hardware [2], [3] vision-based solutions have also benefited from recent advances that effectively map human hand movements to anthropomorphic and/or parallel-jaw grippers via hand-pose prediction algorithms [4], [20], [21], [22], enabling them to be user-friendly and accommodate a variety of body types [23]. Despite these advantages, notable shortcomings persist. Some solutions preserve vision-based convenience but bring in additional bulk to the operator [10]. Another issue is occlusion, which occurs when parts of the hand or arm become hidden from the camera's view due to obstruction by other objects or the angle of observation. This leads to incomplete or corrupted data, making it difficult for the system to accurately estimate hand pose. Furthermore, a combination of factors such as ambient noise, sensor inaccuracies, and limitations in the resolution of the vision device can have non-negligible impacts on overall effectiveness, further complicating vision-based systems.

B. Consumer-grade Controllers

There is active research on teleoperation controllers that do not map joint-by-joint to end-effectors. One main advantage of controllers such as the 3D mouse controller [12], which utilizes a SpaceMouse Compact to telemanipulate TIAGo, is that they are often more affordable and reproducible than other specialized systems that require custom 3D-printed components [24]. A notable challenge remains in the requirement to kinematically retarget robot motion to accommodate diverse form factors of different control interfaces. Solutions from a software direction that attempt to make such abstract mappings more precise have been proposed [25]. However, operators often still lack appropriate awareness of the robot's physical limitations. This hinders precise and dexterous control, adversely impacting task performance. In contrast to these more specific and intricate systems, our solution maintains the advantages of a low-cost accessible system while supporting dexterous control through a lightweight hand pose estimation algorithm combined with visuotactile sensing to give the operator transparency about the robot's state and physical singularities. AirTouch provides an accessible framework for designing portable, user-friendly controllers using affordable materials, while also delivering tactile feedback and ensuring generalizability across different end-effectors, all without over-complicating the system.

TABLE I
COMPARISON OF COMMERCIAL TELEOPERATION DEVICES

Device	Cost	Force-feedback	Weight
Leap Controller 2 [19]	∼$250	N/A	29 g
Apple Vision Pro [26]	$3499	N/A	625 g
Prime 3 Mocap [27]	$3,799	N/A	138 g
SenseGlove Nova 2 [28]	∼$6,300	Yes	350 g
HaptX Gloves G1 [7]	∼$5,495	Yes	570 g
AirTouch (Ours)	∼$39	Yes	100 g

III. HARDWARE SYSTEM

In this section, we provide a comprehensive overview of the system design, which consists of three main components: (1) **AirTouch**, a versatile hardware system for human hand motion capture used to control the end-effector (Section III-A); (2) **Closed-loop pneumatic control system**, which provides haptic feedback through closed-loop pressure regulation (Section III-B). Notably, the AirTouch system supports multiple operational modes, offering flexibility for customization based on specific application requirements; (3) **Customizable control interface**: In Section III-C, we illustrate the integration of the AirTouch with the T265 camera to form a handheld controller, as well as its use as the end-effector controller within the ALOHA manipulation system. Additionally, the AirTouch is paired with a 2-DOF platform to function as a desktop controller for robotic arm movement.

A. AirTouch Bubble

The AirTouch system consists of four main components: an LED light ring, a flexible spherical skin, PLA-printed parts, and an RGB camera, all made from commonly available materials. The LED light ring features a 5V COB strip that provides uniform illumination without hotspots or dark spots, ensuring consistent lighting for the camera while minimizing image quality variations. The flexible spherical skin, made from thermoplastic rubber (TPR), offers elasticity, wear resistance, softness, and comfort. TPR, commonly used in toys and medical products, is ideal for this application. As shown in Fig. 2(a)(b), the TPR skin is positioned at the top of the device and securely attached to the 3D-printed

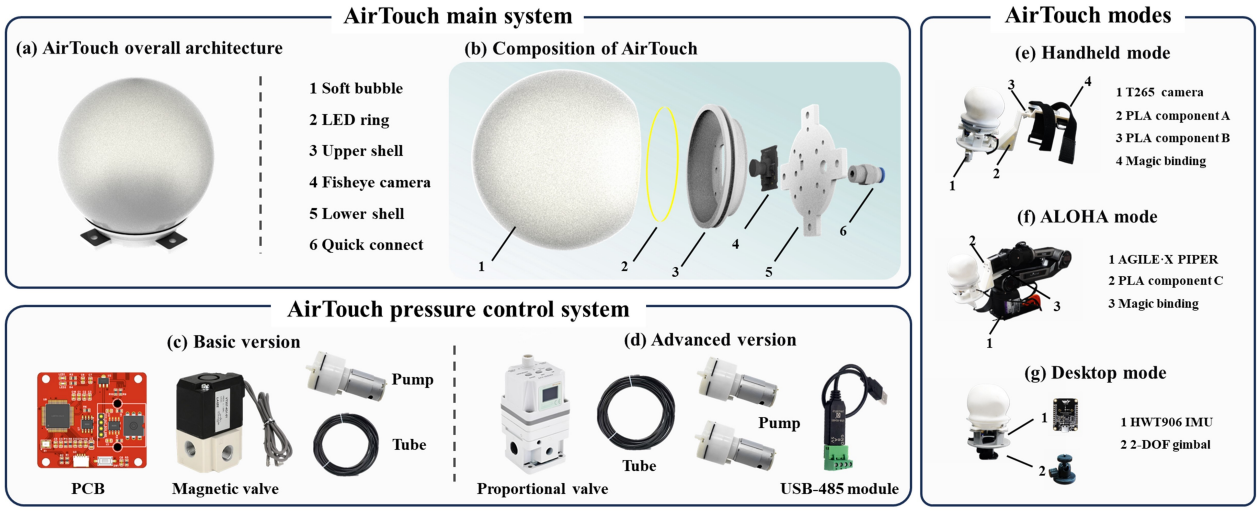

Fig. 2. Overview of the AirTouch hardware system. (a) AirTouch overall system frameworkp; (b) Components of AirTouch; (c) Basic version of the pneumatic closed-loop controller; (d) Advanced version of the pneumatic closed-loop controller; (e) Air pump system; (f) Handheld mode; (g) ALOHA mode; (h) Desktop mode.

components. The LED strip is embedded in the printed parts and evenly distributed across the annular surface. At the bottom, PLA-printed parts hold the camera in place and house air vents connected to an external pneumatic control system. The entire AirTouch Bubble structure is sealed with silicone, except for the air vents, ensuring an airtight enclosure.

For the camera, we select a single-lens 210° fisheye camera from JIERUIWEITONG, with a resolution of 640×480 and a frame rate of 30 frames per second. The wide field of view provided by the fisheye lens offers a significant advantage for close-range hand gesture capture. Despite some degree of image distortion, experimental results demonstrate that the camera performs exceptionally well in feature recognition, making it highly suitable for the specific requirements of this scenario.

During operation, the user places their hand on the surface of the ball skin. The internal RGB camera captures the concurrent changes in the surface and the hand. This data is then processed through a neural network, which maps the observed movements to control the target end-effector.

B. Closed-loop Pneumatic Control System

As shown in Fig. 2(c)(d), the pneumatic closed-loop control system consists of two primary components: the air pump and the pressure controller. The air pump is a 12V DC motorized pump with a flow rate of 25 L/minute and a vacuum level of -60 kPa. Its flow rate is adjustable to accommodate varying device scales across different application scenarios.

The pressure controller is available in two versions: a basic version (Fig. 2(c)) and an advanced version (Fig. 2(d)). The basic version is composed of a pressure detection module and a custom PCB, offering a low-cost solution priced at just $39, capable of providing basic pressure control. The advanced version, in contrast, utilizes the EPV1 industrial pneumatic proportional valve from XINGYU-ELECTRON, which offers higher control precision with a minimum adjustable range of 0.5 kPa.

TABLE II
SYSTEM BOM TABLE

Basic version		Advanced version	
Item	**Price**	**Item**	**Price**
Soft bubble (TPR)	$1	Soft bubble (TPR)	$1
Magnetic valve	$8	Proportional valve	$64
Pump	$6	Pump ×2	$12
Camera	$18	Camera	$18
PCB	$5	USB-485 module	$2
Tube	$1	Tube	$1
Total	**$39**	**Total**	**$98**

The pneumatic proportional valve regulates the supply and exhaust solenoid valves based on input signals, enabling precise control of output pressure. As the input signal increases, the control loop activates the supply valve and deactivates the exhaust valve, allowing supply pressure to enter the pilot chamber and apply pressure to the diaphragm. This upward movement of the diaphragm actuates the pilot valve, adjusting the output pressure. The output pressure is continuously monitored by a pressure sensor, and the feedback is used by the control loop to dynamically adjust the solenoid valve states, maintaining proportionality between input and output pressures. This closed-loop control ensures stable, precise, and linear pressure regulation.

Specifically, we establish a linear mapping between the tactile sensor on the robotic end effector and the air pressure. Through this real-time mapping, any changes in the tactile information of the robotic hand are reflected in the form of air pressure changes, which are then fed back to the AirTouch Bubble, allowing the operator to perceive the contact state of the robotic end effector.

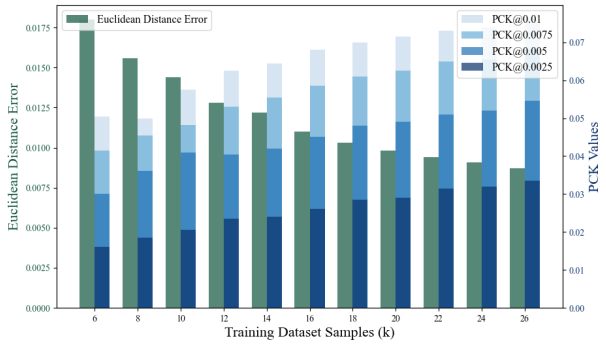

Fig. 3. Varying data scales test. We report the Euclidean distance error, PCK@0.01, PCK@0.0075, PCK@0.005, and PCK@0.0025 for hand pose estimation as the dataset size increases.

C. Customizable Control Interface

The AirTouch operates in three distinct modes: handheld controller mode, ALOHA end-effector mode, and desktop controller mode. In handheld controller mode (Fig. 2(e)), the system uses a T265 camera and 3D-printed components to secure the user's hand, establishing a mapping between the user's hand and the target end-effector. The T265 camera captures pose data, enabling precise control of the robotic arm via redirection.

In ALOHA mode (Fig. 2(f)), AirTouch is mapped to the end-effector, utilizing a master-slave arm configuration to control the robotic arm for task execution.

In desktop controller mode (Fig. 2(g)), the AirTouch Bubble is mounted on a 2-DOF gimbal for spatial movement of the robotic arm. The internal camera continuously monitors the surface of the sphere and the user's hand, detecting the geometric center of the hand's contour. Displacement of this center controls horizontal movement, while the hand's size and rotational orientation control vertical movement and yaw rotation. Pitch and roll are managed by the 2-DOF gimbal, providing flexible, multi-axis control for enhanced versatility.

Finally, the cost of our system is detailed in Table II. All components are low-cost and readily accessible. Our hardware system represents the first **low-cost, universal** teleoperation controller with **tactile feedback** capabilities.

IV. METHOD

A. Dataset Generation

To achieve remote manipulation of a dexterous hand based on the operator's hand posture, it is essential to reflect the operator's hand pose. In this work, we construct our hand pose dataset using two distinct approaches: (a) The data generation scheme proposed by HaMeR et al. [29], which leverages deep learning techniques to generate high-precision hand posture estimates; (b) We use UDEXREAL glove to capture the operator's hand posture in real-time, providing precise measurements of finger and palm movements. While constructing the hand pose dataset using HaMeR, we opt for the commonly used skeletal model with 21 key points, denoted as $X_{\text{hand}} \in \mathbb{R}^{21\times 3}$, to represent the hand posture [29], [30]. For the UDEXREAL, which captures joint angles $R_{\text{hand}} \in \mathbb{R}^{12\times 1}$, we achieve remote manipulation of the dexterous hand by mapping these joint angles to the controlled dexterous hand.

To reduce data redundancy, we collected 29,000 samples each using both left and right hands with the UDEXREAL glove. We then tested the effect of varying the dataset size on model convergence, with intervals of 2,000 samples, using datasets ranging from 6,000 to 28,000 samples. As shown in the Fig. 3, we calculated the Euclidean distance error between the predicted results and labels, as well as the PCK (Percentage of Correct Keypoints) metrics at different thresholds, using a fixed test set and varying training dataset sizes. The PCK is computed using

$$\text{PCK} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left(\|\hat{y}_i - y_i\| \leq \alpha\right), \tag{1}$$

where $\hat{y}_i$ is the predicted keypoint, y_i is the ground truth keypoint, and α is the threshold (e.g., 0.01, 0.0075, 0.005, 0.0025). We found that as the size of the training data set increased, the accuracy of the predictions gradually improved. To strike a balance between data collection efficiency and prediction accuracy, we selected 12,000 samples as the standard dataset size for our subsequent comparative experiments.

B. Control Algorithm

Inspired by the work of Liu et al. [31], we employ ResNet as the backbone for feature extraction. Initially, we identify the centroid of the circular valid region within the camera's field of view and crop the image accordingly. As illustrated in Fig. 4, we perform multi-scale fusion of the features extracted from the RGB and depth channels at different scales. The ResNet backbone outputs a feature map and an initial hand pose for both RGB and depth images. Subsequently, we fuse the RGB and depth features and provide a hand pose output.

The proposed network is designed for hand pose estimation using both RGB and Depth (D) images, with fusion in the RGBD mode. As shown in Fig. 4, while extracting features from RGB and depth information through the Backbone, we concurrently output the predicted hand pose. Therefore, we can estimate the hand pose using three different methods. Given an input image, the network processes the image through a backbone network to extract the hand pose and feature map.

The fusion process in the RGBD mode combines the feature maps extracted from the RGB and Depth images. Specifically, the input feature maps $F_{\text{rgb}} \in \mathbb{R}^{B\times C_{\text{rgb}}\times H\times W}$ and $F_{\text{d}} \in \mathbb{R}^{B\times C_{\text{d}}\times H\times W}$ are fused by concatenating them along the channel dimension.

We conduct comparative experiments to evaluate the performance of multiple AirTouch setups under varying conditions. The experimental setups are designed to systematically assess the impact of camera type (RGB vs. RGBD), membrane transparency (semi-transparent vs. transparent),

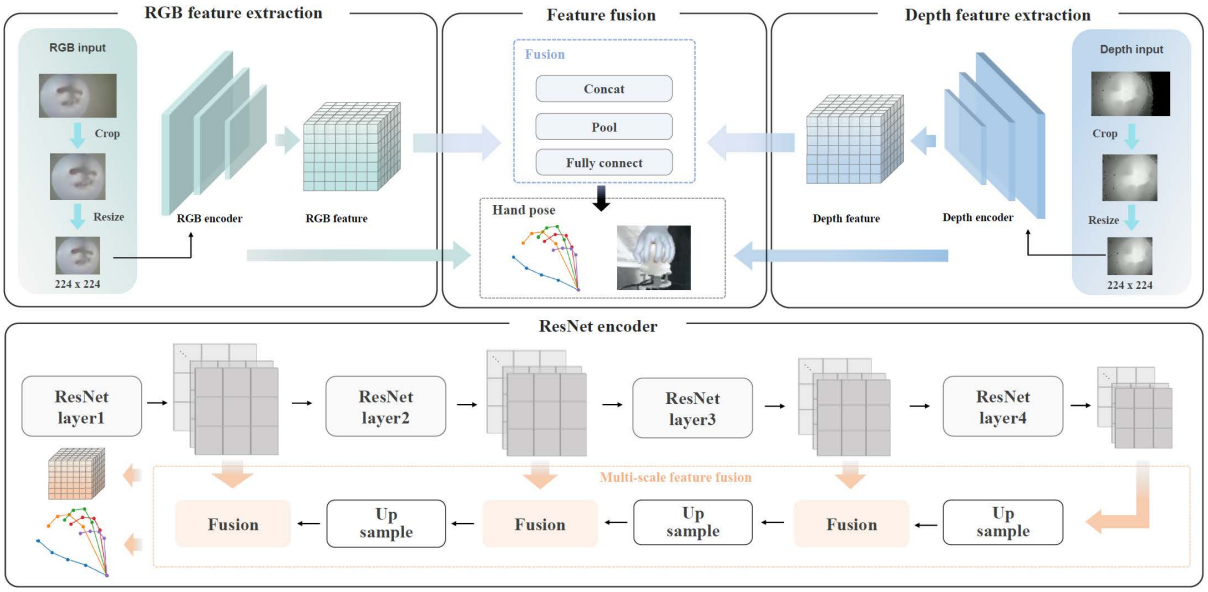

Fig. 4. Hand pose estimation workflow. RGB and depth images output hand pose and feature maps using separate ResNet Encoders. Our workflow fuses the multi-scale features extracted by ResNet to output hand pose and feature maps. The network can estimate hand pose using either RGB or depth images alone, or by fusing both features.

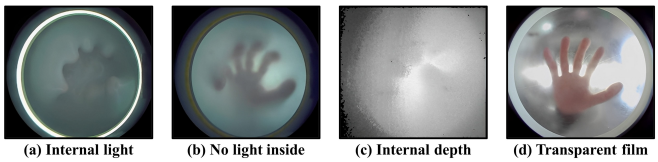

Fig. 5. Image comparison under different membrane and lighting conditions. (a) Translucent membrane, with light; (b) Translucent membrane, without light; (c) Translucent membrane, depth mode; (d) Transparent membrane, without light.

and lighting conditions. Specifically, the four configurations tested were as follows: (a) Bubble without lighting, semi-transparent membrane, RGB camera; (b) Bubble without lighting, semi-transparent membrane, RGBD camera; (c) Bubble with lighting, semi-transparent membrane, RGB camera; (d) Bubble without lighting, transparent membrane, RGB camera. The example images under different settings are shown in Fig. 5, for each of these AirTouch Bubble setups, we collect 12,000 images for training and test the mean error of each keypoint on the corresponding test set.

The error we calculated is the average Euclidean distance between the predicted 3D joint positions and the ground truth positions for each keypoint. The error for each joint is computed using

$$e_j = \sqrt{\sum_{i=1}^{N_d} (p_{j,i} - t_{j,i})^2}, \quad j = 1, 2, \ldots, N_j, \tag{2}$$

where p_j and t_j represent the predicted and true coordinates (or angles) for the j-th joint, respectively. Both p_j and t_j are N_d-dimensional vectors, where N_d denotes the number of dimensions (e.g., 1 for the bending angle of each joint).

In the case where we are predicting the bending angles of each joint collected by UDEXREAL, N_j and N_d would be 12 and 1, respectively. $N_j = 12$ corresponds to the total number of joints, and $N_d = 1$ represents the single bending angle for each joint. The error for each joint is then computed based on these dimensions.

Based on the experimental results obtained from the different setups, we observed significant variations in performance across the configurations, reflecting the effects of camera type (RGB vs. RGBD), membrane transparency (semi-transparent vs. transparent), and lighting conditions. The setup utilizing an RGB camera with a semi-transparent bubble and no additional lighting (Fig. 5(a)) exhibited the lowest mean error of 0.001949m, indicating superior accuracy under these conditions. In contrast, the setup incorporating an RGBD camera with a semi-transparent bubble and no additional lighting (Fig. 5(b)) resulted in a higher mean error of 0.005777m. This increase in error can be attributed to the added complexity of depth data, which may introduce noise during keypoint detection.

Additionally, the setup with supplementary lighting (Fig. 5(c)), featuring an RGB camera and a semi-transparent bubble, produced a mean error of 0.004386m. Although lighting improved visibility, it also introduced potential artifacts or reflections, which may negatively impact accuracy.

The configuration with a transparent bubble and no lighting (Fig. 5(d)) achieved a mean error of 0.002743m. While this setup theoretically outperformed the RGBD and RGBD + lighting configurations, it still exhibited lower accuracy compared to the semi-transparent bubble setup. This can be attributed to the increased susceptibility of transparent membranes to external environmental interference, which affects performance under practical conditions.

C. Direction Control

In this work, we propose three methods for controlling the direction of a robotic arm based on the setup of the bubble mechanism:

(a) Handheld mode: A T265 camera is mounted beneath the AirTouch to capture the pose of the hand relative to its initial position. The captured pose is then synchronized with the robotic arm to control its direction.

(b) Aloha mode: The AirTouch is mounted on the active robotic arm, and joint mapping from the active arm is used to control the movement of the passive robotic arm in space.

(c) Desktop mode: In this setup, the robotic arm moves only within the plane. An IMU is installed beneath the AirTouch, and its X and Y axis angle information is used to control the robotic arm's movement. Specifically, when the angle deviation exceeds a set threshold, the robotic arm moves in the corresponding direction at a constant speed.

V. EXPERIMENTS

We designed a series of experiments to systematically evaluate the performance of the proposed teleoperation system in addressing the following key research questions:

- Can the system achieve high control precision in dexterous manipulation tasks and provide effective tactile feedback?
- Can the system maintain stable and efficient teleoperation performance across three different mounting configurations?
- Is the system adaptable to different types of end-effectors, enabling generalized control?

To assess control precision, we designed a finger-level manipulation task for two dexterous hands, demonstrating that despite the integration of the Soft Bubble layer, the system is still capable of achieving high-precision end-effector control. Additionally, for devices equipped with tactile sensors, we implemented a dexterous hand grasping experiment to validate that the teleoperation system can provide users with tactile feedback, thereby enhancing the perception of physical contact during operation.

To further evaluate the generalizability of the system, we conducted extensive experiments involving a variety of commonly used end-effectors, verifying its adaptability across different robotic platforms. Finally, we demonstrated the teleoperation capabilities of the AirTouch Bubble under various mounting configurations. Experimental results indicate that the system consistently enables precise end-effector control and successfully completes predefined tasks, further confirming its robustness and broad applicability.

A. Control Precision and Tactile Feedback

1) ***Control Precision****:* To evaluate the control accuracy of the AirTouch system in dexterous manipulation tasks, we conducted experiments using the five-finger dexterous hand, selecting both the Xhand and the Inspire dexterous hand for testing.

During the experiments, we designed individual finger control tasks to assess the AirTouch system's ability to independently control each finger of the dexterous hand. As shown in Fig. 6, for the Xhand, the system was able to precisely distinguish and execute independent control commands for all five fingers. Notably, the AirTouch system could also accurately control the sideward movement of the Xhand's index finger. By processing and converting the natural hand movements of the operator, the AirTouch system enabled stable and precise sideward motions of the Xhand's index finger, providing strong support for tasks that require specific finger postures. The Inspire dexterous hand also performed exceptionally well in the experiment. Whether independently flexing and extending each finger, or executing coordinated grasping motions with multiple fingers, the AirTouch system ensured that the Inspire dexterous hand precisely followed the operator's hand motion intentions.

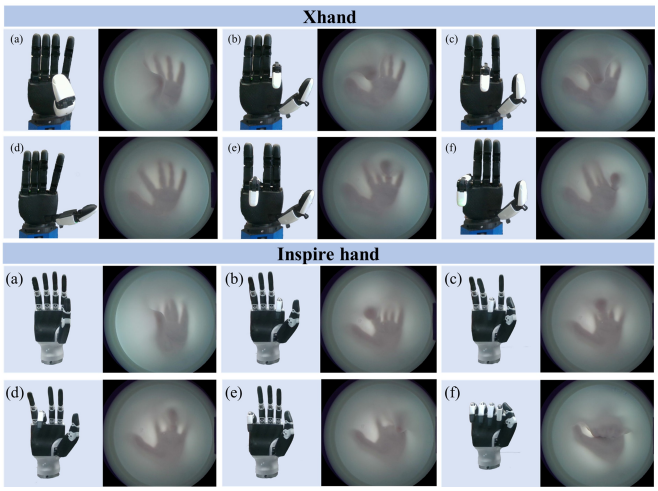

Fig. 6. Precise finger control. It demonstrates the independent control of each finger on a five-finger dexterous hand, as well as the performance of full-hand synchronized control, and is applicable to various types of dexterous hands.

These experimental results clearly demonstrate that the AirTouch system can effectively achieve both independent control of each finger and coordinated grasping control across different types of five-finger dexterous hands, showcasing promising applications in dexterous manipulation tasks.

2) ***Tactile Feedback****:* As shown in Fig. 7, we evaluated the effectiveness of the AirTouch system's pressure-based tactile feedback through experiments using bitter melon, lemon, and mango as test objects for a dexterous hand grasping task. The Xhand, equipped with tactile sensors on all five fingertips, was used. By averaging the fingertip pressures and mapping them to air pressure, we quantified the tactile feedback. The tactile-pressure curve revealed key system characteristics: as the hand made contact with the object, fingertip pressure increased (pink curve), while the air pressure command also rose and stabilized (blue curve). This demonstrates that the sensors capture the average fingertip pressure, which is translated into air pressure adjustments within the AirTouch bubble, providing real-time tactile feedback. When the hand released the object, both fingertip and air pressure decreased,

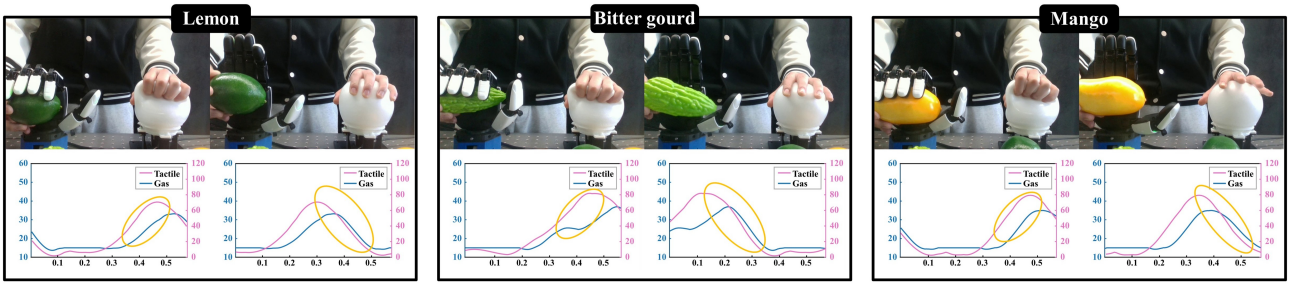

Fig. 7. Tactile closed-loop feedback. A linear mapping is established between the haptic information and the internal air pressure of the sphere. When the actuator applies gripping force, the air pressure at the control end increases, and the operator feels the contact feedback. When the actuator releases, the air pressure decreases, and the operator feels the loss of contact.

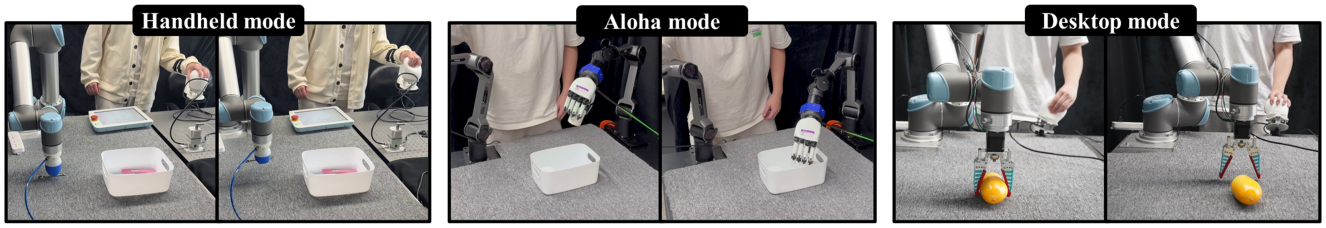

Fig. 8. Cross-platform operation examples. The AirTouch system is integrated into the handheld, ALOHA, and desktop platforms, enabling the control of robotic arms and end-effectors to perform various tasks.

signaling the loss of tactile contact. Furthermore, comparing operator performance with and without tactile feedback showed that, with tactile feedback, the likelihood of object drops during remote operation was significantly reduced. This highlights the AirTouch system's ability to enhance control and precision, improving task performance.

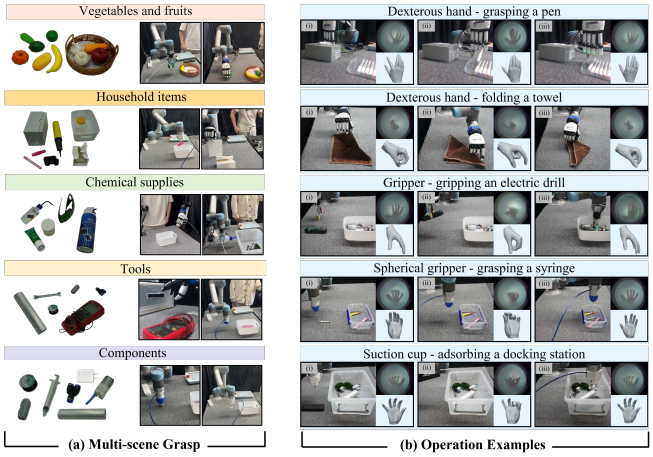

Fig. 9. General operation examples. (a) shows the target test objects, categorized into vegetables and fruits, household items, chemicals, tools and components; (b) illustrates the operational details of some specific tasks.

B. Performance in Different Configurations

To investigate whether the AirTouch system can maintain stable and efficient remote operation performance across different mounting configurations, we conducted experiments in three modes: handheld mode, ALOHA mode, and desktop mode, as shown in Fig. 8. In all three modes, the AirTouch system is responsible for gesture recognition and control of the end-effector, while the remaining components handle the spatial movement control of the robotic arm.

Throughout the operation in these three modes, the system demonstrated excellent dynamic responsiveness, effectively adapting to various changes during operation and ensuring the successful completion of tasks. This underscores the AirTouch system's remarkable ability to adapt to diverse operational demands across different modes.

TABLE III
AVERAGE SUCCESS RATE FOR END-EFFECTORS ACROSS TASKS

	Fruits	Household	Chemical	Tools	Components
XHand	96%	95%	98%	92%	89%
Gripper	95%	95%	92%	90%	90%
Jamming	87%	85%	85%	92%	94%
Suction	93%	84%	90%	83%	87%
Avg	93%	90%	91%	89%	90%
Total Avg		91%			

C. Adaptability to Different End-Effectors

To assess whether the AirTouch system can adapt to various end-effectors and provide generalized control, we conducted a series of experiments. These tested the system's flexibility across different operational scenarios with a range of end-effectors, including five-finger dexterous hands, two- to four-finger grippers, jamming grippers, and suction cups. We designed tasks involving vegetables, fruits, household items, chemicals, tools, and components (see Fig. 9(a)), each with two control examples.

Some of the operational details are shown in Fig. 9(b). In the marker grasping task, the operator controlled the

index finger's lateral motion to move the marker using a pinch between the index and middle fingers. In the towel folding task, the operator coordinated the index and thumb to fold the towel. In the two-finger gripper task, the operator manipulated the gripper to grasp an electric screwdriver and place it in a box. In the syringe needle task, the operator used the jamming gripper to grasp and release the needle. In the docking station task, the operator controlled the suction cup to attach and release it from the target surface.

For each object category, five grasping trials were conducted with different end-effectors, and the average success rates are shown in Table. III. A successful grasp-and-drop involves a first-attempt grasp, uninterrupted transport without premature release, and precise, gentle placement at the target location. The suction cup had a lower success rate (83%) when handling tool-type objects, as the lack of smooth surfaces hindered effective suction. Overall, the system achieved a **91%** success rate, demonstrating excellent adaptability and control across different end-effectors. The results confirm that the AirTouch system can perform precise control and efficiently complete tasks with various end-effectors, making it a versatile solution for a wide range of robotic applications.

VI. CONCLUSION

This paper presents AirTouch, a low-cost ($39), versatile teleoperation system with tactile feedback. By integrating a pneumatic feedback mechanism with a lightweight gesture estimation algorithm, AirTouch addresses limitations in existing visual teleoperation systems, such as occlusion, environmental variability, and lack of tactile feedback. Experimental results demonstrate that AirTouch achieves an average keypoint error of 0.001949 in gesture estimation and a 91% average task success rate across various end-effectors (e.g., dexterous hands, parallel grippers, and suction cups), with the dexterous hand achieving a success rate of 96%. Additionally, AirTouch exhibits exceptional dynamic responsiveness and precision across three operational modes: handheld controller, ALOHA end-effector, and desktop controller. With its low cost, modular design, and high performance, AirTouch provides an expandable, economical, and practical teleoperation solution for industrial, medical, and hazardous environments. Future work will focus on optimizing algorithms and exploring applications in more complex tasks.

REFERENCES

[1] S. P. Arunachalam, I. Güzey, S. Chintala, and L. Pinto, "Holo-dex: Teaching dexterity with immersive mixed reality," 2022.

[2] B. Fang, X. Ma, J. Wang, and F. Sun, "Vision-based posture-consistent teleoperation of robotic arm using multi-stage deep neural network," *Robotics and Autonomous Systems*, vol. 131, p. 103592, 2020.

[3] S. Li, X. Ma, H. Liang, M. Görner, P. Ruppel, B. Fang, F. Sun, and J. Zhang, "Vision-based teleoperation of shadow dexterous hand using end-to-end deep neural network," 2019.

[4] Y. Rong, T. Shiratori, and H. Joo, "Frankmocap: Fast monocular 3d hand and body motion capture by regression and integration," 2020.

[5] C. Lenz and S. Behnke, "Bimanual telemanipulation with force and haptic feedback and predictive limit avoidance," 2021.

[6] "Cyberglove systems." `https://www.cyberglovesystems.com`. Accessed: Jan. 18, 2025.

[7] "Haptx." `https://haptx.com`. Accessed: Jan. 18, 2025.

[8] Z. Q. Chen, K. V. Wyk, Y.-W. Chao, W. Yang, A. Mousavian, A. Gupta, and D. Fox, "Dextransfer: Real world multi-fingered dexterous grasping with minimal human demonstrations," 2022.

[9] O. Taheri, N. Ghorbani, M. J. Black, and D. Tzionas, "Grab: A dataset of whole-body human grasping of objects," *European Conference on Computer Vision (ECCV)*, 2020.

[10] C. Wang, H. Shi, W. Wang, R. Zhang, L. Fei-Fei, and C. K. Liu, "Dexcap: Scalable and portable mocap data collection system for dexterous manipulation," 2024.

[11] S. Yang, M. Liu, Y. Qin, R. Ding, J. Li, X. Cheng, R. Yang, S. Yi, and X. Wang, "Ace: A cross-platform visual-exoskeletons system for low-cost dexterous teleoperation," 2024.

[12] A. Calzada-Garcia, B. Łukawski, J. Victores, and C. Balaguer, "Teleoperation of the robot tiago with a 3d mouse controller," 05 2024.

[13] X. Cheng, J. Li, S. Yang, G. Yang, and X. Wang, "Open-television: Teleoperation with immersive active visual feedback," 2024.

[14] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T.-H. Yu, C.-J. Tai, M. Akbay, Z. Wang, A. Nitzan, G. Dong, Y. Ye, L. Tao, C. Wan, and R. Wang, "Megatrack: monochrome egocentric articulated hand-tracking for virtual reality," *ACM Trans. Graph.*, vol. 39, Aug. 2020.

[15] T. Lin, Y. Zhang, Q. Li, H. Qi, B. Yi, S. Levine, and J. Malik, "Learning visuotactile skills with two multifingered hands," 2024.

[16] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel, "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," 2018.

[17] H. Liu, Z. Zhang, X. Xie, Y. Zhu, Y. Liu, Y. Wang, and S.-C. Zhu, "High-fidelity grasping in virtual reality using a glove-based system," in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 5180–5186, 2019.

[18] J. Wong, A. Tung, A. Kurenkov, A. Mandlekar, L. Fei-Fei, S. Savarese, and R. Martín-Martín, "Error-aware imitation learning from teleoperation data for mobile manipulation," 2021.

[19] "Ultraleap." `https://www.ultraleap.com`. Accessed: Jan. 21, 2025.

[20] S. Li, J. Jiang, P. Ruppel, H. Liang, X. Ma, N. Hendrich, F. Sun, and J. Zhang, "A mobile robot hand-arm teleoperation system by vision and imu," 2020.

[21] T. Ohkawa, K. He, F. Sener, T. Hodan, L. Tran, and C. Keskin, "Assemblyhands: Towards egocentric activity understanding via 3d hand pose estimation," 2023.

[22] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," 2020.

[23] Y. Qin, W. Yang, B. Huang, K. V. Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox, "Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system," 2024.

[24] P. Wu, Y. Shentu, Z. Yi, X. Lin, and P. Abbeel, "Gello: A general, low-cost, and intuitive teleoperation framework for robot manipulators," 2024.

[25] V. Dhat, N. Walker, and M. Cakmak, "Using 3d mice to control robot manipulators," in *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '24, (New York, NY, USA), p. 896–900, Association for Computing Machinery, 2024.

[26] "Apple vision pro." `https://www.apple.com/apple-vision-pro/`. Accessed: Jan. 18, 2025.

[27] "Manus." `https://www.manus-meta.com`. Accessed: Jan. 25, 2025.

[28] "Senseglove nova 2." `https://www.senseglove.com/product/nova-2/`. Accessed: Jan 17, 2025.

[29] G. Pavlakos, D. Shan, I. Radosavovic, A. Kanazawa, D. Fouhey, and J. Malik, "Reconstructing hands in 3d with transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9826–9836, 2024.

[30] R. Yu, H. Yu, S. Li, H. Yan, Z. Song, and W. Ding, "Depth restoration of hand-held transparent objects for human-to-robot handover," 2024.

[31] X. Liu, P. Ren, Y. Gao, J. Wang, H. Sun, Q. Qi, Z. Zhuang, and J. Liao, "Keypoint fusion for rgb-d based 3d hand pose estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, pp. 3756–3764, 2024.